

# PROYECTO FINAL



Commerce Data Analysis and Recommendations

GitHub: Commerce Data Analysis and Recommendations

OCTUBRE 2023



# Sobre nosotros

Somos un grupo de consultores de datos, centrados en el análisis del mercado del turismo estadounidense, presentamos a continuación el desarrollo de este proyecto. Nuestra misión es proporcionar a los clientes finales herramientas que les permitan mejorar sus campañas de marketing, tomar decisiones informadas sobre inversiones y ofrecer recomendaciones basadas en experiencias previas para sus usuarios.

# El Equipo



**Leydy Peñaloza**

Technical Project Manager  
Data Engineer  
Data Scientist



**Camilo Ortiz**

Data Engineer  
Data Analyst  
Data Storyteller



**Aldemar Bohorquez**

Data Engineer  
Data Scientist  
Machine Learning Engineer



**Mayren Silva**

Data Engineer  
Data Analyst  
Machine Learning Engineer



**Yesica Leon**

Data Engineer  
Data Analyst  
Data Scientist

# Nuestros Servicios



Nuestra Aplicación de servicios para nuestros clientes dará información precisa respecto a:

## **Recomendaciones de ubicación para nuevos Hoteles:**

**01**

hacemos uso de análisis de datos y modelos de machine learning para identificar las ubicaciones óptimas para nuevos hoteles de su cadena, considerando variables como la densidad de población, el turismo, la afluencia de viajeros, eventos locales y los precios de la propiedad raíz, maximizando así su éxito en el mercado.

**02**

## **Recomendaciones para segmentación de Anuncios en Google y Redes Sociales:**

Utilizamos análisis de datos y modelos de machine learning para recomendar estrategias de segmentación y palabras clave altamente efectivas en Google y redes sociales, considerando la demografía, el turismo, los viajeros y los eventos locales de ubicaciones específica, optimizando así su presencia en línea y aumentando la conversión.

**03**

## **Outsourcing Logístico:**

identificamos áreas de mejora a partir de opiniones y calificaciones negativas, impulsando una experiencia de hospedaje más satisfactoria.

# Objetivos

## Recopilar

Depurar y disponibilizar la información en un Data Warehouse (proceso de ETL) de forma estática y dinámica.

## Analizar

El conjunto de datos cargados en el Data Warehouse y resumir sus principales características (proceso del EDA).

## Entrenar

Y poner en producción un modelo de Machine Learning que permita predecir cuáles son los rubros del negocio que más crecerán o decaerán y dónde es conveniente emplazar nuevos locales del negocio.

## Generar

A través de Machine Learning un sistema de recomendación del negocio para los usuarios con el propósito de que estos puedan conocer nuevas temáticas basados en sus experiencias previas.

### ÍNDICE DE SATISFACCIÓN DEL CLIENTE :

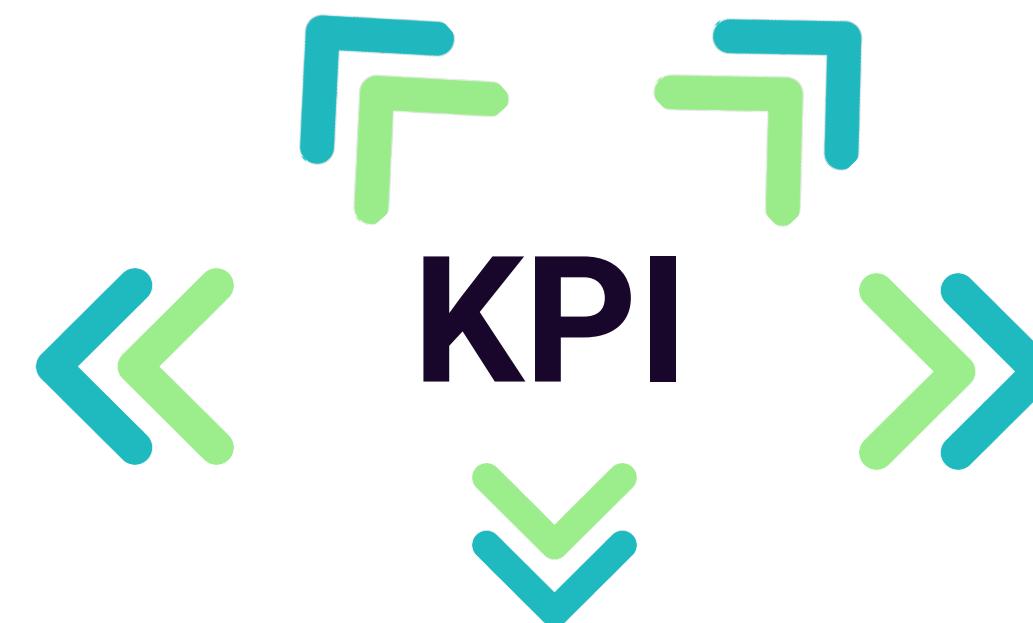
La opinión inmediata de los clientes tras finalizar la interacción con la empresa es una buena forma de conocer su percepción sobre el servicio brindado .

### TASA DE RETENCIÓN DEL CLIENTE:

Se trata de un porcentaje que mide cuántos clientes conserva una empresa al final de un plazo determinado, después de comparar la adquisición de nuevos clientes contra el número de clientes que se perdieron.

### NET PROMOTER SCORE(NPS) PUNTUACIÓN NETA DEL PROMOTOR:

evalúa el grado en que un cliente recomienda un cierto rubro (si un cliente aprecia un servicio lo suficiente como para recomendarlo a otros) % Detractores - % Promotores = NPS



### EL COSTE DE ADQUISICIÓN DEL CLIENTE O CAC:

cuánto dinero has utilizado para capturar a nuevos clientes .

### ÍNDICE DE PENETRACIÓN DEL MERCADO (MPI):

la penetración del mercado se centra en reforzar la relación e interacción de los clientes con el servicio a fin de aumentar el compromiso o engagement de las personas con el servicio prestado. % MPI = Clientes que accedieron al servicio / tamaño total de mercado para este servicio.

# Gestor del Proyecto

cistelsa / Projects / PG - Commerce Data Analysis and Recommendations

Type  to search | > | + | ○ | ⌂ | ⌂ | 🔍

PG - Commerce Data Analysis and Recommendations

Herramientas | Equipos de Trabajo | Campaña | Mis Tareas | Estado | Lista de Tareas | Cronograma | New View

-no:assignee

August 2023 September 2023 October 2023 November 2023

29 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Creación Diagrama de GANTT #2  
Preparación de la plataforma Microsoft Fabric #16  
Documentación - Mayren y Aldemar  
Parámetros generales de presentación  
Bibliografía General  
Gestionar Backup de la nube (MFabric) a local VSC #11  
Extracción, Transformación y Carga de Datos a Datalake (MFabric) #12  
EDA - Análisis Exploratorio de Datos (por sus siglas en inglés, Exploratory Data Analysis) #13  
Creación del DataLake para lectura de datos pesados #17  
Definición del Producto #14  
Búsqueda Nuevos Datasets #15  
ETL completo #22  
Estructura de datos implementada (DW, DL, etc) #23  
Creación del Diagrama Entidad Relación #27  
Filtrado Datasets de YELP #28  
Parámetros Generales de Presentación #29  
Llaves Para Relación Entre Bases de Datos #30  
Automatización #31  
Diccionarios de Bases de Datos #32  
Documento de Presentación de Proyecto #33  
Documentación en Repositorio #34

