

# Trabajo Final

- Grupo 08 -

---

# HENRY

- Edith Cuellar
- Leandro Ibarra
- Matias Baez
- Tinmar Andrade

# Índice

¿Quiénes somos? .....	2
El equipo.....	3
¿Quién, cuándo, dónde y por qué?.....	4
El proyecto .....	5
Alcances.....	5
Objetivos.....	6
KPI's.....	7
Cronograma de trabajo.....	7
Tecnologías y aplicaciones.....	9
Arquitectura .....	9
Bitácora.....	11
Github .....	11
EDA preliminar.....	11
EDA Yelp.....	12
EDA Google maps.....	13
EDA profundo .....	14
Análisis .....	15
Locales seleccionados para evaluación .....	17
ETL.....	18
Diagrama entidad-relación.....	19
Diccionario de datos DER .....	21
Dashboard .....	24
Modelo ML.....	24
Características técnicas .....	26

## ¿Quiénes somos?



**Sinergia Global** somos una consultora líder en el ámbito del análisis de datos y estrategias empresariales. Asistimos a nuestros clientes en el proceso de extraer el valor máximo de sus datos, proporcionando las herramientas esenciales que les permitan tomar decisiones empresariales informadas y efectivas.

En un entorno empresarial cada vez más competitivo, es crucial la toma de decisiones basada en datos. Somos un equipo de profesionales especializados en el análisis de datos de manera integral, ofreciendo a nuestros clientes una visión profunda y significativa de sus operaciones.

En Sinergia Global trabajamos estrechamente con nuestros clientes para identificar las áreas clave donde el análisis de datos puede generar un impacto significativo. Ya sea optimizando procesos internos, mejorando la eficiencia operativa o identificando oportunidades de crecimiento, estamos comprometidos a proporcionar soluciones personalizadas y estratégicas.

En Sinergia Global, entendemos que cada cliente es único y, por lo tanto, adaptamos nuestras soluciones a las necesidades específicas de su empresa. Creemos que la verdadera sinergia se logra cuando combinamos la experiencia técnica con la comprensión profunda del negocio, creando así un impacto positivo en el rendimiento empresarial.

Ya sea que busquen optimizar sus operaciones actuales, desarrollar nuevas estrategias de mercado o abordar desafíos específicos, en Sinergia Global estamos comprometidos a ser su socio confiable en el viaje hacia el éxito empresarial.

## El equipo

En el marco de este proyecto específico, hemos conformado un equipo compuesto por cinco individuos capacitados y comprometidos. Esta sinergia de talento y experiencia se ha propuesto con el propósito de avanzar de manera efectiva hacia la obtención de los objetivos establecidos.

Es importante destacar que, tras la conclusión del segundo sprint, lamentablemente, uno de los integrantes ha decidido abandonar el proyecto por razones personales. A lo largo de su participación, este miembro ha realizado valiosos aportes que han sido fundamentales para el análisis y la definición del rumbo estratégico del proyecto.

La diversidad de habilidades y perspectivas dentro del grupo ha permitido abordar de manera integral los desafíos planteados, aportando soluciones innovadoras y eficientes. Aunque la partida de un miembro representa un cambio en la dinámica del grupo, estamos comprometidos a mantener la excelencia en la ejecución del proyecto. Este desafío inesperado nos brinda la oportunidad de ajustar nuestra estrategia, redistribuir responsabilidades y fortalecer la colaboración entre los miembros restantes.

La adaptabilidad y la capacidad de superar obstáculos son características fundamentales de un equipo exitoso. Aprovecharemos esta experiencia para aprender y crecer, asegurándonos de mantenernos enfocados en nuestros objetivos.

Edith Cuellar  
Bocanegra

Leandro Martín  
Ibarra

Matias Emanuel  
Baez

Nicolas Agustín  
Ibarra

Tinmar Armando  
Andrade Banda



## ¿Quién, cuándo, dónde y por qué?



**Culinary Crossroads** es un grupo empresarial destacado; productos, lugares y servicios que definen el panorama culinario de Indiana y ayudan a hacer de Indiana el lugar donde la gente quiere vivir y trabajar.

Es posible que no hayan pensado en Indiana como un lugar con excelente comida, chefs estrella de rock, productos especiales galardonados y destinos culinarios excepcionales. Pero Culinary Crossroads ayuda a las personas a ver Indiana de una manera que tal vez ni siquiera los residentes más antiguos la hayan visto antes. Después de todo, están en la encrucijada culinaria de Estados Unidos.

A través de Culinary Crossroads se está desarrollando una cultura culinaria más sólida en todo el estado con sus restaurantes, chefs, cocineros competitivos y proveedores que definen la comida de Indiana.

Cuando Culinary Crossroads se pone en contacto con nuestro equipo, no proporciona muchos detalles preliminares. Es durante nuestras reuniones oficiales que comenzamos a comprender la magnitud del desafío que implica esta asociación. Culinary Crossroads, un grupo empresarial ya consolidado en Indiana, con un historial exitoso que incluye diversas cadenas de restaurantes, ha recurrido a Sinergia Global en busca de asesoramiento estratégico. El objetivo es ampliar su cartera de inversiones a otro estado, y están buscando nuestra experiencia para garantizar una inversión exitosa y rentable en este nuevo territorio.

## El proyecto

Nos esforzaremos por identificar oportunidades de crecimiento, analizar la viabilidad de nuevas ubicaciones, y optimizar las operaciones para garantizar el éxito y su presencia en el nuevo mercado

### Alcances

Alcance temporal: el proyecto se llevará a cabo en un período de tres sprints, cada uno con una duración de dos semanas. Este enfoque ágil permitirá una ejecución eficiente y la posibilidad de realizar ajustes en función de los hallazgos y resultados obtenidos en cada sprint.

Alcance geográfico: la focalización del proyecto se centrará en los estados unidos, con un énfasis especial en identificar el estado que se establezca como la opción más viable para la inversión. Esta determinación se basará en análisis exhaustivos que incluirán factores económicos, demográficos y de mercado específicos del estado en consideración.

Alcance de datos: el conjunto de datos a utilizar será proporcionado por henry y abarcará revisiones y comentarios recopilados de plataformas como google maps y yelp. Además, se integrarán datos poblacionales de los estados unidos obtenidos de fuentes confiables mediante apis y archivos csv.

Metodología de análisis: se emplearán técnicas avanzadas de análisis de datos, incluyendo análisis de sentimiento de comentarios, correlación entre opiniones y modelos predictivos si es necesario.

Se realizará un análisis comparativo entre estados para identificar oportunidades de inversión, teniendo en cuenta no solo las tendencias de opiniones sino también factores económicos y poblacionales.

Colaboración continua: se fomentará la colaboración continua con el equipo de culinary crossroads para asegurar que sus objetivos y expectativas estén completamente alineados con los resultados del análisis.

## Objetivos

**Identificar oportunidades estratégicas para nuevas inversiones.** Realizar un análisis de mercado exhaustivo que incluya la evaluación de tendencias y preferencias del consumidor. Incorporar datos poblacionales para identificar áreas con potencial de crecimiento significativo.

Resultado Esperado: Identificar la ubicación ideal para invertir y destacar los locales más atractivos dentro de esa ubicación.

**Analizar datos de usuarios para optimizar el sistema de recomendación.** Aplicar técnicas avanzadas de procesamiento de lenguaje natural para comprender la retroalimentación de los usuarios. Segmentar locales basados en características específicas y preferencias de los usuarios.

Resultado Esperado: Mejorar la precisión y relevancia del sistema de recomendación, proporcionando sugerencias personalizadas en función del menú y la ubicación.

**Implementación del Sistema de Recomendación.** Desarrollar e implementar el sistema de recomendación especificado por el cliente, teniendo en cuenta los requisitos adicionales, como asegurar que la primera recomendación siempre pertenezcan a la cadena del cliente.

