

Etapas I

Exploración Visual de los Datos

Índice

- 01 [Objetivos](#)
- 02 [Contenidos y Herramientas Clave](#)
- 03 [Proceso de Exploración de Datos](#)
- 04 [Visualizaciones Clave](#)
- 05 [Entregables Esperados](#)
- 06 [Evaluación](#)



01

Objetivos

Objetivos

- **Familiarización con el Dataset:** El objetivo principal de esta etapa es que los estudiantes se familiaricen con los datos proporcionados. Comprenderán la estructura del dataset, identificarán las variables clave y explorarán las primeras tendencias y patrones.
- **Aplicación de Técnicas de EDA:** Se espera que los estudiantes utilicen técnicas de Análisis Exploratorio de Datos (EDA) para descubrir insights preliminares y preparar los datos para el análisis más profundo en las etapas posteriores.
- **Creación de Visualizaciones Iniciales:** Los estudiantes desarrollarán visualizaciones que les permitan interpretar los datos de manera más clara, utilizando librerías como Matplotlib y Seaborn.

02

Contenidos y Herramientas Clave

Contenidos y Herramientas Clave

Python y Librerías:

- **Pandas:** Para la manipulación de datos, carga de datasets, y operaciones básicas.
- **Matplotlib y Seaborn:** Para la creación de visualizaciones, como gráficos de líneas, histogramas, boxplots y scatter plots.

Conceptos Estadísticos:

- **Medidas de tendencia central** (media, mediana, moda), **medidas de dispersión** (rango, varianza, desviación estándar), y **correlación**.

