

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Negocios y Administración Pública

**CARRERA DE ESPECIALIZACIÓN EN
MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y
ANÁLISIS DE DATOS EN ORGANIZACIONES**

TRABAJO FINAL INTEGRADOR

**Ubicación de Cocinas Anfitrionas en la Ciudad de Mar
del Plata:**

Estimar Localizaciones Óptimas Aplicando Métodos de
Análisis Cuantitativo

AUTOR: EZEQUIEL JUAN BASSANO

TUTOR: MÓNICA CANTONI RABOLINI

[2023]

Resumen

El objetivo de este trabajo es el de predecir aquellas ubicaciones óptimas para la localización de cocinas anfitrionas en la ciudad de Mar del Plata a través de un modelo de negocio de plataforma que aplique métodos de análisis cuantitativo en base a datos sociodemográficas de cada barrio de la ciudad. El análisis se llevará a cabo a partir de la información disponible producto del censo realizado por el Estado Nacional en el año 2010. Las variables a estudiar son principalmente de carácter cuantitativo abarcando aspectos sociodemográficos, el nivel educativo, el rango etario de los habitantes y el hábito de consumo de alimentos solicitados bajo la modalidad de entregas a domicilio. La población analizada son los diferentes radios censales que representan los conjuntos poblacionales en los que se segmentó a los habitantes de la Ciudad de Mar del Plata durante el censo del año 2010. A través de la investigación, se logró concluir que resulta posible crear un modelo empresarial que prediga aquellos sectores donde se encuentre una brecha entre la oferta y la demanda del mercado. Al haber podido predecir donde existe una demanda insatisfecha de pedidos de comida con entrega a domicilio, se calcularon las ubicaciones óptimas para la localización de cocinas anfitrionas.

Palabras Clave

Cocinas Anfitrionas, Negocio de plataforma, Mar del Plata, Segmentos poblacionales, Análisis predictivo.

Índice

Ubicación de Cocinas Anfitrionas en la Ciudad de Mar del Plata: Estimar Localizaciones Óptimas Aplicando Métodos de Análisis Cuantitativo	3
1. Las Organizaciones de Plataforma	6
1.1 Modelo de Negocios de Hotkit.....	6
1.2 Descentralización y Protección de los Datos.....	13
1.3 Nuevo Paradigma del Big Data	16
2. Análisis de la Demanda Aplicando de Métodos de Análisis Multivariado.....	18
2.1 Ingesta y Almacenamiento de Datos	18
2.2 Manipulación, Transformación y Limpieza de Datos	21
2.3 Aplicación de Métodos de Análisis Multivariado para Caracterizar la Demanda. ...	23
3. Análisis de la Oferta mediante la Implementación de un Modelo Predictivo.....	30
3.1 Diseño del Modelo.....	30
3.2 Proceso de Selección del Modelo.....	33
3.3 Interpretación de Resultados	34
4. Conclusión.....	38
5. Referencias Bibliográficas	40
6. Anexos.....	42
7. Apéndices.....	43

Ubicación de Cocinas Anfitrionas en la Ciudad de Mar del Plata: Estimar Localizaciones Óptimas Aplicando Métodos de Análisis Cuantitativo

Este proyecto tiene su sustento principal a partir de observar la relevancia del mercado gastronómico a domicilio en la ciudad de Mar del Plata y el interés particular del investigador por innovar en la temática. Las *Dark Kitchen* representan un modelo de negocio novedoso y en auge, particularmente impulsado desde la creación de aplicaciones móviles que facilitan su implementación. La pandemia originada en el año 2019 y las consecuentes restricciones gubernamentales incentivaron temporalmente el hábito de consumo de comida a domicilio, sin embargo, han dejado una secuela en el comportamiento de los comensales que se entiende que ha llegado para quedarse.

El comportamiento de los usuarios respecto al consumo de alimentos ha sufrido cambios sustanciales en los últimos años. El arribo de aplicaciones móviles de entrega a domicilio y la búsqueda de satisfacción inmediata, por parte de los individuos pertenecientes a la generación Millennial y Centennial, han creado una sinergia perfecta para impulsar la demanda de comida a domicilio. La solicitud de este tipo de pedidos se ha convertido en un hábito recurrente en pos de la eficiencia de la administración del tiempo de los usuarios que prefieren destinar horas adicionales en su trabajo o priorizar el ocio.

Las plataformas de entrega a domicilio han colaborado en reducir la brecha existente entre el costo de preparar una comida de forma aficionada en el hogar, por sobre la posibilidad de recibir un producto confeccionado por un experto culinario. La variedad de platos disponibles ofrecidos a través de aplicaciones móviles hacen del momento de selección del plato de comida un ritual dentro de un contexto social.

A partir del año 2019, las restricciones producto de la pandemia originada por el COVID-19, actuaron como una suerte de potenciador de una tendencia ya existente de aumento de pedidos de productos alimenticios a domicilio. En este contexto, el sector gastronómico fue afectado, experimentando un fuerte cambio en su paradigma, trasladándose hacia América Latina el curso de acción que ya había tomado forma en países Europeos.

A pesar de que las compañías de entrega de alimentos a domicilio han sabido aprovechar esta oportunidad, en un contexto donde gran parte del mundo se sintió amenazado, existen ciertas limitaciones en su modelo. El obstáculo detectado radica en que los sistemas de entrega

funcionan, sin excepción, dentro un determinado radio urbano preestablecido. Existen entonces zonas con demanda insatisfecha donde los productores no están dispuestos a hacer una inversión mientras una porción del mercado objetivo queda desabastecida.

La brecha entre aquellas zonas marginadas y las áreas habilitadas de reparto podrían ser abarcadas por lo que se lo denomina hoy como *Dark Kitchen*. Esta terminología puede traducirse del inglés como Cocinas Oscuras. Poseen múltiples acepciones tales como cocina oculta, *Dark Kitchen*, *Cloud Kitchen*, *Ghost Kitchen*, restaurante fantasma o restaurante solo por delivery (Romero, 2021). El modelo de negocio de las *Dark Kitchen* responde a un esquema de organización lineal.

Las *Dark Kitchen* son en esencia cocinas comerciales, es decir, locales que se equipan con las instalaciones necesarias para la elaboración de platos de comida con niveles de producción equivalentes a los de un restaurante. La característica principal de las mismas es que la venta se efectúa estrictamente a través de comercio electrónico y que el pedido llega a manos del cliente exclusivamente por servicio de entrega a domicilio. Excepcionalmente existen algunas *Dark Kitchen* que habilitan el sistema de retiro en la puerta del local pero bajo ningún punto de vista el consumidor ingresa a las instalaciones. Si bien la venta se efectúa mediante el comercio electrónico, esto no implica que se trate de un modelo de negocio plataforma.

En este trabajo se desarrollará el modelo de negocio de una *Dark Kitchen* adaptando su funcionamiento de organización lineal hacia el de una empresa de plataforma. El nombre asignado a esta organización es Hotkit y se ha decidido adoptar el término de Cocinas Anfitrionas para distinguirlo del término *Dark Kitchen* el cual hace referencia a un modelo de negocio lineal.

Hotkit busca generar ingresos desde su rol como intermediario y comunicador entre dos tipos de clientes. El activo principal de la organización serán los datos obtenidos de distintas fuentes. Uno de los desafíos más grandes que se presentan para Hotkit es el de lograr utilizar los datos obtenidos para crear valor prediciendo aquellas locaciones con mayor potencial para la instalación de cocinas anfitrionas.

Dado que uno de los servicios que se pretende brindar en la empresa Hotkit, es el de asesorar en cuanto a las mejores localizaciones para la instalación de cocinas anfitrionas, entonces el objetivo de este trabajo es el de predecir aquellas ubicaciones óptimas para la localización de

cocinas anfitrionas en la ciudad de Mar del Plata a través de un modelo de negocio de plataforma que aplique métodos de análisis cuantitativos en base a datos sociodemográficos de cada barrio de la ciudad.

Para abordar este objetivo principal, el presente trabajo se divide en tres apartados. En el primer apartado se desarrollará el concepto de organizaciones de plataforma destacando la relevancia de los efectos de red, se planteará el Plan de Negocios de Hotkit y se elevará la importancia de un sistema de información descentralizado. En el segundo apartado se realizará el diseño de una Arquitectura de Datos en el que se describirá un complejo proceso de ingesta, transformación, manipulación y almacenamiento de datos de diferentes orígenes, y que culminará con la aplicación de métodos de análisis multivariado sobre los datos de la población a analizar. En último lugar, se detallará el proceso de implementación de un modelo analítico predictivo que permita predecir la localización con mayor demanda gastronómica insatisfecha para el cual se realizará la selección de variables relevantes, la asignación de roles, el ajuste de hiperparámetros, la evaluación y selección del modelo más efectivo.

1. Las Organizaciones de Plataforma

1.1 Modelo de Negocios de Hotkit

Las organizaciones de plataforma utilizan a los datos como uno de sus principales activos. Esta afirmación implica que la creación de valor se lleva adelante mediante la transformación de los mismos en acciones inteligentes. Estas acciones inteligentes son descubrimientos o *insights*. Según López (2007), un *insight* es cualquier verdad sobre el consumidor cuya inclusión en un mensaje publicitario hace que este gane en notoriedad, veracidad, relevancia y persuasión a ojos de dicho consumidor.

El modelo de negocios de Hotkit es caracterizado como el de una organización de plataforma. El objetivo de la empresa es poner en contacto a emprendedores gastronómicos que necesitan un espacio acorde donde elaborar sus productos, con personas u organizaciones que cuentan con espacios equipados disponibles y están dispuestos a alquilarlos. Estos espacios equipados funcionan como cocinas comerciales que los emprendedores pueden alquilar por periodos de tiempo determinado, pagando una renta de acuerdo al periodo por el cual se contrate y conforme el equipamiento ofrecido en el mismo. La empresa Hotkit es el mediador en el proceso de búsqueda, selección, contratación, cobranza y seguimiento de clientes, brindando a emprendedores un sistema de recomendación con los espacios disponibles para trabajar.

Los datos son fundamentales en este negocio porque la evaluación que se hace para ponderar el éxito de cada locación se realiza a través de un análisis sociodemográfico de la población. La mayor parte de los datos utilizados para crear el modelo son de carácter público pero en la medida que el negocio capta clientes, se recopilan sus datos para incorporarlos a la base de original. Al ser incorporados, se logra mejorar la efectividad del modelo analítico que prediga el éxito potencial de cada espacio ofrecido disponible en alquiler.

En la medida que se populariza la marca, incrementa la cantidad de espacios disponibles ofrecidos y, en consecuencia, mayor será la cantidad de emprendedores gastronómicos contratando servicios a través de la plataforma. El incremento de la cantidad de suscriptores mejora el sistema para sugerir la locación óptima y estimula la inserción de nuevos usuarios a la plataforma a medida que se consolida la confianza en la misma cada vez que se realizan operaciones.

Entonces, los datos pueden ser utilizados para captar nuevos clientes a través de la lectura de *insights*. Los insights pueden entenderse como una interpretación de los datos de los clientes en beneficio del negocio. Esto tiene sustento en investigaciones como la de Morillas, Soladana y

Mediavilla (2019) quienes concluyen que un buen insight contribuye a conectar la marca con el consumidor.

De acuerdo al artículo de (Lavalle, 2011), Hotkit podría encuadrarse como una *Transformed organization*, es decir, como una empresa con una capacidad fuerte para capturar, manipular y analizar datos. En la medida que se incrementan los datos y se realice una lectura atinada de los insights, se generará un efecto positivo atrayendo a nuevos clientes. Este efecto se lo denomina como Efecto de Red.

El efecto de red refiere al impacto que tiene la cantidad de usuarios en una plataforma sobre los mismos. Los efectos de red pueden ser tanto positivos como negativos. Los efectos de red positivos refieren a la habilidad de producir valor relevante para cada usuario de una plataforma grande y correctamente administrada. Los efectos de red negativos refieren a la posibilidad de que el crecimiento de una comunidad de plataforma administrada pobremente puede reducir el valor producido para cada usuario.