Resultado Esperado: Despliegue exitoso del sistema de recomendación que cumple con las expectativas del cliente y mejora la experiencia del usuario.

*Estos objetivos proporcionan una guía clara para el proyecto, abordando las metas clave relacionadas con la identificación de oportunidades de inversión, la mejora del sistema de recomendación y la implementación exitosa del mismo, con especial atención a los requisitos específicos del cliente.*

## KPI's

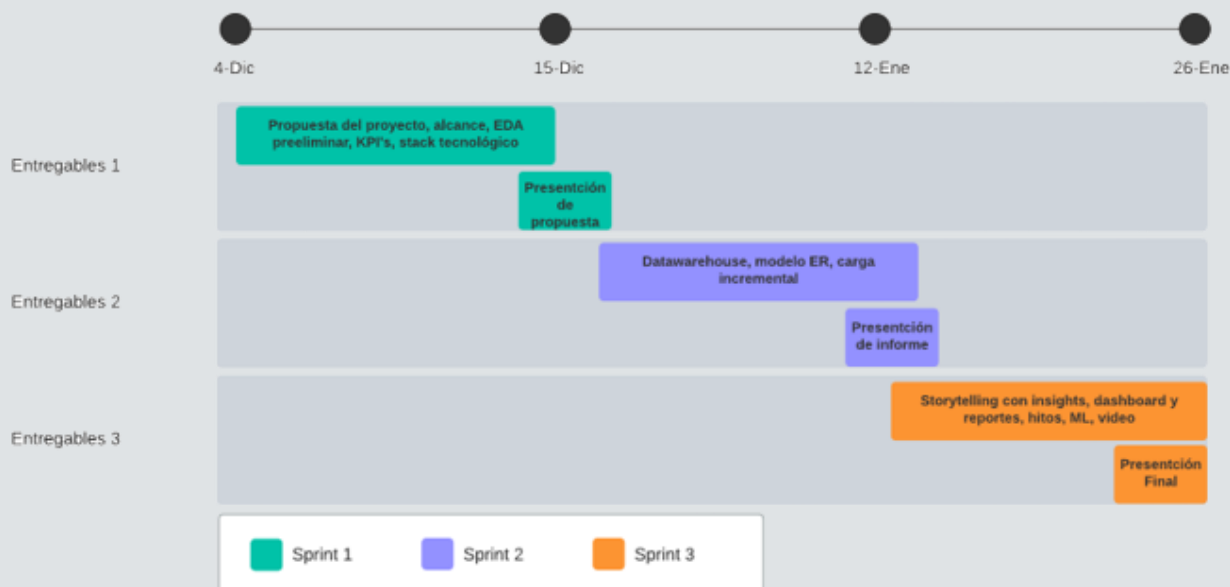
Variación trimestral de rating	Seguimiento del progreso trimestral de las calificaciones de estrellas otorgadas por los clientes a cada local. Permite comprender las variaciones y tendencias a lo largo del año.
Variación trimestral de reseñas	Cálculo de la variación porcentual en el número total de reseñas recibidas por cada local de un trimestre a otro. Permite entender el cambio en la retroalimentación de los clientes.
Reseñas positivas	Clasificación y análisis de las reseñas según su tono para cada local de una cadena, permitiendo identificar áreas de mejora y puntos fuertes específicos. La medida es la variación de un año a otro para evaluar su evolución
Reseñas Negativas	Clasificación y análisis de las reseñas según su tono para cada local de una cadena, permitiendo identificar áreas de mejora y puntos fuertes específicos. La medida es la variación de un año a otro para evaluar su evolución
Presencia	Ratio entre el número de locales de las categorías seleccionadas y el total de locales, esto permite ver la presencia y crecimiento del sector y la recepción del público
Calificación vs competencia	Ratio entre las calificaciones de las categorías seleccionadas y de la competencia permite observar el grado de satisfacción que ofrece el local vs la competencia. Al ponerle objetivo de tiempo podemos ver como evoluciona el grado de satisfacción que ofrecen nuestros locales respecto a la competencia.

## Cronograma de trabajo

El proyecto se extenderá por un período total de seis semanas, organizado en tres sprints de dos semanas cada uno. Esta distribución permite un enfoque ágil y la capacidad de adaptarse a los hallazgos y cambios durante el desarrollo del proyecto.

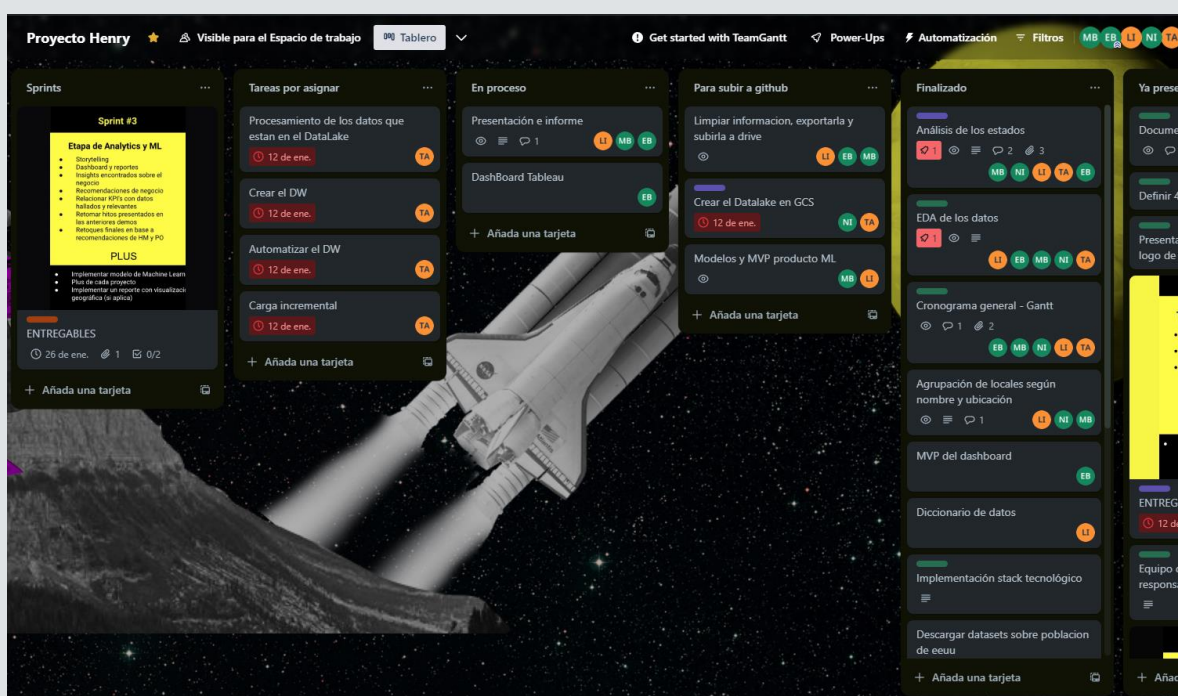
Cada día, los miembros del equipo llevarán a cabo sus tareas asignadas utilizando la plataforma de gestión de proyectos Trello. Al final de cada jornada, se realizarán reuniones diarias para compartir avances, discutir cualquier problema identificado y acordar nuevas tareas.





Se han programado cuatro reuniones con el mentor de Henry, en cada sprint del proyecto. Estas reuniones proporcionarán oportunidades regulares para discutir avances, obtener orientación y abordar cualquier problemática que pueda surgir durante el desarrollo del proyecto.

Al finalizar cada sprint, se llevará a cabo una presentación detallada con el cliente. Durante estas sesiones, se mostrarán los avances logrados hasta la fecha, se discutirán los resultados obtenidos y se establecerán expectativas para el próximo sprint.



## Tecnologías y aplicaciones

Google meet, slack y discord: facilitando la colaboración y comunicación fluida entre los miembros del equipo.

Trello: para gestionar el proyecto y flujo de trabajo, así como supervisar las tareas.

Python: como lenguaje de programación principal para el desarrollo y la manipulación de datos.