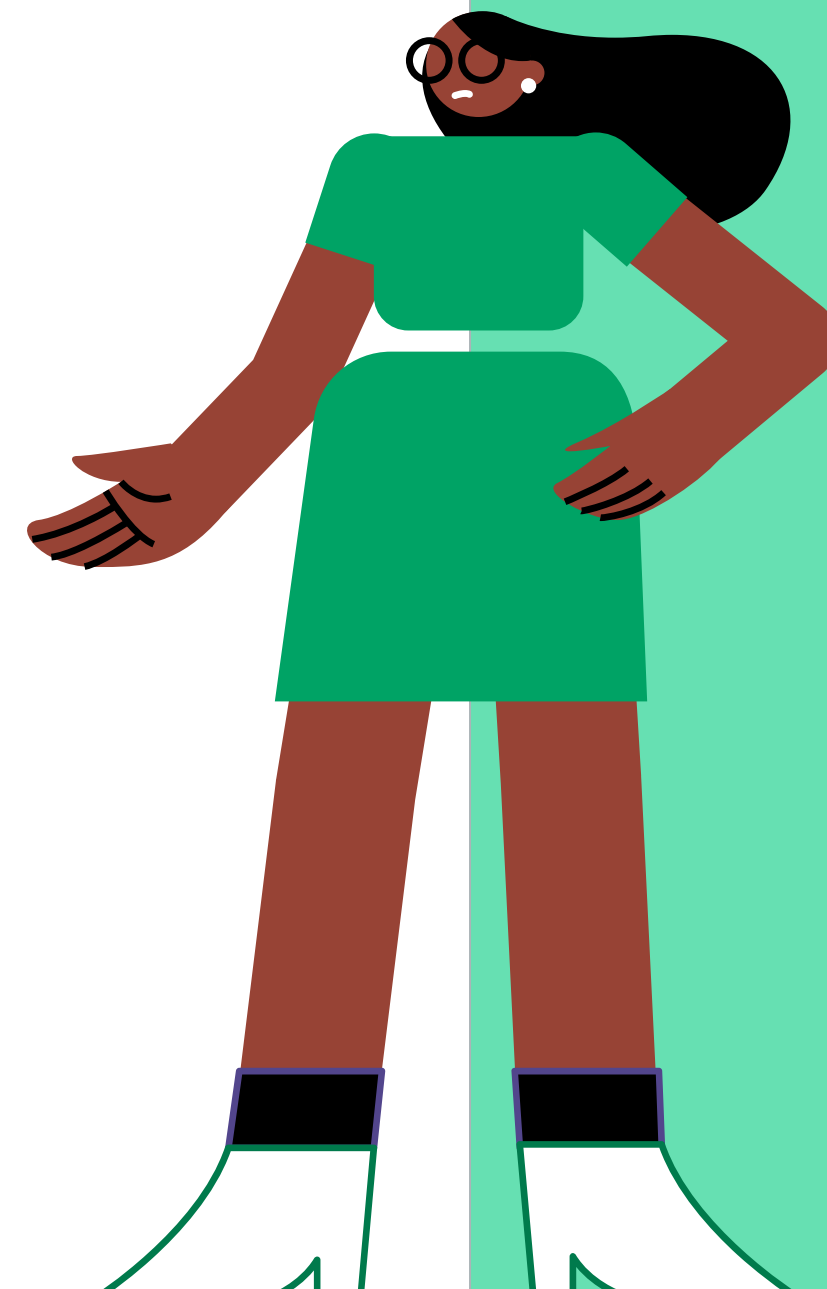
Tips

Documentación

Asegúrate de documentar cada paso del análisis, explicando qué técnicas estás utilizando y por qué.

Exploración Sistemática

Comienza con una visión general (descriptiva) de las variables y luego profundiza en aquellas que parecen más relevantes.



03

Proceso de Exploración de Datos

Proceso de Exploración de Datos

Carga del Dataset

