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

GENERACIÓN DE ESTRATEGIAS DE NEGOCIO MEDIANTE TÉCNICAS DE CLUSTERING CON FOCO EN AUMENTAR LA PARTICIPACIÓN DE USUARIOS EN UNA PLATAFORMA EDUCATIVA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

FELIPE ANDRÉS ARAYA VILLELA

PROFESOR GUÍA:
JUAN ROMERO GODOY

MIEMBROS DE LA COMISIÓN:
JOSE NALDA REYES
ALEJANDRA PUENTE CHANDÍA

SANTIAGO DE CHILE
2024

RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE: Ingeniero Civil Industrial
POR: Felipe Andrés Araya Villela
FECHA: 2024
PROFESOR GUÍA: Juan Romero Godoy

GENERACIÓN DE ESTRATEGIAS DE NEGOCIO MEDIANTE TÉCNICAS DE CLUSTERING CON FOCO EN AUMENTAR LA PARTICIPACIÓN DE USUARIOS EN UNA PLATAFORMA EDUCATIVA

La investigación científica en Latinoamérica es liderada principalmente por universidades según la UNESCO, y este proyecto es un trabajo para una institución que intenta aumentar su impacto en la sociedad mediante una plataforma científico-tecnológica, la cual busca conectar investigadores para resolver retos complejos de la humanidad, aunque el área encargada de diseñar la plataforma escasea de herramientas analíticas dentro de la plataforma, por lo que desconoce sus tipos de usuarios y carece de estrategias personalizadas. Este proyecto se centra en la segmentación y caracterización de usuarios usando datos internos y externos de la plataforma, proponiendo estrategias personalizadas para incrementar la participación de estos. Se emplea la metodología CRISP-DM para guiar el proyecto.

Tras el análisis y preparación de los datos, se utilizan técnicas de clustering como K Means y Fuzzy C Means (FCM), identificando 7 segmentos, destacando los Ganadores, Líderes Multipropuestas, Pregrados y Académicos Externos. La segmentación RFM identifica 4 segmentos adicionales, los Destacados, Participantes, Baja Participación y Espectadores, subsegmentados entre usuarios Conectados y Desconectados. Se eligen los segmentos de K Means por su escalabilidad e interpretabilidad, proponiendo 4 estrategias para participación y 4 para colaboración, además de 6 estrategias personalizadas basadas en RFM para reactivar, convertir y retener usuarios. En un caso esperado, se estima que al utilizar las estrategias con foco en la participación en los segmentos de K Means, aumente la participación y propuestas enviadas promedio por convocatoria en un 25%. Por otra parte, utilizando las estrategias de colaboración se espera un aumento del 6% en la cantidad de propuestas enviadas, y utilizando las estrategias de los segmentos de RFM se espera un aumento del 18% en la participación y propuestas enviadas promedio por convocatoria.

Las principales discusiones asociadas al proyecto son la exclusión de ciertas variables de texto y el no uso de los segmentos de FCM, proponiendo alternativas para su utilización. Finalmente, la conclusión principal es que se cumplen los objetivos del proyecto, proponiendo estrategias personalizadas que pueden incrementar la participación y las propuestas en la plataforma. Se le recomienda a la institución repetir el análisis periódicamente con datos actualizados e integrar proyectos que conecten usuarios y convocatorias para aumentar el impacto de la institución en la sociedad.

Dedicado a mis padres, hermanos y hermana.
Los amo

Agradecimientos

Quisiera agradecer en primer lugar a mi madre Sara, por el inmenso amor y apoyo incondicional en cada momento de mi vida. A mi padre Enrique, por siempre preocuparse de que tenga lo necesario para poder cumplir en mis estudios. A mis hermanos Gabriel, Enrique y Bárbara por estar siempre conmigo cuando los he necesitado, son mi pilar fundamental. Gracias a ustedes por todo el sacrificio que han hecho para lograr esto y por los valores que me han enseñado y me han ayudado a ser una mejor persona.

Agradezco también a mi amor Francisca, por amarme, escucharme, acompañarme en cada momento y por siempre creer en que puedo lograr mis objetivos. Mi compañera de aventuras, seguiremos teniendo muchas más.

A mis amigos y amigas que me han acompañado en todo este proceso, por todas las risas y buenos ratos como también el apoyo en los momentos difíciles.

A Juan Pablo Romero, José Nalda y Felipe Vildoso por sus consejos, buena onda, sabiduría y conocimientos que me compartieron para el desarrollo del proyecto. Su guía ha sido clave para lograr los objetivos.

Finalmente, al equipo de trabajo dentro de Brain Food, quienes me acogieron de una forma excelente, me hicieron sentir uno más de la empresa y me contagiaron de buena onda en un ambiente muy grato para desarrollar el proyecto de título. Son todos unos cracks mvp.

Tabla de Contenido

Índice de Tablas

Índice de Figuras

Capítulo 1: Antecedentes Generales

1.1. La consultoría en Chile

El siguiente proyecto se desarrolla en una empresa de consultoría enfocado en la tecnología. Este tipo de empresas entregan un servicio para aconsejar y asesorar a otras empresas en cómo usar las tecnologías de la información con el fin de alcanzar sus objetivos empresariales (Conasa, s.f.). En Chile, en el 2022 existen cerca de 81 mil empresas relacionadas a la consultoría, en donde se ha visto un incremento de del 260% en comparación con la cifra que había en el 2005. Esto se puede observar en el A.1. (SII, s.f.)

Dentro de los tipos de consultoras, la empresa en la que se desarrolla el proyecto en el presente informe está clasificada como consultora de gestión en el Servicio de Impuestos Internos, las cuales representan el 46% del total de consultoras. A pesar de que la empresa esté catalogada así, también compite directamente con las consultoras informáticas y técnicas, las cuales en conjunto con las de gestión representan los tipos de consultoras más comunes en el año 2022. Esto se puede ver en el A.1. (SII, s.f.)

La industria de la consultoría cuenta tanto con actores locales como internacionales que cuentan con décadas de experiencia en la industria. Las consultoras con mayor presencia internacional y que se encuentran dentro de Chile son Deloitte Touche Tohmatsu Limited, Accenture PLC, PricewaterhouseCoopers LLP, Capgemini SE y McKinsey Company (Mordor Intelligence, s.f.). De estas consultoras las que prestan servicios TI al igual que Brain Food son Accenture PLC y Capgemini SE.

1.2. Las universidades latinoamericanas y la investigación

El cliente con quien se desarrolla el proyecto es una institución educativa latinoamericana con un alto compromiso en la innovación y en la investigación científica. Los países latinoamericanos representan aproximadamente el 2% de la inversión mundial en investigación y desarrollo (I+D) en el año 2021, siendo un número bajo en comparación a Norteamérica (30%), Europa (24%) y Asia (42%). Otro rasgo de la inversión latinoamericana es la disparidad en la región, siendo Brasil el país que es responsable del 62% de la inversión, seguido por México (13%) y Argentina (9%). Esto se puede ver en el Anexo A.2. (RICYT, 2023)

Una parte considerable de la producción de conocimiento científico y tecnológico en América Latina se lleva a cabo en las universidades. Si bien las instituciones universitarias

son importantes en el desarrollo en todo el mundo, en América Latina estas se destacan como actores centrales. Esto se evidencia en el hecho de que su peso específico supera ampliamente al de las universidades en los países industrializados, ya que el 61% de los investigadores medidos en equivalencia a jornada completa latinoamericanos están radicados en las universidades. (UNESCO, 2020)

En relación con la producción bibliográfica en Latinoamérica ha crecido un 84% entre 2008 y 2018, pasando del 3% del total mundial al 5%. En este rol, los casos de Chile, Colombia y Brasil son los más destacados, ya que la participación de los autores radicados en universidades se aproxima al 90% del total de artículos científicos firmados por autores del país. Esto se puede observar en el Anexo A.2. (UNESCO, 2020)

1.2.1. Impacto de la IA en la Investigación

El campo de la inteligencia artificial (IA) ha experimentado un crecimiento notable en la producción científica en la última década, especialmente a partir del 2018. El número de artículos publicados a nivel mundial ha alcanzado un total de 230.000 documentos en 2022 aproximadamente, representando un aumento de casi seis veces en comparación a una década atrás. La IA abarca una gran gama de disciplinas científicas, convirtiéndose en una herramienta poderosa para acelerar la producción de conocimiento en diversas áreas de la ciencia y la tecnología. (RICYT, 2023)

La inclusión de la IA en la ciencia es el resultado de décadas de avances tecnológicos y una creciente comprensión de su potencial, logrando una sinergia entre el humano y la máquina, acelerando y potenciando el progreso científico. Unas de las ventajas del uso de la IA en la investigación, es la colaboración y acceso global, en donde la investigación se ha vuelto más colaborativa gracias a plataformas de investigación colaborativa basadas en IA que permiten a investigadores de distintos lugares trabajar juntos de manera eficiente, compartiendo datos y hallazgos en tiempo real. (Doctrina Qualitas, 2023)

1.3. Descripción general de la empresa

1.3.1. Brain Food

Brain Food es una consultora fundada el año 2015, especializada en transformación digital, analítica avanzada, ingeniería de datos, BI, automatización y software factory. (Brain Food, 2024)

La organización se fundó con el fin de unificar tres áreas que muchas veces parecen no conversar dentro de las empresas, como lo son estrategia y/o negocios, estadística y la tecnología. Creen en un modelo que promueve la colaboración y el intercambio de

perspectivas entre el consultor y el cliente, distinguiéndose por su gran capacidad de integrar la transformación digital y los datos en las estrategias de las compañías.

La oficina principal está ubicada en Santiago de Chile y provee servicios a más de 16 países alrededor del mundo, entre ellos México, Perú y Colombia. Hasta el presente año, Brain Food ha realizado proyectos en más de 15 industrias, contando con más de 40 clientes y habiendo completado más de 150 proyectos. La empresa es considerada como una de mediano tamaño debido a que cuenta con más de 50 empleados y menos de 200. La información complementaria de la empresa se encuentra en el A.3.

1.3.2. Cliente en la industria educativa

Por razones de confidencialidad solicitadas por el cliente de Brain Food, se procederá a anonimizar todos los datos y estrategias relacionadas con dicho cliente en este informe, proporcionando únicamente el contexto esencial para la realización del proyecto.

El cliente opera en el sector de la educación privada y es una institución sin fines de lucro que gestiona múltiples campus. Con una trayectoria de más de 50 años, la institución cuenta con una matrícula de más de 50,000 estudiantes. La visión del cliente se centra en la innovación y el emprendimiento como motores para el progreso humano.

1.3.2.1. Plataformas

La institución educativa cuenta con un área encargada de ser el nodo de innovación capaz de generar nuevas propuestas de valor basadas en el talento de sus estudiantes, ampliando el impacto de la institución a través del desarrollo de nuevas plataformas. Esta área de la institución será llamada durante el informe como “Desarrollo de Plataformas”.

Hasta la fecha se han diseñado varias plataformas en conjunto con otra área de la institución llamada “Desarrollo Tecnológico y Emprendimiento” (DTE), de las cuales se encuentran activas las siguientes:

- Plataforma de tutorías personalizadas impartidas por estudiantes destacados, destinada a reforzar conocimientos en diversas materias para estudiantes de preparatoria.
- Plataforma que conecta a la institución con diversas organizaciones para abordar desafíos académicos y enriquecer la formación estudiantil, fomentando el trabajo colaborativo entre estudiantes, profesores y agentes de diferentes sectores.
- Plataforma científico-tecnológica recientemente lanzada, que facilita convocatorias para abordar desafíos sociales complejos, promoviendo la colaboración entre investigadores, emprendedores e innovadores para desarrollar

soluciones innovadoras que generen un impacto en la sociedad. Para mayor simplicidad a esta plataforma se le llamará “Plataforma X”.

Esta última en la cual se basa el proyecto, en sus inicios, empezó lanzando convocatorias solo para estudiantes y ex alumnos de la institución, pero en sus últimas convocatorias se ha abierto a recibir participantes de otras instituciones, incluyendo instituciones que están fuera del país de origen. Actualmente ha alcanzado más de dos mil usuarios y espera seguir creciendo esta cantidad a través de nuevas convocatorias en distintos idiomas y tipo de público.

La institución monetiza la Plataforma X de dos maneras. La primera es a través del flujo de dinero generado por los premios establecidos en colaboración con industrias o empresas que lanzan las convocatorias, reteniendo un porcentaje para asegurar la autosustentabilidad de la plataforma a corto plazo. La segunda forma consiste en convertir las soluciones desarrolladas en patentes, que pueden venderse, licenciarse o transformarse en startups o empresas. De estas iniciativas, la institución educativa obtiene un porcentaje de las ganancias, al igual que la industria colaboradora y los investigadores que desarrollan las soluciones para los desafíos sociales planteados.

Capítulo 2: Descripción y justificación del proyecto

2.1. Identificación y justificación del problema

La institución educativa busca mejorar sus plataformas mediante la implementación de análisis avanzados con el objetivo de desarrollar mejores productos para sus usuarios. En la actualidad, no cuenta con un entorno en la nube que proporcione herramientas disponibles de forma integrada; en su lugar, dispone de fuentes de datos aisladas en Azure y Google Cloud. Además, cuenta con diccionarios parciales de estas fuentes y realiza análisis superficiales de los datos en algunas de ellas.

Con respecto a la Plataforma X, se dispone de un algoritmo de match entre usuarios basado en embeddings[1], pero la falta de análisis y cruce de datos de diversas fuentes impide comprender las características y las interacciones de los usuarios en las convocatorias lanzadas en la plataforma.

Se comenzó con un levantamiento de información con los owners del negocio con el fin de tener un mayor entendimiento del problema y de las oportunidades de desarrollo, en donde se evidenció que uno de sus dolores es la falta de estrategias de negocio personalizadas que incentiven la participación en la Plataforma X. Esto afecta directamente a la institución ya que uno de sus objetivos al crear la plataforma es aumentar su impacto en la sociedad y a la hora de lanzar retos complejos para la sociedad dentro de la plataforma se espera tener las mejores soluciones innovadoras posibles. La finalidad de la plataforma se ve comprometida ante la falta de estrategias personalizadas, afectando la participación de los usuarios en la plataforma, implicando así menor cantidad de propuestas de solución, disminuyendo el impacto de la institución en la sociedad.

Por otra parte, se comenzaron a explorar los datos, identificando que solo han participado el 40% de los usuarios que se han inscrito en la plataforma en las convocatorias. De estos se tiene un promedio de 160 usuarios que participan en convocatorias, alcanzando hasta mil usuarios en una ocasión específica. Sin embargo, excluyendo esta convocatoria, el promedio se reduce a 55 participantes por convocatoria. Además, se tiene que el 43% los participantes forman equipos en solitario. Estos datos reflejan la baja participación y colaboración que han tenido los usuarios en la plataforma.

[1] Los embeddings son una técnica de procesamiento de lenguaje natural que convierte el lenguaje humano en vectores matemáticos. Estos vectores son una representación del significado subyacente de las palabras, lo que permite que las computadoras procesen el lenguaje de manera más efectiva.

Si no se aplican herramientas analíticas, no se tendrá un conocimiento de los usuarios que participan, dificultando la implementación de estrategias para su retención en futuras convocatorias y la formación de equipos más efectivos, incluso perdiendo la posibilidad de colaboración con usuarios nuevos en la plataforma. Es por esto por lo que buscar una solución a estos problemas podría traer grandes beneficios para la institución educativa, pudiendo generar un mayor impacto en la sociedad a través de la Plataforma X.

2.2. Descripción del proyecto

Brain Food ha acordado con el cliente realizar diversos proyectos de modelos de analítica avanzada para las diversas plataformas con las que cuentan y así potenciar el impacto y capacidades de estas. Para esto se planea crear un nuevo nodo dentro del área de Desarrollo de Plataformas (DP) llamado "Sistema de Conexión de Inteligencia" (SCI), el cual será encargado de procesar múltiples fuentes de datos para generar productos analíticos y modelos para los usuarios. El objetivo principal del SCI será conectar la analítica avanzada e inteligencia artificial para potenciar el talento de cada usuario.

