

Conociendo al Cliente 360°

Proyecto Integrador de Data Science

Análisis integral del ecosistema gastronómico de Chicago combinando datos de usuarios y restaurantes para generar insights estratégicos y sistemas de recomendación personalizados.



Información del Proyecto

- ▶ **Cliente:** InsightReach (Ficticio)
- ▶ **Sector:** Marketing Digital
- ▶ **Enfoque:** Campañas Personalizadas
- ▶ **Ciudad Objetivo:** Chicago, Illinois
- ▶ **Periodo:** Primer Módulo - Henry Bootcamp



Alcance Técnico

- ▶ **Registros procesados:** 30,000 usuarios
- ▶ **API integrada:** Yelp Business API
- ▶ **Restaurantes analizados:** 199
- ▶ **Pipeline:** ETL completo
- ▶ **Metodología:** Data Science 360°



Objetivos del Proyecto

Cuatro pilares estratégicos para el análisis gastronómico integral



Análisis de Usuarios

Limpieza y segmentación estratégica de base de datos de clientes para identificar patrones de comportamiento y preferencias gastronómicas específicas.



Integración de Datos

Obtención e integración de información de restaurantes vía API de Yelp para enriquecer el ecosistema de datos con información actualizada y confiable.



Sistema de Recomendaciones

Desarrollo de modelos predictivos avanzados para matching inteligente usuario-restaurantes basado en preferencias y comportamientos históricos.



Insights Estratégicos

Generación de recomendaciones accionables para campañas de marketing digital personalizadas con alto potencial de conversión.



Pipeline del Proyecto

Metodología ETL robusta y escalable con validación en cada etapa



Avance EDA - Análisis Exploratorio

Limpieza y procesamiento de datos de usuarios

1

Procesamiento integral de 30,000 registros con imputación inteligente por ciudad y estrato socioeconómico. Normalización de rangos de edad (18-90 años) y estandarización completa de tipos de datos.

📁 **Archivos generados:** usuarios_limpios.csv | Datos_usuarios_Chicago_limpios.csv



Avance API YELP - Integración Externa

Consumo de API y procesamiento de datos de restaurantes

2

Conexión exitosa a API de Yelp obteniendo 199 restaurantes de Chicago. Análisis de similitud con FuzzyWuzzy (>70% precisión) e imputación contextual de precios por categoría gastronómica.

📁 **Archivos generados:** chicago_restaurants.csv | Datos_YELP_limpios.csv



Análisis Final - Integración y Recomendaciones

Síntesis de datos y generación de insights









3

Integración completa de datasets usuarios-restaurantes. Desarrollo de sistema de recomendaciones personalizado con visualizaciones avanzadas y documento final de insights estratégicos.

📁 **Archivos generados:** Recomendaciones.pdf | Visualizaciones interactivas

Estructura del Proyecto

Organización completa de notebooks, scripts y datasets

| Archivo | Tipo | Descripción | Estado |
|------------------------------------|-----------|---|--|
| Avance_EDA.ipynb | Notebook | Análisis Exploratorio y limpieza de datos |  Completado |
| Avance_API_YELP.ipynb | Notebook | Consumo de API de Yelp y procesamiento |  Completado |
| Avance_Analisis_Final.ipynb | Notebook | Análisis final y recomendaciones |  Completado |
| mega_funcion.py | Script | Funciones de apoyo (usado en Avance_EDA) |  Funcional |
| base_datos_restaurantes_USA_v2.csv | Dataset | Dataset inicial proporcionado |  Input |
| usuarios_limpios.csv | Dataset | Dataset limpio (Avance_EDA) |  Output |
| chicago_restaurants.csv | Dataset | Datos crudos de API (Avance_API_YELP) |  Output |
| Recomendaciones.pdf | Documento | Insights y recomendaciones finales |  Reporte |



Instalación y Configuración



Requisitos Previos: Python 3.7+ debe estar instalado en el sistema



Instalación de Dependencias

```
pip install -r requirements.txt
```



Librerías Principales Utilizadas



pandas

Manipulación de datos



numpy

Operaciones numéricas



matplotlib

Visualización



seaborn

Gráficos estadísticos



fuzzywuzzy

Similitud de strings



jupyter

Notebooks interactivos



requests

Consumo de APIs



openpyxl

Manejo de Excel



Resultados por Avance

Métricas de calidad y logros técnicos destacados en cada fase



Avance EDA: Usuarios

Registros
procesados 30,000 → 26,847

Compleitud de datos 100%

Rangos
normalizados 18-90 años

Ciudad foco Chicago

Tipos
estandarizados Completo



Avance API: Restaurantes

Restaurantes obtenidos 199

Similitud FuzzyWuzzy >70%

Precisión matching >95%

JSON
normalizados Procesado

Imputación precios Contextual



Calidad Final

Compleitud 100%

Consistencia Estandarizada

Precisión >95%

Validez rangos Realistas

Escalabilidad Pipeline ETL



Enfoques técnicos avanzados aplicados en el proyecto



Imputación Contextual

Valores faltantes imputados considerando ciudad + estrato socioeconómico, garantizando coherencia geográfica y demográfica en los datos.

- Análisis por ciudad específica
- Segmentación socioeconómica
- Validación cruzada de resultados



Análisis de Similitud

FuzzyWuzzy para consolidación inteligente de datos redundantes, eliminando duplicados y unificando entidades similares con >70% de coincidencia.

- Threshold de similitud: >70%
- Consolidación automática
- Preservación de información única



Pipeline ETL Robusto

Procesamiento escalable y reproducible con validación en cada etapa, diseñado para manejar volúmenes crecientes de datos.

- Modularidad y reutilización
- Logging
- Reutilización de código
- Escalabilidad y rendimiento

Conectemos



Correo:

doydurema67@gmail.com



LinkedIn:

Mi Perfil



Desarrollador:

Dody Salim Dueñas Remache