- Importar el dataset utilizando Pandas.
- Verificación del tamaño, estructura, y tipos de datos de las columnas.

Limpieza Preliminar

- Identificación y manejo de valores faltantes.
- Eliminación de duplicados o corrección de errores básicos en los datos.

Análisis Descriptivo

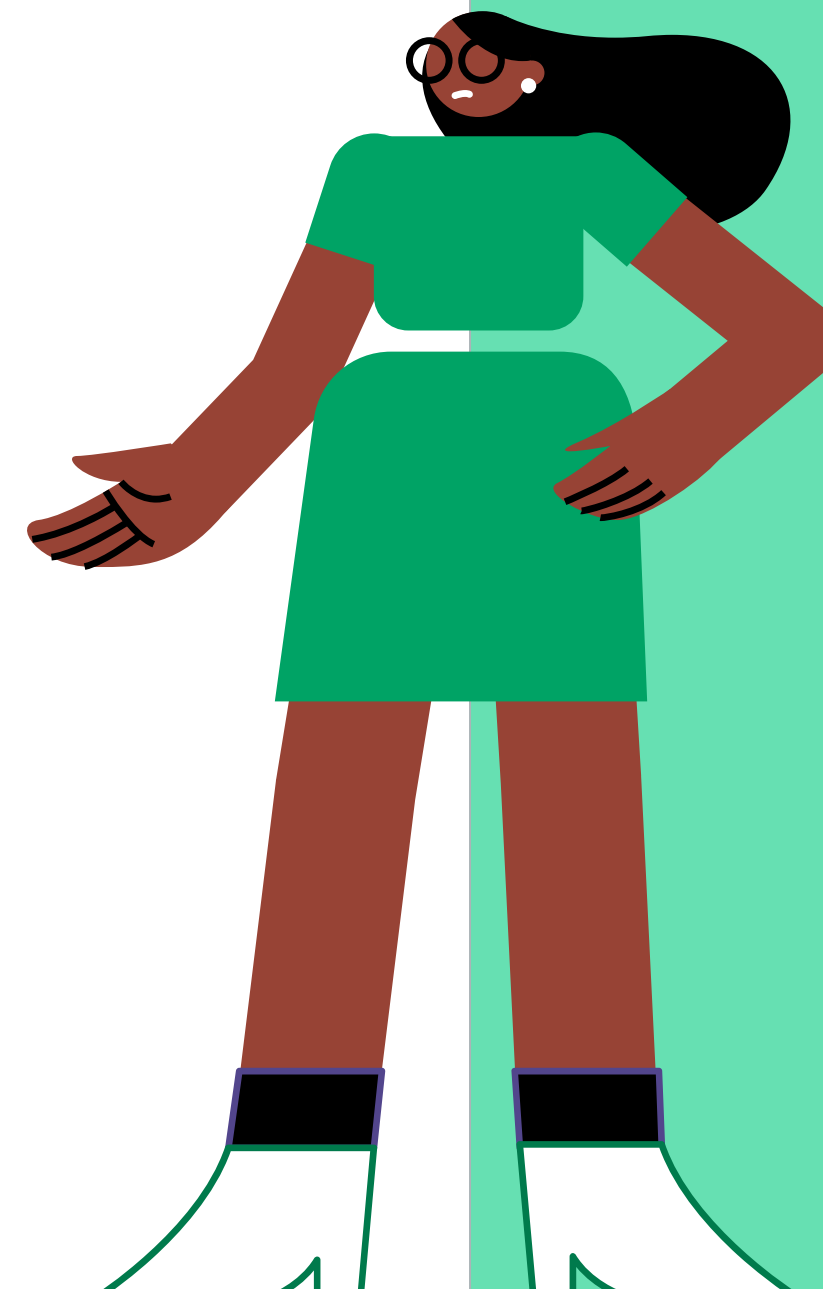
- Cálculo de estadísticas descriptivas para cada variable.
- Identificación de outliers y posibles anomalías.