El modelo de negocios de Hotkit consta de dos segmentos de clientes donde por un lado se encuentran los elaboradores que son los demandantes de espacios para fabricar y distribuir alimentos y, por otro lado, los oferentes de las cocinas equipadas.

En Hotkit se presentan efectos de red positivos, esto quiere decir que en la medida que se incrementa la cantidad de oferentes aumenta entonces la cantidad de demandantes y viceversa. El aumento de la cantidad de oferentes mejora:

- La cantidad de espacios disponibles,
- La diversidad de las cocinas adaptadas para cada necesidad,
- El posicionamiento de la marca. Siguiendo la definición de posicionamiento entendido como la acción de diseñar la oferta y la imagen de una empresa de tal modo que éstas ocupen un lugar distintivo en la mente de los consumidores (Kotler y Keller, 2006), la plataforma logra posicionarse como exitosa en la mente de los clientes demandantes y de nuevos oferentes.

El modelo de negocio a desarrollar se estructura de acuerdo al *framework* desarrollado por Osterwalder y Pigneur (2010) denominado Canvas. Los autores proponen estructurar el modelo de negocios en 9 dimensiones donde la propuesta cobra protagonismo. Las dimensiones a desarrollar son las descriptas a continuación.

1.1.1 Segmento de Clientes

La plataforma crea una conexión mediando entre dos segmentos de cliente. Por un lado, están los elaboradores que buscan espacios de trabajo y, por otro lado, están los oferentes que cuentan con espacios de trabajo disponible. La conexión implica un vínculo comercial entre las partes que se materializa a través de la página web de HotKit. Los espacios de trabajo son exclusivamente aquellos equipados y habilitados con lo necesario para funcionar como cocinas comerciales.

Los elaboradores son marcas fabricantes de productos alimenticios que buscan espacio de trabajo equipados como cocinas ciegas comerciales por tiempo determinado para desarrollar su proceso productivo. Los elaboradores pueden ser restaurantes existentes que buscan expandir su radio de entrega a domicilio accediendo a zonas fuera de su área de cobertura, emprendedores gastronómicos independientes que necesitan un espacio de trabajo a bajo costo o negocios gastronómicos que buscan insertarse en una zona determinada mediante la modalidad *Only-Delivery*.

Los oferentes son aquellos propietarios de inmuebles con instalaciones disponibles suficientes para efectuar actividades de índole gastronómica. Los oferentes pueden ser espacios destinados específicamente para ser arrendados a elaboradores u organizaciones que por diferentes motivos cuentan con cocinas esquipadas operando con capacidad ociosa por periodos prolongados de tiempo de manera que están disponibles para ser utilizados por terceros. Algunos ejemplos de oferentes son las cocinas de hoteles, balnearios y de centros de recreación en temporada baja.

1.1.2 Propuesta de Valor

El modelo de negocio de Hotkit se basa en el alquiler de cocinas, por periodos de tiempo pre-pactado las cuales están equipadas para la elaboración de productos del mercado gastronómico y están instaladas en ubicaciones con carácter estratégico. La empresa de Plataforma actúa como mediador interviniendo en el proceso de búsqueda, selección, contratación, cobranza y seguimiento de clientes brindando a emprendedores un sistema de recomendación con los espacios disponibles de acuerdo a las prestaciones buscadas y ubicados geográficamente cerca del mercado objetivo al que apuntan. Este modelo de negocio entrega valor al cliente a través de la mejora en la *performance*, reducción de costos, reducción del riesgo, accesibilidad y conveniencia.

1.1.2.1 Performance. Esto aplica específicamente en el caso de los pequeños emprendedores elaboradores que habitualmente son quienes no cuentan con equipamiento profesional y a través de la plataforma logran agilizar procesos de producción.

1.1.2.2 Reducción de Costos. Los gastos de luz, gas y mantenimiento de instalaciones propios de un emprendimiento gastronómico se ven acotados al pago de un arancel periódico inferior para los clientes elaboradores. Los clientes oferentes con instalaciones operando a capacidad ociosa recuperan parte de sus costos fijos a través del cobro del alquiler.

1.1.2.3 Reducción del Riesgo. Los clientes elaboradores evaden la necesidad de invertir en equipos e instalaciones que conlleven a un elevado riesgo financiero en el que las ventas obtenidas no justifiquen la inversión realizada. Los clientes oferentes reducen el riesgo operativo de afrontar los costos fijos de sus instalaciones por cuenta propia.

1.1.2.4 Accesibilidad. Permite acceder a clientes a los cuales no se habría podido llegar sin este servicio. Funciona como un amplificador del rango de cobertura de servicio de delivery. Además, la empresa brinda valor levantando barreras de entrada hacia mercados no explorados o inaccesibles para los clientes elaboradores.

1.1.3 Canales de Distribución

1.1.3.1 Para los Elaboradores. A través de la publicidad en la web se produciría la totalidad del proceso de difusión y distribución del servicio abarcando la publicidad, evaluación, adquisición, entrega y servicio post-venta. Adicionalmente, a través de los aliados estratégicos se advierte acerca de la existencia del servicio para canalizar nuevos clientes.

1.1.3.2 Para los Oferentes. Se mantienen los medios de comunicación establecidos para los elaboradores con la diferencia de que se agrega una fuerza de ventas con carácter presencial para atender proceso de evaluación de la propuesta de valor y hacer un seguimiento como parte del servicio post-venta para retener estos clientes.

1.1.4 Contacto con Clientes

1.1.4.1 Para los Elaboradores. Utilizando un sistema de autoservicio la compañía no establece vínculo directo con los clientes, a través el sitio web estos tienen todos los medios necesarios para contratar el servicio, informarse acerca de la compañía, los beneficios de utilizarla y los términos y condiciones previstos de la relación comercial. Accediendo a un sistema automatizado el cliente recibe sugerencias acerca del servicio de forma automática por un sistema de recomendaciones automático.

1.1.4.2 Para los Oferentes. La fuerza de ventas trata con atención dedicada a los oferentes para mejorar la calidad del servicio brindado y mejorar la relación comercial con los elaboradores.

1.1.5 Flujos de Ingreso

1.1.5.1 Cobro por comisión. El alquiler del espacio entre los oferentes y los elaboradores se realiza a través del sitio web de Hotkit. La misma actúa como intermediario de cobranza del alquiler que paga el elaborador, retiene una comisión y paga el saldo al oferente del espacio. El precio del servicio se establece a partir de una lista de precios fija y de acuerdo a las prestaciones del espacio alquilado, la localización del mismo y el periodo de tiempo por el cual se contrate el servicio pudiendo este ser por hora, semana, mes o anual.

1.1.5.2 Entrada sin fricción. Según lo define Parker (2016), se entiende como entrada a la plataforma sin fricción como aquellas en que los usuarios puedan unirse a la plataforma fácil y rápidamente. Este concepto se arraiga a la posibilidad de que los clientes accedan de forma ágil participando en la creación de valor de la organización. La capacidad de *Frictionless entry* es un factor clave que permite a una plataforma crecer rápidamente.

1.1.5.3 Monetización. La plataforma ha optado por aquellas formas de monetización que resultan acordes a la estrategia general del negocio. Si bien la plataforma es de acceso gratuito, es decir, no se generan ingresos por el registro de nuevos usuarios, esta logra monetizar de otras maneras. El flujo de ingresos se genera cobrando una comisión por transacción, cobrando por un acceso mejorado o por una calidad garantizada. Pertenecer a la comunidad Hotkit es gratuito, fácil y rápido.

1.1.5.3.1 Ingresos por comisión. Dado que los clientes acceden a la plataforma de forma gratuita no experimentan ninguna forma de desmotivación para ingresar y registrarse en la misma formando parte de la red. El valor de la comisión resulta difícil de determinar porque se busca el más alto posible para el cual los clientes no se sientan desalentados y abandonen la plataforma. Un particular riesgo de esta forma de ingreso recae en la posibilidad de que los clientes elaboradores y los clientes oferentes se contacten mutuamente por fuera de la plataforma. La reducción de este tipo de interacciones que ocurren sin brindar beneficio a la plataforma se logra consolidando un sistema de recomendaciones que facilite el trabajo para ambas partes de forma tal que el valor que otorga la plataforma represente más que el costo de pagar la comisión.

1.1.5.3.2 Ingresos por acceso mejorado. Los clientes del segmento oferente pueden abonar un importe mensual para obtener mejor posicionamiento en su difusión dentro de la plataforma.

Las ofertas en este caso se las denomina patrocinadas. Los clientes elaboradores pueden visualizar claramente que se trata de un espacio que está siendo patrocinado a través de una cuenta *Premium* y que le ha sido sugerido porque la contraparte ha pagado un diferencial previamente. Tal y como indica Parker (2016) al referirse a la plataforma Yelp, uno de los principios más importantes es asegurar que el consumidor puede distinguir entre el contenido que ha sido destacado como parte de un programa pago y aquel que recibe de forma orgánica.

Cada oferente tiene un límite de tres avisos patrocinados disponibles por año. Se requiere haber realizado al menos una interacción exitosa entre clientes previamente para acceder a avisos patrocinados. El origen del requisito mencionado precedentemente tiene como fundamento la necesidad evitar que unos pocos oferentes acaparen toda la cuota de mercado dentro de la plataforma sin contribuir a la generación de valor dentro de la comunidad.

1.1.5.3.3 Ingresos por una calidad garantizada. Otra forma de ingreso para la plataforma es a través de las reservas prepagas. La contratación de cada espacio puede realizarse sin reserva previa pero el cliente puede pagar un adicional para garantizarse el acceso a un lugar que le resulta de preferencia. Los clientes elaboradores que pertenecen al segmento demandante de espacios pueden reservar una cocina disponible realizando depósitos reembolsables. La capacidad de que este reembolso sea total o parcial es delimitada por el oferente del espacio pero en todos los casos la plataforma retiene una suma mínima para solventar costos de transacción.

1.1.6 Recursos Clave

1.1.6.1 De Carácter Intelectual. Se requieren principalmente en torno a la consolidación de la imagen de la marca, la protección de la propiedad intelectual del proyecto para reducir la entrada de competidores en el mercado que copien el modelo y el fortalecimiento del vínculo con los aliados estratégicos. Adicionalmente se destaca la importancia de la capacidad para administrar bases de datos y capitalizarlos. Este último aspecto tiene relevancia en este modelo de negocio de plataforma dado que los datos son un activo fundamental para conectar a oferentes con elaboradores y consolidar un sistema de recomendación automática. El almacenamiento seguro es esencial para proteger la privacidad de los usuarios garantizando confidencialidad de la información suministrada y ganar confianza de clientes.

1.1.6.2 Recursos Humanos. Tratándose de una organización basada en el conocimiento se requiere conformar un equipo de trabajo al desarrollo de modelos predictivos que requieren una experticia enfocada en lo relativo a las ciencias de la computación y de la ciencia de datos. El

personal a cargo de la manipulación de datos requiere capacitación y actualización constante en lo relativo a nuevas tecnologías.

1.1.7 *Actividades Clave.*

El foco de las actividades clave se encuentra en el mantenimiento de la plataforma y la resolución de problemas. El mantenimiento de la plataforma implica considerar aspectos estéticos y técnicos. Desde su aspecto estético se contemplan actividades que conlleven a realizar cambios visuales que acompañen la imagen de la marca y mejorando la experiencia de usuario. Contemplando un aspecto técnico se hace foco en la actualización del sistema para agilizar el almacenamiento, la administración de los datos obtenidos y el funcionamiento desde una perspectiva operativa de la plataforma en sí. Respecto a la resolución de problemas, a partir de la información obtenida del servicio post-ventas, las quejas y sugerencias de clientes se constituye una fuente de información que retroalimenta la institución para realizar cambios para mejorar la experiencia de usuarios y resolver potenciales disconformidades de ambos tipos de clientes.

