



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

DISEÑO E IMPLEMENTACIÓN DE UN PANEL DE INDICADORES COMO
HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS PARA EL APOYO EN LA
GESTIÓN DENTRO DE IBM CHILE

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

FELIPE ALBERTO BAHAMONDE MUÑOZ

PROFESOR GUÍA:
ANDRÉS CABA RUTTE

MIEMBROS DE LA COMISIÓN:
ALFREDO SCHNELL DRESEL
MARCELO MATUS ACUÑA

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE: INGENIERO CIVIL ELÉCTRICO
POR: FELIPE ALBERTO BAHAMONDE MUÑOZ
FECHA: 2019
PROFESOR GUÍA: ANDRÉS CABA RUTTE

DISEÑO E IMPLEMENTACIÓN DE UN PANEL DE INDICADORES COMO HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS PARA EL APOYO EN LA GESTIÓN DENTRO DE IBM CHILE

El presente trabajo tiene como principal objetivo el diseño e implementación de una plataforma de *Business Intelligence* (BI) para apoyar la toma de decisiones por parte de los ejecutivos o *Delivery Project Executives* (DPE), quienes están a cargo de administrar los servicios de los distintos clientes dentro de GTS - IBM Chile. Para esto, se recopilan distintas fuentes de información (reportes generados de manera semanal) con información relacionada a costos y recursos utilizados, que luego son procesados y modelados para desplegar información útil a través de vistas tipo panel de control en la web.

Se realizan 3 versiones de la plataforma, en donde luego de cada lanzamiento se genera una ronda de comentarios o retroalimentación, los cuales son recogidos para validar hipótesis e iterar nuevamente sobre los problemas encontrados. En la primera versión o *Minimum Product Value* (MVP) se realiza una prueba de concepto en donde se interioriza el proceso que realiza cada DPE al momento de extraer y utilizar la data. Para la segunda versión (1.0) se incorpora la automatización del proceso mediante *scripts* en Python, junto con la integración de la nube (IBM Cloud). Finalmente, en la versión definitiva (2.0), se incorporan modelos predictivos con redes neuronales recurrentes LSTM que permiten predecir los costos y FTE (*Full-Time Equivalent*) cuatro semanas a futuro con un *Mean Absolute Percentage Error* (MAPE) promedio de 24.76 % y 19.07 % respectivamente para 4 DPE.

Para medir tiempo reducido con esta nueva plataforma, se realizan encuestas iniciales en donde los DPE ingresan el tiempo que cada uno de ellos estima que utiliza para manipular y analizar los datos. Luego, estos tiempos son comparados con la duración que tiene el proceso de actualización de la plataforma junto con encuestas finales a los usuarios en donde exponen la optimización en tiempo percibida.

Por otro lado, para cuantificar el impacto de la plataforma, se decide medir la utilización de esta, lo cual se hace generando acortadores de enlaces que permiten registrar la cantidad de ingresos en las diferentes vistas, este indicador es medido desde la puesta en producción (mayo) hasta fines de octubre de 2018.

Los resultados obtenidos permiten concluir que la plataforma cumple con el propósito de reducir la carga de los ejecutivos disminuyendo el tiempo invertido en la extracción, procesamiento y análisis de los datos en un 41.5 %. Además, un 83.33 % de los DPE junto con sus respectivos equipos incorporan la plataforma de manera recurrente en sus labores, por tanto se adopta la plataforma como medio oficial para el acceso a la información. Adicionalmente, la implementación logra aumentar la utilización de los datos respecto a la situación inicial, ya que en promedio los usuarios ingresan más de una vez por semana a la plataforma. Con esto se obtiene un mayor uso, a un menor costo, traduciéndose en un proceso más eficiente.

Tabla de Contenido

1. Introducción	1
1.1. Contexto, motivación y definición del problema	2
1.2. Objetivos	3
1.2.1. Objetivos Generales	3
1.2.2. Objetivos Específicos	3
1.2.3. Objetivos Adicionales	3
1.3. Resultados esperados y alcances	3
1.4. Organización de la memoria	4
2. Marco Teórico y Estado del Arte	6
2.1. Business Intelligence	6
2.1.1. Etapas de un sistema de BI	7
2.1.2. Impacto del uso de BI en organizaciones	8
2.1.3. Principales proveedores y participación	9
2.1.4. BI en Chile	10
2.2. Herramientas a disposición	11
2.2.1. Bluemix o IBM Cloud	11
2.2.2. Watson Studio	11
2.2.3. Watson Analytics	12
2.2.4. DB2: Warehouse	12
2.2.5. Cognos Analytics	12
2.2.6. Open-Source	13
2.3. Modelos de Predicción [1][2]	13
2.3.1. Series de tiempo	13
2.3.1.1. Tipos de series de tiempo	14
2.3.1.2. Objetivo de análisis de series de tiempo	14
2.3.1.3. Métodos de descripción de series	14
2.3.2. Proceso auto regresivo (AR)	15
2.3.3. Proceso de media móvil (MA)	16
2.3.4. Proceso auto regresivo de media móvil (ARMA)	16
2.3.5. Proceso auto regresivo integral de media móvil (ARIMA)	16
2.3.6. Proceso auto regresivo integral de media móvil estacional (SARIMA)	17
2.3.7. Red neuronal artificial [3]	17
2.3.7.1. Capas de una red neuronal	17
2.3.7.2. Perceptrón	18
2.3.8. Redes neuronales recurrentes (RNN)	19

2.3.8.1.	Long-Short Term Memory LSTM[4][5]	20
2.3.8.2.	Paso a paso dentro de una LSTM	21
2.3.9.	Evaluación de la predicción	22
2.3.9.1.	MAPE	22
2.3.9.2.	MSE	23
2.3.9.3.	RMSE	23
3.	Metodología de Implementación	24
4.	Antecedentes para la implementación	26
4.1.	Glosario	26
4.2.	<i>Delivery Project Executive</i>	27
4.3.	Fuentes de información	27
4.4.	Flujo de información	29
5.	Diseño de solución	31
5.1.	Arquitectura de plataforma	31
5.2.	Métricas de evaluación de desempeño	32
6.	Análisis Exploratorio de los datos	33
7.	Implementación	37
7.1.	<i>Minimum Viable Product</i> (MVP)	37
7.1.1.	Preprocesamiento y limpieza de datos	37
7.1.2.	Implementación en Watson Analytics	38
7.1.3.	Distribución de los datos	39
7.1.4.	Resultados Preliminares y Análisis	41
7.2.	Plataforma V1.0	41
7.2.1.	Exploración de alternativas	41
7.2.2.	Propuesta de arquitectura en la nube	42
7.2.3.	Creación de base de datos	43
7.2.4.	Automatización	44
7.2.5.	Implementación en IBM Studio	44
7.2.6.	Distribución de los datos	45
7.2.7.	Resultados preliminares y análisis	45
7.3.	Plataforma V2.0	46
7.3.1.	Reestructuración de arquitectura	46
7.3.2.	Integración datos en formato de texto	47
7.3.3.	Predicción	48
7.3.3.1.	Consideraciones	48
7.3.3.2.	Descripción de Series de Tiempo.	49
7.3.3.3.	Modelos a utilizar	50
8.	Resultados y análisis	52
8.1.	Arquitectura	52
8.2.	Predicción	53
8.2.1.	SARIMA	53
8.2.2.	LSTM	54

8.2.2.1.	Configuración 1	54
8.2.2.2.	Configuración 2	54
8.2.2.3.	Configuración 3	55
8.2.2.4.	Configuración 4	55
8.2.2.5.	Modelo a implementar	55
8.3.	Visualización	57
8.4.	Robustez y confiabilidad de la plataforma	58
8.5.	Utilización de la Plataforma	59
8.6.	Reducción de tiempo	60
9.	Conclusiones y trabajo futuro	61
9.1.	Conclusiones	61
9.2.	Trabajo futuro	62
	Bibliografía	63
	Anexo A. Tablas resumen de cuentas	67
	Anexo B. Gráficos Watson Analytics - MVP	75
	Anexo C. Ejemplo de plataforma	80
C.1.	Video ejemplo Watson Analytics	80
C.2.	Ejemplo Watson Studio	80
	Anexo D. Códigos	81
D.1.	Análisis exploratorio	81
D.2.	Automatización - Proceso ETL	81
D.3.	Modelamiento SARIMA	81
D.4.	Modelamiento LSTM	81
D.4.1.	Modelo 1	81
D.4.2.	Modelo 2	82
D.4.3.	Modelo 3	82
D.4.4.	Modelo 4	82
D.5.	Preddición - Producción	82
	Anexo E. Encuesta de uso y satisfacción	83
	Anexo F. Respaldo Aprobación de plataforma	86

Índice de Tablas

6.1.	Columnas con datos vacíos. Fuente: Elaboración propia.	34
6.2.	Columnas con datos vacíos luego del filtro. Fuente: Elaboración propia.	34
6.3.	Tipos de variables. Fuente: Elaboración propia.	35
7.1.	Comparación Cognos Analytics vs Watson Studio [6][7][8][9][10].	42
7.2.	Columnas escogidas para incorporar en la visualización. Fuente: Elaboración propia.	43
7.3.	Preguntas abiertas en encuesta de satisfacción. Fuente: Elaboración propia.	46
7.4.	<i>Account ID</i> activos seleccionados para las predicciones al 02/11/2018. Fuente: Elaboración propia.	49
7.5.	Configuraciones para redes LSTM. Fuente: Elaboración propia.	51
8.1.	Resumen estadístico de métricas asociadas a modelos generados con SARIMA. Fuente: Elaboración propia.	54
8.2.	Resumen estadístico de métricas asociadas a modelos generados con configuración 1 en LSTM. Fuente: Elaboración propia.	54
8.3.	Resumen estadístico de métricas asociadas a modelos generados con configuración 2 en LSTM. Fuente: Elaboración propia.	55
8.4.	Resumen estadístico de métricas asociadas a modelos generados con configuración 3 en LSTM. Fuente: Elaboración propia.	55
8.5.	Resumen estadístico de métricas asociadas a modelos generados con configuración 4 en LSTM. Fuente: Elaboración propia.	56
8.6.	Detalle de modelamiento por Account ID considerando la configuración 3. Fuente: Elaboración propia.	56
8.7.	Tabla de uso por ejecutivo. Fuente: Elaboración propia.	59
A.1.	Tabla resumen de Work Item Id por contratos, considerando: Primer día, última día y duración en semanas. Fuente: Elaboración propia.	69
A.2.	Tabla resumen de Work Item Id por SIP, considerando: Primer día, última día y duración en semanas. Fuente: Elaboración propia.	72
A.3.	Tabla resumen de Work Item ID por SIP modificada, considerando: Primer día, última día y duración en semanas. Fuente: Elaboración propia.	74
E.1.	Encuesta de satisfacción y uso: Pregunta abierta. Fuente: Elaboración propia.	85

Índice de Ilustraciones

2.1. Cuadro comparativo Gartner 2018 [11].	10
2.2. Descomposición de series de Tiempo[1].	15
2.3. Esquema básico red neuronal [12].	18
2.4. Esquema de un perceptrón simple [13].	19
2.5. Esquema de <i>loop</i> dentro de una red neuronal recurrente [5].	19
2.6. Extensión de idea de <i>loop</i> dentro de una red neuronal recurrente [5].	20
2.7. Módulo repetido en una RNN tradicional que contiene una sola capa [5].	20
2.8. Esquema LSTM [5].	20
2.9. Celda de estado [5].	21
2.10. <i>Gates</i> o puertas en una LSTM [5].	21
2.11. <i>Forget gate layer</i> en LSTM [5].	21
2.12. <i>Input gate layer</i> en LSTM [5].	22
2.13. Nuevos candidatos junto a sus respectivos pesos [5].	22
2.14. Salida de una LSTM [5].	22
4.1. Pestaña del reporte con detalle en columnas. Fuente: Elaboración propia.	29
4.2. Flujo de trabajo inicial. Fuente: Elaboración propia.	30
5.1. Primera aproximación de flujo de trabajo. Fuente: Elaboración propia.	31
5.2. Diseño de arquitectura. Fuente: Elaboración propia.	32
6.1. Porcentaje de valores vacíos por columna. Fuente: Elaboración propia.	33
6.2. Correlación entre columnas numéricas. Fuente: Elaboración propia.	35
7.1. Vista histórica de Costos en USD por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.	39
7.2. Vista histórica de horas por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.	39
7.3. Vista costos en USD de últimas 4 semanas con distinción por tipo de hora. Fuente: Elaboración propia.	39
7.4. Costos últimas 4 semanas por Account ID y con diferenciación en tipo de hora. Fuente: Elaboración propia.	39
7.5. Detalle de horas cargadas por empleado, haciendo la distinción por tipo de hora cargada. Fuente: Elaboración propia.	40
7.6. Distribución de costos en USD por Banda/Empleado. Fuente: Elaboración propia.	40

7.7. Porcentaje de distribución de costo en USD por banda. Fuente: Elaboración propia.	40
7.8. Mapa de relación hora y costo. Fuente: Elaboración propia.	40
7.9. Costo en USD por Work Item ID (servicio). Fuente: Elaboración propia. . . .	40
7.10. Arquitectura en la nube. Fuente: Elaboración propia.	42
7.11. Encuesta realizada sobre reducción de tiempo. Fuente: Elaboración propia. .	45
7.12. Arquitectura final en la nube. Fuente: Elaboración propia.	47
7.13. Reporte formato txt. Fuente: Elaboración propia.	48
7.14. Ejemplo de costos para un Account ID en específico. Fuente: Elaboración propia.	50
7.15. Descomposición de serie de tiempo. Fuente: Elaboración propia.	50
8.1. Arquitectura completa. Fuente: Elaboración propia.	52
8.2. Correlación entre MAPE obtenido y cantidad de datos utilizados. Fuente: Elaboración propia.	57
B.1. Vista histórica de Costos en USD por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.	75
B.2. Vista histórica de horas por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.	75
B.3. Vista costos en USD de últimas 4 semanas con distinción por tipo de hora. Fuente: Elaboración propia.	76
B.4. Costos últimas 4 semanas por Account ID y con diferenciación en tipo de hora. Fuente: Elaboración propia.	76
B.5. Detalle de horas cargadas por empleado, haciendo la distinción por tipo de hora cargada. Fuente: Elaboración propia.	77
B.6. Distribución de costos en USD por Banda en últimas 4 semanas. Fuente: Elaboración propia.	77
B.7. Distribución porcentual de costos en USD por banda considerando las últimas 4 semanas. Fuente: Elaboración propia.	78
B.8. Mapa relación horas cargadas vs costos asociados a cada empleado. Fuente: Elaboración propia.	78
B.9. Costos en USD por Account ID (diferenciado por bandas). Fuente: Elaboración propia.	79
E.1. Mapa relación horas cargadas vs costos asociados a cada empleado. Fuente: Elaboración propia.	83
E.2. Encuesta de uso. Fuente: Elaboración propia.	83
E.3. Encuesta de satisfacción. Fuente: Elaboración propia.	84
E.4. Encuesta de tiempo reducido. Fuente: Elaboración propia.	84
E.5. Encuesta realizada para conocer si los ejecutivos comparten la plataforma. Fuente: Elaboración propia.	84
F.1. Respaldo Delivery Partner Manager.	86
F.2. Respaldo Delivery Project Executive.	86
F.3. Respaldo Delivery Project Executive.	86

Capítulo 1

Introducción

Hoy en día, es difícil quedar indiferente frente a la cantidad de avances tecnológicos que se desarrollan, donde la llamada nueva era cognitiva se adentra cada vez más en todos los ámbitos[14]. Sin embargo, saber aprovechar dichos avances dependerá de cada uno. Es así como las personas y organizaciones tienen la responsabilidad de informarse acerca de las nuevas tendencias y adoptarlas para mantenerse vigentes y competitivos en un mundo cada vez más dinámico y globalizado[15][16].

Es en este contexto, que surge la necesidad de buscar ventajas frente a la competencia que permitan tomar decisiones oportunas. Algunas organizaciones ya han empezado a descubrir el potencial que la tecnología ofrece para poder transformar, revitalizar y hacer crecer su negocio sea cual sea su sector (Gobierno, banca, salud, alimentos, seguridad, entre otros.) mediante la adopción de distintas herramientas, pero aún quedan muchas otras que día a día pierden oportunidades al no incorporarse al avance[17].

Business Intelligence (BI) es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Desde un punto de vista más pragmático, y asociándolo directamente con las tecnologías de la información, se puede definir BI como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa o para su análisis y transformación en conocimiento, dando así soporte a la toma de decisiones sobre el negocio[16].

BI actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio, tales como: Entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costos, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, control de gestión, entre otros ejemplos[18].

1.1. Contexto, motivación y definición del problema

En IBM Chile, en la Unidad estratégica de Negocios GTS (*Global Technology Services*) existe una estructura definida por roles y responsabilidades encargada de velar por la amplia disponibilidad de las plataformas tecnológicas que los clientes confían a IBM. Para la administración de los servicios de GTS, IBM designa a un DPE (*Delivery Project Executive*) para la entrega de los servicios comprometidos por contrato a cada cliente. Este DPE es responsable tanto de la satisfacción del cliente como de los costos incurridos para la entrega del servicio y el margen operacional.

Cada cliente cuenta con un número determinado de profesionales de las diversas áreas de gestión y técnicas que pueden trabajar en forma dedicada o compartidos entre los diversos clientes y a su vez, cada profesional carga sus horas de trabajo dedicadas a cada proyecto, cuenta o actividad que puede ser para uno o varios clientes. En esta materia, se requiere de un exhaustivo control de horas, costos y esfuerzo con el que se entrega el servicio.

Hoy en día los DPE cuentan con un reporte semanal en Excel, este incluye detalladamente la cantidad de horas que los especialistas/técnicos de las diversas áreas del conocimiento dedican y cargan mediante una plataforma llamada CLAIM a las distintas cuentas. Semana a semana los DPE deben trabajar con este reporte en Excel para monitorear el consumo de horas por especialistas y administrar los costos relacionados. El tiempo que cada DPE dedica a esta actividad de control de costos es de 2 a 3 horas a la semana en donde se estima que el 66 % del tiempo se utiliza para extraer y procesar los datos y el otro 33 % en la utilización y control a través de la información¹.

En definitiva, hoy en día en IBM Chile se utilizan herramientas que no son eficientes a pesar de que son ellos mismos quienes desarrollan soluciones analíticas y de control de gestión. En base a esto, el trabajo de título a realizar dentro de IBM consta de conocer y explorar todas las plataformas y nuevas tecnologías que actualmente se utilizan en la industria con el fin de aplicarlas a procesos internos. El objetivo es darle valor agregado a la información existente y hacer más eficiente los procesos, sobre todo de gestión y toma de decisiones.

En particular, se propone el diseño e implementación de una plataforma tipo panel de control orientado a apoyar la gestión realizada por los DPE, considerando desde la extracción de datos (proceso ETL), procesamiento (limpieza de datos y modelamiento) para finalizar con el despliegue de *insights*². Todo esto, enmarcado en un contexto de inteligencia de negocios, dado que el objetivo principal es potenciar y facilitar las decisiones estratégicas en base a la información existente.

Para la realización de dicha labor, se disponen de todas las herramientas y conocimientos adquiridos durante la preparación como ingeniero civil eléctrico, siendo de vital importancia lo aprendido en áreas tales como ciencias de datos (*Data Mining y Machine Learning*) y telecomunicaciones (*Arquitectura Cloud*), que serán utilizadas tanto en el procesamiento y despliegue de la información a utilizar, como en el diseño e implementación de la plataforma.

¹Estimación entregada por los DPE en conversaciones y encuestas iniciales.

²*Insight*: Descubrimiento o revelación no antes vista.

1.2. Objetivos

Dentro del trabajo a realizar, se considerarán los siguientes objetivos tanto generales como específicos:

1.2.1. Objetivos Generales

- Diseñar e implementar una plataforma de control de gestión basándose en tecnologías orientadas a análisis y modelamiento de datos junto con servicios en la nube propios de IBM, con la finalidad de apoyar a la toma de decisiones por parte de los ejecutivos en GTS - IBM Chile.

1.2.2. Objetivos Específicos

- Conocer e interiorizar la estructura y dinámica interna de GTS - IBM Chile en cuanto a la gestión de recursos asociados a los servicios ofrecidos a clientes.
- Comprender el concepto de *Business Intelligence*, su utilidad, sus respectivos procesos o etapas, casos de uso y estado actual en Chile y el mundo.
- Diseñar e implementar *Management System* para los DPE (*Delivery Project Executive*) considerando: horas, costos, personas, distribución de cargas, entre otros.
- Generar *insights* mediante el análisis de la información disponible.
- Diseñar e implementar arquitectura en la nube (IBM Cloud) para automatizar el flujo de actualización y distribución de la información.
- Reducir las horas hombre utilizadas para extraer y analizar la información, lo cual se traduce directamente en reducción de costos y/o aumento de productividad.
- Discutir resultados y entregar posibles vías de desarrollo para un trabajo futuro extendiendo la plataforma.

1.2.3. Objetivos Adicionales

Debido a los resultados obtenidos y en base al tiempo, durante el desarrollo del trabajo se propone extender el alcance y agregar un último objetivo adicional:

- Evaluar modelos estadísticos y de aprendizaje de máquinas para la predicción de costos y horas utilizadas por parte de los ejecutivos.
- Desarrollar un módulo de predicción, el cual será evaluado en 4 DPE.

1.3. Resultados esperados y alcances

Para este trabajo se esperan los siguientes resultados:

1. Descripción de BI y estado del arte.
2. Diseño de un panel de control o *Management System* involucrando desde la extracción de datos hasta su despliegue.
3. Implementación y operatividad del *Management System*.
4. Módulo de predicción para costos en dólares y horas cargadas en FTE.
5. *Pipeline* extensible y replicable en otros escenarios similares.
6. Métricas de uso y reducción de tiempos asociados a la plataforma.

Con respecto a los alcances, el foco del trabajo se encuentra en el diseño, implementación y medición de uso de la plataforma por parte de los usuarios. Métricas asociadas al desempeño de los DPE en cuanto a gestión antes y después de la implementación escapan del alcance de esta memoria.

Por último, es importante destacar que todos los datos y resultados expuestos en este trabajo se encuentran escalados y con nombres modificados por confidencialidad.

1.4. Organización de la memoria

El documento consta de 9 capítulos, cada uno comenzando con una breve introducción a los temas abordados. A continuación, se presenta una síntesis de cada uno de ellos:

- **Capítulo 2: Marco teórico y estado del arte** En este capítulo se revisan los antecedentes atinentes al desarrollo de esta plataforma. Comenzando por definir los conceptos necesarios de *Business Intelligence* (su relevancia y estado actual), plataforma en la nube junto a sus servicios asociados y finalmente distintos modelos de predicción.
- **Capítulo 3: Metodología de implementación** Corresponde al capítulo que detalla la metodología a utilizar en el proceso de diseño e implementación de la plataforma de control de gestión.
- **Capítulo 4: Antecedentes para la implementación** En este capítulo se detallan todos los aspectos necesarios para comprender el problema y la importancia de los datos a utilizar, descripción de parámetros relevantes y flujo de información antes de la implementación.
- **Capítulo 5: Diseño de solución** Este apartado presenta la solución a desarrollar, junto con la presentación de métricas a utilizar para evaluar el impacto de la plataforma.
- **Capítulo 6: Análisis exploratorio de los datos** En este capítulo se realiza un análisis exhaustivo para comprender la naturaleza y utilidad de los datos.
- **Capítulo 7: Implementación** En esta sección se detalla el proceso de diseño e implementación del sistema, contemplando 3 iteraciones dada la metodología a utilizar. Es importante recalcar que al cabo de cada iteración se procede con una ronda de análisis de resultados preliminares, los cuales pretenden revisar objetivos cumplidos; problemas encontrados y cambios a realizar, para evaluar y justificar -o no-, la realización de la siguiente iteración.
- **Capítulo 8: Resultados y análisis** Aquí se describen los resultados obtenidos luego de la implementación, junto con un análisis de métricas relevantes y discusión de las

mismas. Como se menciona en el punto anterior, parte de los resultados y análisis se detallan dentro del capítulo 7 debido a que se consideran parte de la implementación dada la metodología por versiones/iteraciones. Aun así, esta sección considera los resultados más relevantes para el proceso completo, incluyendo resultados de los modelos predictivos, métricas de uso y de reducción de tiempo.

- **Capítulo 9: Conclusiones** En este apartado se enumeran las conclusiones del trabajo realizado, comparando los objetivos iniciales con las metas logradas, y se identifica el trabajo futuro a realizar.

Capítulo 2

Marco Teórico y Estado del Arte

A continuación, se detalla sobre distintos conceptos y definiciones necesarias para comprender los temas tratados en esta memoria. Primero, se profundiza en el concepto de *Business Intelligence*: las etapas de un proceso BI y el estado del arte. Luego, se profundiza sobre las herramientas y servicios a disposición dentro de IBM que se utilizarán en este trabajo; se detalla la función de cada servicio y a que tipo de usuario está orientado. Por último, se presentan distintos algoritmos y modelos para la realización de pronósticos con series de tiempo.

2.1. Business Intelligence

Se denomina inteligencia de negocios o *Business Intelligence* (BI) al conjunto de estrategias, aplicaciones, datos, productos, tecnologías y arquitecturas, los cuales están enfocados a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o corporación[19].

Las funciones comunes de las tecnologías de inteligencia de negocios incluyen informes, procesamiento analítico en línea, análisis, extracción de datos, minería de procesos, procesamiento de eventos complejos, gestión del rendimiento empresarial, evaluación comparativa, minería de textos, análisis predictivo y análisis prescriptivos, entre otros[18].

Las tecnologías de BI pueden manejar grandes cantidades de datos estructurados y no estructurados para ayudar a identificar, desarrollar y, de otro modo, crear nuevas oportunidades comerciales estratégicas. Su objetivo es permitir la fácil interpretación de toda esta información. La identificación de nuevas oportunidades y la implementación de una estrategia efectiva basada en conocimientos puede proporcionar a las empresas una ventaja competitiva en el mercado y estabilidad a largo plazo[20].

La inteligencia de negocios puede ser utilizada por las organizaciones para respaldar una amplia gama de decisiones comerciales, desde operativas hasta estratégicas. Las decisiones operativas básicas incluyen el posicionamiento del producto o el precio. Las decisiones co-

merciales estratégicas implican prioridades, objetivos e instrucciones al más amplio nivel. En todos los casos, BI es más eficaz cuando combina datos derivados del mercado en el que opera una empresa (datos externos) con datos de fuentes internas de la empresa tales como datos financieros y operativos (datos internos). Cuando se combinan, los datos externos e internos pueden proporcionar una imagen completa que, en efecto, crea una inteligencia que no puede derivarse de ningún conjunto de datos singular[16].

2.1.1. Etapas de un sistema de BI

La estructura de un sistema BI comprende 5 etapas que deben cubrirse para diseñar una solución completa que impacte en una organización. A continuación, se procederá a detallar cada una de estas etapas[19]:

- **Fuentes de datos:** La primera etapa en cualquier sistema BI es definir los objetivos, y con eso los datos a utilizar. Hoy en día, cualquier organización tiene una mezcla de sistemas de registro, conocidos normalmente como OLTP (*On-Line Transactional processing Systems*), los cuales son utilizados generalmente para administrar las transacciones, clientes, proveedores, equipos, entre otros. Aquí se encuentran los ERP (*Enterprise Resource Planification*), CRM (*Costumer Relationship Management*), POS (*Point of Sales*), entre otros. Por lo general, estos sistemas mantienen los datos almacenados de forma limitada, y luego son archivados en Data Warehouses. Es con estos datos que, procesándolos de forma oportuna, es posible darles un valor agregado y transformarlos en activos para una empresa.
- **Data Warehouses:** Son sistemas de bases de datos (generalmente relacionales) que poseen como característica principal que almacenan datos históricos de cualquier índole. Esta etapa se encuentra fuertemente ligada a la etapa anterior y al proceso siguiente (ETL). La importancia de esta etapa es la mantención de datos históricos, los cuales deben ser almacenados de forma de poder realizar consultas posteriores que permitan su análisis.
- **Extracción, Transformación y Carga de Datos (ETL):** Proceso fundamental en cualquier proyecto de análisis de datos, y en especial en sistemas de BI. Tal y como su nombre lo dice ETL: *Extract, Transform and Load*, es la extracción de los datos, su posterior procesamiento y carga para una eventual visualización o despliegue de resultados. Este proceso resulta ser el más complejo en un sistema dado que en general se quiere mezclar datos de distintas fuentes, lo cual no siempre es sencillo. Usualmente, en esta etapa se realizan distintas transformaciones a los datos, tanto de limpieza, reducción y corrección [21][22].
- **Análisis de Datos:** Una vez finalizado el proceso anterior, los datos ya se encuentran disponibles para su análisis. Aquí, dependiendo del objetivo (*Forecasting, Clustering, Clasificación*, entre otros.) se aplicarán distintos modelos y técnicas estadísticas que se agrupan en lo que se conoce como *Machine Learning*[23].
- **Resultados y Visualización:** Con los modelos ya definidos y los *outputs* obtenidos, se debe desplegar y entregar la información a quien corresponda, para su posterior aplicación. En este aspecto existen diversas formas de distribución dependiendo del área de la organización, el rol del usuario que utilizará los datos, y lo que se quiera

modificar mediante este sistema. Aquí es importante recordar el perfil de quien utiliza la plataforma, ya que, para obtener un sistema eficiente, la salida generada debe ser fácil de leer y útil para el usuario. Es en este sentido que se utilizan distintas alternativas tales como reportes o *dashboards* que permitan rápidamente entender la situación y tomar decisiones.