Visualización de Patrones y Relaciones

- Creación de gráficos para visualizar relaciones entre variables.
- Análisis de correlaciones utilizando una matriz de correlación.

Tips

- Utiliza `df.info()` y `df.describe()` para una visión rápida del dataset.
- Usa `df.isnull().sum()` para identificar rápidamente las columnas con valores nulos.
- Utiliza visualizaciones como **histogramas** para entender la distribución de variables numéricas.
- Usa **scatter plots** para ver correlaciones entre variables numéricas, y **boxplots** para analizar distribuciones según categorías.



04

Visualizaciones Clave

Visualizaciones Clave

Histogramas

Útiles para entender la distribución de una variable numérica.

Boxplots

Ideales para identificar outliers y analizar la dispersión de los datos.

Scatter Plots

Para explorar la relación entre dos variables numéricas.

Heatmaps de Correlación

Para visualizar relaciones entre múltiples variables numéricas.

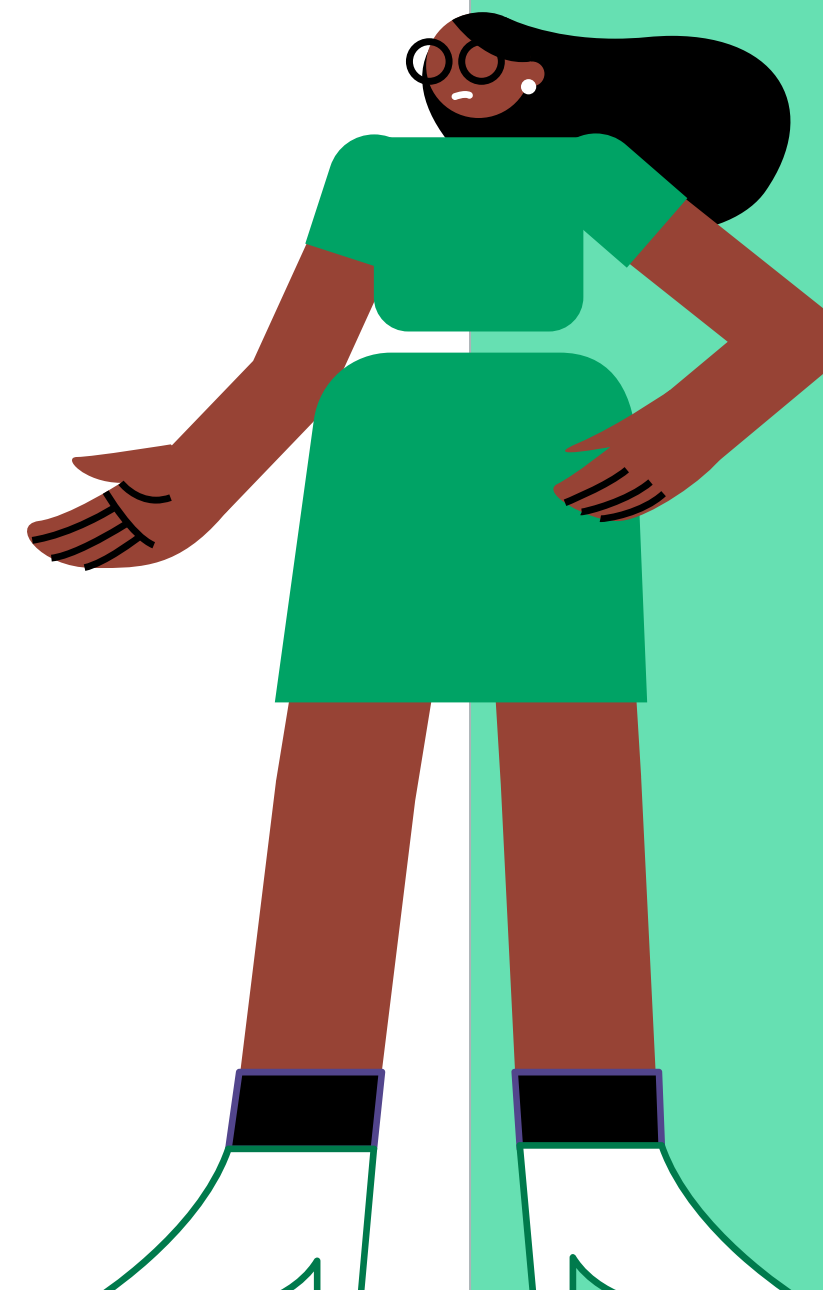
Tips

Claridad en Visualizaciones

Asegúrate de que cada visualización sea clara y fácil de interpretar. Añade **títulos**, **etiquetas de ejes**, y **leyendas** *cuando sea necesario*.

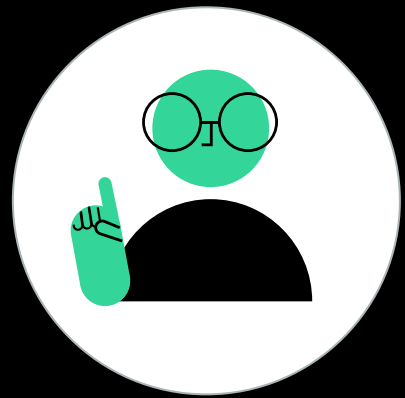
Exploración Iterativa

No te limites a una sola visualización; prueba diferentes formas de representar los datos para descubrir insights ocultos.



05

Entregables Esperados



Entregables Esperados

01

Análisis Exploratorio de Datos (EDA)

