



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**DISEÑO E IMPLEMENTACIÓN DE MÉTRICAS E INDICADORES SOBRE
CONSUMO DE CONTENIDO AUDIOVISUAL EN PLATAFORMAS WEB PARA
EMPRESA DE TELEVISIÓN**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

TOMÁS ESTEBAN MORDOJOVICH RUIZ

**PROFESOR GUÍA:
SEBASTIÁN RÍOS PÉREZ**

**MIEMBROS DE LA COMISIÓN:
CARLOS REVECO DÍAZ
LUCIANO VILLARROEL PARRA**

**SANTIAGO DE CHILE
MAYO 2013**

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
POR: TOMÁS MORDOJOVICH RUIZ
FECHA: 31/05/2013
PROF. GUÍA: SEBASTIÁN RÍOS PÉREZ

**DISEÑO E IMPLEMENTACIÓN DE MÉTRICAS E INDICADORES SOBRE
CONSUMO DE CONTENIDO AUDIOVISUAL EN PLATAFORMAS WEB PARA
EMPRESA DE TELEVISIÓN**

Servicios de televisión Canal del Fútbol LTDA distribuye parte de su contenido audiovisual en distintas plataformas web, con un modelo de negocio asociado a cada una de ellas. La Gerencia Multiplataforma, encargada de la gestión de estas plataformas, desea obtener mejor información respecto al consumo del contenido audiovisual publicado para apoyar la toma de decisiones editoriales y comerciales con el fin de mejorar la experiencia del usuario.

El objetivo general del trabajo de título es mejorar la gestión del contenido audiovisual de la Gerencia Multiplataforma de CDF, para aumentar la rentabilidad de éstos mediante el aumento de su consumo y la conversión de usuarios consumidores de contenido gratuito a clientes de productos de pago.

Se implementó un prototipo funcional que entrega una serie de métricas e indicadores que cumplen con apoyar los procesos de toma de decisiones, para lo cual se realizó un análisis basado en los resultados obtenidos. Además, el prototipo permite la navegación en un cubo OLAP con el fin de analizar con un alto nivel de detalle el consumo de videos en el portal CDF.cl, una de las plataformas web gestionadas por la Gerencia Multiplataforma.

El cálculo de las métricas e indicadores definidos en el trabajo resulta útil para generar propuestas concretas en la mejora del contenido audiovisual en ésta área de la empresa. En particular, se estableció una estrategia que ha logrado aumentar el promedio de reproducciones mensuales de videos en el sitio CDF.cl en un 14%, desde marzo del año 2013.

Como trabajo futuro se propone la utilización de los datos recabados en este trabajo en un proyecto de Data Mining que permita segmentar al público de fútbol chileno de acuerdo a su comportamiento de consumo.

AGRADECIMIENTOS

A mi madre, MARÍA SUSANA, por educarme en los aspectos más importantes de la vida, por apoyarme en cada uno de los tropiezos que he dado, y por nunca rendirse en convertirme en una mejor persona.

A mi padre, ALBERTO PATRICIO, por enseñarme a perseverar en los tiempos difíciles, por buscar siempre lo mejor para nuestra familia y por ayudarme a comprender el mundo en que vivimos.

A mis hermanos GERARDO y ÁLVARO, y a mi hermana, SOFÍA, por todos los momentos maravillosos que vivimos en nuestra infancia, y que espero se sigan repitiendo en nuestra adultez.

A mi profesor guía, SEBASTIÁN A. RÍOS PÉREZ, por su enorme paciencia, confianza y dedicación en apoyarme a sacar esto adelante. Sin su colaboración, no habría sido posible.

A los profesores miembros de la comisión, CARLOS REVECO DÍAZ y LUCIANO VILLARROEL PARRA, por su buena disposición y comentarios a mi trabajo.

A FELIPE GONZÁLEZ BARBIERI, por confiar en mí, por clarificar el camino y por haber construido un excelente lugar de trabajo junto a excelentes personas.

A mis compañeros de trabajo en la Gerencia Multiplataforma de CDF, por haberme integrado a su mundo, por enseñarme a trabajar en equipo y por vivir el fútbol con la pasión que nos caracteriza.

Dedicado a la Memoria de
NICOLÁS MORDOJOVICH RAINERI

Tabla de Contenido

1.	Introducción	1
1.1.	Antecedentes generales.....	1
1.2.	Descripción de productos por plataforma.....	3
1.3.	Justificación del trabajo	5
1.4.	Objetivos	8
1.5.	Resultados Esperados.....	9
1.6.	Alcances del trabajo.....	10
2.	Marco Conceptual	10
2.1.	World Wide Web.....	10
2.2.	Análisis de comportamiento de navegación.....	11
2.3.	Sistemas de Almacenamiento y Procesamiento de Datos	12
2.4.	Cambios de Paradigma en la industria de la Televisión	13
3.	Metodología.....	14
4.	Diseño de métricas e indicadores.....	16
4.1.	Descripción de necesidades generales de información	16
4.2.	Diseño de métricas.....	17
4.2.1.	Métricas orientadas hacia mejorar argumentos de venta de espacios publicitarios asociados al contenido audiovisual	17
4.2.2.	Métrica orientada para evaluar el volumen de consumo por plataforma.....	19
4.2.3.	Métricas orientadas a conocer la forma en que los usuarios consumen el contenido audiovisual del sitio.....	20
4.2.4.	Métricas orientadas a conocer el momento en que los usuarios consumen contenido audiovisual del sitio.....	21
4.3.	Diseño de indicadores	21
5.	Construcción de Repositorio de Datos.....	25
5.1.	Diseño de Data Mart	25
5.1.1.	Modelo de datos para almacenar resultados de plataformas.....	26
5.1.2.	Modelo dimensional de datos para obtener métricas de comportamiento de consumo de usuarios de CDF.cl.....	27
5.2.	Fuentes de datos	28
5.3.	Flujo de Datos	30

5.3.1.	Extracción de datos.....	31
5.3.2.	Transformación.....	31
5.3.3.	Carga de datos.....	32
6.	Implementación de Métricas e Indicadores.....	32
6.1.	Construcción de cubos OLAP.....	32
6.2.	Procesamiento de datos.....	34
7.	Análisis de Resultados.....	34
8.	Conclusiones.....	40
9.	Referencias.....	43
10.	Anexos.....	45
A.	Composición demográfica de los Usuarios.....	45
B.	Resultados de Reproducciones de Videos según atributos demográficos de usuarios.....	47
C.	Resultados Comportamiento de Consumo.....	52
D.	Resultados de Indicadores.....	54
E.	Descripción de Tablas correspondientes a las dimensiones del Data WebHouse.....	56

Índice de Figuras

Ilustración 1: Modelo de tres capas Fuente: Elaboración propia.....	11
Ilustración 2: Ejemplo de un Modelo Multidimensional de datos.....	13
Ilustración 3: Ciclo de Vida de Metodología de Kimball Fuente: Kimball Group.....	15
Ilustración 4: Ejemplo de cálculo de métrica AV1[2=FP][A]: Gráfica de consumo mensual de capítulos de Free Pass por rango de edad de los usuarios.....	18
Ilustración 5: Ejemplo de cálculo de métrica AV1[2][C=f]: Gráfica de consumo mensual de videos por parte del público femenino.....	19
Ilustración 6: Gráfica de consumo mensual por tipo de dispositivo y rango de edad de usuarios.....	20
Ilustración 7: Modelo de datos para almacenamiento de resultados.....	26
Ilustración 8: Modelo Data Mart Reproducciones de Video.....	27
Ilustración 9: Flujo de datos para construcción de Cubo OLAP de Reproducción de videos.....	30
Ilustración 10: Construcción de Data Mart en PowerPivot.....	33
Ilustración 11: Construcción de gráficos en PowerPivot.....	33
Ilustración 12: Indicador IAV1 - Gráfico de total de reproducciones de videos en CDF.cl.....	35
Ilustración 13: Reproducciones de video mensuales por Plataforma.....	36
Ilustración 14: Porcentaje de Usuarios por Género.....	37
Ilustración 15: Reproducciones de video por parte del público femenino.....	37
Ilustración 16: Reproducciones de video por Rango de edad del usuario.....	38
Ilustración 17: Métrica AV3[1][G]: Promedio de reproducciones por día de semana.....	38
Ilustración 18: Métrica AV3[1][C]: Promedio de reproducciones por Hora del día.....	39

Ilustración 19: Métrica AV1[2][A=0-20]: Total de reproducciones por subcategoría de video por usuarios menores de 21 años.....	39
Ilustración 20: Porcentaje de usuarios por Rango de Edad	45
Ilustración 21: Porcentaje de usuarios por género	45
Ilustración 22: Porcentaje de usuarios por Estado Civil	46
Ilustración 23: Porcentaje de usuarios por Ubicación	46
Ilustración 24: Reproducciones de 'Compactos' por Rango de Edad.....	47
Ilustración 25: Reproducciones de 'Gol a Gol' por Rango de Edad	47
Ilustración 26: Reproducciones de 'Mejores Atajadas' por Rango de Edad	48
Ilustración 27: Reproducciones de 'Free Pass' por Rango de Edad.....	48
Ilustración 28: Reproducciones de 'La Fecha a Fondo' por Rango de Edad	49
Ilustración 29: Reproducciones de Videos tipo Programas por Rango de Edad	49
Ilustración 30: Reproducciones de Compilados por Rango de Edad	50
Ilustración 31: Reproducciones de video por Género y Subcategoría de Video.....	50
Ilustración 32: Reproducciones de video por Rango de Edad del usuario	51
Ilustración 33: Reproducciones de video por Estado Civil del usuario	51
Ilustración 34: Reproducciones por Tipo de Dispositivo.....	52
Ilustración 35: Reproducciones por día de semana	52
Ilustración 36: Reproducciones por Hora del día.....	53
Ilustración 37: Gráfico de total de reproducciones de videos mensual en CDF.cl.....	55
Ilustración 38: Evolución de consumo de videos por Plataforma.....	55

Índice de Tablas

Tabla 1: Productos de CDF por plataforma de consumo	3
Tabla 2: Resumen de Derechos por competencia y Plataforma	4
Tabla 3: Descripción de contenido audiovisual disponible en productos Multiplataforma	5
Tabla 4: Definición de objetivos y resultados esperados.....	9
Tabla 5: Diferencias entre modelos Relacional y Multidimensional	12
Tabla 6: Métricas de perfilamiento de contenido.....	18
Tabla 7: Definición métrica de consumo de videos por plataforma	20
Tabla 8: Métricas de tecnología de consumo	20
Tabla 9: Métricas sobre ciclo de vida del contenido	21
Tabla 10: Definición de indicador IAV1	23
Tabla 11: Definición de indicador IAV2	23
Tabla 12: Definición de Indicador IAV3	24
Tabla 13: Definición de indicador IAV4	24
Tabla 14: Resumen de principales fuentes de datos	28
Tabla 15: Resultados indicador IAV1.....	34
Tabla 16: Resultados Indicador IAV2.....	35
Tabla 17: Resultados Indicador IAV1.....	54
Tabla 18: Resultados Indicador IAV2.....	54

1. Introducción

1.1. Antecedentes generales

Servicios de Televisión Canal del Fútbol LTDA (en adelante, CDF) es una sociedad fundada en 2003 por la Asociación Nacional de Fútbol Profesional (ANFP¹) y Gestión de Televisión², con el objetivo de rentabilizar de mejor manera los ingresos percibidos por los clubes profesionales de fútbol por concepto de transmisiones de las competencias oficiales de la ANFP y la Federación de Fútbol de Chile.

En 10 años de operaciones, CDF se ha convertido en la principal fuente de ingresos para la mayoría de los clubes de fútbol profesional chileno (que componen la ANFP), jugando un rol muy importante no sólo para la actividad futbolística, sino también en la cultura y cotidianidad de los chilenos, entendiendo al fútbol como una actividad transversal a todos los sectores de la sociedad.

El principal negocio de CDF es vender su señal mediante abonos mensuales, los cuales son ofrecidos por las empresas operadoras de televisión por cable. Este modelo de negocios establece una relación entre empresas que no permite a CDF conocer quiénes son sus clientes finales. Sin embargo, de acuerdo a proyecciones industriales, la distribución de televisión será en el futuro completamente a través de Internet mediante la arquitectura IPTV³ [1], por lo cual los actores se enfrentan a cambios importantes no sólo a nivel tecnológico, sino paradigmático sobre el consumo de televisión.

Otro modelo de negocio importante para CDF es la licitación de los resúmenes de los partidos a canales de Televisión abierta, dándoles derecho a presentar compactos al finalizar una fecha de la competencia de Primera División del fútbol nacional. Esta licitación se realiza periódicamente cada 2 años. Además, CDF percibe ingresos por venta de espacios publicitarios tanto en sus señales de televisión como en los sitios web y aplicaciones de dispositivos móviles que mantiene.

La irrupción de nuevas tecnologías, en particular en materia de calidad, transferencia y consumo de material audiovisual, ha permitido que CDF cree productos con un nuevo modelo de distribución utilizando distintas plataformas de visualización del contenido, los que son mantenidos y administrados por la gerencia Multiplataforma de la empresa, la cual es evaluada anualmente por el desempeño de estos productos en el mercado.

¹ <http://www.anfp.cl>

² <http://www.claroyasociados.cl>

³ Internet Protocol Television.

En ese contexto, es que CDF mantiene su sitio web y otros productos de consumo multiplataforma donde los usuarios pueden acceder a un conjunto de contenidos, en particular videos. Éstos son compactos de los partidos de fútbol disputados, resúmenes de programas emitidos por la señal televisiva o programas desarrollados especialmente para la distribución en multiplataforma.

Este módulo de negocio abre nuevos canales de distribución buscando no sólo posicionar a la empresa como medio digital, sino también combatir la difusión de señales y videos no autorizados en la web y a establecer una comunicación directa con su público. CDF lanzó recientemente el producto *Estadio CDF*⁴, el que consiste en transmitir vía *streaming* a sus clientes abonados los eventos (partidos) en vivo, además de dejar disponible un repositorio de videos con los programas y partidos transmitidos por la señal televisiva. El comportamiento de consumo de los usuarios hacia el contenido audiovisual de CDF.cl puede dar luces sobre cómo mejorar la oferta en esta plataforma.

La empresa no tiene competidores directos, ya que es el único medio con derechos de distribución de las competencias de fútbol nacional chileno. Sin embargo, pueden considerarse competencia otros canales deportivos que emiten fútbol y/u otros deportes. En particular, *Fox Sports LA*, que transmite torneos internacionales en los que participan equipos chilenos y *ESPN*, que transmite gran variedad de ligas internacionales. Estos canales han comenzado a transmitir vía *streaming*, ofreciendo el servicio a sus clientes abonados a través de algún cable operador. Además, han creado aplicaciones para dispositivos móviles, dejando algunos VODs⁵ disponibles para descarga y reproducción.

Un problema para CDF es la difusión de emisiones por streaming directo o *Peer to Peer* (P2P) [2] sin autorización de su señal durante los eventos en vivo de mayor importancia, como los partidos. También, existe el problema de almacenamiento y distribución de videos que utilizan imágenes de CDF sin autorización, que son editados en su mayoría por personas naturales, pero también por portales web no oficiales que buscan generar contenidos propios, editando imágenes capturadas desde televisores.

La incorporación de nuevas tecnologías tanto a la industria de la televisión como a la práctica deportiva ha modificado la manera en que se consume por televisión los eventos deportivos. En particular, la incorporación de estadísticas en vivo sobre el desempeño individual de cada atleta ha sido beneficiosa no sólo para los televidentes, sino también para los entrenadores, reclutadores y representantes de jugadores.

⁴ www.estadiocdf.cl

⁵ *Video On Demand*, es un video almacenado para ser reproducido cuando el usuario desee.

Dado que Internet apunta a experiencias de uso personalizadas, las empresas de televisión deben comprender esto como un desafío, buscando crear señales en vivo o listas de reproducción particulares para cada usuario.

Problemas similares se han tratado en otros trabajos [3], [4], [5] donde se analiza el comportamiento de navegación de los usuarios en portales web, lo cual es relevante para el interés de este trabajo, ya que se busca identificar la manera en que los usuarios hacen uso del contenido audiovisual disponible.

1.2. Descripción de productos por plataforma

Plataforma	Productos	Descripción	Gestión y mantenimiento
TV	Señal Básica	Señal de CDF que se ofrece en la mayoría de los paquetes básicos de las empresas operadoras de cable.	Gerencia Programación y Gerencia Producción
	Señal Premium	Señal que incluye partidos en vivo, y que se vende como un canal adicional a través de los operadores de cable.	
	Señal HD	Es la misma programación de la señal <i>Premium</i> pero el contenido se emite en alta definición. También se vende mediante las empresas operadoras de televisión por cable.	
	Servicios Smart TV LG	La empresa LG adquirió exclusividad para distribución de contenidos de CDF en aplicaciones de televisores Smart TV. Existe una aplicación para televisores LG que sirve como portal de distribución de los contenidos.	Gerencia Multiplataforma
	Servicios VTR	Los clientes de VTR con decodificador de alta definición pueden cargar VODs de la parrilla programática de CDF de forma gratuita.	
Web	CDF.cl⁶	Portal de CDF, se publica contenido editorial (noticias, columnas de opinión u otros) y audiovisual.	Gerencia Multiplataforma
	Estadio CDF	Servicio web para poder ver contenido de CDF en alta definición, incluyendo programas y partidos en vivo y VOD.	
Telefonía Móvil	Servicios Entel	Entel se adjudicó exclusividad para distribución de las imágenes de la competencia de fútbol en redes móviles. Entel vende a sus clientes servicios de alerta de gol (envío de videos de goles mediante mensajería) y señal Básica y Premium de CDF (llamada CDF Móvil).	
	Apps para iPhone, Android, Blackberry	Aplicaciones para teléfonos del contenido editorial de CDF.cl y algunos videos que no se componen por imágenes de la competencia de fútbol nacional.	
Tablet	App para iPad	Aplicación para iPad del contenido editorial y audiovisual de CDF.cl.	
	Estadio CDF	Estadio CDF está disponible en Tablets, pero no en teléfonos por restricciones de contrato con Entel.	

Tabla 1: Productos de CDF por plataforma de consumo

⁶ www.cdf.cl

A diferencia de los productos gestionados por la Gerencia de Programación, en los productos multiplataforma descritos en la Tabla 1 existe la posibilidad para la empresa de realizar mediciones reales (y no estimativas) en cuanto a consumo del contenido. Estas mediciones sirven para definir métricas e indicadores de desempeño de la gerencia y la empresa, y apoyar el proceso de toma de decisiones.

El ámbito del proyecto considera una comprensión global sobre el desarrollo de los eventos ligados a las competencias deportivas de fútbol profesional en Chile y Sudamérica⁷, los cuales componen las variables más importantes al momento de hacer un análisis del impacto que genera el contenido emitido por CDF, dado que su principal activo es el derecho de transmitir estos eventos que son de interés masivo.

Sin embargo, estos derechos de transmisión tienen bastantes aristas, las cuales se han ido definiendo tanto por las empresas interesadas en adquirir estos derechos, y por las federaciones de fútbol, con el fin de maximizar su rentabilidad. Éstas se transforman en restricciones por Plataforma y por competencia, las cuales se resumen en la Tabla 2.

Competencia de Fútbol	Descripción	Productos con Derechos transmisión en Vivo	Productos con Derechos VoD
Primera A	Competencia de los equipos profesionales en Primera División A.	CDF Premium, CDF HD, Estadio CDF, CDF Móvil (Entel)	Entel, CDF.cl, Estadio CDF y App iPad.
Primera B	Competencia de los equipos profesionales en Primera División B.	CDF Premium, CDF HD, Estadio CDF, CDF Móvil (Entel)	Entel, CDF.cl, Estadio CDF, App iPad y App Smartphones.
Copa Chile	Competencia mixta entre clubes de Primera A y B, Segunda y Tercera División	CDF Premium, CDF HD, Estadio CDF, CDF Móvil (Entel)	Entel, CDF.cl, Estadio CDF, App iPad y App Smartphones.
Partidos Selección Chilena	Distintas competencias en las que participa la selección Chilena de Fútbol, incluyendo partidos amistosos.	CDF Premium, CDF HD, CDF Móvil (Entel)	Ninguno

Tabla 2: Resumen de Derechos por competencia y Plataforma

Las restricciones de estos derechos representan un problema en la comunicación hacia los clientes. Además, han sido variables, lo cual dificulta el análisis al no tener resultados comparables en el tiempo.

⁷ Las bases de las competencias nacionales pueden descargarse en <http://www.anfp.cl/bases>. Las bases de las competencias internacionales en las que compiten los clubes chilenos y la selección nacional de fútbol pueden descargarse en <http://www.conmebol.com/reglamentos>.

En el sitio web se encuentra disponible, para todas las personas, un repositorio de videos que son producidos por CDF, siendo éstos los principales productos audiovisuales de la Gerencia Multiplataforma sobre los cuales se desea analizar el consumo. En la Tabla 3 se describen los videos que componen el conjunto de productos audiovisuales sobre los cuales se realizará el análisis de consumo definido en el trabajo de acuerdo a su categoría.

Título de Videos	Descripción	Categoría
Compactos	Resumen de un partido. Competencia Primera A y competencia Primera B.	Compilados Fútbol Chileno
Gol a Gol	Resumen de goles de una fecha del torneo profesional. Para Primera A, Primera B y Copa Chile	Compilados Fútbol Chileno
Mejores Atajadas	Resumen con las mejores atajadas de una fecha del torneo Primera A.	Compilados Fútbol Chileno
EENF	Resumen de capítulos emitidos por señal televisiva del programa <i>En el nombre del Fútbol.</i>	Programas TV
LFF	Resumen de capítulos emitidos por señal televisiva del programa <i>La Fecha a Fondo.</i>	Programas TV
Show de Goles	Resumen de capítulos emitidos por señal televisiva del programa <i>Show de Goles</i>	Programas TV
CDF Noticias	Resumen del noticiario emitido por la señal televisiva	Programas TV
Marca	Spots publicitarios de CDF	Spots
Free Pass	Programa producido para el sitio web	Programas web
Extra Futbol	Programa producido para el sitio web	Programas web
Otros	Videos producidos para acompañar artículos editoriales específicos, o bien, videos virales para publicar en redes sociales	Otros

Tabla 3: Descripción de contenido audiovisual disponible en productos Multiplataforma

Es importante mencionar que todos los productos de CDF siguen una línea editorial definida por la ANFP, en el interés de salvaguardar los intereses de cada club en particular y del fútbol chileno en general.