2.1.2. Impacto del uso de BI en organizaciones

Actualmente se habla mucho de estas nuevas tecnologías y métodos que se están implementando y lanzando constantemente en el mercado, muchas de estas prometen predecir situaciones, mejorar la toma de decisiones, automatizar procesos y transparentar información entre otros beneficios. A raíz de esto surge naturalmente la pregunta ¿Cuál es el verdadero impacto del BI en las organizaciones?

Diversos estudios se han realizado para cuantificar el impacto de estas herramientas en las organizaciones, evaluando distintos aspectos y funcionalidades de estas plataformas; midiendo el impacto directo e indirecto en la toma de decisiones; y sobre todo el costo-beneficio que implica implementarlas. Para contextualizar y ejemplificar se citan diversos estudios encontrados en la literatura:

- S. Rouhani[24]

Objetivo(s)	Evaluar impacto de herramientas BI en los distintos niveles de una organización.
Experimento	Se plantean 16 hipótesis o beneficios asociados a cómo las distintas funciones de un sistema de BI permiten generar beneficios respecto a las decisiones, y cómo esto a su vez repercute a nivel estructural en una organización. Luego, se les realizan distintas preguntas con escala de 1 a 5 a aproximadamente 250 firmas que han utilizado estos sistemas por al menos un año para descartar cada una de estas hipótesis.
Resultados	Se concluyó que efectivamente las herramientas de BI influyen positivamente en el desempeño de una organización brindando beneficios en cada nivel de una organización.

- L. Calzada and J. L. Abreu [25]

Objetivo(s)	Validar que si se cuenta con información de calidad en los sistemas de información organizacionales, así como con herramientas de inteligencia de negocios que permitan la explotación de la información, los ejecutivos de las empresas podrán tomar decisiones bien fundamentadas e informadas.
Experimento	Revisión bibliográfica de fuentes documentadas (estudio de SAS 2009).
Resultados	El estudio muestra que las prácticas eficaces de administración de la información aunado con una estrategia contundente de inteligencia de negocios, ayuda a lograr un mejor desempeño a nivel de toda la organización. Por otro lado, se detallan y explicitan los obstáculos para lograr una implementación adecuada.

- A. Popovič, R. Hackney P. Simões and J. Jaklič [26]

Objetivo(s)	Examinar las relaciones entre la madurez, la calidad de la información, la cultura analítica de toma de decisiones y el uso de los datos para la toma de decisiones como elementos significativos del éxito en sistemas de inteligencia de negocios.
Experimento	Se realiza un análisis estadístico a partir de datos obtenidos de 181 organizaciones consideradas medianas y grandes.
Resultados	Construir una cultura analítica dentro de una organización es de vital importancia para mejorar el acceso y la madurez de los datos, lo que se traduce en un mayor impacto general y por tanto mayor tasa de éxito.

Estos ejemplos, y tantos otros [27][28][29] dejan en evidencia la necesidad por parte de las organizaciones de integrar rápidamente este tipo de prácticas, o de lo contrario, perderse competitivamente en el mercado[30].

Es importante destacar que se debe contar con un completo plan de integración para incorporar estas herramientas, dado que el no realizarlo efectivamente (sin una ruta definida, capacitaciones y una correcta orientación a las metas de la organización) se puede convertir en un costo innecesario además de un desincentivo a volver a tratar con nuevas iniciativas similares [27].

2.1.3. Principales proveedores y participación

Dentro del mundo del BI, las grandes compañías han destinado esfuerzos y recursos para desarrollar herramientas que permitan facilitar el trabajo para las distintas industrias. Es en este rubro donde hoy en día existen más de 30 grandes compañías dedicadas a desarrollar sistemas completos para análisis y detección de *insights*. Año a año la consultora Gartner evalúa dichas herramientas en base a: su participación en el mercado, ventajas y desventajas. En la figura 2.1 es posible ver el cuadro comparativo realizado el 2018, en donde se aprecian las distintas empresas con sus respectivos productos que destacan en esta industria[11].

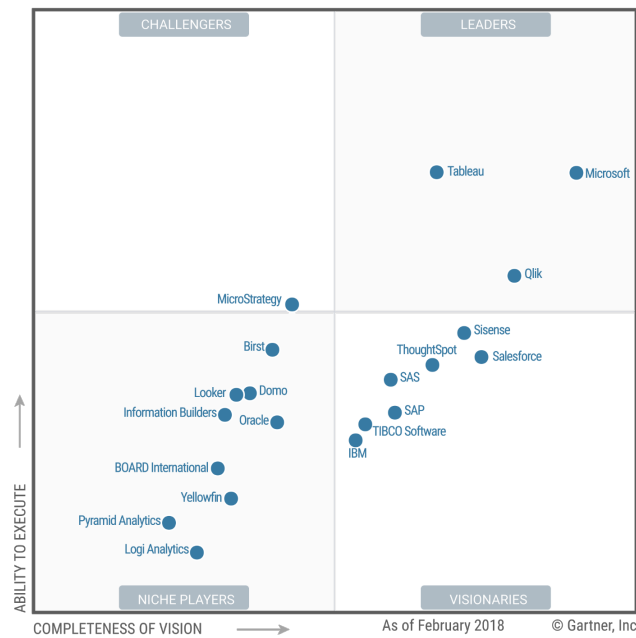


Figura 2.1: Cuadro comparativo Gartner 2018 [11].

Dentro de las principales se pueden encontrar a Microsoft con su software Power BI, el cual resulta ser el complemento perfecto al tradicional Excel. Esta empresa al día de hoy ocupa el quinto lugar en el ranking 2018 creado por Capterra[31].

También se puede encontrar a Tableau, el cual ya lleva cerca de 14 años y actualmente ocupa el primer lugar según la percepción del mercado, debido a su fácil uso y posibilidades a la hora de desplegar los datos.

Más abajo en el cuadro de visionarios, es posible encontrar a SAP, un conocido del mercado desde los años 70 que al día de hoy posee el tercer lugar. En este mismo recuadro es posible ver la empresa IBM, la cual se embarca con 2 productos: Watson Analytics y Cognos Analytics, en donde al día de ocupan el cuarto lugar del mercado en cuanto a software de BI se trata.

Analizando el lado izquierdo del gráfico es posible encontrar múltiples empresas pero que no logran abarcar una parte significativa del mercado, es más, si se excluye Oracle del análisis los demás softwares prácticamente no aparecen en el ranking de Capterra[31].

2.1.4. BI en Chile

Para comprender el mercado de estas herramientas y el nivel de uso en Chile, en el año 2013 se realizó un estudio que pretendía evaluar el nivel de inmersión de estas soluciones en el país[16]. Los resultados del trabajo dejan ver el estado de Chile versus otros países, en donde lentamente las empresas en Chile empiezan a tomar conciencia de qué son estas herramientas y las posibilidades que generan, pero con respecto a otros países aún se encuentra muy atrás, generando desventajas en cuanto a crecimiento.

Aun así, el trabajo mencionado no permite concluir el estado actual de lo que es BI en

Chile, dado que 5 años atrás el panorama era completamente distinto, considerando el auge de la transformación digital que se vive hoy en el país[32].

2.2. Herramientas a disposición

Para la implementación de esta plataforma orientada a gestión, y bajo el contexto de incentivar el uso de herramientas propias de IBM, se utilizará la plataforma en la nube Bluemix o IBM Cloud y todos los servicios adicionales disponibles. En esta sección se describirá brevemente las aplicaciones más relevantes a utilizar.

2.2.1. Bluemix o IBM Cloud

IBM Cloud, más conocido como Bluemix es la nube de IBM, la cual ofrece tanto PaaS (*Platform as a service*) como IaaS (*Infrastructure as a service*). El primero de estos, orientado a que las organizaciones puedan implementar y acceder a servicios virtualizados de TI, como por ejemplo, poder de computo, almacenamiento y *networking*. Y el segundo, que las organizaciones pueden implementar en distintos lenguajes múltiples aplicaciones en una nube pública, o bien, en ambientes *on-premises*¹. Al igual que otras nubes importantes, esta soporta distintos lenguajes de programación tales como Python, Java, PHP, entre otros.

2.2.2. Watson Studio

Watson Studio provee valga la redundancia, un estudio diseñado para la creación de aplicaciones relacionadas a aprendizaje de máquinas o *Machine Learning*, que pretenden mediante la innovación, potenciar los negocios. Su principal ventaja es que proporciona un conjunto de herramientas para los científicos de datos; desarrolladores de aplicaciones y expertos en la materia, lo que les permite conectarse de forma colaborativa a los datos, analizar esos datos y utilizarlos para crear e implementar modelos a escala.

Entre dichas herramientas se encuentran los ya conocidos Notebooks en Python o Rstudio en R, ambos diseñados para analizar y modelar datos. Adicionalmente, este estudio cuenta con la posibilidad de conectarse con otras aplicaciones de IBM, como por ejemplo Cognos Embedded, el cual sirve para el despliegue de la información una vez procesada, permitiendo una rápida visualización dado el motor de consultas que posee internamente, el cual constantemente almacena cachés que permiten interactuar con los gráficos velozmente.

¹*On-premise*: Servicio alojado en el lado del cliente, esta opción usualmente es utilizada cuando existe un alto grado de confidencialidad en los datos/aplicación y por tanto es complejo operar fuera de la red del cliente.

2.2.3. Watson Analytics

Watson Analytics es actualmente el nuevo buque insignia en cuanto a analítica y sistemas cognitivos de IBM. Bajo la premisa de *self-service*², esta plataforma se encuentra enfocada a que el usuario final, sin ser experto en análisis de datos o estadística, pueda encontrar sus propios *insights* de manera fácil e intuitiva.

Para esto, Watson Analytics posee la capacidad de entender el lenguaje humano, brindando la posibilidad de preguntarle a la aplicación información relevante sobre los datos, y esta mediante sus modelos analíticos generará sugerencias, gráficos y estadística que considera relevante.

Dado que está orientada al usuario común, la aplicación cuenta con una serie de restricciones que no permiten sacarle el máximo provecho a los datos, como podría realizarse haciendo un análisis exhaustivo en Python o R. Además, presenta problemas en cuanto a rapidez para bases de datos grandes.

2.2.4. DB2: Warehouse

Al igual que muchas otras compañías dedicadas a las tecnologías de información (TI), IBM posee su propio sistema de base de datos estructurados llamado DB2, el cual interactúa fácilmente con las herramientas internas de IBM, permitiendo conectar rápidamente los datos con los módulos de procesamiento o de despliegue de datos. Una variante de esta base de datos es DB2 Warehouse, el cual está orientado al almacenamiento de grandes cantidades de datos históricos.

2.2.5. Cognos Analytics

Cognos Analytics, a diferencia de Watson Analytics se presenta como una herramienta completa a la hora de realizar reportaría inteligente o tableros de control. Ya que se enfoca en usuarios con un nivel de expertiz medio.

Una de las principales características que posee Cognos Analytics es la gran capacidad que tiene para difundir la información, ya sean reportes u otro tipo de entregas, estos se pueden compartir de manera online, o generar archivos descargarles que pueden ser almacenados.

A diferencia de Watson analytics, esta plataforma presenta una distribución en la nube y otra *on-premise*, debido a la desconfianza que existe por parte de los clientes al subir sus datos en la nube. Otra de las ventajas que posee es la fácil integración dentro de aplicaciones web e incluso aplicaciones móviles, brindando un formato de visualización para clientes o internos dentro de una organización.

²Self-service o autoservicio: El usuario se considera autosuficiente y por tanto puede operar sin ayuda.

2.2.6. Open-Source

En la actualidad, cada vez más organizaciones apuestan por herramientas Open-Source, esto debido a sus grandes comunidades, fácil replicación y versatilidad. Es en ese sentido que IBM dentro de sus servicios agrega la compatibilidad de herramientas de este tipo, permitiendo la utilización por ejemplo de librerías de análisis de datos en Python o en R a disposición para generar modelos.

2.3. Modelos de Predicción [1][2]

Para efectuar la predicción de costos y recursos humanos asociados a cada cliente, se utilizarán distintas aproximaciones de donde se escogerá la mejor. Entre las opciones a utilizar se encuentran aproximaciones de series de tiempo tanto lineales como no lineales, donde se incluyen modelos ARIMA/SARIMA y Redes neuronales recurrentes (RNN).

2.3.1. Series de tiempo

El estudio de series de tiempo ve su aplicación en distintas áreas de interés, como análisis económicos, proyecciones de ventas, planificación de redes, medicina, ingeniería, entre otros. Además, dependiendo de los objetivos, se pueden encontrar distintas técnicas para el tratamiento de los datos y obtener así modelos que permitan tomar acciones de predicción, detección y control, según su aplicación.

Una serie de tiempo es una secuencia de observaciones (muestras) de una variable particular donde cada observación se registra dentro de un tiempo específico. En el caso de las series de tiempo discretas, estas observaciones son registradas en períodos de tiempo regulares.

2.3.1.1. Tipos de series de tiempo

Las series de tiempo pueden clasificarse según sus características de construcción de la siguiente manera:

1. **Muestreadas:** Series obtenidas a partir del muestreo de un conjunto de medidas continuas que pueden ser registradas de manera constante por un periodo de tiempo. Corresponde a una discretización de series continuas.
2. **Agregadas:** Generadas a partir de la acumulación o agregación de datos concatenados, obteniéndose series que representan una condición global.
3. **Discretas:** Series de registros consecutivos obtenidos con periodos regulares de tiempo entre cada muestra.

2.3.1.2. Objetivo de análisis de series de tiempo

Los principales objetivos que se persiguen con el análisis de series de tiempo son:

1. **Descripción:** Caracterizar la serie de tiempo utilizando métodos de descripción estadísticos, gráficos, valores representativos (máx., min., tendencias, ciclos), entre otros.
2. **Modelamiento:** Generar modelos que representen el comportamiento de los datos. Estos modelos pueden ser univariantes, es decir, que consideran solo una variable, o multivariantes, en el caso en que se consideren valores presentes y pasados de la misma u otras variables asociadas al proceso.
3. **Predicción:** Estimar los valores futuros de una serie de tiempo, manteniendo una representación cercana a la realidad, a partir de los valores pasados y presentes.
4. **Control:** A partir de la caracterización del comportamiento de los procesos obtenida desde las series de datos, es posible realizar acciones de control. El control a realizar está estrechamente ligado a la predicción y, por supuesto, al tipo de modelo generado.

2.3.1.3. Métodos de descripción de series

Antes de proceder con el análisis definitivo sobre las series de datos, es necesario procesar la información con el fin de remover errores que se puedan encontrar en esta, como datos faltantes, mediciones incorrectas, valores fuera de escala, entre otros.

En una serie de tiempo se pueden encontrar dos componentes de variación que la conforman, la tendencia y la variación estacional, como muestra la figura 2.2[2][33].

Como expresión matemática, una serie de tiempo está representada por:

$$X_t = f(S_t, T_t, C_t, I_t) \quad (2.1)$$

donde cada una de las componentes se define como sigue:

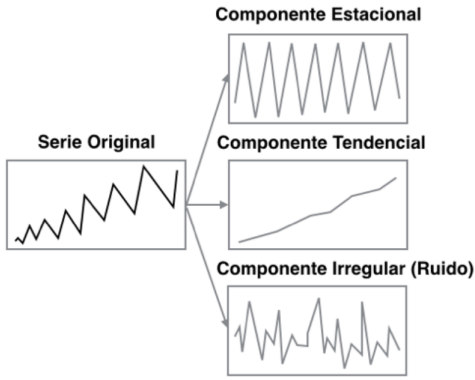


Figura 2.2: Descomposición de series de Tiempo[1].

1. **Variación estacional (S_t):** Variación que presenta un patrón de comportamiento similar dentro de un periodo largo de tiempo. Por ejemplo, oscilaciones anuales, mensuales, semanales, etc.
2. **Tendencia (T_t):** Variación que refleja el incremento o reducción constante de los valores de la serie de tiempo dentro de un periodo de muestra o dentro de una variación estacional.
3. **Variación cíclica (C_t):** Incluye variaciones de los datos en periodos de magnitud mayor. Generalmente se asocia a fenómenos cíclicos como procesos económicos de décadas, fenómenos naturales, comportamientos biológicos, etc.
4. **Fluctuaciones (I_t):** Factor que describe las variaciones remanentes después de remover las componentes tendenciales, estacionales y otros valores sistemáticos. Estas fluctuaciones pueden ser completamente aleatorias, en cuyo caso no es posible pronosticarlas.

2.3.2. Proceso auto regresivo (AR)

Una serie de tiempo Z_t se dice que es un proceso auto regresivo de orden p (AR(p)) si el proceso cumple con la siguiente expresión:

$$Z_t = \sum_{k=1}^p \Phi_k Z_{t-k} + e_t \quad (2.2)$$

donde Φ_k corresponde a un factor escalar y e_t es un ruido de media cero y varianza σ_e^2 . En otras palabras, el valor de la serie de tiempo en el tiempo t es igual a una combinación lineal de p valores anteriores de la serie, más una componente de ruido.

2.3.3. Proceso de media móvil (MA)

Una serie de tiempo Z_t se dice que es un proceso de media móvil de orden q (MA(q)) si el proceso cumple con la siguiente expresión:

$$Z_t = \sum_{k=1}^q \phi_k e_{t-k} + e_t \quad (2.3)$$

donde ϕ_k corresponde a un factor escalar y e_t es un ruido de media cero y varianza σ_e^2 . Por lo tanto, el valor de la serie en el tiempo t es una combinación lineal de los q valores de ruido anteriores más una componente nueva de ruido e_t .

2.3.4. Proceso auto regresivo de media móvil (ARMA)

Cuando se combinan los dos modelos presentados anteriormente se obtiene un modelo auto regresivo de media móvil. Si el modelo cuenta con p términos auto regresivos y q términos de media móvil (ARMA(p,q)), entonces la expresión que lo describe está dada por:

$$Z_t = \sum_{k=1}^p \Phi_k Z_{t-k} + \sum_{k=1}^q \phi_k e_{t-k} \quad (2.4)$$

2.3.5. Proceso auto regresivo integral de media móvil (ARIMA)

Los modelos representados anteriormente funcionan para modelos estacionarios, pero este no es el caso para la mayoría de las series de tiempo a estudiar. Para solucionar este problema, se proponen los modelos auto regresivos integrales de media móvil (ARIMA), en donde a través de la diferenciación de la serie de tiempo se obtiene un proceso estacionario.

La expresión que describe un modelo ARIMA está dada por:

$$\left(1 - \sum_{k=1}^p \phi_k B^k\right)(1 - B)^d Z_t = \sum_{k=1}^q \phi_k e_{t-k} + e_t \quad (2.5)$$

donde B corresponde a un operador de retraso para la serie de tiempo como se describe a continuación:

$$BZ_t = Z_{(t-1)} \quad (2.6)$$

$$B^d Z_t = Z_{(t-d)} \quad (2.7)$$

Con esto, se agrega como nuevo grado de libertad al modelo el valor de d , que corresponde al nivel de diferenciación con el que se genera el modelo.

2.3.6. Proceso auto regresivo integral de media móvil estacional (SARIMA)

Como última generalización, se puede agregar la condición de estacionalidad de la serie de tiempo en estudio, si es que corresponde. generando entonces un modelo ARIMA Estacional o SARIMA. De esta manera se incluye de manera matemática la estacionalidad de los datos en el momento, obteniéndose la siguiente expresión que lo describe:

$$(1 - \sum_{k=1}^p \phi_k B^k)(1 - B)^d(1 - B^s)^D Z_t = \sum_{k=1}^q \phi_k e_{t-k} + e_t \quad (2.8)$$

Además se puede generalizar esta expresión de manera multiplicativa obteniéndose:

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D Z_t = \Theta(B^s)\theta(B)e_t \quad (2.9)$$

donde:

$$\theta(B) = 1 - \sum_{k=1}^p \phi_k B^k \quad (2.10)$$

$$\Theta(B^s) = 1 - \sum_{k=1}^p \Phi_k B^k_s \quad (2.11)$$

$$\theta(B) = 1 + \sum_{k=1}^p \theta_k B^k \quad (2.12)$$

$$\Theta(B^s) = 1 + \sum_{k=1}^p \Theta_k B^k_s \quad (2.13)$$

2.3.7. Red neuronal artificial [3]

Las redes neuronales son un modelo basado en un gran conjunto de unidades neuronales artificiales simples o perceptrones, de forma análoga al comportamiento observado en los axones de las neuronas en los cerebros biológicos. Cada unidad neuronal está conectada con muchas otras y los enlaces entre ellas pueden incrementar o inhibir el estado de activación de las neuronas adyacentes. Cada perceptrón, de forma individual, opera empleando funciones de suma. Puede existir una función limitadora o umbral en cada conexión y en la propia unidad, de tal modo que la señal debe sobrepasar un límite antes de propagarse a otra neurona. Estos sistemas aprenden y se forman a sí mismos, en lugar de ser programados de forma explícita, y sobresalen en áreas donde la detección de soluciones o características es difícil de expresar con la programación convencional.

2.3.7.1. Capas de una red neuronal

Las redes neuronales suelen consistir en varias capas, y la ruta de la señal atraviesa de adelante hacia atrás.

En la figura 2.3 es posible identificar 3 grupos de neuronas o perceptrones: El conjunto coloreado rojo que corresponde al de entrada o *input*, conjunto azul que viene siendo la capa oculta y finalmente el conjunto verde que es la salida u *output*.

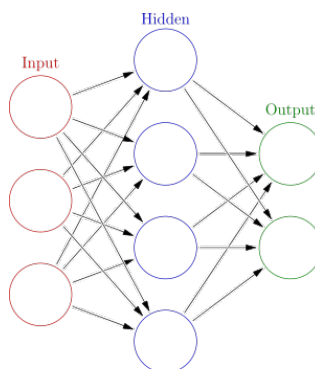


Figura 2.3: Esquema básico red neuronal [12].

- **Capa de entrada:** Encargada de recibir los datos para su procesamiento, normalmente se utiliza una neurona por cada característica.
- **Capa oculta:** El número de neuronas en la capa oculta define la capacidad de la red a ajustarse a los datos de entrenamiento, por lo que sí es muy pequeño (menor al número de entradas), el problema no tendrá flexibilidad para aprender de los datos y definir la frontera de decisión. Por otro lado, si hay muchas neuronas, la red se sobreajustará a los datos memorizando el conjunto de entrenamiento. Otro problema que surge al definir el número de capas ocultas es tener muchas de estas, lo que produce un fenómeno llamado *Vanishing Gradient*, el cual es un efecto de la propagación del gradiente a través de múltiples capas neuronales, que provoca una reducción del impacto del gradiente en capas posteriores, y por tanto aumenta la complejidad de entrenamiento[34].
- **Capa de salida:** Encargada de entregar la salida de la red, normalmente se utiliza codificación *One-hot* para representar la salida lo que se traduce en una neurona por clase.

2.3.7.2. Perceptrón

Un perceptrón es un modelo computacional simplificado de una neurona como unidad procesadora de información, el cual tiene la capacidad de reconocer patrones. La ecuación característica de un perceptrón es

$$y_k = \varphi\left(\sum_{j=1}^n w_{kj}x_j + b_k\right) \quad (2.14)$$

El perceptrón está constituido por un conjunto de sensores de entrada x_k , los cuales son ponderados por distintos pesos modificables w_k llamados pesos sinápticos y una neurona de salida que se encarga de clasificar las respuestas, mediante una función de activación φ a las entradas en distintas clases. Un modelo de perceptrón se muestra en la figura 2.4.

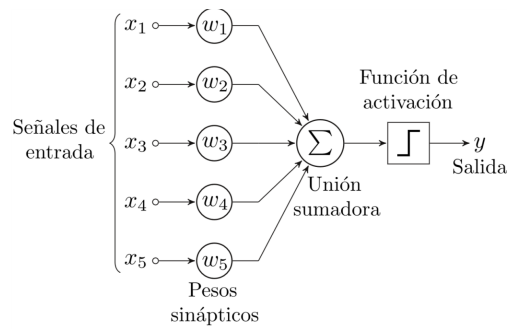


Figura 2.4: Esquema de un perceptrón simple [13].

El algoritmo perceptrón, según el Teorema de Convergencia: "*Converge siempre en tiempo finito para un conjunto linealmente separable y finito de ejemplos, en donde se define que un conjunto es linealmente separable si es posible separar sus clases mediante un hiperplano en el espacio de entradas x* ".

2.3.8. Redes neuronales recurrentes (RNN)

A pesar de la eficiencia y gran utilidad de las redes neuronales tradicionales, aún hay problemas que no pueden ser abordados con estos modelos, como por ejemplo, predecir la siguiente palabra dentro de un contexto, ya que en general estas no pueden usar eventos anteriores para analizar los nuevos.

Las redes neuronales recurrentes vienen a resolver este problema, estas son redes con *loops* dentro de ellas mismas, manteniendo la información persistente.

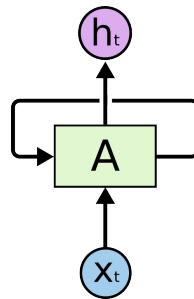


Figura 2.5: Esquema de *loop* dentro de una red neuronal recurrente [5].

En la figura 2.5 se puede ver una parte de una red neuronal **A**, la cual ve una entrada x_t y una salida h_t . El *loop* permite pasar información del estado actual, al siguiente.

Generalizando y extendiendo el esquema de *loop*, puede obtener una figura similar a la de una red neuronal tradicional, donde cada copia pasa el *output* al siguiente.

La figura 2.6 deja ver de manera intuitiva que las redes recurrentes están íntimamente relacionadas con las secuencias, y por tanto, es arquitectura primordial a la hora de trabajar con series de tiempo.

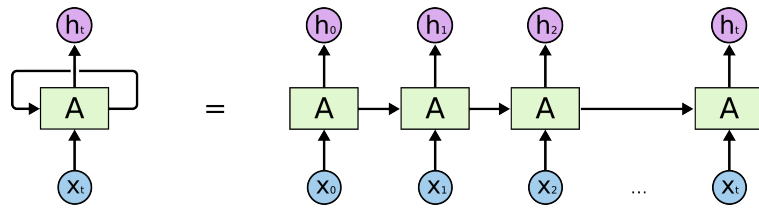


Figura 2.6: Extensión de idea de *loop* dentro de una red neuronal recurrente [5].

Teniendo la idea intuitiva de como funciona una RNN, se tendería a pensar que su capacidad para almacenar eventos anteriores permite que se almacenen sin límite, pero lamentablemente la práctica dice lo contrario, dado que este tipo de redes también presenta el problema de *Vanishing gradient*[34].

Es en respuesta a este problema, que surgen las *Long-Short Term Memory* (LSTM)[35].

2.3.8.1. Long-Short Term Memory LSTM[4][5]

Las LSTM son un tipo especial de RNN, con gran capacidad de aprendizaje de términos anteriores[36]. En una RNN tradicional, el módulo que se repite tiene una estructura como se muestra en la figura 2.7, donde cada módulo contiene una estructura simple.

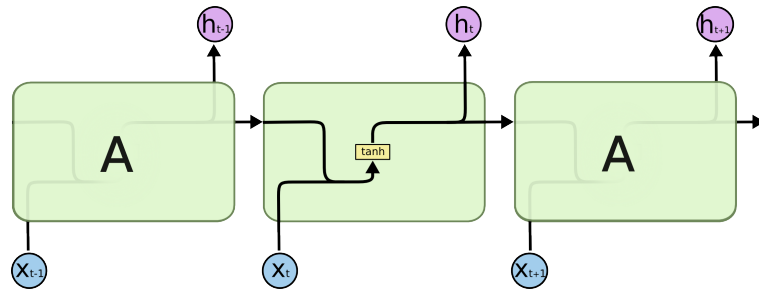


Figura 2.7: Módulo repetido en una RNN tradicional que contiene una sola capa [5].

Una red LSTM contiene la misma estructura de cadena que una RNN, pero el módulo que se repite tiene variaciones, en vez de una sola capa, estas contienen 4 capas, que interactúan de forma especial.

