

El presente documento se entrega como parte de los requisitos formales del Proyecto Integrador, con el propósito de compilar la estructura, organización y desarrollo del proyecto. No obstante, su valor principal radica en el proceso formativo y en la experiencia adquirida al implementar una metodología de trabajo profesional en entornos de Data Science.

Durante el desarrollo del proyecto se incorporaron herramientas de control de versiones y colaboración, particularmente Git y GitHub, que permitieron gestionar el flujo de trabajo, mantener trazabilidad y documentar de manera ordenada cada avance. Esta práctica representa un primer acercamiento al uso real de repositorios como parte de un entorno académico y profesional.

El repositorio personal del proyecto se encuentra disponible en: <https://github.com/cfgarciac/Module-1>

Se extiende la invitación a revisarlo directamente en línea para visualizar correctamente la estructura y la renderización del archivo README.md, el cual refleja de forma integral la documentación técnica y analítica del proyecto.

## # Data Project – Conociendo al Cliente 360°

### ## Descripción General

**“Conociendo al Cliente 360°”** es un proyecto académico desarrollado con enfoque en análisis de datos e integración de información externa, cuyo propósito es construir una **visión integral del cliente** a partir de fuentes internas (base de clientes) y datos externos provenientes de la **API de Yelp**.

El proyecto aplica un flujo completo de ciencia de datos, desde la exploración y limpieza hasta el análisis visual e interpretativo, siguiendo una estructura modular basada en la metodología **\*Cookiecutter Data Science\***.

---

## ## Objetivos del Proyecto

- Analizar las características demográficas, económicas y de consumo de los clientes.
- Integrar información de la API de Yelp para enriquecer el contexto comercial local.
- Identificar patrones de gasto, frecuencia de visita, preferencias y hábitos de consumo.
- Formular conclusiones accionables que orienten estrategias de fidelización y marketing.

---

## ## Estructura del Proyecto

...

```
|— LICENSE
|— README.md
|
|— data
|   |— raw/ <- Datos originales (base de clientes)
|   |— external/ <- Datos descargados desde la API de Yelp
|   |— processed/ <- Datos limpios y consolidados
|   |— interim/ <- Datos intermedios durante el procesamiento
|
|— models/ <- Espacio reservado para modelos futuros
|
|— notebooks/
|   |— 1.0-cfg-EDA.ipynb <- Exploración y limpieza de datos (Avance 1)
|   |— 2.0-cfg-API_Yelp.ipynb <- Integración con la API de Yelp (Avance 2)
|   |— 3.0-cfg-Analisis_Interpretacion.ipynb <- Análisis visual e interpretación (Avance 3)
|
```

```
|— references/ <- Diccionarios, manuales y documentación de apoyo
|
|— reports/ <- Resultados y visualizaciones finales
|   |— figures/ <- Imágenes y gráficos generados
|
|— src/ <- Código fuente modular
|
|— config.py
|— dataset.py
|— features.py
|— plots.py
|— services/
|— init.py
...
---
```

## Configuración del Entorno

### Clonar el repositorio

```
git clone https://github.com/cfgarciac/Module-1.git
cd Module-1
```

### Crear entorno virtual

```
python -m venv .venv
```

### Activar entorno

```
..venv\Scripts\activate (en PowerShell / Windows)
```

### Instalar dependencias

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

---

## Desarrollo por Fases

- Avance 1 – EDA: análisis exploratorio y limpieza de la base de clientes. (Notebook: 1.0-cfg-EDA.ipynb)
- Avance 2 – Integración Yelp: conexión con la API, extracción y normalización de datos externos. (Notebook: 2.0-cfg-API\_Yelp.ipynb)
- Avance 3 – Análisis e interpretación: visualizaciones, correlaciones y conclusiones del comportamiento del cliente. (Notebook: 3.0-cfg-Analisis\_Interpretacion.ipynb)

---

## ## Principales Resultados

- Miami se posiciona como el mercado más rentable, con el mayor gasto promedio mensual y la mayor proporción de membresías premium (~53%).
- Los estratos alto y muy alto concentran los mayores niveles de gasto, sin requerir alta frecuencia de visita.  
Existe una correlación positiva ( $r = 0.64$ ) entre los ingresos mensuales y el gasto promedio en restaurantes.
- Las preferencias alimenticias se distribuyen principalmente entre Carnes (27%) y Vegetariano (23%), mostrando interés por opciones saludables.
- Los clientes premium son adultos con ingresos altos y hábitos de consumo selectivos, representando un segmento ideal para programas de fidelización.

---

## ## Conclusiones

- Miami, San Diego y Seattle destacan como los mercados con mayor potencial comercial.
- La frecuencia de visita no necesariamente determina el gasto: los clientes de alto poder adquisitivo priorizan la calidad sobre la cantidad.
- Las estrategias de negocio deben basarse en segmentación geográfica y socioeconómica.

- Las tendencias saludables representan oportunidades de crecimiento en todos los estratos.
- Se recomienda implementar análisis predictivos para anticipar el comportamiento de compra y la propensión a membresías premium.

---

## ## Próximos Pasos

- Integrar nuevas fuentes de datos (reseñas, clima, tendencias locales).
- Desarrollar modelos predictivos de gasto y fidelización.
- Construir dashboards interactivos con Power BI o Streamlit para la toma de decisiones.

---

## ## Tecnologías Utilizadas

- Lenguaje: Python (3.10+)
- Librerías: Pandas, NumPy, Matplotlib, Seaborn, Requests, Dotenv
- Entorno: Jupyter Notebooks
- Control de versiones: Git y GitHub

---

## ## Autor

Cristian García

Correo: [cfgarciac@unal.edu.co](mailto:cfgarciac@unal.edu.co)

LinkedIn: <https://www.linkedin.com/in/cfgarciac/>

Versión: 1.0 – Octubre 2025