1.1.8 *Aliados Estratégicos*

Se establece una alianza estratégica entre no-competidores con las organizaciones de entrega de comida a domicilio. Los elaboradores fabrican alimentos para vender por medio de las aplicaciones móviles desarrolladas por organizaciones de entrega a domicilio. La plataforma de las Cocinas Anfitrionas requiere de la existencia de un mercado prolífico de entregas de comida a domicilio. La misma otorga valor creando mayores puntos de elaboración y venta, aumentando las entregas de comida a domicilio y, por ende, mejorando los ingresos de las empresas de reparto como PedidosYa de quien se han tomado datos para la confección del modelo.

1.1.9 Estructura de Costos

Se trata de un modelo de negocio impulsado por la creación de valor con poca relevancia en la minimización de costos. El foco principal de la organización es crear valor a partir de sus recursos y actividades, trabajando fuertemente la relación con los clientes.

Los costos fijos de la organización son principalmente de salarios y, en menor medida, impactan los de mantenimiento de la plataforma. La presencia de costos variables detectada tiende a ser nula. Se trata de una organización con economía de escala dado que en la medida que se expande incrementando el nivel de actividad, el costo de concretar relaciones entre oferentes y elaboradores es cada vez menor.

1.2 Descentralización y Protección de los Datos

1.2.1 Acerca de los Datos

Resulta relevante señalar la importancia de resguardar los datos obtenidos de clientes de forma segura para evitar transgresiones a la privacidad de estos. *Existe una creciente preocupación de que las tecnologías avancen de manera responsable. Las tecnologías emergentes entran en un mundo regulado delimitado por la supervisión regulatoria de los problemas sociales.* (Youtie, 2017). Teniendo en cuenta el aspecto mencionado precedentemente, se destaca tanto la oportunidad que representan los datos para tomar decisiones estratégicas como la responsabilidad que implica operar con la privacidad de la información de terceros.

La forma en que las organizaciones administran y almacenan sus datos puede ser tendiendo hacia la centralización o la descentralización de los mismos. Tradicionalmente, las instituciones que manipulan grandes volúmenes de datos lo hacen utilizando un sistema centralizado sobre el cual recae la responsabilidad del resguardo de esta información. Por otro lado, los avances tecnológicos hacen posible el resguardo de los datos utilizando herramientas de almacenamiento con carácter descentralizado.

Desde el año 2008 ha cobrado relevancia la concepción de métodos de descentralización de datos de transacciones a partir de la creación del *white paper* firmado a nombre de Satoshi Nakamoto donde se detalla la idea de un protocolo financiero descentralizado en torno a una nueva moneda denominada *Bitcoin*. Globalmente el *Bitcoin* ha tenido su foco de atención por ser una moneda digital descentralizada que no requiere la intervención de un Banco Central para su funcionamiento. En la siguiente parte del trabajo se hará particular foco en la aplicación práctica de la tecnología *blockchain* detrás de las cryptomonedas.

1.2.2 Tecnología Blockchain

La tecnología *blockchain* funciona como un libro digital compartido e inmutable que facilita el proceso de registro de transacciones y de seguimiento de activos en una red de negocios. De esta tecnología iniciada durante la creación del *Bitcoin* nacieron otras como es el caso de Ethereum. La importancia fundamental detrás de Ethereum radica en la incorporación de los denominados contratos inteligentes o *Smart Contracts* elaborados a partir de la tecnología *blockchain*.

Estos contratos operan de forma similar a uno convencional en el que se indican deberes y obligaciones de las partes con la particularidad de que no están sujetos a intermediarios para su cumplimiento y su funcionamiento se basa en códigos informáticos llamados *script*. Los contratos inteligentes son particularmente eficientes y seguros gracias a su administración de carácter descentralizado. Si bien la explicación técnica del funcionamiento de esta tecnología es extensa de desarrollar, resulta importante destacar que los beneficios de uso están siendo aprovechados en diferentes sectores productivos como en logística, prestación de servicios legales y, principalmente, en el campo de la informática.

1.2.3 La Descentralización de los Datos

Luego de haber expuesto dos posturas de gestión de datos opuestas resulta interesante la posibilidad de plantear la necesidad de incorporar en la empresa Hotkit un sistema de gestión de la información con carácter descentralizado a través de la tecnología *blockchain*. La empresa manipula información de dos tipos de clientes distintos que en ambos casos requiere confidencialidad y deben ser fehacientemente resguardados. A través de la implementación de un sistema descentralizado podrían obtenerse estas ventajas.

Asimismo, se destaca la capacidad de reducir el margen en torno hacia la vulnerabilidad de la difusión de los datos de clientes por fuera de la organización. La privacidad de los datos se alcanzaría trabajando en la reducción del riesgo de sufrir un ciber-ataque. Manteniendo la seguridad de los datos conlleva a incrementar la confianza depositada por terceros en la organización.

De forma tradicional Hotkit operaría con servidores propios almacenando la totalidad de la información en un único lugar. Si bien esto suena interesante, en realidad esto provocaría cantidades incrementales en el tiempo de recursos humanos asignados a tareas de almacenamiento y registración de transacciones. La centralización de los datos pareciera poseer una correlación inversa en cuanto a la eficiencia en la gestión de los mismos. Tal y como

enuncia Moxie (2022), existiría en este caso la necesidad de desarrollo de una amplia infraestructura y habría una tendencia hacia la complicación del desarrollo de soluciones.

1.2.4 Ley 25.326 de Protección de los Datos Personales

El objeto de la Ley de Protección de Datos Personales se encuentra enunciado en su artículo primero entendido este como la protección integral de los datos personales asentados en archivos, registros, bancos de datos, u otros medios técnicos de tratamiento de datos, sean éstos públicos, o privados destinados a dar informes, para garantizar el derecho al honor y a la intimidad de las personas, así como también el acceso a la información que sobre las mismas se registre, de conformidad a lo establecido en el artículo 43, párrafo tercero de la Constitución Nacional. Adicionalmente, es necesario mencionar las distinciones presentes en esta Ley respecto a los Responsables de archivo y los Titulares de los mismos.

— Responsable de archivo, registro, base o banco de datos: Persona física o de existencia ideal pública o privada, que es titular de un archivo, registro, base o banco de datos.

— Titular de los datos: Toda persona física o persona de existencia ideal con domicilio legal o delegaciones o sucursales en el país, cuyos datos sean objeto del tratamiento al que se refiere la presente ley.

1.2.5 Administración y manipulación de datos de terceros

A partir de los conceptos enunciados precedentemente debe entenderse a la Plataforma Hotkit como el agente responsable del archivo, registro y administración de la base de datos que realizará operaciones y análisis de tratamiento de datos de sus dos segmentos de clientes. La transformación de datos será en torno a fines promocionales, comerciales y publicitarios. Para que esto suceda se requiere el consentimiento de los titulares de los datos, que en este caso son los clientes oferentes y clientes elaboradores.

Más allá del consentimiento requerido deberán respetarse los derechos de los titulares de los datos previstos por la Ley de Protección de Datos. El consentimiento debe ser siempre explícito, es decir, no existe en forma tácita. El requerimiento de consenso con las partes debe darse siempre que se realice una cesión o transferencia del banco de datos a un tercero o para el uso de los mismos por la propia compañía. Hotkit deberá indicar al titular del dato conforme el artículo 6 de la Ley 25.326 la finalidad para la que serán tratados y quiénes pueden ser sus destinatarios o clase de destinatarios; la existencia del archivo, registro, banco de datos, electrónico o de cualquier otro tipo, la identidad y domicilio de su responsable; el carácter

obligatorio o facultativo de las respuestas al cuestionario que se le proponga; las consecuencias de proporcionar los datos, de la negativa a hacerlo o de la inexactitud de los mismos; así como la posibilidad del interesado de ejercer los derechos de acceso, rectificación y supresión de los datos.

1.3 Nuevo Paradigma del Big Data

Partiendo de la definición de *Big Data* como un proceso de transformación contemporánea de la economía y la sociedad, el autor Zuboff (2015) se refiere a aquellos grandes manipuladores de datos con prácticas que vulneran a las libertades individuales como *Big Other*. Él entiende a los monopolios acaparadores de datos como propietarios de una arquitectura automatizada y omnipresente. Lo que el autor describe son situaciones de manifestación expresa de posición dominante. Textualmente habla de los mecanismos de las grandes entidades como “unexpected and illegible mechanisms of extraction and control that exile persons from their own behavior”[Mecanismo inesperado e inteligible de extracción y control que exilia a la gente de su propio comportamiento]. Esta frase hace referencia a cómo algunas organizaciones se apropian de grandes volúmenes de registros y operaciones que realizan los usuarios pudiendo manipular y predecir de forma inescrupulosa el comportamiento de los consumidores al punto de vulnerar derechos y libertades personales.

La postura de Zuboff respecto a la manipulación de datos sensibles se percibe radical al contraponer la democracia con lo que él denomina capitalismo de vigilancia. Este último concepto remite a la apropiación de los datos de los titulares por parte de las organizaciones. Zuboff (2015) afirma que no puede haber un contrato social a partir de la ignorancia inducida por el engaño y concluye que la libertad a partir de lo incierto no es libertad cuando dice que “deception-induced ignorance is no social contract, and freedom from uncertainty is no freedom”. Sin duda esto hace referencia al desconocimiento de aquello que aceptan quienes ingresan a ciertas plataformas las cuales ejercen una posición dominante por sobre sus usuarios. Por otro lado, la postura de Varian (2014) quien ilustra sus ideas ejemplificando frecuentemente con la empresa Google, resulta un tanto más optimista. Este entiende las nuevas posibilidades resultantes del incremento del volumen, variedad y velocidad de los datos que definen al Big Data como el acceso a una nueva forma de contratos mientras para Zuboff en realidad implica el fin de los contratos.

La relación comercial entre Hotkit y sus clientes se entiende como una comunicación donde el dato actúa como el mensaje que articula el nexo entre usuarios y la organización. Si bien existe

una comprensión clara de los problemas que señala Zuboff, la visión organizacional se encuentra orientada hacia el optimismo de Varian quien destaca la munificencia positiva de la personalización, la experimentación constante y la posibilidad de realizar monitoreo de transacciones en línea. Sumada a esta línea de pensamiento, Hotkit tiene un fuerte compromiso legal y ético que se hace tangible al alinear su funcionamiento dentro del marco legal argentino en torno a la protección de datos personales.

Visto los argumentos presentados precedentemente resulta posible afirmar que Hotkit puede operar como una organización de plataforma sobre la cual debería aplicarse un sistema de gestión de datos descentralizado. Esta decisión tiene su fundamento en el interés de garantizar la protección de los datos provistos por sus usuarios, garantizando su confianza, y amparando las actividades del negocio dentro del marco legal del país en que se desempeña. Hasta el momento no se detectan perjuicios o motivos suficientes para no utilizar tecnologías como *blockchain* para la administración de datos o mediante el diseño de contratos inteligentes que contribuyan a la seguridad y la eficiencia en las transacciones de la compañía.

2. Análisis de la Demanda Aplicando de Métodos de Análisis Multivariado

2.1 Ingesta y Almacenamiento de Datos

2.1.1 Los Datos

Los datos a utilizar para el análisis fueron obtenidos de diversas fuentes. Se recopiló información del Sitio Web de la Municipalidad del Partido General Pueyrredón¹, de la plataforma de la empresa Pedidosya² y del sitio web del portal de clasificados inmobiliarios Zonaprop³. Del Área de Datos de la Municipalidad se obtuvo información propia del Censo Nacional del Año 2010 que se complementó con datos recopilados por el propio municipio.

Del Área de Datos del Partido General Pueyrredón se descargaron múltiples archivos de bases de datos con información respecto a distintos aspectos de la población de la ciudad. Los datos fueron confeccionados por el Gobierno Nacional como resultado del Censo realizado en el año 2010. La base cuenta con 898 registros los cuales representan fracciones censales, es decir, conjuntos poblacionales en los que se agrupó la población al momento de recopilar la información de los habitantes de la ciudad. Las bases de datos obtenidas son las que se detallan a continuación.

Censo-2010-educación: Contiene información del nivel educativo en cada conjunto poblacional.

Censo-2010-población: Contiene información respecto al parentesco de los habitantes en cada hogar, la nacionalidad, género y edad.