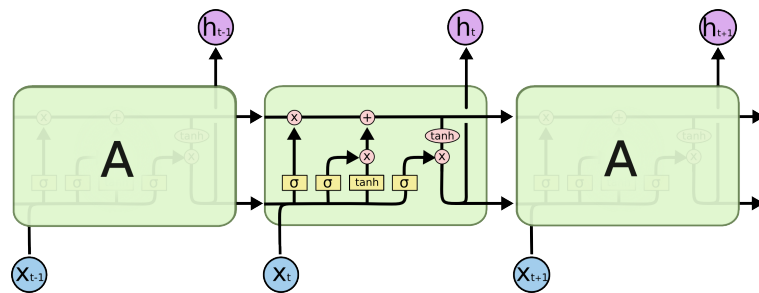


Figura 2.8: Esquema LSTM [5].

La clave de las LSTM es la celda de estado, la cual se puede representar como la línea horizontal que cruza a lo largo del diagrama 2.8, la cual se puede ver también en 2.9. Las

LSTM tienen la capacidad de agregar o remover información en las celdas de estado, a través de estructuras llamadas puertas o *gates*, las cuales se pueden ver gráficamente representadas en la figura 2.10.

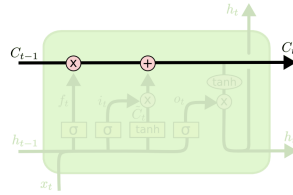


Figura 2.9: Celda de estado [5].

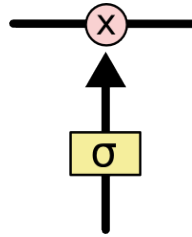


Figura 2.10: *Gates* o puertas en una LSTM [5].

Las puertas o *gates* son una forma de pasar información de forma opcional. Estas puertas se componen de una capa neuronal sigmoide y un operador de multiplicación. En donde la salida de la sigmoide puede ir de valores 0 a 1 indicando cuanto de la componente deja pasar. De la imagen 2.8 se puede ver que una LSTM posee 3 puertas que regulan y controlan la información.

2.3.8.2. Paso a paso dentro de una LSTM

El primer paso dentro de una LSTM es decidir que información va a ser removida de la celda de estado. Esta decisión la toma una de las sigmoides llamada *forget gate layer*. Esta sigmoide recibe como entrada h_{t-1} y x_t y como output un número entre 0 y 1 para cada celda de estado C_{t-1} . Esta etapa puede verse gráficamente en la figura 2.11.

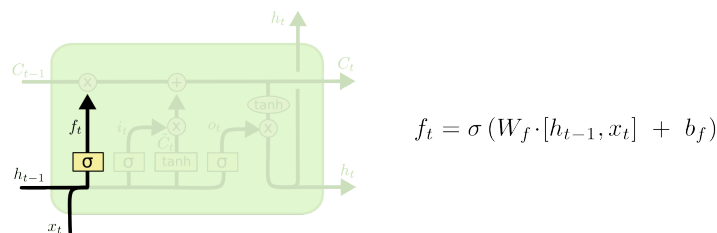


Figura 2.11: *Forget gate layer* en LSTM [5].

El siguiente paso es decidir que información se va a mantener en la celda de estado. Para esto se procede, en primer lugar, con la *input gate layer*, quien decide qué valores actualizar. Luego, la siguiente capa crea un nuevo vector de candidatos \tilde{C}_t , que podría ser agregado al estado.

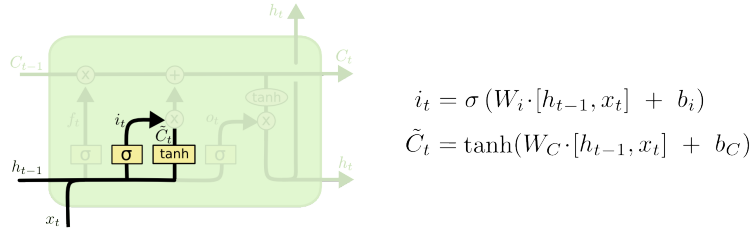


Figura 2.12: *Input gate layer* en LSTM [5].

Ahora, debe ser actualizado la celda de estado antigua C_{t-1} en la nueva celda de estado C_t . Luego, se multiplica el estado antiguo por f_t , a lo que luego se le suma $i_t * \tilde{C}_t$. Esto equivale a los nuevos candidatos valores multiplicados por un peso según cuánto se quiera actualizar.

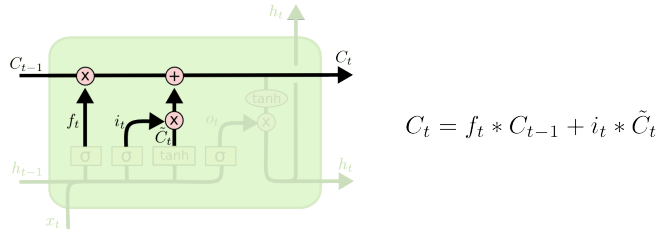


Figura 2.13: Nuevos candidatos junto a sus respectivos pesos [5].

Finalmente, falta decidir cuál va a ser el *output*. Este *output* va a estar basado en la celda de estado, aplicándole ciertos filtros. En primer lugar, la sigmoide final que parte de la celda de estado será el *output gate*. Luego, a la celda de estado se le aplica una función de activación y este resultado es multiplicado por el *output* de la sigmoide para dejar como salida solo la parte de la celda señalada. Tal y como muestra la figura 2.14.

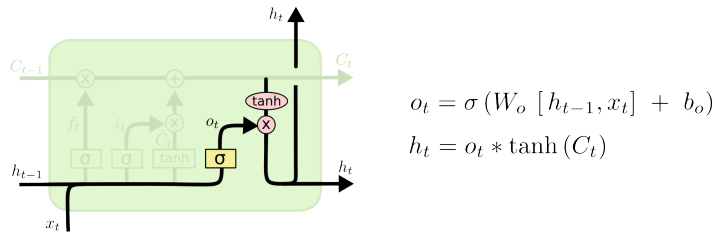


Figura 2.14: Salida de una LSTM [5].

2.3.9. Evaluación de la predicción

2.3.9.1. MAPE

Mean Absolute Percentage Error o MAPE, es la métrica más utilizada para medir error asociado a una predicción, normalmente expresa el error como porcentaje, tal y como muestra la fórmula 2.15.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2.15)$$

Donde A_t es el valor real, F_t el valor de predicción y n la cantidad de puntos.

A pesar de ser simple e intuitivo, esta métrica presenta el problema de que no puede ser usado si hay valores iguales a cero, ya que el denominador se anularía [37].

2.3.9.2. MSE

Mean Squared Error o MSE mide el promedio de los errores al cuadrado, es decir, la diferencia de la predicción y el valor real, su fórmula se puede ver en 2.16.

$$MSE = \frac{1}{n} \sum_{t=1}^n (\hat{Y}_i - Y_i)^2 \quad (2.16)$$

2.3.9.3. RMSE

Root-Mean Squared Error o RMSE básicamente comprende la raíz del MSE, por tanto permite una interpretación más fácil ya que tiene las mismas unidades de medida que la predicción en cuestión. Su fórmula puede verse en 2.17.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{Y}_i - Y_i)^2} \quad (2.17)$$

Capítulo 3

Metodología de Implementación

Como el objetivo de este trabajo es la implementación de un panel de control que permita obtener indicadores relevantes sin mayor esfuerzo para los ejecutivos, se debe trabajar en conjunto con ellos para que la experiencia de usuario sea óptima, ya que por muy potente que sea una herramienta, si el usuario no está interesado en utilizarla o le resulta engorrosa, el desarrollo pierde el sentido.

A continuación, se procede a detallar la metodología a seguir para la implementación del panel de control o *Management system* que servirá a los ejecutivos de IBM Chile.

1. **Definir datos a utilizar:** En primer lugar se deberá determinar qué datos son los que actualmente utilizan los DPE para sus propósitos y solicitar los distintos accesos para poder descargar dichos datos.
2. **Definir métricas de desempeño:** Para evaluar el éxito en la implementación de la herramienta se deberán buscar métricas que permitan al final del proyecto, conocer si efectivamente la plataforma resuelve el problema actual o hay que seguir experimentando con alternativas.
3. **Definir tecnología a utilizar:** En base al objetivo planteado de incentivar y utilizar tecnología propia de IBM, las alternativas a utilizar deberán ser propiedad de la compañía o en su defecto, Open-Source. Aquí se explorarán distintas alternativas en donde se evaluará la que mejor se adapte a la necesidad actual, según distintos parámetros: Velocidad, métodos de difusión, capacidad de análisis, facilidad de uso, entre otros.
4. **Procesar y analizar la Data:** Una vez obtenida la data, se reunirá para formar una gran base de datos (ETL), la cual se procesará con el fin de encontrar indicadores útiles. Aquí es importante destacar que, al ser un caso real de implementación, las fuentes de datos a utilizar presentan múltiples problemas: Datos mal etiquetados, información incompleta, celdas o archivos completos vacíos, entre otros. Para esta etapa, es de suma importancia entender el contexto del problema a resolver y la actividad de la empresa, ya que en general se requiere de información relacionada al negocio para entender cuáles datos se encuentran erróneos y cuáles no. También es importante destacar el valor de la información y el impacto que tiene el equivocarse con dichos datos, ya que cualquier error en el procesamiento y limpieza de los datos implica errores en la toma

de decisiones, lo que se traduce en costos no previstos o incluso despido de personal.

5. **Encontrar y desplegar insights:** Una vez procesada la información, se modelará la información con el fin de encontrar indicadores útiles. Luego, se deberá desplegar dichos datos de forma que sean de fácil comprensión y rápida utilización.
6. **Definir método de distribución:** Al ser información delicada, se debe procurar que solo las personas asignadas puedan acceder a la plataforma (funcionarios específicos de IBM). Adicionalmente, se evaluará distintos formatos de entrega, tratando siempre de buscar la comodidad por parte del usuario final para visualizar y recibir dichos datos.
7. **Montar sistema en la nube:** Para no depender de un equipo local que realice toda esta actividad, se debe procurar enviar a la nube todos -o la mayor cantidad- de procesos para que siempre estén disponibles los equipos locales.
8. **Automatizar procesos para la recolección y publicación de la información:** Uno de los objetivos es reducir horas hombre, por lo que se deberá automatizar el proceso de recolección para así evitar depender de terceros.

Muchos de los pasos anteriormente mencionados, se deberán iterar constantemente en conjunto con los usuarios para ajustar la solución, esto debido a que se trabajará con una metodología ágil o *Agile*, la cual pretende a grandes rasgos servir como apoyo a proyectos dinámicos o cambiantes en el tiempo, dividiendo el proyecto en pequeñas etapas semanales, en donde la iteración es el principio fundamental para una mejora continua[38][39].

Asimismo, se trabajará bajo el concepto de *Design Thinking*, el cual es un conjunto de metodologías o procesos que abordan problemas colocando a los usuarios como foco principal del proyecto, buscando empatizar para entender las necesidades y obtener una experiencia de usuario completa [40].

Capítulo 4

Antecedentes para la implementación

Esta sección tiene como función introducir conceptos necesarios para comprender la problemática y la solución a desarrollar, abordando la terminología propia de IBM y la situación previa a la plataforma.

4.1. Glosario

A continuación, se detalla brevemente siglas y términos propios de IBM que serán utilizados constantemente durante el desarrollo de este trabajo:

- **DPE:** *Delivery Project Executive*, ejecutivo encargado de velar por los recursos asociados a los clientes dentro de IBM.
- **FTE:** *Full Time Equivalent*, métrica que se utiliza para hablar de recursos humanos, 1 FTE equivale a la jornada laboral de una persona durante un mes.
- **Account ID o Work Item ID:** Dentro de cada contrato se encuentran distintos *Account ID* (AI) o *Work Item ID* (WII) donde cada uno de estos representa un centro al cual se cargan costos y horas dependiendo de la actividad realizada.
- **SIP o Contract Number:** Código de contrato entre IBM y el cliente.
- **Service Line:** Se considera un grupo de productos o servicios relacionados a cierta unidad dentro de IBM.
- **CLAIM:** Plataforma en la cual los empleados deben registrar sus horas trabajadas con su respectivo detalle.
- **Skills:** Habilidades, conocimientos o expertiz en algún área o trabajo.
- **Banda del empleado:** Costos asociados a la hora trabajada del empleado (Tanto en Chile como en otros países). Estos costos dependen de múltiples factores, entre los cuales se consideran las *skills* del trabajador y cargo. Mientras más alta la banda, mayor serán los costos de sus horas hombre.
- **Horas Normales:** Horas dentro de la jornada laboral ordinaria.
- **Horas Overtime:** Horas extras, las cuales son horas fuera de la jornada regular.

- **Horas *Stand By*:** Horas que pueden considerarse tiempo de reserva, en el cual el empleado no está trabajando, pero si debe estar disponible para poder atender cualquier incidente que se presente.
- **Línea de Costo:** Presupuesto contemplado para costos implicados para cada servicio dentro del contrato. A medida que avanza el proyecto, se va incurriendo en costos que van aumentando la línea hasta llegar al presupuesto destinado.
- ***Week Ending Date*:** Cada reporte generado contempla información desde el día de cierre de semana o *Week Ending Date*, hasta un día antes del próximo cierre. Estas fechas están estipuladas para ser los viernes de cada semana.
- ***Business Case*:** Documento que sirve de base para valorar la viabilidad financiera, estratégica y organizacional de un proyecto.

4.2. *Delivery Project Executive*

Dentro de GTS IBM (*Global Technology Services*) se debe tener un DPE encargado de velar por los recursos asociados a cada una de las cuentas. Actualmente hay 18 DPE que abordan alrededor de 60 cuentas. Este último número varía constantemente dado que los contratos y clientes van modificándose en el tiempo.

Dependiendo del tamaño de la cuenta y la capacidad del/la DPE, un ejecutivo puede tener a cargo más de una cuenta, asimismo, si una cuenta presenta una gran cantidad de contratos, se le asignará un DPE exclusivo a dicho cliente.

4.3. Fuentes de información

Como el objetivo final de este proyecto es crear una plataforma que les permita a los ejecutivos obtener información útil al instante respecto a sus cuentas, se utilizará información relacionada a todo lo que concierne a sus cuentas semana a semana y a sus respectivos empleados. Algunos de los datos presentes son:

- Información semana a semana de los costos involucrados asociados a horas hombre.
- Personas que trabajaron o cargaron horas a distintos AI.
- Tipo de hora cargada en cada actividad, es decir, si la actividad fue realizada en su horario de trabajo, durante turno extra (*overtime*) o *stand by*.
- Momento en que fueron cargadas dichas horas a la plataforma.
- Banda del empleado que registra horas.
- País de origen, desde donde se cargó la hora, y por tanto, país en donde se encuentra el empleado.
- Nombre de los distintos contratos. Aquí un mismo cliente puede tener más de un contrato dependiendo de cómo se fue acordado el negocio.
- Información del empleado como nombres y códigos asociados.

- *Service Line*: Equivalente al área en donde se desempeña el trabajador.
- Cantidad de horas trabajadas por los distintos empleados en las distintas cuentas, considerando un resumen semanal y el detalle por días.

Actualmente la información con la que se trabaja se puede resumir en aproximadamente 70 características o columnas diferentes, con distintas cantidades de filas que van variando semana a semana. Dichos datos son generados una vez a la semana en IBM Costa Rica, en donde son subidos a un repositorio de datos llamado BOX separados en distintos Excel o planillas, con un aproximado de 60 archivos semana a semana, en donde cada archivo corresponde a un contrato.

En estos archivos -y como en cualquier aplicación real- los datos vienen con problemas, ya sea por duplicidad, celdas corridas, en blanco o valores erróneos, lo cual se debe a que en algún momento del proceso, estos son digitados manualmente. Lo anterior introduce una fuerte componente de error.

Actualmente, los DPE poseen acceso a estas planillas, pero solo se limitan a ver el desempeño de la semana anterior, sin profundizar en antecedentes históricos como podrían ser las tendencias, evolución o análisis similares debido a la complejidad y falta de tiempo para integrar todos los datos.

Los datos a encontrar en el repositorio se pueden dividir en 2 grandes categorías: **Reportes correspondientes a la semana** y **Archivos Contables**.

- **Reportes correspondientes a la semana**: Cada uno de estos reportes contiene información correspondiente a un contrato, que abarca desde el comienzo del mes, hasta la semana que indica su nombre. Por ejemplo, si el archivo se encuentra en la carpeta ‘Octubre ->Semana 3’, este informe contiene el detalle para esta cuenta de desde la primera semana de octubre hasta la tercera.
- **Archivos contables**: A diferencia de los anteriores, los informes contables son archivos en Excel que abarcan información de un contrato desde el inicio del año contable, hasta el cierre del mes en cuestión. Por ejemplo, un archivo contable en la carpeta ‘Octubre ->Contable’ contiene información del contrato desde inicios de año hasta el cierre de octubre. Otra diferencia a considerar respecto a los informes de semana es que en los archivos contables la información encontrada se considera definitiva, no así en los semanales, ya que de una semana a otra la información puede ser corregida. Por tanto, lo que aparezca en contable queda registrado sin posibilidad de cambio.

De ambos tipos de archivos, el detalle de la cuenta se puede encontrar en la hoja ‘DATA_AC’, tal y como muestra la figura 4.1, en donde cada fila representa el costo y horas cargadas por un empleado para dicha cuenta en un día.

	A	B	C	D	E	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL
					Emp Num	Sun Hrs	Mon Hrs	Tue Hrs	Wed Hrs	Thu Hrs	Fri Hrs	Usage Qty	Orig Currency Cd	Plan Exchg Rate	Rate	Rate USD	Cost	Cost USD
1																		
2					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$659.29	\$16,055.41	\$24.35	\$32,110.81	\$4
3					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$659.29	\$16,055.41	\$24.35	\$32,110.81	\$4
4					H40809	0	0	0.4	0.4	0.4	0.4	1.6	CLP	\$659.29	\$16,055.41	\$24.35	\$25,688.65	\$3
5					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$659.29	\$16,055.41	\$24.35	\$32,110.81	\$4
6					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$667.78	\$16,055.41	\$24.04	\$32,110.81	\$4
7					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$667.78	\$16,055.41	\$24.04	\$32,110.81	\$4
8					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$667.78	\$16,055.41	\$24.04	\$32,110.81	\$4
9					H40809	0	0	0.4	0.4	0.4	0.4	1.6	CLP	\$667.78	\$16,055.41	\$24.04	\$25,688.65	\$3
10					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$663.37	\$16,055.41	\$24.20	\$32,110.81	\$4
11					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$663.37	\$16,055.41	\$24.20	\$32,110.81	\$4
12					H40809	0	0	0.4	0.4	0.4	0.4	1.6	CLP	\$663.37	\$16,055.41	\$24.20	\$25,688.65	\$3
13					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$663.37	\$16,055.41	\$24.20	\$32,110.81	\$4
14					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$663.37	\$16,055.41	\$24.20	\$32,110.81	\$4
15					H40809	0	0	0	0.4	0.4	0.4	1.2	CLP	\$673.33	\$16,055.41	\$23.84	\$19,266.49	\$2
16					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$673.33	\$16,055.41	\$23.84	\$32,110.81	\$4
17					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$673.33	\$16,055.41	\$23.84	\$32,110.81	\$4
18					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$673.33	\$16,055.41	\$23.84	\$32,110.81	\$4
19					H40809	0	0.4	0.4	0.4	0.4	0.4	2	CLP	\$670.10	\$16,055.41	\$23.96	\$32,110.81	\$4
20					H40809	0	0.4	0.4	0.4	0	0.4	1.6	CLP	\$670.10	\$16,055.41	\$23.96	\$25,688.65	\$3
21					H40809	0.4	0.4	0.4	0.4	0	0.4	2	CLP	\$670.10	\$16,055.41	\$23.96	\$32,110.81	\$4
22					H40809	0.4	0.4	0.4	0.4	0	0.4	2	CLP	\$670.10	\$16,055.41	\$23.96	\$32,110.81	\$4
23					H66995	0	3	2.9	3	3	3	14.9	CLP	\$683.39	\$38,889.24	\$56.91	\$679,449.67	\$98
24					H66995	0	3	2.5	3	3	3	14.5	CLP	\$656.96	\$38,889.24	\$59.20	\$563,993.98	\$98
25					H66995	0	3	3	3	3	3	15	CLP	\$656.96	\$38,889.24	\$59.20	\$583,338.60	\$98
26					H66995	0	3	3	3	3	3	16	CLP	\$656.96	\$38,889.24	\$59.20	\$583,338.60	\$98
27					H66995	0	3	3	3	4	3	16	CLP	\$656.96	\$38,889.24	\$59.20	\$522,227.04	\$94
28					H66995	0	3	3	2.6	4	3	15.6	CLP	\$659.29	\$38,889.24	\$58.99	\$606,672.14	\$92
29					H66995	0	2	2	2.6	3	2	11.6	CLP	\$659.29	\$38,889.24	\$58.99	\$451,115.18	\$66
30					H66995	0	2	2.1	3	2	2	11.1	CLP	\$659.29	\$38,889.24	\$58.99	\$431,670.56	\$66
31					H66995	0	2	2	2.1	3	2	11.1	CLP	\$659.29	\$38,889.24	\$58.99	\$431,670.56	\$66
32					H66995	0	2	2	2.1	2	2	10.1	CLP	\$667.78	\$38,889.24	\$58.24	\$392,781.32	\$58
33					H66995	0	2	2	2.1	2	2	10.1	CLP	\$667.78	\$38,889.24	\$58.24	\$392,781.32	\$58
34					H66995	0	0	0.4	2.1	2	2	6.1	CLP	\$667.78	\$38,889.24	\$58.24	\$237,224.36	\$38
35					H66995	0	2	2	2.1	2	0	8.1	CLP	\$663.37	\$38,889.24	\$58.62	\$315,002.84	\$41
36					H66995	0	0	3	2.1	2	2	9.1	CLP	\$663.37	\$38,889.24	\$58.62	\$363,892.08	\$51
37					H66995	0	2	3	2.1	2	2	11.1	CLP	\$663.37	\$38,889.24	\$58.62	\$431,670.56	\$66
38					H66995	0	3	2	3	3	3	14	CLP	\$683.39	\$38,889.24	\$56.91	\$544,449.36	\$76

Figura 4.1: Pestaña del reporte con detalle en columnas. Fuente: Elaboración propia.

4.4. Flujo de información

Luego de entender la fuente y naturaleza de los datos a utilizar, se procede a analizar el flujo de información que se sigue actualmente para obtener los datos.

Consultando el origen de los datos, se puede ver que estos son generados por IBM en Costa Rica, lugar designado para generar estos reportes a toda la región de Latinoamérica. Entendiendo que estos datos presentan un carácter sensible (dado que involucran datos personales asociados a rangos de sueldos) no cualquiera puede acceder a dicha información, es por esto que solo personal autorizado en Costa Rica puede ingresar a la base de datos de Claim y generar estos reportes mediante Macros, que luego serán subidos a un repositorio (BOX) para ser distribuidos y utilizados por quienes los necesiten previa autorización (En este caso los DPE y sus equipos). La figura 4.2 ilustra el proceso actual.

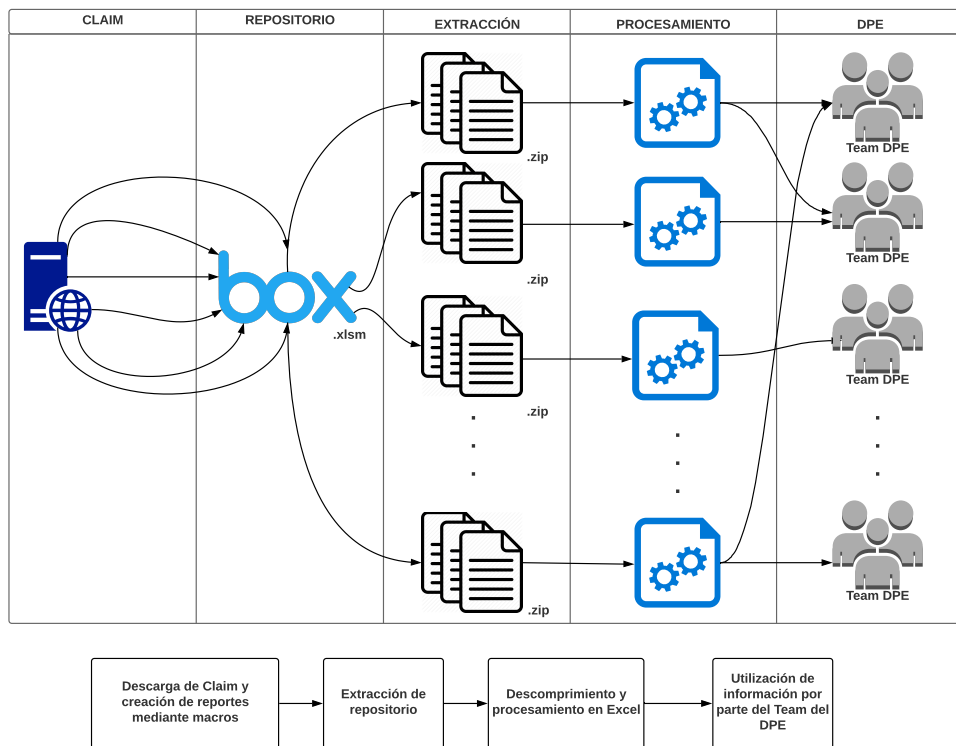


Figura 4.2: Flujo de trabajo inicial. Fuente: Elaboración propia.

Capítulo 5

Diseño de solución

En este capítulo se detalla la solución a implementar, junto con las figuras de mérito escogidas para realizar una posterior evaluación.

5.1. Arquitectura de plataforma

Considerando el flujo de información mostrado en 4.4, la figura 5.1 ilustra el proceso tentativo a implementar, en donde el objetivo principal es realizar la tarea que ejecuta cada DPE con sus respectivas cuentas, en un solo proceso para optimizar y estandarizar el trabajo.

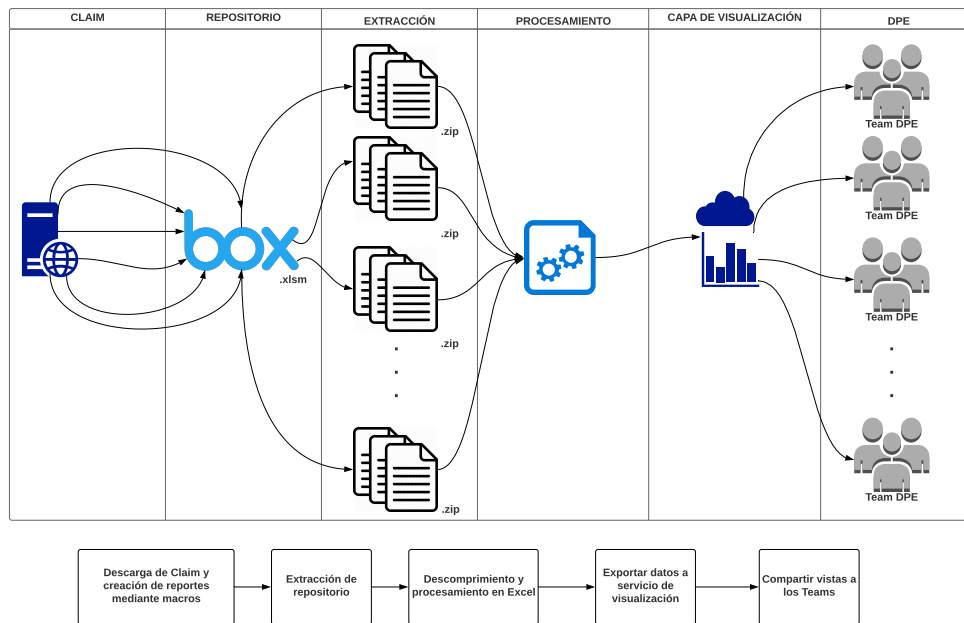


Figura 5.1: Primera aproximación de flujo de trabajo. Fuente: Elaboración propia.

Llevando el esquema anterior a lo que sería la arquitectura, se propone el diseño expuesto

en la figura 5.2, en donde se tiene que:

1. Mediante un *host* local y un *script* en Python se procede a recopilar la información relevante para generar el proceso ETL.
2. Toda la información recopilada y procesada es cargada en diferentes tablas, según ejecutivo, en una base de datos montada en la nube.
3. Las distintas tablas son conectadas a algún servicio de visualización que permita generar vistas para los DPE y sus equipos.
4. Considerando el acceso mediante VPN como método de seguridad, se procede a desplegar y permitir el acceso a las visualizaciones según corresponda.