Sprint #1

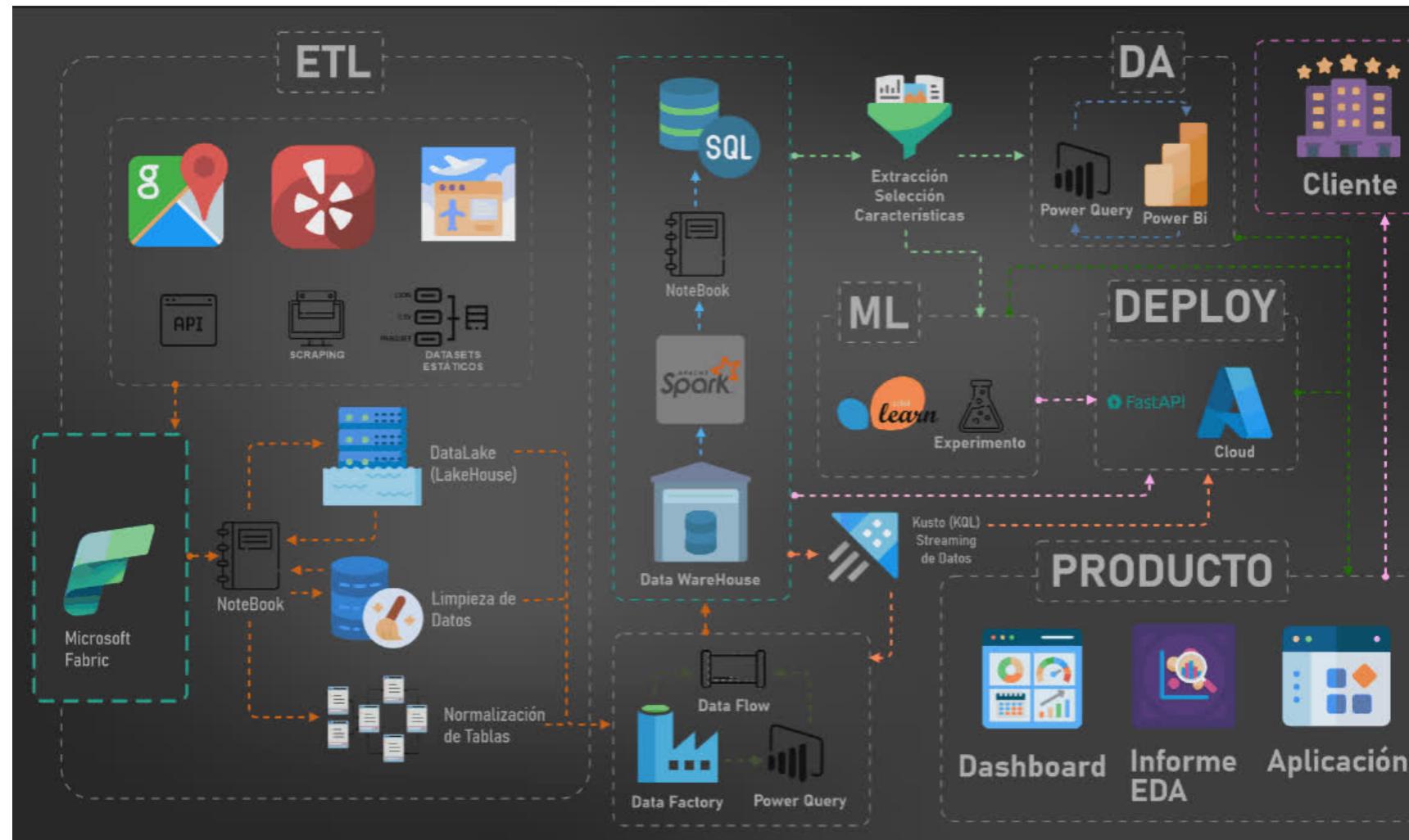
Sprint #2

Markers Priority Date fields Month Today

Save

# Soporte tecnológico

Son diversas herramientas las cuales nos van a ayudar a cumplir nuestros objetivos a nivel Técnico y Profesional.



## Microsoft Fabric:

Es un todo en uno de data, se integran todas las herramientas para ETL, EDA y DA. Se propuso esta herramienta con el fin de adelantarnos a la tendencia, ya que se encuentra en fase Beta y pronto saldrá la versión Oficial, viniendo de Microsoft y la inversión que ha realizado los ultimos años en herramientas de data posicionandose en segundo lugar, tendremos la mejor experiencia en un entorno muy Profesional y nos ayudará para futuros proyectos en diferentes empresas.



## NoteBooks:

Trabajaremos con esta herramienta conectada a Python, SQL y Apache Spark Data Factory. Esta herramienta nos ayuda a tener un flujo de datos entre areas y automatización de tareas, tambien es muy importante para la ingesta de datos en el Data WareHouse.



## Apache Spark:

Hadoop y Spark nos ayuda para el procesamiento de grandes cantidades de datos en forma de nodos, paralelizando el trabajo y siendo más eficiente de acuerdo su configuración y el tipo de archivo que se use, lo podremos gestionar con Python o con SQL.



## Scikit Learn:

Lo usaremos para crear nuestros modelos de ML, también con una herramienta "Experiment" la cual realizamos un seguimiento del desarrollo a los modelos de ML y validar las hipótesis.



## Power Bi:

Aprovecharemos esta herramienta para realizar, el análisis y la visualización de datos creando un Dashboard muy profesional y en la web.



## Matplotlib:

Es indispensable para crear el mejor informe EDA para que nuestros clientes puedan acceder a él de la forma más legible e intuitivo, será necesario traer nuestros modelos de ML allí y exponerlos.



## Kusto (KQL):

Muy posiblemente usaremos streaming de datos, no es algo seguro pero lo proponemos desde el inicio.



## Microsoft Azure:

Usaremos la nube de Microsoft para apoyarnos respecto a Bases de datos de SQL Server de ser necesarias, también para realizar el deploy de la aplicación para nuestros clientes a través de Fast API, usando Docker.

# Los Datos

**Los datos fueron extraídos haciendo uso de las API de las siguientes plataformas:**



Es un servidor de aplicaciones de mapas en la web que pertenece a Alphabet Inc. Ofrece imágenes de mapas desplazables, así como fotografías por satélite del mundo e incluso la ruta entre diferentes ubicaciones.

Programa Google Local Guides: bajo la consigna de compartir tu mundo en Google Maps para ayudar a otros usuarios a encontrar los mejores lugares locales y premiando a los usuarios con beneficios exclusivos.



El sitio web de Yelp, Yelp.com, es un sitio de redes sociales y reseñas de empresas locales de colaboración colectiva. las empresas pueden actualizar la información de contacto, los horarios y otra información básica del listado o agregar ofertas especiales.