Estos productos son aquellos sobre los cuales se realizó el trabajo, analizando el consumo que los usuarios le dan.

1.3. Justificación del trabajo

CDF comienza a explorar nuevos negocios mediante la distribución de su contenido en la red. Para gestionar de mejor manera estos productos, es necesario obtener información que apoye el proceso de toma de decisiones a nivel editorial y comercial.

La Gerencia Multiplataforma necesita, para rentabilizar de mejor manera sus productos, que éstos sean más consumidos. Es decir, que los usuarios permanezcan más tiempo navegando por los sitios y mirando videos. De lograr dichos objetivos, los beneficios para la empresa serían: obtener mayores ingresos por venta de espacios publicitarios, mejorar el posicionamiento de la marca en Internet y aumentar la captación de clientes para productos de pago en plataformas online, descritos en la Tabla 1.

Al disponer de contenidos audiovisuales almacenados en Internet y al cual el usuario puede acceder de la forma que prefiera, se marca una gran diferencia con el modelo tradicional de televisión. En este caso, no se emite una señal masiva, sino que es el usuario quien llega a buscar un producto específico en el momento que él disponga o prefiera.

Una de las principales ventajas de una plataforma web de distribución de contenidos es que permite recoger información sobre los usuarios y además, sobre el uso que éstos le dan al producto, lo que representa una fuente de información muy poderosa para realizar un mejoramiento continuo sobre los contenidos y desarrollar estrategias de marketing orientadas hacia la fidelización y captación de clientes.

En ese sentido, es necesario crear contenido audiovisual que logre diferenciarse de lo que el público pueda encontrar en otros sitios de videos o portales de fútbol, para lograr aumentar el volumen de usuarios cautivos. Sin embargo, actualmente no es posible evaluar, por ejemplo, si un proyecto de series de videos ha logrado crear un nicho de usuarios que regresan al sitio para ver las nuevas publicaciones, ya que dicha información no está disponible. A pesar que se sabe que los contenidos de CDF son consumidos principalmente por hombres de todas las edades, es necesario saber con mayor detalle qué tipo de contenidos es mejor o peor recibido por determinados segmentos de la sociedad.

Al mismo tiempo, la Gerencia Comercial de la empresa carece de argumentos de venta potentes para los productos audiovisuales en multiplataforma, ya que no se sabe con certeza si un determinado conjunto de videos es consumido preferentemente por un segmento específico del público del portal web.

La falta de información referente al consumo de las piezas audiovisuales publicadas en el sitio representa entonces un problema para la Gerencia Multiplataforma de CDF, dado que no posee una base sobre la cual tomar decisiones que apunten a mejorar la gestión de los videos.

Las decisiones que se busca apoyar mediante la entrega de nueva información son:

- **Determinar estrategia de publicación de una pieza audiovisual:** ¿En qué plataforma(s) se publicará primero el video? ¿Qué día y a qué hora conviene

publicar un video perteneciente a una determinada categoría? ¿Cuánto tiempo el video será destacado en el portal?

- **Evaluar continuidad de producción de categoría de videos:** El conjunto de videos, ¿ha cumplido los objetivos propuestos? ¿Agregan valor al sitio? ¿Las series de videos, tienen un nicho de usuarios que consumen los videos frecuentemente?
- **Aprobar o rechazar nuevos proyectos audiovisuales:** La propuesta editorial, ¿satisface una necesidad existente en el público de fútbol chileno? ¿Se orienta hacia un segmento del público objetivo que no se ha logrado cautivar?

Otras preguntas que se busca responder son las siguientes:

- ¿En qué momento de la semana se consume mayor cantidad de videos?
- ¿Qué diferencias de comportamiento de consumo existen, según los atributos demográficos de los usuarios?
- ¿Qué usuarios consumen videos con mayor frecuencia?
- ¿Cuáles son los videos más vistos por los usuarios que cumplen cierto perfil demográfico?

Adicionalmente, se apunta a descubrir comportamientos de consumo para cada usuario en particular, para así, por ejemplo, poder aumentar la duración de la experiencia del usuario proponiéndole otros videos de acuerdo sus gustos personales, los que serán determinados de acuerdo a su historial de consumo con el fin de aumentar la duración de la experiencia del usuario, tal como lo hace *YouTube* y la mayoría de los sitios de almacenamiento y reproducción de videos o música.

Por tanto, el interés de este trabajo es responder: quién, cuándo, cómo y qué contenido audiovisual consume. Debido a limitantes en la entrega de información para los productos de cada plataforma, el trabajo se centra en aquellos donde se puede determinar los atributos demográficos de los usuarios, es decir, el portal CDF.cl (ya que se dispone de la base de datos de los usuarios registrados). Además, se pretende presentar y analizar los resultados generales de todos los productos multiplataforma, donde no siempre se dispone de todas las fuentes de datos deseables.

Es importante considerar que este tipo de información no se encuentra disponible para la transmisión televisiva tradicional, ya que para la empresa es imposible establecer una comunicación directa con el televidente. La única herramienta de medición disponible para esta plataforma es el sistema *Time Ibope*⁸, el cual no entrega datos con el nivel de detalle que exige la televisión moderna. El sistema de estimación

⁸ <http://www.peoplemeter.cl/timeibope/rating.asp>

de rating para televisión consiste en la instalación de un dispositivo de registro junto a cada televisor de la muestra, y es el usuario quien debe indicar manualmente qué canal está viendo en cada momento, por lo tanto, a pesar que es útil para realizar estimaciones, no es del todo confiable.

Otra de las oportunidades de la distribución *online* de contenidos es la oportunidad de aplicar técnicas de Data Mining sobre los atributos demográficos de los consumidores para vender publicidad de forma segmentada, ya sea asociada a cada producto, o bien distribuir publicidad diferente para cada individuo [6], dependiendo de la estrategia de venta de publicidad que adopte cada canal.

En [7] se discute sobre la implementación tecnológica de un eventual canal de Televisión cuyo contenido sea personalizado. La industria se acerca a esa realidad a medida que la tecnología se desarrolla y los usuarios se adaptan a ella. Los autores critican proyectos fallidos en este ámbito, debido a la falta de entendimiento del producto deseado por los usuarios, como por malas aplicaciones.

Para obtener información consistente respecto al consumo individual de los usuarios, se trabajó en la implementación de un *Data Web Warehouse*, o simplemente, *Data Webhouse*. Esta herramienta es una base de datos que contiene registros de las reproducciones de video particulares de los usuarios con un alto nivel de descripción de los hechos, lo que permite posteriormente hacer distintos análisis de acuerdo a las necesidades específicas del cliente [8].

1.4. Objetivos

Objetivo General

Mejorar la gestión del contenido audiovisual de la Gerencia Multiplataforma de CDF, para aumentar la rentabilidad de éstos mediante el aumento de su consumo y la conversión de usuarios consumidores de contenido gratuito a clientes de productos de pago.

Objetivos Específicos

- A. Analizar el entorno de la empresa, la dinámica de la industria y determinar necesidades específicas de información que aporten a comprender las variables del mercado.

- B. Crear un listado de métricas e indicadores útiles para analizar el comportamiento de consumo de los usuarios e indicadores que sirvan para evaluar el desempeño de la gerencia Multiplataforma.
- C. Diseñar e implementar a nivel prototipo un sistema computacional que permita almacenar y visualizar las métricas e indicadores.
- D. Generar propuestas de mejora en la gestión del contenido editorial y audiovisual en base a los resultados obtenidos.

1.5. Resultados Esperados

Objetivo General	Mejorar la gestión del contenido audiovisual de la Gerencia Multiplataforma de CDF, para aumentar la rentabilidad de éstos mediante el aumento de su consumo y la conversión de usuarios consumidores de contenido gratuito a clientes de productos de pago.						
Objetivos Específicos	Analizar entorno y estado del arte de la empresa	Crear un listado de métricas e indicadores útiles para analizar el comportamiento de consumo de los usuarios e indicadores que sirvan para evaluar el desempeño de la gerencia Multiplataforma.		Diseñar e implementar a nivel de prototipo un sistema computacional que permita almacenar y visualizar las métricas e indicadores			Generar propuestas de mejora en la gestión del contenido audiovisual
Resultados Esperados	Un análisis de las variables del mercado que afectan el negocio de la empresa.	Definición de las necesidades de información generales y específicas de la Gerencia Multiplataforma.	Un set de métricas e indicadores para apoyar la toma de decisiones con el fin de mejorar la oferta de valor de los productos Multiplataforma.	Diseñar modelo de base de datos para consolidar información recabada.	Un repositorio centralizado de los datos originados mediante el consumo de contenido multiplataforma.	Un prototipo de la herramienta diseñada.	Un análisis de los resultados obtenidos desde el prototipo. Un conjunto de propuestas de mejora en la gestión del contenido audiovisual y nuevos productos para la gerencia Multiplataforma

Tabla 4: Definición de objetivos y resultados esperados

1.6. Alcances del trabajo

El trabajo se enfoca al módulo del sitio más relevante y que representa su diferenciación con otros medios digitales de fútbol chileno: contenido audiovisual capturado durante los partidos y las producciones propias de programas, los cuales se componen de imágenes cuyos derechos son propiedad exclusiva de CDF. Además, se realizará un análisis de la base de datos de usuarios registrados en el sitio, para saber cómo se compone el público de CDF.cl.

Debido a que CDF no dispone de la base de datos de usuarios de los servicios ofrecidos por terceros, las métricas que incluyen componentes demográficas de usuarios serán calculadas solo para la base de reproducciones realizadas por los visitantes del sitio CDF.cl que estén registrados en la base de datos del sitio.

La plataforma web Estadio CDF no es parte de este trabajo, ya que no se dispone de una fuente de datos confiable sobre la navegación de los usuarios registrados.

No se considerará implementar una automatización de los procesos de extracción, transformación y carga de datos para la construcción del prototipo.

Los datos utilizados para la construcción del Data Mart y presentación de resultados del uso de CDF.cl de los usuarios registrados fueron capturados durante el mes de Septiembre de 2012.

Se presentarán los resultados generales de los productos multiplataforma entre los meses de Enero 2011 y Septiembre 2012.

2. Marco Conceptual

2.1. World Wide Web

La web “es un sistema de documentos de hipertexto a la cual se accede a través de Internet” [3]. Estos documentos son textos, sonidos, videos, imágenes u otros. Este trabajo se enfoca en el contenido audiovisual, el cual se masificó en la red con la implementación de reproductores que pueden ejecutarse sobre los navegadores, y al surgir servicios *online* de almacenamiento y distribución de videos como *YouTube*⁹.

La navegación por sitios web se realiza por el usuario, enviando solicitudes generalmente usando un software llamado navegador. Las solicitudes viajan por las rutas de la red hacia los servidores donde se aloja la aplicación del sitio, quien a su vez resuelve la solicitud del usuario, y responde entregando los paquetes de datos correspondientes, que son descargados a la terminal del usuario de Internet y se

⁹ <http://www.youtube.com>

despliegan en el navegador de acuerdo al código HTML¹⁰ de la página. Este es el modelo llamado de tres capas [9], [10].

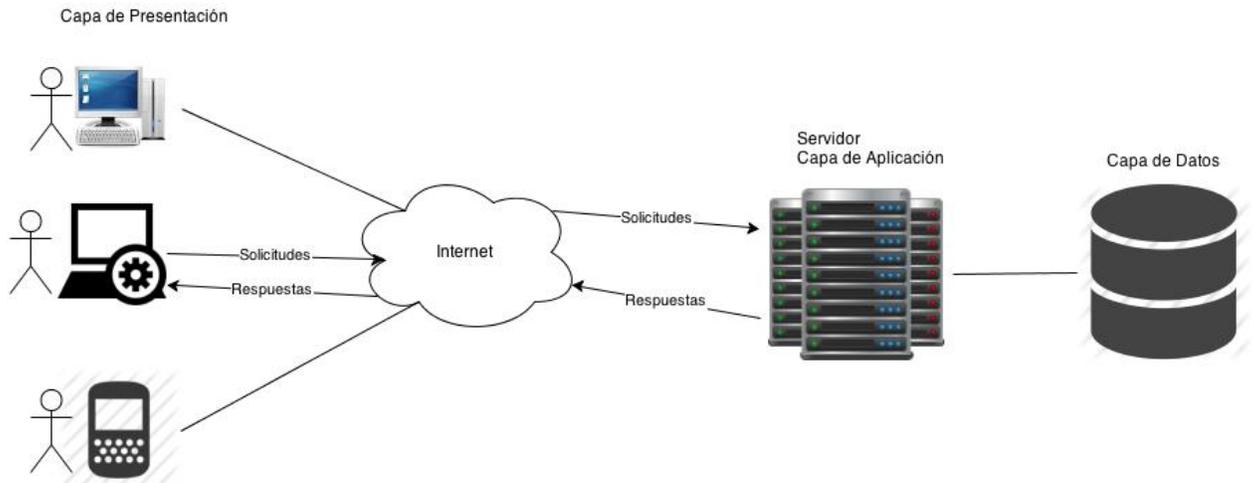


Ilustración 1: Modelo de tres capas

Fuente: Elaboración propia

La evolución histórica de la red ha integrado nuevas tecnologías que permiten navegar desde redes de telefonía móvil, lo cual cambió un paradigma, apuntando a que los usuarios estén siempre conectados a Internet. En general, el uso que se le da a la red varía de acuerdo a la terminal que utilice el usuario. Por ejemplo, un *Smartphone* es muy útil para utilizar aplicaciones de mensajería instantánea, pero la pantalla, al ser pequeña, no resulta cómoda para la reproducción de videos o visualización de documentos de trabajo.

2.2. Análisis de comportamiento de navegación

El aumento de transacciones online ha motivado la implementación de sistemas que permiten analizar las visitas a los sitios web. La justificación radica en que una lectura fina de las acciones que ejecuta un visitante puede responder interrogantes como por ejemplo, la página en la secuencia de navegación en la que desisten de realizar una compra, cuáles son los productos cuyas descripciones han sido leídas más veces, en qué página los clientes pasan mayor tiempo, en qué página terminan la visita, etcétera [8].

Para poder hacer este análisis, se requiere realizar una captura de datos que identifique los eventos que el usuario ejecuta en el sitio web, como por ejemplo, la apertura de una página o el inicio de la reproducción de un video. Así, se construye una fuente de datos confiable que contenga todas estas acciones, para luego procesarla y construir la sesión (visita) [8] de un usuario al sitio. Este registro de datos se llama *web log*, y se compone de escrituras hechas por el servidor web, en la cual escribe registros de lo que cada usuario le solicita.

¹⁰ Hypertext Markup Language

Posteriormente, se procesan los datos recabados en estas fuentes para componer un *Data WebHouse* [8], repositorio de datos que almacena esta información para su posterior análisis, con el fin de mejorar la rentabilidad del e-commerce o sitio web.

2.3. Sistemas de Almacenamiento y Procesamiento de Datos

Las tecnologías de información ofrecen sistemas de almacenamiento de datos que permiten generar información desde los distintos sistemas operacionales, con el objetivo de apoyar de forma confiable y oportuna al proceso de toma de decisiones [11], sin entorpecer el funcionamiento de los sistemas operacionales con los usuarios. Uno de estos sistemas de almacenamiento corresponde a la herramienta llamada Data Warehouse, que se estructura de forma tal que logra solucionar los problemas de heterogeneidad de datos provenientes de distintas fuentes operacionales y del alto costo que significa agregar datos a estas, a causa del enfoque relacional utilizado para guardar datos, y a la diversidad de arquitecturas tecnológicas que conviven en el entorno de las empresas. Un Data Warehouse soluciona estos problemas almacenando en forma consolidada los datos que componen cada unidad de negocio de la empresa.

La estructura de datos que utiliza un Data Warehouse se conoce como “Modelamiento Multidimensional de Datos” [11] y permite realizar consultas flexibles, jerarquizando la información requerida por el analista. En efecto, consiste en un registro de hechos con un alto nivel de descripción de variables. Este sistema se caracteriza entonces por calcular resultados en un mejor rendimiento, y con una mejor calidad y nivel de confianza sobre los datos, resumiéndolo en un solo repositorio de datos provenientes de diversas fuentes [12].

Sistema de datos Relacional	Sistema de datos Multidimensional
Datos no redundantes	Datos con alta redundancia
Muchos usuarios concurrentes del sistema	Pocos usuarios concurrentes
Datos actualizados de acuerdo a la actividad del negocio.	Datos históricos, rara vez modificados. Busca registrar hechos históricos y no actuales, la actualización del sistema es periódica.
Sirve para gestionar operaciones inmediatas de la empresa.	Sirve para obtener información sobre las operaciones históricas de la organización.

Tabla 5: Diferencias entre modelos Relacional y Multidimensional

Fuente: Apuntes del Curso Data Warehousing

Como se mencionó anteriormente, las bases de datos operacionales se estructuran de manera relacional mientras que un Data Warehouse visualiza los datos de manera multidimensional. Por lo tanto es necesario implementar un esquema que permita usar conceptos de análisis multidimensional, construcción de cubos OLAP [11] y aplicarlos

en una base de datos relacional [13]. Este esquema es conocido como Modelamiento Estrella y su esquema general considera una tabla central llamada *Fact Table* (o Tabla de Hechos) y tablas dimensionales, como se puede observar en la Ilustración 2.

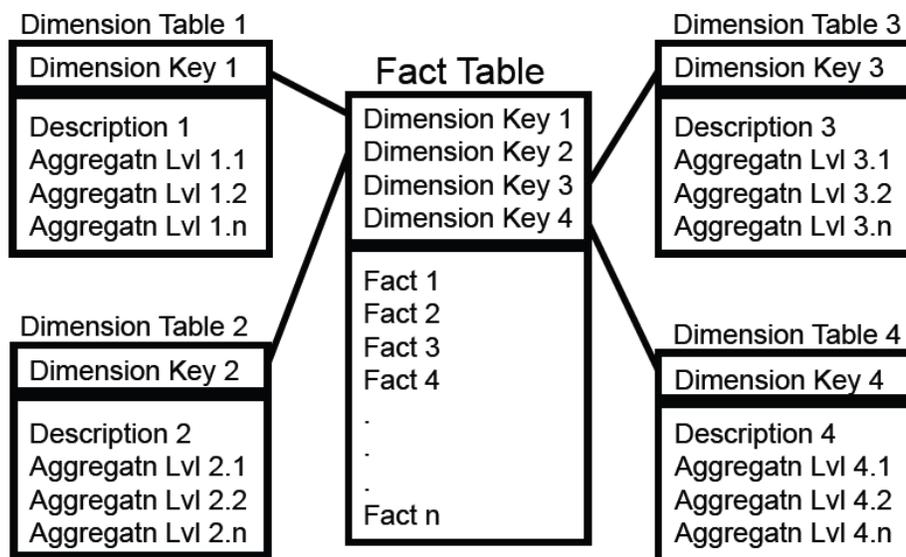


Ilustración 2: Ejemplo de un Modelo Multidimensional de datos

Fuente: Apuntes del curso Data Warehousing

Generalmente los Data Warehouse pueden ser construidos siguiendo dos metodologías distintas. En el primero se construye un gran y único repositorio que centraliza todos los datos provenientes de todas las fuentes operacionales de una organización. Debido a esto, el proceso de construcción es largo y costoso. El segundo enfoque, también conocido como metodología de Kimball, propone la construcción de Data Marts, siendo estos últimas representaciones de los datos de un único proceso de negocios [11]. Para esto, los distintos Data Marts deben ser construidos utilizando dimensiones comunes a distintas tablas de hechos. Por razones de complejidad, de presupuesto y de tiempo se empleó este último enfoque.

2.4. Cambios de Paradigma en la industria de la Televisión

El crecimiento al acceso a banda ancha y redes móviles, acompañado además por las mejoras tecnológicas que permiten una transferencia de datos cada vez más veloz, han permitido la masificación en el consumo de videos en plataformas online.

Como principal cambio paradigmático en la industria de la televisión, esto permite liberar al usuario de las restricciones impuestas por la parrilla programática de la televisión tradicional, debido a que es posible descargar y reproducir un *Video on Demand* (Video en demanda, en adelante VOD) en el momento que el consumidor desee.

Este cambio produce oportunidades y riesgos para cualquier empresa de televisión. En primer lugar, permite establecer un nuevo tipo de relación con sus clientes, mucho más cercana al formato de televisión tradicional. Sin embargo, también prolifera la distribución no autorizada del contenido, las cuales pueden tener incluso fines comerciales.

3. Metodología

El trabajo se compone de una metodología para definir un plan de evaluación de aplicaciones web, y una para el diseño y construcción de repositorios de información. La cual se basa en determinar cuáles son las necesidades de información que tiene la Gerencia Multiplataforma de CDF, las cuales se determinan mediante entrevistas con los editores y jefes del área.

El trabajo debe satisfacer nuevas necesidades de información generadas al interior de la empresa para mejorar la gestión del contenido audiovisual publicado en los productos multiplataforma. Para esto, se necesita obtener datos originados en la web respecto al uso que le dan los usuarios al conjunto de videos producidos por CDF para estos fines.

La metodología aplicada a este trabajo de título fue inspirada, fundamentalmente, en la propuesta por Ralph Kimball en [11] para el diseño e implementación de repositorios de información para Data Warehousing. Esta metodología consiste en centralizar los datos relevantes a un proceso de negocio particular para construir Data Marts. Además, en los trabajos [3], [4] y [5], los autores adaptan ésta y otras metodologías para obtener información a partir de datos generados en la web.

La metodología de Kimball se basa en lo que el autor llama Ciclo de Vida Dimensional del Negocio, y se basa en la comprensión estructural de un proceso de negocio y el mejoramiento continuo de la información entregada por el sistema. En la Ilustración 3 se muestra este ciclo que va desde la comprensión y planificación del proyecto hasta la implementación y mantención del sistema, lo cual generalmente entrega nuevas necesidades de información (debido a la evolución y mejoramiento de la gestión del negocio), lo que significa volver a planificar y ejecutar las etapas de desarrollo necesarias para su obtención.

