Proyecto Yelp & Google Maps



Sprint 2 - Data Engineering

El cliente

'Active Life Strategies' es una firma de inversiones en busca de oportunidades estratégicas en el sector de servicios orientados a una vida sana, en el estado de Florida, USA.

Los objetivos

Analizar el mercado para determinar la presencia o ausencia de establecimientos que ofrezcan servicios dirigidos a promover la vida saludable, en el estado de Florida, USA.

Evaluar las reseñas de los clientes, registradas en Google Maps y Yelp, sobre este tipo de negocios para entender la percepción de estos sobre los servicios recibidos.

Ofrecer a nuestro cliente recomendaciones sobre ubicaciones estratégicas, en las cuales posicionar sus locales y ofrecer sus servicios, buscando así asegurar el éxito en la inversión realizada.

Comprender la interacción de factores sociales, las tendencias emergentes y las preferencias del consumidor en esta región.

Alcance del producto

Negocios objetivo: aquellos enfocados en el bienestar y mejora de la vida saludable de los clientes, ubicados en el estado de la Florida, Estados Unidos. Fuentes de información las bases de datos de Yelp y Google Maps

Elementos entregables

Informe de análisis de la situación del mercado

Valoración del comportamiento de los usuarios frente a los servicios recibidos en el rubro analizado

Modelo de predicción y recomendación de posibles ubicaciones con buen potencial de éxito.

Tablero de métricas y KPI's

Data Engineering

El trabajo de ingeniería de datos en esta segunda semana fue extenso y muy importante para el desarrollo del proyecto. Se trabajó sobre el flujo de los datos, trasladando a la nube los procesos de ETL que se habían realizado de manera local durante la semana 1, y llevando a cabo todos los procesos necesarios para que los datos quedarán disponibles en un Data Warehouse y su carga se realizará de manera incremental y automática.

Data pipeline - Flujo de Trabajo



- Primera etapa de Captura.
- Tenemos datos y metadatos json provenientes de reviews de Yelp y Google maps
- Contamos con una herramienta de google: Cloud Functions

Cloud Functions

Tenemos implementada una función que reconoce la llegada de un archivo nuevo, reconoce el patrón y lo carga en una tabla en data storage.



Si no reconoce la etructura del archivo nuevo, la función puede crear una tabla nueva infiriendo el schema en base a lo que recibe para almacenar los datos. Con esta implementación vamos a automatizar la captura y carga de datos en el storage.

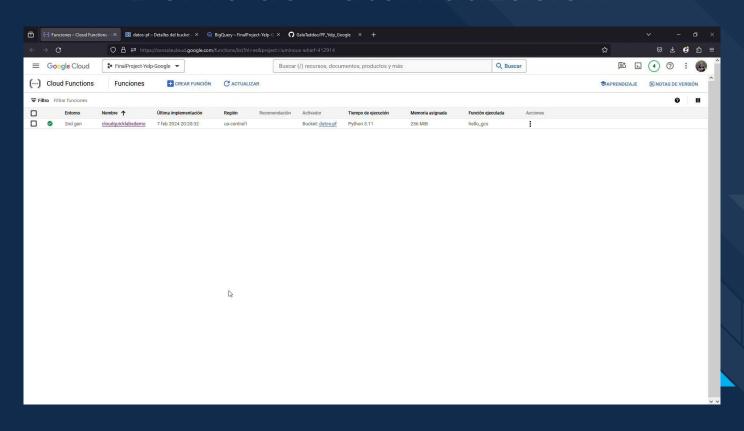
Tenemos probado exitosamente su funcionamiento, y preveemos que esta automatización se seguirá refinando en la siguiente semana

Cloud Storage

A solicitud de las analistas y para que puedan comenzar con sus pruebas y visualizaciones, procedimos a cargar el datawarehouse con los datos procesados y listos para consumir.

Tenemos el EDA y ETL completo que Felipe a continuación les va a compartir en detalle.