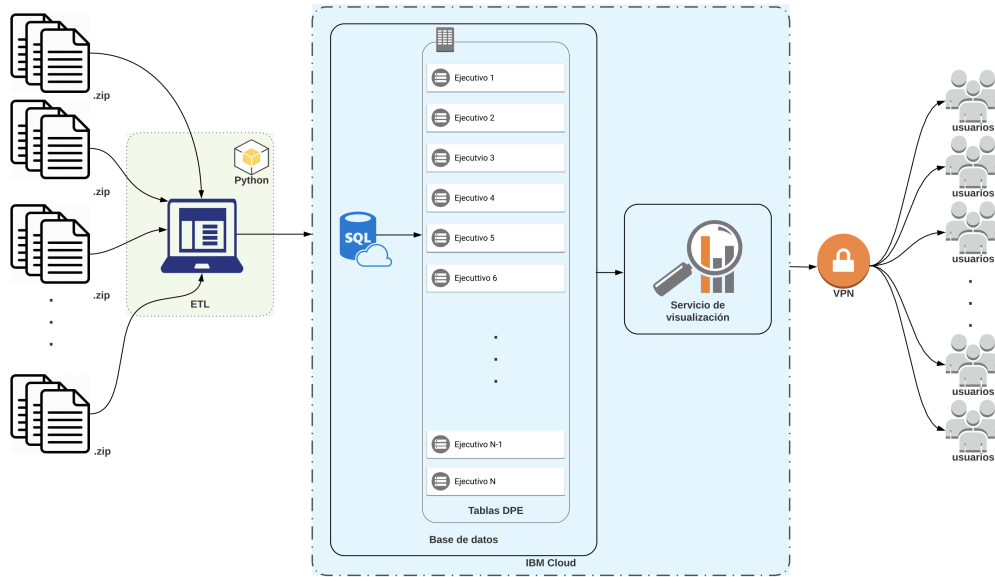


Figura 5.2: Diseño de arquitectura. Fuente: Elaboración propia.

5.2. Métricas de evaluación de desempeño

Para evaluar la correcta implementación y utilización de esta plataforma, se medirá la cantidad y frecuencia con que los DPE ingresan a sus respectivas vistas. Para realizar esta medición se utilizará un acortador de URL que se adiciona en la salida del servicio de visualización mostrado en 5.2, con el fin de entregar un enlace fácil de memorizar y que además permita monitorear su uso.

Otra métrica de desempeño será la retroalimentación constante brindada por parte de los DPE, ya que al trabajar en conjunto con ellos en el desarrollo se estará recibiendo constantemente la opinión y sugerencias respecto a la experiencia en la plataforma.

Adicionalmente, se medirán los tiempos de ejecución de la plataforma, en cuanto al despliegue de los datos, frecuencia de actualización y procesamiento de los datos, con el fin de realizar una comparación con los tiempos que les tomaba a los DPE realizar un análisis similar sin la plataforma.

Capítulo 6

Análisis Exploratorio de los datos

Antes de adentrarse en el desarrollo de la plataforma, es necesario analizar y entender la información con la que se va a trabajar. Para esto, se procede a descargar todos los informes existentes desde que se empezó a recopilar la información hasta la fecha (2016 al 2018), con el fin de ver en que estado llega la data bruta, con lo que se obtiene una gran tabla de 70 x 291123 con distintos valores y significados.

Continuando con la exploración, se identifica la cantidad de valores en blanco que presenta el dataset, tal y como muestra la figura 6.1. De aquí es posible ver hay columnas que se encuentran totalmente vacías, y por otro lado que todas las columnas presentan datos vacíos en mayor o menor medida.

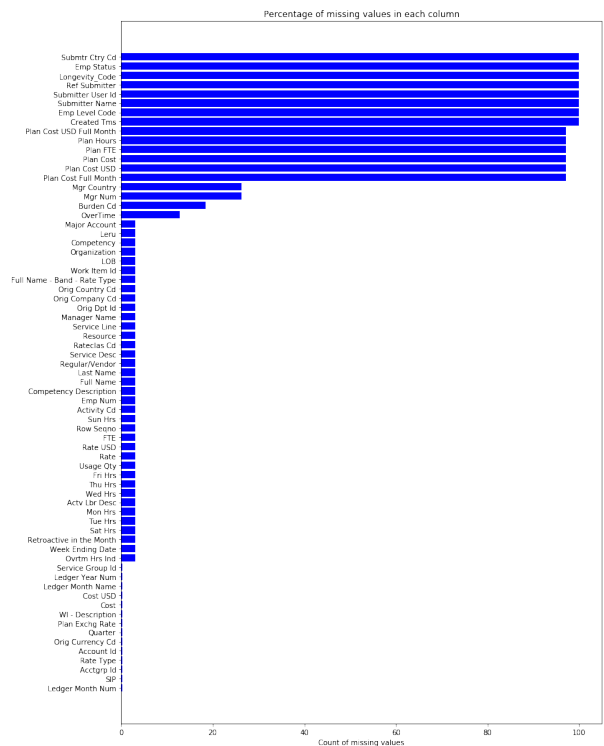


Figura 6.1: Porcentaje de valores vacíos por columna. Fuente: Elaboración propia.

Un análisis más exhaustivo permite dar a conocer que la información bruta que se recibe y lee en los archivos .xslm, contiene una gran cantidad de filas vacías, las cuales son parte del formato del informe del archivo. Dado que para efectos prácticos no interesan dichas filas, se eliminan sin mayores problemas, reduciendo la cantidad de columnas con datos parcialmente vacíos a 4 y columnas completamente vacías a 14, tal y como muestra la tabla 6.1.

Column	Total	percentage
Plan Cost USD Full Month	282218	100.00
Emp Status	282218	100.00
Submitter Name	282218	100.00
Submitter User Id	282218	100.00
Submtr Ctry Cd	282218	100.00
Ref Submitter	282218	100.00
Longevity_Code	282218	100.00
Emp Level Code	282218	100.00
Created Tms	282218	100.00
Plan Hours	282218	100.00
Plan FTE	282218	100.00
Plan Cost Full Month	282218	100.00
Plan Cost USD	282218	100.00
Plan Cost	282218	100.00
Mgr Num	67626	23.96
Mgr Country	67626	23.96
Burden Cd	44550	15.79
OverTime	28318	10.03

Tabla 6.1: Columnas con datos vacíos. Fuente: Elaboración propia.

Siguiendo con la inspección, se eliminan todas las columnas que poseen todos sus datos en blanco, dado que no aportan información relevante al problema, obteniendo solo 4 columnas que aún se mantienen con datos parcialmente en blanco, tal y como muestra la tabla 6.2.

Column	Total	Percent
Mgr Country	67626	23.96
Mgr Num	67626	23.96
Overtime	28318	10.03
Burden Cd	44550	15.79

Tabla 6.2: Columnas con datos vacíos luego del filtro. Fuente: Elaboración propia.

Resumiendo el tipo de datos que contiene el set de datos generados, se obtiene que hay 35 tipo *string*, 1 *date* y 20 *float*, tal y como muestra la tabla 6.3.

Tipo	Cantidad
String	35
datetime64[ns]	1
Float64	20

Tabla 6.3: Tipos de variables. Fuente: Elaboración propia.

Para proceder con las columnas relevantes antes de incluir las reglas de negocio propias del problema a resolver, se analiza la correlación entre columnas, con el fin de encontrar posible información redundante. El resultado se puede ver en la figura 6.2, dando cuenta que efectivamente hay información fuertemente correlacionada y por tanto podría eliminarse.

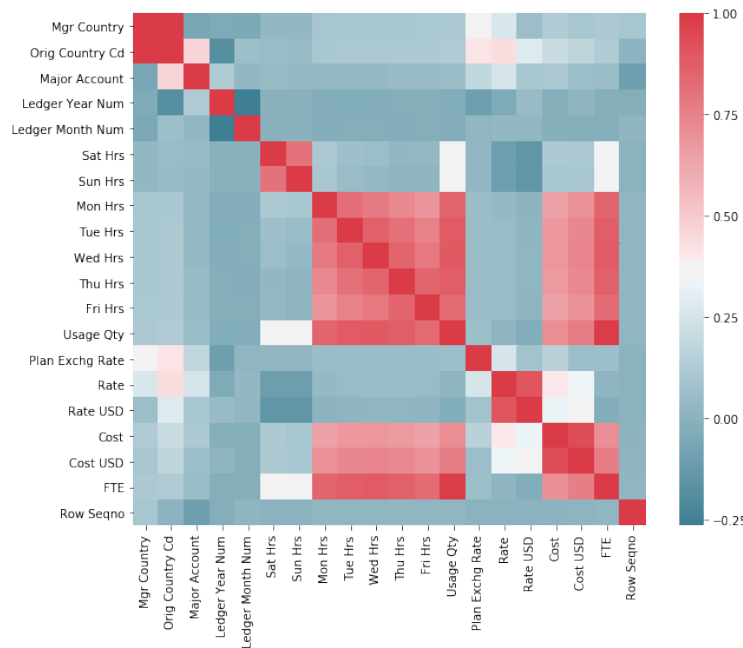


Figura 6.2: Correlación entre columnas numéricas. Fuente: Elaboración propia.

De las columnas restantes, se pueden identificar las mencionadas en el capítulo de antecedentes 4.3, donde aún se mantienen las que contienen mayor información para la plataforma de gestión.

Con los datos medianamente limpios, se procede a verificar cuántos clientes y contratos existen en este archivo histórico. En la tabla A.1 es posible ver un resumen de los datos considerando la cantidad de fuentes de información o reportes con distinto código, cantidad de AI¹, fecha de inicio del contrato, la última fecha donde hubo un movimiento asociado a ese contrato y la duración en semanas.

Esta tabla permite ver que hay 125 fuentes distintas o informes con código diferente, dando a entender que existen 125 contratos. Para corroborar esto, se procede a generar una vista similar, pero ahora por SIP, con lo que se obtiene la tabla A.2, en donde la cantidad de

¹Recordar que un cliente puede tener múltiples contratos y que a su vez, un contrato puede tener múltiples AI o Account ID

registro se reduce a 107, lo cual se puede atribuir en parte a que muchos informes no cuentan con la información del SIP.

Para corregir esto se extrae el código del nombre del informe y se reemplaza en los datos que faltan, con lo cual se obtiene la tabla A.3, que muestra los registros corregidos y la cantidad de registros o contratos (116), dando a entender que hay diferentes fuentes de información con códigos distintos pero que tienen un mismo SIP, y por tanto corresponden al mismo contrato.

El detalle de estas tablas se puede encontrar en el anexo A.

Capítulo 7

Implementación

En este capítulo se expondrá detalladamente la implementación de la plataforma desde su concepción, pasando por el desarrollo, para terminar con la operatividad de la misma. Para estructurar y organizar la información, se separará el desarrollo en 3 hitos, que contemplan entregas parciales de la plataforma en donde al finalizar cada iteración, se redefinen los objetivos y pasos a seguir.

7.1. *Minimum Viable Product* (MVP)

En esta primera iteración de la plataforma, se busca crear un primer MVP funcional para dar a conocer el objetivo al cual se quiere llegar, analizar la recepción por parte de los usuarios y los requisitos necesarios para lograr la producción de la herramienta.

7.1.1. Preprocesamiento y limpieza de datos

Para poder utilizar estos datos, primero deben ser trabajados, ya que como se vio en la sección 6, se reciben datos en bruto (informes en Excel) a los cuales se les debe dar significado para su utilización.

En primer lugar, se procede a repetir la etapa del análisis exploratorio en donde se eliminan columnas y datos vacíos, asimismo, se procede a eliminar las columnas que no aporten información al problema. Seguido de esto, se procede a limpiar y dar formato a los datos siguiendo reglas de negocios particulares del problema a resolver. Aquí se consideran:

- Fechas: Debido a que IBM trabaja por años contables, la información importante a considerar corresponde a la del año en curso, es decir, solo información del año 2018. Por tanto, información relacionada de años anteriores debe ser descartada.
- Overtime: Luego del primer proceso de limpieza, se identifica que el 10.03% de los datos de la columna 'Overtime' se encuentran vacíos. Lo cual, según reglas de negocio

internas de IBM, deben ser reemplazados por horas tipo ‘normal’.

- Eliminar duplicados: Una porción de los clientes cuentan con servicios tanto en IBM Chile como en IBM India. Los recursos utilizados en India deben ser cargados de igual forma a las cuentas aquí en Chile, para esto, se crean ‘centros de costos espejo’ en India, los cuales tienen el mismo AI que el centro de costo original en Chile para poder hacer el traspaso posteriormente a pesos chilenos. Este proceso se ve reflejado en los informes semanales duplicando la información debido a que aparece los recursos utilizados en India y en Chile, con la diferencia de que en la columna ‘Orig Company Cd’ se indica el país de origen. Aquí deben ser considerados los valores cargados en India, eliminando los otros para no reflejar los valores duplicados en las cifras.
- Agregar fechas contables, las cuales difieren de las reales.
- Agregar el nombre del archivo como columna.
- Realizar el cruce ejecutivos-clientes para estructurar y separar los distintos informes por DPE.

Este proceso en primera instancia se realiza en conjunto con personal de IBM especialista en el tema (tanto de Chile como de Costa Rica) para ajustar y afinar las reglas de limpieza a seguir. Esto debido a que cualquier error en estos datos, puede implicar grandes pérdidas y descuadramientos en las cuentas de los DPE al momento de tomar decisiones.

Cabe destacar que este proceso se realiza de forma manual durante 3 semanas para ajustar día a día mediante prueba-error y comparación con datos oficiales para no desviarse bajo ningún aspecto.

7.1.2. Implementación en Watson Analytics

Utilizando los servicios propios de IBM, se decide utilizar en primera instancia Watson Analytics, donde se cargan los datos en formato .csv (previamente procesados) para luego ser filtrados según ejecutivo, generando así 18 visualizaciones con información relevante.

Aquí, se procede a crear distintas vistas con los datos según los requerimientos y posible información relevante para que los ejecutivos puedan tomar decisiones.

Antes de empezar a graficar, se procede a estructurar la vista para cada uno de ellos como **Histórica - Detalle de horas -Distribución de Bandas**, en donde la primera proporciona un resumen general en cuanto a horas cargadas por los especialistas y funcionarios asociados a dicha cuenta y también sus costos; la segunda entrega un detalle completo con información relacionada únicamente a FTE u horas cargadas por empleado, permitiendo un detalle completo de como se comportan las horas cargadas; y finalmente la vista de costos, la cual es un complemento similar a la de FTE pero para los costos asociados a las cuentas.

En esta etapa se trabaja fuertemente con los ejecutivos para entender sus necesidades con el fin de generar vistas útiles para ellos. Dentro de los aspectos útiles y diarios para los cuales ellos necesitan esta información se encuentran:

- Evolución de costos y FTE para evaluar la realización de cambios en las cláusulas y

cantidad de recursos en caso de ser necesario.

- Poder llevar un control semana a semana de las horas efectivas que se están cargando a las distintas cuentas/contratos/AI, permitiendo monitorear que no aparezcan horas registradas no trabajadas.
- Controlar los excesos de horas *overtime*, dado que la inspección del trabajo permite un límite mensual que no debe ser superado por ningún empleado no sujeto al artículo 22[41].
- Control de costos y FTE en general para los distintos aspectos de una cuenta, comparar con el caso de negocios y lo previsto para efectuar medidas en caso de ser necesario.

Considerando los puntos anteriores se generan distintas visualizaciones tentativas para cada uno de los DPE. En las figuras 7.1 a la 7.9 es posible ver los distintos gráficos generados que permiten -en conjunto- abarcar cada uno de los puntos mencionados anteriormente. Para un mayor detalle de los gráficos generados ver anexo B y C.1.

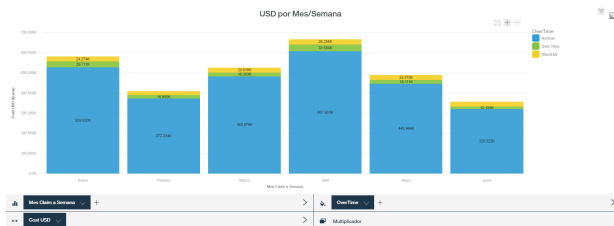


Figura 7.1: Vista histórica de Costos en USD por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.

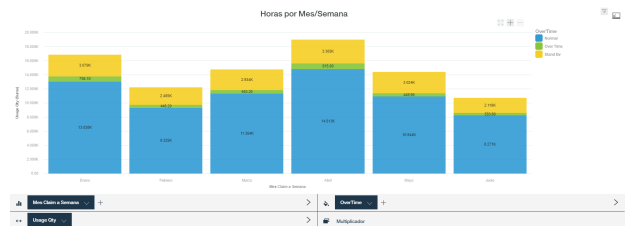


Figura 7.2: Vista histórica de horas por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.

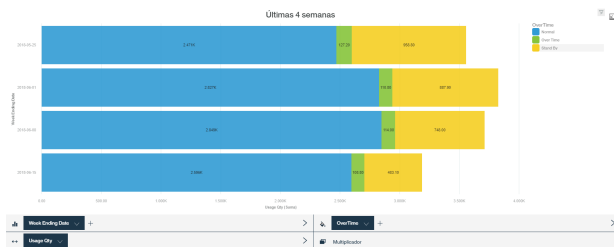


Figura 7.3: Vista costos en USD de últimas 4 semanas con distinción por tipo de hora. Fuente: Elaboración propia.

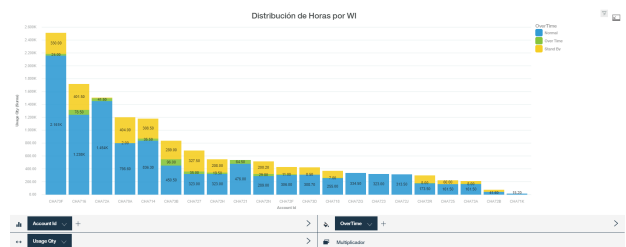


Figura 7.4: Costos últimas 4 semanas por Account ID y con diferenciación en tipo de hora. Fuente: Elaboración propia.

Es importante destacar que esta plataforma permite la interacción entre distintos gráficos, permitiendo seleccionar uno y actualizar los demás en base a la elección, o permitir distintas vistas con bajadas y subidas de lo general a lo particular (vista a nivel cliente ->contrato ->AI ->empleado).

7.1.3. Distribución de los datos

Para la distribución de datos, se escoge un método tradicional: El correo electrónico. Se decide enviar semanalmente -luego de la actualización de la plataforma- un correo avisando

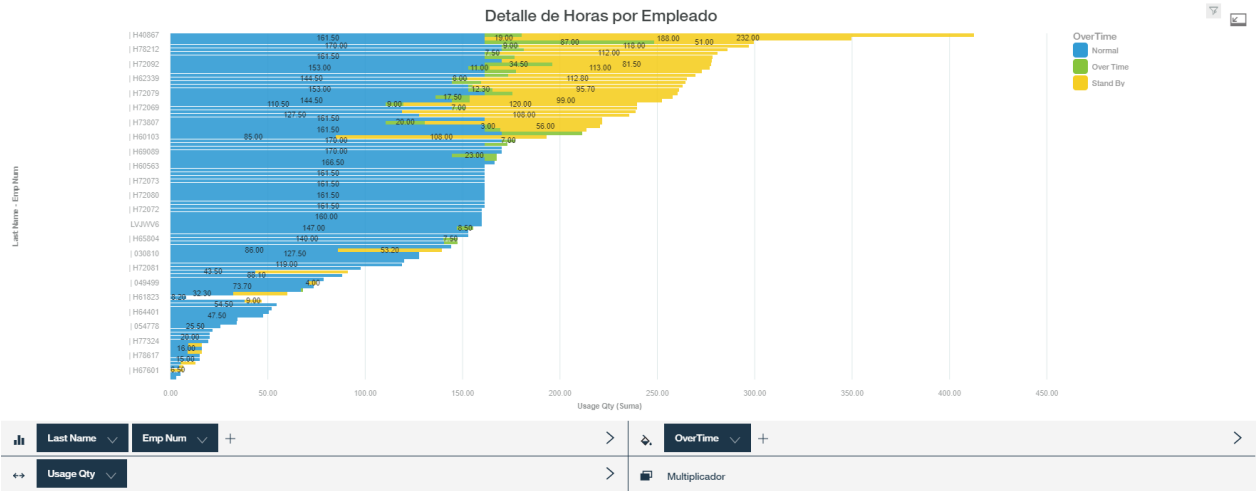


Figura 7.5: Detalle de horas cargadas por empleado, haciendo la distinción por tipo de hora cargada. Fuente: Elaboración propia.

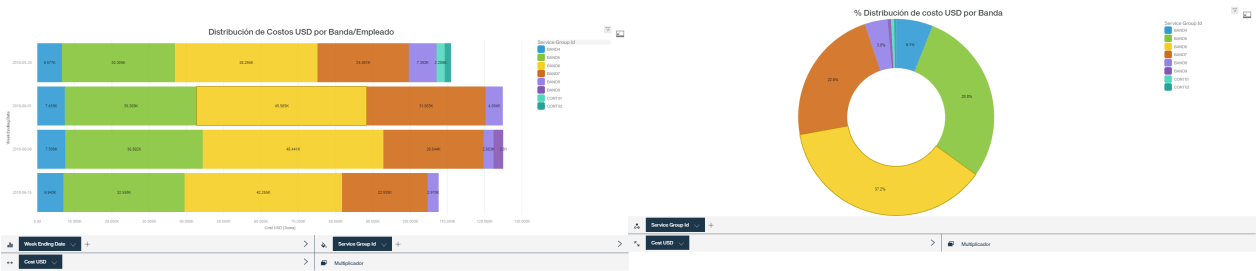


Figura 7.6: Distribución de costos en USD por Banda/Empleado. Fuente: Elaboración propia.

Figura 7.7: Porcentaje de distribución de costo en USD por banda. Fuente: Elaboración propia.



Figura 7.8: Mapa de relación hora y costo. Fuente: Elaboración propia.

Figura 7.9: Costo en USD por Work Item ID (servicio). Fuente: Elaboración propia.

de la actualización en conjunto con los respectivos enlaces para el ingreso. Aquí es importante destacar que el ingreso a las vistas está restringido por permisos además del requisito de estar dentro de la red de IBM o en su defecto VPN. Por lo tanto, cada vez que un miembro nuevo del equipo de algún ejecutivo quiere acceder a los datos se debe comunicar para su previa autorización.

7.1.4. Resultados Preliminares y Análisis

Luego de las primeras entregas mediante este formato, se recogen opiniones y experiencias de uso por parte de los usuarios. Es aquí donde se obtienen las métricas e información relevante para reorientar y planificar la siguiente etapa dentro de la plataforma.

En conversaciones con los usuarios se extrae que la visualización en Watson Analytics resulta lenta, en promedio tarda 8,3 segundos en cargar la información y al momento de cambiar o interactuar con los gráficos el sistema vuelve a cargar (8-10 segundos de latencia) o arroja un error solicitando recargar la página. También, se detecta un problema en cuanto a compatibilidad con los distintos navegadores y tamaños de pantalla, dado que en algunos funciona mejor que en otros y dependiendo de la resolución, la plataforma puede desconfigurarse ya que el sistema no se desarrolló de forma responsiva.

7.2. Plataforma V1.0

En base a la retroalimentación obtenida en la iteración anterior, se propone una nueva meta que implica modificar la plataforma por completo. En esta nueva meta se contemplan los siguientes aspectos:

- Disminuir latencia de la plataforma.
- Estandarizar y automatizar el proceso ETL que se genera (extracción de datos; procesamiento de datos y carga a la plataforma destinada a visualización).
- Minimizar el tiempo de actualización.

En donde adicionalmente se itera sobre los objetivos principales, donde se proponen nuevas visualizaciones y posible información adicional para complementar el sistema.

7.2.1. Exploración de alternativas

En base a los nuevos objetivos propuestos, se plantea buscar una alternativa para manejar la latencia y frecuentes caídas del sistema. Para esto, se exploran distintos servicios tanto de IBM como de terceros.

Frente a esto, se llega a 2 soluciones factibles: Watson Studio y Cognos Analytics, ambas soluciones de IBM por lo que interactúan sin problemas en Bluemix.

Evaluando en detalle y comparando ambas plataformas se obtiene la tabla 7.1, en donde se exponen los ventajas y desventajas de cada una pensando en los objetivos y experiencias de usuarios.

De la tabla anterior, se puede obtener como resumen que Cognos Analytics es una plataforma orientada a reportería y despliegue de información, por tanto cumple con todo los

Solución	Ventajas	Desventajas
Cognos Analytics	Orientado a reportaría tradicional	Lenta curva de aprendizaje
	Facilidad en la creación de insights	Entorno cerrado
	Entorno de escritorio y web	Largo proceso de instalación
	Posibilidad de generar reportes interactivos descargables	Autorización requerida para su instalación
	Pipelines completos para reportaría	Disponibilidad de un servidor o espacio en la nube para instalación
Watson Studio	Baja latencia	Orientado a público experto en analítica
	Cuenta con microservicio de Cognos embebido	
	Compatibilidad con herramientas de ML (Python, R, Spark, etc)	Pocos recursos asociados a la creación de insights
	Acceso fuera de la red de IBM	
	Fácil creación del servicio	Pipelines deben ser creados desde cero
	API para insertar dashboard en aplicaciones web	
No requiere de autorización para su utilización		

Tabla 7.1: Comparación Cognos Analytics vs Watson Studio [6][7][8][9][10].

requisitos mínimos para la actividad en cuestión. Por otro lado, Watson Studio se encuentra más enfocada al análisis de datos y modelamiento, sin embargo, cuenta con un microservicio de Cognos embebido que permite utilizar ciertas funcionalidades de Cognos para explotar el análisis. Es este último punto el que se utiliza para inclinarse por Watson Studio, además de los otros beneficios mencionados en la tabla 7.1.

7.2.2. Propuesta de arquitectura en la nube

Dado que en esta etapa se utilizan nuevos servicios, además de considerar la automatización, se propone una arquitectura en la nube que permita flexibilidad, disponibilidad del 100 %, estándares de seguridad, capacidad de automatización y escalabilidad en caso de crecer. Para esto, la arquitectura a considerar se ilustra conceptualmente en la figura 7.10.

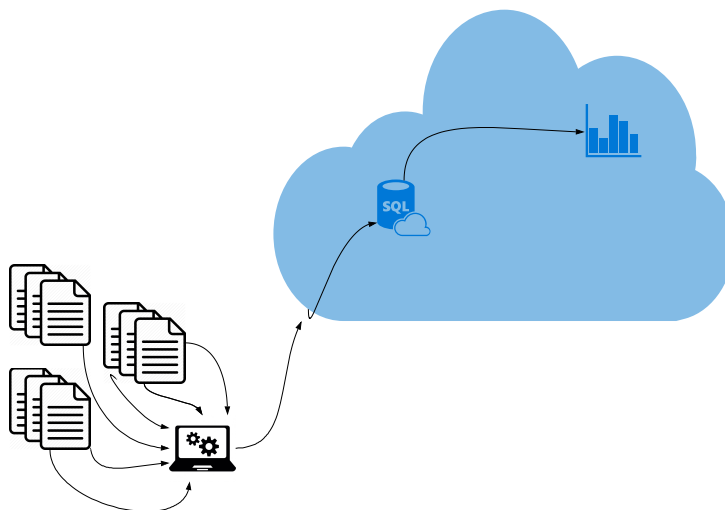


Figura 7.10: Arquitectura en la nube. Fuente: Elaboración propia.

En donde el sistema se compone de un equipo local desde donde se debe ejecutar el proceso de actualización de forma semanal dado que los datos no se pueden obtener de forma automática o desde el origen (DB) por los permisos mencionados en el capítulo 4.4, luego estos son procesados en la maquina local a través de Python, automatizando el proceso de

limpieza realizado en la primera iteración. Luego de esto los datos son enviados a la nube para continuar el proceso desde ahí, liberando el equipo utilizado para no consumir recursos innecesarios.

En el lado *cloud* se cuenta con una base de datos a definir donde se almacenarán los datos históricos asociados que luego será conectada a Watson Studio, en específico, Cognos Dashboard Embedded para su posterior publicación.

7.2.3. Creación de base de datos

Para el almacenamiento de la data contable, se procede con DB2 Warehouse. Aquí a través de un *script* en Python y utilizando la API de DB2 se crean distintas tablas asociadas a los distintos ejecutivos en donde se consideran 22 columnas con información esencial, las cuales pueden verse en la tabla 7.2.

Columna
Account_Id
Cost_USD
Emp_Num
FTE
Fri_Hrs
Fuente
Full_Name
Last_Name
Manager_Name
Mon_Hrs
Overtime
SIP
Sat_Hrs
Service_Group_Id
Sun_Hrs
Thu_Hrs
Tue_Hrs
Usage_Qty
Wed_Hrs
Week_Ending_Date
Semanas_a_la_Fecha
Mes_de_Claim

Tabla 7.2: Columnas escogidas para incorporar en la visualización. Fuente: Elaboración propia.

7.2.4. Automatización

Utilizando Python como lenguaje base, se genera un código con el objetivo de:

1. Abrir y descomprimir carpetas descargadas de BOX.
2. Abrir carpetas descomprimidas y volver a descomprimirlas hasta llegar a los archivos .xlsm.
3. Integrar todos los archivos .xlsm.
4. Procesamiento de los datos (limpieza y aplicación de reglas de negocios).
5. Separación de contratos según ejecutivo.
6. Limpieza de data en DB.
7. Carga de nuevos datos en DB.