Censo-2010-vivienda: Contiene información respecto al tipo de vivienda en que habitan los ciudadanos, la cantidad de personas por vivienda y la tenencia o no de determinados bienes como heladeras, computadora, celular o telefonía fija en cada hogar.

Rural: Presenta información sobre el porcentaje de población urbana y rural dentro del conjunto poblacional.

Radios_censales_geojson: En este archivo con formato geojson se obtuvieron los radios censales. Cada conjunto poblacional se encuentra segmentado territorialmente en radios

¹ <https://datos.mardelplata.gob.ar/>

² <https://www.pedidosya.com.ar/>

³ <https://www.zonaprop.com.ar/>

censales. Estos radios están expresados como múltiples pares de coordenadas geográficas que conforman polígonos que delimitan el territorio en que se emplaza cada conjunto poblacional.

Playas: Del mismo sitio web se obtuvo información recopilada por el Municipio en relación a la localización de las playas de la ciudad. Estos datos se encuentran en el archivo `playas.csv`.

Respecto a la información referida a la oferta gastronómica de cada radio censal, esta se obtuvo ingresando al sitio web de la plataforma PedidosYa simulando la realización de pedidos para dentro de cada radio censal para conocer la oferta existente en cada uno de ellos. Finalmente, se incorporó al análisis el valor medio de venta del metro cuadrado de las propiedades publicadas en cada barrio de Mar del Plata obteniendo esta información del sitio Zonaprop.

De las bases de datos mencionadas precedentemente se trabajó con variables de diferente carácter. Algunas de ellas con carácter categórico binomial, multinomial y ordinal, así como variables cuantitativas discretas y continuas. Las variables se encuentran descriptas en su totalidad en el [Apéndice II](#).

2.1.2 Programas y Lenguajes Empleados para la Investigación

Para desarrollar la investigación se utilizaron varios programas y lenguajes de programación en las distintas etapas. A su vez, de estas herramientas se aprovecharon servicios y librerías específicos según fuera necesario. Las herramientas utilizadas fueron las siguientes:

Google Drive: Se utilizó para almacenar las bases de datos en un *datalake*.

Microsoft Excel: La planilla de Cálculo de Microsoft se utilizó para previsualizar y hacer una primera limpieza de datos.

Google Colaboratory: se utilizó la notebook de Google con lenguaje Python para realizar el proceso de limpieza de datos avanzado y llevar a cabo el proceso de manipulación de datos hasta la exportación de la nueva base de datos en formato `xlsx`. Las librerías empleadas fueron las siguientes:

Pandas: se empleó para cargar, limpiar y preprocesar los datos utilizados en el trabajo final. Facilitó la manipulación de conjuntos de datos tabulares y la realización de análisis exploratorios.

Numpy: Esencial para realizar cálculos numéricos avanzados en el proyecto. Se utilizó para operaciones matriciales y manipulación de datos numéricos.

Matplotlib se utilizó para crear visualizaciones y gráficos que ayudaron a presentar los resultados y hallazgos de manera efectiva en el trabajo final.

StandardScaler se aplicó para estandarizar los datos antes de entrenar los modelos de aprendizaje automático, garantizando así la adecuada preparación de los datos.

PCA se utilizó para reducir la dimensionalidad de los datos vinculados a la distancia a las playas.

Kmeans, implementado con Sklearn, se aplicó en el trabajo final para llevar a cabo análisis de clústeres en función de similitudes.

Missingpy: se usó para abordar y manejar de manera efectiva los valores faltantes en los conjuntos de datos utilizados en el proyecto, evitando así la pérdida de información crítica.

Nominatim: se utilizó para la geocodificación de direcciones y la obtención de coordenadas geográficas, lo que enriqueció los datos con información espacial en el trabajo final.

Point: se empleó en análisis geoespaciales, permitiendo la representación y manipulación de puntos en el espacio geográfico en el contexto del proyecto.

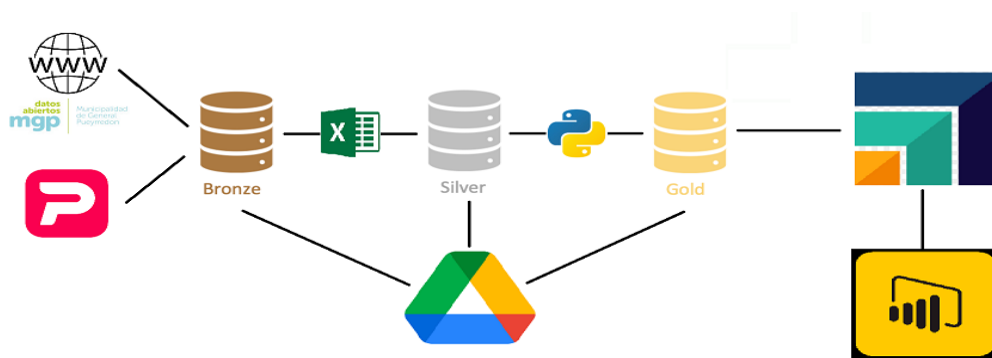
Rapidminer: software utilizado para implementar algoritmos de métodos de aprendizaje automático con carácter supervisado.

Microsoft Power BI: se empleó para realizar representaciones gráficas de aquellos datos considerados relevantes a lo largo de la investigación.

2.1.3 Almacenamiento

Los datos fueron almacenados en diferentes carpetas denominadas Bronze, Silver y Gold. Cada una con un propósito distinto, en la carpeta Bronze se encuentran las bases originales previo a sufrir cualquier tipo de modificación. En la carpeta Silver se almacenaron las bases posteriormente a ser modificadas utilizando Microsoft Excel. Finalmente, en la Carpeta Gold se almacenó la salida del proceso de manipulación de datos efectuado en la notebook de Google Colaboratory. A continuación se visualiza un esquema que representa lo que sería el proceso de extracción, transformación y carga de datos.

Gráfico 1 – Proceso ETL (Extracción, Transformación y Carga de Datos).



Elaboración propia

2.2 Manipulación, Transformación y Limpieza de Datos

Se comienza trabajando a partir de los datos de referencia geográfica de cada segmento poblacional importando la base de datos bajo el nombre *latlon*. Esta base de datos contiene información acerca de los límites territoriales en los que se ubica cada conjunto poblacional. Con esta información se pretende para cada fracción censal un único valor de latitud y longitud. El objetivo es obtener el centroide en que se ubica cada radio censal.

Para ello se calculó, para cada registro, la media de los valores de latitud y la media de los valores de longitud. Estos nuevos pares de datos representan el centroide de cada polígono geográfico expresado en coordenadas de latitud y longitud. Estos valores se incorporaron a la base de datos en las columnas denominadas *lat* y *lon* respectivamente.

En el gráfico se visualizan los puntos geográficos utilizando Microsoft Power BI.

Gráfico II – Representación de los diferentes segmentos poblacionales distinguiendo con distintos colores el nivel de oferta.

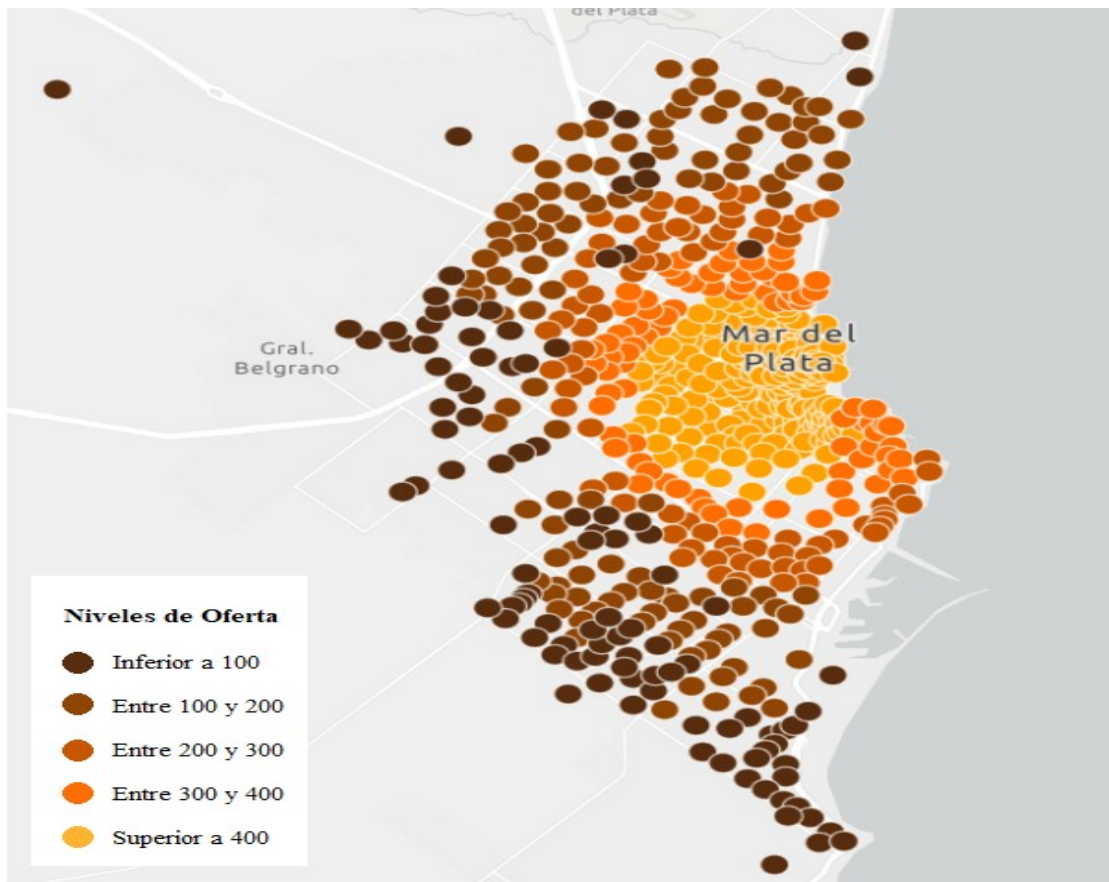


Gráfico de elaboración propia utilizando Microsoft Power BI

Se calculó el total de la población en cada segmento y se incorporó en una nueva columna denominada *pob_total*. Este valor se calculó sumando el total de individuos del total de columnas que muestran el rango etario por fracción censal. Esto representa el total de personas encuestadas durante el censo.

Se agrupó cada variable edad a partir de la teoría generacional de Pendergast (2008). Basado en las afirmaciones del autor se arribó a 6 segmentos poblacionales de acuerdo a su rango etario. Se agruparon como *gen_z* a los individuos menores de 7 años, en la variable *gen_x* a aquellos individuos de entre 8 y 28 años, como *gen_y* a la población de entre 29 y 49 años, como *gen_boomer* a aquellas personas de entre 50 y 67 años, en la variable *gen_silent* a quienes tienen entre 68 y 85 años y, finalmente, se agruparon las personas de más de 85 años en la variable *gen_greatest*. Posteriormente, las columnas que representaban la cantidad de individuos en cada segmento poblacional según la edad fueron eliminadas.

De los datos presentes en el dataframe denominado *rural* se extrajeron solamente los datos almacenados en la columna *r_porcrur*. Esta columna indica el porcentaje del territorio del radio

censal que pertenece a un área rural. Se creó una nueva base de datos denominada *data*. El objetivo es consolidar múltiples bases de datos en una única para continuar realizando análisis subsiguientes. La base *data* implica la unión de las bases hasta el momento denominadas *edu*, *pob*, *viv*, *latlon*, *rural* y *oferta*.

Se detectaron algunas variables que evidencian una colinealidad mutuamente inversa respecto de otra. Se prevé que al momento de aplicar un modelo predictivo en etapas posteriores esto brindaría información redundante que podría entorpecer los resultados a obtener. Se decidió remover una de cada par de estas variables relacionadas. Las variables eliminadas fueron: *leeyesc_si*, *heladera_si*, *computadora_si*, *celular_si* y *telefono_si*.