Para abordar el problema mencionado anteriormente, este proyecto se centra en el desarrollo de unas de las soluciones analíticas que será parte del SCI en la Plataforma X, la cual es la segmentación de usuarios utilizando técnicas de clustering. Para esto, se dispone de tres fuentes internas otorgadas por el área de Desarrollo de Plataformas, una de estas bases cuenta con información de los usuarios, otra con información de las convocatorias y, finalmente, una con información de los equipos que participan en las convocatorias. Además, se utilizarán fuentes de datos externas que complementen las existentes en la institución que permitan caracterizar mejor a los usuarios de la plataforma.

Tras realizar la segmentación de los usuarios de la plataforma, estos se caracterizan con el fin de proponer estrategias con foco en el aumento de la participación de usuarios en la Plataforma X. Estas estrategias serán propuestas al área de Desarrollo de Plataformas y serán estos quienes tomen la decisión de si las estrategias son implementadas o no.

Se espera que con este proyecto se pueda solucionar el problema mencionado y justificado anteriormente, consiguiendo así un mayor conocimiento de los usuarios con todo el proceso analítico que conlleva la segmentación de los usuarios y también se podrá solventar la falta de estrategias personalizadas que aumenten la participación de los usuarios dentro de la Plataforma X. Además, se espera que este nuevo nodo creado dentro del área de DP pueda continuar con la actualización periódica de los modelos analíticos creados en este proyecto y mejorarlos continuamente.

2.3. Objetivos del proyecto

2.3.1. Objetivo General

Desarrollar estrategias de negocio dirigidas a los segmentos de investigadores que se identificarán utilizando técnicas de clustering, que permitan aumentar la participación y colaboración de los usuarios en una plataforma científico-tecnológica.

2.3.2. Objetivos Específicos

- Desarrollar análisis de datos de las diversas fuentes de datos internas y externas de la Plataforma X.
- Diseñar e implementar una base de datos de usuarios con variables relevantes para el negocio a través de cruces entre las fuentes de datos disponibles.
- Desarrollar modelos de segmentación de usuarios de la Plataforma X utilizando técnicas de clustering en Azure Databricks[2].
- Caracterizar los clústeres encontrados a través del uso de algoritmos de clustering.
- Proponer estrategias personalizadas para los distintos segmentos encontrados en la Plataforma X.

2.4. Alcances

El proyecto de colaboración entre la institución educativa y Brain Food aborda diferentes aspectos en cada una de las tres plataformas activas del Área de Desarrollo de Plataformas. Por esta razón, dentro del equipo se establecen divisiones para atender cada uno de estos aspectos. Específicamente para la Plataforma X, se han identificado cinco soluciones destinadas a mejorar la experiencia de usuario y el rendimiento de la plataforma. Estas soluciones son:

1. Segmentación de usuarios
2. Recomendación de usuarios
3. Analítica de retos y soluciones

[2] Azure Databricks es una plataforma de análisis unificada y abierta para crear, implementar, compartir y mantener soluciones de datos, análisis e IA de nivel empresarial a escala.

4. Ciclo de vida de los usuarios

5. Match de soluciones

Dado que no es posible abarcar cada una de estas soluciones en el proyecto que se detalla en el presente informe, el alcance del proyecto se restringe únicamente a la segmentación de los usuarios y la formulación de estrategias de negocio personalizadas. Esto debido a que es el primer paso para empezar a entender y conocer a los usuarios dentro de las plataformas y puede servir como base para cada uno de los próximos desarrollos y también porque el tiempo destinado al desarrollo del proyecto solo podrá incorporar la segmentación de usuarios.

Una vez completada la propuesta de estrategias para aumentar la participación y colaboración de los usuarios en las convocatorias, éstas se pondrán a disposición de la contraparte. Será responsabilidad de la institución educativa implementar y evaluar estas estrategias a lo largo del tiempo. Por lo tanto, las etapas de implementación y evaluación quedan fuera del alcance del presente proyecto.

Capítulo 3: Marco Conceptual

Este proyecto se vincula directamente con dos áreas de la Ingeniería Civil Industrial. La primera es el área de Tecnologías de la Información, ya que el proyecto emplea la minería de datos para extraer información de los usuarios de la plataforma, la cual es utilizada para generar conocimiento sobre estos. La segunda área es la de Marketing, debido a la propuesta de estrategias personalizadas dirigidas a los segmentos identificados mediante técnicas de clustering, con el objetivo de aumentar su participación en la Plataforma X.

3.1. Minería de datos

La minería de datos es el proceso de descubrir patrones y conocimientos interesantes a partir de grandes cantidades de datos. Las fuentes de datos pueden ser bases de datos, almacenes de datos, la Web, otros repositorios de información o datos que se introducen en el sistema de forma dinámica. (Han, Kamber, & Pei, 2012)

La toma de decisiones, el control de procesos, la gestión de la información y el procesamiento de consultas son sólo algunas de las aplicaciones de la experiencia recién descubierta. Como resultado, la minería de datos se reconoce como una de las tecnologías de bases de datos modernas más apasionantes de la industria de la información, así como una de las fronteras más importantes de los sistemas de bases de datos. (Ogunleye, 2021)

3.2. Clustering

El análisis de clústeres, o simplemente clustering, es el proceso de dividir un conjunto de objetos de datos (u observaciones) en subconjuntos. Cada subconjunto es un clúster, de forma que los objetos de un clúster son similares entre sí, pero diferentes de los objetos de otros clústeres. El conjunto de clústeres resultante de un análisis de clústeres puede denominarse clustering. En este contexto, diferentes métodos de clustering pueden generar diferentes agrupaciones en el mismo conjunto de datos. La partición no la realiza el ser humano, sino el algoritmo de clustering. Por lo tanto, el clustering es útil en la medida en que puede conducir al descubrimiento de grupos previamente desconocidos dentro de los datos. (Han, Kamber, & Pei, 2012)

El clustering se ha utilizado ampliamente en muchas aplicaciones, como la inteligencia empresarial, el reconocimiento de patrones de imagen, la búsqueda en Internet, la biología y la seguridad. En inteligencia empresarial, el clustering puede utilizarse para organizar un gran número de clientes en grupos, en los que los clientes de un grupo comparten

características muy similares. Esto facilita el desarrollo de estrategias empresariales para mejorar la gestión de las relaciones con los clientes. (Han, Kamber, & Pei, 2012)

3.3. Algoritmos de clustering utilizados

3.3.1. K Means

Supongamos que un conjunto de datos, D, contiene n objetos en un espacio euclidiano. Los métodos de partición distribuyen los objetos de D en k conglomerados, C_1,..., C_k, es decir, $C_i \subset D$ y $C_i \cap C_j = \emptyset$ para ($1 \leq i,j \leq k$). Se utiliza una función objetivo para evaluar la calidad de la partición, de forma que los objetos de un clúster sean similares entre sí, pero disímiles a los objetos de otros clústeres. Es decir, la función objetivo busca una alta similitud intra-clúster y una baja similitud inter-clúster.

Una técnica de partición basada en el centroide utiliza el centroide de un clúster, C_i, para representarlo. Conceptualmente, el centroide de un clúster es su punto central. El centroide puede definirse de varias formas, como la media o medoide de los objetos (o puntos) asignados al clúster. La diferencia entre un objeto $\rho \in C_i$ y c_i, el representante del clúster se mide por $dist(\rho, c_i)$, donde $dist(x, y)$ es la distancia euclidiana entre dos puntos x e y. La calidad del clúster C_i puede medirse por la variación dentro del clúster, que es la suma del error cuadrático (SSE) entre todos los objetos en C_i y el centroide c_i, esto queda definido como se muestra en la Ec. 1.

$$E = \sum_{i=1}^{k} \sum_{\rho \in C_i} dist(\rho, c_i)^2 \qquad \textit{Ec. 1}$$

Donde E es la suma del error al cuadrado para todos los objetos del conjunto de datos, ρ es el punto en el espacio que representa un objeto determinado; y c_i es el centroide del clúster C_i (tanto ρ como c_i son multidimensionales). En otras palabras, para cada objeto de cada clúster, se eleva al cuadrado la distancia del objeto a su centro de clúster y se suman las distancias. Esta función objetivo intenta que los k clústeres resultantes sean lo más compactos y separados posible. (Han, Kamber, & Pei, 2012)

3.3.2. Fuzzy C Means

En situaciones comunes, un dato puede estar cercano a dos clústeres, dificultando su etiquetado en uno u otro debido a la frecuencia con la que presenta características de distintos clústeres. El algoritmo Fuzzy C Means (FCM) fue diseñado para abordar este problema.

El algoritmo FCM asigna a cada dato un valor de pertenencia dentro de cada clúster y por consiguiente un dato específico puede pertenecer parcialmente a más de un clúster. A diferencia del algoritmo K Means clásico que trabaja con una partición dura, FCM realiza una partición suave del conjunto de datos, en tal partición los datos pertenecen en algún grado a todos los clústeres; una partición suave se define formalmente como sigue: (Rojas, Chavarro, & Moreno, 2008)

Sea X un conjunto de datos y x_i un elemento perteneciente a X. Se dice que una partición $P = \{C_1, C_2, \dots, C_c\}$ es una partición suave de X si y solo si las condiciones de la Ec. 2 y Ec. 3 se cumplen:

$$\forall x_i \in X, \qquad \forall C_j \in P \quad 0 \leq \mu c_j(x_i) \leq 1 \qquad \textit{Ec. 2}$$

$$\forall x_i \in X, \qquad \exists C_j \in P \quad tal\ que\ \ \mu c_j(x_i) > 0 \qquad \textit{Ec. 3}$$

Donde $\mu c_j(x_i)$ denota el grado en el cuál x_i pertenece al cluster C_j.

FCM produce una partición suave restringida y para hacer esto la función objetivo E se extiende de dos maneras, por un lado, en la Ec. 4 se incorporan los grados de pertenencia difusos de cada dato en cada clúster, por otro lado, se introduce un parámetro adicional m que sirve de peso exponente en la función de pertenencia, así la función objetivo extendida E_m es:

$$E_m = \sum_{i=1}^{k} \sum_{x_k \in X} (\mu c_j(x_k)^m) \cdot dist(x_k, c_i)^2 \qquad \textit{Ec. 4}$$

Donde P es una partición difusa del conjunto de datos X formada por $C_1, C_2, \dots, C_k$. El parámetro m es un peso que determina el grado en el cual los miembros parciales de un clúster afectan el resultado. Al igual que K Means clásico, FCM también intenta encontrar una buena partición mediante la búsqueda de los prototipos c_i que minimicen la función objetivo E_m y adicionalmente, FCM también debe buscar las funciones de pertenencia μc_i que minimicen a E_m.

3.3.3. RFM

La segmentación RFM es una técnica utilizada en marketing y análisis de clientes para clasificar a los clientes en función de su historial de compra y comportamiento. RFM son las iniciales de tres dimensiones clave Recency, Frequency y Monetary definidas en la Tabla 1. (Khajvand & Jafar Tarokh, 2010).

Para este proyecto en específico se utilizará esta técnica, pero redefiniendo el significado del RFM. Los parámetros adaptados se pueden observar en la Tabla 1.

Tabla 1: Definición de parámetros de RFM y redefinición de los parámetros para el proyecto en específico.

Fuente: Elaboración propia.

Parámetro	Definición	Parámetro adaptado
Recencia (R)	Última fecha de compra en un periodo determinado	Días desde la última conexión a la plataforma
Frecuencia (F)	Número de compras en un periodo determinado	Cantidad de equipos en los que participa en la plataforma
Monetary (M)	Valor de las compras en un periodo determinado	Cantidad de equipos con los que manda propuestas en la plataforma

3.4. Marketing personalizado

El marketing o mercadotecnia personalizados es una estrategia que aprovecha los datos y análisis para brindar experiencias más personales a los clientes o en este caso los usuarios. Su objetivo es responder a las necesidades e intereses individuales de los usuarios para crear relaciones más valiosas. (Sordo, 2023)

El marketing personalizado ha traído muchos beneficios para las empresas. Incluso los especialistas de marketing afirman ver un aumento en sus ventas cuando utilizan experiencias personalizadas. Su aplicación también puede incrementar la participación del cliente con la marca y generar más ingresos. (Sordo, 2023)

Esto nos lleva a entender que la comprensión de los clientes es clave para impulsar las empresas y que, a pesar de ser una tarea compleja, aplicando marketing personalizado se puede fortalecer la empresa.

Para poder aplicar estrategias de marketing personalizado en primer lugar se debe conocer a los clientes y usuarios, tanto sus preferencias como su comportamiento. Para esto se deben organizar los datos con los que se cuenten de estos usuarios para poder segmentarlos, lo cual será una parte clave de establecer estrategias personalizadas. Posteriormente a la segmentación se deben establecer los canales de comunicación adecuados para llegar a los distintos usuarios y finalmente crear las estrategias o campañas para lograr los objetivos establecidos. (Sordo, 2023)

Capítulo 4: Metodología

El problema anteriormente expuesto se aborda siguiendo la metodología CRISP-DM (IBM), diseñada para guiar proyectos de minería de datos. Se decide utilizar esta metodología ya que a diferencias de otras comúnmente utilizadas (SEMMA, Catalyst y KDD) plantea una constante retroalimentación entre las diferentes etapas, lo que permite avanzar o retroceder en cada etapa sin mayores inconvenientes. Además, cuenta con una etapa importante para el proyecto, como lo es la comprensión del negocio y de los datos. CRISP-DM describe fases, tareas y relaciones entre ellas, como se muestra en el Anexo B.

En particular dada las características del problema, la metodología a utilizar se estructura de la siguiente forma:

1. **Entendimiento del negocio:** Comprender cómo opera la Plataforma X, incluyendo la formación de convocatorias y la invitación de usuarios.

2. **Comprensión de los datos:** Analizar los datos utilizando Python, mediante gráficos y tablas, para evaluar su calidad y obtener información relevante tanto de fuentes internas como externas.

3. **Preparación de los datos:** Cruzar las distintas fuentes para crear una base de usuarios únicos con variables relevantes, aplicando supuestos y transformaciones necesarias.

4. **Modelado:** Utilizar modelos de machine learning (K Means y Fuzzy C Means) y RFM para segmentar a los usuarios en grupos con características y comportamientos similares.

5. **Evaluación:** Evaluar la coherencia de los modelos de clustering mediante métricas y conocimiento experto para seleccionar el mejor enfoque de segmentación.

6. **Despliegue:** Utilizar los resultados para diseñar estrategias personalizadas que aumenten la participación de los usuarios, proponiendo estas estrategias a la institución educativa para su futura implementación.

Para la realización del proyecto utilizando la metodología CRISP-DM, se cuenta con insumos relevantes, como fuentes de datos internas y externas a la Plataforma X. Estos insumos son esenciales para el entendimiento del negocio, la comprensión de los datos y el modelado. Los datos se trabajarán en la plataforma Azure Databricks, donde se crearán distintos cuadernos para manipularlos utilizando el lenguaje Python.

Capítulo 5: Desarrollo y Resultados

5.1. Entendimiento del negocio

5.1.1. Plataforma X

La Plataforma X es co-diseñada por Desarrollo de Plataformas (DP) y Desarrollo Tecnológico y Emprendimiento (DTE). DP maneja el desarrollo técnico, mientras que DTE genera convocatorias, establece relaciones con industrias para identificar desafíos científicos-tecnológicos, y determina los premios para las mejores soluciones. Un ejemplo de convocatoria se puede encontrar en el Anexo C.1. Las convocatorias se comunican a través de:

- Avisos a los usuarios registrados.
- Canales de comunicación de la institución educativa.
- Promoción en congresos de investigación educativa.