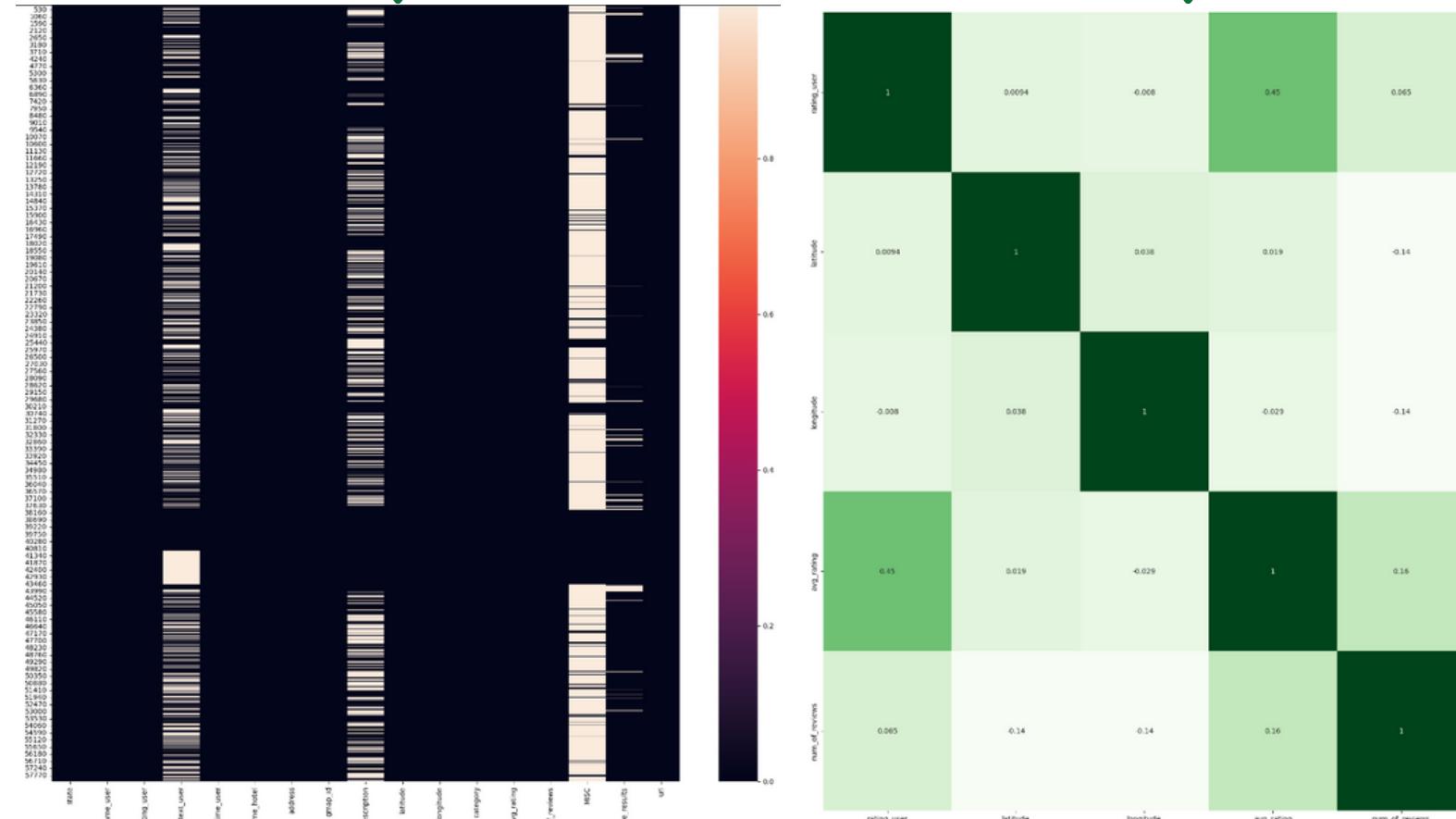
Además de escribir reseñas, los usuarios pueden reaccionar a las reseñas, planificar eventos o discutir sus vidas personales.



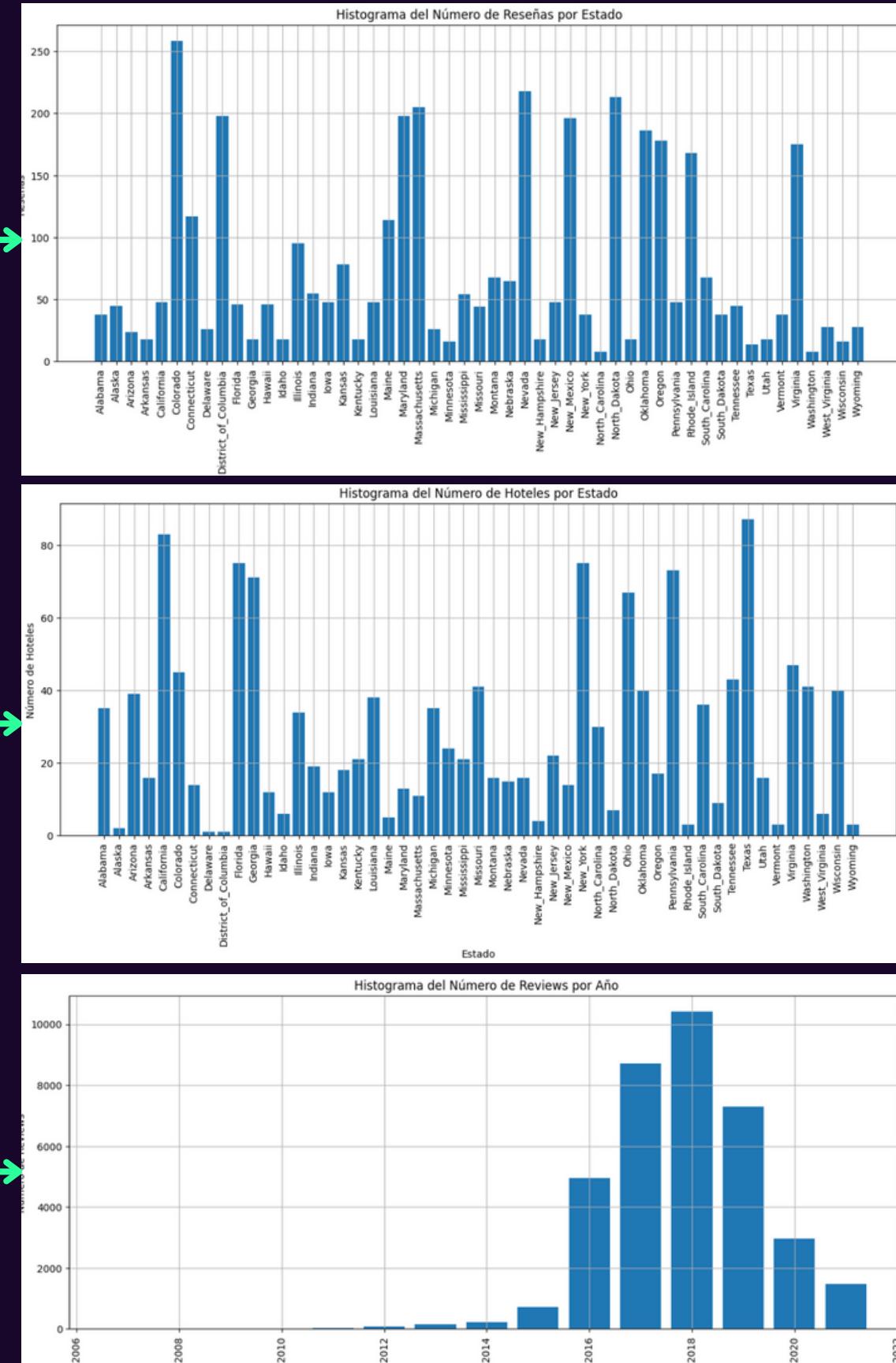
Es una plataforma proveedora de datos para negocios y portales web que venden servicios de hospedaje similar a booking.com funciona a través de una API y podemos gestionar una aplicación de reservas de habitaciones haciendo escritura, pero para nuestro proyecto solo lo necesitamos como lectura.