02

Scripts de Código en Python

03

Visualizaciones

04

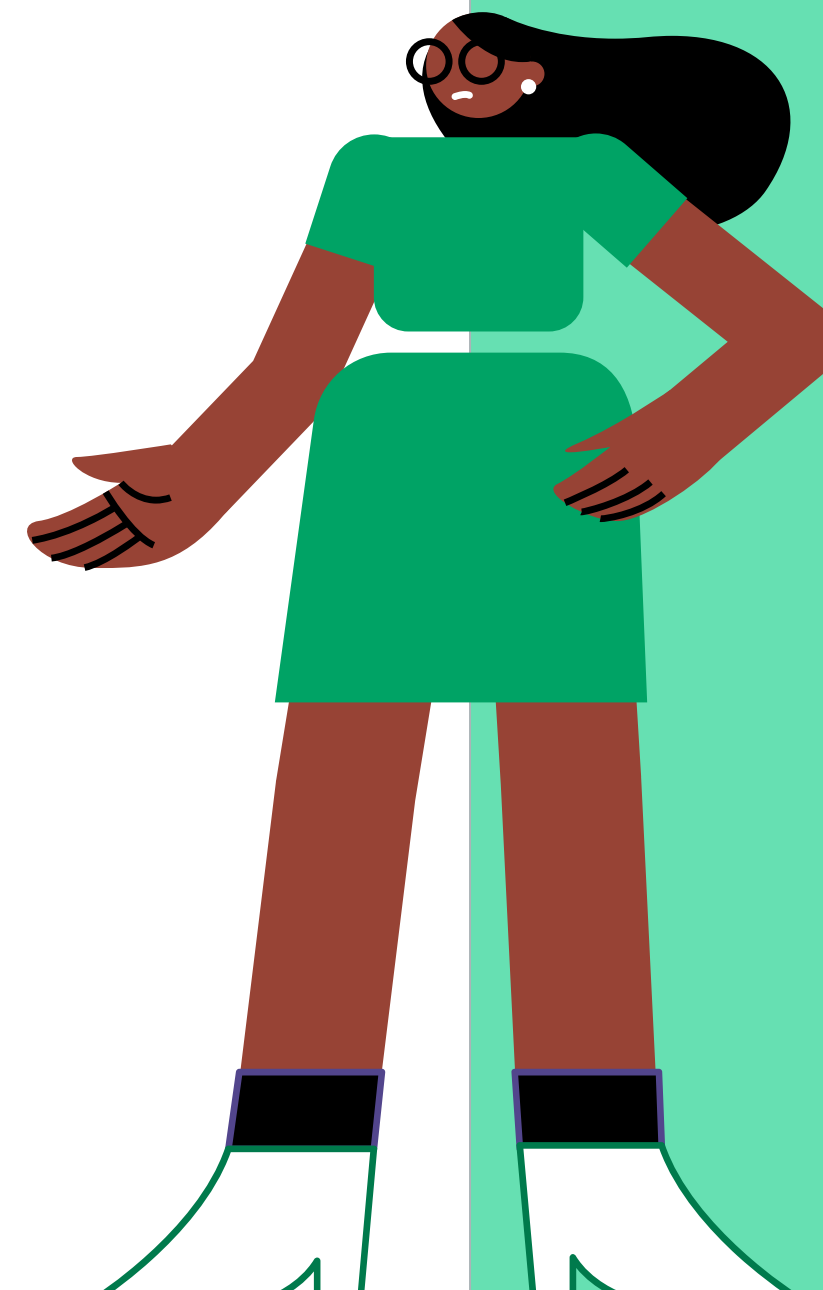
Documentación del Proceso

Análisis Exploratorio de Datos (EDA)

Documento que detalle los hallazgos iniciales, incluyendo estadísticas descriptivas, identificación de outliers, y cualquier patrón o tendencia relevante.

Tips

Explica en detalle cada paso, incluyendo las decisiones que tomaste y por qué. Incluye visualizaciones para respaldar tus hallazgos.