Apache airflow: para la automatización y gestión de flujos de trabajo, permitiendo una programación eficiente de tareas y procesos.

Google cloud storage: para el almacenamiento seguro y escalable de datos en la nube.

Google bigquery: como una solución de almacenamiento y análisis de datos a gran escala.

Databricks: como plataforma unificada para el procesamiento y análisis de datos a gran escala, facilitando la colaboración y ejecución de proyectos de datos complejos.

Tableau: como herramienta de visualización de datos para presentar información de manera intuitiva y comprensible.



## Arquitectura

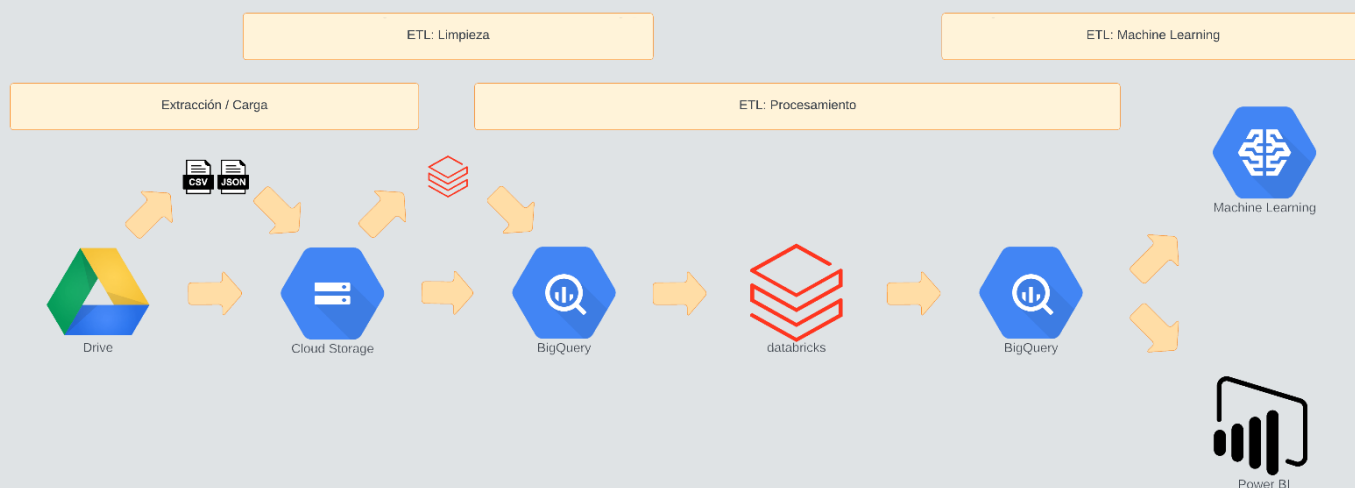
Comenzaremos con información almacenada en drive, la cual abarca formatos como JSON, CSV y Parquet. Para gestionar todo este conjunto de datos, los llevaremos al entorno de Google Cloud Storage.

Con Databricks, realizaremos un proceso de limpieza y preparación de datos. Utilizaremos sus potentes herramientas y entorno de ejecución para realizar tareas como la identificación y manejo de valores nulos, la normalización de datos y la corrección de posibles inconsistencias.

Una vez que los datos estén pulidos y listos, los transferiremos a BigQuery, un confiable espacio de almacenamiento y análisis a gran escala en la nube. BigQuery ofrece también la capacidad de realizar consultas SQL complejas y eficientes en conjuntos de datos masivos.

Retomaremos Databricks para emprender un proceso de ETL profundo y exhaustivo. Aquí la transformación de datos de será más avanzada, aplicando algoritmos y técnicas específicas para extraer y formatear los datos. Esta fase proporcionará una capa adicional de refinamiento a nuestros datos antes de su traslado a nuestra base de datos en BigQuery.

Los datos estarán listos para ser sometidos a un análisis más detallado y específico. Nuestro destino final será tableau, una herramienta de visualización de datos potente y versátil. Aquí, utilizaremos las capacidades gráficas y analíticas para crear informes comprensibles, que brindarán una representación clara y accesible de los insights extraídos de nuestros datos.



## Bitácora

### Github

Con el objetivo de centralizar y facilitar la gestión de avances del proyecto, hemos establecido un repositorio dedicado en GitHub. Este espacio digital actuará como el núcleo central donde se registrarán y almacenarán todas las actualizaciones, cambios y el progreso general del desarrollo.

El repositorio en GitHub también será el lugar designado para publicar los resultados finales del proyecto. Esto incluirá informes, presentaciones, visualizaciones y cualquier otro material relevante generado durante el desarrollo.

La adopción de GitHub como plataforma central para la gestión de avances promueve la colaboración eficiente y la transparencia en todo el equipo. Cada contribución se documenta de manera sistemática, lo que facilita la revisión y la coordinación entre los miembros del equipo y cualquier parte interesada en el proyecto.



### EDA preliminar

Para realizar un análisis preliminar exhaustivo de los datos, emprendimos, de forma independiente, un análisis de las fuentes de datos de Yelp y Google, cada uno con sus propias particularidades.

En el caso de los datos de Yelp, optamos por un enfoque manual utilizando Python y bibliotecas especializadas. Este análisis nos permitió abordar problemas de duplicación de información, asegurándonos de seleccionar los datos pertinentes para nuestra evaluación.

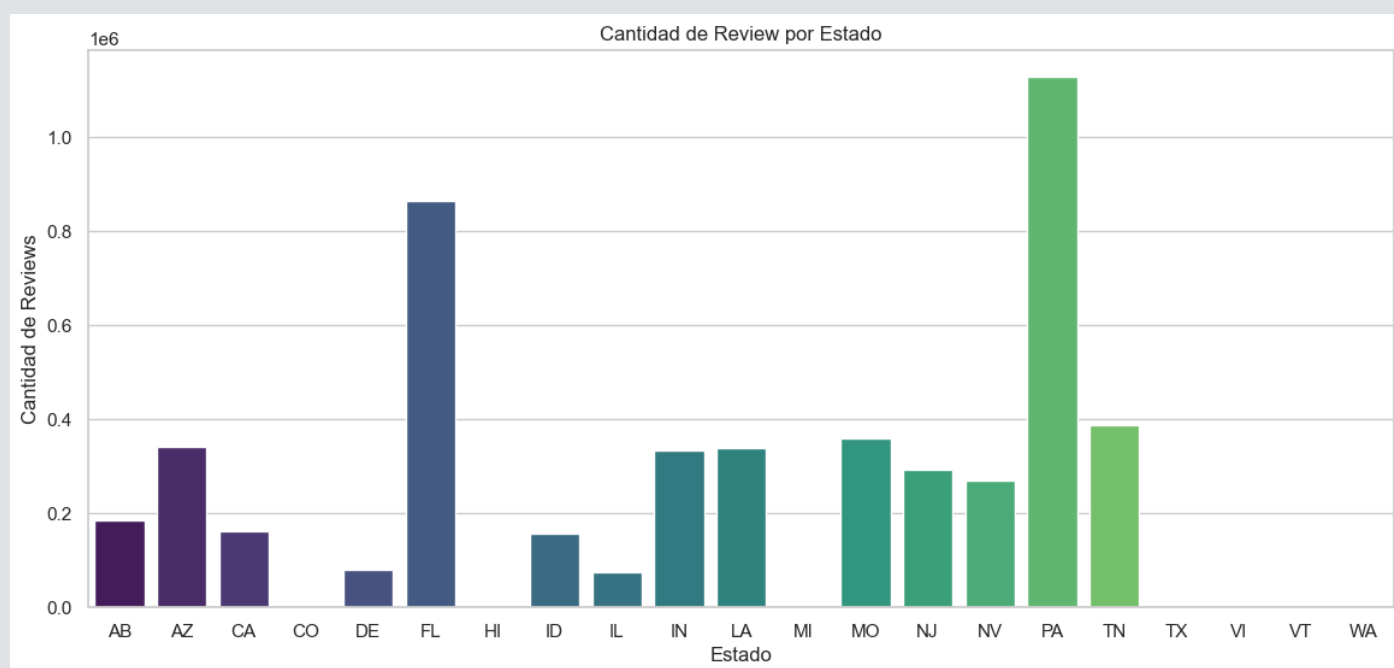
En cuanto a los datos de Google, nos enfrentamos a un conjunto organizado por estados, lo que generaba un gran número de archivos para gestionar. Decidimos implementar un Análisis Exploratorio de Datos (EDA) automático. Inicialmente, lo aplicamos a un estado en particular y encontramos que los datos estaban prácticamente limpios y estructurados de manera eficiente.