2.3 Aplicación de Métodos de Análisis Multivariado para Caracterizar la Demanda.

2.3.1 Método de Análisis de conglomerados para segmentar la población por grupo etario

El Análisis de conglomerados, al que también se denomina comúnmente Análisis de Clúster, es una técnica diseñada para clasificar distintas observaciones en grupos, de forma tal que cada grupo sea homogéneo respecto a las variables utilizadas para caracterizarlo. Por un lado, las observaciones contenidas en un grupo son parecidas entre sí. Por otro lado, los grupos son lo más distintos posible uno de otros respecto a las variables consideradas (Aldás y Uriel, 2017).

2.3.1.1 Definición de la Cantidad de Grupos. La función a utilizar para implementar el modelo es *KMeans* de la librería *Scikit-learn*. Previamente se comenzó normalizando las variables utilizando la función *StandardScaler* de la misma librería. Esto conlleva a que las variables pasen a tener una media igual a 0 y una dispersión de 1.

Para aplicar este algoritmo se define previamente un rango de valores K entendiendo a K como el número de *clusters* a definir. En este caso se seleccionó al azar un rango de entre 1 y 10 donde este último es el valor máximo de grupos a generar. Los resultados obtenidos se plasmaron en un gráfico en que se evidencia la relación entre el puntaje del modelo y la cantidad de grupos a realizar.

Gráfico III – Gráfico de Sedimentación utilizado para determinar el número óptimo de conglomerados.

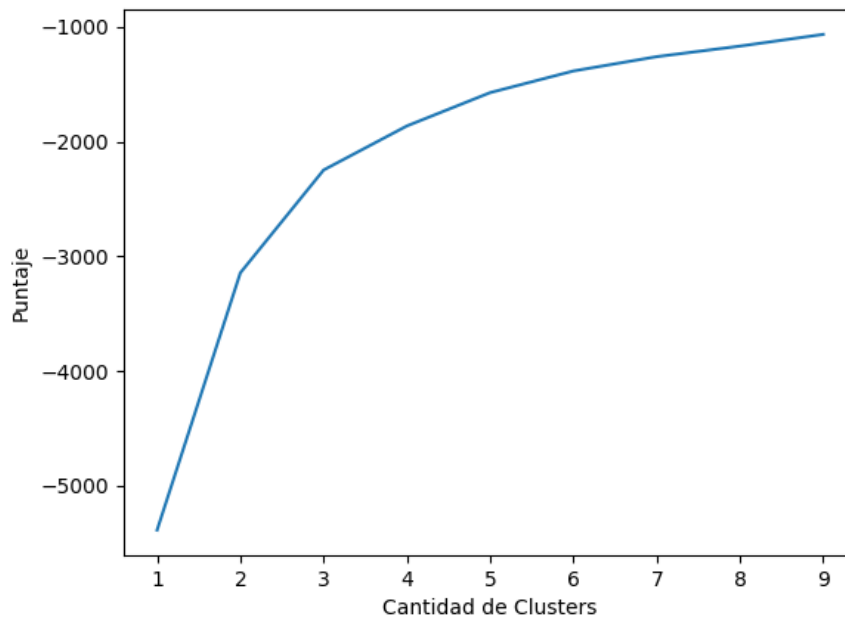


Gráfico de elaboración propia utilizando lenguaje Python en Google Colaboratoy

La curva representa la cantidad de grupos que propone el algoritmo para cada puntaje obtenido. La métrica utilizada para comparar resultados es la distancia media entre los puntos de datos y su centroide. El valor de la media disminuirá en la medida que aumenta el valor de k y cuando la media comienza a decrecer a tasa decreciente es donde se genera un punto de inflexión. El valor de K para el cual se genera esa inflexión en el gráfico indica la cantidad de clústers a definir. En este caso el K seleccionado es de 3.

2.3.1.2 Aplicación del Modelo. Se aplicó el modelo de forma que se agrupen las variables en 3 clústers. El modelo se implementó con una semilla con el valor 2022 para mantener constante los valores obtenidos. Se agregó una nueva variable en la base de datos denominada *cluster* la cual indica el grupo al que pertenece cada registro. El modelo de conglomerados se aplicó sobre la totalidad de los registros. Se observaron la cantidad de registros en cada grupo, la cantidad de individuos en cada generación por grupo, la mediana de individuos de cada generación, la proporción de individuos de cada generación dentro de cada Cluster y la proporción de individuos de cada generación entre los tres clústers.

2.3.1.3 Caracterización. Se caracterizaron entonces cada clúster a partir de los aspectos más destacados de cada uno. El cluster 0 es aquel que agrupa la mayor cantidad de segmentos poblacionales, sin embargo, se trata del segmento con la menor cantidad total de individuos. El Clúster 1 es aquel con la mayor cantidad de total individuos pero la cantidad de segmentos

ubicados dentro de esta categoría son minoría. Este Clúster concentra la mayor proporción de miembros de las generaciones más jóvenes, es decir, de la Generación Z, Generación Y y Generación X. El Clúster 2 se caracteriza por representar aquellos segmentos poblacionales con mayor proporción de individuos pertenecientes a las generaciones más adultas, es decir, de la generación Silent y Generación Greatest.

Según Ramírez Morla (2019) en “Análisis del uso de aplicaciones de delivery de alimentos mediante el modelo de aceptación de tecnología (TAM)” observa que el grupo de usuarios más jóvenes utiliza significativamente con mayor frecuencia las aplicaciones de delivery que el grupo de mayores. A su vez, se considera que el Clúster 1 agrupa aquellos segmentos con mayor cantidad de habitantes y con grupos etarios más jóvenes. Es por esto que los segmentos clasificados dentro del Cluster 1 son entonces aquellos a considerar como parte del Mercado Objetivo de los oferentes de comida a domicilio, principales aliados estratégicos de Hotkit.

2.3.2 *Aplicación de Método de Análisis de Componentes Principales (PCA) - para medir la distancia de cada barrio a las playas.*

Se considera importante para el análisis el contemplar la distancia existente desde el centro de cada radio censal respecto a la playa. Este aspecto se considera relevante desde el momento que Mar del Plata es una ciudad balnearia, siendo uno de sus principales atractivos su costa. Desde este punto de vista, se plantea la hipótesis de que es posible sintetizar la información relativa de la ubicación de cada radio censal respecto a las playas de la ciudad. Se pretende arribar a que cada segmento poblacional puede ser clasificado categóricamente como cercano o alejado de las playas.

Para medir esto se partió de la base de datos *playas* que contiene el id de cada playa, su nombre y la referencia geográfica. A partir de la librería *scikitlearn*, utilizando la distancia *Manhattan* se calcula la distancia existente desde el centroide de cada radio censal y hasta cada una de las playas de la ciudad. El resultado obtenido fue una nueva base de datos denominada *df_dist_playas* donde se almacenan tantos registros como conjuntos poblacionales y tantas variables como playas. Las variables son de carácter cuantitativo continuo mostrando valores que reflejan la distancia, utilizando kilómetros lineales como unidad de medida.

Se arribó a una base de datos con 898 filas y 107 columnas que resultan en 96.086 posibles distancias a cada una de las playas. Se considera necesario, con el objetivo de agilizar los tiempos computacionales y simplificar la interpretación de los datos, reducir la cantidad de

columnas de esta base de datos aplicando el Método de Análisis de Componentes Principales (PCA).

Utilizando la función *StandardScaler* y *PCA* de la librería *Scikit-Learn* se procedió a implementar la reducción de dimensionalidad planteada. La primera librería mencionada previamente se utilizó para estandarizar los datos y la segunda se utiliza para aplicar el método en sí que reduce la cantidad de variables. Previo a la aplicación del método se confeccionó una matriz de correlación que evidencie la existencia de correlación entre pares de variables.

En el gráfico a continuación se evidencia la correlación entre variables. Se observa que efectivamente existe correlación directa e inversa fuerte en la base de datos. Para aplicar el Método de Análisis de Componentes Principales es importante que al observar la matriz de correlación la misma no se asemeje a la matriz identidad. En este caso algunas variables cuentan con correlación positiva cercana con un coeficiente de 1 y correlación inversa parcialmente alta con un coeficiente de correlación de $-0,75$.

Gráfico IV – Matriz de Correlación de la distancia de la totalidad de los segmentos a las playas.

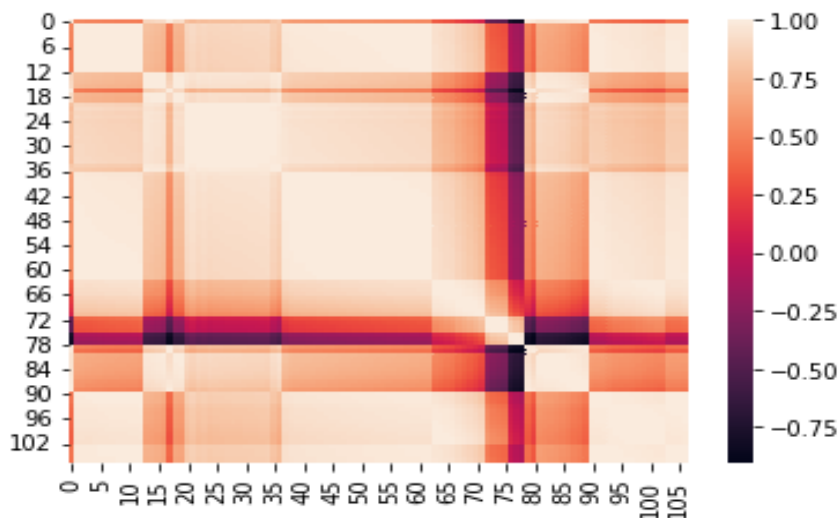


Gráfico de elaboración propia utilizando lenguaje Python en Google Colaboratoy

En principio se planteó la función *PCA* procurando que se mantenga una proporción de la varianza explicada cercana al 99% para ver la cantidad de componentes planteados por el modelo. El resultado obtenido fue una sintetización de las 107 columnas a solo 4.

La primera componente explica el 79,83% de la varianza total, la siguiente el 16,81%, la tercera un 1,56% y la cuarta apenas un 1,04%. Dado que conservando únicamente las dos primeras componentes se logra explicar un 96,64% de la varianza total se decide reducir aplicar el

método únicamente utilizando 2 componentes principales las cuales se incorporaron a la base de datos principal. Las nuevas columnas se agregaron con la denominación *PCI* y *PC2*.

Las variables *PCI* y *PC2* se establecieron como variables que reflejan el grado de cercanía a las playas. Se utilizó la distancia intercuartil como parámetro para segmentar aquellos registros con mayor o menor distancia a la playa. Si los valores de variable de *PCI* son menores a los del primer cuartil y/o si los valores de variable de *PC2* son mayor al tercer cuartil entonces se considera que el segmento es cercano a la playa. Aquellos segmentos poblacionales que no reúnen ninguna de las dos condiciones se los considera alejados de las playas. Para expresar este concepto en la base de datos principal, se agregó una nueva variable categórica denominada *cat_playa* considerando que aquellos registros que cumplen con el requisito de mayor cercanía a la playa pertenecen a la categoría A y el resto a la categoría B. A partir de esta segmentación se arribó a 351 registros categorizados como cercanos a la playa y otros 547 registros como aquellos alejados a las mismas. En análisis subsiguientes se hará especial foco en aquellos segmentos poblacionales con mayor cercanía a la playa, dado que se considera que están emplazados en sectores geográficos de carácter estratégicos de la ciudad.

2.3.3 Identificación de aquellos Segmentos Poblacionales con Mayor Poder Adquisitivo

2.3.3.1 Identificar el barrio de cada segmento poblacional. A partir de los datos de latitud y longitud se aplica un proceso de geocodificación inversa para obtener el domicilio más cercano al centroide de cada radio censal. Para esto se utilizó la librería de *Python* llamada *GeoPy*. La salida obtenida se descargó y almacenó como un nuevo *dataframe* denominado *dirNom*. El objeto *dirNom* es un *dataframe* de dos columnas donde la primera posee el código que representa cada segmento poblacional y el segundo refiere a la dirección que representa el centroide del radio censal.