Este proceso se crea pensando en una actualización con periodicidad semanal, considerando la actualización de información de todas las cuentas en el repositorio. El *script* utilizado puede encontrarse en anexos D.

Para medir el tiempo de ejecución del proceso automático se toma el tiempo desde que empieza el proceso hasta que termina la carga en las bases de datos. Es importante destacar que el tiempo aumentará en cada actualización debido a que semana a semana se agregan más datos, pero este no representa tiempo de usuario. El código completo de limpieza puede ver en anexos D.

7.2.5. Implementación en IBM Studio

Utilizando las bases de datos previamente creadas y cargadas, se procede a crear vistas interactivas siguiendo la misma idea realizada con Watson Analytics pero ahora en Watson Studio y Cognos Embedded.

Respecto a la versión anterior, se modifican las pestañas dejando las siguientes 5:

- **Resumen Histórico:** Resumen del año contable con datos históricos de costos en USD, horas cargadas y FTE.
- **Detalle Horas:** Vista detallada con información respecto a horas cargadas para las últimas 4 semanas.
- **Costo USD por Banda:** Vista detallada con información respecto a Costos para las últimas 4 semanas considerando las bandas.
- **FTE por Banda:** Análisis de horas cargadas en FTE pero ahora considerando las bandas.
- **Tutorial:** Vídeo tutorial para introducir el funcionamiento de la plataforma.

Un ejemplo interactivo de la vista creada puede encontrarse en el anexo C.2.

7.2.6. Distribución de los datos

Para facilitar el ingreso a las distintas vistas por partes de los usuarios finales y poder llevar un seguimiento respecto al uso, se procede a utilizar acortadores de URL. En este caso se utiliza IBM BIZ, con el cual se generan 18 nuevos enlaces que poseen la siguiente estructura: **ibm.biz/+ ‘cargo’+‘nombre ejecutivo’**.

7.2.7. Resultados preliminares y análisis

Considerando una segunda ronda de retroalimentación, se recogen nuevamente opiniones y experiencias de uso por parte de los usuarios. Adicionalmente, se les realiza una encuesta para medir la satisfacción y uso, entre otras cosas, donde respondió el 83.3 % de los usuarios¹.

En este segundo proceso, se obtiene por parte de las experiencias que aún hay modificaciones en cuanto a visualizaciones que agregar. Adicionalmente, se solicita por parte de los DPE la integración de clientes internos de IBM, que funcionan bajo el mismo principio, pero que los informes son obtenidos en formato .txt semanalmente dado que el origen es distinto.

En cuanto a la latencia, se logra disminuir a tiempos menores a 2 segundos con el cambio de servicio de visualización, lo cual mejora drásticamente la percepción del usuario respecto a la plataforma.

Para medir la reducción de tiempo que permite esta plataforma, se les consulta directamente su percepción a los usuarios. La figura 7.11 muestra las respuestas obtenidas en donde se puede ver que el 50 % de los consultados² afirma haber reducido entre 1 y 2 horas su tiempo en análisis y gestión, y un 28.6 % dice entre 2 y 3 horas.

¿Cuánto tiempo has reducido con la implementación de esta herramienta? (Comparado con el método de control que hacías antes)

14 respuestas

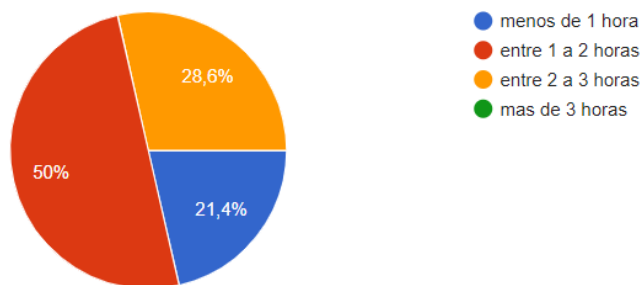


Figura 7.11: Encuesta realizada sobre reducción de tiempo. Fuente: Elaboración propia.

¹18 ejecutivos en total.

²La figura 7.11 muestra las respuestas de las personas que en la pregunta previa incluida en la misma encuesta afirmaron utilizar la plataforma.

Asimismo, se les realiza una pregunta abierta sobre que modificaría/agregaría en la plataforma, tal y como muestra la tabla 7.3.

¿Qué le agregarías/cambiarías en esta plataforma?
“capacidad de analítica y proyección”
“que se pudiera cargar información del BC para hacer la comparación entre BC y Actual”
“tener información de la línea de costo ”
“Información respecto a los tickets generados”
“cruzar información con ticket de especialistas ”
“Agregar Business Cases”
“línea de costo y tener una vista actualizada en tiempo real”
“Pronósticos de costos”
“Pestaña de análisis de ticket, para conocer el rendimiento de cada especialista”
“Business Case par cada cuenta”
“Análisis de incidentes para ver el impacto que este produce ”
“El poder hacer cruces de información para poder determinar incidencias repetitivas y como estas impactan en el costo.”
“Dicha Plataforma debería complementarse con los Business Case de cada DPE”

Tabla 7.3: Preguntas abiertas en encuesta de satisfacción. Fuente: Elaboración propia.

De las respuestas entregadas se desprenden 3 grandes puntos:

- Integración del Business Case.
- Pronóstico de costos y horas cargadas.
- Cruce de información con los incidentes para obtener información en cuanto a desempeño.

Analizando los 3 puntos, viendo las limitaciones, alcances del trabajo y tiempos asociados, se decide evaluar la integración del segundo punto: Pronóstico de costos y horas cargadas en una iteración final.

7.3. Plataforma V2.0

En esta versión final de la plataforma, se consideran cambios a nivel de visualizaciones desplegadas, la integración de cuentas internas importantes para IBM, así como también un módulo de predicción para anticiparse en cuanto a recursos utilizados³.

7.3.1. Reestructuración de arquitectura

Antes de empezar con el desarrollo del módulo de predicción, deben tomarse ciertas consideraciones a nivel de arquitectura para compatibilizar el desarrollo actual con el nuevo y desplazar la mayor cantidad de computo posible a la nube, con el fin de aprovechar los recursos y dejar disponibles los equipos locales. En la figura 8.1 se ilustra conceptualmente la nueva arquitectura a utilizar.

³Si bien este módulo no se considera dentro de los objetivos iniciales, se decide realizar debido al valor agregado que puede aportar.

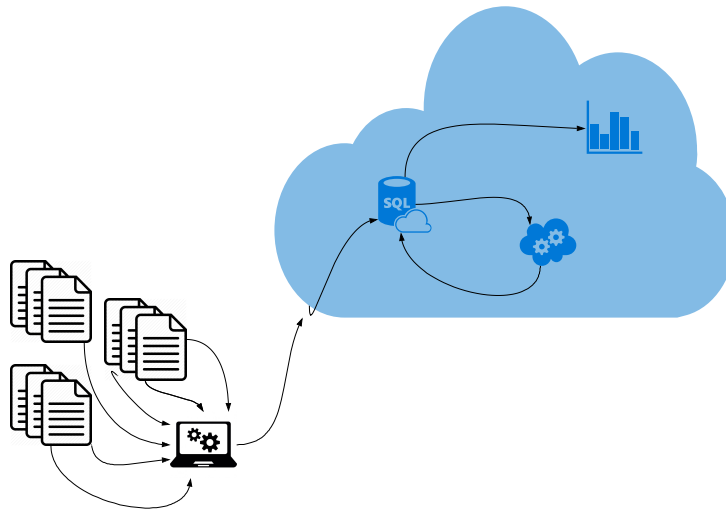


Figura 7.12: Arquitectura final en la nube. Fuente: Elaboración propia.

En esta nueva arquitectura, se propone generar una nueva base de datos tipo *warehouse* donde se almacenan datos históricos desde que se tienen registro (2016), que luego serán utilizados en el modelamiento de los datos, el cual se realizará en la nube en un servidor capaz de ejecutar una rutina en Python de forma automática para realizar la actualización del modelo de forma periódica.

En esta nueva arquitectura el flujo es:

1. Empezar el proceso desde un equipo local en donde se recopilarán los archivos `.xlsm` y `.txt`.
2. El *host* comienza la limpieza de datos y actualiza la base de datos con información del año contable.
3. A su vez, el mismo equipo inserta la nueva información en la base de datos histórica.
4. De forma automática el servidor ejecutará el proceso de modelamiento para alimentarse de la base de datos histórica y así actualizar las predicciones. Para ejecutar este proceso, se utiliza una instancia en la nube con Python 3.5, 16vCPU y 64 GB RAM.
5. Los resultados son desplegados de manera dinámica en Cognos Embedded.

7.3.2. Integración datos en formato de texto

En primer lugar, se procede a la integración de los archivos en formato de texto, siguiendo un procedimiento análogo al de los `.xlsm`, en donde se extrae la información relevante del archivo de texto para luego ser enviado a la base de datos.

En la figura 7.13 se ilustra un ejemplo de cómo son los reportes en formato `.txt`.

Como se aprecia en la imagen, este reporte contiene menos columnas y por tanto, menos

COUN CODE	WEEK ENDING DATE	EMP SER NUM	WORK ITEM ID	EMP LAST NAME	EMP INI	ACCOUNT ID	OVERTIME IND	SUM		COMPONENT ID	ACTIVITY CD	SUBMITT USER ID	EMP DEPT NUMBER
								HRS EXPENDE	TOTAL				
655	2018-11-02	H74679	8BI05CG6	RC	J	CIAAAX	Y	5.3	DI		GB0020	H74679	00HDS5
655	2018-11-02	H74679	8BI05CG6	RC	J	CIAAAX	S	2.0	DI		GB0020	H74679	00HDS5
655	2018-11-02	H74679	8BI05CG6	RC	J	CIAAAX	N	4.5	DI		GB0020	H74679	00HDS5
655	2018-10-26	H66687	8BI05B91	GC	D	CIAAAW	S	2.5	GSNI -IRM		GB0020	H66687	00623G
655	2018-10-26	H66687	8BI05B91	GC	D	CIAAAW	N	2.5	GSNI -IRM		GB0020	H66687	00623G
655	2018-10-26	H76778	8BI05B91	CH	JE	CIAAAW	N	2.5	GSNI		GB0020	H76778	001165
655	2018-10-26	H63826	8BI05B91	SU	RA	CIAAAW	Y	0.7	GSNI		GB0020	H63826	000857
655	2018-10-26	H63826	8BI05B91	CH	RA	CIAAAW	N	0.5	GSNI		GB0020	H63826	000857
655	2018-10-26	H40563	8BI05B91	GC	HE	CIAAAW	Y	2.0	General Billable		GB0020	H40563	000857
655	2018-10-26	H40563	8BI05B91	GC	HE	CIAAAW	N	5.0	General Billable		GB0020	H40563	000857
655	2018-10-26	H67649	8BI05B91	HE	A	CIAAAW	N	0.5	General Billable		GB0020	H67649	000607
655	2018-10-26	H78617	8BI05B91	VI	L	CIAAAW	N	0.5	GSNI		GB0020	H78617	000857
655	2018-10-26	033157	8BI05B91	SA	JA	CIAAAW	N	1.5	GSNI		GB0020	033157	000607
655	2018-10-26	082997	8BI05CG6	GA	FJ	CIAAAX	S	1.0	General Billable		GB0020	082997	000857
655	2018-10-26	082997	8BI05CG6	GA	FJ	CIAAAX	N	-5.5	General Billable		GB0020	082997	000857
655	2018-10-26	082997	8BI05CG6	GA	FJ	CIAAAX	Y	3.0	General Billable		GB0020	082997	000857
655	2018-10-26	082997	8BI05CG6	GA	FJ	CIAAAX	N	2.5	General Billable		GB0020	082997	000857
655	2018-10-26	082997	8BI05CG6	GA	FJ	CIAAAX	S	8.0	General Billable		GB0020	082997	000857
655	2018-10-26	H40885	8BI05CG6	AL	BA	CIAAAX	N	-13.0	DI		GB0020	H40885	00HDD5
655	2018-10-26	H40885	8BI05CG6	AL	BA	CIAAAX	Y	0.9	DI		GB0020	H40885	00HDD5

Figura 7.13: Reporte formato txt. Fuente: Elaboración propia.

información, ya que no se consideran temas relacionados a costos.

7.3.3. Predicción

Tal y como se mencionó anteriormente, se procede a la creación del módulo de predicción de costos y horas cargadas para un AI. A continuación, se detalla las consideraciones a tomar antes de empezar y los distintos modelos a utilizar.

7.3.3.1. Consideraciones

Antes de comenzar con la generación de modelos, es importante definir los alcances del módulo piloto a desarrollar. Puesto que cada ejecutivo debe velar por los costos de cada AI en cada contrato, y como se ve en la tabla A.1 en anexos, cada contrato o SIP puede llegar a tener más de 40 AI **activos**, y si por cada AI hay que generar 2 modelos (Costos y Horas) la cantidad de predicciones a generar supera con creces los 500, lo cual escapa del alcance si se considera además que cada Account ID presenta un comportamiento distinto y por tanto los modelos no se pueden reutilizar.

Es en base a esto, que se decide privilegiar a los DPE que más utilización presentan según las métricas de ingreso presentes en los acertadores de enlaces, se seleccionan los 4 DPE que más utilizan la plataforma para la realización de los modelos, en donde se incluyen aproximadamente 34 AI (los cuales pueden ir variando en cada actualización). En la tabla 7.4 es posible ver los códigos de los SIP seleccionados al 02/11/2018.

Indice	DPE	SIP	Account ID	First Date	Last Date	Num of Weeks
0	m_XZZ	CLNMTA_XXXXXX	CHA72A	2016-11-04	2018-11-02	104
1	m_XZZ	CLNMTA_XXXXXX	CHA716	2016-11-04	2018-11-02	104
2	m_XZZ	CLNMTA_XXXXXX	CHA714	2016-11-04	2018-11-02	104
3	m_XZZ	CLNMTA_XXXXXX	CHA71K	2017-09-29	2018-11-02	57
4	m_XZZ	CLNMTA_XXXXXX	CHA73F	2016-11-04	2018-11-02	104
5	m_XZZ	CLNMTA_XXXXXX	CHA727	2016-11-11	2018-11-02	103
6	m_XZZ	CLNMTA_XXXXXX	CHA72J	2016-11-11	2018-11-02	103

Tabla 7.4 continuación de la página anterior

Indice	DPE	SIP	Account ID	First Date	Last Date	Num of Weeks
7	m_XZZ	CLNMTA_XXXXXX	CHA72B	2016-11-18	2018-11-02	102
8	m_XZZ	CLAEAL_XXXXXX	CHA7ZG	2016-11-04	2018-11-02	104
9	m_XZZ	CLG7KO_XXXXXX	CHA70A	2016-11-04	2018-11-02	104
10	m_XZZ	CLNMTA_XXXXXX	CHA73B	2016-11-04	2018-11-02	104
11	m_XZZ	CLNMTA_XXXXXX	CHA73D	2016-11-04	2018-11-02	104
12	m_XZZ	CLNMTA_XXXXXX	CHA718	2016-11-04	2018-11-02	104
13	m_XZZ	CLNMTA_XXXXXX	CHA723	2016-11-04	2018-11-02	104
14	m_XZZ	CLNMTA_XXXXXX	CHA72F	2016-11-11	2018-11-02	103
15	m_XZZ	CLNMTA_XXXXXX	CHA721	2016-11-11	2018-11-02	103
16	m_XZZ	CLNMTA_XXXXXX	CHA72R	2016-11-11	2018-11-02	103
17	m_XZZ	CLNMTA_XXXXXX	CHA72H	2016-11-11	2018-11-02	103
18	m_XZZ	CLNMTA_XXXXXX	CHA725	2016-11-11	2018-11-02	103
19	m_XZZ	CLNMTA_XXXXXX	CHA72N	2016-11-11	2018-11-02	103
20	m_XZZ	CLNMTA_XXXXXX	CHA73A	2017-12-29	2018-11-02	44
21	j_XZZ	IO9383_XXXXXX	CHA4KW	2016-01-01	2018-11-02	148
22	j_XZZ	IO9383_XXXXXX	CHA4KE	2016-01-01	2018-11-02	148
23	j_XZZ	CL0540_XXXXXX	CHA7GL	2017-03-31	2018-11-02	83
24	j_XZZ	CL0540_XXXXXX	CHA7G3	2016-12-30	2018-11-02	96
25	f_XZZ	CHE590_XXXXXX	CHAQSU	2016-01-01	2018-11-02	148
26	f_XZZ	CHE590_XXXXXX	CHAQST	2016-01-01	2018-11-02	148
27	f_XZZ	CHE590_XXXXXX	CHAQSW	2016-01-01	2018-11-02	148
28	f_XZZ	CHE590_XXXXXX	CHA1HM	2016-01-01	2018-11-02	148
29	f_XZZ	IP8352_XXXXXX	CHA87E	2017-07-07	2018-11-02	69
30	f_XZZ	IP8352_XXXXXX	CHA88K	2017-07-07	2018-11-02	69
31	f_XZZ	IP8352_XXXXXX	CHA9EK	2017-09-15	2018-11-02	59
32	f_XZZ	IP8352_XXXXXX	CHA8U3	2017-08-04	2018-11-02	65
33	f_XZZ	IP8352_XXXXXX	CHA8GG	2018-05-18	2018-11-02	24
34	f_XZZ	IP8352_XXXXXX	CHA967	2018-03-09	2018-11-02	34
35	f_XZZ	IP8352_XXXXXX	CHBAGJ	2018-06-22	2018-11-02	19
36	f_XZZ	IP8352_XXXXXX	CHA97H	2018-03-02	2018-11-02	35
37	f_XZZ	IP8352_XXXXXX	CHA76K	2018-06-15	2018-11-02	20
38	f_XZZ	IP8352_XXXXXX	CHA6US	2018-06-15	2018-11-02	20
39	f_XZZ	IP8352_XXXXXX	CHBAGK	2018-10-12	2018-11-02	3
40	p_XZZ	IO9500_XXXXXX	CHA435	2016-01-01	2018-11-02	148
41	p_XZZ	IO9500_XXXXXX	CHBADW	2018-05-04	2018-11-02	26
42	p_XZZ	IO9500_XXXXXX	CHBADZ	2018-09-14	2018-11-02	7
43	p_XZZ	IO9500_XXXXXX	CHBAD1	2018-05-11	2018-11-02	25
44	p_XZZ	IO9500_XXXXXX	CHBADS	2018-08-03	2018-11-02	13

Tabla 7.4: *Account ID* activos seleccionados para las predicciones al 02/11/2018. Fuente: Elaboración propia.

7.3.3.2. Descripción de Series de Tiempo.

Una vez seleccionados los datos a utilizar, se procede a analizar las series de tiempo, para evaluar su comportamiento y determinar que modelo es conveniente utilizar.

Para obtener series útiles para la predicción, se suman todas las horas y costos cargados por los empleados dentro de una misma semana, para así obtener una sola fila que represente el valor neto de esa semana para ese AI. En la figura 7.14 es posible ver un ejemplo de serie de tiempo para los costos de un AI específico.

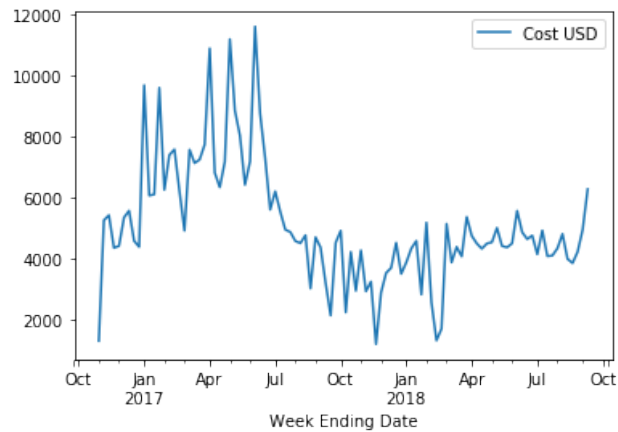


Figura 7.14: Ejemplo de costos para un Account ID en específico. Fuente: Elaboración propia.

Realizando la separación por tendencia y estacionalidad, se obtiene la figura 7.15, la cual permite ver que esta serie en particular presenta ciclos de 4 semanas.

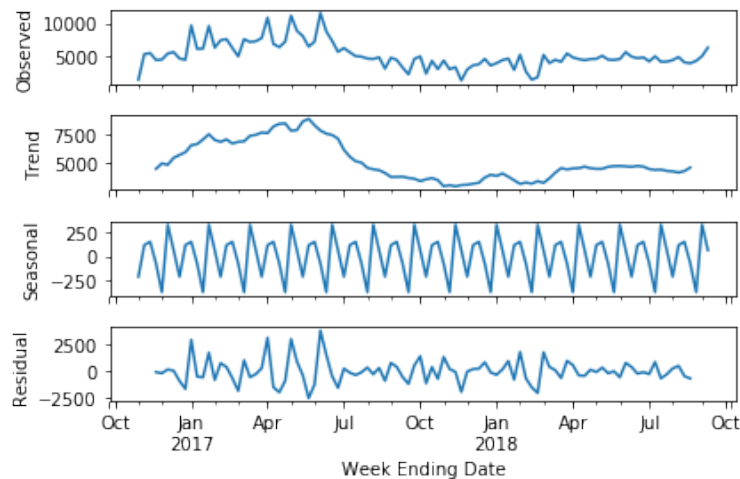


Figura 7.15: Descomposición de serie de tiempo. Fuente: Elaboración propia.

Realizada esta etapa con todas las series de tiempo tanto de horas como de costos, se decide eliminar el primer mes, esto debido a que como se ven en las figuras 7.14 y 7.15, el comportamiento inicial no es representativo de la serie, y por tanto, puede influir de forma negativa en el modelamiento.

7.3.3.3. Modelos a utilizar

Se prueba con distintos algoritmos de predicción y sus distintos parámetros:

- **Proceso Auto Regresivo Integral de Media Móvil Estacional:** De la descomposición de la serie de tiempo realizada en 7.3.3.2 se ve que la serie presenta una componente estacional, por tanto se decide utilizar SARIMA como uno de los métodos de predicción mediante la librería *auto-arima* en Python y una malla de parámetros p,d,q,P,D,Q a utilizar para encontrar el mejor modelo en cada AI.
- **Long Short Term Memory (Redes neuronales recurrentes):** Considerando la cantidad de datos históricos, se procede a modelar con redes neuronales LSTM usando las configuraciones de la tabla 7.5. En este proceso, se debe agregar una normalización a los datos y una redimensión de las series de tiempo para adaptarse a las configuraciones propias de una red recurrente.

Versión	Cantidad de capas	Núm. de nodos	Núm. de épocas	Tam. de batch	Fun. activación	dropout
1	2	4	100	1	linear	0
2	1	4	10000	1	linear	0
3	1	4	400	1	tanh	0
4	2	4	100	1	linear	0.4

Tabla 7.5: Configuraciones para redes LSTM. Fuente: Elaboración propia.

Puntos importantes a tener en cuenta en el modelamiento mediante LSTM son:

- **Ventana de tiempo:** Se considera una ventana de tiempo de 4 semanas para realizar la predicción, esto debido a que más allá de eso los valores pierden representatividad.
- **Métrica a maximizar:** Para escoger el mejor predictor, se utiliza principalmente la métrica *Mean Absolute Percentage Error* (MAPE) junto con *Root Mean Square Error* (RMSE) y *Mean Square Error* (MSE). Aquí, se tomará el promedio de estas métricas teniendo en cuenta que se generarán modelos para cada AI.
- **Cantidad mínima de datos:** Considerando que no se puede realizar una predicción asertiva sin datos históricos, se decide no modelar AI que presenten menos de 16 semanas de información.
- **Última Fecha de Actualización:** Se considerará como **cerrado** un AI si en la última actualización no se cargaron horas ni costos. En caso de estar cerrado, se excluirá de la predicción.

Los resultados finales de este módulo y la aplicación completa se expondrán en la siguiente sección.

Capítulo 8

Resultados y análisis

8.1. Arquitectura

Incluyendo los cambios finales mencionados en la versión 2.0, la arquitectura queda como se ilustra en la figura 8.1.

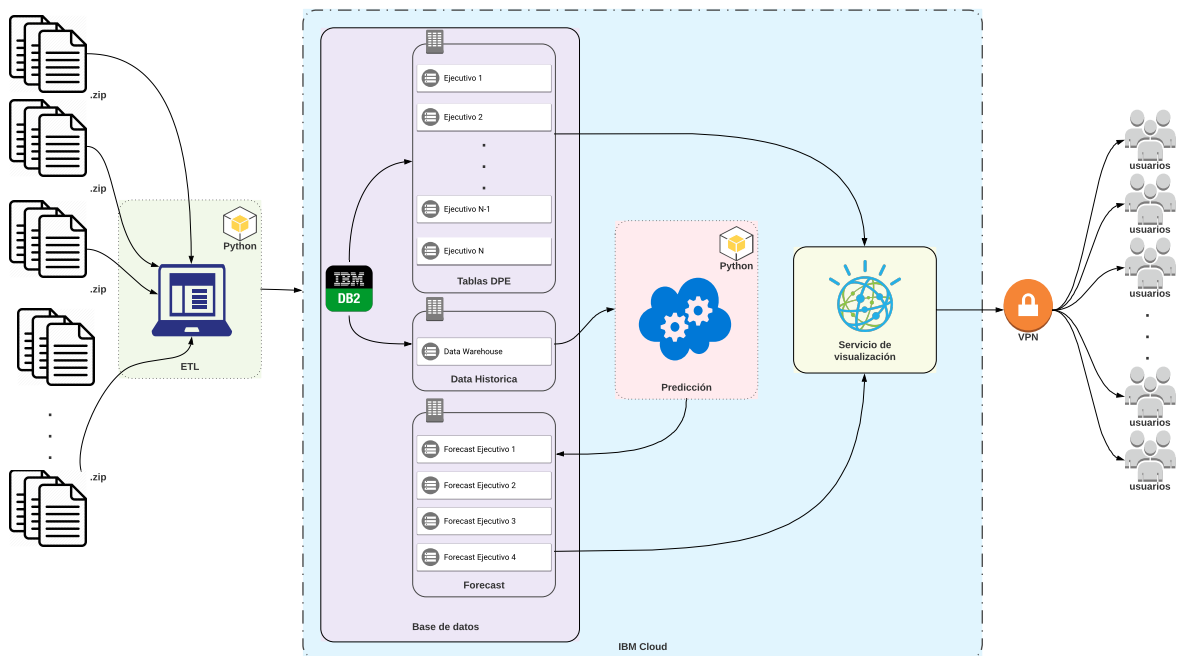


Figura 8.1: Arquitectura completa. Fuente: Elaboración propia.

Este diseño considera al igual que en la versión anterior, un equipo local que desencadena el proceso de automatización al empezar la recolección de los datos y su posterior procesamiento

para poder ser consumidos en la nube.

Una vez que termina el proceso ETL, los datos extraídos de los reportes .xslm y .txt son subidos a una base de datos histórica, en donde se almacena toda la información existente desde la creación del repositorio en BOX, y a una segunda instancia de bases de datos en donde se almacena información correspondiente al año contable.

Luego de almacenar los datos, las bases contables son consumidas inmediatamente por el servicio de visualización para que los usuarios puedan acceder a los datos (mediante VPN). Por otro lado, los datos históricos pasan a un segundo proceso de cálculo en el cual se modelan las predicciones según corresponda.

El resultado de este cómputo corresponde las fechas de las siguientes cuatro semanas junto con la información de FTE y costos en dólares para los respectivos AI. Luego, esta información es almacenada en un set de tablas llamadas ‘forecast + nombre ejecutivo’, que luego serán utilizadas por el servicio de visualización para que se desplieguen los gráficos.

Este proceso de modelamiento y predicción en la nube ocurre todos los días jueves a las 23:00, esto debido a que la actualización de información se realiza a más tardar el jueves a las 18:00, por tanto a las 23:00 ya se cuenta con la información actualizada para poder entrenar los modelos, además de ser un horario fuera de la jornada laboral que no interrumpe las actividades diarias.

8.2. Predicción

Se realizan pruebas utilizando los diferentes algoritmos en el mismo set de datos, considerando las últimas 4 semanas de cada Account ID como set de prueba.

8.2.1. SARIMA

Para el primer método con SARIMA, se utiliza la librería pyramid-auto arima, la cual permite, entre otras cosas, encontrar el mejor modelo para cada serie de tiempo utilizando un arreglo de parámetros. Aquí la librería prueba con todos hasta encontrar el con mejor AIC (Criterio de información de Akaike)[42].

Luego de la realización de los modelos, se obtiene la tabla 8.1, la cual muestra el resumen en cuanto a métricas de desempeño obtenidas para la realización de modelos tanto de costos como de horas para los 34 AI. De la tabla se puede ver un MAPE promedio de 27.55 % para los modelos de costos y un 30.67 % para los de horas. Adicionalmente, es posible ver de los RMSE promedio que la diferencia en magnitud entre costos y horas es de al menos 4 órdenes de magnitud, lo cual se condice con las magnitudes de cada serie.