Con esta validación, extendimos el mismo proceso al resto de los estados, asegurándonos de mantener la consistencia en todo el conjunto de datos.

Adicionalmente, llevamos a cabo una investigación detallada sobre la calidad de los datos disponibles, teniendo en cuenta la presencia de valores nulos, duplicados y evaluando cómo se trataban estos datos respectivamente.

## EDA Yelp

Los datos de locales presentaban duplicidades en las columnas, aunque no en los registros. Notamos que Florida poseía la segunda mayor cantidad de registros en la base de datos. Las categorías de los restaurantes abarcaban un amplio espectro, desde "Acupuncture" hasta "Electronics", lo que nos llevó a realizar un exhaustivo filtrado de datos para focalizarnos únicamente en aquellos pertenecientes al sector gastronómico.



Los datos de usuarios poseían duplicados, contabilizando un total de 117.700 registros duplicados. Después de realizar una limpieza, eliminamos estos duplicados, dejándonos con un conjunto final de 1.987.897 registros únicos. Que luego fueron filtrados para coincidir con los locales pertenecientes a Florida.

En una fase posterior de análisis, nos enfocamos en evaluar el comportamiento de los usuarios, específicamente en sus revisiones de locales. Identificamos los cinco usuarios que realizaron la mayor cantidad de revisiones y notamos que, en promedio, otorgaron una calificación de 3.69 estrellas a los locales.

Usuarios con mayor cantidad de reviews realizadas

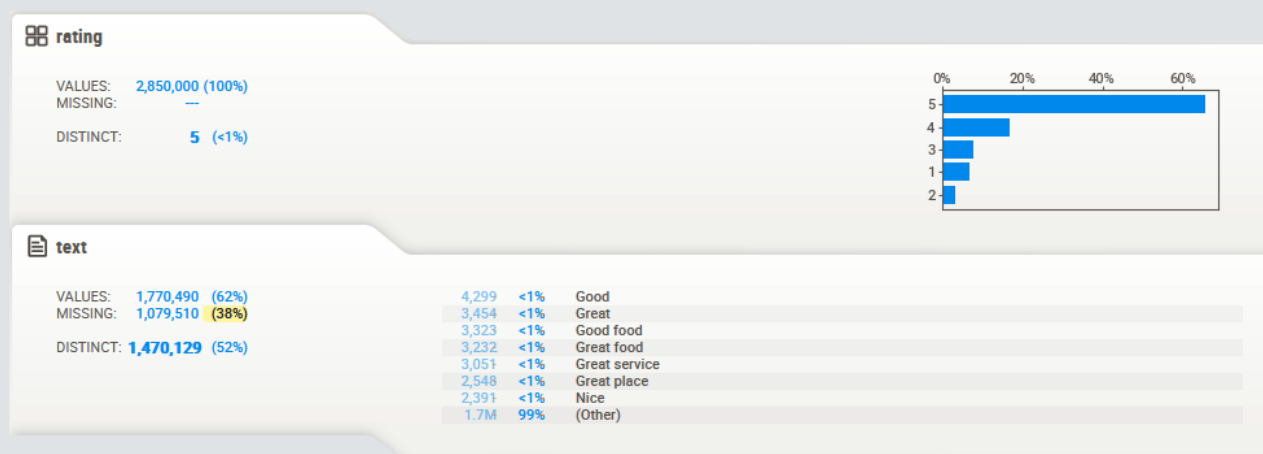
	user_id	name	review_count	yelping_since	useful	funny	cool	elite
17169	Hi10sGSZNxQH3NlyWSZ1oA	Fox	17473	2009-05-26 11:33:58	206296	185823	195814	2014,2015,2016,2017,2018,2019,20,20,2021
795339	8k3aO-mPeyhbR5HUucA5aA	Victor	16978	2007-12-08 14:56:45	152979	84992	125516	2008,2009,2010,2011,2012,2013,2014,2015,2016,2...
4723	hWDybu_KvYLSdEFzGmiTw	Bruce	16567	2009-03-08 21:47:44	173089	98459	144849	2010,2011,2012,2013,2014,2015,2016,2017,2018,2...
606393	RtGqdDBvvBCjcu5dUqwfzA	Shila	12868	2010-10-17 06:35:06	11109	3747	3292	2012,2013,2014,2015,2016,2017,2018,2019,20,20,...
152	P5bUL3Engv-2z6kKohB6qQ	Kim	9941	2006-05-31 21:27:42	27549	10772	21375	2006,2007,2008,2009,2010,2011,2012,2013,2014,2...

5 rows × 22 columns

## EDA Google maps

Los datos obtenidos de Google, en realidad, se encontraban notablemente limpios en comparación con los datos de Yelp.

Al realizar un análisis detallado de las revisiones de cada estado, observamos patrones consistentes en las valoraciones otorgadas. Aproximadamente el 60% de las revisiones recibieron una calificación de 5 estrellas, seguido por 4 estrellas, luego 3 estrellas, y finalmente, las valoraciones de 1 y 2 estrellas, siendo estas últimas las menos frecuentes. Además, notamos que alrededor del 41% de las revisiones carecían de una descripción. Esta observación destaca la tendencia de muchos usuarios a proporcionar únicamente una calificación numérica sin agregar comentarios adicionales.



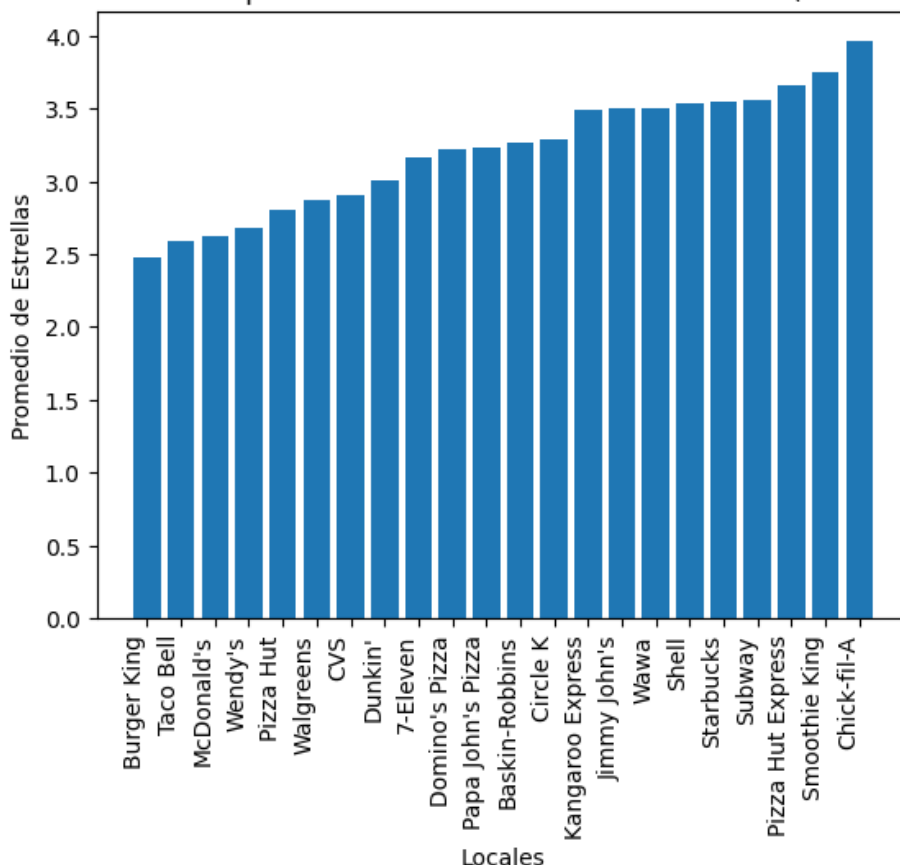
## EDA profundo

Nos enfocarnos exclusivamente en los datos de Florida y realizamos un análisis detallado de los restaurantes en el estado. Resulta evidente que las cadenas dominan el panorama gastronómico, siendo Subway la líder con 532 sucursales, seguida por Circle K (264), Starbucks (251), McDonald's (200) y Dunkin' (177). Esta concentración de sucursales sugiere una presencia significativa de cadenas reconocidas en el mercado local.