Se incorpora como una nueva columna del *dataframe* *data* las direcciones almacenadas en *dirNom*. Se llevó a cabo un proceso de *tokenización*, es decir, se separaron los componentes de la cadena de texto de las direcciones utilizando la función *str.strip* de la librería *pandas*. Del resultado producto de la *tokenización* se conservaron únicamente los valores de las columnas *barrio0*, *barrio1* y *barrio2*, quitando todos aquellos valores que no representan barrios. Del total de 898 segmentos, hubo 8, para los cuales el barrio no pudo ser identificado. Aquellos barrios con una frecuencia de aparición menor a 2 fueron agrupados bajo el valor de variable *Otros*. Como resultado del proceso se logró identificar el barrio de 851 del total de segmentos poblacionales

2.3.3.2 Obtención del precio medio del metro cuadrado en cada barrio. A partir de los datos obtenidos de la página web de Zonaprop se confeccionó un *dataframe* en el que se incluyó como variable cada barrio. Con esta información se calculó la mediana del valor del metro cuadrado de cada barrio. El objetivo es comprender cuales segmentos poblacionales son aquellos con mayor poder adquisitivo utilizando como indicador el precio de venta del metro cuadrado de las propiedades.

Se comenzó realizando un proceso de manipulación de datos en el cual se corrigieron errores ortográficos o de denominación de los barrios. Aquellos barrios con una frecuencia inferior a 6 fueron reemplazados por la categoría *Otros*. Se efectuó un proceso de limpieza de datos en el que se quitaron aquellos registros con nombre de barrio *Mar del Plata*, aquellos que hacían referencia a venta de terrenos, aquellos con dimensiones inferiores a 15 m² o superiores a 3.000 m² y aquellos con precios incongruentes menores a \$10.000 dólares o superiores a \$2.000.000 dólares. Realizado esto, se calculó la mediana del precio del metro cuadrado de cada barrio y se incorporó en la base de datos principal como un nuevo dato en una columna denominada “m²”.

Se estableció que los segmentos con precio medio del m² superior a \$1.000 dólares pertenecen a aquellos barrios con mayor poder adquisitivo. Al entender que el consumo de alimentos por entrega a domicilio como un bien de lujo se asocia que este hábito se profundiza en aquellos de barrios con mayor poder adquisitivo. Se concluye que los barrios con valores medios del precio del metro cuadrado superiores a \$1.000 dólares son aquellos con susceptibles de atender al mercado objetivo de las aplicaciones de comida a domicilio.

2.3.4 Caracterización de la demanda gastronómica

A partir de los análisis realizados se determinó que aquellos segmentos con mayor demanda de productos gastronómicos bajo la modalidad de delivery son aquellos pertenecientes al Cluster 1, categorizados con una categoría de cercanía a las playas tipo A, es decir, aquellos cercanos a las playas, y que pertenecen a los barrios para los cuales el valor del metro cuadrado es superior a \$1.000 dólares. Al aplicar un filtro en Python sobre la base de datos actual se cuenta con un listado de 20 registros que son aquellos segmentos poblacionales con mayor propensión a demandar comida a domicilio por aplicaciones de delivery. Resulta posible afirmar que es posible identificar al menos 18 segmentos poblacionales con características propias de los grupos de individuos con tendencia alta de demanda de productos alimenticios bajo la modalidad de entrega a domicilio.

Tabla 1 – Tabla con los 20 segmentos poblacionales con características propias del mercado objetivo de las aplicaciones de entrega de comida a domicilio.

Código	cluster	cat_playas	preciom2	oferta
63570108	1	A	1753.846154	392.0
63570306	1	A	1904.761905	367.0
63571009	1	A	3651.980930	234.0
63571010	1	A	3651.980930	256.0
63571102	1	A	1904.761905	370.0
63571103	1	A	1753.846154	392.0
63571104	1	A	1753.846154	395.0
63571105	1	A	1753.846154	405.0
63571107	1	A	1598.750000	400.0
63571109	1	A	1598.750000	407.0
63571112	1	A	1753.846154	421.0
63571201	1	A	1753.846154	410.0
63571203	1	A	1598.750000	417.0
63571207	1	A	1753.846154	413.0
63571208	1	A	1753.846154	421.0
63571210	1	A	1753.846154	400.0
63571213	1	A	1753.846154	384.0
63571306	1	A	1115.317029	128.0
63571308	1	A	1115.317029	137.0
63571701	1	A	1904.761905	328.0

Tabla de elaboración propia en Google Colaboratory utilizando lenguaje Python.

Habiendo aplicado métodos de manipulación de datos en Python es posible hallar cuales registros tienen las características donde se concentran mayores niveles de demanda. Estos segmentos representan los puntos geográficos o localizaciones óptimas donde resulta en principio conveniente posicionar los centros de distribución de Hotkit desde una lectura de análisis de la demanda. En el apartado subsiguiente se buscará que sectores pudieran estar manifestando una oferta insatisfecha del servicio.

3. Análisis de la Oferta mediante la Implementación de un Modelo Predictivo

3.1 Diseño del Modelo

A partir del *dataframe* resultante del proceso de manipulación de datos realizado, se buscó implementar un modelo de aprendizaje automático que permita predecir la oferta gastronómica de cada radio censal. Como resultado de este modelo predictivo se espera contar con los niveles originales de oferta y aquellas predicciones realizadas a partir de la información en el *dataframe*. Las diferencias entre ambos valores serán objeto de análisis en esta sección para detectar localizaciones geográficas con oferta gastronómica insuficiente.

Antes de crear un modelo predictivo sobre el *dataframe* producto del proceso de Manipulación de Datos del apartado anterior, se eliminaron aquellos registros cuya oferta resultó nula. Esto se dio en aquellas locaciones para las cuales no fue posible obtener la oferta gastronómica. Del total de 898 registros se mantuvieron 602 registros para confeccionar el modelo. Se definieron las variables *Código* como *ID* y la variable *Oferta* como *Label*. La variable *ID* se utilizó como clave primaria para identificar a cada conjunto poblacional. Se entiende como *Label* a aquella variable objetivo a predecir.

La implementación de los modelos de aprendizaje automático supervisado se realizó en el software Rapidminer. Los métodos de aprendizaje supervisado son una categoría de técnicas en el campo del aprendizaje automático que involucra la construcción de modelos predictivos o descriptivos basados en la relación entre variables predictoras y una variable a predecir (*label*). Se trata de un proceso en el que el modelo busca capturar patrones y relaciones en los datos para poder hacer predicciones. Los métodos de aprendizaje supervisado pueden dividirse en dos categorías principales siendo estos de regresión o de clasificación. En este caso se trabajará con distintos modelos de regresión.

Los modelos de regresión son una categoría fundamental dentro de los métodos de análisis predictivo en estadística. Estos modelos se utilizan para comprender y cuantificar la relación entre una variable dependiente y una o más variables independientes. El objetivo principal es estimar el valor de la variable dependiente (*label*) en función de los valores de las variables predictoras.

Las variables se normalizaron utilizando puntuaciones *Z*. La estandarización de variables utilizando puntuaciones *Z* es un proceso en el que se transforman los valores de una variable para que tengan una media de 0 y una desviación estándar de 1. Esto se realizó en Rapidminer al inicio del Proceso utilizando el módulo denominado *Normalize*. Según afirma Field, A.

(2000), al estandarizar una variable utilizando puntuaciones Z, se logra que los valores tengan una interpretación más intuitiva y comparable, ya que están en una escala común. Esto es especialmente útil al comparar variables con diferentes unidades o escalas originales como es en el caso de la base de datos bajo análisis.

El método de evaluación el modelo a utilizar es el de validación cruzada (*Cross Validation*). La validación cruzada consiste en realizar múltiples particiones de los datos para luego estimar la métrica en cada partición y finalmente promediar los resultados (Santiago, 2022).

Gráfico V – Esquema del Modelo de Aprendizaje Automático implementado en Rapidminer.

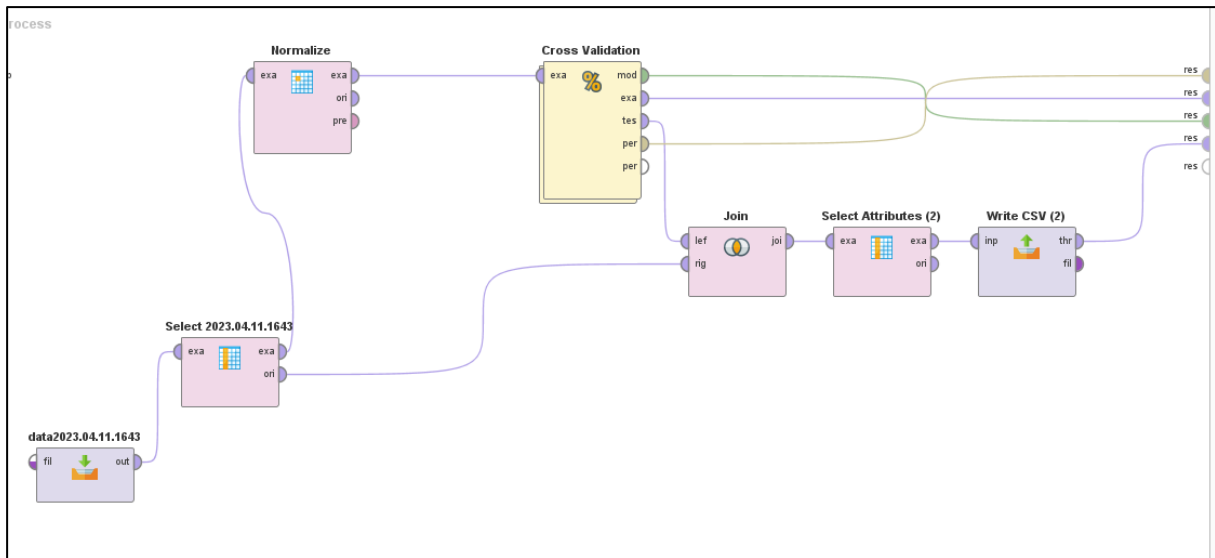


Imagen de elaboración propia

3.1.1 Clase de Modelos

En todos los casos se compararon los resultados tanto con parámetros por defecto como optimizando los mismos. Los modelos utilizados fueron *Gradient Boosted Trees*, *Decision Tree*, *Random Forest* y *Vote*. Este último se aplicó combinando los algoritmos mencionados precedentemente en forma conjunta tanto con sus parámetros estándar como optimizados.

3.1.1.1 *Gradient Boosted Trees (GBT)*. Los *Gradient Boosted Trees*, o Árboles Potenciados por Gradiente, son un conjunto de modelos de aprendizaje automático utilizados para la regresión y clasificación. Están basados en árboles de decisión, pero en lugar de construir un solo árbol, se construye una secuencia de árboles en la que cada uno se ajusta a los errores del modelo anterior. En cada paso, el modelo se enfoca en los errores que quedaron sin explicar por los árboles anteriores y ajusta un nuevo árbol para corregir esos errores. Luego, los resultados de todos los árboles se combinan para producir una predicción final. El proceso de construir y combinar estos árboles potenciados permite mejorar la precisión predictiva y es especialmente efectivo en conjuntos de datos complejos.

3.1.1.2 *Decision Tree*. Un *Decision Tree* o Árbol de Decisión es una estructura jerárquica en forma de árbol utilizada en aprendizaje automático y análisis de datos para tomar decisiones. Cada nodo interno del árbol representa una pregunta o una característica del conjunto de datos, mientras que las ramas representan las posibles respuestas o valores de esa característica. A medida que se desciende por el árbol, se toman decisiones en función de las respuestas a las preguntas, lo que finalmente lleva a una hoja que contiene la decisión o predicción. Los árboles de decisión son fáciles de interpretar y pueden manejar datos numéricos y categóricos. Es un aspecto desfavorable que tienden a ser susceptibles al sobreajuste en relación a otros modelos similares como el Random Forest.

3.1.1.3 *Random Forest*. Un *Random Forest* o Bosque Aleatorio es un conjunto de árboles de decisión independientes que se construyen utilizando muestras aleatorias del conjunto de datos original y técnicas de selección de características aleatorias. Cada árbol en el bosque opera de manera independiente y produce una predicción. El Bosque Aleatorio combina las predicciones de todos los árboles mediante votación para determinar la predicción final. Este enfoque ayuda a reducir el sobreajuste y aumenta la generalización del modelo, ya que se basa en la diversidad de árboles. Otro aspecto favorable es su robustez ante el ruido y los valores atípicos.