Los equipos, que pueden formarse con investigadores internos o externos, tienen fechas límites para enviar sus propuestas. Los evaluadores, que son investigadores de la institución educativa o de otras universidades, califican las propuestas según criterios de innovación, factibilidad y otros. Los ganadores reciben financiamiento para desarrollar y patentar sus soluciones junto a las instituciones colaboradoras.

5.2. Comprensión de los datos

En la fase inicial del proyecto, se tres bases de datos principales dentro de la plataforma: usuarios, equipos y convocatorias. En esta sección se destacarán los hallazgos más relevantes del EDA realizado a las bases mencionadas junto a las fuentes de datos externas utilizadas.

5.2.1. Fuentes Internas

5.2.1.1. Usuarios

Antes de comenzar con el EDA desarrollado en esta base es importante mencionar que el equipo recibió una base de datos de usuario con 2.578 usuarios en donde se hicieron las siguientes definiciones:

- Usuario registrado: Usuarios que se registró en la plataforma y firmó el aviso de privacidad. (1.740 usuarios)

- Usuario activo: Usuario registrado que ha ingresado alguna vez a la plataforma tras su registro o ha participado en un equipo. (1.113 usuarios)

Tras esta explicación, la base de datos a la que se le realiza el EDA es solo de los usuarios activos dentro de la plataforma. Las variables que contiene esta base de datos se pueden observar en el Anexo D.1. De esta base se encontraron los siguientes insights:

- A pesar de poder tener usuarios investigadores, innovadores y emprendedores, casi en la totalidad de los usuarios son investigadores. Esto se ve en la Figura 1.

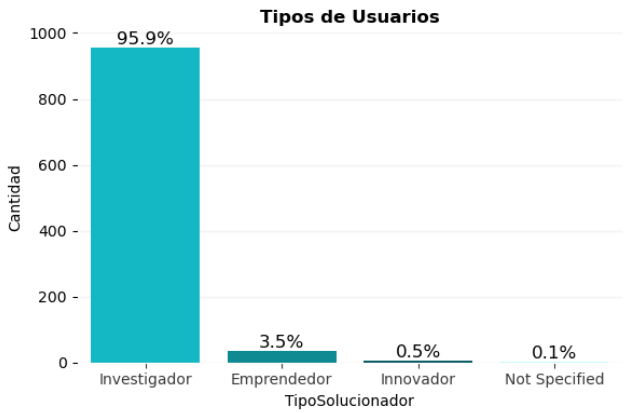


Figura 1: Tipos de Usuarios en la Plataforma X. Fuente: Elaboración propia.

- La mayoría de los usuarios activos se han inscrito (que han mandado propuestas en alguna convocatoria). Esto se ve en la Figura 2.

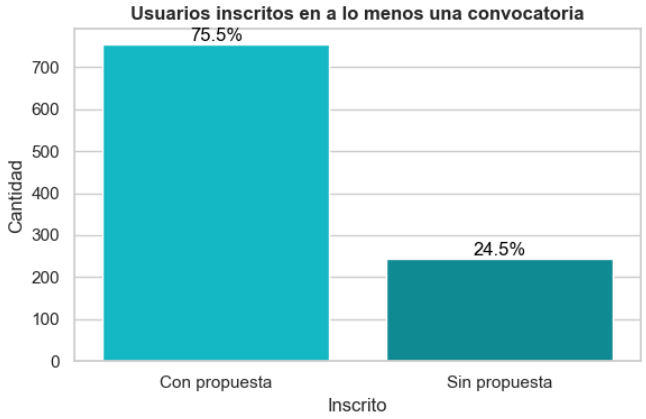


Figura 2: Usuarios inscritos en la Plataforma X. Fuente: Elaboración propia.

- 73% de los usuarios activos han formado un equipo en alguna convocatoria y el 48% de estos se ha inscrito en más de un equipo.
- El 15% de los usuarios activos han participado en más de una convocatoria.

5.2.1.2. Convocatorias

Al realizar el EDA en esta base se pudo identificar que se cuenta con información de 10 convocatorias, de las cuales solo 4 han finalizado y 4 han sido abiertas al público externo a la institución. Generalmente, duran entre 5 y 69 días en donde se han inscrito entre 4 y 266 equipos. Además, se ha identificado que en las convocatorias finalizadas han ganado 3 equipo en cada una en promedio.

De esta base se obtuvieron siguientes insights:

- En promedio participan 49 equipos por convocatoria, pero sin contar la 10 que tuvo aproximadamente 250 equipos, el promedio disminuye a 25 equipos por convocatoria. Esto se analiza del gráfico de la Figura 3.

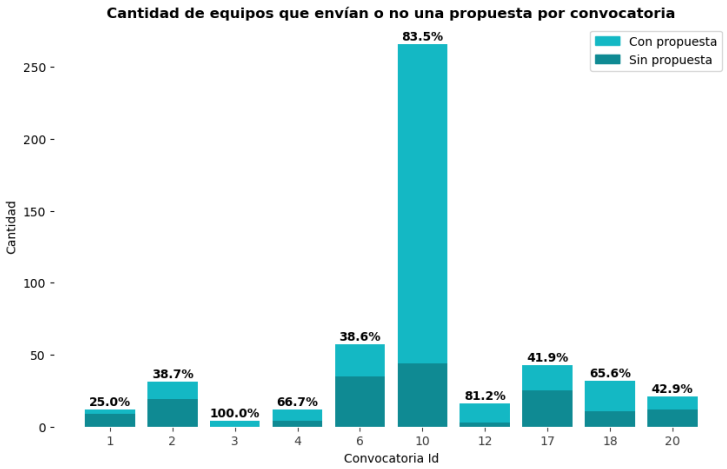


Figura 3: Cantidad de equipos con propuesta o no por convocatorias. Fuente: Elaboración propia.

- En las convocatorias abiertas al público externo se ha encontrado la mayor participación de los usuarios, pero se observó que, a pesar de ser abiertas al público externo, casi en su totalidad son de la institución. Esto se ve en la Figura 4.

Figura 4: Proporción de participantes en convocatorias abiertas. Fuente: Elaboración propia.

5.2.1.3. Equipos

En relación con los equipos se observa que en total se han inscrito 494 equipos en las convocatorias y solo un 12% corresponde a convocatorias finalizadas. De esta base se encuentran los siguientes insights:

- El 67% de los equipos inscritos han mandado al menos una propuesta en las convocatorias. Esto se ve en la Figura 5.

Figura 5: Proporción de equipos que envían o no una propuesta. Fuente: Elaboración propia.

- El 25% de los usuarios ha sido líderes de un equipo y de estos el 46% ha sido líder en más de una ocasión.
- Los equipos tienen en promedio 3.4 usuarios inscritos.
- El 57% de los equipos son de múltiples integrantes, los cuales suelen mandar más propuestas que los equipos en solitario. Esto se ve en la Figura 6.

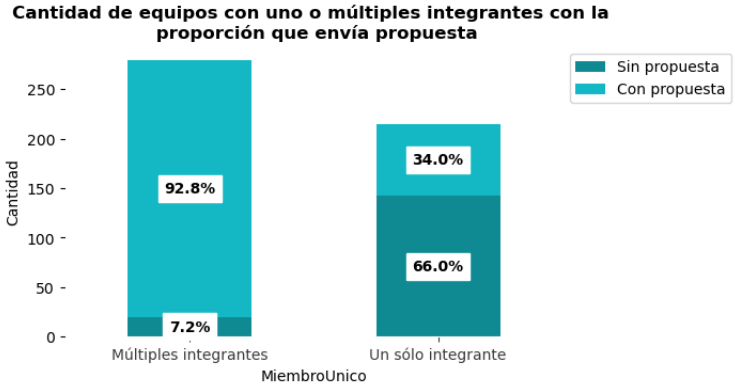


Figura 6: Cantidad de equipos con uno o múltiples integrantes con la proporción que envían propuestas. Fuente: Elaboración propia.

5.2.2. Fuentes Externas

El uso de las fuentes externas es principalmente para obtener nuevas variables que ayuden a perfilar a los usuarios que más relacionados están en el mundo de la investigación, lo que puede ser de mucho aporte considerando que se tienen muchos investigadores en la plataforma.

5.2.2.1. Scopus

De la Fuente externa de Scopus[3] se tienen tres tablas, una es *author* que es información de los autores y la cantidad de publicaciones que tienen, *affiliation* que contiene

[3] Scopus es una base de datos bibliográfica iniciada en 2004, de resúmenes y citas de artículos de revistas científicas.

información de las afiliaciones (universidades) en las que están asociados los autores y finalmente *publication* que tiene información de más de 33.000 publicaciones de los autores que están en la Plataforma X.

De la tabla *author* se tiene un cruce con 62% de los usuarios activos en donde se encuentran variables descriptivas como los temas principales de los autores y variables estadísticas como la cantidad de publicaciones.

En relación con la base *affiliation* solo tiene 3 columnas y ninguna aporta demasiada información a lo que contienen las fuentes internas, de esta base solo se obtiene que los países más frecuentes en los autores son México, Chile y España.

En la base de datos *publication*, el 49% de los usuarios activos tienen información registrada. En esta tabla se encuentran datos interesantes sobre la cantidad de autores, la cantidad de citas que tienen sus publicaciones, las fechas de las publicaciones y los tipos de publicaciones que tienen los autores.

En conclusión, de esta fuente externa se encuentran datos y variables interesantes para completar y entender mejor los investigadores con los que se cuenta en la plataforma, añadiendo información de la relevancia que tienen sus publicaciones y el tiempo que llevan publicando en Scopus. El análisis más detallado de las variables relevantes encontradas en esta fuente se encuentra en el Anexo D.2.

5.2.2.2. Orcid

Orcid[4] tiene una base de datos en donde se obtienen principalmente variables descriptivas de los autores, como lo es la biografía, keywords y el país de donde provienen. Se le veía potencial en cuanto a la información que se podría complementar a las fuentes internas, pero al realizar el cruce con estas, solo el 2% de los usuarios activos tenían información en esta fuente, por lo que se decide descartar el uso de esta base por ahora.

5.2.2.3. Crossref

Crossref[5] alberga información que guarda similitudes con *publication* de Scopus en cuanto a la naturaleza de sus registros. Ambas plataformas proporcionan fundamentalmente datos sobre títulos de publicaciones, autores (que pueden ser múltiples), referencias bibliográficas, citas recibidas, número de páginas, entre otros. Sin embargo, es importante destacar que Crossref cuenta con un total de 3.405 registros de publicaciones únicas, lo que representa solo el 10% aproximadamente de la cantidad de publicaciones que se

[4] ORCID (Open Research and Contributor ID) es un identificador único que tiene como principal finalidad proporcionar a los investigadores un código de autor persistente e inequívoco que distinga claramente su producción científica y evite confusiones vinculadas con la autoría científica. (ORCID, s.f.)
[5] Organización cuyo objetivo es hacer que los resultados de todas las investigaciones no solo sean fáciles de encontrar, citar y vincular, sino que también sea sencillo evaluarlos y reutilizarlos. (Crossref, s.f.)

encuentran disponibles en Scopus. Al realizar el cruce con la Plataforma X se obtuvo que el 29% de los usuarios activos poseen información de Crossref.

Estas variables pueden ser útiles para complementar las de Scopus y mejorar la calidad de las variables relacionadas a sus publicaciones para los usuarios que posean esta información de Crossref. El análisis más detallado de estas variables se encuentra en el Anexo D.2.

5.3. Preparación de los datos

En esta etapa se realizarán los cruces entre las distintas fuentes para conseguir una base de usuarios únicos y se tomarán supuestos para completar los datos nulos existentes.

En primer lugar, se filtró por el rol del usuario, quedando solamente con los solucionadores y quitando a los administradores, pasando así de 1.113 a 1.093 usuarios. Con el objetivo de crear una base de datos con registros únicos por usuario con variables relevantes para el negocio, se realizaron merge entre distintas fuentes internas y externas mencionadas anteriormente, la relación entre las distintas tablas se puede ver en el Anexo E.1. Además, se crearon nuevas variables numéricas a partir de fechas participaciones, de publicaciones y variables de texto. La base *user* con usuarios activos queda con un total de 84 columnas y 1093 filas que corresponde cada una a un usuario.

Tras la creación de esta base, se eliminan las variables de las fuentes externas con poco cruce y las variables de texto que no pueden ser utilizadas en los modelos de clustering en Python, dentro de las cuales se encuentran variables como nombres de los usuarios, experiencia académica y títulos de las publicaciones, identificadores de las distintas fuentes y variables de fechas. Con esto la base queda con 28 columnas y 1093 filas.

Después de realizar los distintos cruces entre las fuentes, crear nuevas variables numéricas y eliminar las que no son útiles en este momento, se toman 10 supuestos basados en la racionalidad para completar los datos nulos de la base de usuarios que se segmentará. Estos supuestos se encuentran en el Anexo E.2.

Para evitar que las variables externas de publicaciones eclipsen a las relevantes de la Plataforma X. Se ha decidido agrupar 12 variables de publicaciones en una única variable llamada "ImportanciaPub", que reflejará la importancia de los usuarios en el mundo de las publicaciones. Esto se puede observar en el Anexo E.3.

Con esto queda lista la base de datos, sin datos nulos y con 16 variables numéricas listas para aplicar en los algoritmos de clustering. Las variables y su descripción se pueden encontrar en el Anexo E.4.

5.4. Modelado

5.4.1. K Means

En primer lugar, se normalizan las variables para que tengan la misma escala y evitar la sensibilidad del K Means al tener escalas distintas entre las variables. Para esto se ocupa la función *StandardScaler* de la librería *sklearn*.

Tras la estandarización se observa la correlación entre las variables a través de una matriz de correlaciones, en donde mientras el valor esté más cercano a 1 más correlacionadas están y más cercano a 0 lo contrario. Esto se puede encontrar en el Anexo F.1.

Se eliminan las variables "NumEquipos", "ProporcionConvGanadas", "CantidadConv", "NivelAcademico" y "ProfileCompletion" debido a una alta correlación (superior a 0,8) y que pueden ser explicadas también por otras variables. Con esto quedan 10 variables a utilizar sin contar la id de los usuarios.

Después de eliminar las variables correlacionadas, se utiliza el método del codo y el método de la silueta para escoger la cantidad de clústeres antes de ejecutar el modelo, la visualización de este método se puede ver en las Figura 7 y Figura 8.

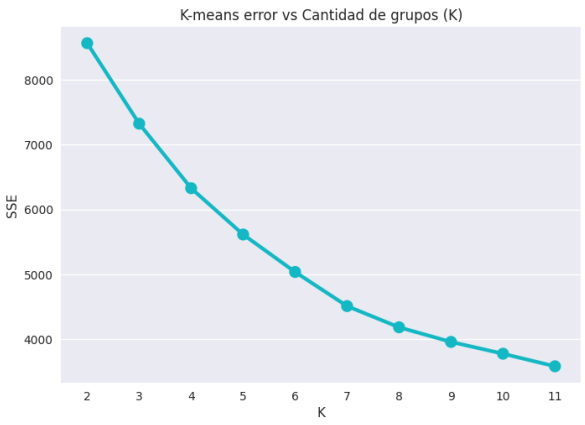


Figura 7: Visualización método del codo. Fuente: Elaboración propia.

Figura 8: Visualización método de la silueta. Fuente: Elaboración propia.

Al evaluar ambos métodos se ha llegado a la conclusión de ocupar 7 clústeres que es donde se piensa que está el codo y es la cantidad con mejor coeficiente de la silueta. La cantidad de usuarios por clúster se puede observar en la Tabla 2.

Tabla 2: Cantidad de usuarios por clúster en K Means. Fuente: Elaboración propia.