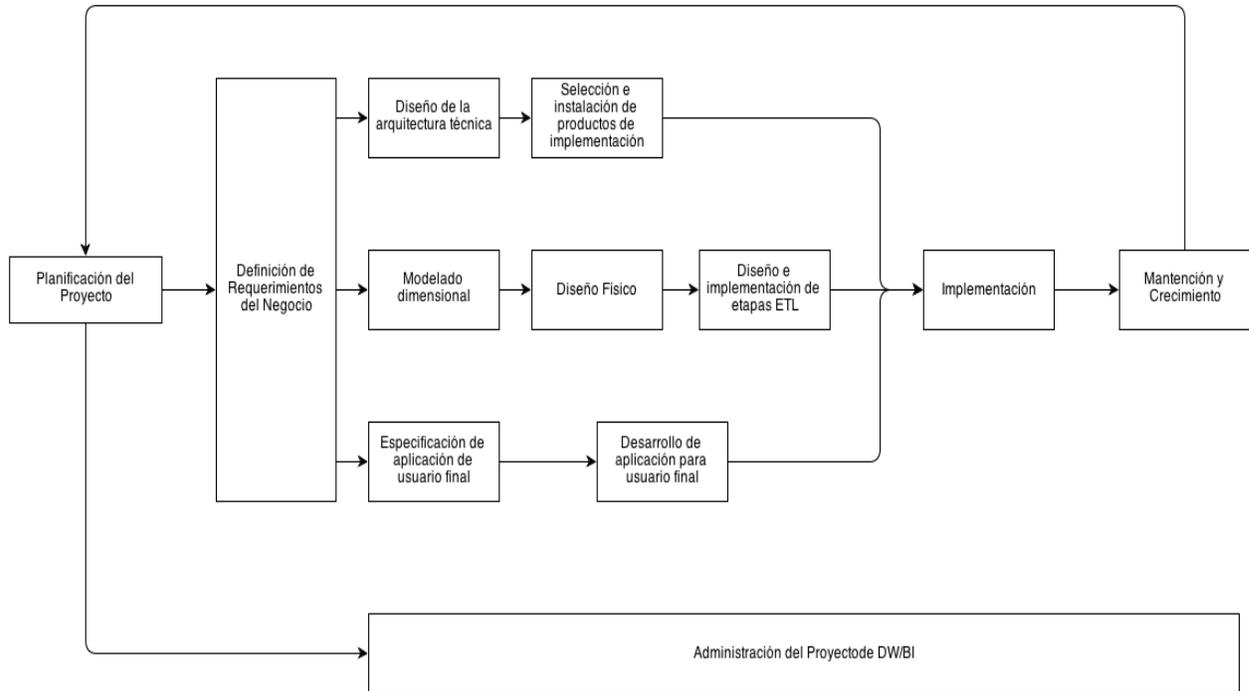


Ilustración 3: Ciclo de Vida de Metodología de Kimball

Fuente: Kimball Group.

Adaptando la metodología de Kimball al marco del proyecto, las etapas de desarrollo del trabajo de título son las siguientes:

- Análisis de la industria de la Televisión y del entorno de la empresa: Etapa introductoria para entender la dinámica de la industria; La misión, visión y valores de CDF, la estrategia y estado del arte de la empresa y su proyección en plataformas web. Para esto, se investiga la historia de la empresa y de la Gerencia Multiplataforma, los cambios en la industria de la Televisión y se analiza la oferta de otros canales de televisión, en particular de deportes, en productos multiplataforma.
- Levantamiento de necesidades de información: Mediante entrevistas al cliente, se determina qué tipo de información es necesaria, es decir, definir el ámbito del proyecto. Posteriormente, se definen métricas e indicadores que respondan a estas necesidades, procurando que éstas sean comparables y consistentes [14].
- Diagnóstico de la situación actual: Esta etapa consiste en crear un mapa general de las fuentes de datos disponibles ligadas a la unidad de negocio, las cuales serán de utilidad para la construcción de las métricas e indicadores propuestas.
- Garantizar captura y transformación de datos necesarios para la construcción del sistema: Asegurar la existencia y accesibilidad de los datos necesarios para la

composición del repositorio de datos, con el fin de tener a disposición todos los elementos necesarios para el cálculo de las métricas e indicadores.

- Desarrollo de Modelo Estructural del Repositorio de Datos: Corresponde al diseño del modelo multidimensional, es decir, el conjunto de tablas y campos que componen el Data Webhouse para la construcción de cubos OLAP y almacenamiento de resultados.
- Procesamiento de Datos: Consiste en realizar todas las transformaciones necesarias a las fuentes de datos para consolidarlos en un repositorio único de información.
- Implementación de un prototipo funcional de la herramienta: Es una maqueta de lo que la herramienta debe satisfacer, incluyendo los reportes en base a una carga parcial de datos.
- Evaluación de resultados obtenidos: Consiste en entrevistar, con el fin de obtener una evaluación cualitativa, a los consumidores finales de la información generada, en este caso, Director Ejecutivo de CDF, Gerente de Multiplataforma y Editor Multiplataforma.

4. Diseño de métricas e indicadores

4.1. Descripción de necesidades generales de información

Como en todo negocio, es importante entender a los consumidores del producto o servicio. Esta comprensión no se acota simplemente a determinar un perfil demográfico específico de los usuarios, sino a analizar el comportamiento de consumo, es decir, poder responder quién, cómo, cuándo, dónde y por qué los clientes o usuarios escogen hacer uso del contenido.

El consumo de contenido televisivo tiene tres dimensiones donde centrar el análisis: Contenido, Espectadores y Tecnología [15].

La primera dimensión corresponde a la oferta de videos, y se describe de acuerdo a la catalogación de cada pieza. Los espectadores son los usuarios del sitio, y se describen de acuerdo a los datos demográficos que han proporcionado en el formulario de registro del sitio. Algunos de los atributos relevantes son: Edad, Ubicación (país, región y comuna), equipo de fútbol del cual es hinchista y género. Finalmente, la tecnología corresponde al dispositivo que utiliza para realizar la navegación, ya sea un

computador, un teléfono móvil, un Tablet o incluso un Smart TV. Además, se incorpora la dimensión temporal, la cual va ligada a los eventos particulares de fútbol chileno, principalmente los partidos del campeonato local y de la selección chilena. Esto hace posible advertir los momentos en los que el contenido se consume para definir su ciclo de vida, es decir, el lapso de tiempo en que un video permanece publicado en el sitio y cuánto permanece destacado en cada nivel.

Un análisis compuesto por estas dimensiones permite concluir no sólo qué segmento de usuarios consume un determinado producto, sino también la manera en que lo hace.

La utilidad final de esta herramienta puede desglosarse en las siguientes necesidades de información, que serán abordadas por las métricas diseñadas en la sección 4.2:

- Conocer el perfil demográfico de los usuarios que consumen videos desde el portal CDF.cl, para vender campañas de publicidad asociadas a cada categoría de videos y perfilar su contenido.
- Detectar patrones de comportamiento de consumo asociados a los atributos demográficos de los usuarios.
- Conocer evolución en el tiempo de consumo del contenido mediante nuevas tecnologías como *Tablets*, *SmartPhones* o *Smart TV*.
- Saber el grado de cumplimiento de la meta asociada al objetivo estratégico de disminuir la producción y distribución no autorizada de videos con contenido cuyos derechos son propiedad exclusiva de CDF.

4.2. Diseño de métricas

Se presentan las métricas diseñadas para satisfacer las necesidades de información descritas en el capítulo 4.1.

4.2.1. Métricas orientadas hacia mejorar argumentos de venta de espacios publicitarios asociados al contenido audiovisual

El objetivo de estas métricas es perfilar el contenido para realizar ventas de campañas publicitarias asociadas a una categoría de videos, como por ejemplo el video de resumen de goles en una fecha del torneo.

ID conjunto de Métricas	Necesidad de Información	Atributo	SubAtributo	Métrica
AV1	Conocer composición demográfica de los consumidores de piezas audiovisuales CDF.cl desde perspectiva de venta de publicidad y perfilamiento de contenido	[1] Reproducciones de video por Categoría de video; [2] Reproducciones de video por Subcategoría.	[A] Rango edad de usuario; [B] Ubicación Geográfica; [C] Género; [D] Equipo favorito; [E] Hora del día.	Semanal
				Mensual
				Anual

Tabla 6: Métricas de perfilamiento de contenido

Como ejemplo, se presentan gráficas de cálculo de algunas de estas métricas a nivel mensual. Se calculó la cantidad de reproducciones de videos hecha por los usuarios registrados del sitio sobre capítulos del programa web *Free Pass* y la subcategoría de los videos vistos por el público femenino.

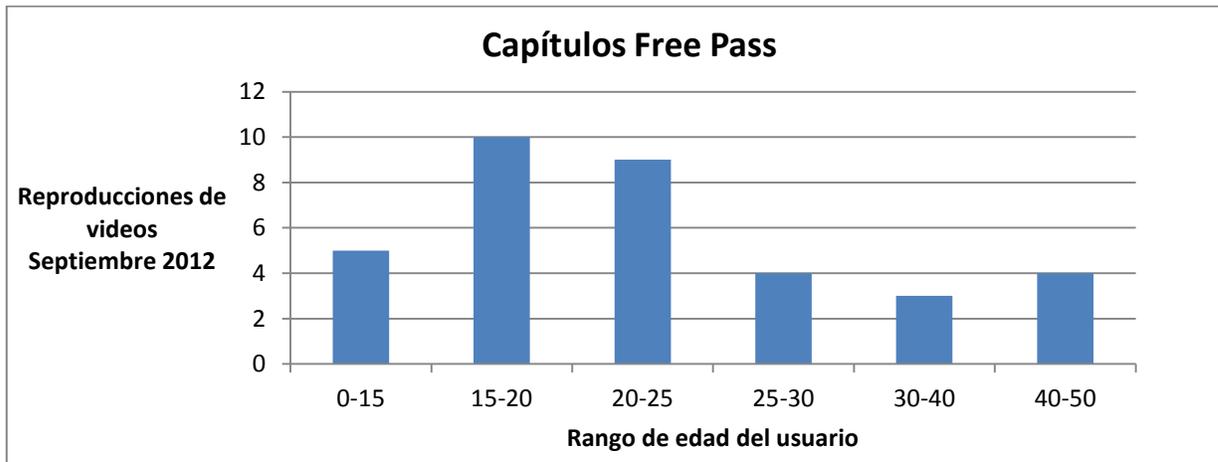


Ilustración 4: Ejemplo de cálculo de métrica AV1[2=FP][A]: Gráfica de consumo mensual de capítulos de Free Pass por rango de edad de los usuarios



Ilustración 5: Ejemplo de cálculo de métrica AV1[2][C=f]: Gráfica de consumo mensual de videos por parte del público femenino

La implementación de estas métricas se justifica al entregar información relevante respecto a la información para tomar decisiones editoriales, así como también para entregar mejores argumentos de venta de espacios publicitarios, ya que se puede identificar qué contenidos son preferidos por algún segmento específico de la población.

Por ejemplo, en el caso de la Ilustración 5 se puede observar que el público femenino reproduce principalmente videos de compactos de goles, y luego videos que son categorizados como “Otros”, que no corresponden a videos que se producen recurrentemente. Estos videos corresponden a pequeñas producciones realizadas con el fin de publicarlas en redes sociales o en notas editoriales particulares que buscan resaltar un hecho puntual ocurrido en la competencia de fútbol.

4.2.2. Métrica orientada para evaluar el volumen de consumo por plataforma

Esta métrica busca mostrar la evolución del uso de cada plataforma, para así adecuar los productos a demandas variables. Esta métrica cobra vital importancia, ya que es una manera representativa de observar cómo la tecnología penetra en el comportamiento de consumo de los usuarios, y da indicaciones sobre qué productos nuevos desarrollar para las plataformas con mayor aumento de demanda.

ID Métrica	Necesidad de Información	Atributo	SubAtributo	Métrica
AVP1	Conocer evolución del volumen de reproducciones por plataforma	[1]Reproducciones de video	[A] Plataforma de reproducción	Semanal
				Mensual
				Anual

Tabla 7: Definición métrica de consumo de videos por plataforma

4.2.3. Métricas orientadas a conocer la forma en que los usuarios consumen el contenido audiovisual del sitio

Para obtener información sobre la manera en que el contenido se distribuye y finalmente, se consume por los usuarios, es necesario analizar el comportamiento a nivel de la tecnología utilizada para reproducir el video, puesto que no es la misma experiencia reproducir un video en el computador que en el teléfono o un televisor.

ID Métrica	Necesidad de Información	Atributo	SubAtributo	Métrica
AV2	Saber la tecnología que usan los usuarios al consumir el contenido audiovisual del sitio	[1]Reproducciones de video [2] Visitas	[A] Categoría de dispositivo de navegación (Agente); [B] App o Web;	Semanal
				Mensual
				Anual

Tabla 8: Métricas de tecnología de consumo

Como ejemplo, se presenta el número de reproducciones realizadas por tipo de dispositivo.

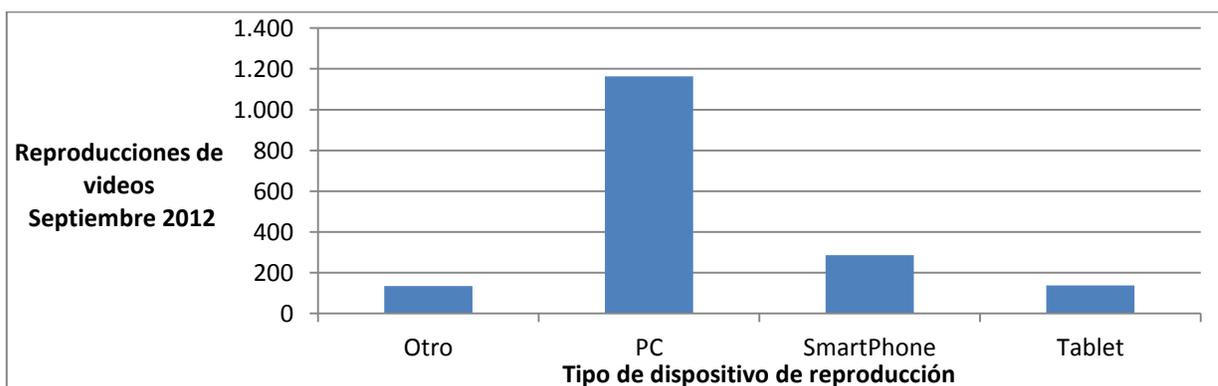


Ilustración 6: Gráfica de consumo mensual por tipo de dispositivo y rango de edad de usuarios

Un análisis más acabado sobre estas métricas incluye el desagregar esta información de acuerdo a los atributos demográficos de los clientes. Algunos de los resultados relevantes a estas métricas se encuentran en el Anexo C.

4.2.4. Métricas orientadas a conocer el momento en que los usuarios consumen contenido audiovisual del sitio

Conocer los momentos en que los usuarios consumen la mayor cantidad de videos permite entender de mejor manera el ciclo de vida de un evento futbolístico o un producto televisivo. Por ejemplo, el consumo de un video compacto de un partido particular, va asociado tanto al partido en sí y a los equipos que compiten, como también a las incidencias particulares que pueden haber ocurrido en dicho evento, como lo son un resultado abultado, grescas u otros acontecimientos.

Esto permitiría optimizar, a largo plazo, el ciclo de vida de los videos en el sentido que pueden derivar tanto en una exposición aumentada o disminuida de acuerdo a las etiquetas de catalogación del video (*tags*).

ID Métrica	Necesidad de Información	Atributo	SubAtributo	Métrica
AV3	Conocer ciclo de vida de los videos de acuerdo a su catalogación y <i>tags</i> .	[1] Número de Reproducciones de video [2] Número de Visitas al Sitio	[A] Fecha; [B] Bloque horario; [C] Hora; [D] Fecha publicación video;	Semanal
			[E] Tags de video; [F] Evento destacado; [G] Día de la Semana.	Mensual

Tabla 9: Métricas sobre ciclo de vida del contenido

Esta métrica permite identificar los momentos en que los usuarios demuestran mayor interés por el contenido, en particular, en los momentos en que se registran los *peaks* y valles de concurrencia.

4.3. Diseño de indicadores

El desempeño del sitio CDF.cl debe ser evaluado desde la totalidad de los visitantes, y no sólo sobre el consumo que hacen sobre el sitio los usuarios registrados en la base de datos. Por lo mismo, se obtienen algunas métricas desde *Google Analytics*, que rastrea la actividad de todas las visitas del sitio, para hacer esta evaluación.

De todas formas, es necesario considerar en el análisis que algunas temporadas generan más visitas y reproducciones de videos que otras. Por ejemplo, se sabe que si

el club Colo Colo (que tiene el mayor volumen de hinchas) queda fuera de competencia en instancias preliminares, mermará los resultados generales para esa temporada.

También es importante destacar que el campeonato cambia su formato de acuerdo a decisiones tomadas por la ANFP, lo cual también debe considerarse, ya que varía la cantidad de partidos y su impacto.

En resumen, las variables más importantes a considerar, por ahora desde el punto de vista cualitativo, son:

- Equipos que se mantienen en competencia.
- Rendimiento deportivo de los equipos protagonistas al momento de protagonizar un evento (partido).
- Tipo de partido (Amistoso, partido de selección, final, regular)
- Formato del campeonato.
- Instancia del campeonato.
- Impacto del partido previo a su realización.
- Impacto del partido posterior a su realización (significancia del resultado, de polémicas originadas, declaraciones a la prensa por parte de los jugadores y miembros de cada club, etcétera).

Estas variables no dependen de la gestión de CDF, pero deben ser incorporadas al análisis comparativo año a año.

Para evaluar el contenido audiovisual del sitio se consideran las siguientes métricas:

Métrica MIAV1: Total de reproducciones mensuales de videos en el sitio CDF.cl.

Métrica MIAV2¹¹: Suma de reproducciones mensuales detectadas de VODs con contenido de CDF subidos sin autorización en sitios de almacenamiento y reproducción de videos, como *YouTube*.

Métrica MIAV5: Total de reproducciones mensuales en todas las plataformas de terceros.

Métrica MIS1: Total de visitas que recibe CDF.cl en un mes.

Métrica MIS2: Total de sesiones de usuarios registrados en el sitio en un mes.

¹¹ Las reproducciones se asocian al mes en las que el video es publicado, ya que no es posible determinar el momento en que se ejecuta cada reproducción, solo se puede conocer el total de reproducciones acumuladas al momento de la denuncia.

Para evaluar si el resultado de cada indicador es considerado bueno, regular o malo, se observó la información del sitio web y otros resultados históricos, además de entrevistar al gerente del área.

Se definen entonces los siguientes indicadores de evaluación:

Indicador IAV1: Variación de reproducciones mensuales en CDF.cl respecto al mismo mes del año anterior

Nombre Indicador	Métrica	Indicador	Resultado	
IAV1	MIAV1	Variación respecto al resultado del mismo mes del año anterior	Menor a -10%	Malo
			Entre -10% y 10%	Regular
			Superior a 10%	Bueno

Tabla 10: Definición de indicador IAV1

Se compara el resultado mensual con el resultado del mismo mes del año anterior para incorporar a la evaluación la estacionalidad de la competencia de fútbol profesional, variable que es fundamental ya que los partidos disputados en cierta fase del torneo generan más impacto que otros.

Indicador IAV2: Diferencia de Porcentaje de reproducciones piratas sobre total de reproducciones

Nombre Indicador	Métrica	Indicador	Resultado	
IAV2	$MIAV3 = \frac{MIAV2}{MIAV1 + MIAV2} \times 100$	Diferencia respecto al mes anterior	Menor a -10%	Bueno
			Entre -10% y 10%	Regular
			Superior a 10%	Malo

Tabla 11: Definición de indicador IAV2

Este indicador busca medir el desempeño en cuanto a la gestión anti piratería y el posicionamiento de CDF.cl como medio digital de fútbol chileno, lo que consiste esencialmente en detectar y denunciar los videos alojados en repositorios de videos públicos que contienen contenido audiovisual de CDF sin autorización (proceso ejecutado por empresa externa que presta servicios tecnológicos a CDF).

Además, la métrica MIAV3 indica si CDF se posiciona como medio único de distribución de contenido de fútbol chileno en internet.

Indicador IAV3: Porcentaje de usuarios que navegan estando registrados

Nombre Indicador	Métrica	Indicador	Resultado	
IAV3	$MIAV4 = \frac{MIS2}{MIS1} \times 100$	Variación respecto al mes anterior	Menor a 1%	Malo
			Entre 1% y 5%	Regular
			Superior a 5%	Bueno

Tabla 12: Definición de Indicador IAV3

Este indicador se asocia a una meta estratégica que apunta a establecer una relación “uno a uno” entre CDF y el público del fútbol nacional. La “Cuenta CDF” será la manera que CDF tendrá para establecer una línea de comunicación con el consumidor final de sus productos y servicios. Hacer un seguimiento a las acciones que el usuario haga en las distintas plataformas permite conocerlo de mejor manera, y así poder diseñar nuevos productos a la medida de cada persona. La importancia de este indicador radica es que a medida que la herramienta de seguimiento capture mayor volumen de datos, más certera es la información que entrega.

Indicador IAV4: Variación de suma de reproducciones en plataformas de terceros:

Este indicador busca evaluar el impacto que tienen los productos audiovisuales de CDF en todas las plataformas, siendo utilizado para evaluar la gestión de las empresas externas que gestionan la oferta de contenidos en otras plataformas (LG, VTR y Entel). El objetivo de CDF es que estos productos se rentabilicen, y por tanto, deben ser cada vez más utilizados por los usuarios. Sin embargo, muchos aspectos, como la comunicación y promoción de tales productos, no son gestionados por la empresa.