Para el desarrollo de esta fase del proyecto se realizó el Análisis Exploratorio de Datos (Exploratory Data Analysis o EDA, por sus siglas en inglés)

## Heatmaps

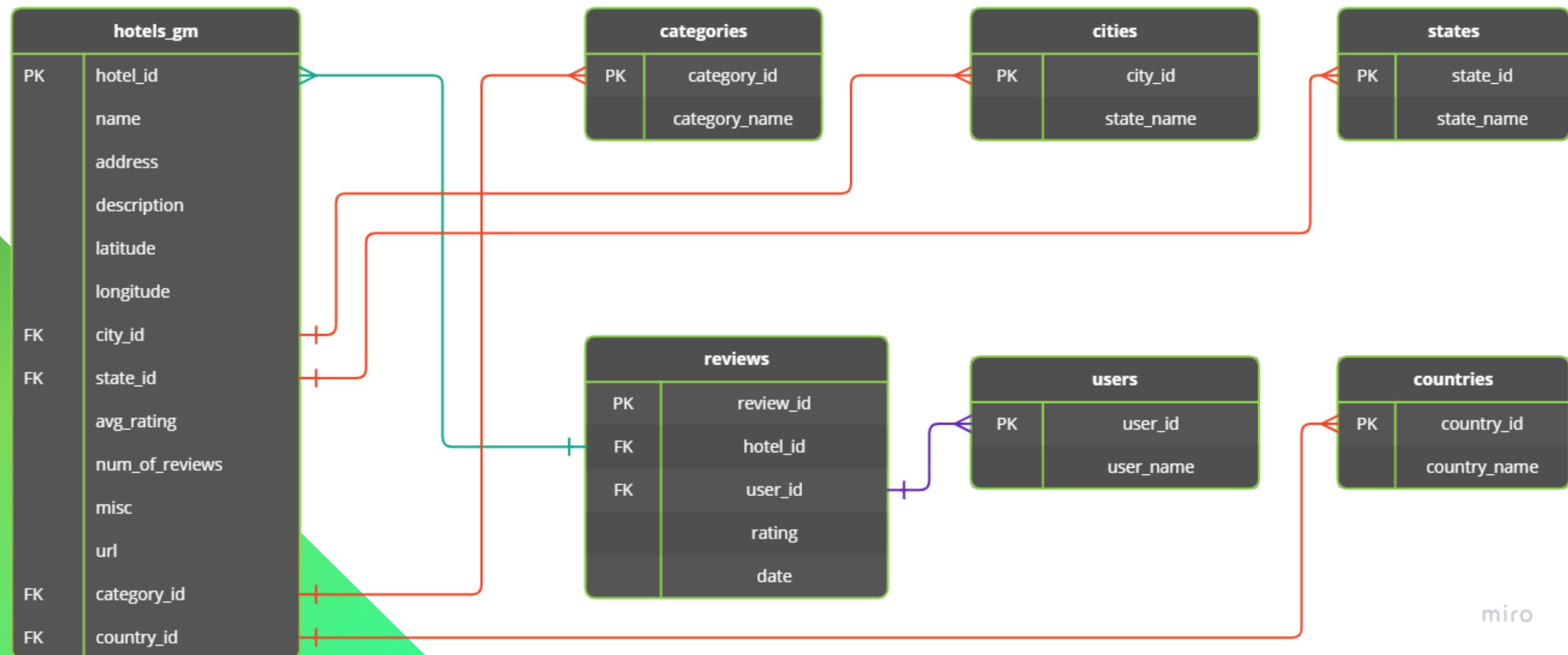