	Clústeres							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
# Usuarios	495	58	160	242	42	39	57	**1093**
% Usuarios	45%	5%	15%	22%	4%	4%	5%	**100%**

Finalmente, se utiliza la función *KMeans* de *sklearn* con 7 clústeres, se crea una columna llamada "label" en la cual se le asigna el número de su clúster a cada usuario y se aplica PCA (Principal Component Analysis) para reducir la cantidad de variables y poder visualizar los clústeres. Estos se pueden observar en el F.1.

5.4.2. Fuzzy C Means

Para implementar este modelo en Python, se emplea la biblioteca *fuzz* de *skfuzzy* y se utilizan las mismas variables que en el algoritmo K Means, lo que implica mantener el mismo número de clústeres previamente seleccionados. El modelo se ejecuta mediante la función *fuzz.cluster.cmeans* ocupando 7 clusteres y el parámetro de difusividad (m) igual a 1.5 siendo este el valor que maximiza el coeficiente de la silueta. Además, se le asignan las etiquetas del clúster al cual tienen mayor grado de pertenencia a cada usuario. Esta biblioteca proporciona el coeficiente de partición fuzzy (FPC), que da como resultado 0,73, dando un valor más cercano a 1 que a 0 lo que indica una mejor partición. Para ver la difusividad de los usuarios se decide crear una base auxiliar con 7 columnas, las cuales indicarán el porcentaje de pertenencia de un usuario a cada clúster. A los usuarios que tienen una pertenencia a su clúster asignado menor a 0,7 se le dirá que son usuarios difusos, teniendo que el 25% de los usuarios son difusos (273 usuarios). Además, se ha decidido también crear una nueva columna que indique el segundo clúster de mayor pertenencia. Esto servirá para tener un mejor entendimiento de cómo son los usuarios y aprovechando uno de los beneficios que nos da este algoritmo al no tener una etiqueta fija para cada usuario.

Con la definición anterior se encuentra que 5 de los clústeres tienen un porcentaje de usuarios difusos mayor a 25%. Esto se puede observar en la Tabla 3.

Tabla 3: Porcentaje de usuarios difusos por clústeres. Fuente: Elaboración propia.

	Clústeres							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
# Usuarios	24	58	239	134	255	87	296	**1093**
% Usuarios difusos	29%	7%	29%	29%	14%	39%	28%	**25%**

Los clústeres generados con el algoritmo FCM se pueden apreciar en el F.2. Para esto se aplica PCA para reducir la dimensionalidad de la base de datos y de las coordenadas de los centroides que se pueden ver reflejado con una cruz en el gráfico.

5.4.3. RFM

En este caso no es un algoritmo como tal lo que se aplica, sino que es una técnica de segmentación la cual se adaptará para los datos disponibles en la Plataforma X. Para esto se crean las siguientes variables:

- Recency (R): Cantidad de días desde el último acceso de un usuario a la plataforma.
- Frequency (F): Cantidad de equipos en los que ha participado un usuario.
- Monetary (M): Cantidad de equipos con propuesta en los que ha participado un usuario.

Tras la definición de los parámetros adaptados para este caso, se realizan los segmentos de RFM definiendo un valor entre 1 a 4 para R, F y M. Para F y M mientras mayor sea el valor, significa que tiene mayor F y/o M en comparación a los demás usuarios. Caso contrario para R, ya que mientras menor sea la recencia se le asignará un valor más alto, indicando que su conexión es más reciente.

Para definir los valores que se les asignan a estas variables generalmente se utilizan los cuartiles, en donde cada número entre 1 y 4 indica en el cuartil en el que está. En este caso esto solo se pudo aplicar a la variable R, ya que se tiene mucha más variedad de valores para los distintos días de conexión entre los usuarios. Esto no sucede con F y M, ya que muchos usuarios repiten la cantidad de equipos en los que participan y el número de equipos en los que mandan propuestas, por lo que los límites entre cuartiles son los mismos, impidiendo que se pueda ocupar este método. Por esto se deciden reglas de negocio para definir el valor de F y M, las cuales se pueden encontrar en el Anexo F.3.

Con estas definiciones se obtienen 31 combinaciones de RFM que van desde el 111 hasta el 444, en donde cada número es el valor asignado para cada letra que representa una de las variables. Debido a que se tienen muchas combinaciones, se debe realizar un siguiente paso el cual es agrupar estas combinaciones en segmentos con sentido de negocio. Este agrupamiento se puede observar en el Anexo F.3.

Las reglas principales para definir estos segmentos se basan en la recencia, donde los usuarios se clasifican como conectados o desconectados según su última interacción en la plataforma. Aquellos con valores de recencia 1 y 2 se consideran desconectados, lo que significa que no han accedido a la plataforma en más de 110 días, un período que coincide con el tiempo transcurrido desde la finalización de la convocatoria con mayor

participación hasta la fecha de los datos más recientes. Los usuarios con valores de recencia entre 3 y 4 se consideran conectados, lo que indica que han accedido a la plataforma en los últimos 110 días. Además, para determinar si un usuario es un espectador en la plataforma, se verifica si los valores de ciertas métricas, etiquetadas como "F" y "M", son ambos iguales a 1. Aquellos con al menos un valor igual a 2 se clasifican como de baja participación, mientras que aquellos con valores de 3 se consideran participantes. Finalmente, aquellos que presentan al menos un valor igual a 4 en alguna de las métricas se consideran destacados, lo que sugiere una alta contribución en la conformación de equipos o envío de propuestas.

5.5. Evaluación de segmentos

Tras el modelado con técnicas de clustering, se encontraron 7 clústeres en K Means y Fuzzy C Means y 4 segmentos en RFM divididos en subsegmentos entre conectados y desconectados según su recencia. En esta sección se evalúa la importancia de las variables en cada clúster y se caracteriza cada segmento.

5.5.1. K Means

Para comprender el perfil de usuarios en cada clúster, se analiza su comportamiento en diferentes variables. Se calculan los promedios de las variables para los usuarios de cada clúster y, en algunos casos, se observa el porcentaje de usuarios en cada clúster. Además, se asigna un color verde al valor más alto y rojo a los más bajos en las comparaciones entre clústeres. Esto se observa en el Anexo G.1.

Con esto se describen los segmentos de la siguiente manera:

- **Clúster 0 - Pregrados** (495 usuarios): Usuarios que son principalmente hasta de pregrado, no suelen ser profesores ni exalumnos. Participan poco en convocatorias y en su mayoría cuando participan no suelen ser lideres, la mayoría son investigadores y tienen poca importancia en las publicaciones.
- **Clúster 1 – Solitarios** (58 usuarios): Usuarios que participan en convocatorias principalmente como líderes de su propio equipo en solitario, no suelen enviar muchas propuestas y nunca han ganado una convocatoria, su nivel académico es diverso y tienen una importancia media en las publicaciones.
- **Clúster 2 - Exalumnos** (160 usuarios): Usuarios que son exalumnos investigadores y en su mayoría profesores, suelen participar en convocatorias en equipo y no han ganado convocatorias. En su mayoría tienen postgrado y tienen una importancia media en las publicaciones.
- **Clúster 3 – Académicos Externos** (242 usuarios): Usuarios que no son exalumnos, en su mayoría son investigadores y profesores, suelen participar en convocatorias

en equipo, pero no suelen liderarlos y no han ganado ninguna. Casi en su totalidad tienen postgrado y tienen importancia media-alta en las publicaciones.

- **Clúster 4 – Innovadores y Emprendedores** (42 usuarios): Usuarios que son principalmente emprendedores o innovadores. Participan poco en convocatorias, y cuando lo hacen, recurrentemente es en solitario y tienen una importancia más baja en las publicaciones en promedio que los demás.
- **Clúster 5 – Ganadores** (39 usuarios): Usuarios que han ganado por lo menos una convocatoria, no suelen ser exalumnos y participan en más convocatorias en promedio que la mayoría de los usuarios. Muchos de estos han sido líderes de equipos y algunos han participado en solitario.
- **Clúster 6 – Líderes Multipropuestas** (57 usuarios): Usuarios que han mandado más de una propuesta, pocos de estos han ganado convocatorias y casi en su totalidad han sido líderes de equipos. La mayoría de estos son profesores y pocos son exalumnos. En su gran mayoría tienen postgrado y en promedio tienen una importancia en las publicaciones más altas que los demás usuarios.

5.5.2. Fuzzy C Means

De igual forma que en el modelo de K Means, se caracterizan los clústeres analizando sus comportamientos en cada variable utilizada en el modelo. Esto se observa en el Anexo G.2.

Analizando la importancia de cada variable y los usuarios que son más representativos de cada clúster (mayor grado de pertenencia) en los clústeres se han encontrado los siguientes segmentos:

- **Clúster 0 - Lideres Influyentes** (24 usuarios): Principalmente investigadores que son profesores y algunos exalumnos, casi en su totalidad tienen postgrados. Participan mucho más en convocatorias, suelen ganar bastantes, son generalmente líderes y participan tanto en equipo como en solitario. En general tienen una importancia en las publicaciones alta. Un 29% de los usuarios del segmento son difusos.

- **Clúster 1 – Multipropuestas Poco Ganadores** (58 usuarios): Usuarios que en su mayoría tienen postgrado, casi en su totalidad son investigadores y pocos son exalumnos. Se caracterizan por mandar muchas propuestas en las convocatorias, generalmente en equipo sin ser líderes y ganan pocas convocatorias. Solo el 7% de estos usuarios son difusos.

- **Clúster 2 – Académicos** (239 usuarios): Principalmente investigadores que son profesores y exalumnos, en su mayoría tienen Postgrado. Participan poco en las convocatorias en donde no suelen ser líderes y cuando lo son, es en

equipos en solitario. No tienen una alta importancia en las publicaciones. Un 29% de los usuarios son difusos.

- **Clúster 3 – Participativos y Colaborativos poco Ganadores** (134 usuarios): Usuarios principalmente profesores investigadores y pocos son exalumnos. Se caracterizan por mandar múltiples propuestas y no suelen ser líderes, son los usuarios que más participan en equipo, en su mayoría tienen postgrado y poseen una importancia media en las publicaciones. Un 29% de los usuarios son difusos.

- **Clúster 4 – No Académicos sin Propuestas** (255 usuarios): Principalmente usuarios que no son profesores ni tienen postgrado. No han participado en las convocatorias. Se tiene que un 14% de estos son difusos.

- **Clúster 5 – Líderes Solitarios Participativos** (87 usuarios): Usuarios principalmente profesores investigadores y pocos son exalumnos. Se caracterizan por participar y ser lideres de equipos, muchos de estos usuarios participan en equipos en solitario, en su mayoría tienen postgrado y suelen tener una mayor importancia en las publicaciones que el promedio de los usuarios. Un 39% de los usuarios son difusos.

- **Clúster 6 – Pregrados** (296 usuarios): Principalmente usuarios que no son profesores y que poseen hasta pregrado. En su mayoría son investigadores, cuando participan en convocatorias suelen mandar propuestas y formar equipo con múltiples integrantes y normalmente no ganan. Un 28% de los usuarios son difusos.

También se ha decidido observar en qué otro clúster se identifican los usuarios difusos. Para esto se ve en qué porcentaje se reparten entre otros clústeres. Esto se puede observar en el Anexo G.2.

Al observar la similitud entre los clústeres más difusos (0, 2, 3, 5 y 6), se puede notar que los usuarios difusos del clúster 0 pertenecen entre el clúster 1 y 5. Esto tiene sentido, ya que son los usuarios más destacados en términos de participación comparándolos con los demás. De la misma forma el clúster 1 solo tiene usuarios difusos entre el 0 y el 5. En cambio, los difusos del clúster 5 se distribuyen en todos los clústeres, siendo el 1 en donde más se encuentran. En el clúster 2 están distribuidos casi de la misma forma entre el 2, 4 y 6, principalmente puede ser por la característica de no ganar muchas convocatorias. Los del clúster 6 son principalmente del clúster 3, compartiendo la característica de no tener una alta reputación en publicaciones, participar en equipo y ganar poco.

5.5.3. RFM

Tras identificar los distintos segmentos utilizando RFM, basados principalmente en su conexión a la plataforma y participación en las convocatorias, se describen de la siguiente manera:

- **Destacados Conectados** (91 usuarios): Participantes que entraron a la plataforma hace 2 meses en promedio y que han participado en 6 equipos y mandado 5 propuestas en promedio.
- **Destacados Desconectados** (11 usuarios): Participantes que entraron a la plataforma hace más de 3 meses en promedio y que han participado en 5 equipos y mandado 5 propuestas en promedio.
- **Participantes Conectados** (194 usuarios): Participantes que entraron a la plataforma hace entre 2 y 3 meses en promedio y que han participado con 2 equipos en promedio, mandando 2 propuestas en promedio.
- **Participantes Desconectados** (46 usuarios): Participantes que entraron a la plataforma hace más de 3 meses en promedio y que han participado con 3 equipos en promedio, mandando 2 propuestas en promedio.
- **Baja Participación Conectados** (256 usuarios): Usuarios que entraron hace menos de 3 meses en promedio y que han participado con 1 equipo y mandado 1 propuesta en promedio.
- **Baja Participación Desconectados** (204 usuarios): Usuarios que entraron en la plataforma hace más de 3 meses en promedio y que han participado con 1 equipo y mandado 1 propuesta en promedio.
- **Espectadores Conectados** (162 usuarios): Participantes que han entrado a la plataforma hace 2 meses en promedio y que nunca han participado en las convocatorias.
- **Espectadores Desconectados** (129 usuarios): Participantes que han entrado a la plataforma hace más de 3 meses promedio y que nunca han participado en las convocatorias.

Para ver el comportamiento de estos segmentos en sus tres dimensiones, se estandarizan las variables de RFM para poder estar en la misma escala. Esto se puede ver de mejor forma en el Anexo G.3.

5.5.4. Utilización de modelos

Tras experimentar con distintos modelos y caracterizar segmentos de la Plataforma X, se decide enfocar el uso del RFM en estrategias de activar usuarios desconectados, convertir usuarios que no participan en participantes y de retener a los usuarios que más participan. Por otra parte, se plantea ocupar K Means o Fuzzy C Means para establecer estrategias personalizadas según distintas características de los clústeres enfocadas en la participación y colaboración. Para decidir qué modelo utilizar se evaluaron beneficios y costos de cada uno que se pueden observar en la Tabla 4.

Tabla 4: Comparación de costos y beneficios de K Means vs Fuzzy C Means. Fuente: Elaboración propia.

	K Means	**Fuzzy C Means**
Costos	Sensibilidad a los outliers	Complejidad Computacional Interpretabilidad más compleja
Beneficios	Simplicidad y rapidez Escalabilidad Interpretabilidad	Flexibilidad Asignación suave

Tras analizar junto a la contraparte se llegó a la decisión de elegir K Means debido a la simplicidad de ejecución del modelo y la escalabilidad que este puede tener al trabajar con conjuntos grandes de datos, pensando en su posible replicación para las distintas plataformas que se tienen dentro de la institución.

5.6. Estrategias personalizadas

Tras la caracterización de los clústeres encontrados, se seleccionan los segmentos más relevantes para la plataforma con la finalidad de proponer estrategias personalizadas focalizadas en estos segmentos con el fin de aumentar la participación de estos en la plataforma.

5.6.1. Selección de segmentos

5.6.1.1. K Means

Los segmentos más relevantes y la explicación del por qué son los siguientes:

- **Ganadores**: Es el segmento en donde se encuentran los usuarios más ganadores de la plataforma, es importante mantener el conocimiento y experiencia que tienen para generar propuestas ganadoras dentro de la plataforma.

- **Líderes Multipropuestas**: Es el segmento en donde se encuentran los usuarios más líderes y participativos de la plataforma, por lo que es prioritario mantener a estos usuarios activos y colaborativos.

- **Pregrados**: Es el segmento más grande dentro de la plataforma. Es el público al que apunta la institución en general y es el tipo de usuarios que más pueden conseguir en sus canales de comunicación internos.