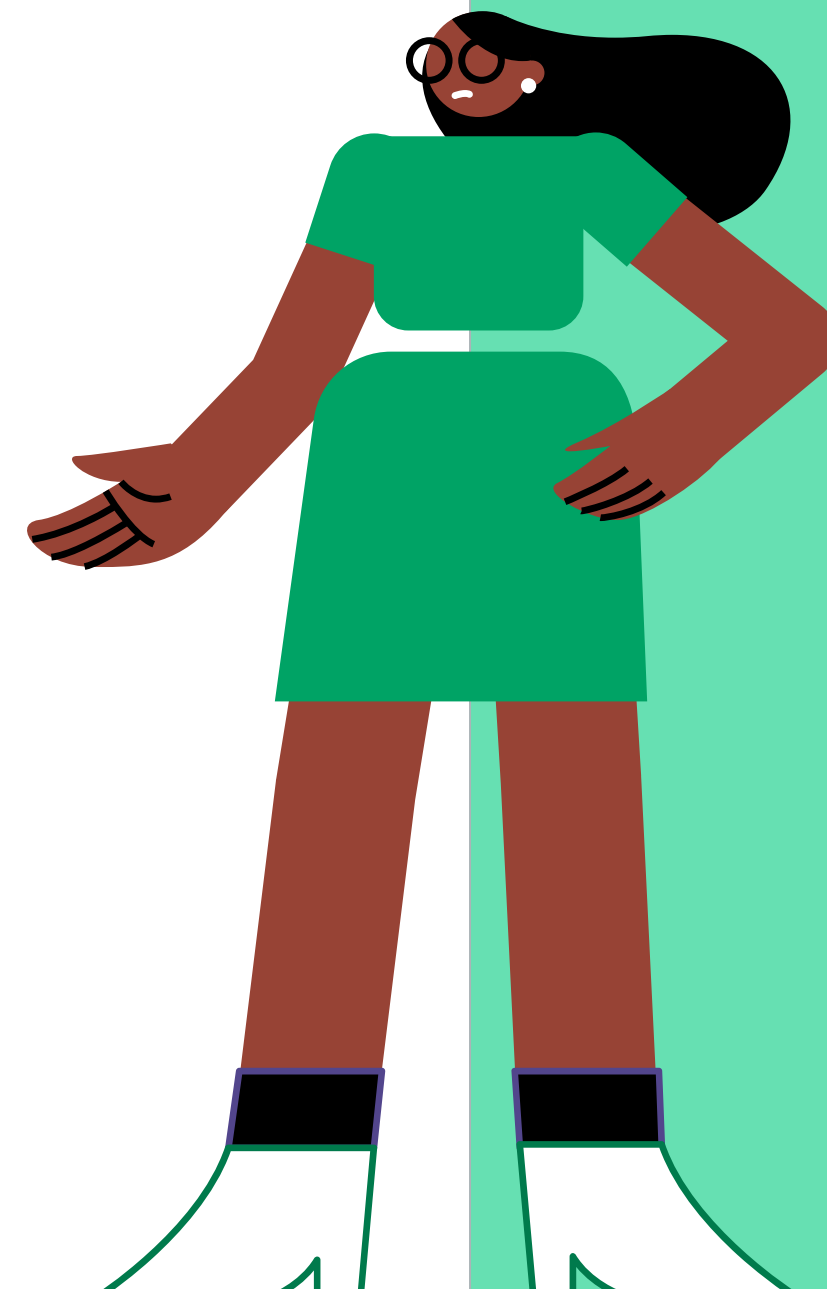


Scripts de Código en Python

Código utilizado para cargar, manipular, y explorar el dataset.

Tips

Asegúrate de que tu código esté bien organizado y comentado, facilitando su comprensión y reproducibilidad.