De los restaurantes analizados, observamos que el 85.5% se clasifica como "abierto normalmente", mientras que menos del 1% se encuentra "cerrado temporalmente" y el 13.6% está catalogado como "cerrado permanentemente". Estas cifras proporcionan una visión general del estado operativo de los locales en Florida.

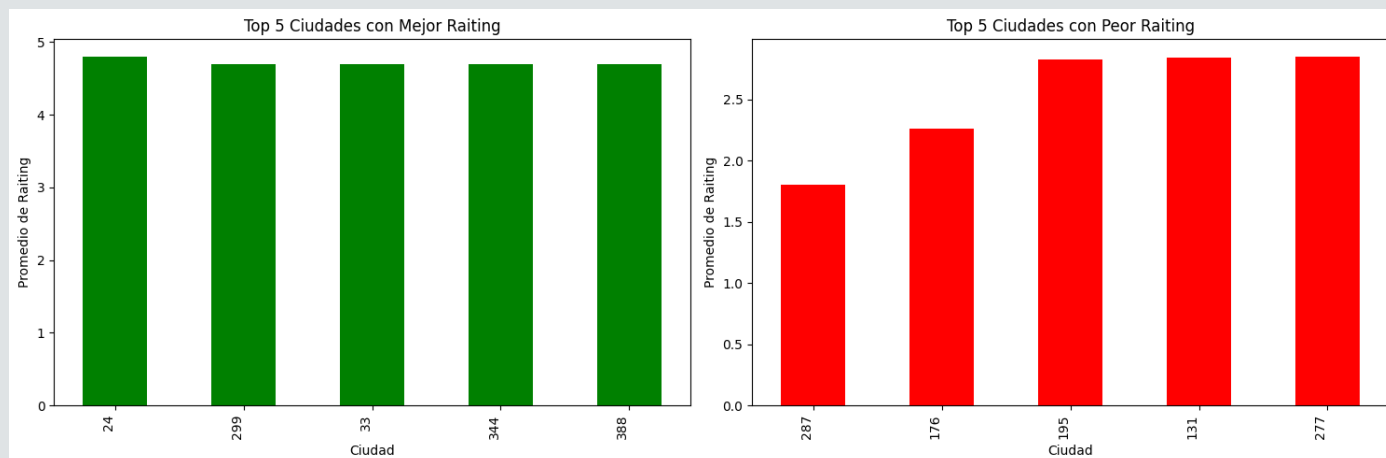
Al examinar las calificaciones, notamos que las cinco puntuaciones más comunes son 4.0, 4.5, 3.5, 3.0 y 5.0. Sin embargo, surge una observación interesante al analizar los locales con más de 50 sucursales, ya que tienden a estar por debajo del promedio de 4 estrellas. Este patrón sugiere que, a medida que la cadena de sucursales se expande, la calidad percibida por los usuarios tiende a disminuir.

Promedio de Estrellas para Locales con más de 50 Sucursales (Orden Ascendente)





Identificamos las cinco ciudades con las mejores y peores puntuaciones, proporcionando información valiosa sobre la percepción general de los locales en diferentes áreas geográficas de Florida.



24 – Belleair	287 – Otter Creek
299 – Palmetto Bay	176 – Jennings
33 – Bonifay	195 – Lake Hamilton
344 – Shalimar	131 – Greensboro
388 – Welaka	277 – Ocean Ridge

En el proceso de recomendación para nuestro cliente, nos centramos en sugerir sucursales que no solo acumulan muchas revisiones, sino que también mantienen altos promedios de calificaciones. Además, consideramos estratégico sugerir locales con una cantidad moderada de sucursales, ya que, como observamos, la calidad percibida tiende a disminuir a medida que la presencia de la cadena se expande. Esta estrategia busca asegurar que las recomendaciones no solo estén respaldadas por la cantidad de revisiones, sino también por la calidad consistente de la experiencia del cliente.

## Análisis

Se realizan análisis que abarcan consideraciones poblacionales, económicas y turísticas de cada estado con el fin de evaluar la mejor opción de inversión. Se concluye que Florida destaca como la elección más prometedora y equilibrada en términos de cualidades estratégicas para el proyecto y la expansión de inversiones.

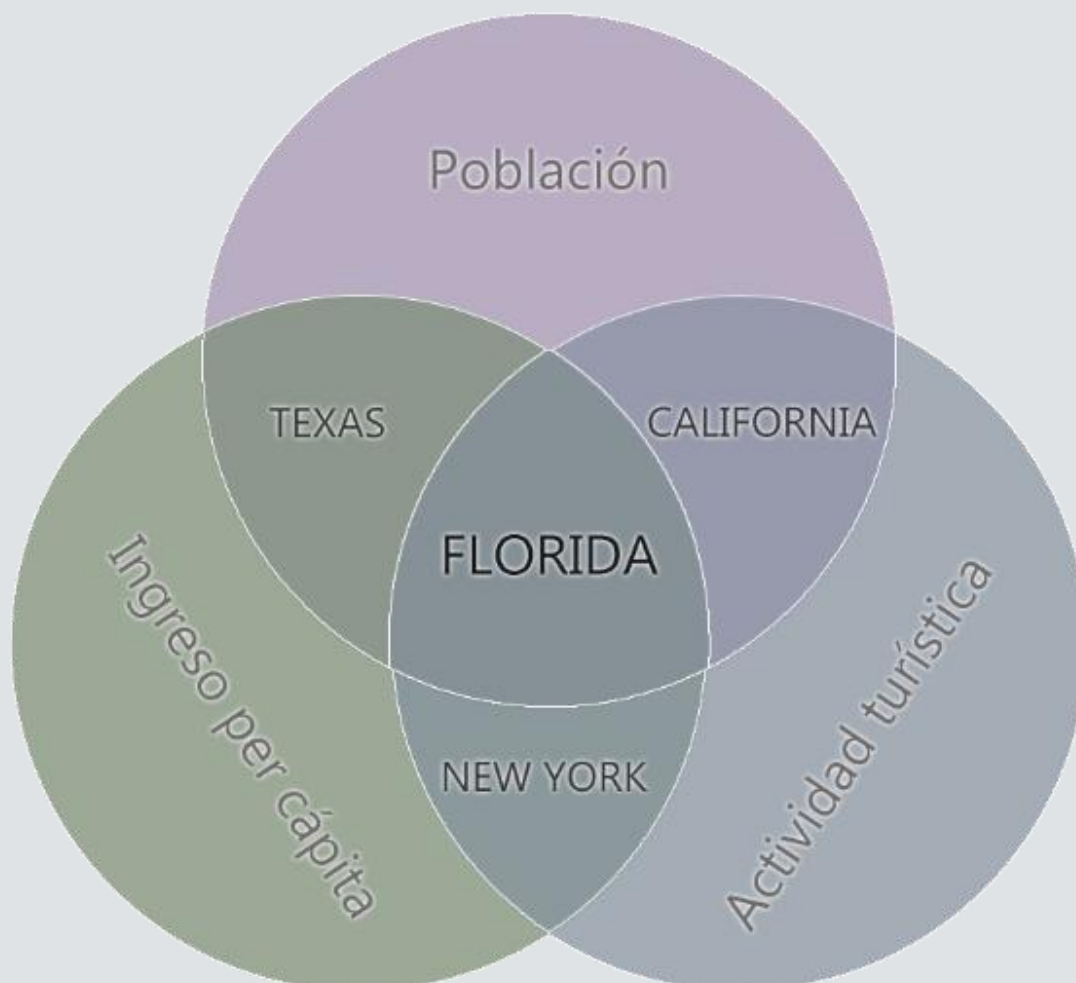


	id_estado	estado	avg_pop	avg_pop_grow	per_capita_income
19	6	California	39029342	-0.597204	63133
10	48	Texas	30029572	1.380412	55734
39	12	Florida	22244823	1.554003	55859
37	36	New York	19677151	-1.191607	60451
24	42	Pennsylvania	12972008	-0.067751	55624