- **Académicos Externos**: Es el segundo segmento más grande de la plataforma y donde se encuentran usuarios con gran potencial para crear propuestas de calidad debido a su alto nivel académico e importancia de sus publicaciones.

5.6.1.2. RFM

Los Segmentos más relevantes escogidos y su porqué son los siguientes:

- **Destacados Conectados:** Es un segmento destacado en participación y que está conectado en la plataforma en comparación a los demás, por lo que es valioso retener a este tipo de usuario.

- **Destacados Desconectados:** Es un segmento de pocos usuarios pero que es importante de reactivar debido a sus destacadas participaciones.

- **Participantes Conectados:** Es un segmento que suele participar y estar conectado de forma más reciente en la plataforma. Es importante aprovechar este interés para incentivar la participación continua.

- **Participantes Desconectados:** A pesar de ser un segmento pequeño, es importante para la plataforma recuperar a estos tipos de usuarios que en algún momento fueron participantes recurrentes en la plataforma.

- **Baja Participación Conectados:** Es el segmento con mayor cantidad de usuarios y este subsegmento de conectados indica interés dentro de la plataforma, por lo que puede ser una gran cantidad de usuarios que se podrían transformar en participantes o destacados.

- **Espectadores Conectados:** Es un segmento que no ha participado en convocatorias aún pero que se ha conectado en la plataforma últimamente, por lo que es importante aprovechar el interés mostrado para que empiecen a participar.

5.6.2. Estrategias segmentos K Means

Para el diseño de las estrategias para estos segmentos, se focalizó en dos objetivos en particular, aumentar la participación y aumentar la colaboración entre los usuarios. Se diseñaron 4 estrategias para cada objetivo las cuales se encuentran en la Tabla 5 y Tabla 6.

Tabla 5: Estrategias a segmentos K Means con foco en la participación. Fuente: Elaboración propia.

Objetivo	Descripción	Segmentos que aplican
Incentivar la participación	Diseñar **sistema de beneficios de participación** otorgando beneficios al mandar propuestas cada 5 convocatorias distintas	Líderes Multipropuestas Pregrados Académicos Externos
	Crear un **sistema de beneficios de retención** ofreciendo un premio extra si vuelve a ganar una convocatoria	Ganadores
	Convalidación de propuestas de convocatorias con proyectos de títulos en convocatorias que cumplan los requisitos	Pregrados
	Sistema de referidos dentro de la plataforma, otorgando un monto o descuento a los usuarios que inviten a uno nuevo y que este mande una propuesta	Ganadores Líderes Multipropuestas Pregrados Académicos Externos

Tabla 6: Estrategias a segmentos K Means con foco en la colaboración. Fuente: Elaboración propia.

Objetivo	Descripción	Segmentos que aplican
Aumentar la colaboración	Establecer **programa de mentores** ganadores, donde estos puedan guiar a participantes que no han ganado o que son nuevos, compartiendo experiencia y estrategias	Ganadores
	Crear un **programa de Líderes** en donde se Invite a participar como líderes de nuevos equipos en nuevas convocatorias	Ganadores Líderes Multipropuestas
	Elaborar un **programa de pupilos** en donde se les invite a conocer estrategias de los usuarios mentores como de otros investigadores reconocidos dentro de la institución	Pregrados
	Establecer **talleres de liderazgo** para incentivar a los usuarios participar como líderes en nuevas convocatorias	Académicos Externos

5.6.3. Estrategias segmentos RFM

Para las estrategias de RFM se basarán principalmente en informar sobre convocatorias, la activación, conversión y retención de usuarios dependiendo del comportamiento que han tenido en la plataforma. Estas se pueden observar en la Tabla 7.

Tabla 7: Estrategias a segmentos RFM. Fuente: Elaboración propia.

Objetivo	Descripción	Segmentos que aplican
Informar sobre nuevas convocatorias	**Programa de acumulación de puntos** canjeables por leer lanzamientos de nuevas convocatorias y responder si le parece interesante o no	Destacados Conectados Participantes Conectados Baja Participación Conectados Espectador Conectados
Reactivar usuarios importantes para la plataforma	**Mailing o correos dirigidos** a todos los usuarios indicando nuevas convocatorias, fechas de participación e incentivos que ofrecen	Destacados Desconectados Participantes Desconectados
Convertir en participantes	**Programa de incentivo a participar por primera vez**, como descuentos a cursos o beneficios económicos	Espectador Conectado
Retención de usuarios importantes	**Programa de embajadores** en donde se invite a los usuarios más participativos a ir a eventos exclusivos dándoles financiamiento adicional o puntos canjeables por cursos de la institución	Destacados Conectados Participantes Conectados
Convertir en participantes frecuentes	**Programa de Re-participación** que incentive con beneficios canjeables a medida que los usuarios manden más propuestas en convocatorias distintas	Participantes Conectados Baja Participación Conectados
Incentivar la participación continua	**Talleres presenciales**, en donde se invite a los usuarios a compartir lluvias de ideas de propuestas y formar contactos con otros usuarios	Destacados Conectados Destacados Desconectados Participantes Conectados Participantes Desconectados Baja Participación Conectados Espectador Conectados

5.6.4. Impacto estimado

En esta sección se propone las métricas a utilizar para medir el impacto de las estrategias de participación y colaboración del K Means y las de retención, reactivación y conversión del RFM. Además, se plantea un impacto estimado en base a supuestos y datos obtenidos de la comprensión de los datos.

Para evaluar el impacto de las estrategias de participación diseñadas a los segmentos encontrados utilizando K Means, se tendrá como supuesto un caso optimista en que el 15% de los usuarios de los grupos objetivo respondan positivamente a las estrategias personalizadas, esto considerando que actualmente el 15% de los usuarios ha vuelto a participar en una convocatoria. El resultado esperado será de 10% y el pesimista el 5%.

Se propone utilizar como métricas la **cantidad de equipos conformados** y la **cantidad de propuestas enviadas**. Actualmente se tiene un promedio de 49 equipos y 33 propuestas por convocatoria.

$$\#\ Equipos\ estimados = \%_{captación} \cdot grupo\ objetivo\ \cdot 73\% \qquad Ec.\ 5$$

$$\#\ Propuestas\ estimadas = \#\ Equipos\ estimados\ \cdot 67\% \qquad Ec.\ 6$$

En la Ec. 5 se muestra el cálculo para estimar la cantidad de usuarios que participan en las convocatorias aplicando las estrategias, utilizando el supuesto de que el 73% de los usuarios forman un equipo. En la Ec. 6 se muestra el cálculo de la estimación de las propuestas enviadas, tomando el supuesto de que el 67% de los equipos mandan una propuesta en la actualidad.

El impacto estimado de las estrategias se puede ver en la Tabla 8.

Tabla 8: Estimación de impacto de las estrategias de participación utilizando K Means. Fuente: Elaboración propia.

Caso	# Equipos promedio estimado	# Propuestas promedio estimado	Impacto estimado
Pesimista	30	20	0%
Esperado	61	41	+25%
Optimista	91	61	+86%

Por otra parte, para las estrategias de colaboración se propone utilizar las métricas de **porcentaje de equipos con múltiples integrantes** y la **cantidad de propuestas enviadas**. Actualmente se tiene que un 56% de los equipos son de múltiples integrantes. Además, se tiene que el 93% de los equipos con múltiples integrantes envían propuestas y que el 34% de los equipos de un solo integrante envían propuestas. En estas estrategias se tendrá un caso pesimista de subir el porcentaje de equipos de múltiples integrantes a 60%, en el esperado a un 65% y el optimista a un 70%. Para calcular la cantidad de propuestas promedio estimadas se utiliza la Ec. 7.

$$\begin{aligned} &\#\ Propuestas\ estimadas \\ &= 93\% \cdot (\#\ Equipos\ promedio \cdot \%_{Eqiupos\ Integrantes\ Multiples}) \\ &+34\% \cdot (\#Equipos\ promedio \cdot \%_{Equipos\ unico\ integrante}) \end{aligned} \qquad Ec.\ 7$$

El impacto estimado para cada caso se puede observar en la Tabla 9.

Tabla 9: Estimación de impacto de las estrategias de colaboración utilizando K Means. Fuente: Elaboración propia.

Caso	# Equipos promedio	# Propuestas promedio estimado	Impacto estimado
Pesimista	49	34	+3%
Esperado	49	35	+6%
Optimista	49	37	+12%

Finalmente, para las estrategias de RFM las métricas propuestas son la **tasa de retención** para los segmentos más participativos, **tasa de reactivación** a los segmentos desconectados, **tasas de conversión** de usuarios de baja participación a participantes y de participantes a destacados. Para este caso se vuelve a tomar como supuesto que el 15% de los usuarios participan en más de una convocatoria, por lo que el caso pesimista es una captación del 5% de los grupos objetivo, en el esperado un 10% y el optimista un 15%.

Utilizando las Ec. 5 y Ec. 6 nuevamente se observa el impacto en la Tabla 10.

Tabla 10: Estimación de impacto de las estrategias de participación utilizando RFM. Fuente: Elaboración propia.

Caso	# Equipos promedio estimado	# Propuestas promedio estimado	Impacto estimado
Pesimista	29	19	0%
Esperado	58	39	+18%
Optimista	87	58	+78%

Capítulo 6: Discusiones

En esta sección se discutirán las principales decisiones que pudieron ser distintas dentro del proyecto, analizando posibles alternativas a las decisiones tomadas y por qué pueden ser buenas soluciones. Una de estas decisiones es no realizar un análisis de texto a profundidad debido a que se contaba con variables interesantes para caracterizar a los usuarios que estaban en formato texto y la otra decisión es elegir K Means sobre Fuzzy C Means para la realización de estrategias. La primera decisión mencionada fue tomada por el equipo de trabajo (estudiante y Data Scientist de Brain Food) y la segunda decisión fue conversada en conjunto con la contraparte.

Durante el análisis de datos, se identificaron variables como la biografía, experiencia académica, habilidades y palabras clave de los usuarios. Estas variables son ingresadas voluntariamente por los usuarios en la plataforma, por lo que no es obligatorio completarlas. Debido a esto, entre el 50% y el 60% de los datos están incompletos, es decir, los campos no han sido llenados en los perfiles de los usuarios. Además, al ser información proporcionada por los propios usuarios, no hay una estandarización en el idioma, y existen errores ortográficos, así como inconsistencias en las habilidades o palabras clave declaradas. Estos factores dificultan significativamente el análisis y uso de estas variables en los modelos empleados. Por esta razón, se decidió trabajar únicamente con la variable de experiencia académica. Se identificaron palabras relacionadas con los niveles de estudiantes, licenciados, magíster y doctorados para crear una variable de grado académico, completando los valores nulos con base en los supuestos mencionados anteriormente. En cambio, las otras variables fueron descartadas para la caracterización y segmentación de los usuarios.

La decisión de descartar estas variables de texto es tomada principalmente por la gran cantidad de datos nulos, considerando que son pocos usuarios activos en la plataforma, y la nula estandarización que poseen, dificultando el análisis e interpretación de las variables.

Si se comenzara el proyecto nuevamente, como alternativa a la decisión tomada durante el proyecto, Al notar el problema descrito en estas variables de texto, se solicita anticipadamente a la contraparte que pida completar las variables de texto con respuestas predefinidas por la institución en las habilidades, palabras clave y nivel académico. Además, pedir incluir obligatoriamente la biografía en sus perfiles. Con esto se puede empezar a recolectar más información en las variables de texto y tenerlas estandarizadas. Complementando la información con la que se empieza el proyecto más las nuevas respuestas

de los usuarios en sus perfiles, se pueden aplicar técnicas de LLM (*Large Lenguage Models*[6]) las cuales pudieran permitir analizar estas variables de texto con el fin de extraer información relevante de los usuarios y poder agrupar ciertas habilidades o palabras clave de los usuarios para crear nuevas variables que pueden ser útiles tanto para la segmentación y caracterización de los usuarios. También puede ser útil para otras aplicaciones como como un algoritmo de match entre usuarios y convocatorias de su interés. Esta alternativa en la práctica puede ser complicado conseguir que todos los usuarios completen total o conscientemente estos campos, pero con el tiempo y con mayor cantidad de usuarios se puede conseguir una buena base para analizar este tipo de variables y darle un uso que beneficie a la plataforma.

Por otra parte, se decide utilizar los segmentos de K Means para la generación de estrategias debido a su facilidad de implementación y replicación en otras plataformas de la institución, en comparación con Fuzzy C Means (FCM).

Si se reiniciara el proyecto, una alternativa sería no descartar por completo el uso de los segmentos de FCM. Se pueden seleccionar segmentos similares a los de K Means y complementar las estrategias para los usuarios difusos que pertenezcan a estos segmentos, involucrándolos en las estrategias de su segundo segmento con mayor grado de pertenencia. Otra opción es proponer el uso de segmentos de FCM en ciertas convocatorias específicas, lo que permite a la institución comparar el funcionamiento de ambos modelos al implementar estrategias personalizadas y decidir si vale la pena utilizar un modelo más complejo.

De esta manera, se pueden aprovechar los beneficios del FCM, trayendo consecuencias positivas como el diseño de estrategias aún más personalizadas y haciendo partícipes a ciertos usuarios en distintas campañas lanzadas por la institución.

[6] Los LLM consisten en tecnología de IA avanzada que se centra en comprender y analizar texto. Son más precisos que los algoritmos de aprendizaje automático tradicionales, porque pueden comprender las complejidades del lenguaje natural.

Capítulo 7: Conclusiones

Al finalizar el proyecto, se concluye que la segmentación de usuarios obtenida proporciona un conocimiento más detallado de los usuarios de la Plataforma X. Las estrategias propuestas tienen el potencial de ser un motor clave para incentivar la participación y colaboración en las convocatorias lanzadas. Se estima que al aplicar las estrategias enfocadas en la participación de los segmentos identificados mediante K-Means, se podría lograr un aumento del 25% en la participación de equipos y propuestas mandadas promedio por convocatoria. Asimismo, al utilizar las estrategias de colaboración, se proyecta un incremento del 6% en la cantidad promedio de propuestas por convocatoria. Además, al aplicar las estrategias diseñadas para los segmentos identificados con la técnica RFM, se espera un aumento del 18% en la participación de equipos y en el número de propuestas enviadas por convocatoria. Este impacto podría traer grandes beneficios a la plataforma, haciendo que sea más activa y participativa, lo que a su vez permitiría negociar de manera más efectiva con las industrias o empresas que lancen convocatorias. Además, aumentaría la probabilidad de generar propuestas ganadoras que puedan convertirse en productos o licencias con un impacto significativo en la sociedad, generando ingresos futuros para la institución.

En cuanto a los objetivos, se cumplió el objetivo principal del proyecto: generar estrategias de negocio personalizadas para segmentos utilizando técnicas de clustering. Este logro se alcanzó cumpliendo cada objetivo específico propuesto, comenzando con un análisis de las fuentes internas y externas disponibles. Se entregaron a la institución una presentación y un notebook de Python con los principales insights obtenidos. Posteriormente, se diseñó una base de datos con un registro único por usuario, que incluye las variables internas y externas más relevantes para el negocio. Esta base de datos fue fundamental para aplicar las técnicas de clustering K Means, FCM y RFM. Todo el desarrollo de los modelos se documentó en un notebook de Python entregado a la contraparte. Utilizando las variables más importantes de cada clúster, se caracterizó cada segmento, seleccionando los más relevantes para la generación de estrategias. La caracterización de los segmentos se le facilitó a la contraparte. Además, se propusieron estrategias específicas con objetivos diferenciados según la técnica de clustering utilizada.

7.1. Limitaciones y Alcances

Uno de los principales desafíos es la escasez de variables descriptivas de los usuarios, muchas de las cuales no estaban estandarizadas y contenían valores nulos. Esto limitó la cantidad y calidad de las segmentaciones. Otras de las limitaciones importantes durante el proyecto es la cantidad de convocatorias realizadas hasta la fecha, siendo esta

solo 10 y en donde solo un 40% de estas habían finalizado, lo que puede conllevar sesgos en el análisis y los segmentos identificados.