	MAPE horas	MAPE costos	RMSE horas	RMSE costos	MSE horas	MSE costos
mean	27.55	30.67	0.13	754.03	0.039	979171
std	34.51	61.08	0.14	650.41	0.110	1599279
min	4.00	0.00	0.00	6.57	0.000	43
25 %	12.25	9.00	0.05	312.09	0.003	974173
50 %	15.00	13.50	0.08	688.61	0.007	474192
75 %	20.75	27.00	0.16	995.33	0.028	990723
max	170.00	346.00	0.79	2712.64	0.629	7358431

Tabla 8.1: Resumen estadístico de métricas asociadas a modelos generados con SARIMA. Fuente: Elaboración propia.

8.2.2. LSTM

Tal y como se mencionó en 7.3.3.3, se procede con redes LSTM considerando 4 configuraciones.

8.2.2.1. Configuración 1

Utilizando la primera configuración con 2 capas de 4 nodos, 100 épocas, batch de tamaño 1 y función de activación lineal se obtiene como resultado la tabla 8.2, la cual entrega un MAPE de 40.06 % para horas y 44.71 % para costos, lo cual es considerablemente peor que el resultado obtenido con SARIMA. Continuando el análisis en base a RMSE y MSE, se puede ver que en todos los aspectos este modelo se comporta peor que el anterior.

	MAPE horas	MAPE costos	RMSE horas	RMSE costos	MSE horas	MSE costos
mean	40.06	44.71	0.18	1055.74	0.06	1787374
std	54.74	69.51	0.17	832.57	0.12	2479572
min	8.15	4.25	0.00	79.36	0.00	6298
25 %	13.61	12.15	0.07	410.31	0.00	173207
50 %	21.13	19.75	0.14	856.74	0.02	734077
75 %	32.53	46.82	0.22	1698.45	0.05	2925770
max	279.65	376.73	0.81	3384.91	0.65	1145764

Tabla 8.2: Resumen estadístico de métricas asociadas a modelos generados con configuración 1 en LSTM. Fuente: Elaboración propia.

8.2.2.2. Configuración 2

En esta segunda configuración se utiliza 1 capa con 4 nodos, 10000 épocas con batch de tamaño 1 y nuevamente una función de activación lineal. La tabla 8.3 muestra los resultados asociados, en donde se puede ver un desempeño notablemente mejor al modelo anterior, en donde el MAPE se reduce aproximadamente a la mitad. Comparando con SARIMA se ve que este tiene un menor promedio y desviación para MAPE, no así para RMSE y MSE.

	MAPE horas	MAPE costos	RMSE horas	RMSE costos	MSE horas	MSE costos
mean	24.92	20.62	0.15	800.71	0.05	1165084
std	19.60	14.98	0.16	734.72	0.12	2005330
min	7.90	5.38	0.00	19.85	0.00	394
25 %	12.82	10.57	0.06	254.40	0.00	65428
50 %	16.77	16.99	0.09	562.43	0.00	316441
75 %	29.99	26.67	0.20	1063.87	0.04	1139465
max	100.30	64.16	0.81	2926.92	0.66	8566902

Tabla 8.3: Resumen estadístico de métricas asociadas a modelos generados con configuración 2 en LSTM. Fuente: Elaboración propia.

8.2.2.3. Configuración 3

Para esta nueva configuración, se plantea mantener la cantidad de capas, modificar el número de épocas a 400 y cambiar la función de activación a Tanh. En la tabla 8.4 es posible ver los resultados obtenidos para el conjunto de modelos, en donde se obtiene un desempeño mejor al modelo 2 en todos los aspectos y con un número de épocas considerablemente menor. Este modelo también presenta mejor MAPE en costos y horas tanto en promedio y desviación que el modelo con SARIMA y desempeño similar en RMSE y MSE.

	MAPE horas	MAPE costos	RMSE horas	RMSE costos	MSE horas	MSE costos
mean	24.76	19.07	0.15	770.65	0.04	1136681
std	18.94	15.14	0.15	747.80	0.10	2021906
min	4.92	5.34	0.00	21.82	0.000	476
25 %	12.36	7.96	0.07	236.06	0.00	55737
50 %	16.78	14.26	0.09	475.59	0.00	226429
75 %	32.91	24.32	0.21	899.807	0.04	809649
max	92.03	74.05	0.76	2958.52	0.58	8752894

Tabla 8.4: Resumen estadístico de métricas asociadas a modelos generados con configuración 3 en LSTM. Fuente: Elaboración propia.

8.2.2.4. Configuración 4

Utilizando la misma configuración del modelo 1 pero ahora con un dropout de 0.4, se obtiene la tabla 8.5, de la cual se puede ver un peor MAPE que en casos anteriores, aun así, se aprecia que la desviación para los casos de MSE y RMSE resulta ser menor que en el caso anterior.

8.2.2.5. Modelo a implementar

De lo anterior, es posible ver que la mejor configuración dentro de las utilizadas es el modelo 3 con LSTM, dado que presenta un mejor MAPE para ambos casos y una menor

	MAPE horas	MAPE costos	RMSE horas	RMSE costos	MSE horas	MSE costos
mean	45.42	41.76	0.17	995.35	0.05	1515914
std	64.94	55.31	0.14	735.59	0.08	1762347
min	7.27	5.22	0.00	25.26	0.00	638
25 %	14.39	12.72	0.08	370.16	0.00	138919
50 %	20.55	19.21	0.13	811.41	0.01	660048
75 %	40.44	48.77	0.21	1633.61	0.04	2669034
max	328.06	240.38	0.64	2427.49	0.42	5892744

Tabla 8.5: Resumen estadístico de métricas asociadas a modelos generados con configuración 4 en LSTM. Fuente: Elaboración propia.

desviación, junto con un RSME y SME similar al obtenido con SARIMA. Es en base a este resultado y la bibliografía relacionada que se escoge continuar con el modelo 3 [43].

Indagando en el detalle particular de cada modelo generado con la configuración 3, se obtiene la tabla 8.6, donde se puede apreciar los resultados obtenidos para cada modelo en particular junto con la cantidad de datos utilizados.

indice	Account Id	MAPE horas	MAPE costos	RMSE horas	RMSE costos	MSE horas	MSE costos	Num of Weeks
0	CHA7ZG	13.020419	6.845947	0.082114	486.832405	0.006743	2.370058e+05	97.0
1	CHA70A	11.904373	9.255357	0.316556	1242.801088	0.100208	1.544555e+06	97.0
2	CHA73B	8.172788	12.150371	0.109365	1179.299332	0.011961	1.390747e+06	97.0
3	CHA73D	9.877009	6.253553	0.088464	179.815928	0.007826	3.233377e+04	97.0
4	CHA714	21.310326	10.318004	0.323980	730.029467	0.104963	5.329430e+05	97.0
5	CHA727	17.859534	24.903792	0.184108	1225.828880	0.033896	1.502656e+06	96.0
6	CHA718	53.448085	47.384130	0.087932	377.547471	0.007732	1.425421e+05	97.0
7	CHA72A	9.780191	6.897767	0.235137	961.005104	0.055289	9.235308e+05	97.0
8	CHA723	13.949264	10.637777	0.098774	507.928140	0.009756	2.579910e+05	97.0
9	CHA72F	24.275850	16.485152	0.209390	570.647777	0.043844	3.256389e+05	96.0
10	CHA73F	14.646472	9.834777	0.567810	2153.311148	0.322408	4.636749e+06	97.0
11	CHA72H	52.257574	59.165500	0.341792	1819.981688	0.116822	3.312333e+06	96.0
12	CHA72B	92.116503	32.326380	0.091429	123.942586	0.008359	1.536176e+04	95.0
13	CHA72J	16.831573	11.739520	0.107217	706.289047	0.011495	4.988442e+05	96.0
14	CHA72I	17.723627	10.839448	0.124263	313.362689	0.015441	9.819617e+04	96.0
15	CHA716	39.636597	19.203971	0.767100	2535.486065	0.588442	6.428690e+06	97.0
16	CHA73A	29.384332	23.424771	0.066689	397.471969	0.004447	1.579840e+05	37.0
17	CHA72N	31.624038	17.423330	0.217319	803.596623	0.047228	6.457675e+05	96.0
18	CHA725	19.491551	14.858069	0.074423	447.480926	0.005539	2.002392e+05	96.0
19	CHA71K	9.990572	5.805735	0.003210	21.829436	0.000010	4.765243e+02	50.0
20	CHA7GL	12.747440	28.294105	0.187042	2987.815499	0.034985	8.927041e+06	76.0
21	CHA4KW	42.801902	17.210262	0.030033	103.102976	0.000902	1.063022e+04	141.0
22	CHAQSU	8.052563	7.630332	0.173284	923.414398	0.030027	8.526942e+05	141.0
23	CHAQST	21.321612	19.406211	0.010351	241.639517	0.000107	5.838966e+04	141.0
24	CHAQSW	8.776379	5.182340	0.012745	64.381903	0.000162	4.145029e+03	141.0
25	CHA87E	27.635110	25.569975	0.136549	704.396689	0.018646	4.961747e+05	62.0
26	CHA044	54.684315	66.782124	0.188985	1627.146206	0.035715	2.647605e+06	141.0
27	CHA8U3	14.951747	28.297515	0.019657	227.093775	0.000386	5.157158e+04	58.0
28	CHA967	17.748444	20.586943	0.081646	819.740305	0.006666	6.719742e+05	27.0
29	CHA9EK	46.405592	36.599038	0.029917	220.832984	0.000895	4.876721e+04	52.0
30	CHA88K	12.306115	11.108163	0.070439	301.467407	0.004962	9.088260e+04	62.0
31	CHA97H	39.929204	27.784700	0.202360	1639.275052	0.040949	2.687223e+06	28.0
32	CHBAD1	12.381876	9.966778	0.020319	125.559287	0.000413	1.576513e+04	18.0
33	CHBADT	10.265832	5.248437	0.030731	154.591396	0.000944	2.389850e+04	18.0

Tabla 8.6: Detalle de modelamiento por Account ID considerando la configuración 3. Fuente: Elaboración propia.

Para analizar la correlación entre rendimiento y cantidad de datos utilizados para el modelamiento, se procede a calcular la matriz de correlación entre los MAPEs obtenidos y la cantidad de datos o Num. of Weeks, el resultado puede verse en la figura 8.2, en donde se ve que para este caso particular, no se puede determinar una relación entre eficiencia y cantidad de datos, esto debido a que cada modelo es muy distinto al anterior por la naturaleza de cada serie de tiempo utilizada.

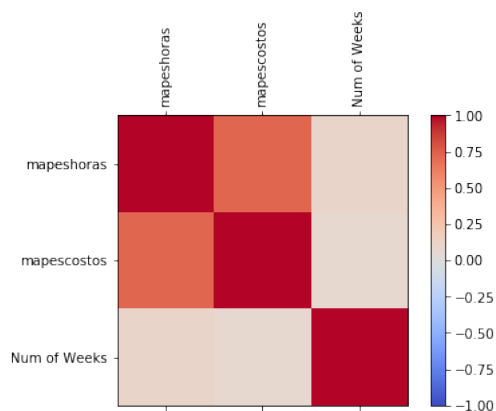


Figura 8.2: Correlación entre MAPE obtenido y cantidad de datos utilizados. Fuente: Elaboración propia.

En cuanto a validación y utilidad, se les presenta el resultado de la configuración 3 como piloto a los 4 DPE escogidos inicialmente (ver sección 8.5). Los resultados son avalados por los ejecutivos dado que les permite obtener tendencias y estimados respecto al comportamiento del siguiente mes. Es importante destacar que al momento de desplegar los datos, se les avisa y recalca la desviación del modelo junto con una alerta indicando resultados experimentales.

8.3. Visualización

Considerando los últimos cambios mencionados en la sección 7.3 se obtiene la versión final de la plataforma, la cual incluye:

- **Resumen Histórico:** Resumen de costos, FTE y horas cargadas según tipo: *overtime*, Normal y *stand by*. Posibilidad de seleccionar contrato o Account ID en particular.
- **Detalle de Horas:** Detalle específico de horas cargadas para las últimas 4 semanas con posibilidad de cambiar la ventana de tiempo. Se cuenta con el detalle por empleado, Account ID y contrato.
- **Costos USD por Bandas:** Detalle específico de costos en dólares para las últimas 4 semanas con posibilidad de cambiar la ventana de tiempo. Se cuenta con el detalle por empleado, Account ID y contrato. Adicionalmente se agrega la distinción por banda y mapa de relación costos versus horas cargadas.
- **FTE por Banda:** Información de FTE cargadas a lo largo del año contable, con distinción por tipo de banda.

- **Cuentas Internas:** Detalle en cuanto a horas cargadas por empleado para cuentas internas de IBM que funcionan con Claim.
- **Predicción:** Pestaña de predicción en donde se encuentra el detalle en costos y FTE para las siguientes 4 semanas a partir de la última actualización.
- **Tutorial:** Pestaña tutorial con información relevante para la utilización de la plataforma.

Un ejemplo con datos escalados y nombres cambiados puede verse en el siguiente enlace:

<https://ibm.co/2QgVhrP>

Cabe destacar que no todas las vistas presentan las 7 pestañas anteriormente mencionadas, ya que dependerá de que cuentas y contratos maneja cada ejecutivo.

8.4. Robustez y confiabilidad de la plataforma

Para medir la robustez y confiabilidad de los datos, se realizan comparaciones con el proceso manual de carga y procesamiento de datos en Excel, tal y como se realizó en la implementación del MVP. A partir de esto, se puede ver una diferencia o desviación del 0% de los datos, es decir, el proceso se logra estandarizar y automatizar sin incurrir en errores.

Junto con lo anterior, y a lo largo de la tercera iteración, se presentan diversas instancias para comparar los resultados obtenidos a partir del procesamiento de los datos, debido a que en paralelo se generan otros reportes (por otras áreas con propósitos diferentes) que contienen la misma información, pero extraída desde un origen y método diferente. Estas instancias permiten dar cuenta de la robustez y estandarización del proceso ya que la verificación advierte errores cometidos en los reportes generados de las otras áreas. En base a esto, IBM decide utilizar la plataforma como origen oficial para la generación de esos reportes mediante la descarga de los datos en formato Excel, permitiendo así, la extensión de su uso.

El punto mencionado anteriormente representa una gran aprobación de la plataforma, puesto que el funcionamiento y utilización se valida por empleados de otras áreas. Junto con eso, se obtienen validaciones o testimonios de algunos usuarios (DPE) y del gerente (Manager) encargado de los ejecutivos.

Pablo Gonzalez, Delivery Partner Executive:

“...Agradezco la iniciativa de crear esta aplicación para facilitar nuestra tarea como DPE...”

Miguel Padilla, Delivery Partner Executive:

“...Quiero agradecer y destacar la herramienta Análisis de Claim, la cual me ha permitido tener la información disponible en todo momento y por tanto la considero un apoyo en la gestión...”

Carlos Lambert, Delivery Partner Manager:

“...Quiero destacar la automatización realizada la cual nos permite tener un Management System de control de labor...”

En el Anexo F es posible encontrar las validaciones citadas mediante correos electrónicos por parte de usuarios de la plataforma, así como también del gerente encargado de los ejecutivos.

8.5. Utilización de la Plataforma

El detalle de ingresos por cada DPE se puede ver en la tabla 8.7, la cual muestra el uso por parte de usuarios desde el mes de mayo hasta octubre. En esta tabla se ve que de los 18, los últimos 3 en la tabla solo representan el 5% de los ingresos, dando a entender que no se logró la penetración en esos ejecutivos y sus respectivos equipos. Por otro lado, se ve como los primeros 4 ejecutivos de la tabla abarcan el 35.86% del uso, indicando una fuerte adhesión (Son estos 4 ejecutivos quienes se escogen para realizar el módulo de predicción).

Vista por DPE	Núm. de visitas	Visitas promedio por semana
Ejecutivo 1	55	2.75
Ejecutivo 2	45	2.25
Ejecutivo 3	42	2.10
Ejecutivo 4	42	2.10
Ejecutivo 5	38	1.90
Ejecutivo 6	34	1.70
Ejecutivo 7	31	1.55
Ejecutivo 8	29	1.45
Ejecutivo 9	28	1.40
Ejecutivo 10	27	1.35
Ejecutivo 11	25	1.25
Ejecutivo 12	25	1.25
Ejecutivo 13	23	1.15
Ejecutivo 14	22	1.10
Ejecutivo 15	22	1.10
Ejecutivo 16	10	0.50
Ejecutivo 17	8	0.40
Ejecutivo 18	7	0.35

Tabla 8.7: Tabla de uso por ejecutivo. Fuente: Elaboración propia.

El desuso registrado responde en gran medida al tipo y tamaño de las cuentas que maneja cada ejecutivo, ya que en cuentas pequeñas, la cantidad de AI y empleados a cargo resulta ser menor, por lo cual los DPE argumentan no sentir la necesidad de una plataforma adicional para llevar el control y gestión de sus cuentas.

Adicionalmente, la tabla 8.7 entrega los ingresos promedios por semana que cada vista tuvo, en ese sentido, se puede ver que -sin considerar a los últimos 3 ejecutivos que presentaron menor participación- todos ingresaron a la plataforma en promedio más de una vez a la semana, dando cuenta de que aumentó la utilización de los datos respecto a la situación inicial en donde los DPE afirmaban revisar estos reportes en el momento en que los procesaban, es decir, una vez a la semana. Esto se traduce en un incremento del uso promedio de los datos en un 42,5 %, dejando ver la utilidad adicional que se le extrae a los datos con el análisis y procesamiento realizado.

8.6. Reducción de tiempo

Antes de poder hablar de reducción de tiempo es necesario conocer todos los tiempos y procesos involucrados en esta labor:

Por el lado de los ejecutivos, se tiene el tiempo de extracción y procesamiento que utilizan los DPE para poder obtener la información relevante, que luego será utilizada en una segunda etapa que involucra tiempo de análisis para realizar el control de costos y recursos.

Por el lado de la plataforma, se tiene un tiempo asociado al proceso de extracción; procesamiento y carga de datos a la nube, el cual se ejecuta en segundo plano en un equipo local sin interferir ni depender de terceros. Adicionalmente, se tiene el tiempo de entrenamiento de las redes LSTM que semana a semana reciben nuevos datos y son actualizadas.

A pesar de que la extracción y procesamiento de los datos, junto con el entrenamiento de las redes neuronales toma un tiempo considerable (20 a 30 minutos la extracción – procesamiento y 2 - 3 horas para el entrenamiento completo de todos los modelos), ninguno de estos procesos inhabilita o interfiere en las actividades diarias del personal de IBM, ya que uno se realiza en segundo plano en un equipo local y el otro se ejecuta en la nube fuera de la jornada laboral. Por tanto, si se realiza la comparación entre tiempo de extracción y procesamiento antes y después de la implementación se puede hablar de una reducción del 100 % del tiempo, es decir, el usuario percibe el servicio como automático y siempre disponible. Lo cual cumple con el objetivo principal de aliviar la carga y reducir los costos, ya que una hora de ejecutivo tiene costos elevados asociados.

Considerando el tiempo completo que destinaban los ejecutivos entre extracción, procesamiento y análisis, se tiene que los usuarios dedicaban entre 2 a 3 horas, por tanto se tienen al menos 36 horas de ejecutivo en esta labor. Ahora, tomando los resultados expuestos en la encuesta 7.11 es posible ver que se logran reducir en base a sus propias palabras al menos 15 horas con la implementación de la plataforma, lo cual equivale a una reducción del 41 % del tiempo total destinado a esta labor. Con esto, se logra redistribuir el tiempo de los ejecutivos junto a sus equipos permitiendo una mayor capacidad de análisis y de realizar otras actividades.

Capítulo 9

Conclusiones y trabajo futuro

9.1. Conclusiones

El trabajo presentado se centró en el diseño e implementación de una herramienta tipo panel de indicadores en la nube con acceso controlado para apoyar la gestión interna dentro de IBM Chile – GTS, esto a través de la recolección, procesamiento y despliegue inteligente de la información. Lo anterior, como iniciativa enmarcada en un contexto de innovación dentro la empresa.

Sobre el análisis exploratorio visto en el capítulo 6, se destaca el levantamiento de información previo a la implementación, el cual aseguró contar con los datos debidamente tratados para las siguientes etapas. En ese sentido, un desarrollo de la plataforma sin haber comprendido los datos en su totalidad hubiese resultado en errores futuros, como inconsistencias respecto a la información mostrada en las vistas, lo cual podría terminar en acciones erróneas.

Con relación a la arquitectura montada, se pudo ver que la implementación final dista de la presentada inicialmente en el capítulo 5, dado que se le agregaron tablas y servicios adicionales para la predicción de recursos y horas. Sin embargo, este hecho no afectó el desarrollo del proyecto debido a que el diseño propuesto contempló flexibilidad y escalabilidad, por tanto, no fue necesario modificar lo existente, sino que solo se debieron anexar los recursos faltantes. En ese sentido, se postula trabajar en lo posible con servicios en la nube, los cuales permiten cierta modularidad que facilita y agiliza el desarrollo.

En cuanto a la plataforma en sí, se logró diseñar e implementar un tablero de indicadores capaz de extender las funcionalidades que poseían los reportes usados antiguamente, brindando la capacidad de: Análisis histórico, vista de tendencias y cambios a lo largo del año, detalle específico por empleado; Account ID o incluso contrato, agrupación de recursos en base a las bandas (o costos), cruce de información entre distintas cuentas según DPE, registro de gastos mensuales y semanales, entre otras funcionalidades que permite el dinamismo de la plataforma al ser interactiva.

Con respecto a la utilización, se logra incorporar la herramienta a las labores diarias de los

ejecutivos y sus respectivos equipos en un 83.33 % de los casos, obteniendo una reducción del 100 % en cuanto a tiempo de extracción y procesamiento de la información para los usuarios activos, junto con la disminución de al menos 15 horas (reducción de un 41 % del tiempo total) si se consideran las horas utilizadas por todos los DPE (incluyendo los que no usan la plataforma) en el proceso completo de monitoreo de sus cuentas. Este punto representa el principal indicador de éxito de la plataforma al validar el desarrollo mediante su uso y adopción oficial por parte de IBM Chile – GTS. Asimismo, las métricas obtenidas indican que aumentó la utilización semanal promedio de los datos respecto a la situación inicial en un 42.5 %, lo que se traduce en una mayor explotación de la información a un menor costo, es decir, un proceso más eficiente.

Por otro lado, los resultados obtenidos utilizando LSTM indican que la idea de integrar un módulo de predicción a la plataforma de control es factible, permitiendo obtener resultados con un MAPE promedio en torno al 24 % para horas y 19 % para costos sin optimizar completamente los modelos utilizados. Sin embargo, para llegar a todos los ejecutivos es necesario seguir desarrollando los modelos para disminuir el error, ya que como se observó en 8.2.2.5, las desviaciones estándar se encuentran en torno al 18 % e incluso existen IA que obtuvieron un MAPE sobre el 70 % en costos y horas.

Analizando de forma íntegra el desarrollo realizado, se destaca la metodología presentada en el capítulo 3, la cual permitió probar rápidamente el impacto sin gastar recursos excesivos y mucho menos disponer de la plataforma completa. En este contexto, la realización de una versión preliminar (MVP) resultó fundamental para interiorizar el proceso manual que realizaban los ejecutivos junto con la obtención de experiencias por parte de los usuarios con el fin de entender como debía ser la plataforma. En base a lo anterior, se postula replicar la forma de proceder en proyectos similares en donde se deba trabajar junto con usuarios para así alinear los objetivos.

Por último, se enfatiza el valor del trabajo realizado, en donde el rol principal no fue desarrollar nuevas tecnologías, sino que integrar soluciones existentes a un problema cotidiano pero de gran importancia dentro de una organización, brindando un valor agregado a los datos existentes y sobretodo aumentando la eficiencia, permitiendo una redistribución del tiempo utilizado por parte de los ejecutivos y sus respectivos equipos.

9.2. Trabajo futuro

En base a los resultados, se propone como trabajo futuro:

- Estudiar el impacto de esta plataforma en la forma de trabajar de los ejecutivos en cuanto a rendimiento y resultados. Para lo anterior, se propone analizar el desempeño de los DPE con y sin la implementación considerando grupos de control, para así medir el impacto de las visualizaciones generadas.
- Personalizar la experiencia en las distintas vistas para los usuarios, a modo de mejorar la experiencia y así penetrar en los usuarios que aun no incorporan esta herramienta a sus actividades diarias.

- En el trabajo presentado, se realiza una primera aproximación a lo que es un módulo de predicción para costos y horas cargadas. En futuras versiones se propone extender los pronósticos a todos los usuarios y mejorar estos modelos, agregando un análisis exhaustivo y personalizando los modelos para cada Account ID con el fin de optimizar los resultados. Asimismo, se propone realizar un estudio para evaluar el tamaño ideal de la ventana de tiempo que se adecua mejor al problema para que los valores sean lo más representativos posibles y sean útiles para los DPE. Para lo anterior, es necesario trabajar directamente con los usuarios para saber que ventana de tiempo les resulta útil.
- Extender esta estructura y replicarla en otros escenarios tanto para procesos internos como externos en IBM, brindando la posibilidad de sacar el máximo provecho a la información que se genera. pño
- Ver la posibilidad de modificar la extracción de los datos, evitando tener que utilizar los informes y conectándose directamente a la base de datos para lograr una actualización en tiempo real de la información. Para la realización de esta mejora y soportar el stream de datos que implica tener la información en tiempo real, se deberá proponer una nueva arquitectura en donde el proceso completo se encontrará en la nube.
- Incorporar alertas a la plataforma frente a anomalías, con el fin de abordarlas en el menor tiempo posible.
- Realizar la integración de otras fuentes de información que permitan dar un valor agregado a la plataforma, como puede ser el cruce de información con los incidentes que ocurren asociados a los distintos clientes, con lo cual se podría tener en tiempo real las situaciones que están ocurriendo y, por otro lado, la posibilidad de cruzar trabajos e incidentes resueltos con actividad registrada en Claim, para poder abordar temas de productividad y eficiencia.

Respecto a este último punto mencionado, al momento de finalizar la experiencia de trabajo de memoria, IBM Chile dispone de un nuevo equipo encargado de darle mantenimiento y extender la plataforma desarrollada. Este equipo queda encargado de realizar una extensión mediante el cruce de información con los tickets de servicio asociados a incidentes y requerimientos, hecho que avala la utilidad de la herramienta y permite continuidad para futuras versiones.

Bibliografía

- [1] J. G. Chávez Barbaste, “Análisis y modelos de datos de redes para seguridad informática,” *Universidad de Chile*, 2016.
- [2] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: Forecasting and control: Fourth edition*. 2013.
- [3] C. Gershenson, “Artificial Neural Networks for Beginners,” *arXiv*, no. April, p. 8, 2003.
- [4] P. A. Granell, “Redes Neuronales Recurrentes: Una aplicación para los mercados bursátiles,” 2017.
- [5] “Understanding lstm networks.” <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Último acceso: 02-01-2019.
- [6] “Overview of cognos dashboard embedded.” <https://console.bluemix.net/docs/services/cognos-dashboard-embedded/index.html#overview-of-cognos-dashboard-embedded>. Último acceso: 02-01-2019.
- [7] “Cognos analytics - overview | ibm.” <https://www.ibm.com/products/cognos-analytics>. Último acceso: 02-01-2019.
- [8] E. Team, “Envisn’s ibm cognos blog.” <https://www.envisn.com/envisn-cognos-blog/cognos-analytics-versus-watson-analytics>. Último acceso: 02-01-2019.
- [9] “Compare ibm watson analytics vs ibm cognos 2018.” <https://comparisons.financesonline.com/ibm-watson-analytics-vs-ibm-cognos>. Último acceso: 02-01-2019.
- [10] “Watson studio.” <https://www.ibm.com/cloud/watson-studio>. Último acceso: 02-01-2019.
- [11] V. Summary and E. C. Definitions, “Magic Quadrant for Business Intelligence Platforms Market Definition / Description,” *Intelligence*, no. February, pp. 1–32, 2018.
- [12] “Kdnuggets - artificial intelligence, deep learning neural networks explained.” <https://www.kdnuggets.com/2016/10/artificial-intelligence-deep-learning-neural-networks-explained.html>.

Último acceso: 02-01-2019.