## Histogramas



# Google Maps

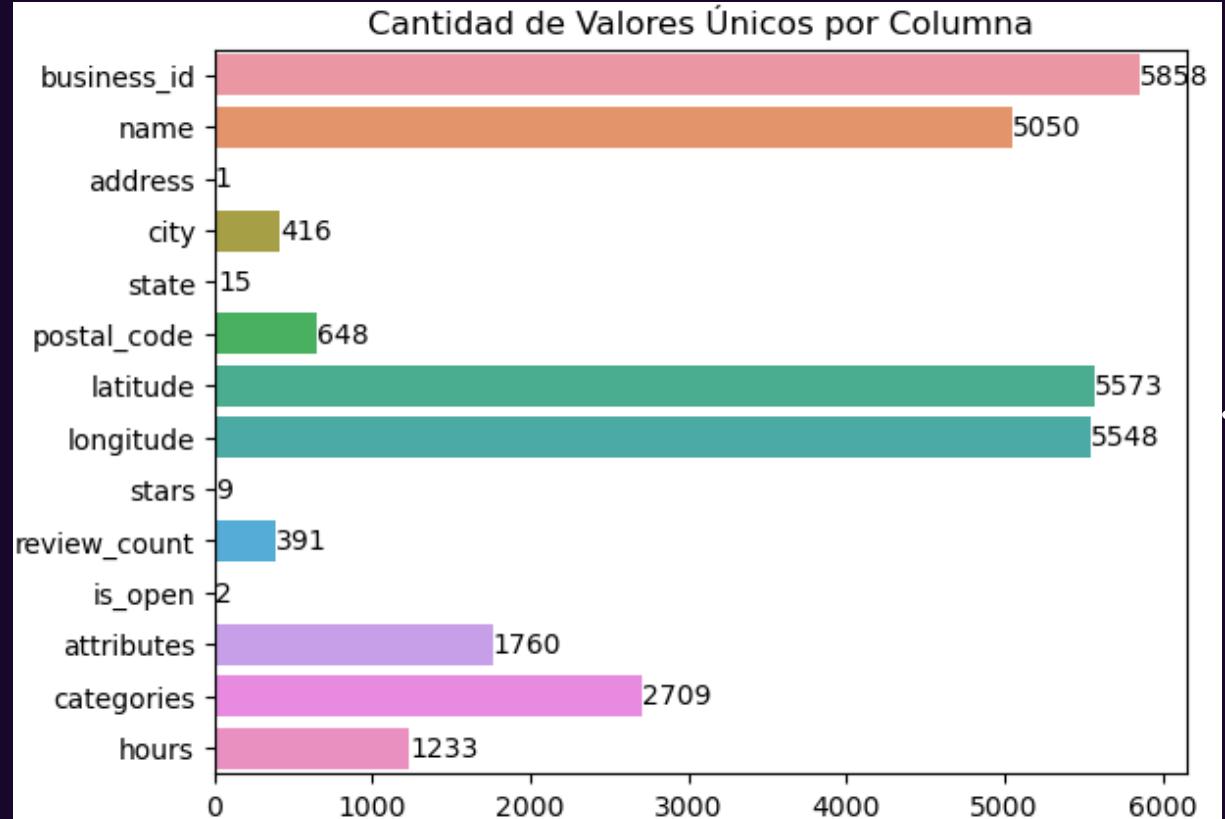
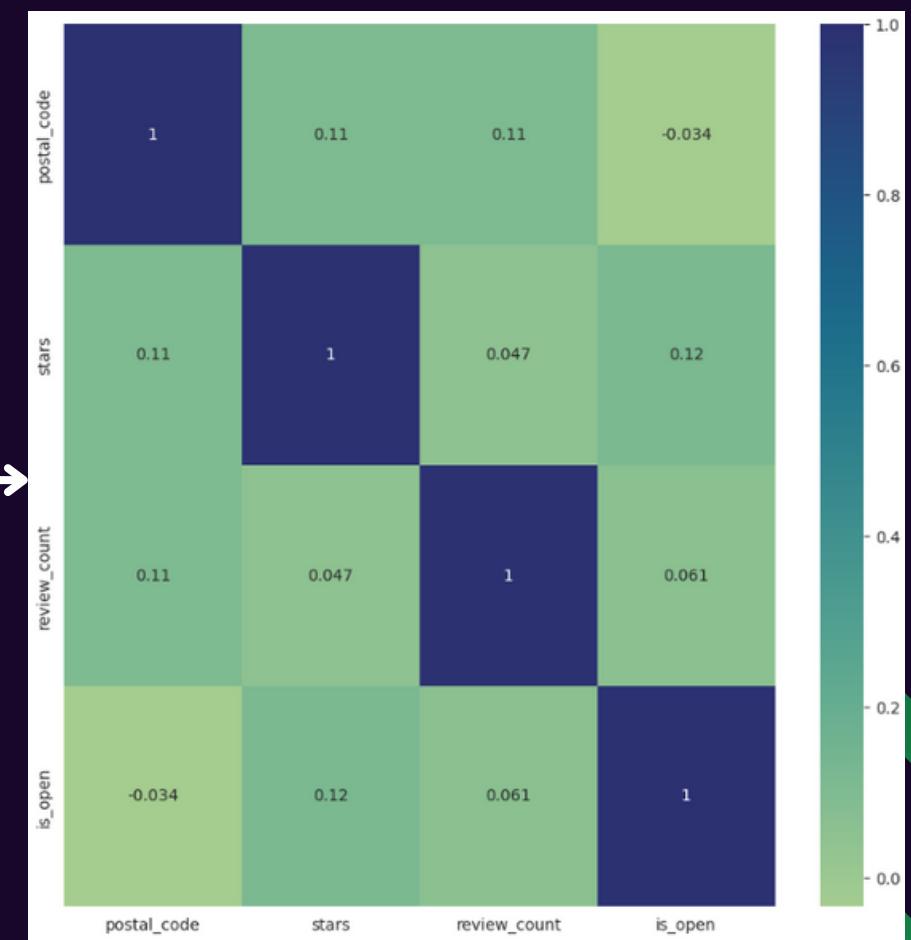
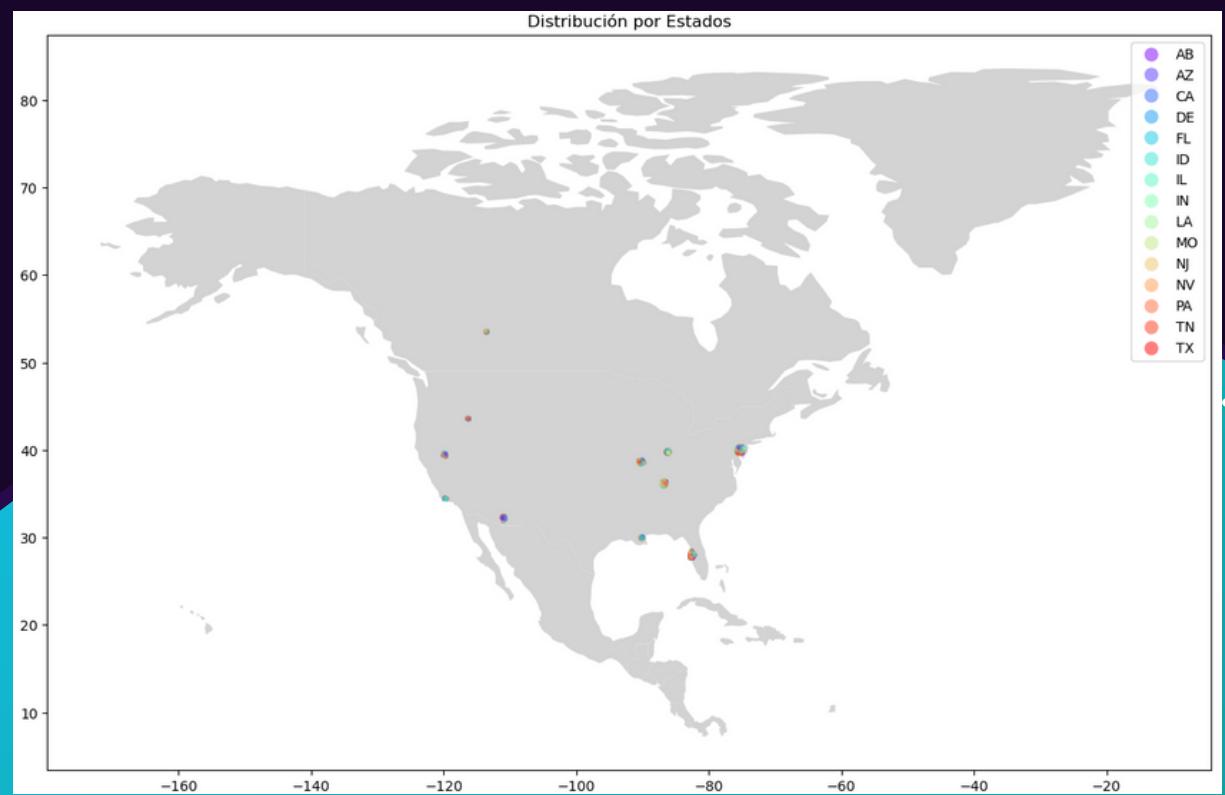
# DER