Nombre Indicador	Métrica	Indicador	Resultado	
IAV4	$MIAV5$	Variación respecto al mismo mes del año anterior	Menor a 5%	Malo
			Entre 5% y 10%	Regular
			Superior a 10%	Bueno

Tabla 13: Definición de indicador IAV4

5. Construcción de Repositorio de Datos

5.1. Diseño de Data Mart

Se presenta el modelo de base de datos, buscando proporcionar toda la información necesaria para el cálculo de métricas e indicadores propuestos en el capítulo 4. Se ilustran las tablas y sus relaciones, y se define la granularidad de los registros, lo que permite comprender el origen de las métricas que son presentadas en los reportes después de procesar los datos.

Se construyen dos esquemas. El primero de ellos sirve para almacenar los resultados obtenidos en las distintas plataformas, en cuanto a visitas y reproducciones de video, entre otras variables que no forman parte de este trabajo. La idea es tener esta información para poder comparar la evolución del consumo en cada plataforma en función del tiempo.

Por otro lado, se diseñó un Data Mart cuyo grano (nivel de registro) representa una reproducción de video realizada por un usuario registrado en el sitio CDF, con el fin de poder analizar el comportamiento de consumo incorporando la dimensión de usuarios y sus atributos demográficos.

5.1.1. Modelo de datos para almacenar resultados de plataformas

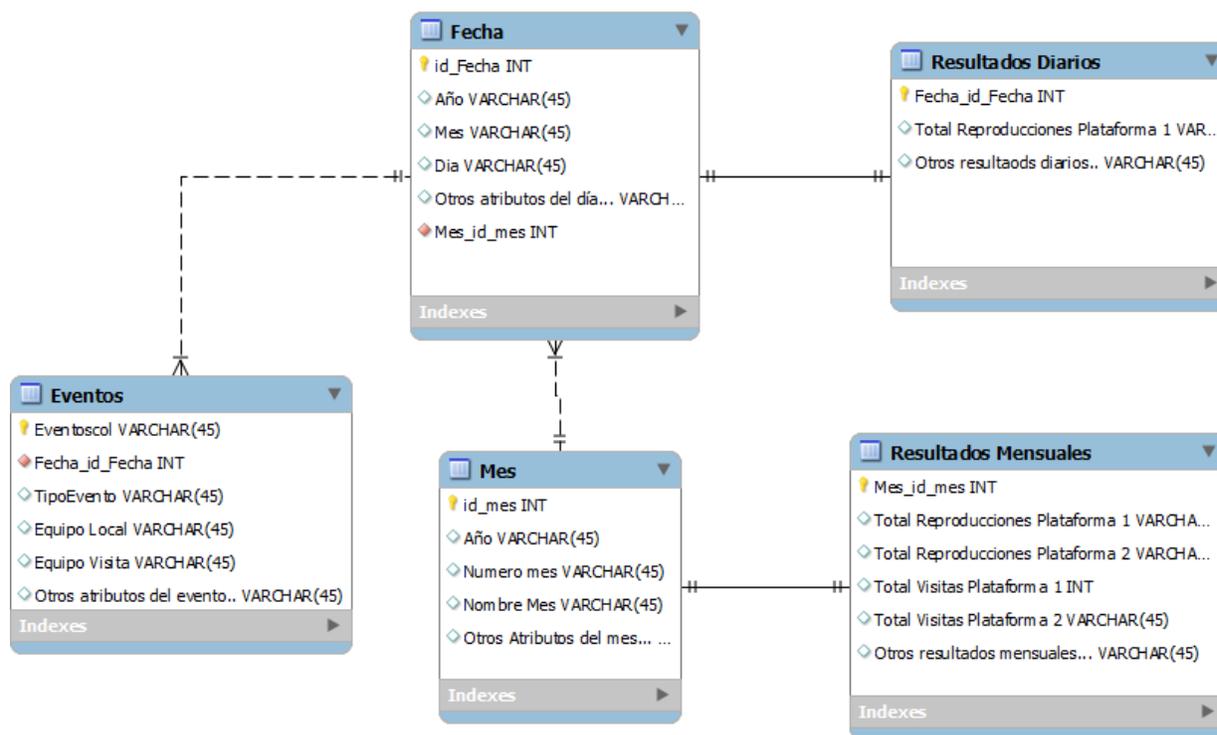


Ilustración 7: Modelo de datos para almacenamiento de resultados

Este modelo de tablas busca consolidar en un mismo repositorio los resultados diarios y mensuales obtenidos en todas las plataformas. En el caso de los resultados diarios, se busca asociarlos a eventos deportivos particulares desarrollados en cada fecha, para poder posteriormente deducir el impacto que éstos eventos representan en el nivel de consumo de las plataformas de reproducción.

5.1.2. Modelo dimensional de datos para obtener métricas de comportamiento de consumo de usuarios de CDF.cl

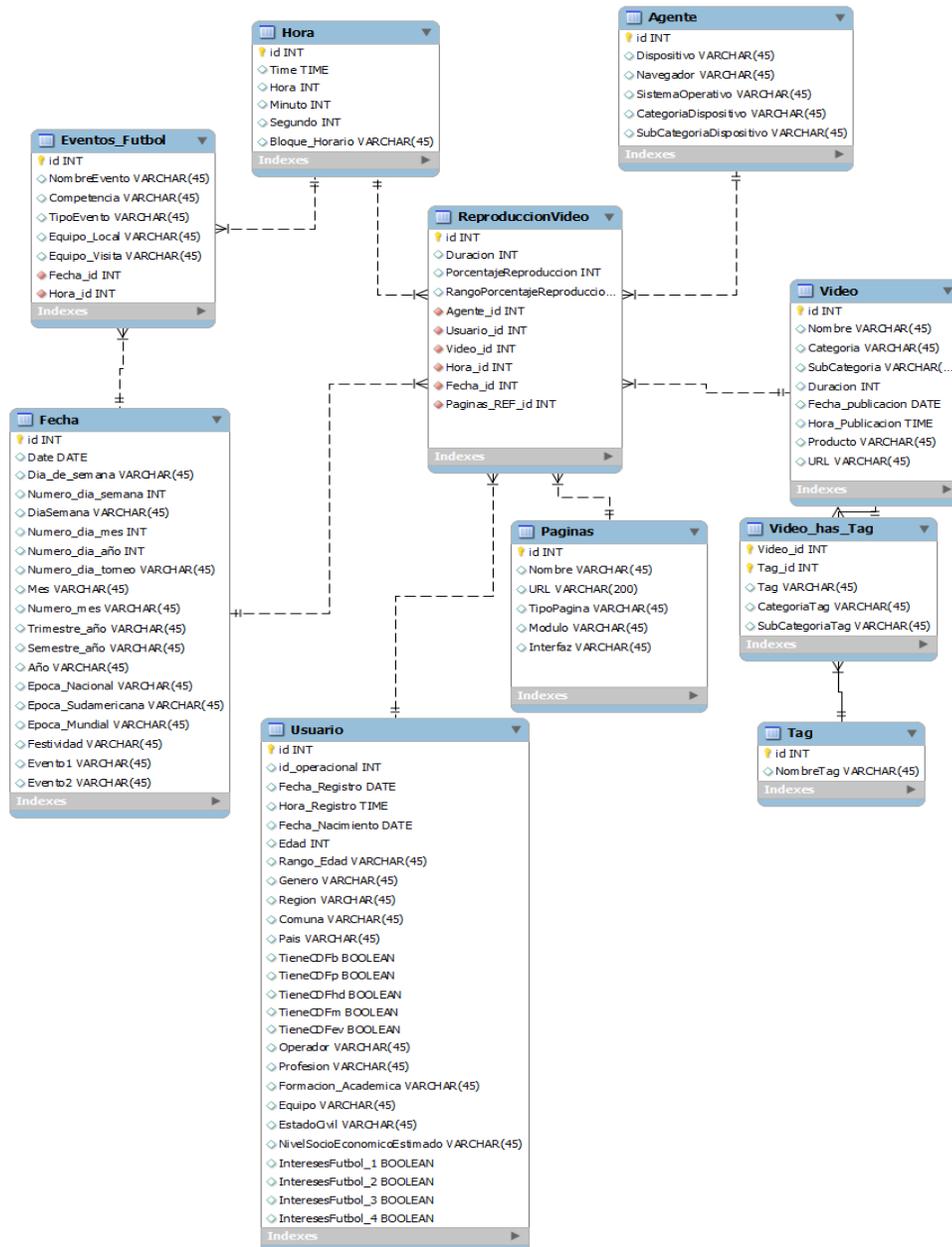


Ilustración 8: Modelo Data Mart Reproducciones de Video

Este es un “Modelo Estrella” [11], el cual se compone de una tabla central que registra cada reproducción de video ejecutada, y de las tablas relacionadas que describen los componentes de cada registro.

Este mismo modelo se replica para otros módulos de análisis, como por ejemplo, las vistas de páginas en el sitio o las sesiones de los visitantes. Para las reproducciones de videos en otras plataformas, aún no se dispone de información de los usuarios, por lo

tanto, pierde sentido realizarlo ya que existen herramientas como *Google Analytics* u otros sistemas de reportería que registran información general de consumo, pero que no permiten analizar según atributos demográficos.

La descripción de cada tabla y sus atributos se encuentra en el Anexo E.

5.2. Fuentes de datos

El sistema debe alimentarse de datos verídicos, los cuales se obtienen desde distintas fuentes. En la Tabla 14 se describe en forma resumida el origen de los datos que componen el repositorio diseñado en el capítulo 4.

Fuente	Origen de los datos	Datos	Utilidad
Google Analytics	Registra actividad del total de visitantes del sitio y de aplicaciones móviles.	Total de visitas, visitantes exclusivos, páginas vistas, duración media de la visita, agente, reproducciones de video.	Implementación de indicadores.
Administrador de Videos	Catalogación interna de videos al momento de su publicación.	Título, tags, fecha y hora subida, duración, URL	Obtener registros y metadatos de los videos.
Base de datos de Usuarios	Formulario de registro, completado por los usuarios al registrarse.	Nombre, fecha registro, fecha nacimiento, equipo favorito, género, región, comuna	Obtener información demográfica de los usuarios que ejecutan reproducciones de video.
Weblogs (reproducciones de videos)	Servidor del reproductor de videos escribe registros cuando un usuario ejecuta una reproducción.	Usuario, video, Agente (terminal donde se realiza la reproducción), Fecha, Hora, Tiempo de reproducción	Registros de una reproducción, que permite construir la tabla de hechos del Data Mart.
Reportes mensuales de servicios externos	Reportes mensuales de empresas que distribuyen contenido de CDF en sus plataformas.	Total de visitas a plataforma de distribución y de reproducciones de videos.	Conocer nivel de uso de estos servicios.
Reportes de gestión Anti Piratería	Reporte periódico por parte de empresa externa.	Título del video, fecha de subida, fecha de bajada, reproducciones hasta la fecha de bajada.	Cálculo de métrica MIAV2.

Tabla 14: Resumen de principales fuentes de datos

Se identifican 6 grupos de fuentes de datos:

- *Google Analytics*: Contiene información referente a la navegación del total de usuarios, incluyendo secuencia de páginas vistas, fecha y hora de las visitas, fuentes de tráfico (de dónde vienen los visitantes), duración de la visita, entre otros. No entrega datos demográficos de los usuarios, más allá de su distribución geográfica (deducida de acuerdo a la dirección IP) e idioma de instalación del sistema operativo
- Administrador de contenido audiovisual: Corresponde a los datos y metadatos de los videos publicados en el sitio, como por ejemplo, el título del video, su duración, categorías y etiquetas.
- Base de datos de usuarios: Corresponde a los datos de los Usuarios registrados en el sitio. Cada usuario debe entregar los datos completando un formulario de registro que solicita información personal de los Usuarios necesaria para el desarrollo del Data Warehouse.
- *Web logs*: Contiene todos los datos correspondientes a la navegación dentro de CDF.cl. Cada registro contiene, entre otras cosas, una identificación (en caso de existir) del usuario que realiza la acción, la fecha y hora al momento de solicitar algún objeto web, el objeto web que solicita el usuario (se identifican aquellas acciones que corresponden a reproducciones de video). Existe también un registro de la hora en que termina la transferencia de datos, lo que permite calcular posteriormente la duración de la reproducción.
- Reportes de servicios de terceros: Son reportes que se reciben mensualmente respecto al uso de las plataformas de terceros (Entel, VTR, LG) por parte de la empresa externa que distribuye contenido de CDF.
- Reportes de Gestión Antipiratería: Son reportes que llegan periódicamente y que informan a CDF de los contenidos audiovisuales distribuidos sin autorización por terceros, los cuales son denunciados y eventualmente dados de baja. La información incluye el listado de videos, junto con la cantidad de reproducciones hasta el momento de realizada la denuncia.

5.3. Flujo de Datos

Para la construcción del repositorio de datos diseñado, es necesario capturar datos que se generan mediante la administración del sitio web y la navegación de los usuarios sobre el mismo.

En la Ilustración 9: Flujo de datos para construcción de Cubo OLAP de Reproducción de videos se presenta el flujo de los datos, desde su origen, ya sea por la escritura de un archivo de log por parte del servidor, o almacenamiento de los metadatos en el administrador de contenido audiovisual y base de datos de usuarios, hasta su almacenamiento en el repositorio de datos.

Por otro lado, los resultados diarios y mensuales de CDF se recopilan desde otras fuentes de datos, las cuales se originan en la web (*Google Analytics* para el sitio web como para aplicaciones móviles) o de reportes que provienen desde empresas externas que distribuyen contenido audiovisual de CDF.

Estos datos también deben ser procesados para consolidarlos en el mismo sistema de información y trabajar en un solo ambiente.

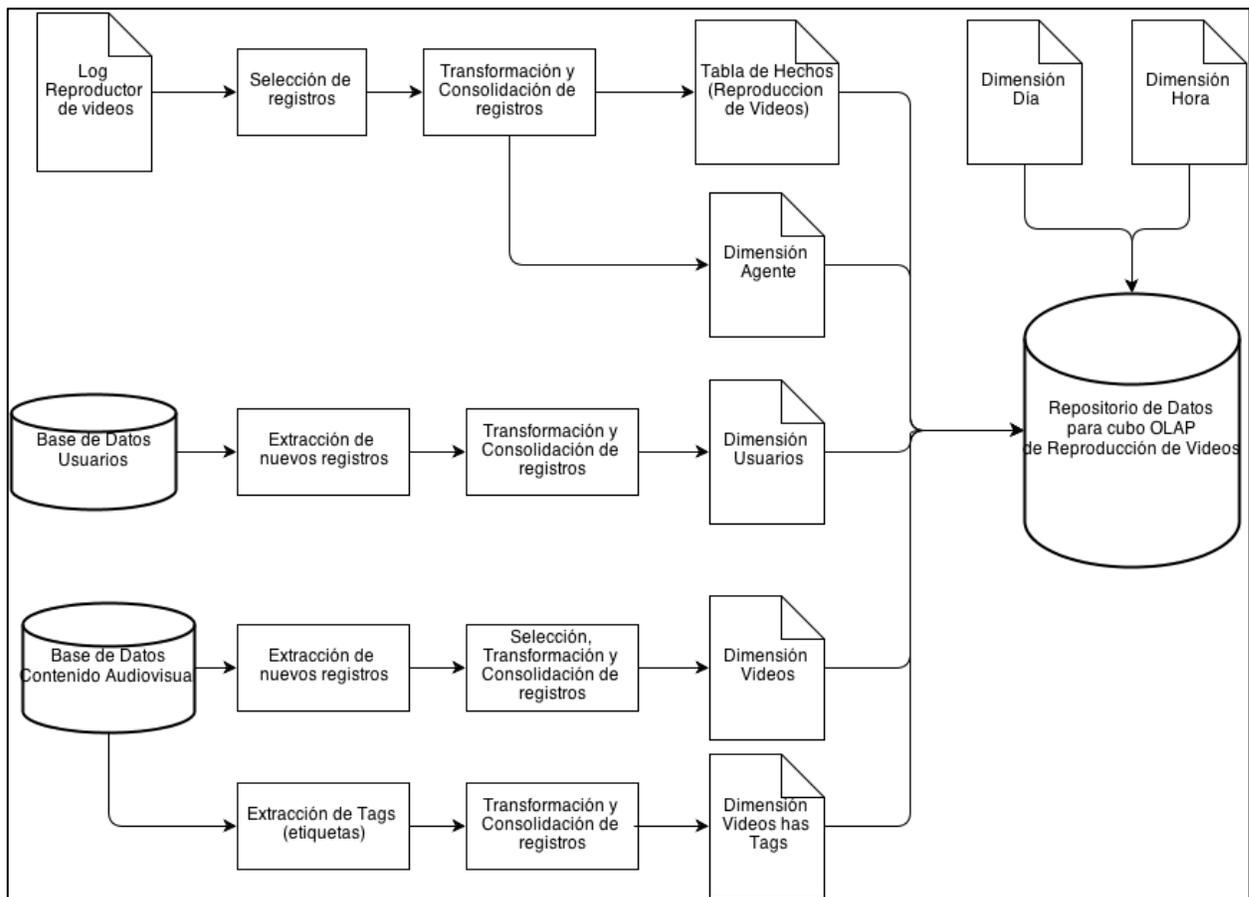


Ilustración 9: Flujo de datos para construcción de Cubo OLAP de Reproducción de videos

Para calcular las métricas e indicadores propuestos en 5.1, es necesario obtener fuentes de datos y procesarlas para que finalmente se consoliden en el repositorio de datos. Estas etapas son por lo general las más costosas de los proyectos de *Data Warehousing*, y significan aproximadamente el 80% del costo [13].

A continuación se hace un breve repaso de las etapas ETL (*Extract, Transformation and Load*) sin entrar en detalles de la implementación técnica. Para esto, se recomienda estudiar los trabajos [3], [5], [16] donde se detallan los módulos informáticos necesarios para garantizar el flujo de los datos en forma automatizada.

5.3.1. Extracción de datos

Para implementar el prototipo de la herramienta, se instaló un sistema de captura de datos de navegación en cada página del sitio y, en particular, en el reproductor de videos (*jw-player*¹²), desde donde se originan los datos más relevantes a este trabajo. También se modificaron campos en el formulario de registro de los usuarios a la base de datos y se consolidó información obtenida desde *Google Analytics* y otras fuentes de datos, como los reportes de gestión anti piratería o informes mensuales respecto al consumo de productos de CDF a través de pantallas de terceros (VTR, Entel, LG).

La estrategia adoptada para completar las etapas de extracción es crear copias periódicas de los nuevos registros en estas bases de datos y cargarlas en el *Data Staging Area*, desde donde se transformarán los datos. Este proceso tiene lugar una vez al día, la cual se ejecuta a las 04 AM. De esta forma, se busca no entorpecer el funcionamiento normal del sitio, el cual tiene menor flujo de visitantes a esa hora.

5.3.2. Transformación

El proceso de transformación no implica mayores cambios a los datos para poder crear las tablas que conforman la estructura del Data Warehouse dado que la calidad de los datos por defecto es satisfactoria y la cantidad de registros nuevos en el *weblog* alcanza como máximo un total de 300 mil reproducciones cada mes.

La transformación más significativa que se realiza consiste en la eliminación de registros correspondientes a videos que fueron cargados en el sitio a modo de prueba. A su vez, se eliminan aquellos registros con errores, con tal de mantener una calidad de datos elevada, y se asocian los atributos que corresponden a llaves foráneas de otras tablas.

¹² <http://www.longtailvideo.com/jw-player/>

Para la implementación del prototipo, se trabajó realizando modificaciones de formato de los campos en Excel, generando finalmente archivos de tipo CSV¹³ que componen las tablas del Data Mart y de resultados generales.

5.3.3. Carga de datos

El proceso de carga es sencillo en comparación a otros casos de Data Warehouse, ya que no requiere comparar la información en el Data Warehouse con la información almacenada en el DSA para cargar los registros nuevos. Esto sucede debido a que los datos extraídos en el DSA corresponden a los nuevos registros y una vez que estos han sido cargados en el Data Warehouse, el DSA es borrado [8].

La carga se realiza periódicamente cada noche (para evitar saturar los servidores de administración del contenido). Así, el cliente del sistema de información disponga todos los días de los datos consolidados hasta el día anterior.

6. Implementación de Métricas e Indicadores

6.1. Construcción de cubos OLAP

Utilizando la herramienta *PowerPivot*, extensión de *Microsoft Excel 2010*, se construyeron los cubos OLAP de acuerdo al modelo de base de datos descrito en el capítulo 5.

En efecto, después de cargar los archivos con las tablas que componen la base de datos, se configura la estructura de la misma de forma a replicar el modelo estrella (Ilustración 10).

¹³ *Comma Separated Values*, es un archivo de texto que separa los valores de cada celda usando comas.

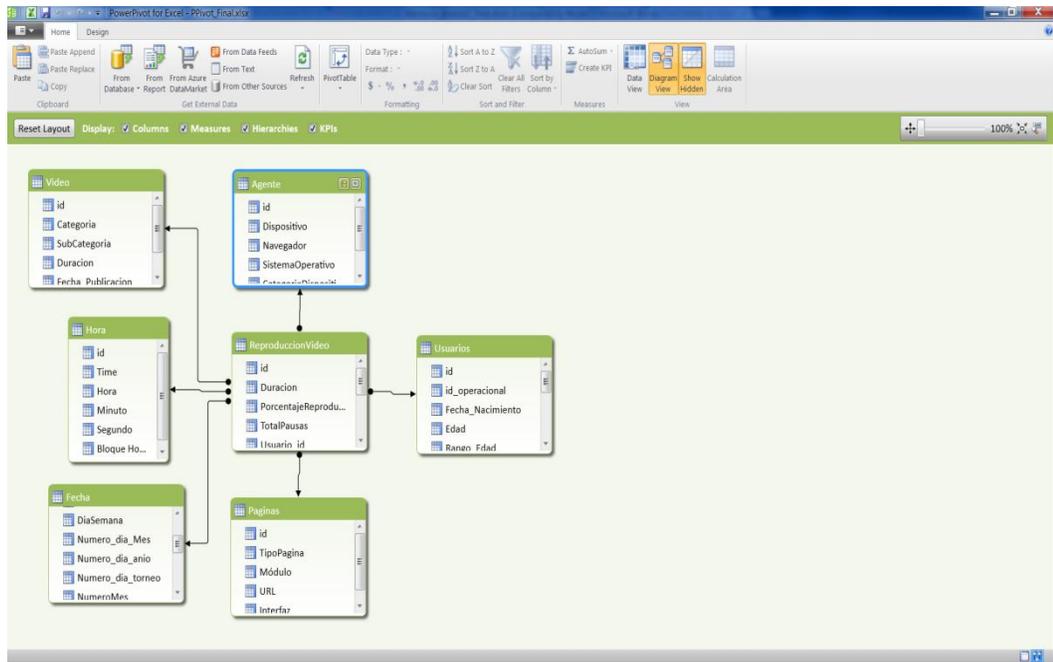


Ilustración 10: Construcción de Data Mart en PowerPivot

Posteriormente, se crean los gráficos y tablas para obtener las métricas definidas en el capítulo 4.2. En la interfaz de Excel, se seleccionan los atributos, ya sean como filtros, medidas o campos del eje, desde la lista de campos del cubo OLAP (desplegada a la derecha en la Ilustración 11) para agregarlos en el gráfico y así construir las métricas.

