



soyHenry / PF\_DS



<> Code

Issues 1

Pull requests

Actions

Projects

Security

Insights



FULL-TIME

PF\_DS / Proyectos / yelp-goople.md



Go to file

t



HX-FNegrete Update yelp-goople.md

bffa6b6 · 2 months ago



History



Preview

Code

Blame

62 lines (33 loc) · 5.91 KB

Raw



## PROYECTO GRUPAL N°1

YELP & GOOGLE MAPS - REVIEWS AND  
RECOMMENDATIONS





# Google Maps

## Contexto

---

"La opinión de los usuarios es un dato muy valioso, que crece día a día gracias a plataformas de reseñas. Su análisis puede ser determinante para la planificación de estrategias. Yelp es una plataforma de reseñas de todo tipo de negocios, restaurantes, hoteles, servicios entre otros. Los usuarios utilizan el servicio y luego suben su reseña según la experiencia que han recibido. Esta información es muy valiosa para las empresas, ya que les sirve para enterarse de la imagen que tienen los usuarios de los distintos locales de la empresa, siendo útil para medir el desempeño, utilidad del local, además de saber en qué aspectos hay que mejorar el servicio. Además, Google posee una plataforma de reseñas de todo tipo de negocios, restaurantes, hoteles, servicios, entre otros integrada en su servicio de localización y mapas, Google Maps. Los usuarios utilizan el servicio y luego suben su reseña según la experiencia vivida. Muchos usuarios leen las reseñas de los lugares a los que planean ir para tomar decisiones sobre dónde comprar, comer, dormir, reunirse, etc. Esta información es muy valiosa para las empresas, ya que les sirve para enterarse de la imagen que tienen los usuarios de los distintos locales de la empresa, siendo muy útil para medir el desempeño, utilidad del local, además de identificar los aspectos del servicio a mejorar.""

## Rol a desarrollar

---

"Como parte de una consultora de data, nos han contratado para poder realizar un análisis del mercado estadounidense. Nuestro cliente es parte de un conglomerado de empresas de restaurantes y afines, y desean tener un análisis detallado de la opinión de los usuarios en Yelp y cruzarlos con los de Google Maps sobre hoteles, restaurantes y otros negocios afines al turismo y ocio, utilizando análisis de sentimientos, predecir cuáles serán los rubros de los negocios que más crecerán (o decaerán). Además, desean saber dónde es conveniente emplazar los nuevos locales de restaurantes y afines, y desean poder tener un sistema de recomendación de restaurantes para los usuarios de ambas plataformas para darle, al usuario por ejemplo la posibilidad de poder conocer nuevos sabores basados en sus experiencias previas. Pueden cambiar el tipo de comercio (no es necesario que sean restaurantes). "

# Propuesta de trabajo

---

Recopilar, depurar y disponibilizar la información: Creación de una base de datos (DataWarehouse) de diferentes fuentes, tanto provistas por Henry como incorporadas por ustedes, corriendo en local o alojada en proveedores en la nube. La base de datos depurada deberá contemplar por lo menos dos tipos diferentes de extracción de datos, ejemplo: datos estáticos, llamadas a una API, scrapping, entre otros.

Reporte y análisis significativos de la(s) línea(s) de investigación escogidas: El análisis debe contemplar las relaciones entre variables y concluir, si es que existe, una relación entre estas, y los posibles factores que causan dicha relación en la realidad.

Entrenamiento y puesta en producción de un modelo de machine learning de clasificación no supervisado o supervisado: El modelo debe resolver un problema y conectar globalmente con los objetivos propuestos que se propongan como proyecto. .

## Ideas de análisis e implementación

---

Mejoramiento de estrategias de marketing: campañas microsegmentadas

Sistemas de recomendación: Sobre algún lugar en particular, pueden ser comercios como restaurantes, hoteles, entre otros.

Datos adicionales a cruzar: Cotizaciones en bolsa, tendencias en redes sociales y medios de comunicación sobre comercios en expansión.

## Datasets y fuentes complementarias

---

"Los datos son extraídos de la plataforma de reseñas Yelp y de Google Maps, para Estados Unidos. Tiene información sobre la ubicación de los comercios, su categoría, puntajes promedios, si están abiertos o no, sobre los usuarios, las reseñas que hicieron, cuántas reseñas hicieron, cuántos votos han recibido esas reseñas entre otros. Se puede complementar el análisis con datasets adicionales que ofrezcan información que consideren pertinente al pedido (valores de acciones de las empresas, información geográfica adicional de los locales, etc). Recomendamos fuertemente el uso de NLP para procesar la información de las reseñas."

### Fuentes de datos

- [Diccionario de Datos](#)

Fuentes de datos obligatorias:

- [Dataset de Google Maps](#)
- [Dataset de Yelp!](#)

## Disclaimer

---

De parte del equipo de Henry se quiere aclarar y remarcar que los fines de los proyectos propuestos son exclusivamente pedagógicos, con el objetivo de realizar proyectos que simulen un entorno laboral, en el cual se trabajen diversas temáticas ajustadas a la realidad. No reflejan necesariamente la filosofía y valores de la organización. Además, Henry no alienta ni tampoco recomienda a los alumnos y/o cualquier persona leyendo los repositorios (y entregas de proyectos) que tomen acciones en base a los datos que pudieran o no haber recabado. Toda la información expuesta y resultados obtenidos en los proyectos, nunca deben ser tomados en cuenta para la toma real de decisiones (especialmente en la temática de finanzas, salud, política, etc.).