Demo de Automatización



ETL















Tratamiento de los datos

- 1. Eliminación de datos nulos, duplicados y outliers.
- 2. Desarrollo de nuevas columnas
- 3. Conversión de formatos
- 4. Filtro de los datos

text	sentimiento
If 0 stars was an option, i	0
wouldnt even give the place	
a star! There was garbage	
everywhere, clothes people	
had left behind. It was so	

text	sentimiento
Prices are too high	1
considering how economy is	
at this point in time	



text	sentimiento	
Absolutely AMAZING	2	
service! The staff was		
amazing friendly and very		
helpful with helping me		
choose what i wanted! And		/





Tratamiento de los datos

address

Porter Pharmacy, 129 N Second St, Miami, FL 950021



dirección	ciudad	estado	codigo_postal
Porter Pharmacy, 129 N Second St	Miami	FL	950021

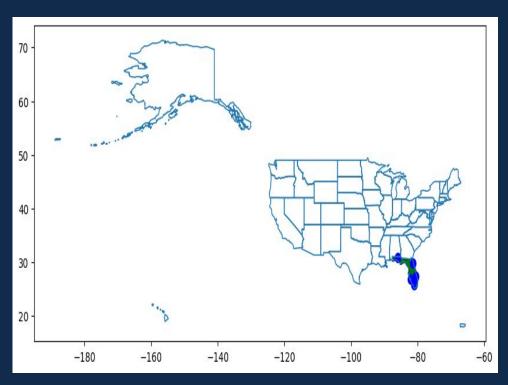
(1,54

latitude	longitude
1,5498132	88,456123



Location	
98132 , 88,456123)	

Filtro por ubicación geográfica



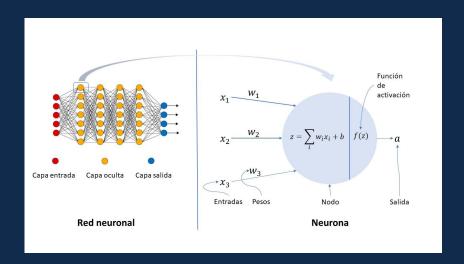
Filtro por categoría de negocio

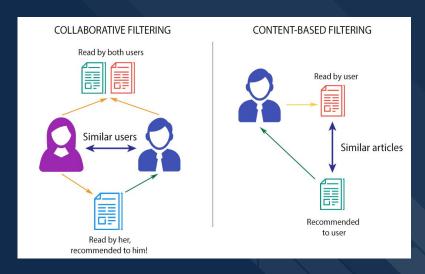
```
vegetarian restaurant
medical
cal center physical therapy
salon' day therapist' physical
                   store' health
                                                                                                                                                   hair removalwomen health
                        day span redical supply sports massage sports massage sports
                                 hair salon'goods store'
                                                                                                                              skin care therapist reflexologist
                   waxing hair health care supplements store spar beauty store inharmony salon' spa' life insurance conter sym medicine producine producine producine producine donut shop the salon's salon's shop the salon's salon's shop the salon's 
                    therapy clinic store organic
              health food gmm personal loss service care service
                                                                                 beauty shop'
                        staurant' delivery beauty salon' center wellness salon beauty chir
                                                                                                                                chiropractor chiropractor
                  medical clinic mental healthtrainer physical restaurant takeout center personal physical arts school
                                 removal service massage
                                 fitness center 'vegan restaurant 'shop' juice &
      spa' hair therapist' massage coffee shop's
                                               alternative medicine wellness program
```

Diagrama Entidad - Relación (DER)

Metadata		Reviews	
name	STRING	user_id ∅	STRING
gmap_id 🖉	STRING *	nombre	STRING
category	STRING	rating	INTEGER
ocation	STRING	text	STRING
direccion	STRING	resp	BOOLEAN
ciudad	STRING	1 gmap_id	STRING
estado	STRING	sentimiento	FLOAT
codigo_postal	STRING	sentiment_analysis	INTEGER

Modelo de Predicción y Recomendación

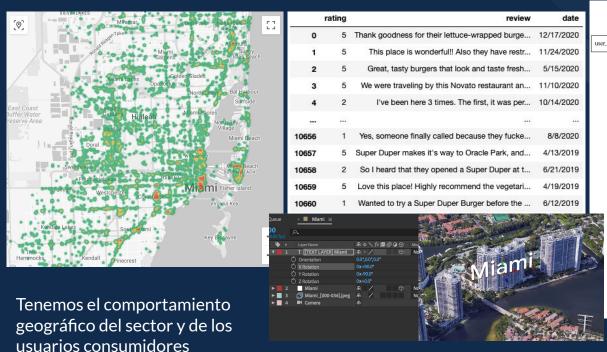




La inteligencia Artificial es el campo de las ciencias informáticas dedicado a la resolución de problemas cognitivos comúnmente asociados a la inteligencia humana, como el aprendizaje, la resolución y el reconocimiento de patrones.