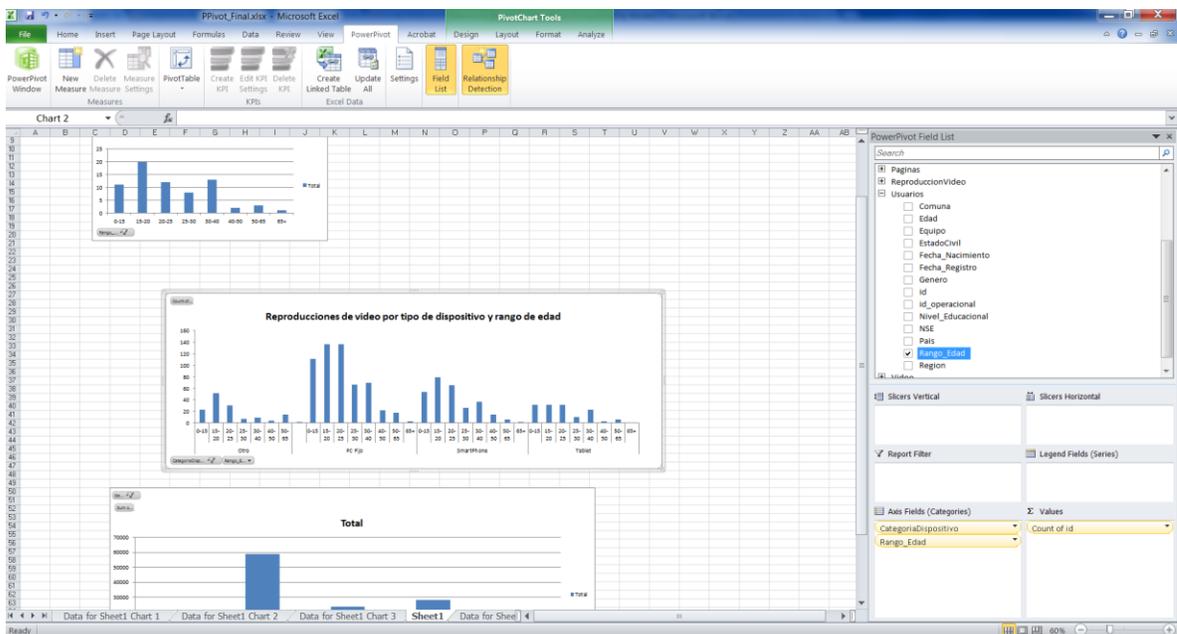


Ilustración 11: Construcción de gráficos en PowerPivot

La implementación de estos cubos OLAP permite al analista realizar cruces de información a su medida, atendiendo necesidades de información específicas que pueden ser variables en el tiempo.

6.2. Procesamiento de datos

De acuerdo a la definición de las métricas, el programa procesa los cálculos generando los gráficos que representan los resultados del trabajo. Estos cálculos corresponden conteos, sumas o porcentajes, resultados de los valores encontrados en las distintas tablas que forman parte de la base de datos, de acuerdo a los parámetros fijados por el analista para obtenerlos (por ejemplo, sumar las reproducciones de video en un mes).

De esta forma, se despliegan los resultados que permiten hacer el análisis y obtener las métricas e indicadores diseñados, además de permitir al analista navegar por el cubo OLAP, lo cual permite flexibilizar la demanda de información y adaptarse al entorno variable de la empresa.

7. Análisis de Resultados

En este capítulo se busca discutir los resultados más relevantes obtenidos durante el trabajo.

En primer lugar, se observa la evolución de vistas de videos en el sitio CDF.cl durante el años 2012, mediante el cálculo del indicador IAV1:

	Reproducciones de Video en el sitio (2011)	Reproducciones de Video en el sitio (2012)	Variación	Resultado
enero	54.621	44.290	-19%	Malo
febrero	89.072	81.753	-8%	Malo
marzo	78.581	83.939	7%	Bueno
abril	79.532	76.652	-4%	Regular
mayo	112.064	104.872	-6%	Malo
junio	84.869	62.308	-27%	Malo
julio	26.021	91.836	253%	Bueno
agosto	93.591	89.722	-4%	Regular
septiembre	97.179	87.518	-10%	Malo

Tabla 15: Resultados indicador IAV1

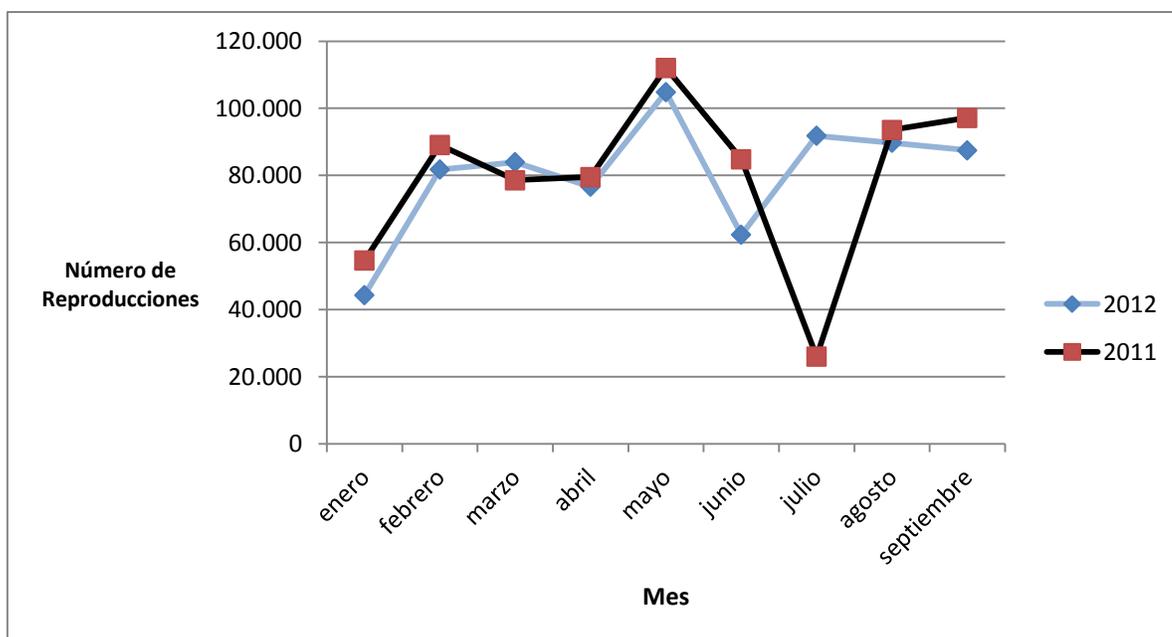


Ilustración 12: Indicador IAV1 - Gráfico de total de reproducciones de videos en CDF.cl

Los resultados muestran que la estacionalidad de las competencias es el factor más relevante respecto al volumen de reproducciones de videos en el sitio. Es por esto que los *peaks* se dan en los meses de Mayo, cuando culmina la competencia de Apertura. El valle marcado en el mes de Julio de 2011 se debe a que no hubo competencia local, debido al desarrollo de una competencia de selecciones que fue transmitida por la televisión abierta (Copa América Argentina 2011), provocando una fuga temporal de usuarios.

Mes	Reproducciones "Piratas"	Reproducciones en CDF.cl	Valor Métrica MIAV3	Valor Indicador IAV2	Resultado
ene-12	91.703	44.290	67%	-	-
feb-12	1.219.051	81.753	94%	26%	Malo
mar-12	1.155.406	83.939	93%	0%	Regular
abr-12	952.648	76.652	93%	-1%	Regular
may-12	849.796	104.872	89%	-4%	Regular
jun-12	677.327	62.308	92%	3%	Regular
jul-12	389.648	91.836	81%	-11%	Bueno
ago-12	1.498.538	89.722	94%	13%	Malo
sep-12	433.594	87.518	83%	-11%	Bueno

Tabla 16: Resultados Indicador IAV2

Estos resultados comprueban que a CDF le perjudica en gran medida la producción y distribución de su contenido mediante terceros no autorizados, ya que en promedio,

tiene una participación del 13% frente a los sitios de almacenamiento y reproducción de videos. Este es un problema transversal en la industria de la Televisión, no sólo para eventos deportivos ya que no existe una solución tecnológica que evite la captura de la señal, la distribución de los contenidos y su eventual difusión por redes sociales o páginas de personas naturales.

Se ha detectado, además, que las cuentas con las que estos videos son subidos están asociadas a sitios web no oficiales de clubes de fútbol chileno, mantenidos por hinchas. Esto demuestra que el público, en general, prefiere consumir el material desde una perspectiva parcial, lo cual es desfavorable para CDF, ya que por definición editorial debe mantener una posición neutra entre los clubes.

Los resultados de la métrica MIAV3 prueba que CDF no ha logrado posicionarse en internet como el único medio con las imágenes de las competencias de fútbol chileno, a pesar que la gestión para evitar la difusión de contenido no autorizado ha mejorado.

Al ver el detalle de las reproducciones mensuales de videos por plataforma, se observa que en la demanda no se conserva una relación entre ellas mes a mes, lo que puede explicarse por la naturaleza de los productos disponibles en cada producto debido a la diferencia en los derechos de exhibición de cada una de ellas.

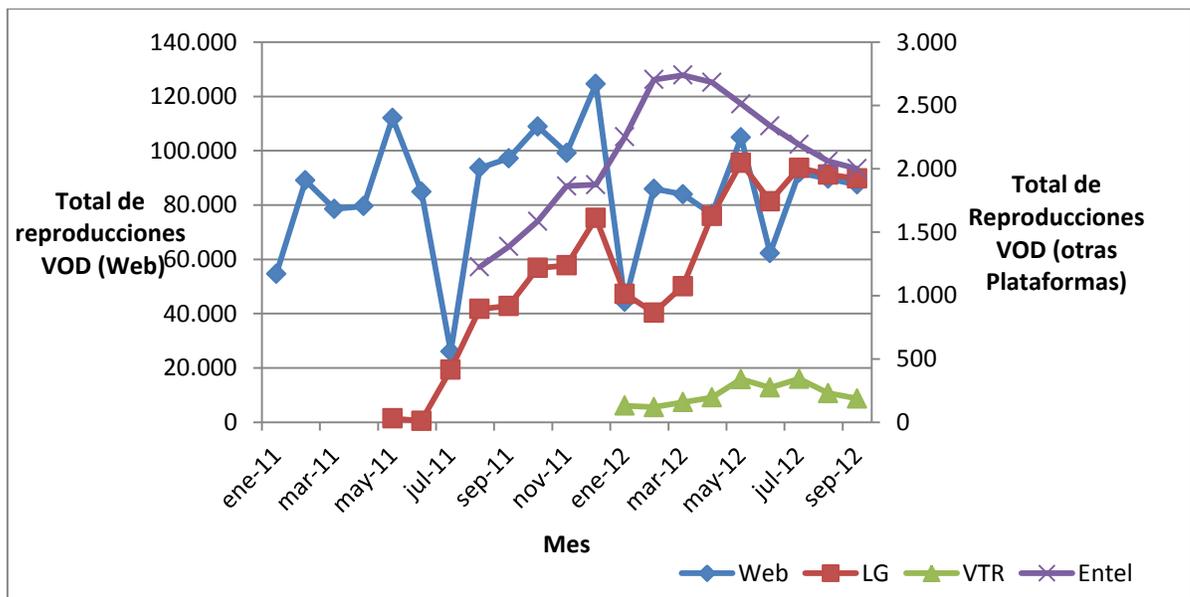


Ilustración 13: Reproducciones de video mensuales por Plataforma

De igual modo, esto representa diferencias en el comportamiento de consumo entre la mayor parte de los usuarios (que consumen vía web), y los usuarios que utilizan servicios externos, los que representan nichos de consumo diferenciado.

Respecto a la composición demográfica de los usuarios en la base de datos de CDF.cl, se observa un porcentaje considerable de mujeres, público que no ha sido

considerado como un segmento relevante en la generación de contenidos audiovisuales o editoriales.

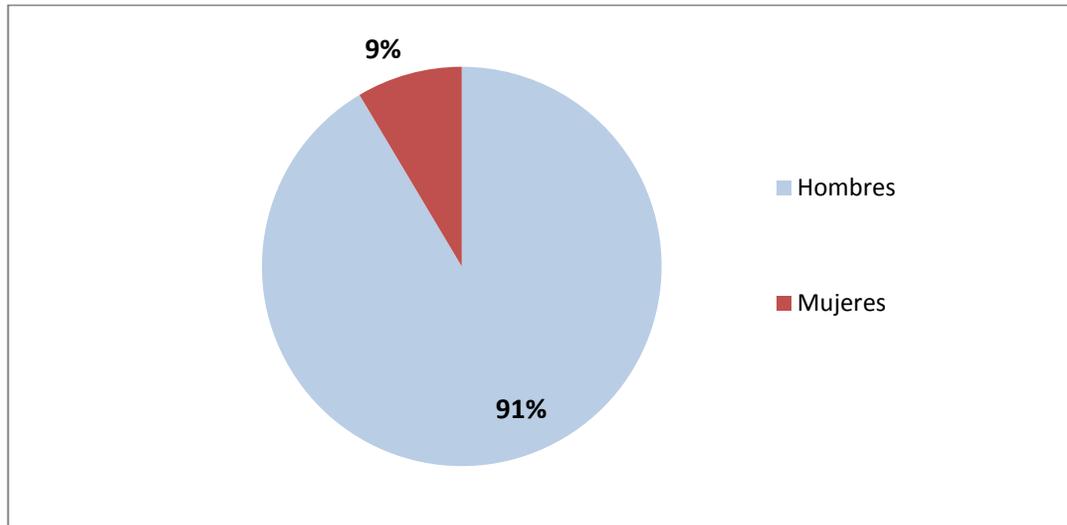


Ilustración 14: Porcentaje de Usuarios por Género

Al filtrar las reproducciones de video y centrarse sólo en el público femenino, se puede observar qué tipo de videos prefieren consumir:

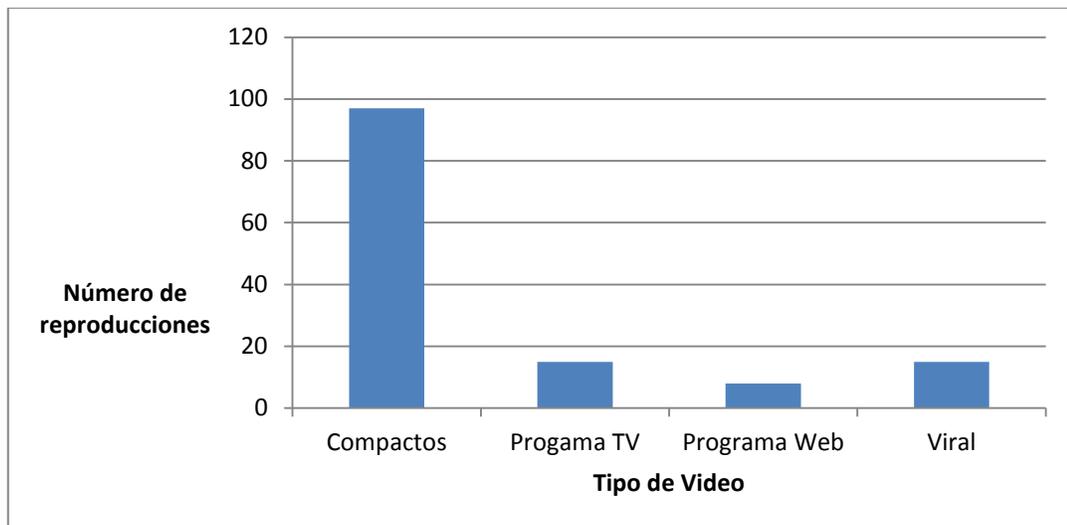


Ilustración 15: Reproducciones de video por parte del público femenino

Los resultados muestran que la mayor parte de las reproducciones corresponden a videos de compactos, lo que es similar al uso que le da el público masculino. Sin embargo, se observa una tendencia a los videos virales. Esto puede ser debido a que estos videos son generalmente promocionados a través de la página de Facebook de CDF, donde se ha apreciado cualitativamente que la participación de mujeres es mayor que la de hombres, lo que ha sido comprobado por estudios internacionales [17].

Otro resultado de interés ha sido conocer el total de reproducciones de video por rango de edad de los usuarios, obteniendo los siguientes resultados:

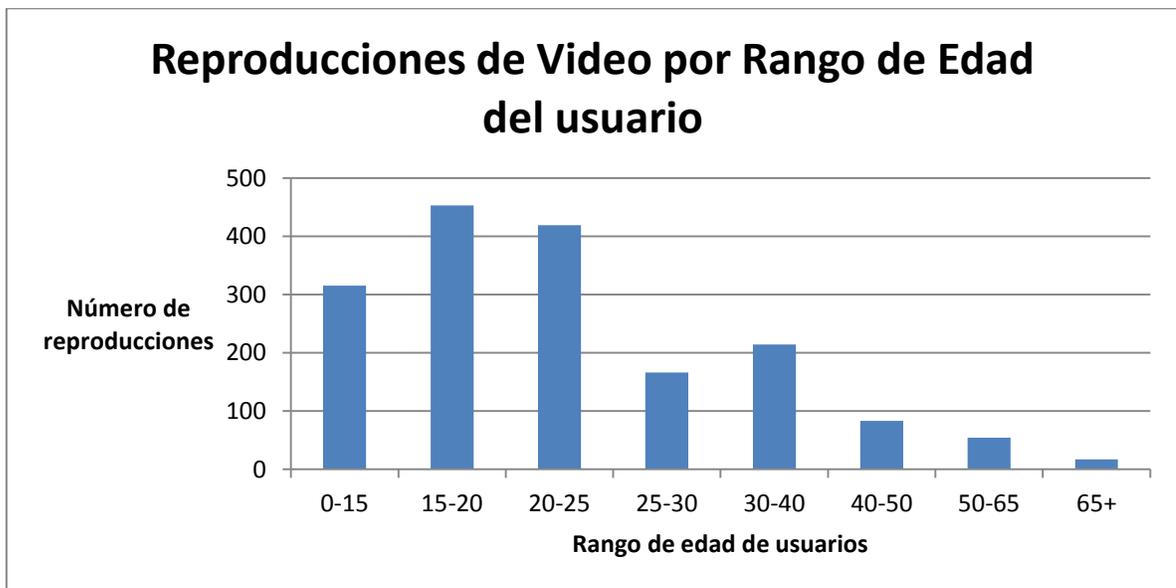


Ilustración 16: Reproducciones de video por Rango de edad del usuario

Esto muestra que la mayor parte de los usuarios que consumen videos son menores de 25 años. Esto parece estar relacionado a una brecha generacional, donde los usuarios más jóvenes tienen la costumbre adquirida de encontrar el contenido en Internet, mientras que el público adulto acostumbra a ver los compactos de goles por televisión.

Respecto al comportamiento de consumo, se observa que los días donde más se producen reproducciones los días sábado, domingo y lunes, lo cual tiene lógica de acuerdo a la naturaleza del negocio, debido a que los partidos de fútbol profesional se disputan durante el fin de semana.

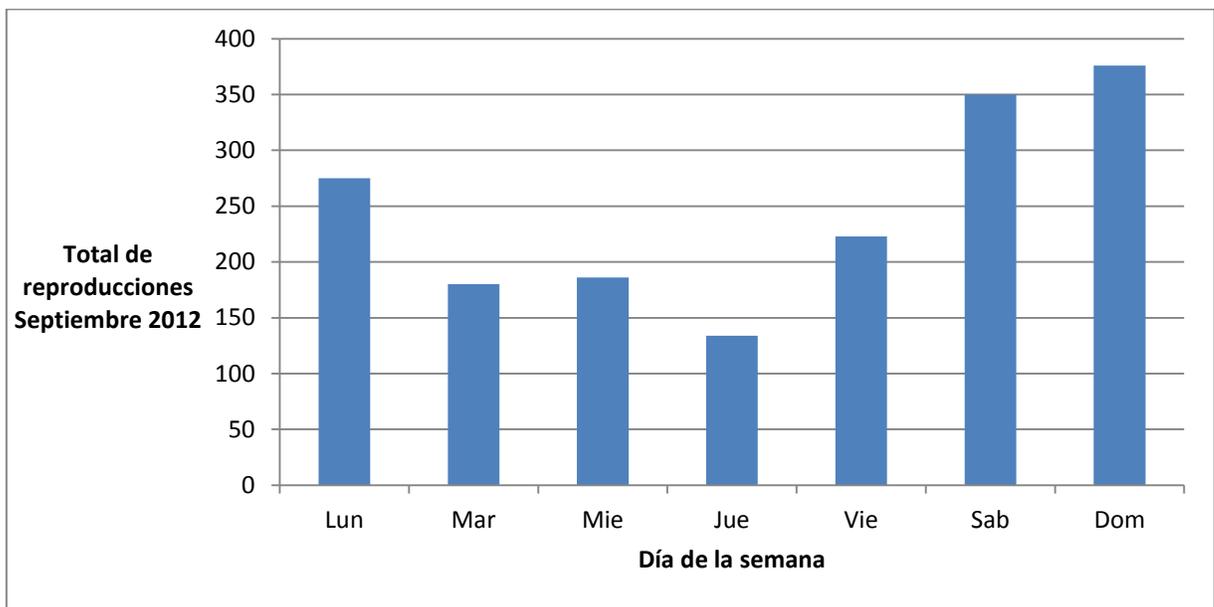


Ilustración 17: Métrica AV3[1][G]: Promedio de reproducciones por día de semana

También se observa que los horarios de mayor concurrencia son acordes al uso general de Internet.