## Análisis Multivariable y Correlación

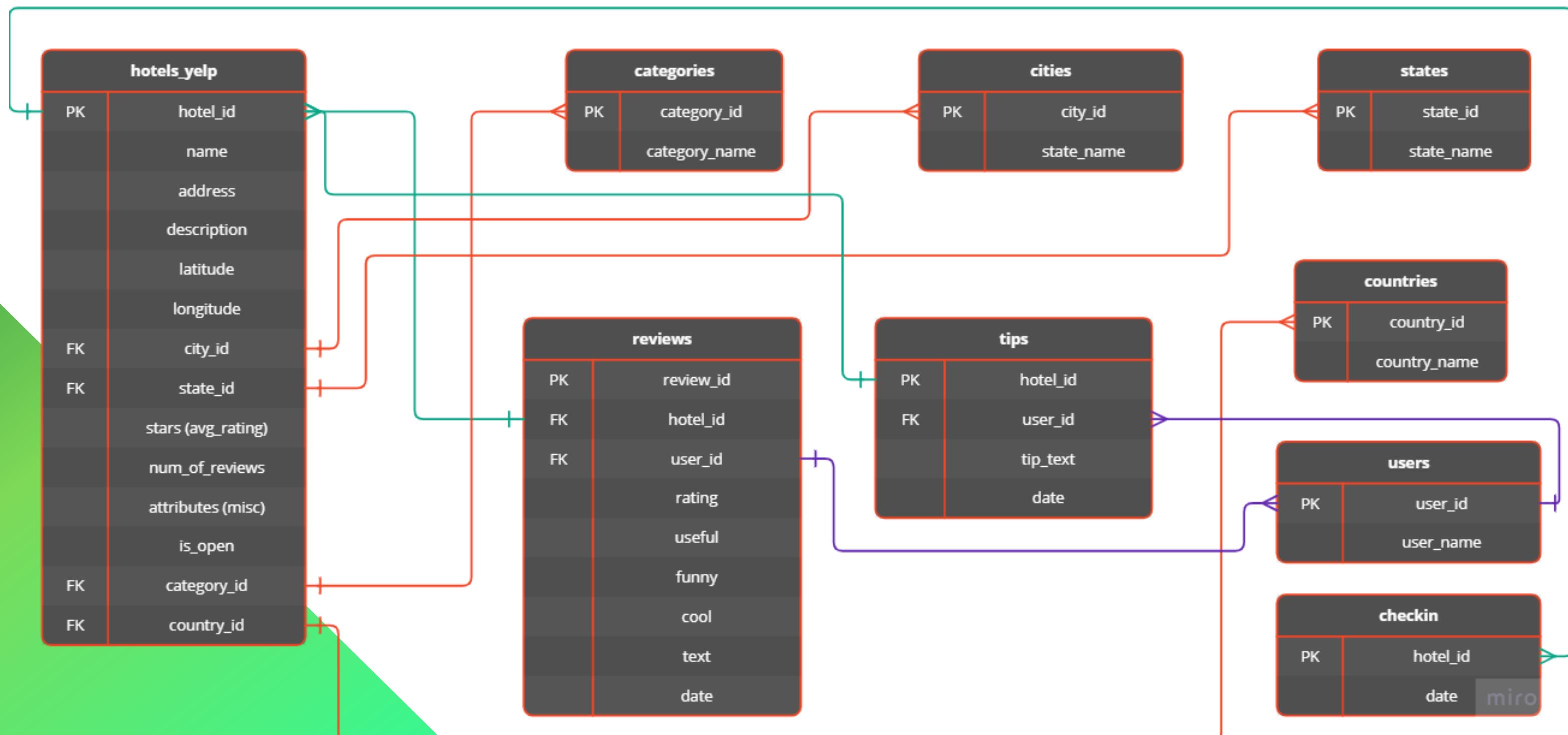


## Análisis Univariante



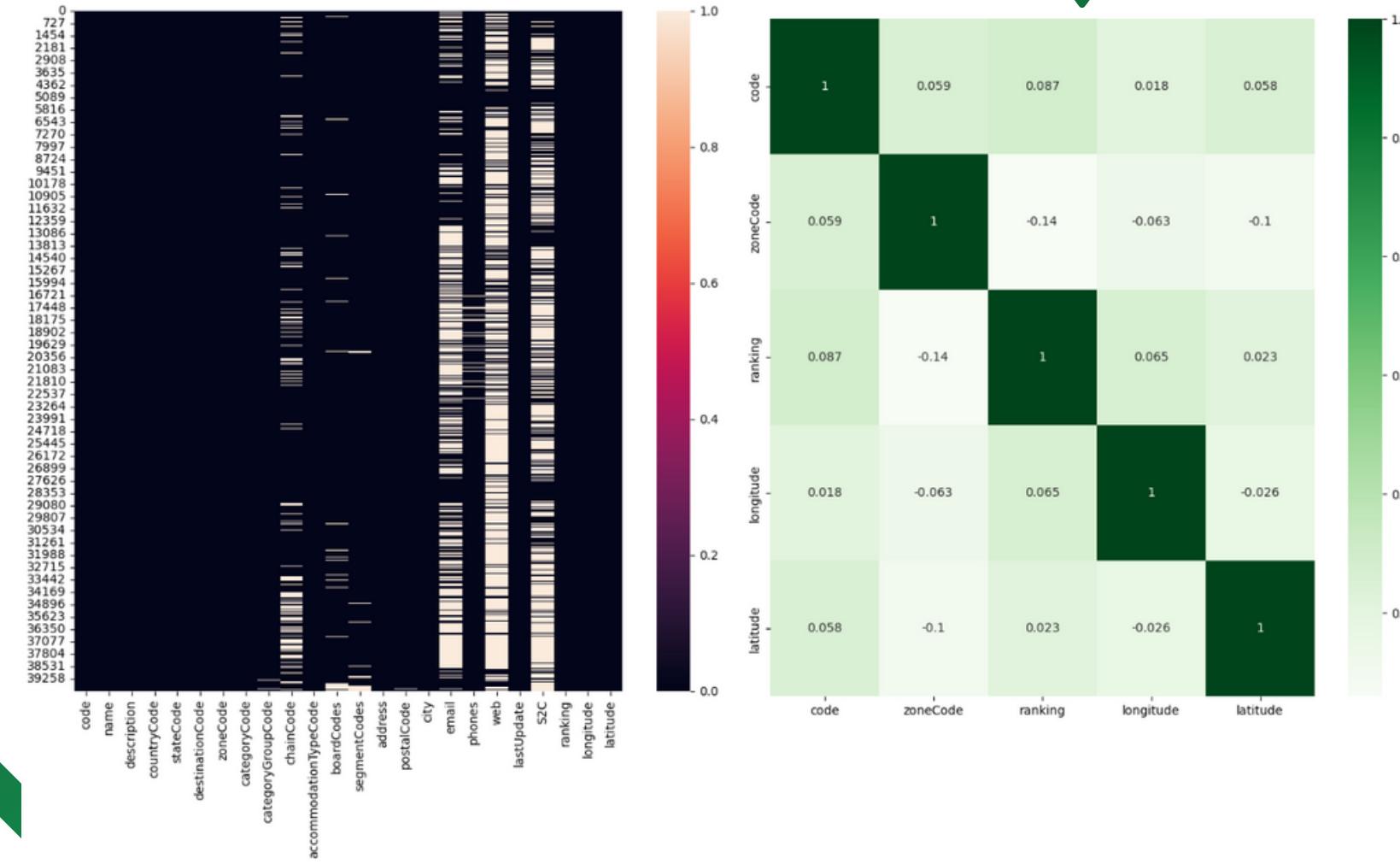
# Yelp

# DER

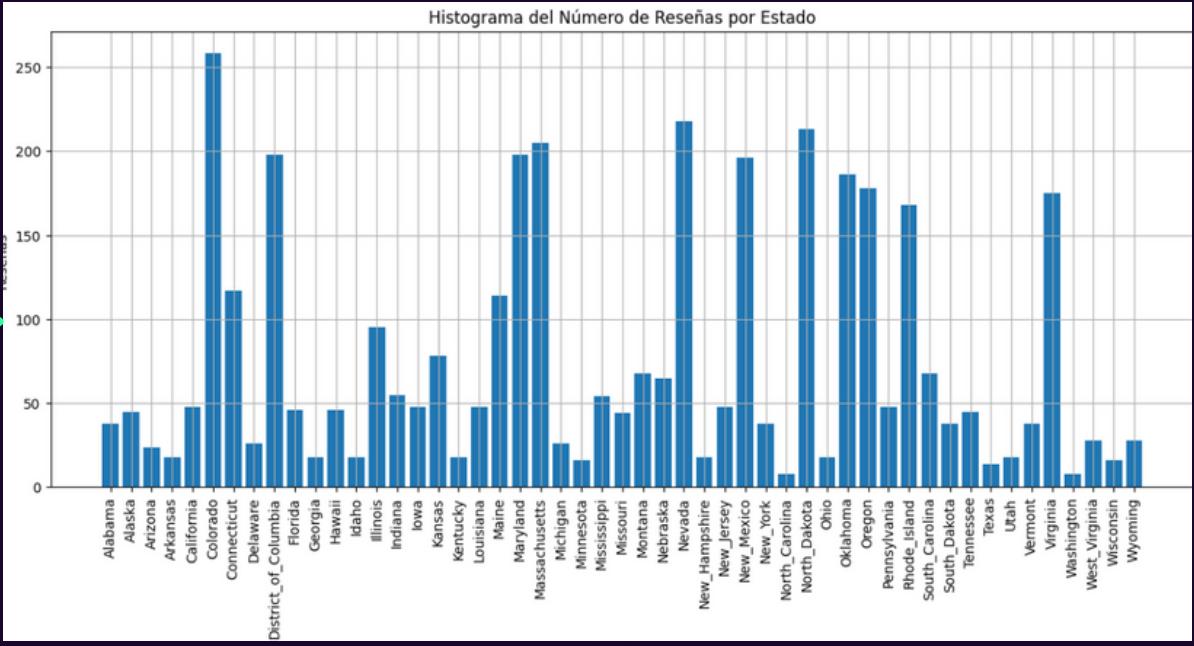


Para el desarrollo de esta fase del proyecto se realizó el Análisis Exploratorio de Datos (Exploratory Data Analysis o EDA, por sus siglas en inglés)