- [13] “Introducción al perceptron con python.” <https://ernestocrespo13.wordpress.com/2018/01/07/introduccion-al-perceptron-con-python/>, Mar 2018. Último acceso: 02-01-2019.
- [14] C. Business, “Welcome to the Cognitive Era . A new era in technology , a new era in business .,” *Cognitive Business*, no. Produced by IBM Blobal in the United States of America November 2015, pp. 1–9, 2015.
- [15] F. Azma and M. A. Mostafapour, “Business intelligence as a key strategy for development organizations,” *Procedia Technology*, vol. 1, pp. 102–106, 2012.
- [16] S. Quintana Lecaros, *Business intelligence en la empresas chilenas*. Springer, 2013.
- [17] C. Matt, T. Hess, and A. Benlian, “Digital Transformation Strategies,” *Business and Information Systems Engineering*, vol. 57, no. 5, pp. 339–343, 2015.
- [18] R. Chugh and S. Grandhi, “Why Business Intelligence?,” *International Journal of E-Entrepreneurship and Innovation*, vol. 4, no. 2, pp. 1–14, 2013.
- [19] B. Nedelcu, “Business Intelligence Systems,” *Database Systems Journal*, vol. IV, no. 4/2013, pp. 12–20, 2012.
- [20] R. van der Lans, *Data Virtualization for Business Intelligence Systems*. 2012.
- [21] A. Maria, I. Florea, V. Diaconita, and R. Bologa, “Data integration approaches using ETL,” *Database Systems Journal*, 2015.
- [22] J. Conesa, C. Isabel, and G. Hormigo, “Procesos ETL,” pp. 1–24.
- [23] S. Angra and S. Ahuja, “Machine learning and its applications: A review,” *Proceedings of the 2017 International Conference On Big Data Analytics and Computational Intelligence, ICBDACI 2017*, pp. 57–60, 2017.
- [24] S. Rouhani, A. Ashrafi, A. Zare Ravasan, and S. Afshari, “The impact model of business intelligence on decision support and organizational benefits,” *Journal of Enterprise Information Management*, 2016.
- [25] L. Calzada and J. L. Abreu, “El impacto de las herramientas de inteligencia de negocios en la toma de decisiones de los ejecutivos.,” *Revista Daena (International Journal of Good Conscience)*, vol. 4, no. 2, 2009.
- [26] A. Popovič, P. S. Coelho, and J. Jaklič, “The impact of business intelligence system maturity on information quality,” 2009.
- [27] R. S. Sharma and V. Djiaw, “Realising the strategic impact of business intelligence tools,” *VINE*, 2011.
- [28] S. Williams and N. Williams, “The business value of business intelligence,” in *The Profit*

Impact of Business Intelligence, 2007.

- [29] S. Negash, “Business intelligence,” *Communications of the association for information systems*, vol. 13, no. 1, p. 15, 2004.
- [30] M. Z. Elbashir, P. A. Collier, and M. J. Davern, “Measuring the effects of business intelligence systems: The relationship between business process and organizational performance,” *International Journal of Accounting Information Systems*, vol. 9, no. 3, pp. 135–153, 2008.
- [31] S. Categories, A. Products, and M. Popular, “Business Intelligence Software. THE TOP 20 MOST POPULAR Business Intelligence Software,” pp. 1–9, 2018.
- [32] C. Maulén, “Adopción & Brechas para Transformación Digital,” *Universidad de Chile*, vol. 1, p. 16, 2018.
- [33] C. Chatfield, “Time-series forecasting,” *Significance*, 2005.
- [34] S. Hochreiter, “The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 1998.
- [35] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, 1997.
- [36] J. T. Connor, R. D. Martin, and L. E. Atlas, “Recurrent Neural Networks and Robust Time Series Prediction,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 240–254, 1994.
- [37] S. Makridakis, “Accuracy measures: theoretical and practical concerns,” *International Journal of Forecasting*, 1993.
- [38] A. Moran, “Agile software development,” in *SpringerBriefs in Computer Science*, 2014.
- [39] D. McKenna, *The Art of Scrum: How Scrum Masters Bind Dev Teams and Unleash Agility*. Berkely, CA, USA: Apress, 1st ed., 2016.
- [40] M. Vianna, *Design thinking: inovação em negócios*. Design Thinking, 2012.
- [41] “Código del trabajo, artículo 22,” *DT - Consultas*. Último acceso: 02-01-2019.
- [42] D. Dickey, P. Brockwell, and R. Davis, “Time Series Theory and Methods,” *Technometrics*, 1989.
- [43] S. Siami-Namini and A. S. Namin, “Forecasting Economics and Financial Time Series: ARIMA vs. LSTM,” pp. 1–19, 2018.

Apéndice A

Tablas resumen de cuentas

	Fuente	Num of Work Item Id	First Date	Last Date	Num of Weeks
0	A3998_XXXX	2	2016-01-01	2016-12-23	51.0
1	A3998_XXXX	1	2016-12-30	2018-06-22	77.0
2	CH1100_XXXX	2	2016-01-01	2016-12-23	51.0
3	CH1120_XXXX	9	2016-01-01	2016-12-23	51.0
4	CH1120_XXXX	6	2016-12-30	2018-06-22	77.0
5	CH1138_XXXX	17	2016-05-06	2018-06-22	111.0
6	CH1142_XXXX	4	2016-05-27	2018-06-22	108.0
7	CH1146_XXXX	3	2016-08-12	2017-12-22	71.0
8	CH1149_XXXX	9	2016-09-23	2018-06-22	91.0
9	CH1152_XXXX	4	2016-11-11	2018-06-22	84.0
10	CH1159_XXXX	1	2016-10-21	2016-12-23	9.0
11	CH1172_XXXX	4	2017-02-03	2018-06-22	72.0
12	CH1173_XXXX	4	2017-03-03	2018-06-22	68.0
13	CH1179_XXXX	2	2016-12-30	2018-03-02	61.0
14	CH1180_XXXX	2	2018-03-16	2018-06-22	14.0
15	CH1185_XXXX	22	2017-01-06	2018-06-22	76.0
16	CH1193_XXXX	2	2017-07-07	2018-06-01	47.0
17	CH1194_XXXX	2	2017-05-12	2018-06-22	58.0
18	CH1197_XXXX	4	2017-06-09	2018-06-22	54.0
19	CH1201_XXXX	1	2017-09-01	2018-06-22	42.0
20	CH1203_XXXX	1	2017-09-01	2018-06-22	42.0
21	CH1204_XXXX	1	2017-09-01	2017-12-22	16.0
22	CH1205_XXXX	1	2017-12-29	2018-06-22	25.0
23	CH1206_XXXX	1	2017-09-08	2018-01-26	20.0
24	CH1207_XXXX	1	2017-12-29	2018-06-01	22.0
25	CH1208_XXXX	5	2017-12-29	2018-06-22	25.0
26	CH1212_XXXX	1	2017-12-29	2018-06-22	25.0
27	CH1215_XXXX	37	2017-12-29	2018-06-22	25.0
28	CH1216_XXXX	2	2018-01-12	2018-06-22	23.0
29	CH1218_XXXX	1	2018-06-01	2018-06-08	1.0
30	CH1220_XXXX	2	2018-03-09	2018-06-22	15.0
31	CHAAB8_XXXX	13	2016-01-01	2016-12-23	51.0
32	CHAACE_XXXX	1	2016-01-08	2016-12-23	50.0

	Fuente	Num of Work Item Id	First Date	Last Date	Num of Weeks
33	CHAACG_XXXX	1	2016-10-28	2016-12-02	5.0
34	CHAACG_XXXX	1	2016-01-01	2016-12-23	51.0
35	CHAACI_XXXX	1	2016-09-16	2016-12-02	11.0
36	CHAACI_XXXX	20	2017-12-29	2018-01-05	1.0
37	CHAACJ_XXXX	6	2017-12-29	2018-06-22	25.0
38	CHAACN_XXXX	1	2017-12-29	2017-12-29	0.0
39	CHAACO_XXXX	4	2016-10-07	2016-12-23	11.0
40	CHAACS_XXXX	1	2017-10-27	2017-12-22	8.0
41	CHE590_XXXX	20	2016-01-01	2018-06-22	129.0
42	CHMAWP_XXXX	2	2016-01-01	2018-06-22	129.0
43	CL0004_XXXX	2	2016-01-01	2016-09-09	36.0
44	CL0004_XXXX	1	2017-12-01	2018-06-22	29.0
45	CL0023_XXXX	2	2016-01-01	2016-12-23	51.0
46	CL0119_XXXX	3	2016-01-08	2018-06-22	128.0
47	CL0129_XXXX	2	2016-01-01	2016-12-23	51.0
48	CL0129_XXXX	3	2016-01-01	2018-06-22	129.0
49	CL0260_XXXX	33	2016-01-08	2018-06-22	128.0
50	CL0371_XXXX	4	2016-04-08	2018-06-22	115.0
51	CL0406_XXXX	1	2016-05-27	2016-05-27	0.0
52	CL0416_XXXX	1	2016-05-27	2016-12-16	29.0
53	CL0430_XXXX	7	2016-05-20	2016-12-16	30.0
54	CL0439_XXXX	2	2017-02-03	2018-01-26	51.0
55	CL0451_XXXX	4	2016-06-17	2016-12-23	27.0
56	CL0526_XXXX	6	2016-10-07	2018-06-22	89.0
57	CL0532_XXXX	6	2016-09-09	2016-12-23	15.0
58	CL0540_XXXX	18	2016-12-30	2018-06-22	77.0
59	CL0605_XXXX	1	2016-09-30	2018-06-22	90.0
60	CL310_XXXX	1	2016-04-15	2016-10-28	28.0
61	CLAEAL_XXXX	1	2016-11-04	2018-06-22	85.0
62	CLG7KO_XXXX	5	2016-11-04	2018-06-22	85.0
63	CLNMTA_XXXX	44	2016-11-04	2018-06-22	85.0
64	CLWKMH_XXXX	1	2018-06-01	2018-06-01	0.0
65	HC0003_XXXX	1	2018-04-13	2018-05-25	6.0
66	HV007Y_XXXX	4	2016-01-01	2018-06-22	129.0
67	HV064L_XXXX	1	2016-01-01	2018-06-22	129.0
68	IG8698_XXXX	1	2016-01-01	2016-12-23	51.0
69	IJ6098_XXXX	3	2016-01-01	2017-12-29	104.0
70	IJ7054_XXXX	33	2016-01-01	2018-06-22	129.0
71	IJ8050_XXXX	12	2016-01-01	2018-06-22	129.0
72	IJ8202_XXXX	4	2016-01-15	2018-06-01	124.0
73	IJ8900_XXXX	2	2016-01-01	2016-12-23	51.0
74	IJ8993_XXXX	14	2016-01-01	2018-06-22	129.0
75	IJ9147_XXXX	1	2016-01-01	2016-12-23	51.0
76	IL8486_XXXX	3	2016-01-01	2016-12-23	51.0
77	IM0470_XXXX	1	2016-01-01	2016-12-23	51.0
78	IM0470_XXXX	1	2016-12-30	2018-06-22	77.0
79	IM0647_XXXX	1	2016-01-01	2016-12-23	51.0
80	IM0647_XXXX	3	2016-12-30	2018-06-22	77.0
81	IM4022_XXXX	2	2016-01-01	2016-12-23	51.0

	Fuente	Num of Work Item Id	First Date	Last Date	Num of Weeks
82	IM4022_XXXX	3	2016-12-30	2018-06-22	77.0
83	IM5307_XXXX	24	2016-01-01	2016-12-02	48.0
84	IM6553_XXXX	12	2016-12-30	2018-06-22	77.0
85	IN5779_XXXX	15	2016-01-01	2018-06-22	129.0
86	IN7088_XXXX	2	2016-01-01	2016-12-16	50.0
87	IN7088_XXXX	3	2016-01-01	2018-03-23	116.0
88	IN7294_XXXX	2	2016-01-01	2016-12-23	51.0
89	IN8010_XXXX	1	2016-01-01	2016-12-23	51.0
90	IN8165_XXXX	23	2016-01-01	2018-06-22	129.0
91	IN8218_XXXX	4	2016-01-01	2016-12-23	51.0
92	IN8523_XXXX	7	2016-01-01	2018-06-22	129.0
93	IN8691_XXXX	3	2016-01-01	2018-06-22	129.0
94	IN8714_XXXX	38	2016-01-01	2018-06-22	129.0
95	IN8888_XXXX	24	2016-01-01	2018-06-22	129.0
96	IN8948_XXXX	66	2016-01-01	2018-06-22	129.0
97	IO9145_XXXX	12	2016-01-01	2018-06-22	129.0
98	IO9276_XXXX	3	2016-01-01	2018-06-22	129.0
99	IO9277_XXXX	1	2016-02-19	2016-12-23	44.0
100	IO9345_XXXX	54	2016-01-01	2017-12-22	103.0
101	IO9383_XXXX	5	2016-01-01	2018-06-22	129.0
102	IO9432_XXXX	17	2016-01-01	2018-06-22	129.0
103	IO9433_XXXX	3	2016-01-01	2018-06-22	129.0
104	IO9459_XXXX	6	2016-01-01	2016-12-23	51.0
105	IO9500_XXXX	27	2016-01-01	2018-06-22	129.0
106	IO9565_XXXX	9	2016-01-01	2018-06-22	129.0
107	IP0131_XXXX	7	2016-01-01	2018-06-22	129.0
108	IP4776_XXXX	41	2016-01-01	2018-06-22	129.0
109	IP5409_XXXX	6	2016-01-01	2018-06-22	129.0
110	IP5609_XXXX	10	2016-01-01	2016-12-23	51.0
111	IP5609_XXXX	2	2017-12-29	2018-06-22	25.0
112	IP7443_XXXX	6	2016-01-01	2018-06-22	129.0
113	IP7990_XXXX	9	2016-01-01	2016-12-23	51.0
114	IP7991_XXXX	37	2016-01-01	2016-12-23	51.0
115	IP8352_XXXX	23	2016-01-01	2018-06-22	129.0
116	IP8489_XXXX	41	2016-01-01	2018-06-22	129.0
117	IP8543_XXXX	1	2017-03-03	2017-12-29	43.0
118	IT0189_XXXX	42	2016-01-01	2018-06-22	129.0
119	IT0425_XXXX	3	2016-01-01	2018-06-22	129.0
120	MC2055_XXXX	4	2016-01-01	2016-12-23	51.0
121	MC2061_XXXX	1	2016-07-29	2016-12-23	21.0
122	MC2069_XXXX	2	2016-01-01	2016-12-23	51.0
123	MC2080_XXXX	3	2016-01-22	2016-12-23	48.0
124	MC2121_XXXX	1	2018-05-18	2018-06-22	5.0
125	MC2276_XXXX	1	2016-12-30	2018-04-20	68.0

Tabla A.1: Tabla resumen de Work Item Id por contratos, considerando: Primer día, última día y duración en semanas. Fuente: Elaboración propia.

	SIP	Num of Work Item ID	First Date	Last Date	Num of Weeks
0	-	49	2016-01-01	2018-06-22	129.0
1	A3998	2	2016-01-01	2018-06-22	129.0
2	CH1100	2	2016-01-01	2016-12-23	51.0
3	CH1120	11	2016-01-01	2018-06-22	129.0
4	CH1138	17	2016-05-06	2018-06-22	111.0
5	CH1142	4	2016-05-27	2018-06-22	108.0
6	CH1146	3	2016-08-12	2017-12-22	71.0
7	CH1149	9	2016-09-23	2018-06-22	91.0
8	CH1152	4	2016-11-11	2018-06-22	84.0
9	CH1159	1	2016-10-21	2016-12-23	9.0
10	CH1172	4	2017-02-03	2018-06-22	72.0
11	CH1173	4	2017-03-03	2018-06-22	68.0
12	CH1179	2	2016-12-30	2018-03-02	61.0
13	CH1180	2	2018-03-16	2018-06-22	14.0
14	CH1185	22	2017-01-06	2018-06-22	76.0
15	CH1193	2	2017-07-07	2018-06-01	47.0
16	CH1194	2	2017-05-12	2018-06-22	58.0
17	CH1197	4	2017-06-09	2018-06-22	54.0
18	CH1201	1	2017-09-01	2018-06-22	42.0
19	CH1203	1	2017-09-01	2018-06-22	42.0
20	CH1204	1	2017-09-01	2017-12-22	16.0
21	CH1205	1	2017-12-29	2018-06-22	25.0
22	CH1206	1	2017-09-08	2018-01-26	20.0
23	CH1207	1	2017-12-29	2018-06-01	22.0
24	CH1208	5	2017-12-29	2018-06-22	25.0
25	CH1212	1	2017-12-29	2018-06-22	25.0
26	CH1215	37	2017-12-29	2018-06-22	25.0
27	CH1216	2	2018-01-12	2018-06-22	23.0
28	CH1218	1	2018-06-01	2018-06-08	1.0
29	CH1220	2	2018-03-09	2018-06-22	15.0
30	CHE590	20	2016-01-01	2018-06-22	129.0
31	CHMAWP	2	2016-01-01	2018-06-22	129.0
32	CL0004	2	2016-01-01	2018-06-22	129.0
33	CL0023	2	2016-01-01	2016-12-23	51.0
34	CL0119	3	2016-01-08	2018-06-22	128.0
35	CL0129	3	2016-01-01	2018-06-22	129.0
36	CL0260	33	2016-01-08	2018-06-22	128.0
37	CL0371	4	2016-04-08	2018-06-22	115.0
38	CL0406	1	2016-05-27	2016-05-27	0.0
39	CL0416	1	2016-05-27	2016-12-16	29.0
40	CL0430	7	2016-05-20	2016-12-16	30.0
41	CL0439	2	2017-02-03	2018-01-26	51.0
42	CL0451	4	2016-06-17	2016-12-23	27.0
43	CL0526	6	2016-10-07	2018-06-22	89.0
44	CL0532	6	2016-09-09	2016-12-23	15.0
45	CL0540	18	2016-12-30	2018-06-22	77.0
46	CL0605	1	2016-09-30	2018-06-22	90.0
47	CL310	1	2016-04-15	2016-10-28	28.0

	SIP	Num of Work Item ID	First Date	Last Date	Num of Weeks
48	CLAEAL	1	2016-11-04	2018-06-22	85.0
49	CLG7KO	5	2016-11-04	2018-06-22	85.0
50	CLNMTA	44	2016-11-04	2018-06-22	85.0
51	CLWKMH	1	2018-06-01	2018-06-01	0.0
52	HC0003	1	2018-04-13	2018-05-25	6.0
53	HV007Y	4	2016-01-01	2018-06-22	129.0
54	HV064L	1	2016-01-01	2018-06-22	129.0
55	IG8698	1	2016-01-01	2016-12-23	51.0
56	IJ6098	3	2016-01-01	2017-12-29	104.0
57	IJ7054	33	2016-01-01	2018-06-22	129.0
58	IJ8050	12	2016-01-01	2018-06-22	129.0
59	IJ8202	4	2016-01-15	2018-06-01	124.0
60	IJ8900	2	2016-01-01	2016-12-23	51.0
61	IJ8993	14	2016-01-01	2018-06-22	129.0
62	IJ9147	1	2016-01-01	2016-12-23	51.0
63	IL8486	3	2016-01-01	2016-12-23	51.0
64	IM0470	1	2016-01-01	2018-06-22	129.0
65	IM0647	3	2016-01-01	2018-06-22	129.0
66	IM4022	3	2016-01-01	2018-06-22	129.0
67	IM5307	24	2016-01-01	2016-12-02	48.0
68	IM6553	12	2016-12-30	2018-06-22	77.0
69	IN5779	15	2016-01-01	2018-06-22	129.0
70	IN7088	3	2016-01-01	2018-03-23	116.0
71	IN7294	2	2016-01-01	2016-12-23	51.0
72	IN8010	1	2016-01-01	2016-12-23	51.0
73	IN8165	23	2016-01-01	2018-06-22	129.0
74	IN8218	4	2016-01-01	2016-12-23	51.0
75	IN8523	7	2016-01-01	2018-06-22	129.0
76	IN8691	3	2016-01-01	2018-06-22	129.0
77	IN8714	38	2016-01-01	2018-06-22	129.0
78	IN8888	24	2016-01-01	2018-06-22	129.0
79	IN8948	66	2016-01-01	2018-06-22	129.0
80	IO9145	12	2016-01-01	2018-06-22	129.0
81	IO9276	3	2016-01-01	2018-06-22	129.0
82	IO9277	1	2016-02-19	2016-12-23	44.0
83	IO9345	54	2016-01-01	2017-12-22	103.0
84	IO9383	5	2016-01-01	2018-06-22	129.0
85	IO9432	17	2016-01-01	2018-06-22	129.0
86	IO9433	3	2016-01-01	2018-06-22	129.0
87	IO9459	6	2016-01-01	2016-12-23	51.0
88	IO9500	27	2016-01-01	2018-06-22	129.0
89	IO9565	9	2016-01-01	2018-06-22	129.0
90	IP0131	7	2016-01-01	2018-06-22	129.0
91	IP4776	41	2016-01-01	2018-06-22	129.0
92	IP5409	6	2016-01-01	2018-06-22	129.0
93	IP5609	11	2016-01-01	2018-06-22	129.0
94	IP7443	6	2016-01-01	2018-06-22	129.0
95	IP7990	9	2016-01-01	2016-12-23	51.0
96	IP7991	37	2016-01-01	2016-12-23	51.0

	SIP	Num of Work Item ID	First Date	Last Date	Num of Weeks
97	IP8352	23	2016-01-01	2018-06-22	129.0
98	IP8489	41	2016-01-01	2018-06-22	129.0
99	IP8543	1	2017-03-03	2017-12-29	43.0
100	IT0189	42	2016-01-01	2018-06-22	129.0
101	IT0425	3	2016-01-01	2018-06-22	129.0
102	MC2055	4	2016-01-01	2016-12-23	51.0
103	MC2061	1	2016-07-29	2016-12-23	21.0
104	MC2069	2	2016-01-01	2016-12-23	51.0
105	MC2080	3	2016-01-22	2016-12-23	48.0
106	MC2121	1	2018-05-18	2018-06-22	5.0
107	MC2276	1	2016-12-30	2018-04-20	68.0

Tabla A.2: Tabla resumen de Work Item Id por SIP, considerando: Primer día, última día y duración en semanas. Fuente: Elaboración propia.

	SIP	Num of Work Item ID	First Date	Last Date	Num of Weeks
0	A3998	2	2016-01-01	2018-06-22	129.0
1	CH1100	2	2016-01-01	2016-12-23	51.0
2	CH1120	11	2016-01-01	2018-06-22	129.0
3	CH1138	17	2016-05-06	2018-06-22	111.0
4	CH1142	4	2016-05-27	2018-06-22	108.0
5	CH1146	3	2016-08-12	2017-12-22	71.0
6	CH1149	9	2016-09-23	2018-06-22	91.0
7	CH1152	4	2016-11-11	2018-06-22	84.0
8	CH1159	1	2016-10-21	2016-12-23	9.0
9	CH1172	4	2017-02-03	2018-06-22	72.0
10	CH1173	4	2017-03-03	2018-06-22	68.0
11	CH1179	2	2016-12-30	2018-03-02	61.0
12	CH1180	2	2018-03-16	2018-06-22	14.0
13	CH1185	22	2017-01-06	2018-06-22	76.0
14	CH1193	2	2017-07-07	2018-06-01	47.0
15	CH1194	2	2017-05-12	2018-06-22	58.0
16	CH1197	4	2017-06-09	2018-06-22	54.0
17	CH1201	1	2017-09-01	2018-06-22	42.0
18	CH1203	1	2017-09-01	2018-06-22	42.0
19	CH1204	1	2017-09-01	2017-12-22	16.0
20	CH1205	1	2017-12-29	2018-06-22	25.0
21	CH1206	1	2017-09-08	2018-01-26	20.0
22	CH1207	1	2017-12-29	2018-06-01	22.0
23	CH1208	5	2017-12-29	2018-06-22	25.0
24	CH1212	1	2017-12-29	2018-06-22	25.0
25	CH1215	37	2017-12-29	2018-06-22	25.0
26	CH1216	2	2018-01-12	2018-06-22	23.0
27	CH1218	1	2018-06-01	2018-06-08	1.0
28	CH1220	2	2018-03-09	2018-06-22	15.0
29	CHAAB8X	13	2016-01-01	2016-12-23	51.0
30	CHAACEN	1	2016-01-08	2016-12-23	50.0
31	CHAACGQ	1	2016-10-28	2016-12-02	5.0
32	CHAACGR	1	2016-01-01	2016-12-23	51.0

	SIP	Num of Work Item ID	First Date	Last Date	Num of Weeks
33	CHAACI1	1	2016-09-16	2016-12-02	11.0
34	CHAACIL	20	2017-12-29	2018-01-05	1.0
35	CHAACJD	6	2017-12-29	2018-06-22	25.0
36	CHAACNZ	1	2017-12-29	2017-12-29	0.0
37	CHAACOR	4	2016-10-07	2016-12-23	11.0
38	CHAACSC	1	2017-10-27	2017-12-22	8.0
39	CHE590	20	2016-01-01	2018-06-22	129.0
40	CHMAWP	2	2016-01-01	2018-06-22	129.0
41	CL0004	2	2016-01-01	2018-06-22	129.0
42	CL0023	2	2016-01-01	2016-12-23	51.0
43	CL0119	3	2016-01-08	2018-06-22	128.0
44	CL0129	3	2016-01-01	2018-06-22	129.0
45	CL0260	33	2016-01-08	2018-06-22	128.0
46	CL0371	4	2016-04-08	2018-06-22	115.0
47	CL0406	1	2016-05-27	2016-05-27	0.0
48	CL0416	1	2016-05-27	2016-12-16	29.0
49	CL0430	7	2016-05-20	2016-12-16	30.0
50	CL0439	2	2017-02-03	2018-01-26	51.0
51	CL0451	4	2016-06-17	2016-12-23	27.0
52	CL0526	6	2016-10-07	2018-06-22	89.0
53	CL0532	6	2016-09-09	2016-12-23	15.0
54	CL0540	18	2016-12-30	2018-06-22	77.0
55	CL0605	1	2016-09-30	2018-06-22	90.0
56	CL310	1	2016-04-15	2016-10-28	28.0
57	CLAEAL	1	2016-11-04	2018-06-22	85.0
58	CLG7KO	5	2016-11-04	2018-06-22	85.0
59	CLNMTA	44	2016-11-04	2018-06-22	85.0
60	CLWKMH	1	2018-06-01	2018-06-01	0.0
61	HC0003	1	2018-04-13	2018-05-25	6.0
62	HV007Y	4	2016-01-01	2018-06-22	129.0
63	HV064L	1	2016-01-01	2018-06-22	129.0
64	IG8698	1	2016-01-01	2016-12-23	51.0
65	IJ6098	3	2016-01-01	2017-12-29	104.0
66	IJ7054	33	2016-01-01	2018-06-22	129.0
67	IJ8050	12	2016-01-01	2018-06-22	129.0
68	IJ8202	4	2016-01-15	2018-06-01	124.0
69	IJ8900	2	2016-01-01	2016-12-23	51.0
70	IJ8993	14	2016-01-01	2018-06-22	129.0
71	IJ9147	1	2016-01-01	2016-12-23	51.0
72	IL8486	3	2016-01-01	2016-12-23	51.0
73	IM0470	1	2016-01-01	2018-06-22	129.0
74	IM0647	3	2016-01-01	2018-06-22	129.0
75	IM4022	3	2016-01-01	2018-06-22	129.0
76	IM5307	24	2016-01-01	2016-12-02	48.0
77	IM6553	12	2016-12-30	2018-06-22	77.0
78	IN5779	15	2016-01-01	2018-06-22	129.0
79	IN7088	3	2016-01-01	2018-03-23	116.0
80	IN7294	2	2016-01-01	2016-12-23	51.0
81	IN8010	1	2016-01-01	2016-12-23	51.0

	SIP	Num of Work Item ID	First Date	Last Date	Num of Weeks
82	IN8165	23	2016-01-01	2018-06-22	129.0
83	IN8218	4	2016-01-01	2016-12-23	51.0
84	IN8523	7	2016-01-01	2018-06-22	129.0
85	IN8691	3	2016-01-01	2018-06-22	129.0
86	IN8714	38	2016-01-01	2018-06-22	129.0
87	IN8888	24	2016-01-01	2018-06-22	129.0
88	IN8948	66	2016-01-01	2018-06-22	129.0
89	IO9145	12	2016-01-01	2018-06-22	129.0
90	IO9276	3	2016-01-01	2018-06-22	129.0
91	IO9277	1	2016-02-19	2016-12-23	44.0
92	IO9345	54	2016-01-01	2017-12-22	103.0
93	IO9383	5	2016-01-01	2018-06-22	129.0
94	IO9432	17	2016-01-01	2018-06-22	129.0
95	IO9433	3	2016-01-01	2018-06-22	129.0
96	IO9459	6	2016-01-01	2016-12-23	51.0
97	IO9500	27	2016-01-01	2018-06-22	129.0
98	IO9565	9	2016-01-01	2018-06-22	129.0
99	IP0131	7	2016-01-01	2018-06-22	129.0
100	IP4776	41	2016-01-01	2018-06-22	129.0
101	IP5409	6	2016-01-01	2018-06-22	129.0
102	IP5609	11	2016-01-01	2018-06-22	129.0
103	IP7443	6	2016-01-01	2018-06-22	129.0
104	IP7990	9	2016-01-01	2016-12-23	51.0
105	IP7991	37	2016-01-01	2016-12-23	51.0
106	IP8352	23	2016-01-01	2018-06-22	129.0
107	IP8489	41	2016-01-01	2018-06-22	129.0
108	IP8543	1	2017-03-03	2017-12-29	43.0
109	IT0189	42	2016-01-01	2018-06-22	129.0
110	IT0425	3	2016-01-01	2018-06-22	129.0
111	MC2055	4	2016-01-01	2016-12-23	51.0
112	MC2061	1	2016-07-29	2016-12-23	21.0
113	MC2069	2	2016-01-01	2016-12-23	51.0
114	MC2080	3	2016-01-22	2016-12-23	48.0
115	MC2121	1	2018-05-18	2018-06-22	5.0
116	MC2276	1	2016-12-30	2018-04-20	68.0

Tabla A.3: Tabla resumen de Work Item ID por SIP modificada, considerando: Primer día, última día y duración en semanas. Fuente: Elaboración propia.