Se considera la dimensión poblacional, se valora la diversidad y el dinamismo demográfico de Florida, lo que proporciona un mercado amplio y variado.

Se examinan las condiciones económicas, donde Florida se destaca por su robusta economía, sectores de crecimiento constante y una infraestructura empresarial favorable.

En el ámbito turístico, Florida se establece como un destino turístico de renombre mundial, con ciudades como Miami y Orlando atrayendo a millones de visitantes anualmente.



Se llevó a cabo un análisis exhaustivo de restaurantes para proporcionar al cliente opciones de inversión en diferentes localidades de Florida. Se consideraron características que podrían influir en el éxito potencial de la inversión, abordando distintos aspectos del mercado de restaurantes.

- Se identificaron restaurantes que a lo largo del tiempo han mantenido un rating alto, ofreciendo al cliente opciones consolidadas y reconocidas por su calidad.
- Se analizaron restaurantes que han experimentado un crecimiento constante en los últimos años, representando oportunidades de inversión en mercados en expansión y con demanda en aumento.
- Se consideraron establecimientos que han demostrado un rendimiento excepcional a pesar de no contar con sucursales, brindando al cliente la oportunidad de respaldar la expansión de negocios exitosos.

## Locales seleccionados para evaluación

**Kobe Japanese Steakhouse.** Reconocido por mantener un rating alto a lo largo del tiempo.

Sucursales		
Dirección	Estrellas	Opiniones
2773 66th St N, St. Petersburg	4,5	91
468 FL-436, Altamonte Springs	4,7	5.548
2210 W Brandon Blvd, Bradenton	4,5	2.425

**Bolay.** Reconocido por mantener un rating alto a lo largo del tiempo.

Sucursales		
Dirección	Estrellas	Opiniones
1330 4th St N, St. Petersburg	4,5	25
7880 SW 104th St #101, Miami	4,7	98
5030 Champion Blvd G1D #G1D, Boca Raton	4,6	748
1880 Okeechobee Blvd Suite A Suite A, West Palm Beach	4,6	951
250 S State Rd 7 Suite 100 Suite 100, Royal Palm Beach	4,6	1161

**Joe's Stone Crab Restaurant.** Sin sucursales, pero siempre un alto rating. Ha experimentado un crecimiento constante de valoraciones positivas en los últimos años.

Sucursales		
Dirección	Estrellas	Opiniones
11 Washington Ave, Miami Beach	4,6	6.388

**Marie Patties.** Sin sucursales, pero con rendimiento excepcional.

Sucursales		
Dirección	Estrellas	Opiniones
9493 SW 160th St, Palmetto Bay	4,9	68

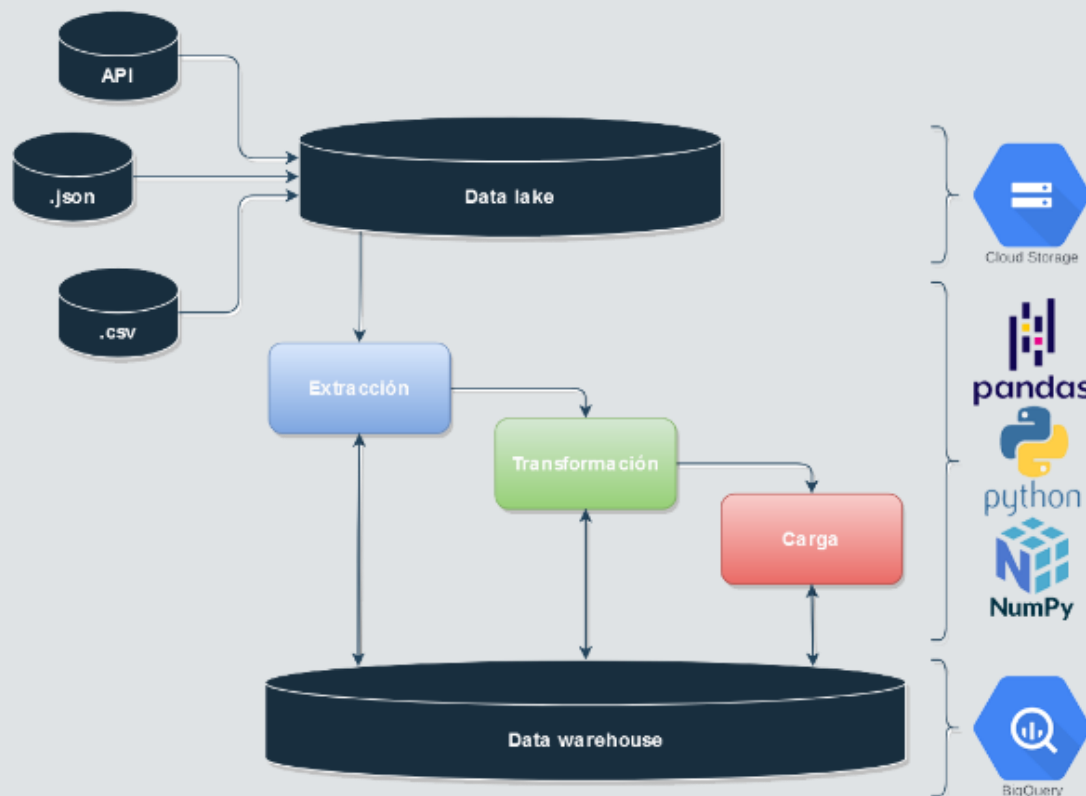
*Este análisis ha revelado una variedad de oportunidades de inversión en restaurantes, desde opciones consolidadas hasta estrategias de adquisición y expansión. Se recomienda al cliente considerar cuidadosamente cada opción en función de sus objetivos y estrategias de inversión, aprovechando las fortalezas de cada local para maximizar el potencial de retorno de la inversión.*

## ETL

En la fase inicial de nuestro proyecto, llevamos a cabo la extracción de datos desde diversas fuentes, incluyendo conjuntos de datos de Yelp, Google Maps y datos poblacionales del gobierno de los Estados Unidos. Esta recopilación nos proporcionó una base sólida para nuestro análisis.

A continuación, nos sumergimos en la aplicación de transformaciones esenciales para satisfacer los requisitos específicos de nuestro cliente. Este proceso incluyó una limpieza meticulosa, eliminando datos irrelevantes, nulos, duplicados y valores faltantes. Además, implementamos filtros precisos para focalizarnos exclusivamente en los datos relacionados con el sector gastronómico, asegurándonos de que estos también estuvieran vinculados al estado de Florida, conforme a las necesidades del proyecto.

Una vez completada la fase de transformación y depuración, procedimos a cargar estos datos transformados en la nube, optando por Google Cloud Storage como nuestro repositorio preferido. Esta elección garantiza la seguridad y accesibilidad de los datos, así como también sienta las bases para las etapas posteriores de nuestro proyecto.