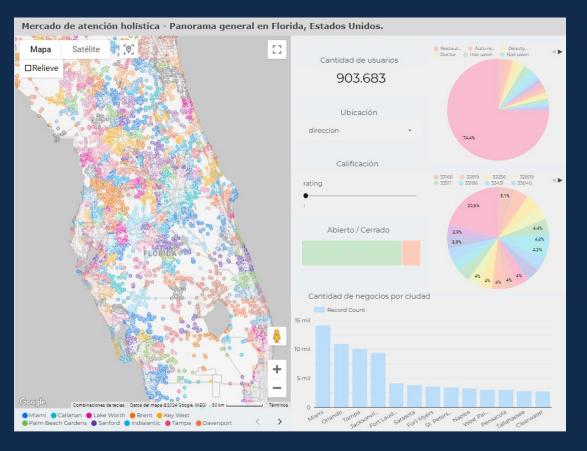
Modelo de Predicción y Recomendación



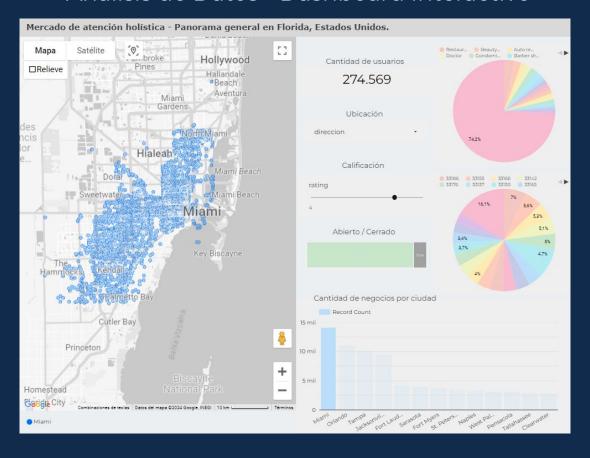
user_input: InputLayer restaurant_input: InputLayer user_bias: Embedding user_factor: Embedding restaurant_factor: Embedding restaurant_bias: Embedding flatten: Flatten flatten 1: Flatten user bias flat: Flatten concatenate: Concatenate dense: Dense restaurant_bias_flat: Flatten dense 1: Dense dense_2: Dense dense_3: Dense regression: Add Sigmoid_Range: Dense

La IA aprende y entrena como los seres Humanos!!

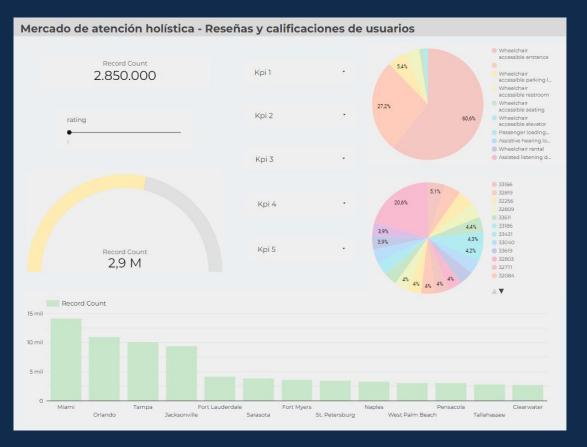
Análisis de Datos - Dashboard Interactivo



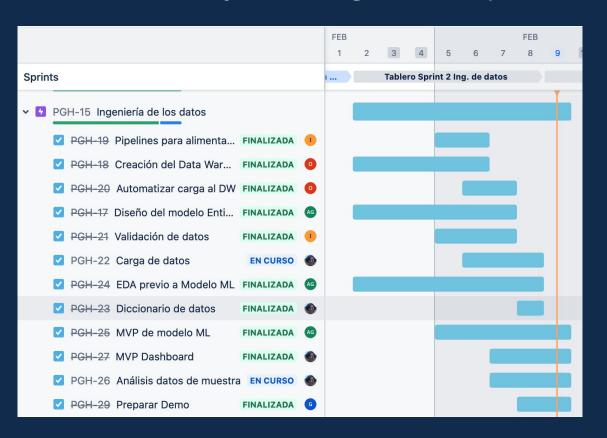
Análisis de Datos - Dashboard Interactivo



Análisis de Datos - Dashboard Interactivo



Planificación y Cronograma. Sprint 2



¡Muchas gracias!