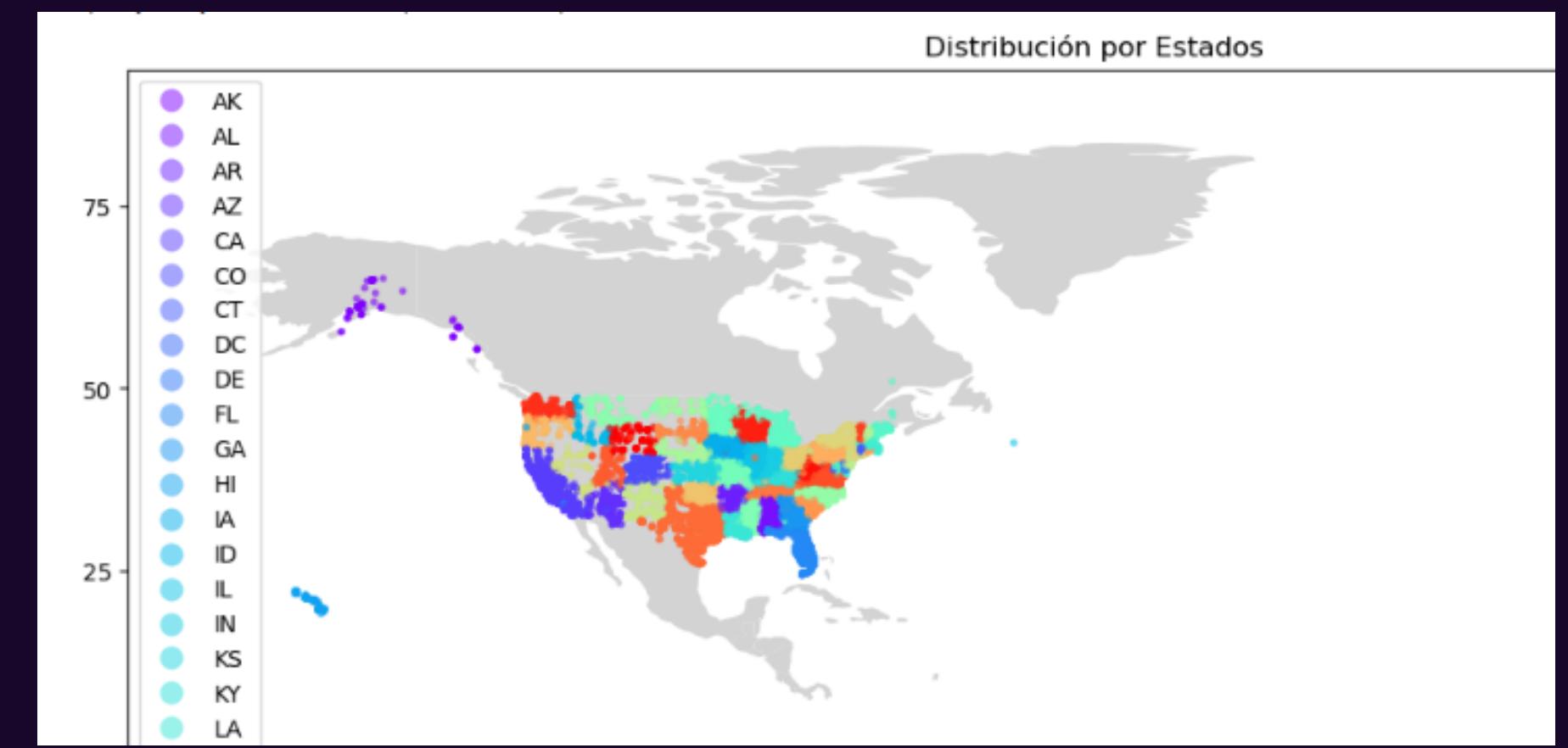
## Heatmaps



## Histograma

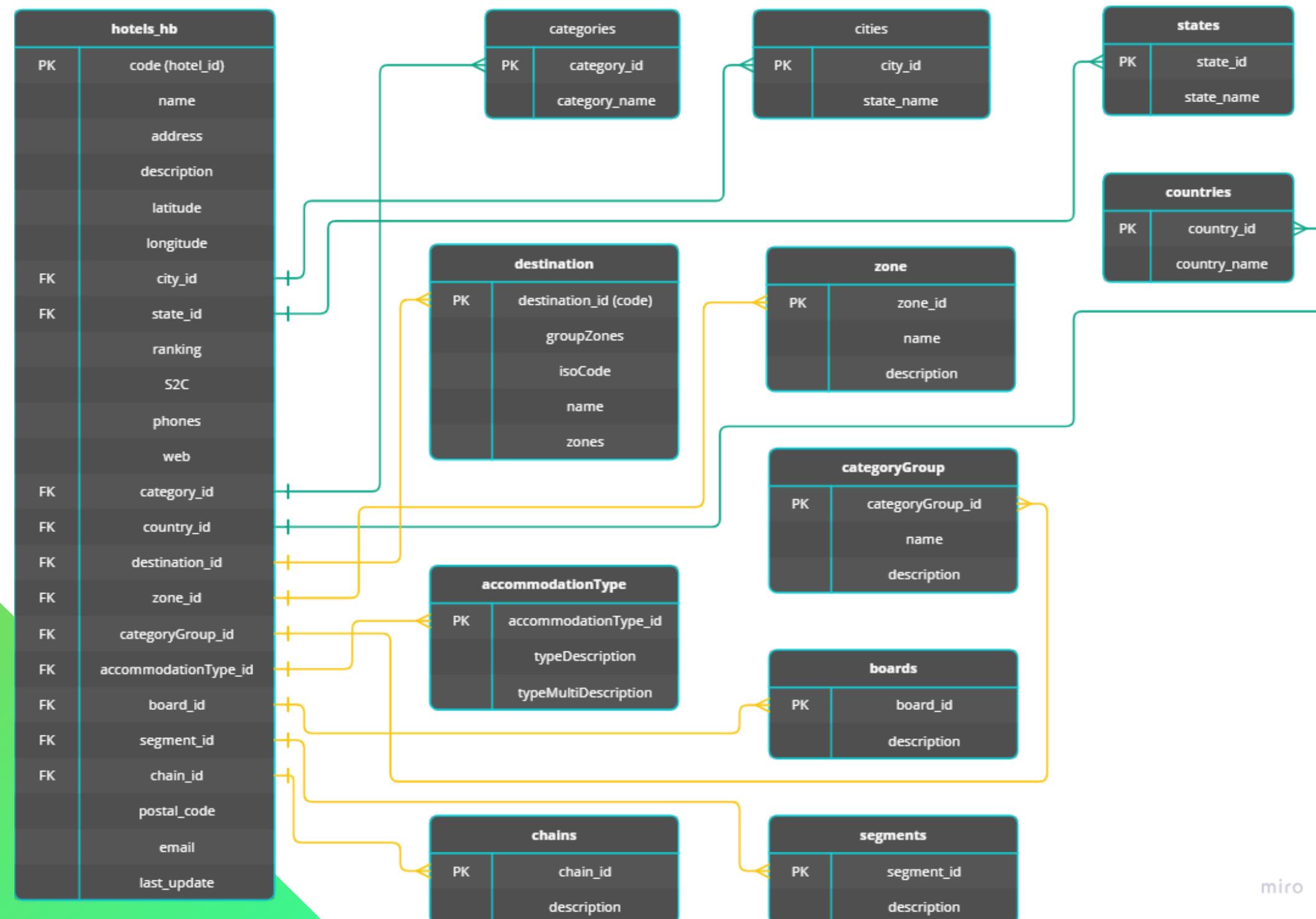


## Geolocalización por Estados



# HotelBeds

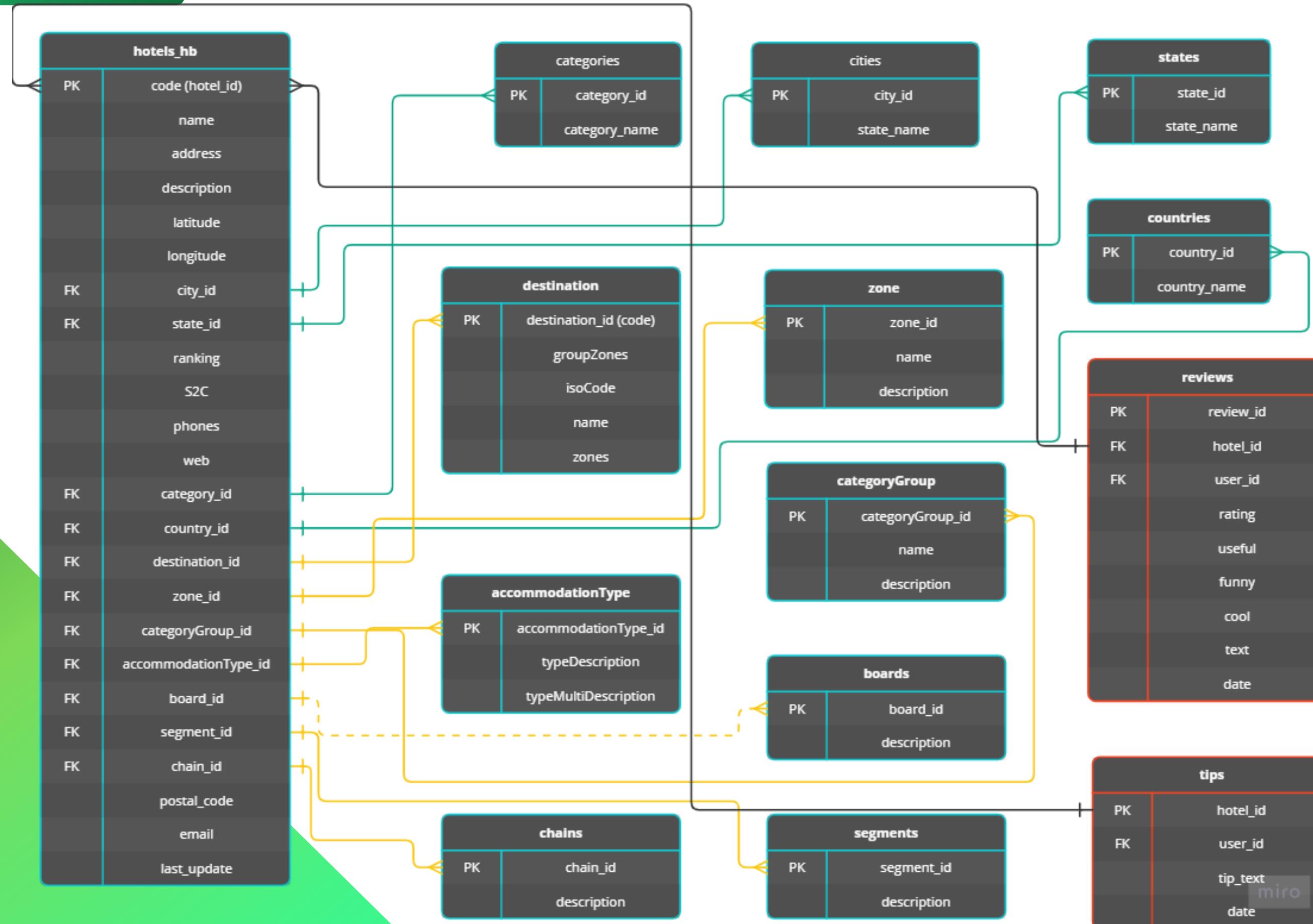
# DER



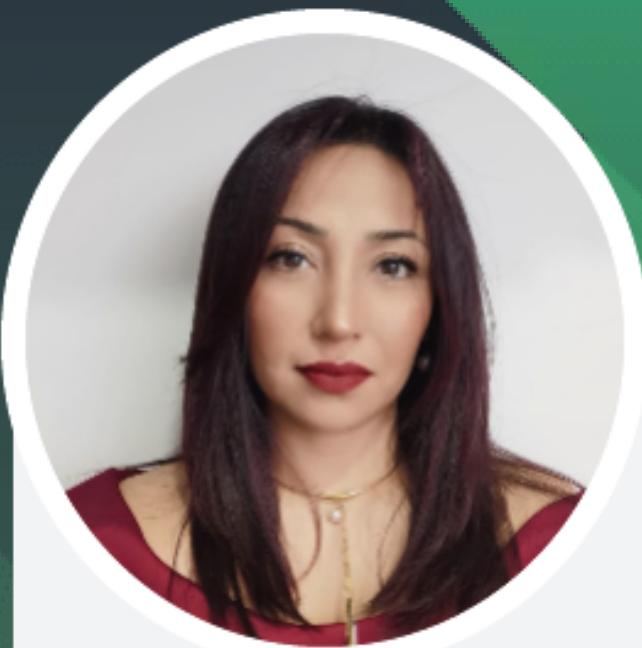
# Automatización

**MFabric: Commerce Data Analysis and Recommendations**





# CONTACTO

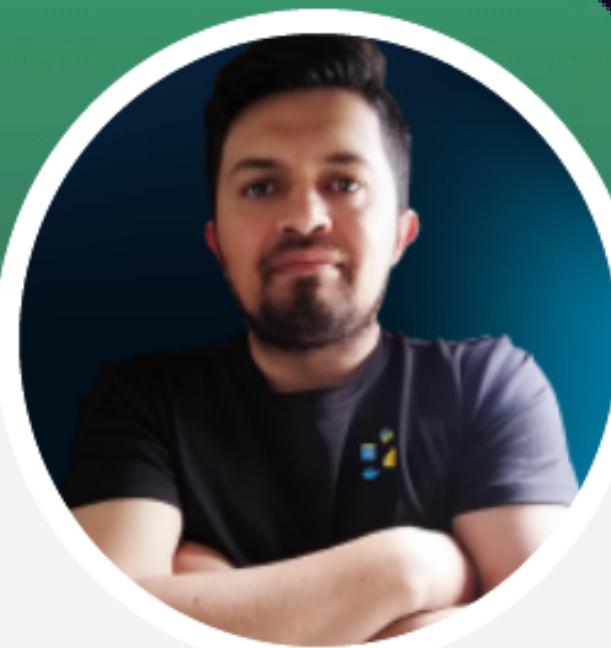


**Leydy Peñaloza**

✉ leydy.penaloza@gmail.com

👤 leydypenaloza  
GitHub

🔗 leydy-penaloza