3.1.1.4 *Vote*. En el contexto de los modelos predictivos, la votación (*vote*) se refiere al proceso de combinar las decisiones o predicciones de varios modelos para llegar a una decisión final. En la votación mayoritaria, cada modelo emite su predicción y el criterio que recibe más votos se elige como la predicción final. Esto se utiliza en enfoques como los Bosques Aleatorios, donde varios árboles emiten sus predicciones y la clase más votada se selecciona. La votación es una técnica eficaz para aumentar la robustez y la precisión de los modelos, ya que aprovecha las fortalezas de múltiples modelos.

3.1.2 Performance para la selección

Para seleccionar el mejor modelo, es decir, aquel que arriba a resultados más precisos se tuvo en cuenta como indicador el Error Cuadrático Medio o RMSE de sus siglas en inglés *Root Mean Square Error*. El RMSE es una métrica comúnmente utilizada en análisis de datos y *machine learning* para evaluar qué tan bien se ajusta un modelo a los datos observados. En esencia, el RMSE es una medida de la diferencia entre los valores predichos por un modelo y los valores reales observados. Los valores posibles de RMSE son siempre superiores a 0.

A partir de esta definición entonces cuanto menor sea el valor obtenido del RMSE mayor será el grado de acierto del modelo predictivo aplicado. Se establecieron como aceptables a los fines de la investigación aquellos valores de RMSE inferiores a 30. En caso de haber más de un algoritmo con valores dentro de lo que se consideraría aceptable se seleccionará aquel más cercano a 0 para realizar la interpretación de resultados.

3.2 Proceso de Selección del Modelo

3.2.1 Evaluación de los Resultados Obtenidos

La tabla a continuación representa los resultados obtenidos a lo largo del proceso de implementación:

Tabla 2 – Cuadro resumen con los resultados obtenidos de la aplicación de distintos algoritmos con parámetros por defecto y optimizados.

Modelo	1. Cross Validation - RMSE Parámetros por defecto.	2. Cross Validation - RMSE Parámetros optimizados
Gradient Boosted Trees	98.627+/-5.547	31.414+/-8.899
Decision Tree	38.031+/-8.469	37.652+/-8.901
Random Forest	28.459+/-8.226	27.919+/-8.337
Vote	45.052+/-7.011	29.456+/-9.003

Tabla de elaboración propia

De entre los ocho resultados se obtuvieron tres algoritmos que alcanzaron valores de RMSE inferiores al establecido como objetivo. Estos son *Random Forest* con parámetros por defecto, *Random Forest* con parámetros optimizados y *Vote* con parámetros optimizados. Los tres modelos serían aquellos considerados aceptables a los fines de la investigación. El modelo

seleccionado es el de *Random Forest* con parámetros optimizados por ser aquel con el valor de RMSE más cercano a 0.

3.2.2 Elección del Modelo

El RMSE del modelo seleccionado tiene un puntaje de 27,919 superando los resultados del modelo *Random Forest* con parámetros por defecto y a la aplicación un modelo *Vote* optimizado el cual agrupaba a los tres modelos precedentes. Para obtener este resultado se seleccionó un algoritmo de *Random Forest*, se ajustaron la cantidad de árboles de 100 a 900 y se elevó la profundidad máxima de 10 a 16.

Se deja constancia de que el RMSE obtenido incluye una variabilidad de la predicción de +/- 8,337. Este término representa la dispersión o variabilidad en las predicciones. Esto significa que, en la mayoría de los casos, el error de predicción estará en el rango de 19,582 (27,919 - 8,337) a 36,256 (27,919 + 8,337) unidades.

A partir del modelo predictivo seleccionado se detectaron aquellas variables tenidas en cuenta por los oferentes gastronómicos de la ciudad de Mar del Plata. Las variables con mayor peso en el modelo predictivo fueron el grado de cercanía a la playa expresado en las variables PC1 y PC2, la cantidad de usuarios masculinos en cada segmento y las referencias geográficas de latitud y longitud.

3.3 Interpretación de Resultados

3.3.1 Comparación entre los resultados de la predicción y los valores originales

Habiendo evaluado diferentes modelos y habiendo seleccionado aquel con mejor *performance* resta analizar las diferencias entre las predicciones realizadas y los valores originales de la variable objetivo. Para determinar si la diferencia es significativa entre el valor real y el estimado se tomó como parámetro toda diferencia superior a la mitad del RMSE resultante. Dado que el RMSE obtenido es de 27,919, siempre que la diferencia entre los valores sea superior a 14 entonces se consideró una diferencia elevada.

Para medir la diferencia se exportaron las predicciones del modelo junto con una serie de variables en formato .csv. Las variables exportadas fueron *Código*, *lat*, *lon*, *prediction*, *oferta*, *cluster*, *cat_playas*, *barrios* y *Mercado Objetivo*. Se utilizó Microsoft Excel para observar aquellos registros con mayor diferencia. La tabla analizada es la presentada a continuación:

Tabla 3 – Tabla exportada de Rapidminer con información de los registros caracterizados como parte del Mercado Objetivo.

Código	lat	lon	prediction(oferta)	oferta	diferencia	cluster	cat_playas	barrios	Mercado Objetivo
63570108	-38.0117378	-57.541901	394	392	2	1	A	Vieja Terminal	1
63570306	-38.0100032	-57.536278	346	367	-21	1	A	Lomas de Stella Maris	1
63571009	-38.02775517	-57.53438017	271	234	37	1	A	Alem	1
63571010	-38.02764129	-57.53648986	260	256	4	1	A	Alem	1
63571102	-38.01402809	-57.53985555	376	370	6	1	A	Lomas de Stella Maris	1
63571103	-38.01343367	-57.5420055	388	392	-4	1	A	Vieja Terminal	1
63571104	-38.0129062	-57.5431228	396	395	1	1	A	Vieja Terminal	1
63571105	-38.0118638	-57.5449583	404	405	-1	1	A	Vieja Terminal	1
63571107	-38.0103378	-57.5424644	398	400	-2	1	A	Centro	1
63571109	-38.0094288	-57.5442626	406	407	-1	1	A	Centro	1
63571112	-38.00844771	-57.54833786	416	421	-5	1	A	Vieja Terminal	1
63571201	-38.010919	-57.546918	413	410	3	1	A	Vieja Terminal	1
63571203	-38.0079106	-57.5472612	415	417	-2	1	A	Centro	1
63571207	-38.0097136	-57.5459038	409	413	-4	1	A	Vieja Terminal	1
63571208	-38.0111324	-57.550141	412	421	-9	1	A	Vieja Terminal	1
63571210	-38.01313571	-57.54568414	404	400	4	1	A	Vieja Terminal	1
63571213	-38.01532082	-57.54230782	388	384	4	1	A	Vieja Terminal	1
63571306	-38.043022	-57.548053	134	128	6	1	A	Puerto	1
63571308	-38.039776	-57.54584971	171	137	34	1	A	Puerto	1
63571701	-38.0180055	-57.53265131	320	328	-8	1	A	Lomas de Stella Maris	1

Tabla de elaboración propia

Se incorporó como una nueva variable la diferencia entre la variable oferta y la variable predicción bajo el nombre *diferencia*. Luego se aplicó un filtro para observar aquellos valores que pertenecen al Mercado Objetivo definitivo en el apartado anterior. A partir de la variable denominada “diferencia” se observaron en la tabla aquellos segmentos donde la predicción resultó en una sobreestimación relativamente alta, aquellos en los que se produjo una subestimación relativamente alta y en tercer lugar aquellos casos en que se estimaron valores cercanos a los reales.

A los fines del presente trabajo se presta especial atención a aquellos registros cuyo valor de la variable de diferencia son superiores a 14 unidades. Esto sucede en los casos en que se predijeron niveles de oferta superiores a los reales. Sería factible pensar que desde una perspectiva del análisis cuantitativo la oferta gastronómica estaría siendo insuficiente en estas ubicaciones geográficas de la ciudad por lo que se trataría de segmentos del mercado con demanda insatisfecha.

Para visualizar los segmentos poblacionales según su ubicación geográfica, se utilizó, Microsoft Power BI para confeccionar un gráfico que plasma los puntos geográficos de la Tabla II en el mapa de la ciudad. El gráfico es el presentado a continuación:

Gráfico VI – Representación gráfica de la ubicación de los segmentos que comparten características propias del Mercado Objetivo.

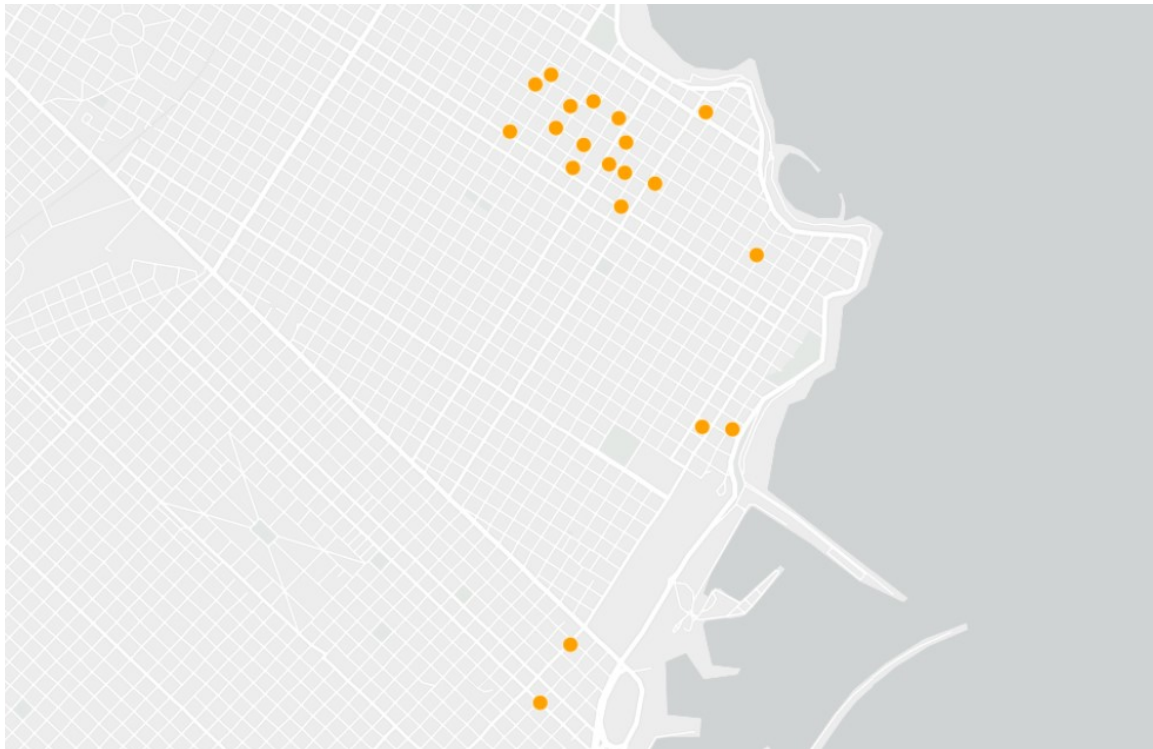


Gráfico de elaboración propia utilizando Microsoft Power BI

La mayoría de los puntos se concentran en el barrio denominado Vieja Terminal. En la parte sur del mapa se destacan 2 segmentos pertenecientes al barrio Alem y otros dos segmentos del barrio Puerto.

3.3.2 Localizaciones óptimas para la ubicación de las Cocinas Anfitrionas de Hotkit

Del listado de registros mencionados precedentemente se encontraron dos los cuales comparten características propias del Mercado Objetivo de acuerdo a lo establecido en el apartado 2 y para los cuales se estima que la oferta gastronómica en la zona debería ser superior a la real. Estos segmentos son los identificados con el Código 63571009 y 63571308, ubicados en los barrios Puerto y Alem respectivamente.