Con relación a los alcances, el proyecto se limitó a la propuesta de estrategias personalizadas, dejando fuera la implementación y evaluación de estas. Por tanto, la institución es responsable de seleccionar las estrategias más adecuadas y evaluar la participación de los usuarios a lo largo del tiempo.

7.2. Recomendaciones

Para continuar con este proyecto, se recomienda a la empresa repetir el análisis con datos más actualizados, considerando convocatorias finalizadas y nuevas abiertas. Esto permitiría analizar el movimiento entre segmentos a lo largo del tiempo. Además, se sugiere estandarizar ciertos campos que completan los usuarios, ofreciéndoles una gama de habilidades o palabras clave con las cuales se identifiquen, y añadir más variables demográficas como edad, ocupación y situación laboral para obtener segmentos más diferenciados.

En cuanto a las estrategias, se recomienda estar constantemente evaluándolas y diseñando nuevas estrategias personalizadas que puedan tener un impacto positivo en la participación de los usuarios.

Por otra parte, se recomienda realizar el análisis de texto mencionado en las discusiones, ya que puede aportar un gran valor en la caracterización de segmentos y en la dirección de estrategias específicas.

7.3. Trabajo Futuro

La segmentación de usuarios puede ser el inicio de una serie de proyectos destinados a mejorar la experiencia de los usuarios de la Plataforma X. Entre los proyectos futuros, se puede considerar establecer redes de usuarios para visualizar las interacciones entre ellos, ya sea conformando equipos o colaborando en publicaciones externas. También se podrían desarrollar proyectos enfocados en hacer un match entre usuarios y convocatorias basándose en los conocimientos y habilidades requeridas, invitando de forma selectiva a los usuarios con mayor probabilidad de participar.

Finalmente, se sugiere extender el uso de fuentes externas, como LinkedIn, para obtener información sobre la carrera profesional de los usuarios y buscar nuevos usuarios con perfiles similares a los existentes en la plataforma, aumentando así la cantidad de usuarios y obteniendo propuestas innovadoras que generen un impacto positivo en la sociedad.

Bibliografía

Brain Food. (2024). *Linkedin*. Obtenido de https://www.linkedin.com/company/brain-food-spa/about/

Brain Food. (s.f.). *Areas de servicios: Brain Food*. Obtenido de https://brainfood.cl/areas-de-servicios/

Conasa. (s.f.). *Conasa*. Obtenido de https://conasa.grupocibernos.com/blog/que-es-y-para-que-sirve-una-consultoria-it#:~:text=La%20consultor%C3%ADa%20tecnol%C3%B3gica%20es%20un,de%20alcanzar%20sus%20objetivos%20empresariales.

Crossref. (s.f.). *About us - Crossref*. Obtenido de https://www.crossref.org/about/

Doctrina Qualitas. (2023). *Estudiar Energias Renovables*. Obtenido de https://estudiarenergiasrenovablesonline.es/impacto-de-la-inteligencia-artificial-en-la-investigacion-cientifica/#ftoc-heading-5

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques.* Elsevier Inc.

IBM. (s.f.). *Guía de CRISP-DM de IBM SPSS Modeler.*

Khajvand, M., & Jafar Tarokh, M. (2010). Estimating customer future value of different customer segments. En *Procedia Computer Science* (Vol. 3, págs. 1327-1332).

Mordor Intelligence. (s.f.). *Tamaño del mercado de servicios de consultoría y análisis de participación tendencias de crecimiento y pronósticos (2024-2029): Mordor Intelligence*. Obtenido de https://www.mordorintelligence.com/es/industry-reports/consulting-service-market

Ogunleye, J. O. (2021). The Concept of Data Mining.

ORCID. (s.f.). *Nuestra Empresa - ORCID*. Obtenido de https://info.orcid.org/es/what-is-orcid/

RICYT. (2023). *El Estado de la Ciencia.*

Rojas, J., Chavarro, J., & Moreno, R. (2008). Tecnicas de logica difusa aplicadas a la mineria de datos. En U. T. Pereira, *Scientia et Technica* (Vol. 3, págs. 1-6).

SII. (s.f.). *Estadísticas de Empresa: SII*. Obtenido de https://www.sii.cl/sobre_el_sii/estadisticas_de_empresas.html

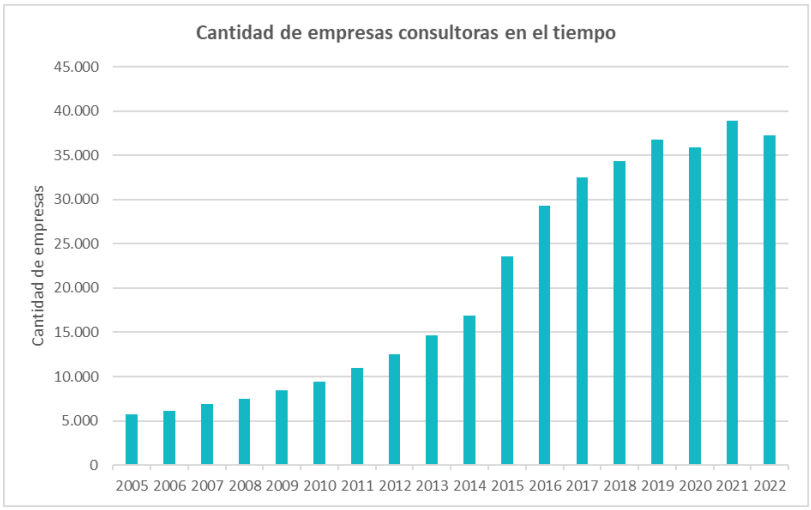
Sordo, A. I. (20 de Enero de 2023). *Hubspot.* Obtenido de https://blog.hubspot.es/marketing/marcas-ejemplos-marketing-personalizado#que-es

UNESCO. (2020). *Investigación y vínculo con la sociedad en universidades de América Latina.*

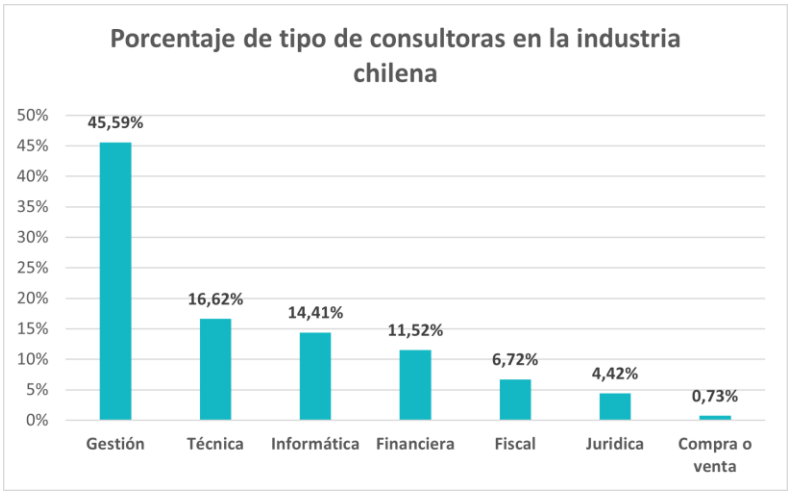
Anexos

Anexo A: Antecedentes Generales

A.1. La consultoría en Chile

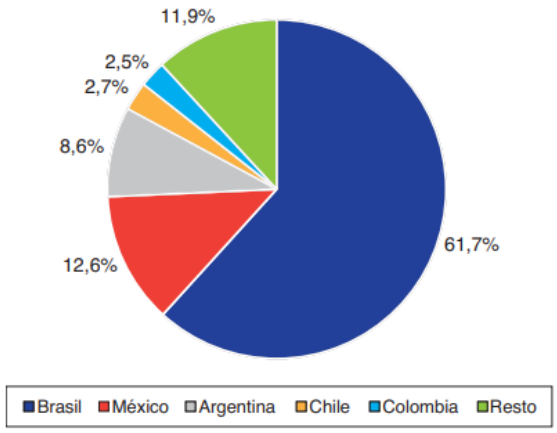


Cantidad de empresas consultoras en el tiempo. - Elaboración propia. Fuente: SII

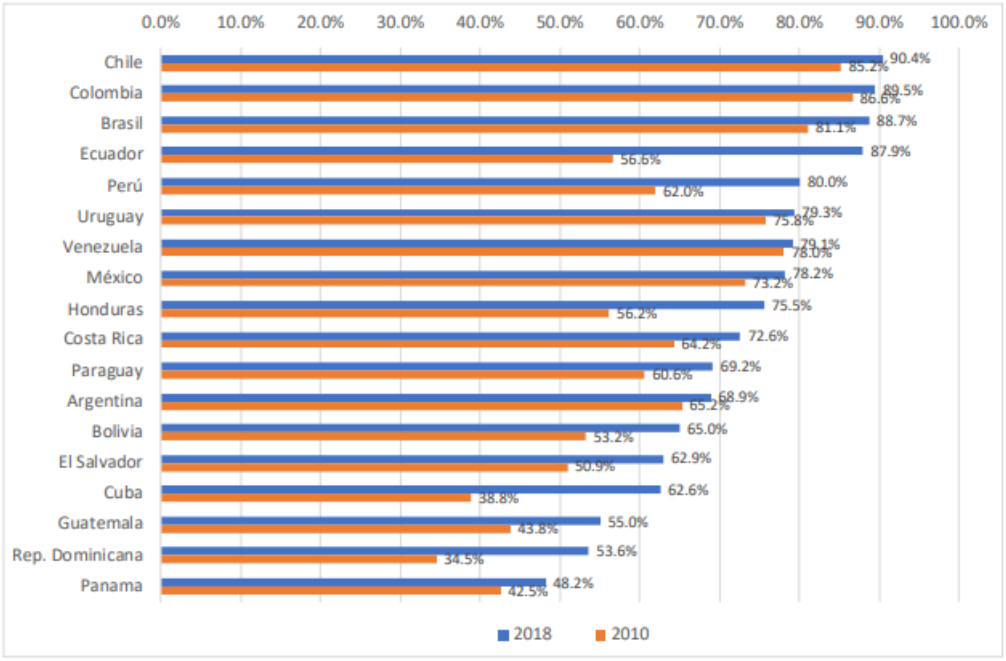


Porcentaje de tipo de consultoras en la industria chilena. - Elaboración propia. Fuente: SII.

A.2. Las universidades latinoamericanas y la investigación



Distribución de la inversión en I+D en ALKC en 2020 (dólares PPC). Fuente: RICYT 2023.



Participación de universidades en la producción científica. Fuente: Red INDICES 2018.

A.3. Brain Food

Tamaño de las empresas en Chile. Elaboración propia. Fuente: Ley N° 20.416

Tamaño Empresa	Clasificación por ventas	Clasificación por empleo
Micro	0 -2.400 UF	0 – 9
Pequeña	2.400,01 – 25.000 UF	10 – 25
Mediana	25.000,01 – 100.000 UF	25 – 200
Grande	100.000,01 UF y más	200 y más

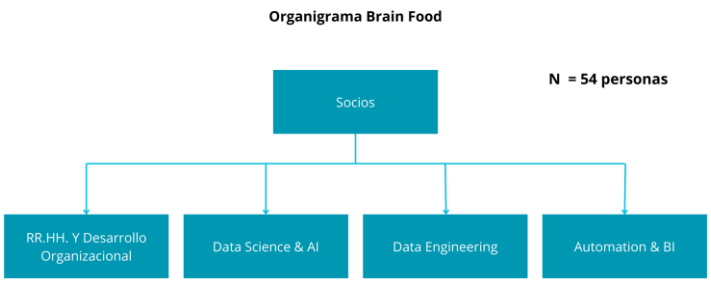
Valores

Brain Food tiene 6 valores principales que se dividen en tres áreas: Personas, Trabajo y Equipo. En el ámbito de Personas, los valores son Integridad y Empatía. En el área de Trabajo, los valores son Rigurosidad y Proactividad. En el ámbito de Equipo, los valores son Comunicación y Compromiso.

Estos valores se promueven a través de diversas instancias como evaluaciones de desempeño, feedback semanal y actividades que fortalecen la cultura organizacional.

Estructura Organizacional

Brain Food actualmente posee 54 trabajadores que se distribuyen entre Socios, RR.HH. y Desarrollo Organizacional, Data Science & AI, Data Engineering y Automation & BI. La relación se observa en el siguiente organigrama.



Organigrama Brain Food. Fuente: Elaboración propia.

Los cargos dentro de Brain Food se pueden apreciar en forma de pirámide junto a la descripción de cada cargo en el siguiente esquema.

Pirámide de cargos Brain Food. Fuente: Elaboración propia.

• Data Analysts: Aportar en el desarrollo de entregables, entregar conocimiento cuando sea necesario y participar de iniciativas internas.

• Consultores: Liderar el desarrollo de entregables específicos, entregar conocimiento cuando sea necesario y participar activamente de iniciativas internas.

• Consultores Senior: Liderar frentes de trabajo específicos, guiar el trabajo de miembros más nuevos del equipo, levantar posibles oportunidades en clientes y participar activamente de iniciativas internas y desarrollar cultura Brain Food.

• Jefes de proyectos/Tech Lead: Liderar proyecto, incluyendo relación con cliente, liderar el desarrollo de los miembros del equipo de trabajo, buscar activamente oportunidades en clientes y participar activamente de iniciativas internas y desarrollar cultura Brain Food.

• Gerentes de proyectos / Specialists por área: Supervisar de forma cercana varios proyectos y llevar relación con clientes, supervisar el desarrollo y el ambiente del equipo de trabajo, buscar oportunidades comerciales de forma activa, liderar iniciativas internas y desarrollar cultura Brain Food.

• Heads áreas de práctica: Asegurar el estándar de entregables relevantes y supervisar proyectos específicos, supervisar el desarrollo del equipo del área de práctica en línea con el plan desarrollado, levantar posibles oportunidades en clientes y liderar el desarrollo de nuevo conocimiento dentro de la empresa.

• Directores de proyectos: Supervisar a alto nivel varios proyectos, supervisar el desarrollo y el ambiente del equipo de trabajo, buscar y desarrollar oportunidades en clientes nuevos y viejos y liderar iniciativas internas y desarrollar cultura Brain Food.

Servicios

En cuanto a los servicios, Brain Food divide sus servicios en cuatro áreas, las cuales interactúan durante los proyectos para desarrollar soluciones integrales. Estas son: (Brain Food, s.f.)

- Estrategia Digital: "Transformamos tu visión en resultados reales".
- Inteligencia Artificial: "Modelamos para entender, predecir e impactar los resultados de tu organización".
- Automatización y BI: "Transformamos los datos en insights, liberando tiempo automatizando procesos manuales y repetitivos".
- Ingeniería de Datos: "Extraemos, preparamos y damos sentido a tus datos sin importar el volumen o formato".

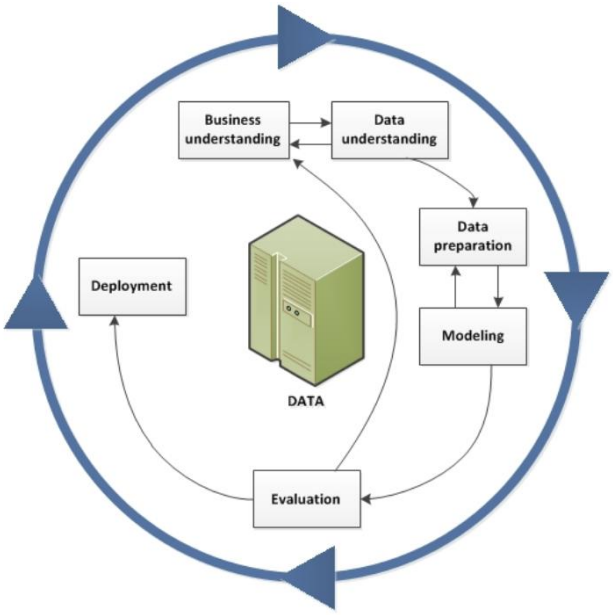
Cartera de clientes

La cartera de clientes de la consultora, en la actualidad está compuesta por empresas de distintos tamaños en diversas industrias y países. Alguna de las industrias en las que se han realizado proyectos son consumo masivo, telecomunicaciones, servicios financieros, retail, educación, salud, automotriz, etc. Esta presencia global con sus colaboradores se puede observar de mejor forma en la siguiente imagen.