**Camilo Ortiz**

✉ eortiz@cistelsa.com

👤 cistelsa  
GitHub

🔗 camilo-ortiz-cistelsa



**Aldemar Bohorquez**

✉ abr942010@gmail.com

👤 aldemarbr94  
GitHub

🔗 aldemar-bohorquez-rodriguez

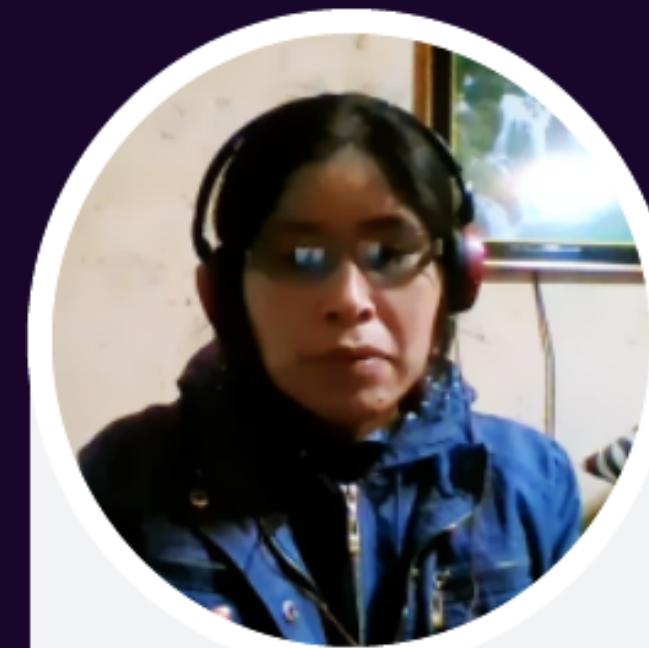


**Mayren Silva**

✉ mayrensilva95@gmail.com

👤 MayrenS95  
GitHub

🔗 mayren-gabriela-silva-basto



**Yesica Leon**

✉ agrostopo@gmail.com

👤 yesicamilagros  
GitHub

🔗 yesica-leon-ccahuana

MFabric: Commerce Data Analysis and Recommendations

MUCHAS  
GRACIAS