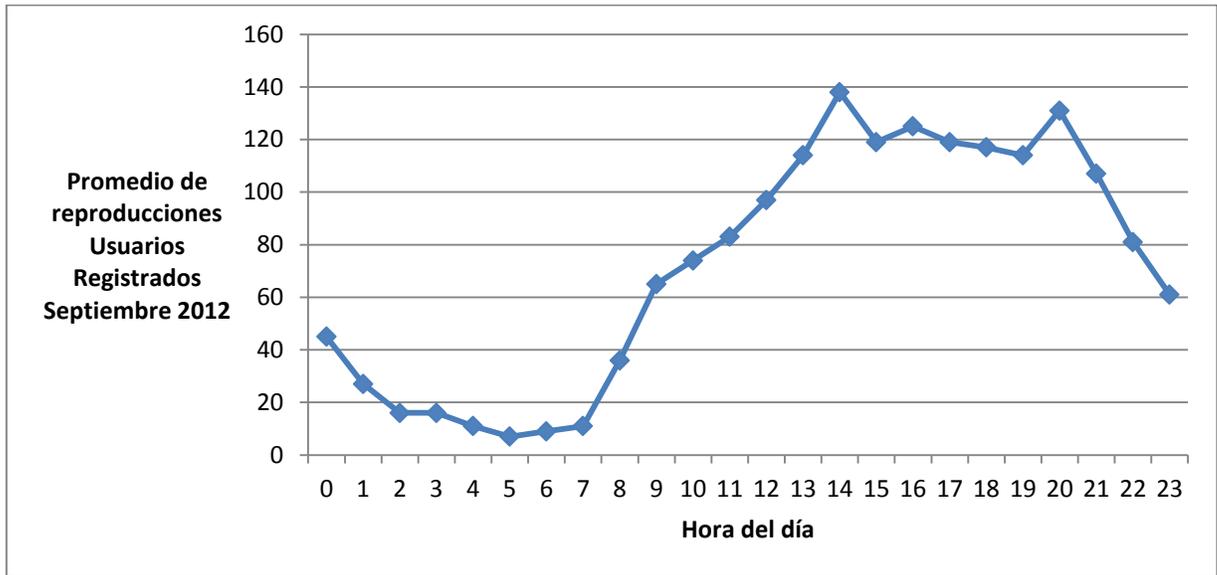


Ilustración 18: Métrica AV3[1][C]: Promedio de reproducciones por Hora del día

Al filtrar en el cubo OLAP las reproducciones de video por subcategoría y mirar el volumen de reproducciones realizadas por los usuarios menores de 21 años, se observa que en su mayoría consumen videos relacionados a la competencia, mucho más que a los programas.

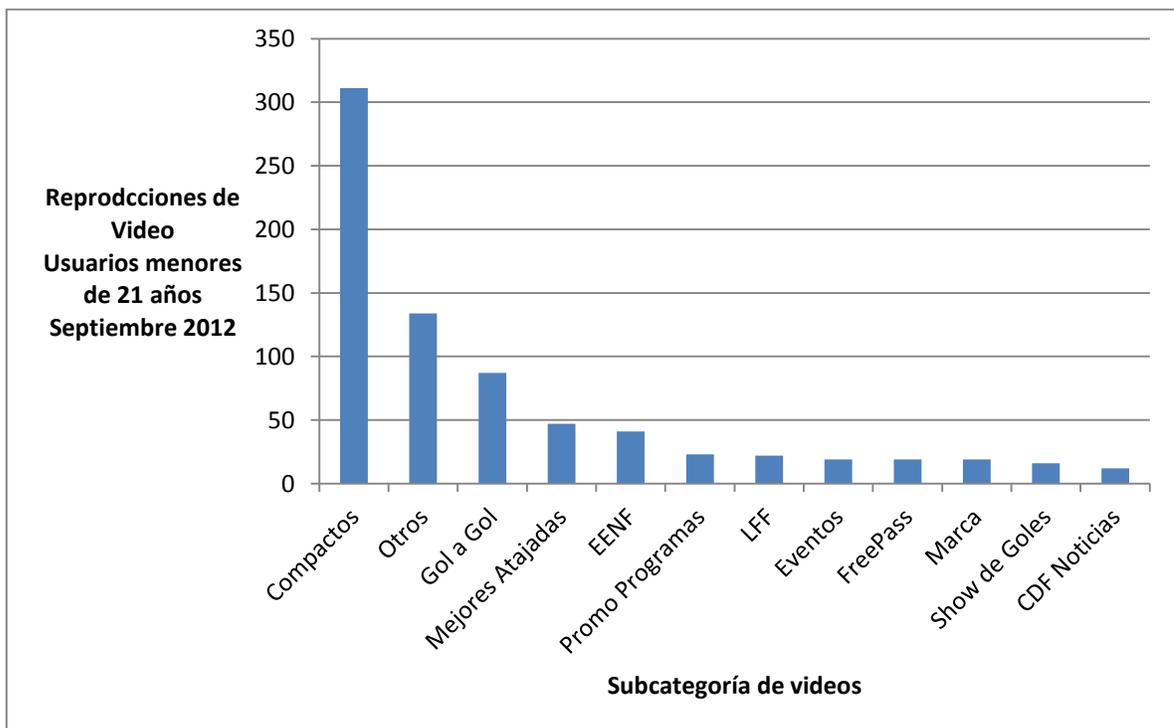


Ilustración 19: Métrica AV1[2][A=0-20]: Total de reproducciones por subcategoría de video por usuarios menores de 21 años.

Esto puede significar que la oferta de videos complementaria a la competencia no resulta atractiva para el público de esta edad, o bien, que los canales de distribución de este contenido no están alineados con el consumo habitual de los más jóvenes.

Finalmente, se destaca el resultado del indicador IAV5: Durante el mes de Septiembre de 2012, el porcentaje de visitas que realizan su navegación registrándose en el sitio es de 0,9%. Este es un porcentaje muy bajo, dado los objetivos de la Gerencia. Esto se debe, principalmente, a que no existe mayor valor agregado por navegar registrado, ya que los únicos diferenciales son la posibilidad de ingresar comentarios a los foros de discusión y a las notas editoriales, de jugar mini-juegos (aplicaciones flash en la web) y de participar en el juego CDF Manager¹⁴.

8. Conclusiones

A partir de múltiples necesidades de información para poder apoyar la toma de decisiones editoriales y comerciales relativas a la gestión del contenido audiovisual de la gerencia Multiplataforma de CDF, se analizó el entorno de la empresa con el fin de definir métricas e indicadores relevantes para dar solución al problema de gestión de contenidos audiovisuales de esta área.

Se planteó como solución la creación de un repositorio de datos que permite realizar análisis multidimensional sobre el consumo de videos y el impacto de éstos en el público de CDF.

La información recabada resulta útil para el proceso de toma de decisiones, ya que está basada en mediciones reales de consumo de videos de los usuarios. En efecto, en base a los resultados se realizó un análisis orientado a dar lineamientos sobre posibles mejoras de estos productos.

En primer lugar, se comprueba una fuerte fuga de usuarios durante los meses donde no hay competencia local de fútbol. En efecto, las reproducciones de videos disminuyen hasta el 32% en comparación con un mes promedio. Se recomendó la producción de piezas audiovisuales orientadas a compilar los momentos más relevantes de la última competencia. Esto con el fin de mantener activo el uso del sitio y aumentar la recurrencia de visitas durante estos períodos de baja demanda. Además, se recomendó la creación progresiva de piezas audiovisuales independientes de los partidos, que apunten a cautivar al público que consume material audiovisual de fútbol no necesariamente ligado a la competencia (por ejemplo, reportajes a la historia de los

¹⁴ <http://manager.cdf.cl> : Juego donde se simula ser un director técnico de fútbol, seleccionando un plantel de jugadores (dado un presupuesto acotado) que entregan puntaje de acuerdo a su desempeño en la competencia real de fútbol profesional chileno.

clubes, vida personal de los jugadores, ambiente de la ciudad o barrio donde juegan de local, etcétera).

Además, se creó una serie de videos (*Resumen Noticioso*, que compila tres eventos interesantes del día) que se publica cada día de semana, con el fin de aumentar la reproducción de videos en los momentos donde no se desarrollan partidos para disminuir la dependencia a la competencia de fútbol y diversificar la oferta de videos disponibles en el sitio. Estos videos se comenzaron a publicar en Marzo de 2013, y en promedio tienen 700 reproducciones diarias, lo que significa un aumento del 14% en el total de reproducciones mensuales del sitio CDF.cl.

Los resultados entregados por el Cubo OLAP de reproducciones de video por usuarios registrados muestran que algunos segmentos del público prefieren cierto contenido por sobre otro. En particular, se destaca que los menores de 21 años generalmente no consumen videos de capítulos de programas, sino que prefieren observar resúmenes de los partidos y compilados de goles, razón por la cual se recomendó estudiar si publicar los videos de los programas diferentes a compilados de la competencia de fútbol en redes sociales (Facebook) logra promover el consumo de estos videos por el público más joven. De todas formas, este tipo de información resulta útil tanto para perfilar el contenido a nivel editorial, como para mejorar los argumentos de venta de los espacios publicitarios de los videos, logrando así rentabilizar de mejor manera las producciones de la Gerencia Multiplataforma de CDF.

Se generó una serie de indicadores para evaluar el desempeño del área en cuanto a gestión de contenidos audiovisuales. El resultado más relevante en este aspecto es el volumen de reproducciones de contenido de CDF subido por terceros sin autorización, en desmedro del uso del portal web de la empresa. Para resolver este problema, se propone el desarrollo de un sistema tecnológico que opere navegando los sitios de videos (en particular *YouTube*, donde se ejecutan el 80% del total de reproducciones “piratas”) y que busque periódicamente videos por palabras clave asociadas al fútbol chileno, generando alertas al equipo de gestión antipiratería para que revisen con mayor rapidez los videos sospechosos y se acelere la denuncia para eliminar ese contenido de la red. Además, se propuso generar una campaña publicitaria en *YouTube*, definiendo tanto el público como el contenido objetivo (la idea es que la publicidad promocióne a CDF.cl como centralizador de contenido audiovisual de la competencia local de fútbol, en todos los videos relacionados con fútbol vistos por personas ubicadas en Chile).

Finalmente, se propuso agregar valor a la navegación dentro del sitio estando registrado. Para esto, el enfoque es personalizar la navegación del usuario generando una vista de diseño gráfica diferenciada de acuerdo al equipo del cual es hincha el usuario. Además, se propuso destacar comentarios emitidos en los foros de discusión en la portada del sitio, como una manera de incentivar a los usuarios a participar de ellos.

Como trabajo futuro, se recomienda realizar proyectos de Data Mining con el fin de poder segmentar a los usuarios y desarrollar estrategias comerciales diferenciadas para la conversión de usuarios de productos gratuitos a clientes de productos de pago de CDF. En efecto, el conjunto de datos almacenados en el Data Mart se convierte en una importante fuente de datos para desarrollar proyectos de ésta índole.

Además, se puede mejorar la oferta mediante la construcción de una navegación orientada a la personalización de su experiencia en los productos multiplataforma de CDF [18], mediante la implementación de un panel de navegación de videos que recomiende contenidos audiovisuales en función del historial de consumo de cada usuario particular. Para esto, es necesaria la implementación de un sistema que lea en línea los registros del usuario y determine, de acuerdo a las etiquetas (tags) de los videos, cuáles son los más probables de cautivar a cada usuario del sitio.

Otro desafío futuro consiste en integrar otras fuentes de información. En particular, es interesante la incorporación de datos ligados a la actividad de CDF en redes sociales (principalmente Facebook y Twitter), para analizar cómo impactan en esos medios los eventos deportivos, por una parte, y la publicación de videos por otra.

Las empresas de televisión tienen una gran oportunidad al poder establecer una relación directa con el consumidor final, lo que abre muchas oportunidades en la industria, siendo éstas una de ellas.

9. Referencias

- [1] A. YARALI, «Internet Protocol Television (IPTV),» Melbourne, 2005.
- [2] P. BERNSTEIN, F. GIUNCHIGLIA, A. KEMENTSIETSIDIS, J. MYLOPOULOS, L. SERAFINI y I. ZAIHRAYEU, «Data Management for Peer-to-Peer Computing: A Vision,» 2002.
- [3] CERCÓS, ROBERT, «Diseño y Construcción de Data Web Warehouse para almacenar información extraída a partir de datos originada en la web,» Santiago de Chile, 2008.
- [4] MUÑOZ, FELIPE, «Diseño e implementación de métricas e indicadores para apoyar la gestión en una comunidad virtual de práctica,» Santiago, Chile, 2012.
- [5] REBOLLEDO, VÍCTOR, «Plataforma para la extracción y almacenamiento del conocimiento extraído desde los Web Data,» Santiago de Chile, 2009.
- [6] B. KITTS, L. WEI, D. AU, S. ZLOMEK, R. BROOKS y B. BURDICK, «Targeting Television Audience using Demographic Similarity,» Seattle, WA, USA, 2010.
- [7] K. CHORIANOPOULOS, G. LEKAKOS y D. SPINELLIS, «The Virtual Channel Model for Personalized Television,» Athens, 2003.
- [8] KIMBALL, RALPH; MERZ, RICHARD, *The Data Webhouse Toolkit*, Nueva York: Wiley, 2000.
- [9] TANENBAUM, A, «Computer Networks,» New Jersey, 1996.
- [10] YEAGER, N.; McGRATH, R., «Web Server Technology: The Advanced Guide World Wide Web Information Providers,» San Francisco, CA, USA, 1996.
- [11] KIMBALL, RALPH, *The Data Warehouse Toolkit*, Wiley, 1996.
- [12] MCFADDEN, F.R., «Data Warehouse for EIS: Some issues and impacts.,» Wailea, HI, USA, 1996.
- [13] VELÁSQUEZ SILVA, JUAN DOMINGO; ROMÁN ASENJO, PABLO, *Apuntes de Curso IN4523, Data Warehousing, Departamento de Ingeniería Industrial Universidad de Chile, Santiago de Chile, Semestre Otoño 2011.*
- [14] L. OLSINA, F. PAPA y H. MOLINA, «How To MEasure and Evaluate Web Applications in a Consistent Way,» de *Web Engineering: Modelling and Implementing Web Applications*, General Pico, La Pampa, Argentina, 2007.
- [15] SIMONS, NELE, «Television Audience Research in the Age of Convergence: Challenges and Difficulties,» Antwerp, Bélgica, 2011.

- [16] L. DUJOVNE, «Diseño y Aplicación de una Metodología para el mejoramiento del contenido de sitios web mediante la identificación de website keyobjects,» Universidad de Chile, Santiago, Chile, 2010.
- [17] E. GONZALEZ, «PRWeb,» 22 9 2011. [En línea]. Available: <http://www.prweb.com/releases/prweb2011/9/prweb8819065.htm>. [Último acceso: 10 2 2013].
- [18] J. VELÁSQUEZ y V. PALADE, Adaptive Web Sites: A Knowledge Extraction From Web Data Approach, IOS Press, 2008.
- [19] SINGH, AMRINDER; MEHTA, SUSHIL KUMAR; MISHRA, HARI GOVIND, «TRP as a Measure of Visual Communication: A study of Jammu City, India,» Jammu, India, 2011.
- [20] G. RIVADERA, «La metodología de Kimball para el diseño de almacenes de datos,» de *Cuadernos de la Facultad n.5*, Salta, Argentina, Universidad Católica de Salta, 2010, pp. 56-71.