Presencia de Brain Food con clientes a nivel global. Fuente: Brain Food.

Debido a los buenos resultados obtenidos en los proyectos realizados, la empresa actualmente tiene una tasa de retención de clientes del 85%, esta tasa representa el porcentaje de los clientes que deciden contratarlos para un nuevo proyecto tras finalizar un primer proyecto.

Anexo B: Metodología



Ciclo de vida de un proyecto de minería de datos con CRISP-DM. Fuente: IBM SPSS Modeler.

Anexo C: Entendimiento del negocio

C.1. Ejemplo de convocatoria

- Objetivos:
    - Incentivar el desarrollo de proyectos de base científica tecnológica, orientados a resolver problemas en el área de cambio climático.
    - Establecer las bases para un potencial emprendimiento con base científica-tecnológica.
    - Culturizar a líderes de proyectos acerca de emprendimiento basado en evidencia, escalamiento tecnológico y descubrimiento del cliente.
- Reto por resolver
    - Esta convocatoria apoyará proyectos de base científica-tecnológica (PBCT) orientadas a resolver problemas en el área del cambio climático.
    - Los PBCT deberán tener un nivel de desarrollo máximo de prototipo funcional en laboratorio.
- Quiénes pueden participar
    - Solo la comunidad científica-innovadora de la institución educativa.
- Premio Total
    - $2.000 USD a repartir entre los mejores PBCT.
- Si el proyecto es seleccionado el equipo debe comprometerse a cumplir:
    - Utilizar los fondos para escalar en el desarrollo del PBCT en un plazo de 6 meses.
    - Participar en reuniones periódicas para mostrar avances.
    - Generar pruebas y prototipos con resultados.
    - Al llegar al MVP el equipo se compromete a buscar fondeo externo para un escalamiento.
- Recepción de propuestas

- Fecha de inicio: 01 de enero de 2024
- Fecha de cierre: 27 de febrero de 2024

Anexo D: Comprensión de los datos

D.1. Variables de base de usuarios

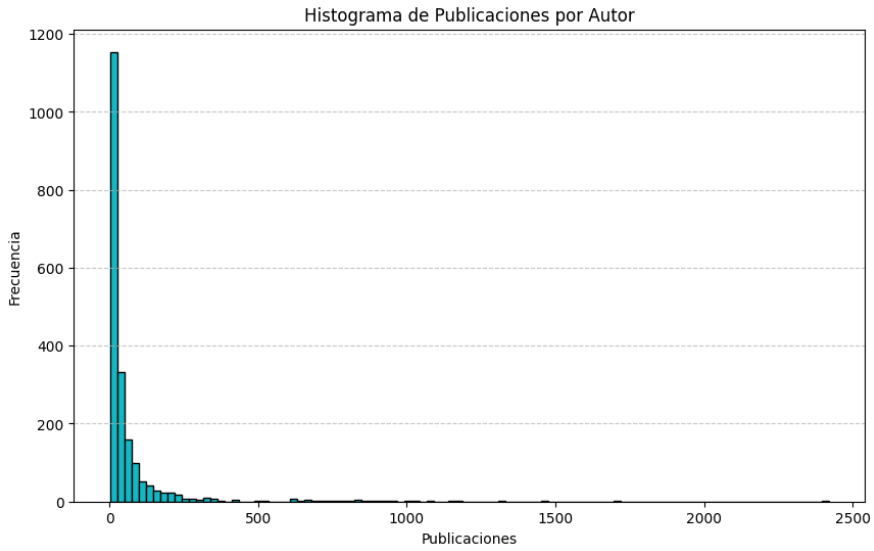
Variable	Tipo	Descripción
Id_user	String	Id de los usuarios
NombreMostrar	String	Nombre que se muestra en la plataforma
Nombre	String	Nombre del usuario
Apellido	String	Apellido del usuario
Correo	String	Correo del usuario
Rol	String	Rol del usuario (Solucionador o Administrador)
VersionAviso	Int	Versión de aviso firmada
Registro	Datetime	Fecha de registro en la plataforma
TipoSolucionador	String	Tipo de solucionador (Investigador, Innovador o Emprendedor)
País	String	País del usuario
Estado	String	Estado en el que vive
Ciudad	String	Ciudad en la que vive
IsProfesor	Boolean	Si es profesor o no
ExperienciaAcademica	String	Experiencia académica escrita por el usuario
ExAlumno	Boolean	Si es exalumno o no.
Biografía	String	Biografía escrita por el usuario
Skills	String	Skills declaradas
Keywords	String	Keywords declaradas

Created	Datetime	Fecha de creación del usuario
Modified	Datetime	Fecha de modificación del perfil
VersionTerminos	Int	Version de términos firmada
ScopusId	String	Id de scopus
FechaUltimoAcceso	Datetime	Fecha de último acceso
UbicacionId	Int	Id de su ubicación
Activo	Boolean	Si es un usuario activo o no
Campus	String	Campus de la institución en la que está
Universidad	String	Universidad del usuario
UniversidadSiglas	String	Siglas de la universidad
EscuelaInstituto	String	Escuela o instituto en el que estudió
GrupoInvestigacion	String	Grupo de investigación al que pertenece
AreaInvestigacion	String	Área de investigación a la que se dedica
GradoAcademico	String	Grado académico del usuario
EquipoId	String	Id de los equipos en los que ha participado
NumEquipos	Int	Numero de equipos en los que ha participado

D.2. Fuentes Externas

SCOPUS

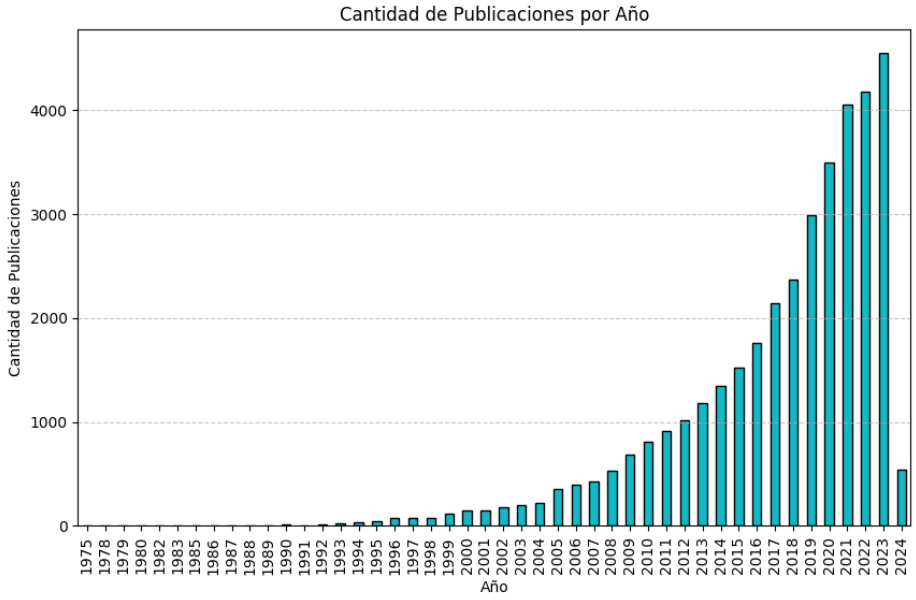
En la tabla *author* la variable más importante que se pudo extraer es "document_count" que indica la cantidad de publicaciones que tienen los autores. Esta muestra que todos los autores que tienen este campo tienen por lo menos una publicación y hay casos en que un autor tiene casi 2400 publicaciones (pueden ser propias o una colaboración), pero en promedio se tienen usuarios con 62 publicaciones. Esto se puede ver en el siguiente gráfico.



Histograma de publicaciones por autor. Fuente: Elaboración propia.

Además, se tiene otra variable interesante llamada "TemasPublicaciones" donde están los principales temas tomados por los autores y de esta se crea una variable llamada "CantidadTemas" para ver los usuarios que tocan temas más diversos que otros, teniendo que en promedio un autor domina 4 temas distintos, teniendo un mínimo de 2 y máximo de 25.

En la tabla *publications*, durante el análisis exploratorio de datos (EDA), se identificaron 33,655 publicaciones distintas, de las cuales aproximadamente 7,000 tienen más de un autor, siendo la mayoría coautorías entre 2 y 3 personas. Las fechas de publicación abarcan desde 1975 hasta 2024, mostrando un aumento considerable en la cantidad de publicaciones a lo largo de los años, como se observa en el siguiente gráfico (exceptuando 2024, que aún está en curso).



Cantidad de publicaciones de los autores por año. Fuente: Elaboración propia.

Se analizaron los tipos de publicaciones que hacen los autores en donde los que más destacan son los *journal*[7] y *conference proceeding*[8].

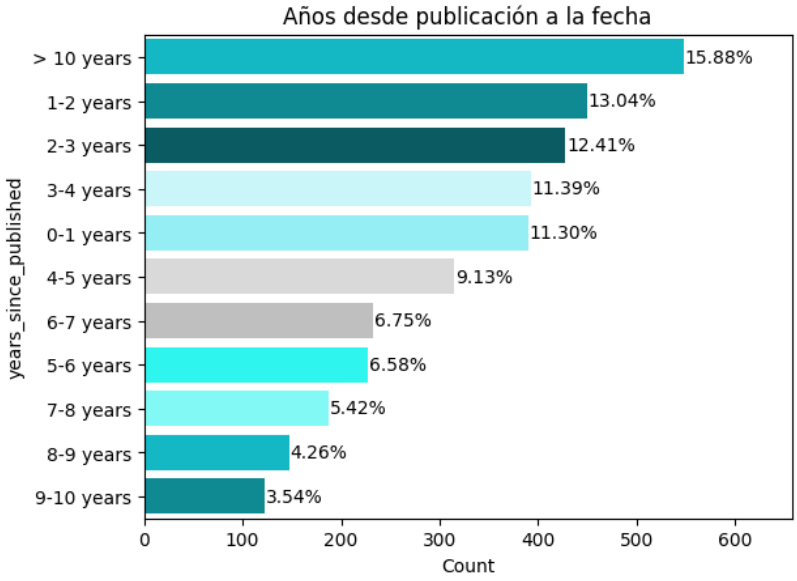
Un aspecto interesante observado fue la extensión de las publicaciones, medida mediante la variable "pageRange", que indica el número de inicio y fin de un documento. Inicialmente, esta variable no estaba estandarizada y contenía rangos en numeración romana, con episodios y números de página. Se aplicaron funciones para estandarizarla lo más posible y se creó la variable "n_pages" para reflejar la cantidad de páginas escritas por los autores en investigaciones científicas. Se identificó que, en promedio, un autor ha escrito 12 páginas, y solo 8 autores han escrito más de 500 páginas.

En cuanto a las veces que han sido citados los autores, 491 autores que se encuentran en la Plataforma X han sido citados alguna vez. Estos han sido citados en promedio 16 veces y habiendo usuarios con más de 400 citaciones en sus publicaciones.

CROSSREF

Las publicaciones que se tienen en esta base han sido publicadas entre 1970 y 2024, Aunque el 99% de las publicaciones que se tienen en la base fueron publicadas posterior al año 2000. Además, se tiene que el 50% de las publicaciones han sido publicadas dentro de los últimos 5 años, lo que nos da una idea de la recencia de publicación que tienen los usuarios. Esto se puede observar en el siguiente gráfico.

[7] Incluye artículos publicados en revistas, así como entradas específicamente etiquetadas como 'revista'.
[8] Comprende artículos presentados en actas de conferencias.

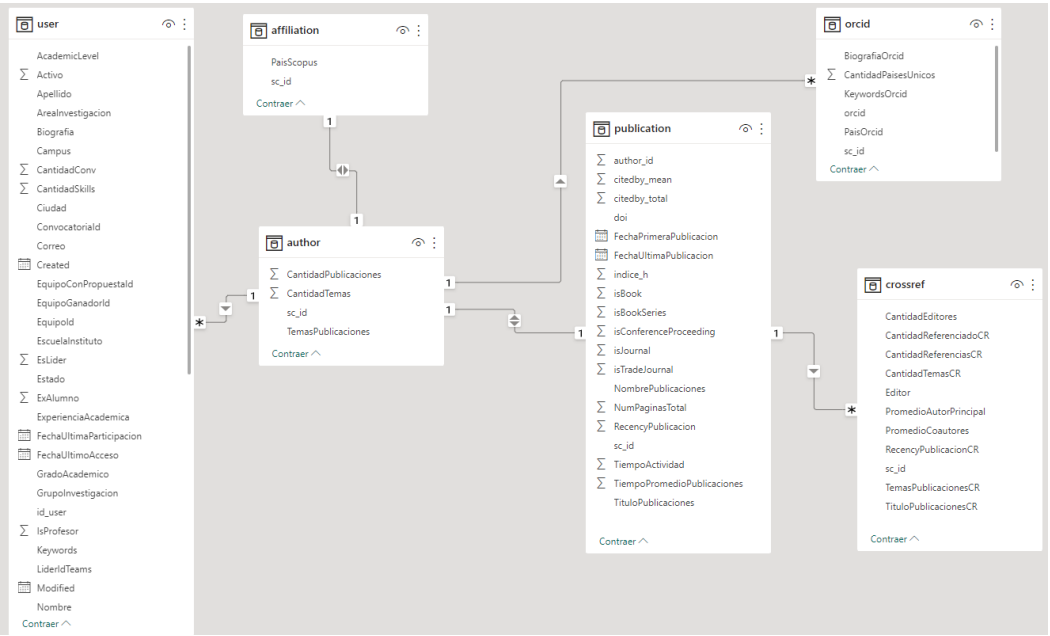


Años desde publicación a la fecha. Fuente: Elaboración propia.

Se encontró además que el 86% de las publicaciones son de tipo *journal* seguidas por *conference proceedings* con un 8%. Otras variables importantes analizadas en el EDA son la cantidad de referencias de los usuarios activos que poseen esta información, en donde han sido referenciados en promedio 15 veces y han referenciado en promedio a 58 autores. Los usuarios activos tienen 13 temas distintos en promedio en sus publicaciones y han tenido 4 editores en promedio. En términos de actividad en las publicaciones, los usuarios que cuentan con esta información no han publicado algo hace 700 días en promedio, con una desviación estándar de 800 días.

Anexo E: Preparación de los datos

E.1. Modelo Relacional entre las fuentes internas y externas



Fuente: Elaboración propia.

E.2. Supuestos para eliminar NA

Fuente: Elaboración Propia.

N°	Descripción de Supuestos
1	Si no tiene equipos con propuesta, se reemplazan los NA por 0, por consecuencia de la misma forma en los equipos ganadores y la proporción de equipos ganadores que tienen.
2	Si un autor tiene una recencia de las publicaciones negativa (Su última publicación aún no se publica en Scopus) se reemplazarán por 0.
3	Se asume que los usuarios sin publicaciones en Scopus no han realizado publicaciones, por lo tanto, no cuentan con ninguna información respecto a las publicaciones.
4	Los usuarios que no declaran si son exalumnos, se asume que no lo son.
5	Los usuarios que no declaran si son profesores o no, se asumirá que no lo son.
6	Si un usuario es profesor, se asumirá que es por lo menos es licenciado.
7	Los usuarios que no tienen nivel académico y son exalumnos se les dará el nivel de licenciado, los que tienen algún dato en Escuela/Instituto o Universidad, se le dará el nivel de estudiante.
8	Los usuarios que no tienen nivel académico, ni registro de la universidad o escuela en la que estuvo, pero tienen publicaciones se asumirá que por lo menos son licenciados.
9	Los usuarios que no declaran nada respecto de su nivel académico y no tienen publicaciones (41 usuarios) se asumirá que son estudiantes.
10	A los usuarios que tienen NA en "RecencyParticipacion" (días desde la última vez que participaron en las convocatorias) se les reemplazará por un número mayor al máximo de los que si han participado, para que así estos datos sean tratados de manera adecuada y se entienda su falta de participación.