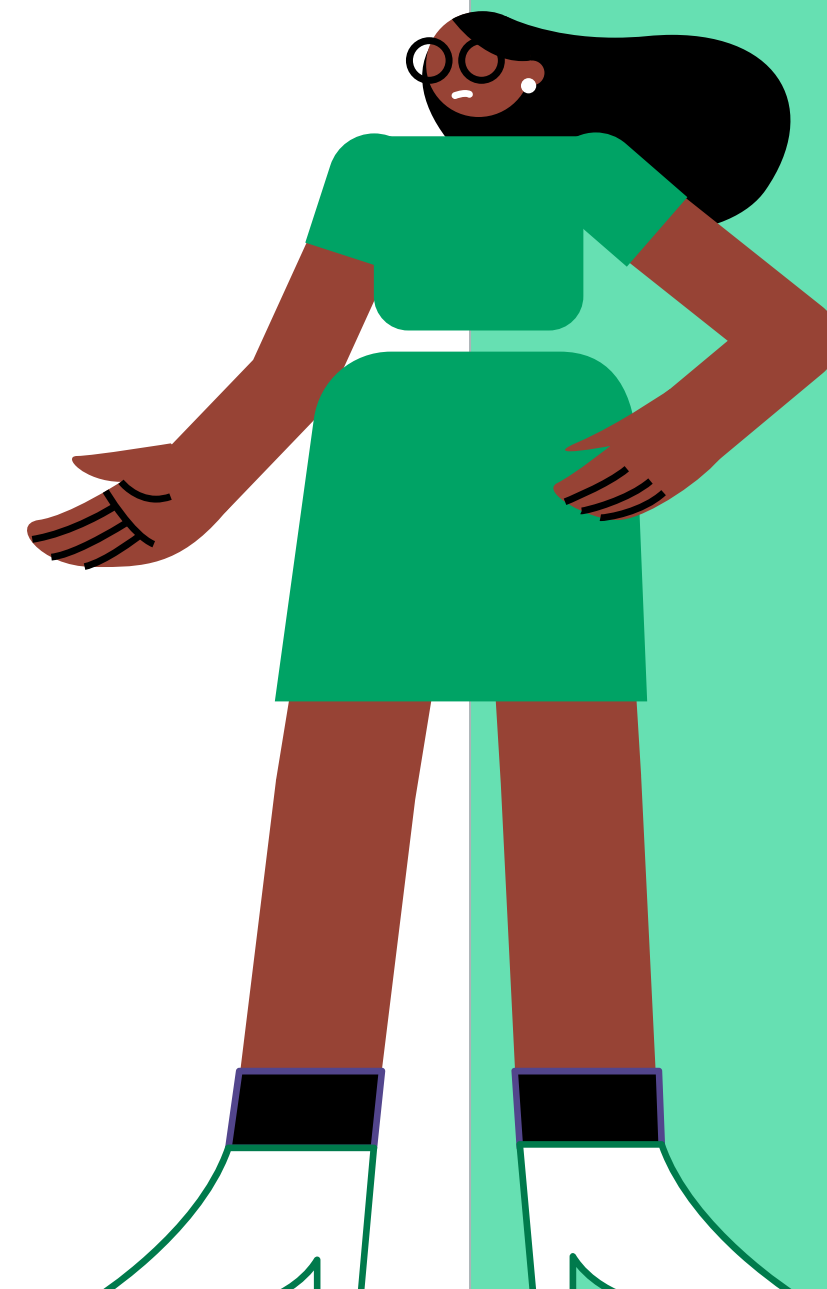


Visualizaciones

Conjunto de gráficos y visualizaciones creados durante el análisis.

Tips

Presenta las visualizaciones más relevantes y acompáñalas con una breve explicación de lo que representan y por qué son importantes.