10. Anexos

A. Composición demográfica de los Usuarios

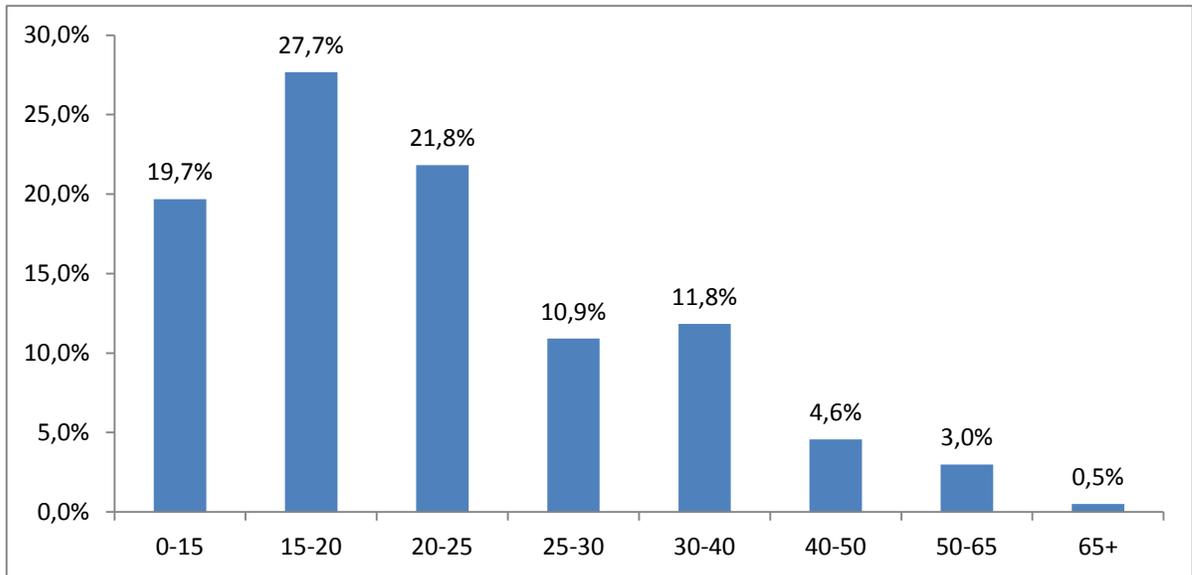


Ilustración 20: Porcentaje de usuarios por Rango de Edad

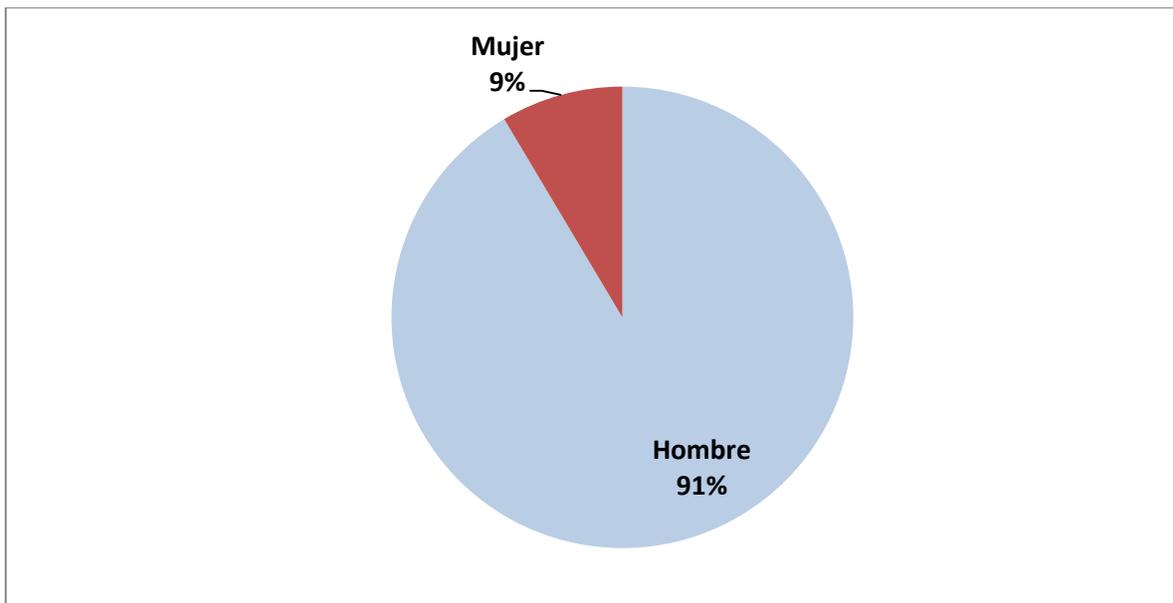


Ilustración 21: Porcentaje de usuarios por género

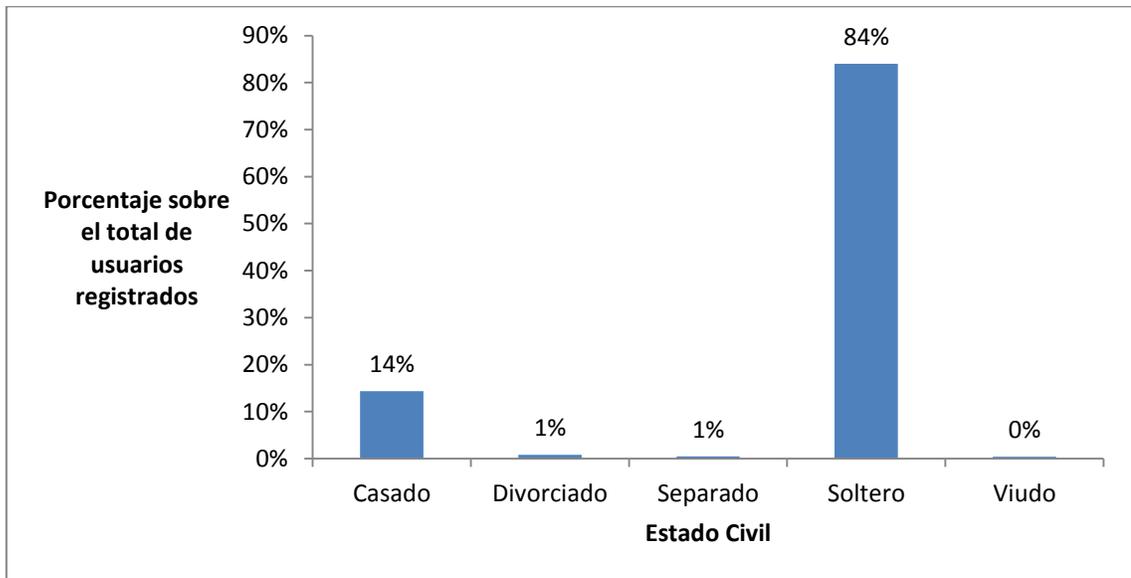


Ilustración 22: Porcentaje de usuarios por Estado Civil

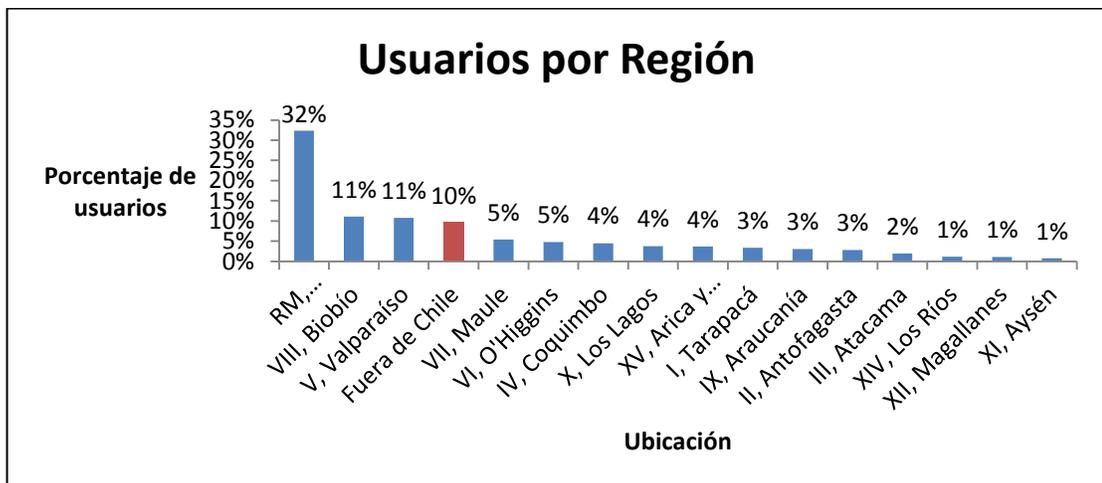


Ilustración 23: Porcentaje de usuarios por Ubicación

B. Resultados de Reproducciones de Videos según atributos demográficos de usuarios

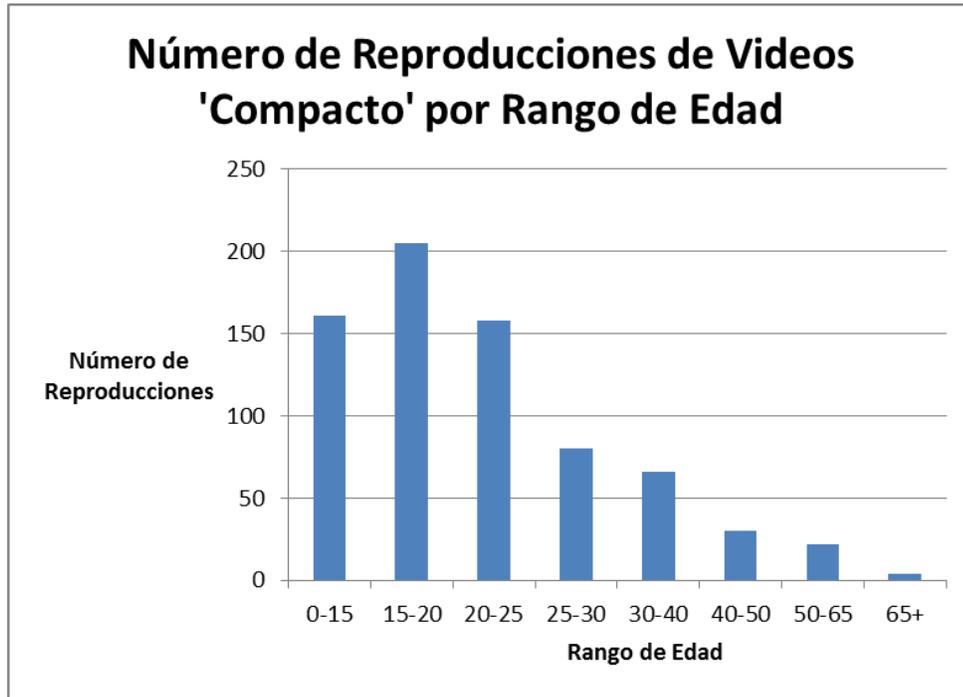


Ilustración 24: Reproducciones de 'Compactos' por Rango de Edad

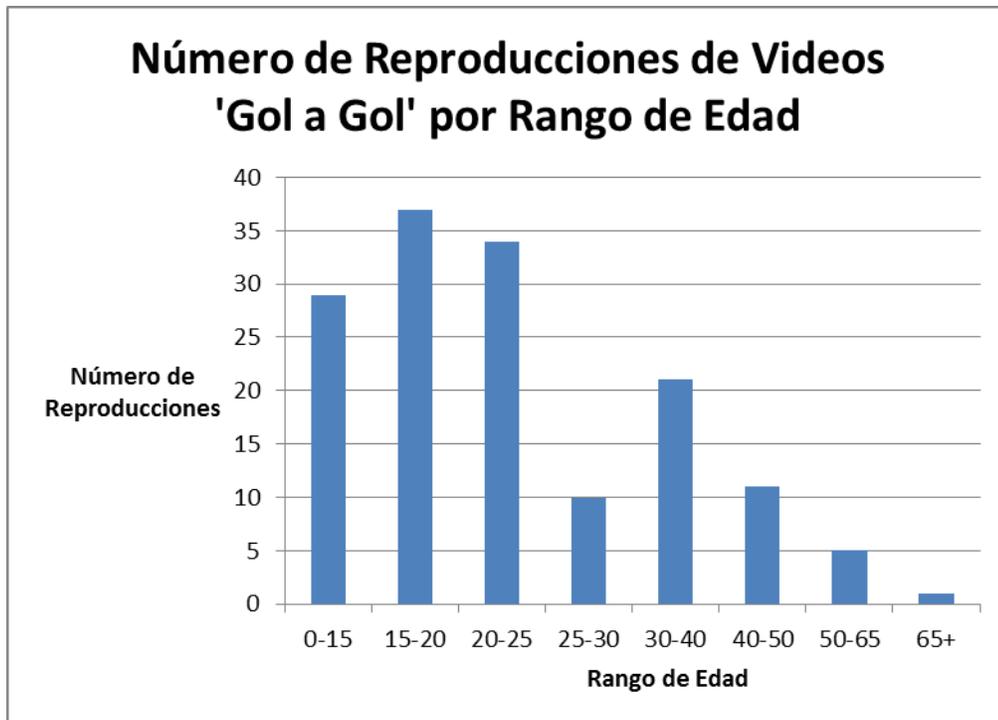


Ilustración 25: Reproducciones de 'Gol a Gol' por Rango de Edad

Número de Reproducciones de Videos 'Mejores Atajadas' por Rango de Edad

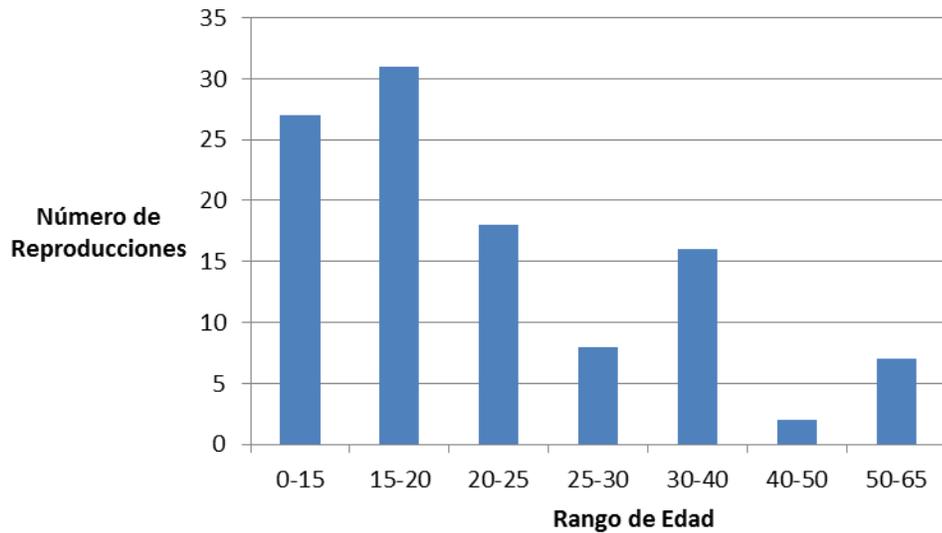


Ilustración 26: Reproducciones de 'Mejores Atajadas' por Rango de Edad

Número de Reproducciones de Videos 'Free Pass' por Rango de Edad

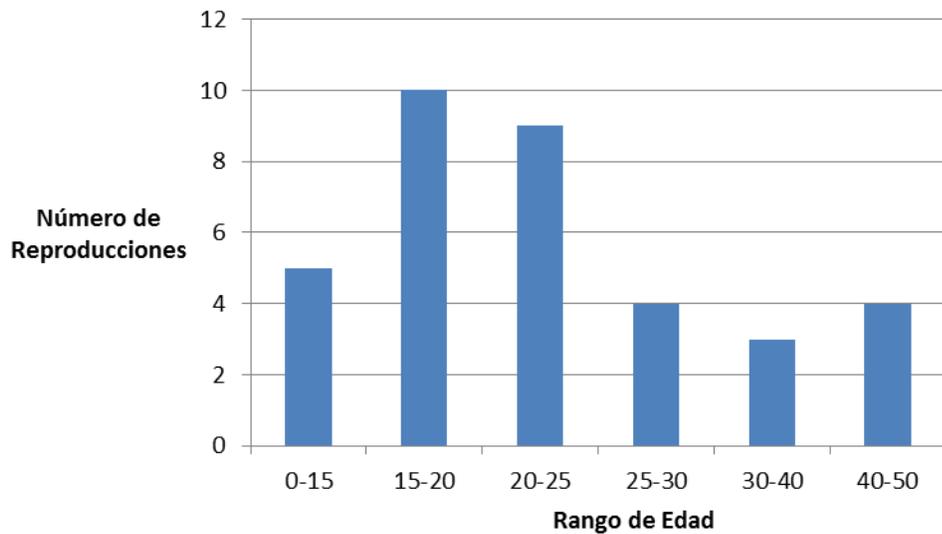


Ilustración 27: Reproducciones de 'Free Pass' por Rango de Edad

Número de Reproducciones de Videos 'La Fecha a Fondo' por Rango de Edad

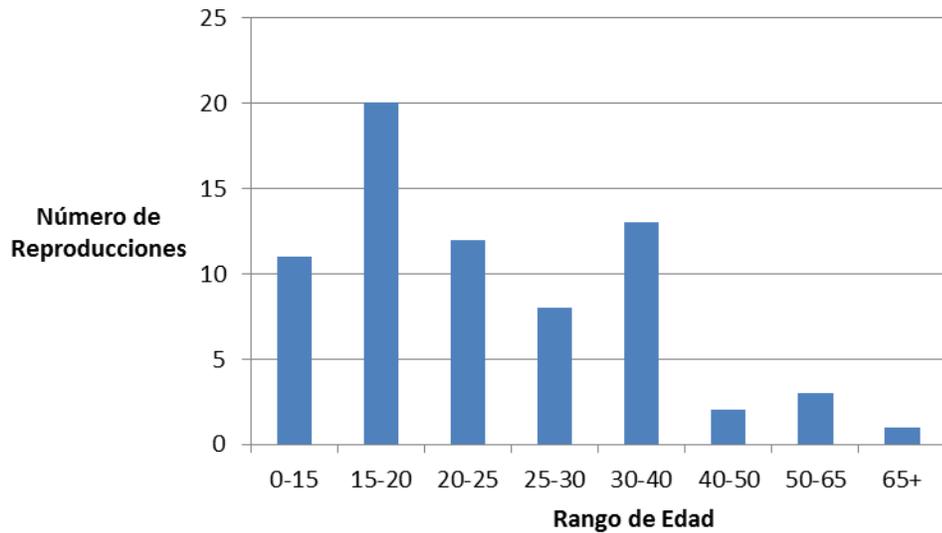


Ilustración 28: Reproducciones de 'La Fecha a Fondo' por Rango de Edad

Número de Reproducciones de Videos de Programas por Rango de Edad

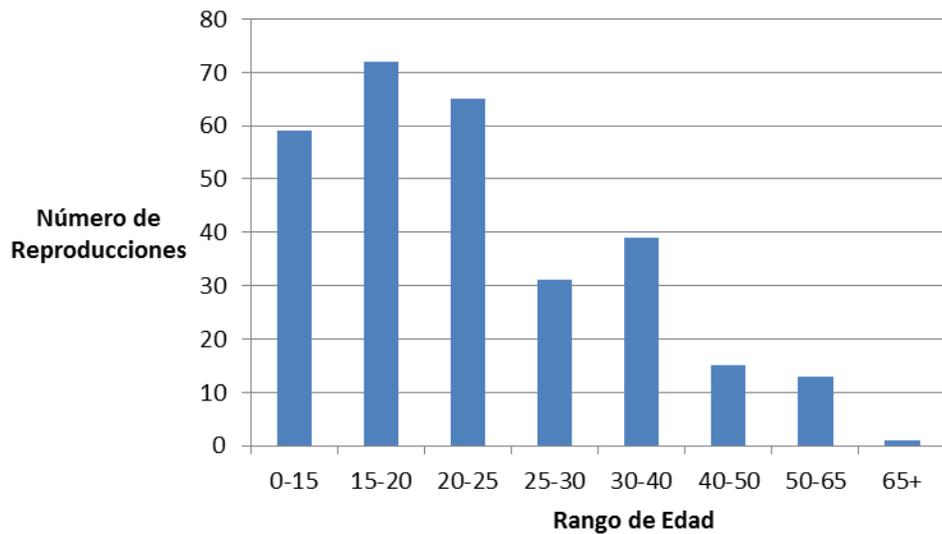


Ilustración 29: Reproducciones de Videos tipo Programas por Rango de Edad

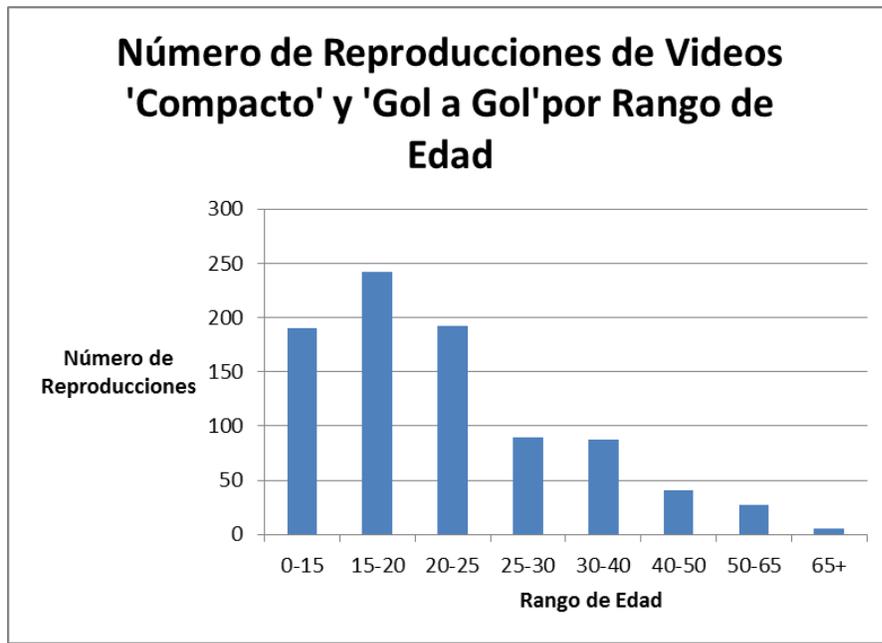


Ilustración 30: Reproducciones de Compilados por Rango de Edad

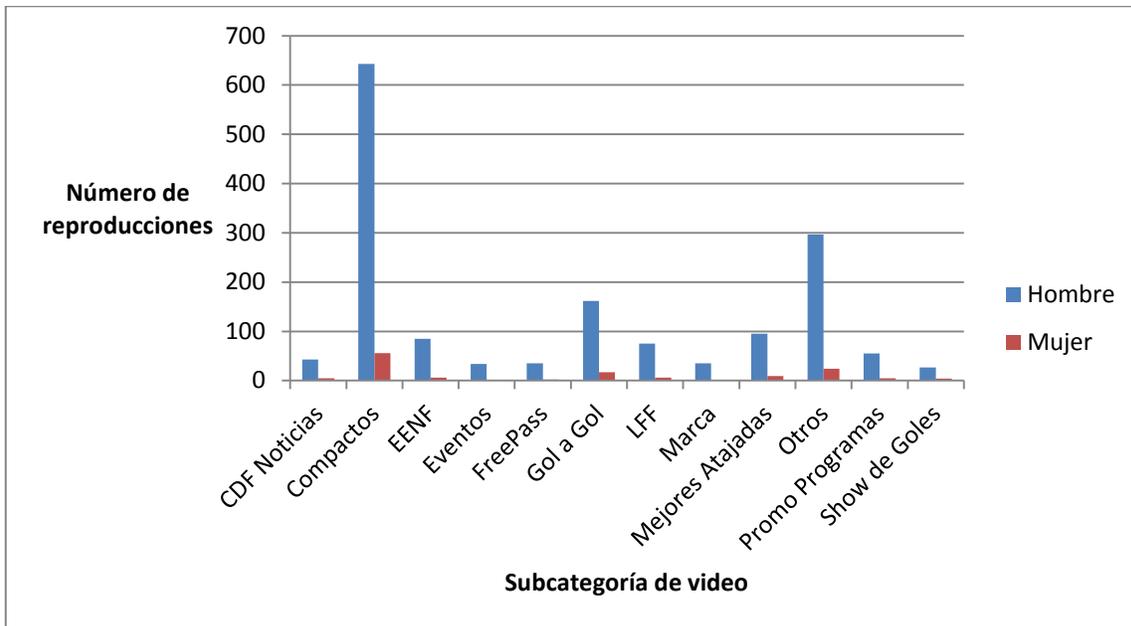


Ilustración 31: Reproducciones de video por Género y Subcategoría de Video

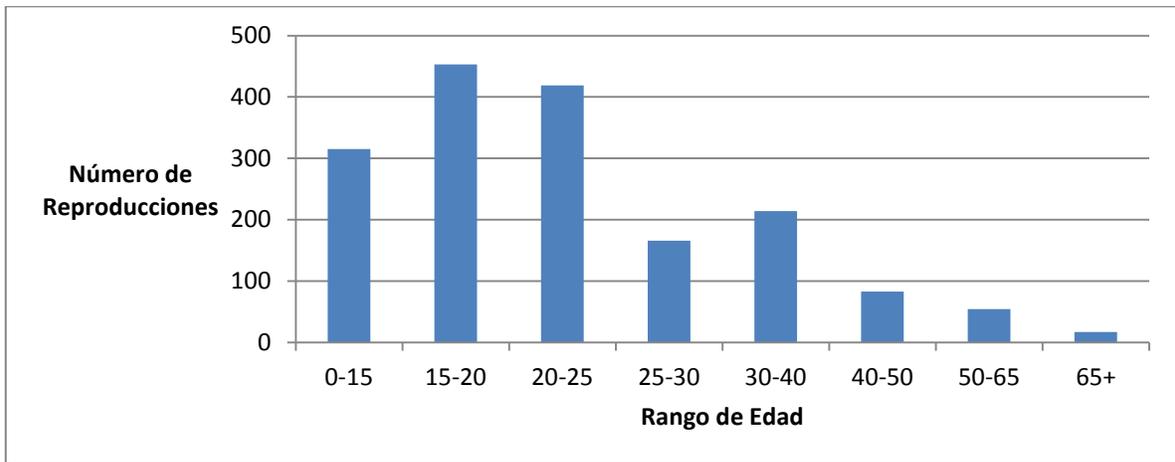


Ilustración 32: Reproducciones de video por Rango de Edad del usuario

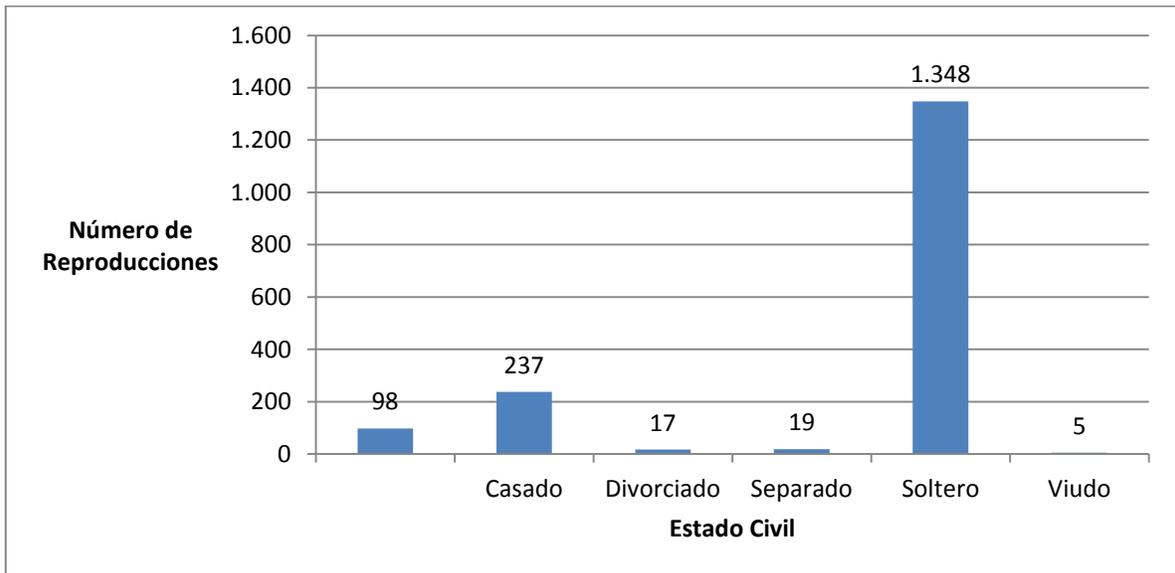


Ilustración 33: Reproducciones de video por Estado Civil del usuario

C. Resultados Comportamiento de Consumo

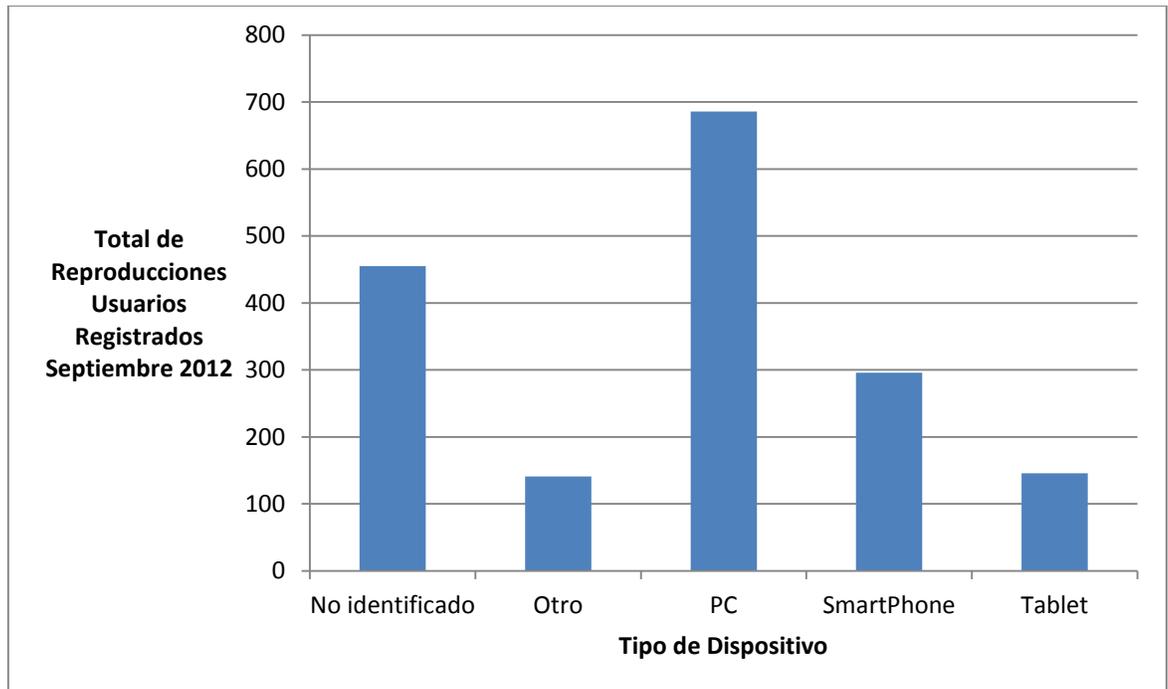


Ilustración 34: Reproducciones por Tipo de Dispositivo

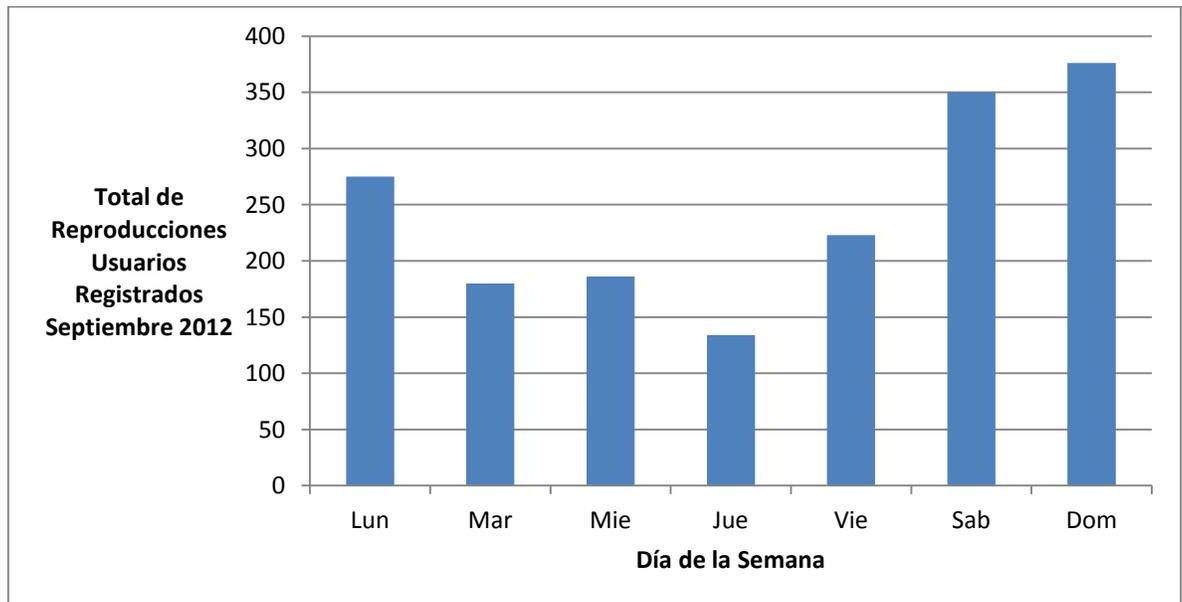


Ilustración 35: Reproducciones por día de semana

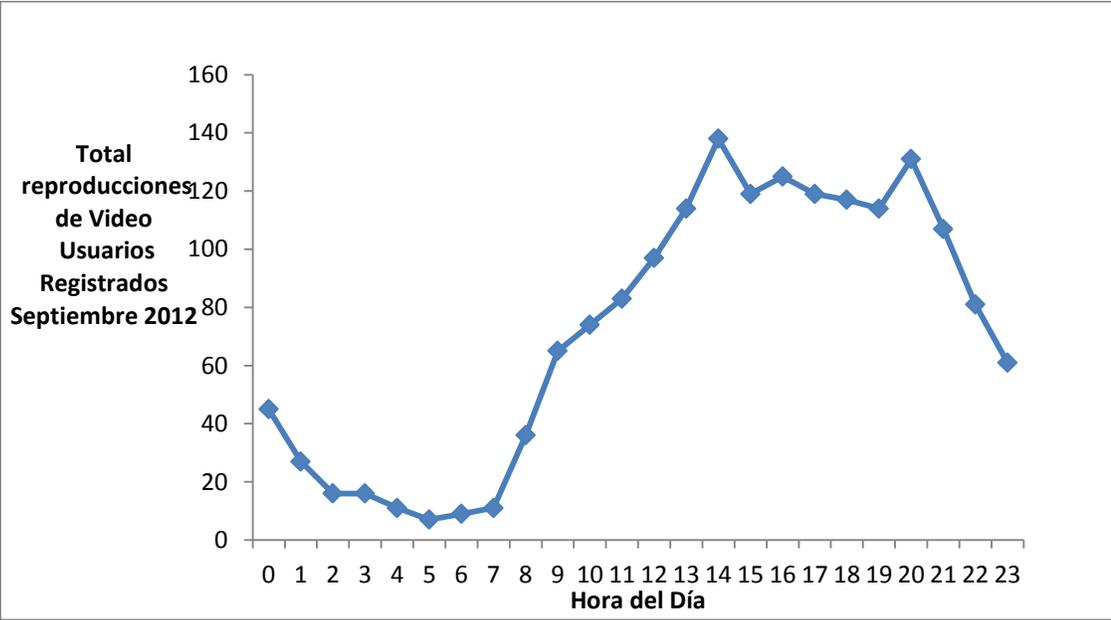


Ilustración 36: Reproducciones por Hora del día

D. Resultados de Indicadores

Mes	Reproducciones de Video en el sitio (2011)	Reproducciones de Video en el sitio (2012)	Variación	Resultado
enero	54.621	44.290	-19%	Malo
febrero	89.072	81.753	-8%	Malo
marzo	78.581	83.939	7%	Bueno
abril	79.532	76.652	-4%	Regular
mayo	112.064	104.872	-6%	Malo
junio	84.869	62.308	-27%	Malo
julio	26.021	91.836	253%	Bueno
agosto	93.591	89.722	-4%	Regular
septiembre	97.179	87.518	-10%	Malo

Tabla 17: Resultados Indicador IAV1

Mes	Reproducciones "Piratas"	Reproducciones en CDF.cl	Valor Métrica MIAV3	Valor Indicador IAV2	Resultado
ene-12	91.703	44.290	67%	-	-
feb-12	1.219.051	81.753	94%	26%	Malo
mar-12	1.155.406	83.939	93%	0%	Regular
abr-12	952.648	76.652	93%	-1%	Regular
may-12	849.796	104.872	89%	-4%	Regular
jun-12	677.327	62.308	92%	3%	Regular
jul-12	389.648	91.836	81%	-11%	Bueno
ago-12	1.498.538	89.722	94%	13%	Malo
sep-12	433.594	87.518	83%	-11%	Bueno

Tabla 18: Resultados Indicador IAV2

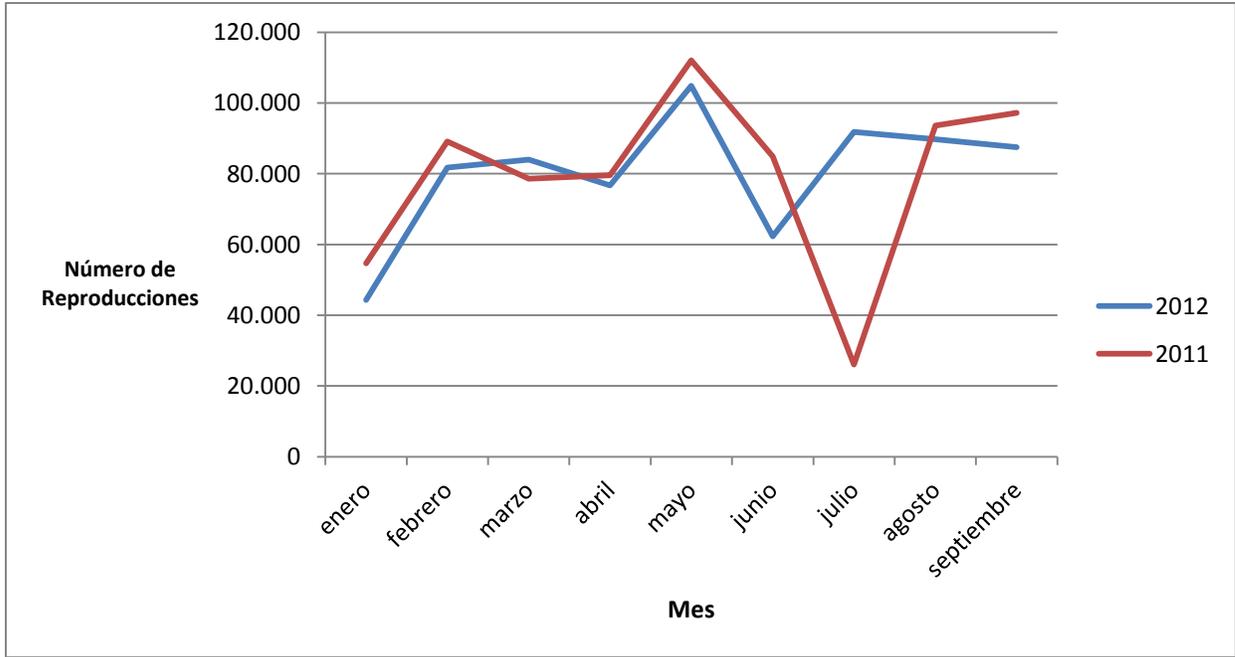


Ilustración 37: Gráfico de total de reproducciones de videos mensual en CDF.cl

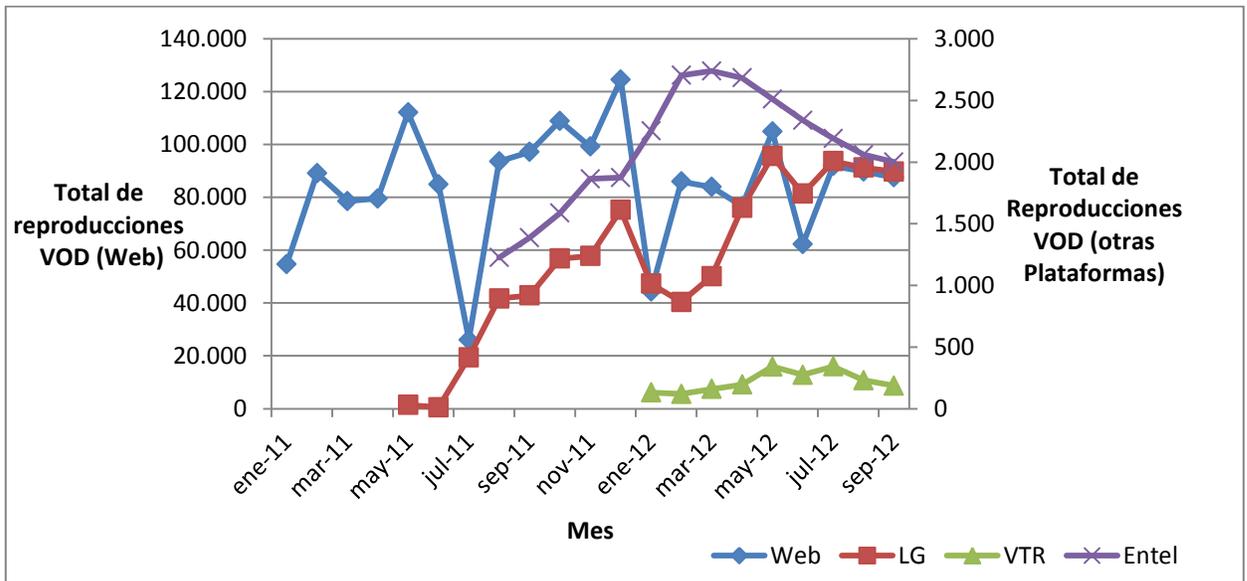


Ilustración 38: Evolución de consumo de videos por Plataforma

E. Descripción de Tablas correspondientes a las dimensiones del Data WebHouse

Tabla: Usuario.

Es la dimensión que describe al usuario registrado en el sitio CDF.cl

Los atributos son los siguientes:

- **id:** Es el identificador surrogante del usuario en el Data Warehouse.
- **Fecha_Nacimiento:** Fecha de nacimiento en formato DATE SQL.
- **Edad:** Se calcula en base a Fecha_Nacimiento
- **Rango_Edad:** Se determina en base a la edad (0-15; 15-20; 20-25; 25-30; 30-40; 40-50;50-65; 65+)
- **Género:** “Masculino”, “Femenino”.
- **Pais:** Corresponde al país de residencia del usuario.
- **Region:** Corresponde al número romano de la región en que habita un usuario de Chile. En caso que el usuario viva en otro país, se marca “No Aplica”.
- **Comuna:** Corresponde al nombre de la comuna. En caso que el usuario viva en otro país, se marca “No Aplica”.
- **TieneCDFtv:** Indica la mejor señal de CDF a la cual tiene acceso: (“No Tiene”, “Básico”, “Premium”, “HD”).
- **Tiene CDFm:** Corresponde a si el usuario tiene acceso directo a CDF móvil. Marca “Si”, “No”, “Indeterminado”.
- **Tiene CDFev:** Indica si el usuario ha comprado un abono del producto Estadio CDF.
- **Nivel_Educacional:** Corresponde al mayor grado de educación del usuario. (Básica incompleta; Básica Completa; Media incompleta; Media completa; Superior profesional incompleta; Superior profesional completa; Técnica incompleta; Técnica completa; Postgrado).

- **Profesión:** Corresponde a la profesión actualmente ejercida por el usuario.
- **Equipo:** Equipo del cual es hincha el usuario. Puede tomar cualquier valor de los predefinidos o "Ninguno".
- **Fecha_Registro:** Fecha de registro del usuario.
- **Fecha_modificacion:** Fecha de la última modificación a los datos del usuario.