Apéndice B

Gráficos Watson Analytics - MVP

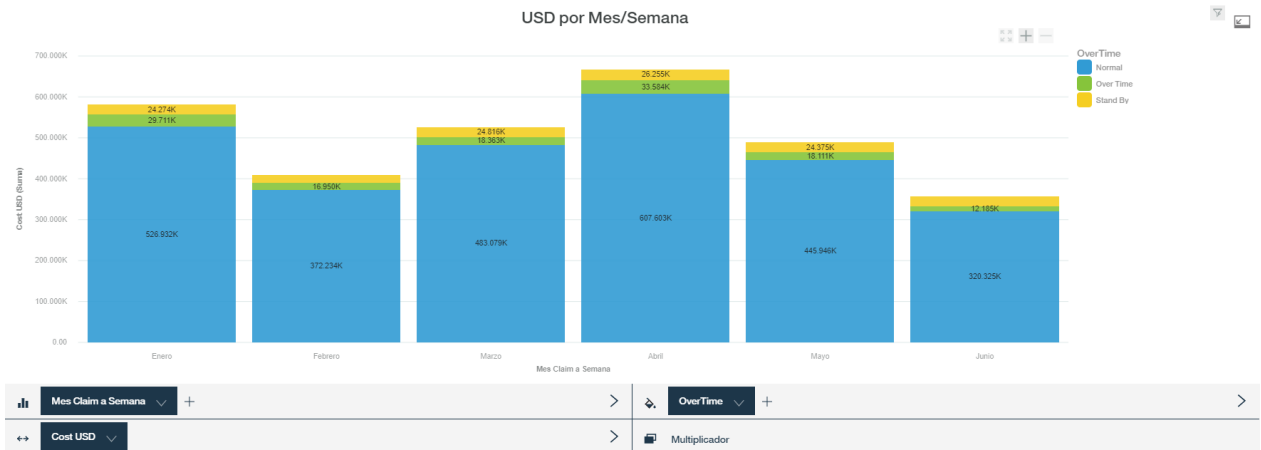


Figura B.1: Vista histórica de Costos en USD por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.

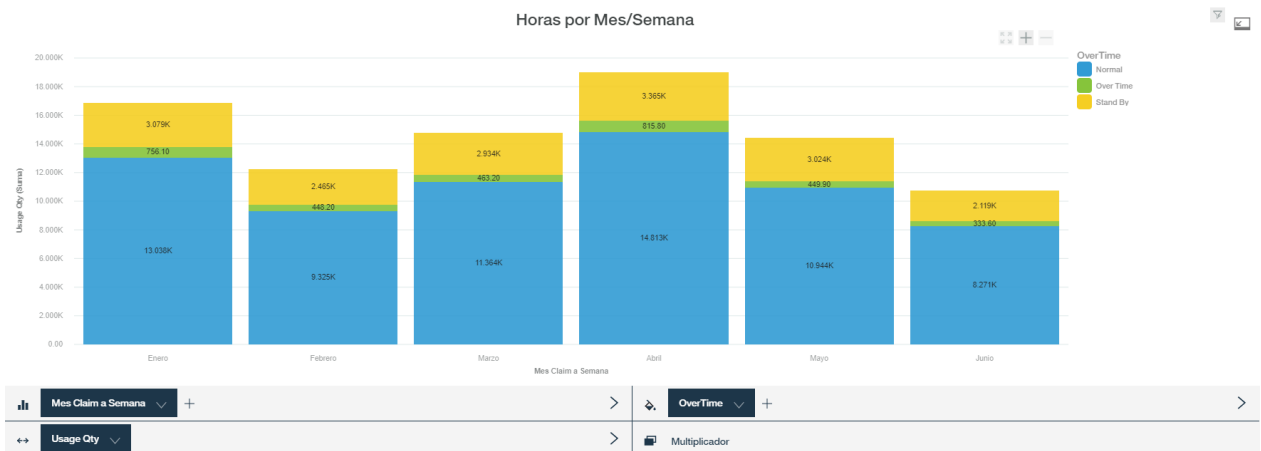


Figura B.2: Vista histórica de horas por mes incluyendo la distinción por tipo de hora. Fuente: Elaboración propia.

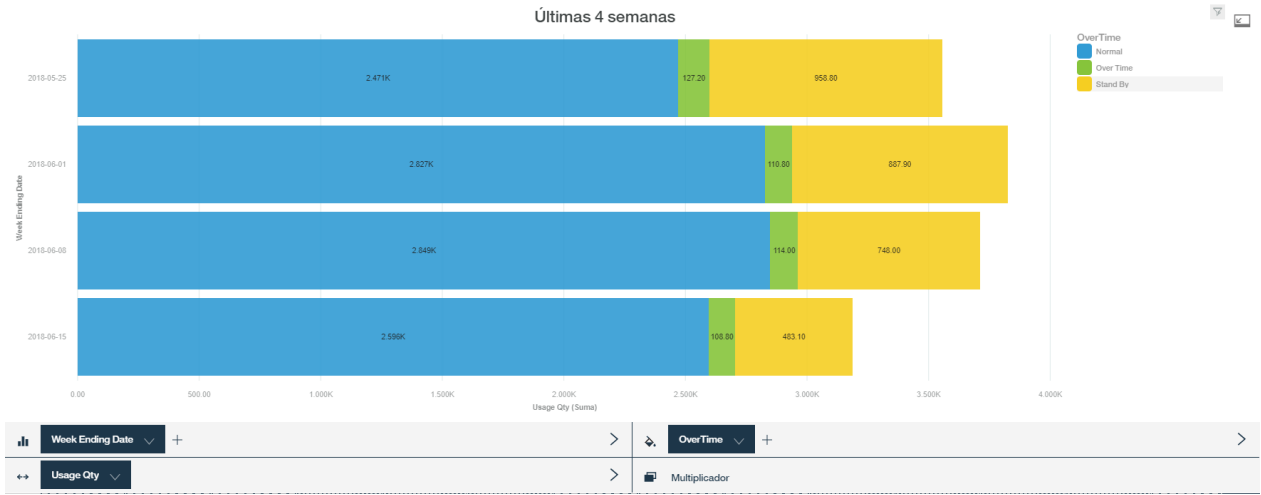


Figura B.3: Vista costos en USD de últimas 4 semanas con distinción por tipo de hora.
Fuente: Elaboración propia.

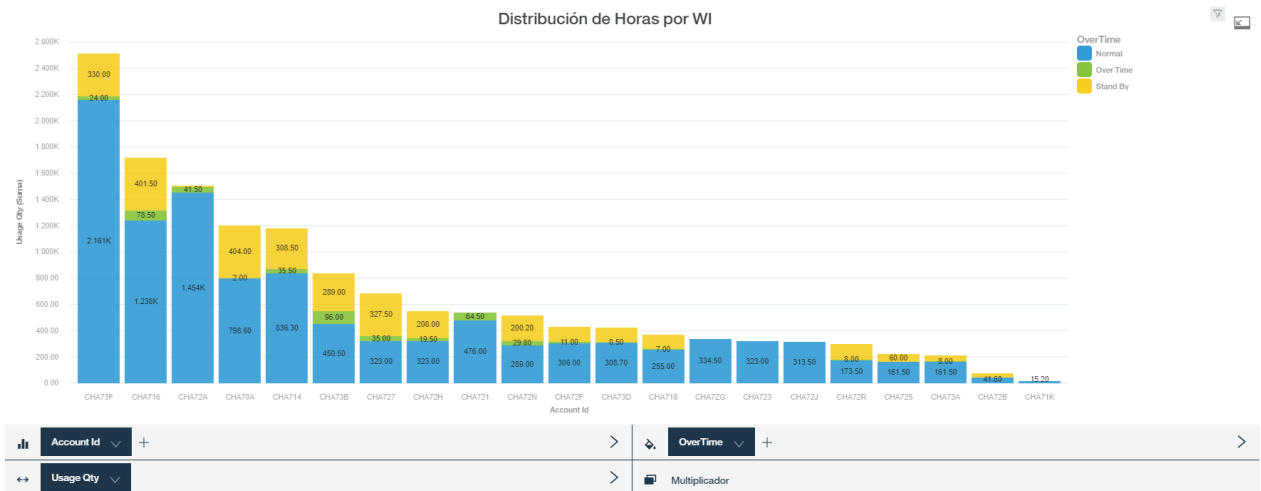


Figura B.4: Costos últimas 4 semanas por Account ID y con diferenciación en tipo de hora.
Fuente: Elaboración propia.

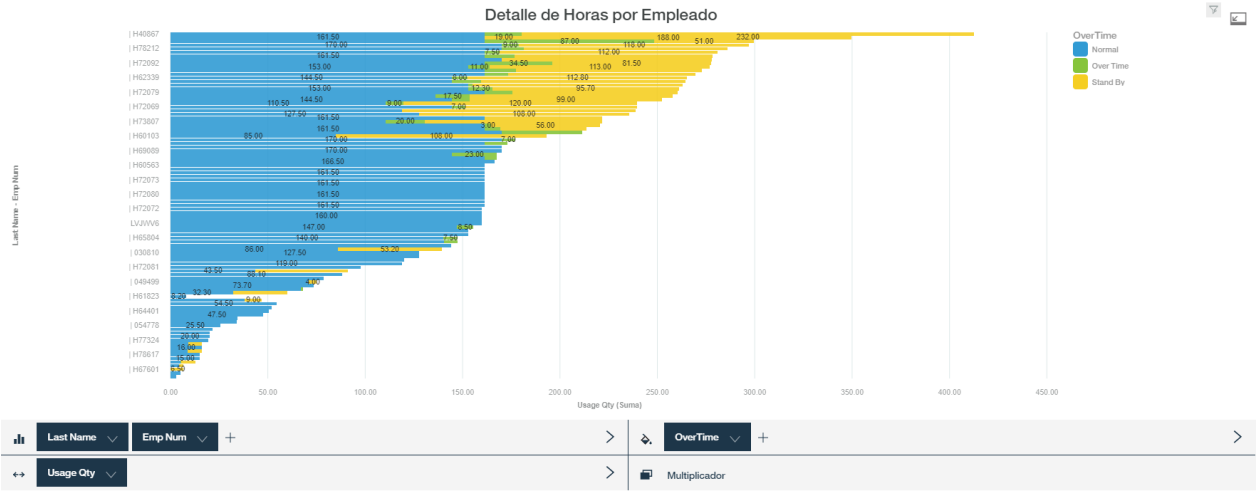


Figura B.5: Detalle de horas cargadas por empleado, haciendo la distinción por tipo de hora cargada. Fuente: Elaboración propia.

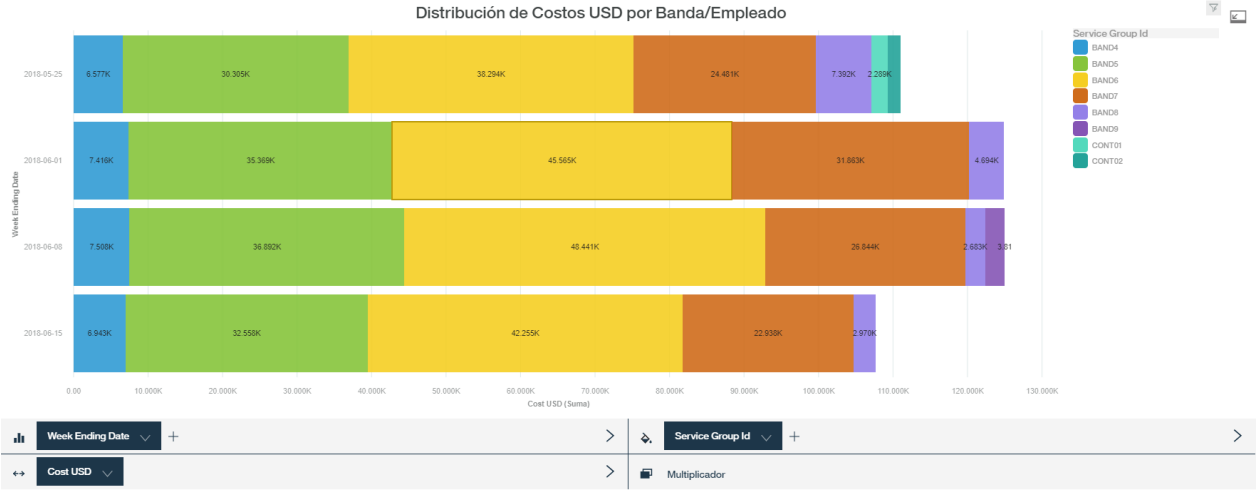


Figura B.6: Distribución de costos en USD por Banda en últimas 4 semanas. Fuente: Elaboración propia.

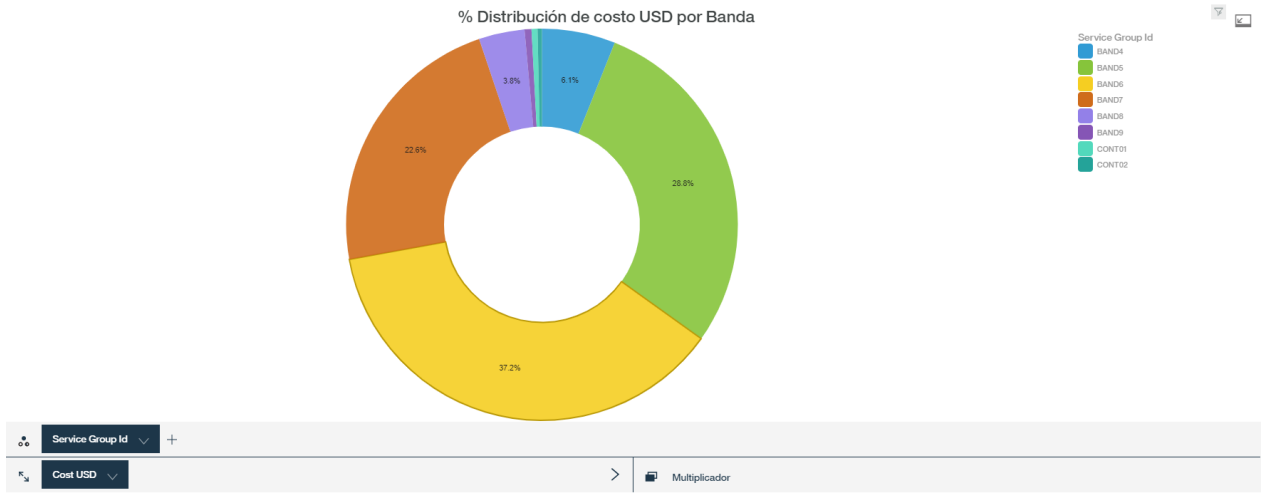


Figura B.7: Distribución porcentual de costos en USD por banda considerando las últimas 4 semanas. Fuente: Elaboración propia.

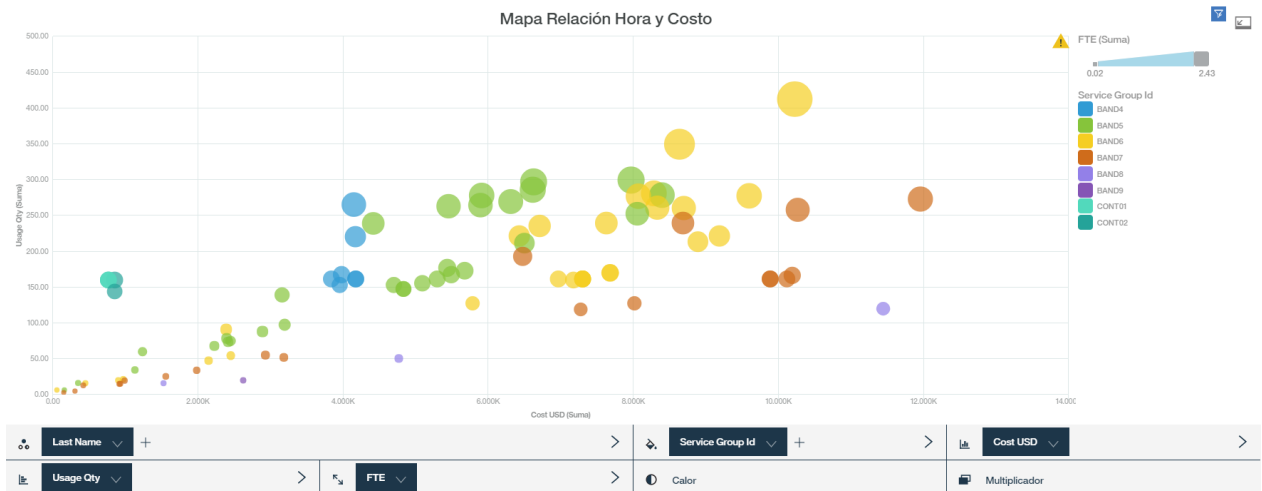


Figura B.8: Mapa relación horas cargadas vs costos asociados a cada empleado. Fuente: Elaboración propia.

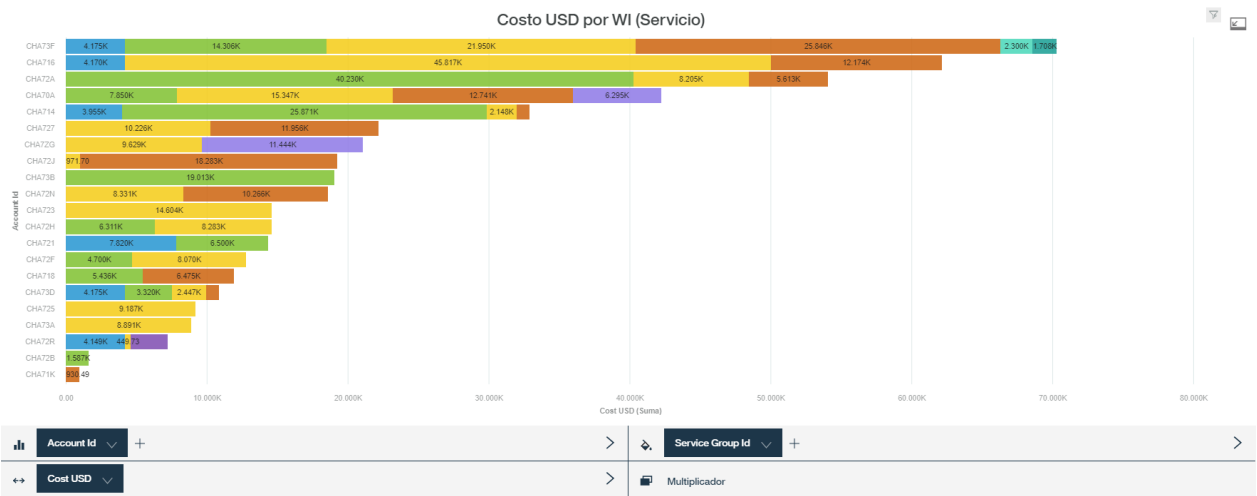


Figura B.9: Costos en USD por Account ID (diferenciado por bandas). Fuente: Elaboración propia.

Apéndice C

Ejemplo de plataforma

C.1. Video ejemplo Watson Analytics

<http://bit.ly/2Vre9bo>

C.2. Ejemplo Watson Studio

<https://ibm.co/2C2oPo6>

Apéndice D

Códigos

D.1. Análisis exploratorio

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/Exploratory%20analysis.ipynb>

D.2. Automatización - Proceso ETL

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/automatizaci%C3%B3n-ETL.py>

D.3. Modelamiento SARIMA

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/SARIMA.ipynb>

D.4. Modelamiento LSTM

D.4.1. Modelo 1

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/MAPE%20RNN-%20Version%201.ipynb>

D.4.2. Modelo 2

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/MAPE%20RNN-%20Version%202.ipynb>

D.4.3. Modelo 3

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/MAPE%20RNN-%20Version%203.ipynb>

D.4.4. Modelo 4

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/MAPE%20RNN-%20Version%204.ipynb>

D.5. Predicción - Producción

<https://github.com/fbahamonde/Trabajodetitulo/blob/master/Modeloproduccion.ipynb>

Apéndice E

Encuesta de uso y satisfacción

¿Qué tan a menudo ocupas esta plataforma?

15 respuestas

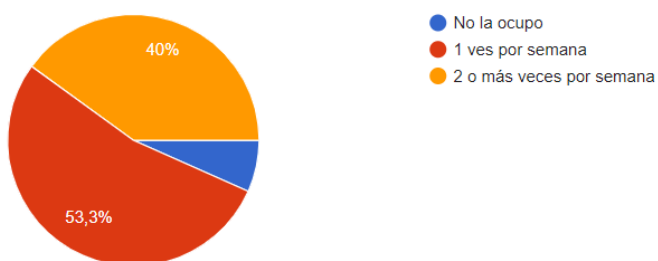


Figura E.1: Mapa relación horas cargadas vs costos asociados a cada empleado. Fuente: Elaboración propia.

Si tu respuesta fue NO, ¿Por qué no?

1 respuesta

nunca logré entender bien como funciona

Figura E.2: Encuesta de uso. Fuente: Elaboración propia.

¿Qué tan satisfecho te encuentras con la plataforma Control Claim?

14 respuestas

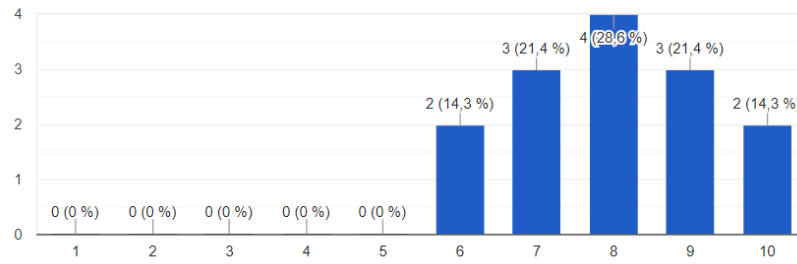


Figura E.3: Encuesta de satisfacción. Fuente: Elaboración propia.

¿Cuánto tiempo has reducido con la implementación de esta herramienta? (Comparado con el método de control que hacías antes)

14 respuestas

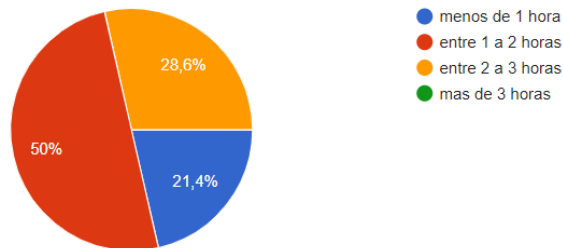


Figura E.4: Encuesta de tiempo reducido. Fuente: Elaboración propia.

¿Compartes el link con tu equipo de trabajo?

14 respuestas

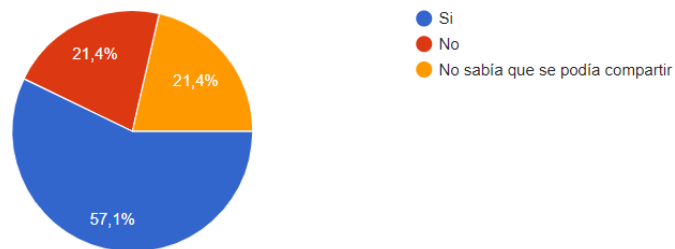


Figura E.5: Encuesta realizada para conocer si los ejecutivos comparten la plataforma. Fuente: Elaboración propia.

¿Qué le agregarías/cambiarías en esta plataforma?
<p>capacidad de analítica y proyección</p> <p>que se pudiera cargar información del BC para hacer la comparación entre BC y Actual</p> <p>tener información de la línea de costo</p> <p>Información respecto a los tickets generados</p> <p>cruzar información con tickets de especialistas</p> <p>Agregar Business Cases</p> <p>línea de costo y tener una vista actualizada en tiempo real</p> <p>Pronósticos de costos</p> <p>Pestaña de análisis de tickets, para conocer el rendimiento de cada especialista</p> <p>Business Case por cada cuenta</p> <p>Análisis de incidentes para ver el impacto que este produce</p> <p>El poder hacer cruces de información para poder determinar incidencias repetitivas y cómo estas impactan en el costo.</p> <p>Dicha Plataforma debería complementarse con los Business Cases de cada DPE</p>

Tabla E.1: Encuesta de satisfacción y uso: Pregunta abierta. Fuente: Elaboración propia.

Apéndice F

Respaldo Aprobación de plataforma

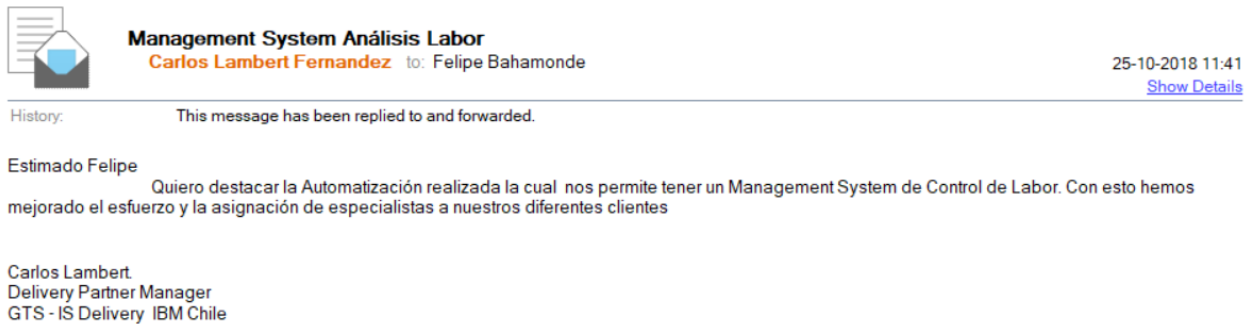


Figura F.1: Respaldo Delivery Partner Manager.

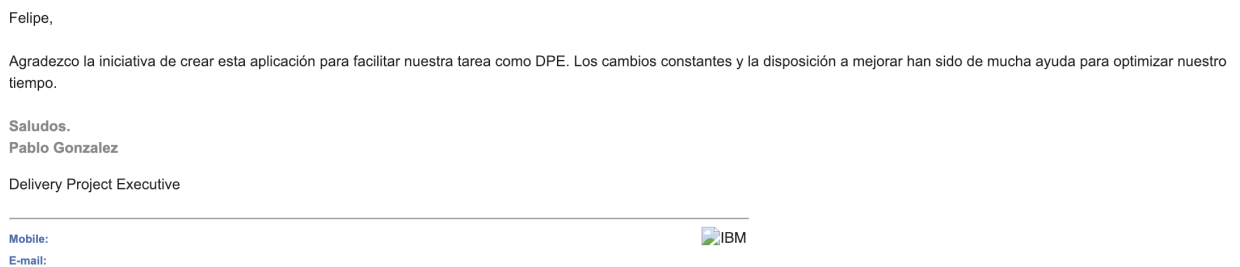


Figura F.2: Respaldo Delivery Project Executive.

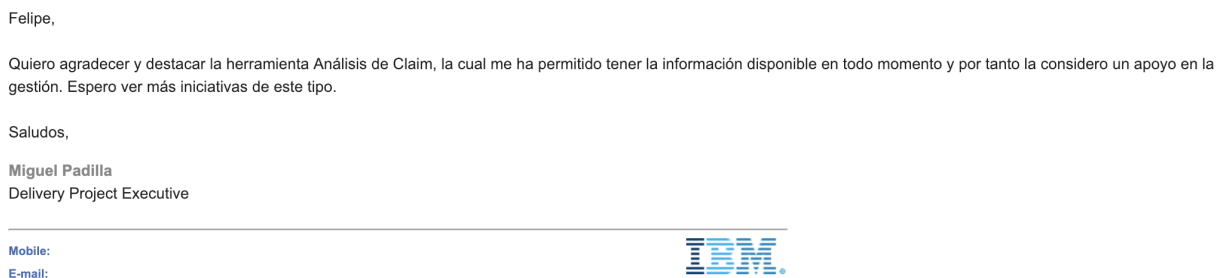


Figura F.3: Respaldo Delivery Project Executive.