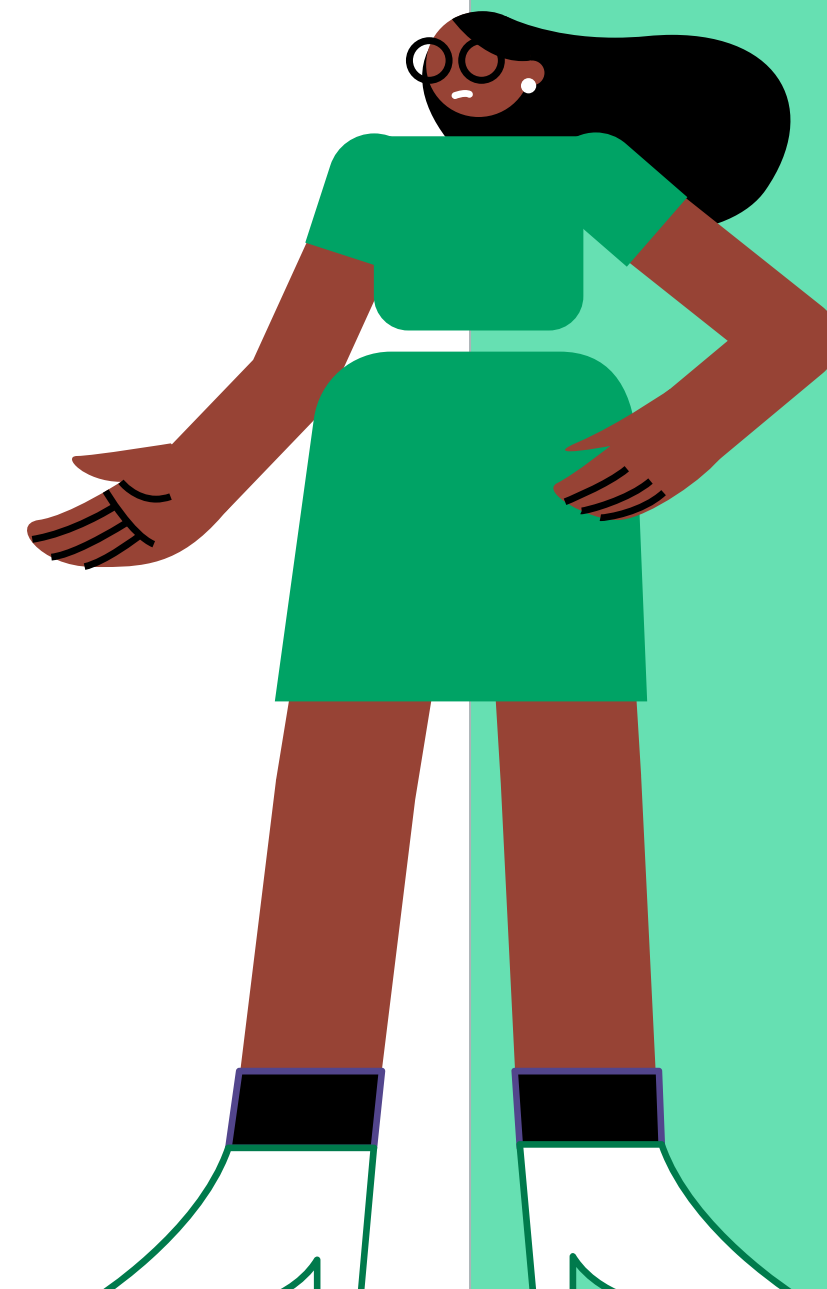


Documentación del Proceso

Un documento que detalle el proceso seguido, las técnicas utilizadas, y las justificaciones detrás de cada decisión tomada durante la exploración de datos.

Tips

Mantén una narrativa coherente, guiando al lector a través de tu proceso de pensamiento desde la carga del dataset hasta las conclusiones iniciales.



06

Evaluación

Criterios de Evaluación

- Profundidad y exhaustividad del análisis exploratorio.
- Calidad y claridad de las visualizaciones.
- Eficiencia en el uso de herramientas (Pandas, Matplotlib, Seaborn).
- Documentación y justificación de las decisiones tomadas durante el análisis.



Después de completar la Etapa 1, estarás listo para profundizar en la limpieza y transformación de los datos en la Etapa 2. ¡Sigue adelante y no dudes en volver a revisar tus hallazgos iniciales si encuentras algo nuevo en las próximas etapas! ✨



La exploración de datos es el primer paso crucial para cualquier proyecto de análisis de datos. Tómate tu tiempo para entender los datos en profundidad, ya que esto sentará las bases para el éxito en las etapas posteriores.



¡Muchas gracias!