- **Hora_Registro:** Hora de registro del usuario.
- **Grupo_Familiar:** Indica cuántas personas forman parte del grupo familiar del usuario. (1,...,10, 10+)
- **Estado_Civil:** Indica el Estado Civil del usuario (Soltero, Casado, Viudo, Divorciado)
- **NivelSocioeconómicoEstimado:** Clasificación interna que depende del valor en los campos "Estado Civil", "Comuna", "Nivel Educativo", "Profesión" y "Grupo Familiar".
- **Intereses Futbolísticos:** Indica si el usuario manifestó intereses por practicar fútbol, usar videojuegos de fútbol o ver partidos nacionales e internacionales (checklist).

Tabla: Fecha.

Un registro caracteriza un día. Poblado hasta 31 diciembre 2017.

- **id:** Identifica un registro, que corresponde a un día.
- **Date:** Fecha del día en formato SQL.
- **Dia_de_semana:** Nombre del día de la semana en 3 caracteres (Lun, Mar, ... Dom)
- **Numero_dia_semana:** Numero del día de la semana (1,..., 7)
- **Numero_dia_mes:** Numero del día en el mes (1,...31).

- **Numero_dia_año:** Numero dia en el año (1,...366).
- **DiaSemana:** “Dia de semana” o “Fin de semana”.
- **Mes:** Nombre del mes en 3 caracteres (ene, Feb,...,Dic)
- **Numero_Mes:** (1,...,12)
- **Trimestre_Año:** Indica trimestre y año del día en cuestión (Q1_2012, Q2_2012,...Q42017).
- **Semestre_Año:** Indica semestre y año (S1_2012,...,S2_2017).
- **Año:** Indica el año del dia (2012,...2017).
- **Epoca_Nacional:** Corresponde al estado del torneo nacional. (Regular_Apertura, Playoffs_Apertura, Receso_Vacaciones_invierno, Receso_Vacaciones_veranos , Regular_Clausura, Playoffs_Clausura, Receso_Seleccion).
- **Epoca_Sudamericana:** (Receso_Verano; Grupo_Libertadores; PlayOffs_Libertadores, Receso_Invierno, Primeras_Fases_Sudamericana, Fases_Finales_Sudamericana, CopaAmerica).
- **Epoca_Mundial:** (Olimpiadas, MundialFutbol, MundialJuvenil, Eurocopa, Grupo_Champions, PlayoffsChampions, MundialClubes).
- **Festividad*:** Marca días festivos, como 18 septiembre o Navidad.
- **Evento1:** Corresponde al principal evento específico de un día. Este campo debe ser escrito manualmente. Ejemplo: Final Copa Libertadores, U. de Chile vs Colo Colo, Semifinal PlayOffs, o algún hecho noticioso de relevancia.
- **Evento2:** Cumple la misma función del campo anterior, pero en segundo orden de importancia.

Tabla: Hora.

- **id:** Identifica un registro, el que corresponde a un segundo.

- **Time:** Registro en formato TIME SQL.
- **Hora:** (0,...,23)
- **Minuto:** (0,...,59)
- **Segundo:** (0,...,59)
- **Bloque_Horario:** Corresponde a un bloque o módulo horario
 - 00:00:00 a 02:59:59 : Madrugada1
 - 03:00:00 a 05:59:59 : Madrugada2
 - 06:00:00 a 08:59:59 : Mañana1
 - 09:00:00 a 11:59:59 : Mañana2
 - 12:00:00 a 14:59:59 : Almuerzo
 - 15:00:00 a 17:59:59 : Tarde
 - 18:00:00 a 20:59:59 : Tarde-Noche
 - 21:00:00 a 23:59:59 : Noche.

Tabla: Agente

- **id:** Identificador del registro.
- **Dispositivo:** Nombre genérico del dispositivo sobre el cual se realiza la navegación. Ejemplo: “iPhone” agrupa a todas las versiones de iPhone, lo mismo para “Samsung Galaxy”, etcétera.
- **Navegador:** Nombre genérico del navegador.
- **SistemaOperativo:** Nombre genérico del S.O. (windows agrupa todas las versiones de windows: XP, Vista, 7 , 8, etcétera, por ejemplo).
- **CategoriaDispositivo:** (PC fijo, PC portatil, Mac fijo, Mac portátil, Teléfono, Tablet, TV, Otro).

Tabla: Video

Esta tabla es un subconjunto de la tabla anterior, que sólo considera los videos.

- **id:** identificador en esta tabla.

- **Categoría:** Indica la categoría del video de acuerdo a la competencia de fútbol al cual hace referencia el video.
- **SubCategoría:** Subcategoría del video. Corresponde al título del programa o tipo de compilado de jugadas.
- **Duración:** Indica la duración en segundos del video.
- **Fecha_publicacion:** Fecha de subida del material audiovisual.
- **Hora_publicacion:** Hora de subida del material audiovisual.

Tabla: Video has tag

- **id:** identificador del registro.
- **video_id:** Llave foránea de la tabla Videos.
- **Tag:** Nombre de la etiqueta asociada al video.
- **CategoriaTag:** Categoría de la etiqueta.
- **SubCategoriaTag:** SubCategoría de la etiqueta.

Tabla: TipoPáginas

- **id:** Identificador de un registro, que corresponde a un conjunto de páginas del sitio.
- **TipoPagina:** (Home, Home_Modulo, Busqueda, Nota Editorial, Registro, Login...).
- **URL:** Dirección URL de la página.
- **Módulo:** Indica a qué Módulo (segmento editorial del sitio) pertenece la página.

Descripción de Tabla de Hechos:

Tabla ReproduccionVideo

Guardará los eventos que correspondan a una reproducción completa de un video.

Llaves foráneas a las siguientes dimensiones:

- Usuario
- Agente
- Video
- Hora
- Fecha
- Paginas_REF (página en la cual el video se encuentra embedido y se realiza la reproducción).

Medidas:

- Duración: Es la duración de la reproducción, medida en segundos.
- PorcentajeReproducción: Corresponde al porcentaje de video que fue reproducido.

$$\text{PorcentajeReproducción} = \frac{\text{duración de reproducción}}{\text{duración de video}} \times 100$$