Gráfico VII – Representación gráfica de los segmentos que representan localizaciones óptimas para la ubicación de Cocinas anfitrionas en Mar del Plata

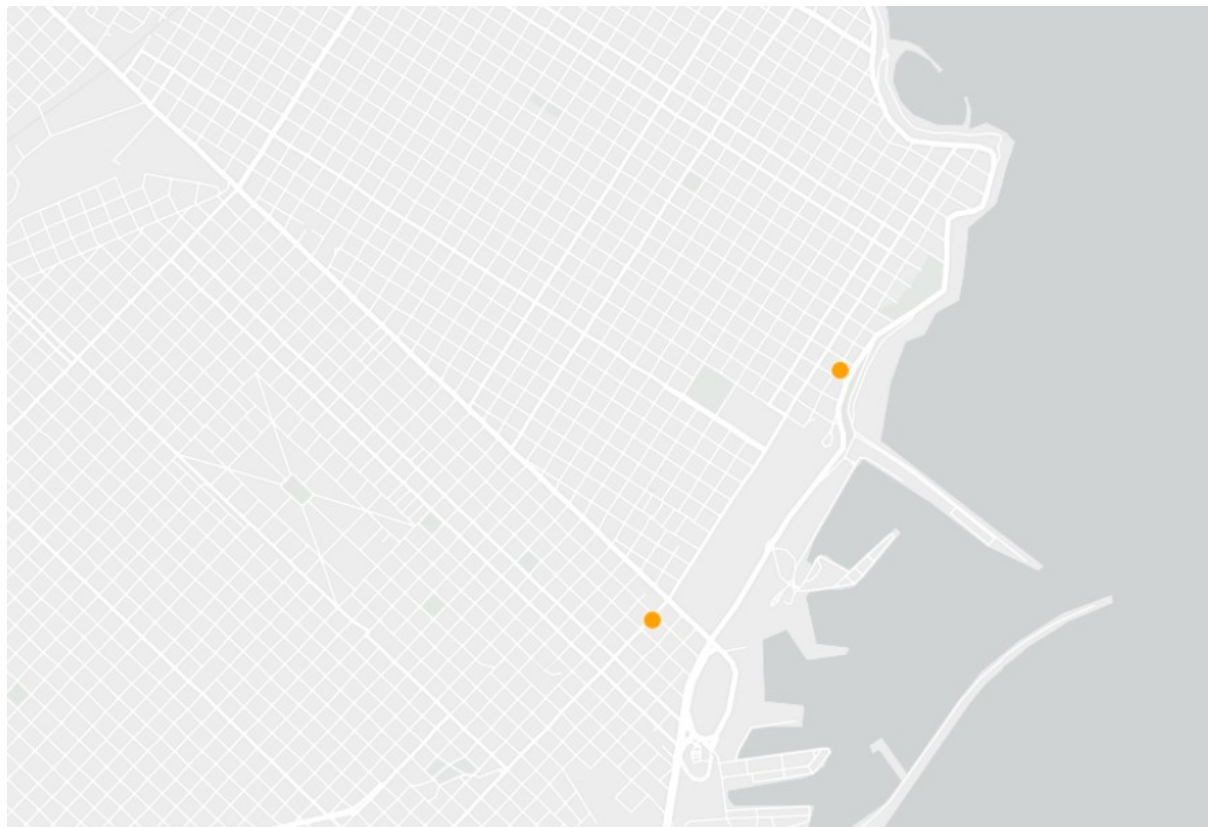


Gráfico de elaboración propia utilizando Microsoft Power BI

Ambos segmentos pertenecen al Cluster 1, forman parte de los barrios más costosos de la ciudad y se encuentran dentro de aquellos con mayor grado de cercanía a la playa. A partir del análisis realizado se determina que la ubicación geográfica donde se encuentran estos segmentos representan aquellas localizaciones óptimas para la ubicación de Cocinas Anfitrionas en la Ciudad de Mar del Plata.

4. Conclusión

Los objetivos de este trabajo pudieron ser abordados en tres apartados distintos. En el primer apartado se explicó el concepto de organizaciones de plataforma, se desarrolló el modelo de negocios de Hotkit y se destacó la importancia de emplear sistemas descentralizados de información. En el segundo apartado se efectuó un abordaje de la población bajo análisis en el que se describió el complejo proceso de ingesta, transformación y almacenamiento de datos de diferentes orígenes. El objetivo de este apartado fue manipular variables para segmentar aquellos registros con características propias del mercado objetivo de Hotkit haciendo uso de métodos de análisis multivariado. En el último apartado, se presentaron las etapas diseño, selección e interpretación de resultados de un modelo de aprendizaje automático que permita predecir la oferta gastronómica de cada segmento poblacional.

Resultó posible demostrar la posibilidad de adecuar un plan de negocios de una Cocina Anfitriona entendida como organización lineal hacia una con carácter de plataforma. Se confeccionó un modelo de negocios definiendo el segmento de clientes, la propuesta de valor, los canales de distribución, el contacto con clientes, los flujos de ingreso, los recursos clave, las actividades clave, los aliados estratégicos y la estructura de costos. Además se logró arribar a los fundamentos necesarios para proyectar la necesidad de aplicar un sistema de administración de datos con carácter descentralizado que garantice el resguardo de información sensible.

Analizando datos de origen público y privado se articuló una base de datos a partir del cual se llevó a cabo un análisis de la demanda del mercado de Hotkit. Para ello se logró segmentar la población de la Ciudad de Mar del Plata encontrando aquellos registros con características propias del mercado objetivo por medio del establecimiento de tres criterios claros. El primer criterio se estableció a partir del agrupamiento de la población en segmentos según el rango etario de los individuos utilizando un método de análisis de conglomerados. El segundo criterio consistió en aplicar el método de análisis de componentes principales para considerar si cada registro de la base de datos posee una cercanía alta o baja a las playas de la ciudad. El tercer criterio definido fue en base al poder adquisitivo del segmento poblacional. Este último criterio se midió en base al valor medio del precio del metro cuadrado del barrio en que se ubica cada fracción censal.

Posteriormente se diseñaron múltiples modelos de aprendizaje automático. De entre estos modelos se seleccionó uno capaz de predecir la oferta gastronómica de la ciudad con el menor

valor de error cuadrático medio (RMSE) posible dentro del rango establecido previamente como aceptable. El modelo seleccionado fue el de *Random Forest* con parámetros optimizados que arrojó un valor de RMSE de 27.919+/-8.337. Como resultado de la implementación de este modelo de aprendizaje automático se observaron los registros para los cuales se estima que exista una considerable demanda insatisfecha de servicios de entrega de comida a domicilio. De esos segmentos poblacionales, aquellos identificados con el código 63571009 y 63571308 reúnen características propias del mercado objetivo de Hotkit. Estos dos registros sirven de prueba para afirmar que resulta posible estimar localizaciones óptimas aplicando métodos de análisis cuantitativo para la ubicación de cocinas anfitrionas en la ciudad de Mar del Plata.

El desarrollo de este trabajo permitió alcanzar conclusiones satisfactorias a los fines de las hipótesis planteadas, sin embargo, aún se consideran ciertos aspectos para continuar estudiando en el futuro. Las bases de datos de origen público que se utilizaron para el análisis podrían actualizarse por unas pertenecientes a periodos de tiempo más recientes. Esto permitiría comprender con mayor precisión el grado de pertenencia de cada segmento poblacional al mercado objetivo de Hotkit.

Asimismo se contempla la posibilidad de incorporar en investigaciones subsiguientes un desglose mayor de las características de la oferta gastronómica. En este trabajo se analizó la misma sin distinción del tipo de producto alimenticio ofrecido. Se considera la posibilidad de realizar investigaciones venideras distinguiendo el carácter de los platos que venden los restaurantes ponderando, en términos de estrategia de marketing, la diferenciación del servicio ofrecido. Esto ampliará las dimensiones del estudio pudiéndose crear modelos predictivos distintos para la oferta de cada tipo de producto.

5. Referencias Bibliográficas

Romero, D. (24 de febrero de 2021). Only delivery: ¿es factible este modelo de negocio en el sector gastronómico? ESAN Graduate School of Business.

<https://www.esan.edu.pe/conexion/actualidad/2021/02/24/only-delivery-es-factible-este-modelo-de-negocio-en-el-sector-gastronomico/>

López, B. (2007). Publicidad Emocional. Estrategias creativas. Madrid: ESIC.

Sebastián-Morillas, A.; Martín-Soladana, I.; Clemente-Mediavilla, J. (2019) Importancia de los insights en el proceso estratégico y creativo de las campañas publicitarias.

<https://doi.org/10.5209/esmp.66570>

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT sloan management review, 52(2).

<https://sloanreview.mit.edu/article/big-dataanalytics-and-the-path-from-insights-to-value/Ley>

Kotler, P. y Keller, K. (2006). Dirección de Marketing. (12^o Edición). Pearson Education

Osterwalder, A., & Pigneur, Y. (2010). Business model generation: A handbook for visionaries, game changers, and challengers (Vol. 1). John Wiley & Sons.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). Platform revolution: How networked markets are transforming the economy and how to make them work for you. WW Norton & Company.

Youtie, J., Porter, A. L., & Huang, Y. (2017). Early social science research about Big Data. Science and Public Policy, 44(1), 65-74. <https://doi.org/10.1093/scipol/scw021>.

Moxie. (2022). My first impressions of web3. ABOUT LIFE PROJECTS BLOG. <https://moxie.org/2022/01/07/web3-first-impressions.html>

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. Journal of Information Technology, 30(1), 75-89.

<https://doi.org/10.1057/jit.2015.5>

Varian, H. R. (2014). Beyond big data. Business Economics, 49(1), 27-31.

<https://doi.org/10.1057/be.2014.1>

Ramirez Morla A. (2019). Análisis del uso de aplicaciones de delivery de alimentos mediante el modelo de aceptación de tecnología (TAM).

Field, A. (2020). Discovering Statistics Using IBM SPSS Statistics. Editorial.

Santiago, F. (2022). Implementación de modelos de aprendizaje automático, Capítulo 6, Modelado, 62.

6. Anexos

ANEXO I - Censo-2010-educación

(https://drive.google.com/file/d/1VGVgfLyvbUm0Dn5IsjWBHoZgXR2PXUC8/view?usp=drive_link)

ANEXO II - Censo-2010-población

(https://drive.google.com/file/d/1Q17BJ1247bND7u3aV3Mybp9OmPTbw0DL/view?usp=drive_link)

ANEXO III - Censo-2010-vivienda (https://drive.google.com/file/d/1DO_p8fy14NeL-JFPHM8j-5FSff0JcXzc/view?usp=drive_link)

ANEXO IV - Rural (https://docs.google.com/spreadsheets/d/1yXgemAYJ49tPHxO-gd1pb-CjRIGAIUzMGpWLOecjO8/edit?usp=drive_link)

ANEXO V - Radios_censales_geojson (https://drive.google.com/file/d/1htai-hlLthKwyph1pjQfTTb301roat52/view?usp=drive_link)

ANEXO VI - Playas (https://docs.google.com/spreadsheets/d/1K5rJkfmJ2ZOJ-TFzfBxcYMkOm7IMrc42/edit?usp=drive_link&oid=108420908985304856568&rtpof=true&sd=true)

ANEXO VII – Barrios

(https://docs.google.com/spreadsheets/d/12_4CVeF7IL6k5xjXRurWZxMXxGapBLDFfwArQbeA2g8/edit?usp=drive_link)

7. Apéndices

APÉNDICE I – SCRIPT DE PYTHON UTILIZADO EN GOOGLE COLABORATORY.

APÉNDICE II –ANÁLISIS DESCRIPTIVO DE LAS VARIABLES.

APÉNDICE III – BASE DE DATOS UTILIZADA PARA EL ANÁLISIS PREDICTIVO.

APÉNDICE IV – MODELO EN RAPIDMINER.

APÉNDICE V – TABLA RESULTANTE DE APLICACIÓN DEL MODELO EN RAPIDMINER

APÉNDICE VI – GRÁFICOS REALIZADOS EN POWER BI.