## Diagrama entidad-relación

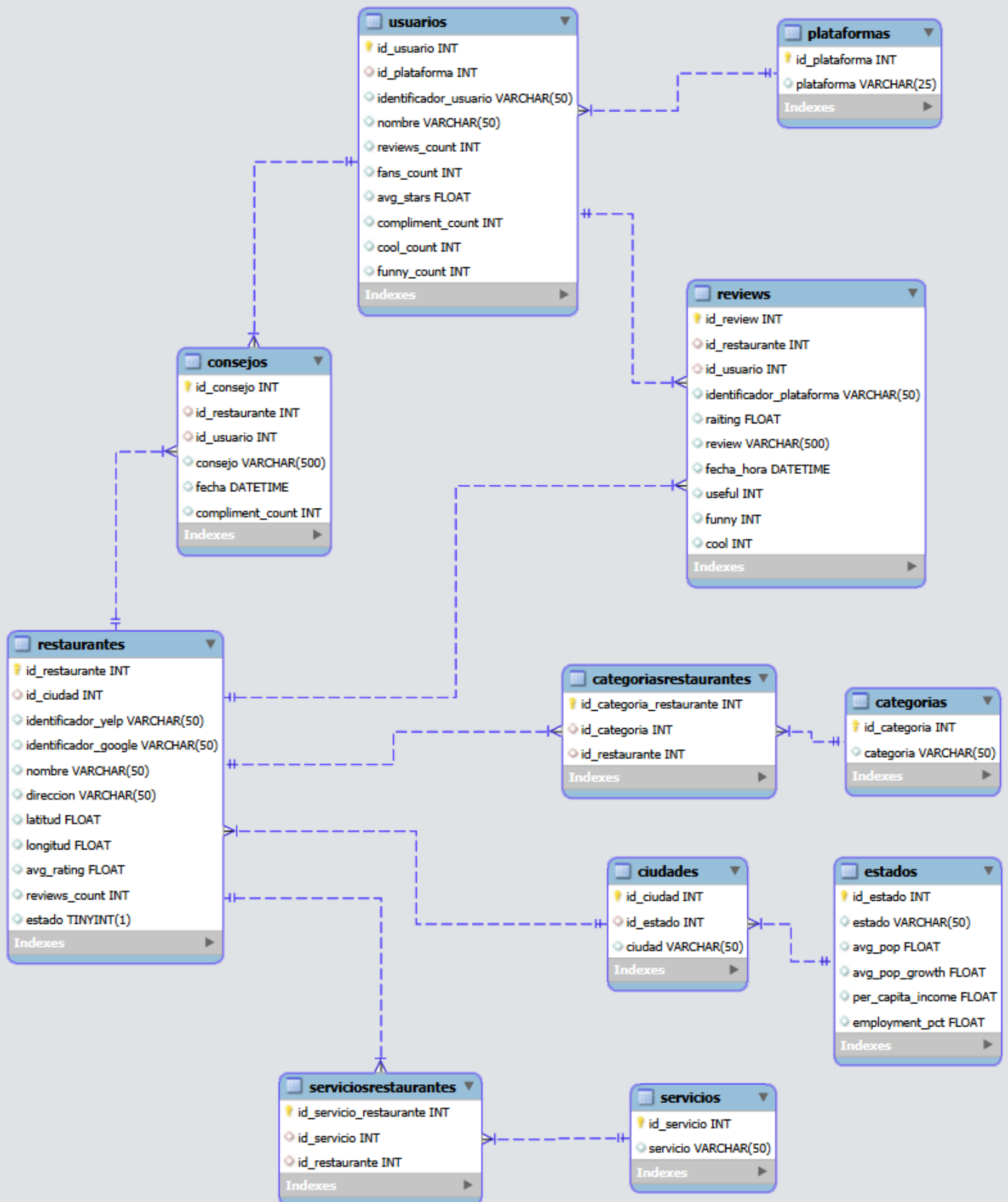
Este diagrama consta de dos tablas de hechos principales, denominadas 'consejos' y 'reviews', junto con otras nueve tablas de dimensiones que proporcionan contextos detallados y enriquecen la comprensión de los datos.

En el proceso de diseño, hemos incorporado identificación de entidades, atributos y relaciones, proporcionando una representación completa de la estructura de datos.

Además, hemos especificado las claves primarias (PK) para cada tabla, asegurando la unicidad y la identificación única de cada registro. También hemos definido claves foráneas (FK) para establecer relaciones coherentes entre las tablas, permitiendo referencias cruzadas.

En esta oportunidad elegimos un modelo de diagrama llamado StarFLake.

Es un esquema que combina elementos de esquemas de estrella y copo de nieve. Adopta la desnormalización de jerarquías dimensionales de los esquemas de estrella y la normalización de dimensiones compartidas mediante estabilizadores, aprovechando así los beneficios de ambos en términos de eficiencia y redundancia reducida.



## Diccionario de datos DER

### plataformas

id\_plataforma (PK)

- plataforma: A que plataforma corresponde (google / yelp)

### categorías

id\_categoria (PK)

- categoría: Categoría con la cual se identifica el comercio.

### servicios

id\_servicio (PK)

- servicio: Servicios que brinda el comercio.

### estados

id\_estado (PK)

- estado: Nombre del estado (USA)
- avg\_pop: Población del estado.
- avg\_pop\_growth: Índice de crecimiento de la población con respecto a años anteriores.
- per\_capita\_income: Ingresos de los habitantes por cabeza.

### ciudades

id\_ciudad (PK)

id\_estado (FK)

ciudad: Nombre de la ciudad (USA)

**usuarios**

id\_usuario (PK)

id\_plataforma (FK)

- identificador\_usuario: Código interno de la plataforma, que identifica al usuario.
- nombre: Nombre del usuario.
- reviews\_count: Cantidad de reviews realizadas por el usuario.
- fans\_count: Cantidad de fans que posee el usuario.
- avg\_stars: Promedio de estrellas en las reseñas realizadas.
- compliment\_count: Cantidad de cumplidos que tienen las reviews del usuario.
- cool\_count: Cantidad de “cool” que tienen las reviews del usuario (buena onda).
- funny\_count: Cantidad de “funny” que tienen las reviews del usuario (gracioso).

**restaurantes**

id\_restaurante (PK)

id\_ciudad (FK)

- identificador\_yelp: Si el restaurante está en yelp, aquí aparecerá su identificador.
- identificador\_google: Si el restaurante está en google, aquí aparecerá su identificador.
- nombre: Nombre del comercio.
- direccion: Dirección del comercio.
- latitud: latitud geográfica para geolocalización.
- longitud: longitud geográfica para geolocalización.
- avg\_rating: Promedio de los rating recibidos.
- reviews\_count: Conteo de cantidad de reviews recibidas.
- estado: Si el local actualmente funciona o está cerrado.

**categoriasrestaurantes**

id\_categoria\_restaurante (PK)

id\_categoria (FK)

id\_restaurante (FK)

**serviciosrestaurantes**

id\_servicio\_restaurante (PK)

id\_servicio (FK)

id\_restaurante (FK)

**reviews**

id\_review (PK)

id\_restaurante (FK)

id\_usuario (FK)

- identificador\_plataforma: Código interno de la plataforma, que identifica la review..
- rating: Cantidad de estrellas que se otorgó.
- review: Texto de la review.
- fecha\_hora: Fecha y hora que se registró.
- useful: Cantidad de veces que la review fue marcada como “useful” (útil).
- funny: Cantidad de veces que la review fue marcada como “funny” (gracioso).
- cool: Cantidad de veces que la review fue marcada como “cool” (buena onda).