E.3. Creación de variable de importancia de publicaciones

Se estandarizaron las 12 variables relacionadas a las publicaciones utilizando *MinMaxScaler* de *sklearn* y se les asignó un peso según se muestra en la Tabla.

Fuente: Elaboración Propia.

Variable	**Peso**	**Variable**	**Peso**
indice_h	0,35	isJournal	0,02
CantidadTemas	0,05	isTradeJournal	0,02
TiempoPromedioPublicaciones	0,1	CantidadReferenciasCR	0,02
isBook	0,1	CantidadEditores	0,02
isBookSeries	0,08	RecencyPub	0,1
isConferenceProceeding	0,02	TiempoActividad	0,12

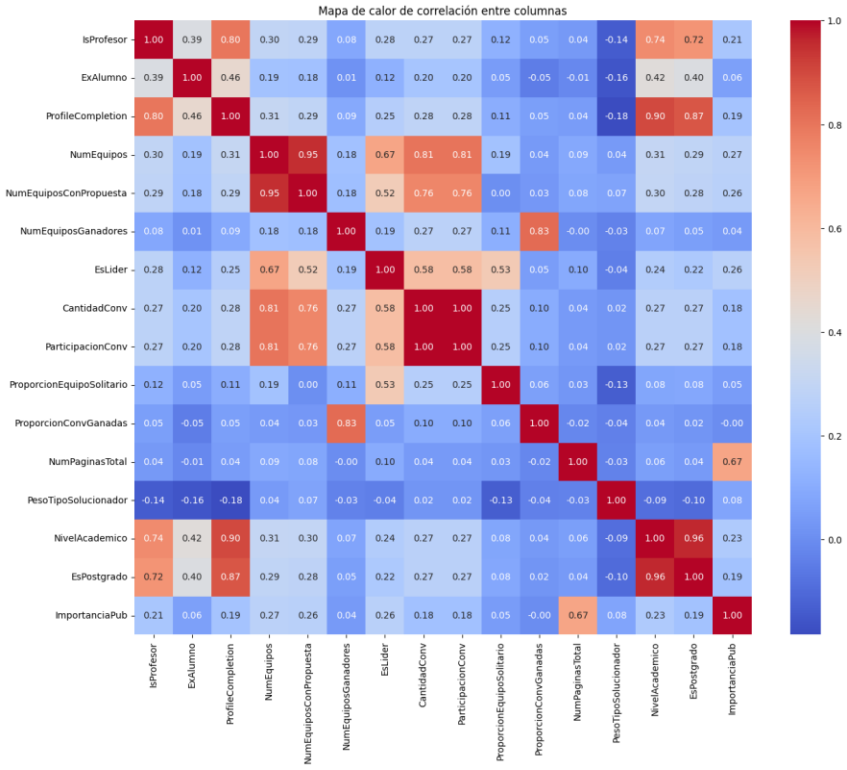
E.4. Variables de base de usuarios a segmentar.

Fuente: Elaboración Propia.

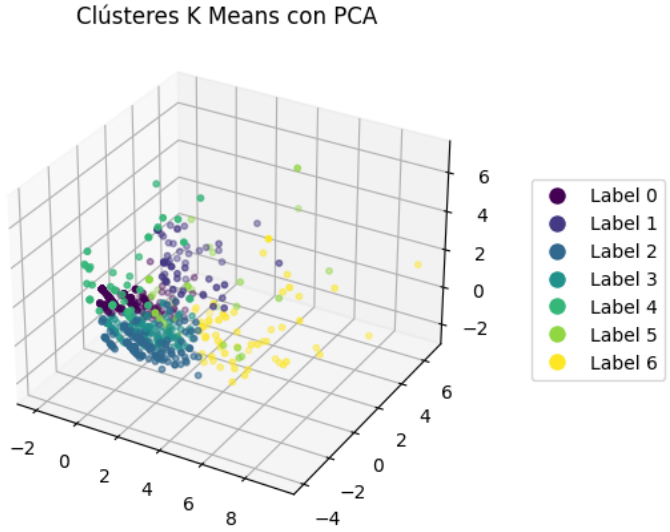
Variable	Tipo	Descripción
id_user	String	Id de los usuarios de la plataforma.
IsProfesor	Boolean	Indica si un usuario es profesor o no.
ExAlumno	Boolean	Indica si un usuario es exalumno o no.
ProfileCompletion	Int	Indica la cantidad de campos completados en perfil de un usuario en la plataforma (Los campos son Skills, Keywords, Biografía y Experiencia Académica).
NumEquipos	Int	Cantidad de equipos en los que ha participado
NumEquiposConPropuesta	Int	Cantidad de equipos con los que ha mandado propuesta.
NumEquiposGanadores	Int	Cantidad de equipos ganadores en los que ha participado.
EsLider	Int	Cantidad de veces que ha sido líder.
CantidadConv	Int	Cantidad de convocatorias en las que ha participado.
ParticipacionConv	Int	Proproción de convocatorias en las que ha participado.
ProporcionEquipoSolitario	Float	Proporción de equipos que ha formado en solitario.
ProporcionConvGanadas	Float	Proporción de convocatorias ganadas.
PesoTipoSolucionador	Float	Peso de tipo de solucionador, calculado como la frecuencia del tipo de solucionador dividido en la cantidad de solucionadores totales.
ImportanciaPub	Float	Variable que indica la importancia que tiene un usuario en el mundo de las publicaciones.
NivelAcademico	Int	Variable categórica que indica si un usuario es 1: Estudiante, 2: Licenciado, 3: Maestro, 4: Doctor.
EsPostgrado	Boolean	Variable booleana que indica si un usuario posee un postgrado o no.

Anexo F: Modelado

F.1. K Means

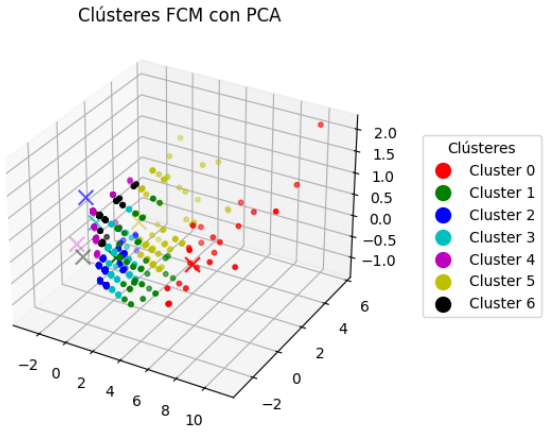


Matriz de correlación de las variables. Fuente: Elaboración propia.



Visualización de clústeres ocupando K Means. Fuente: Elaboración propia.

F.2. Fuzzy C Means



Visualización clústeres Fuzzy C Means. Fuente: Elaboración propia.

F.3. RFM

Definición de las variables R, F y M del RFM. Fuente: Elaboración propia.

Valor	Rango Recency	Rango Frequency	Rango Monetary
1	[113, 168]	0	0
2	[111, 112]	[1, 3)	1
3	[73, 110]	[3, 6)	[2, 4)
4	[0,72]	[6, ∞)	[4, ∞)

Combinaciones de Segmentos RFM. Fuente: Elaboración propia.

Nombre de Segmento	**Combinaciones RFM**	**# Usuarios**
Espectadores Desconectados	111, 211	129
Espectadores Conectados	311, 411	162
Baja Participación Desconectados	121, 122, 222	204
Baja Participación Conectados	321, 322, 422	256
Participantes Desconectados	123, 133, 223, 232, 233	46
Participantes Conectados	323, 331, 332, 333, 423, 432, 433	194
Destacados Desconectados	134, 234, 244	11
Destacados Conectados	334, 343, 344, 434, 443, 444	91

Anexo G: Evaluación de segmentos

G.1. K Means

Características de clústeres K Means por variable. Fuente: Elaboración propia.

	Clústeres						
	0	**1**	**2**	**3**	**4**	**5**	**6**
% IsProfesor	1%	47%	79%	78%	76%	54%	86%
% ExAlumno	1%	19%	100%	0%	50%	15%	35%
NumEquiposConPropuesta	0,78	0,79	1,64	1,57	0,7	2	5,6
NumEquiposGanadores	0	0	0	0	0,04	1,15	0,14
EsLider	0,08	1,81	0,4	0,29	0,45	0,85	3
ParticipacionConv	0,05	0,12	0,09	0,08	0,06	0,14	0,26
ProporcionEquipoSolitario	0	0,87	0,02	0,01	0,2	0,16	0,18
PesoTipoSolucionador	0,96	0,96	0,96	0,96	0,03	0,94	0,94
% EsPostgrado	4%	50%	93%	94%	74%	56%	88%
ImportanciaPub	0,04	0,08	0,07	0,09	0,04	0,08	0,14

G.2. Fuzzy C means

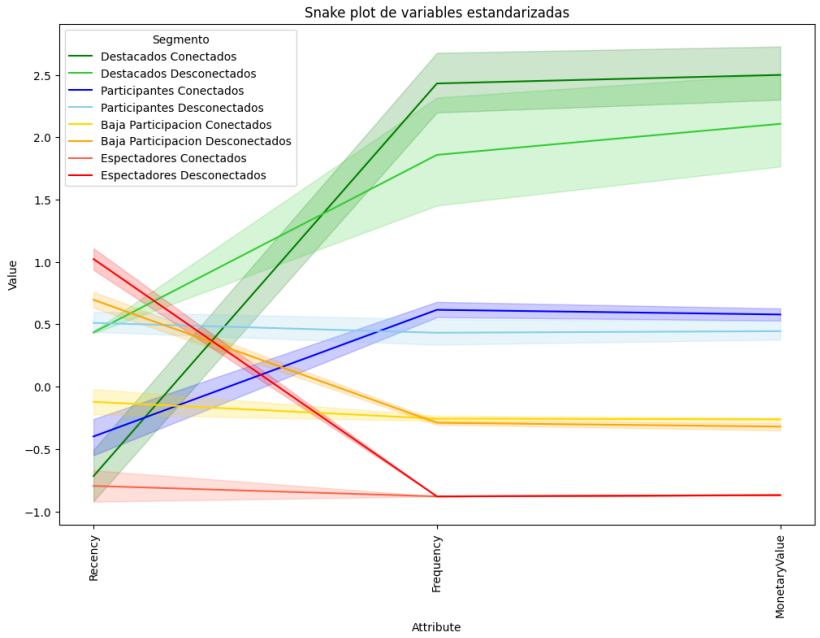
Características de clústeres FCM por variable. Fuente: Elaboración propia.

	Clústeres						
	0	**1**	**2**	**3**	**4**	**5**	**6**
% IsProfesor	88%	48%	92%	72%	3%	78%	3%
% ExAlumno	46%	33%	44%	31%	6%	36%	2%
NumEquiposConPropuesta	7,62	4,74	0,7	2,34	0	2,61	1,14
NumEquiposGanadores	0,25	0,13	0,05	0,02	0	0,15	0,04
EsLider	3,75	0,57	0,24	0,29	0,07	2,7	0,11
ParticipacionConv	0,3	0,17	0,07	0,12	0,01	0,16	0,08
ProporcionEquipoSolitario	0,16	0,02	0,11	0,01	0,06	0,34	0,02
PesoTipoSolucionador	0,96	0,96	0,87	0,94	0,92	0,92	0,95
% EsPostgrado	96%	64%	98%	83%	15%	78%	6%
ImportanciaPub	0,16	0,11	0,06	0,09	0,05	0,12	0,04

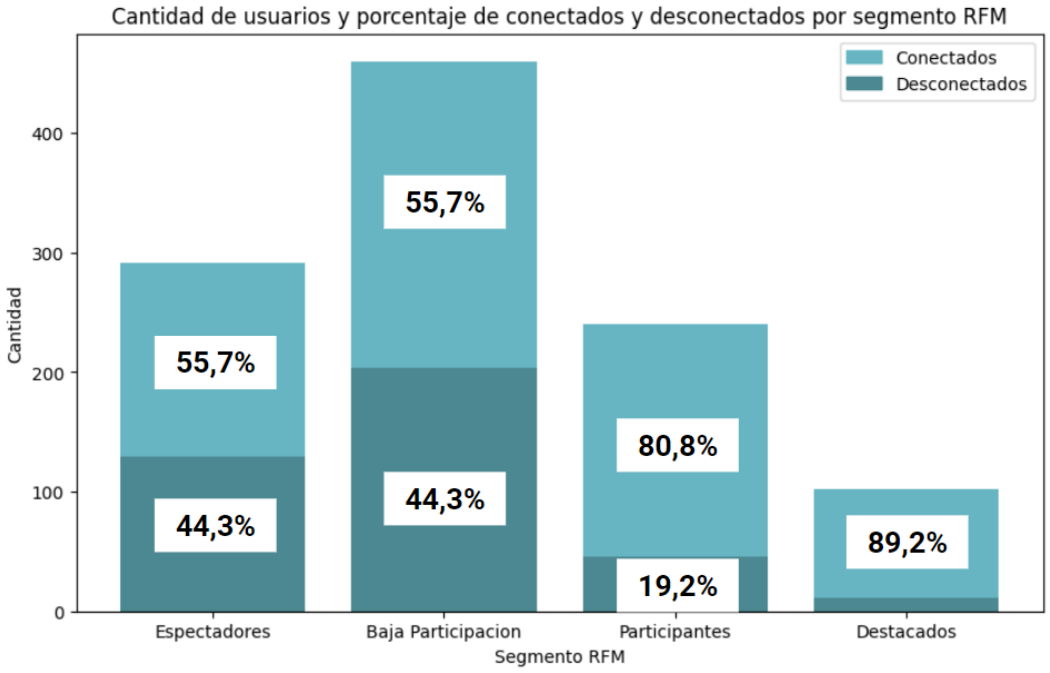
Proporción de segmento secundario de usuarios difusos. Fuente: Elaboración propia.

Segmento FCM	Segmento 2 FCM	% Usuarios Difusos	Total
Académicos	No Académicos sin Propuestas	30%	100%
	Participativos y Colaborativos poco Ganadores	34%	
	Pregrados	36%	
Líderes Influyentes	Líderes Solitarios Participativos	43%	100%
	Multipropuestas Poco Ganadores	57%	
Líderes Solitarios Participativos	Académicos	23%	100%
	Líderes Influyentes	9%	
	Multipropuestas Poco Ganadores	35%	
	No Académicos sin Propuestas	3%	
	Participativos y Colaborativos poco Ganadores	15%	
	Pregrados	15%	
Multipropuestas Poco Ganadores	Líderes Influyentes	25%	100%
	Líderes Solitarios Participativos	75%	
No Académicos sin Propuestas	Académicos	54%	100%
	Pregrados	46%	
Participantes Colaborativos poco Ganadores	Académicos	10%	100%
	Líderes Solitarios Participativos	13%	
	Multipropuestas Poco Ganadores	18%	
	Pregrados	59%	
Pregrados	Académicos	32%	100%
	Líderes Solitarios Participativos	7%	
	No Académicos sin Propuestas	10%	
	Participativos y Colaborativos poco Ganadores	51%	

G.3. RFM



Snake plot de segmentos RFM. Fuente: Elaboración propia.



Cantidad de usuarios y porcentaje de conectados y desconectados por segmento RFM. Fuente: Elaboración propia.