**consejos**

id\_consejo (PK)

id\_restaurante (FK)

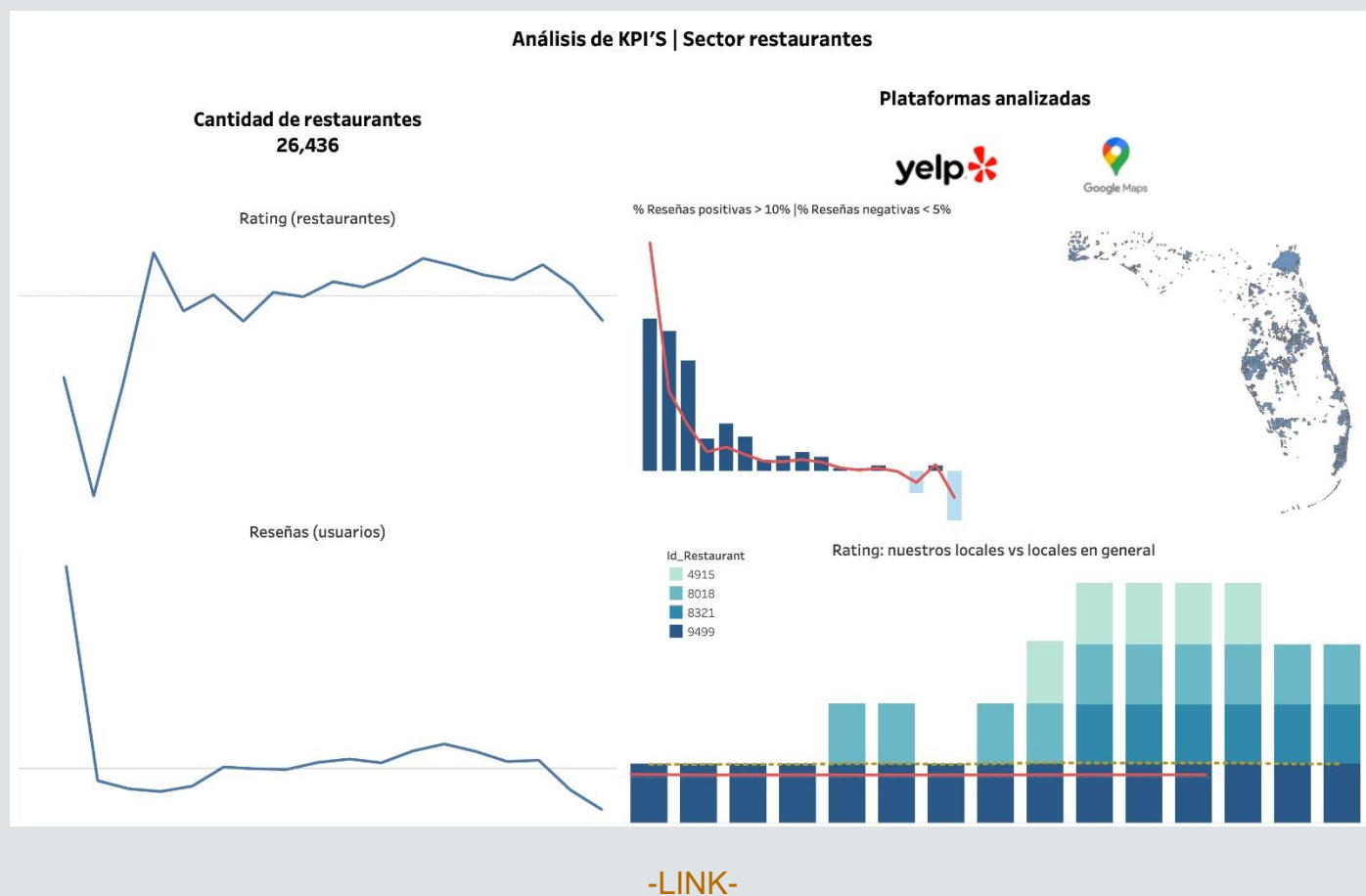
id\_usuario (FK)

- consejo: Texto del consejo.
- fecha: Fecha en que se registró.
- compliment\_count: Cantidad de usuarios que le marcaron un cumplido.



## Dashboard

Para la creación del dashboard, optamos por utilizar Tableau debido a sus funciones avanzadas y la capacidad de compartir y publicar fácilmente los resultados. En las diversas pestañas se presentan los KPIs previamente definidos para evaluar el desempeño de los restaurantes y seguir de cerca la evolución del mercado.



## Modelo ML

Para el modelo de machine learning seleccionado por nuestro cliente, hemos optado por centrarnos en los nombres y categorías de los restaurantes, así como en su ubicación. La función principal de este modelo será proporcionar recomendaciones de cinco locales en la ciudad elegida, que se alineen con los antojos del usuario final. En una estrategia de marketing y mejora de ventas, nuestro cliente ha decidido destacar un restaurante perteneciente a su cadena en el primer lugar de cada lista de recomendación, siempre que haya uno presente en la ciudad seleccionada.


En pro de la presentación, hemos trabajado con datos reducidos, enfocándonos exclusivamente en las localidades asociadas a las recomendaciones de inversión presentadas anteriormente al cliente.

### Ventajas para el Usuario:


- Acceso a recomendaciones personalizadas basadas en sus preferencias gastronómicas y ubicación, permitiéndoles descubrir nuevos lugares que se alineen con sus gustos particulares.
- Facilita el descubrimiento de locales bien valorados y cercanos, mejorando la experiencia culinaria del usuario al proporcionar opciones de alta calidad y convenientes.


### Ventajas para el Cliente:

- Estratégica promoción de los locales asociados al cliente, generando mayor visibilidad y popularidad en la plataforma.
- Facilita la llegada de usuarios a los establecimientos de nuestro cliente al proporcionar recomendaciones personalizadas, lo que se traduce en un aumento de tráfico.



## Recomendación Restaurantes






### ¿En qué ciudad de Florida te encuentras?

Confirma tu ubicación

Miami Beach

Miami Beach , increíble decisión.




### ¿Qué menú se te antoja hoy?

Ej.: Café, Smoothies, Pizza, Tacos, Hot Dog, Chicken, Sandwich, Burgers, Seafood, Chinese, Thai, Italian, Argentinian

Pizza

Recomendar

	nombre	direccion	ciudad	avg_rating
0	Joe's Stone Crab Restaurant	11 Washington Ave	Miami Beach	4.6
1	B&B Haulover Beach Cafe	10800 Haulover Park	Miami Beach	4.6
2	Fresh Restaurant	4441 Collins Ave [Poolside]	Miami Beach	4.2
3	Burgers And More	6995 Collins Ave	Miami Beach	3.9
4	Miami Grill South Beach	1541 Washington Ave	Miami Beach	4.1



Con la implementación de nuestro sistema de recomendación basado en aprendizaje automático, buscamos lograr un equilibrio entre las necesidades tanto de los usuarios como de nuestro cliente. Ofrecemos recomendaciones personalizadas que mejoran la experiencia del usuario final al tiempo que destacamos estratégicamente los locales asociados al cliente. Este enfoque integral no solo garantiza la satisfacción del usuario, sino que también fortalece la presencia y competitividad de nuestro cliente en el dinámico mercado gastronómico.

## Características técnicas

Trabajamos con scikit-learn y scipy para nuestro modelo de recomendación basado en categorías de restaurantes por varias razones:

- Ambas librerías, scikit-learn y scipy, son ampliamente utilizadas en la comunidad de ciencia de datos y aprendizaje automático. Esto asegura una abundancia de recursos, documentación y ejemplos disponibles.
- Scikit-learn proporciona implementaciones eficientes y de fácil uso de diversas métricas de similitud. En particular, el índice de Jaccard compara conjuntos de elementos, como las categorías de restaurantes en nuestro caso.
- Scipy ofrece funciones especializadas para calcular distancias pair-wise y convertirlas en una matriz cuadrada. Esta capacidad es esencial para evaluar la similitud entre puntos en un espacio n-dimensional.
- Ambas librerías están optimizadas para un rendimiento eficiente, lo que las hace capaces de manejar conjuntos de datos de tamaño considerable sin sacrificar velocidad.

