

Análisis de datos

Nivel Básico – Explorador

Misión 1

1. Cargar Datos

a. Desde un Archivo CSV

```
import pandas as pd

# Cargar datos desde un archivo CSV
df = pd.read_csv('archivo.csv')

# Mostrar las primeras filas del DataFrame
print(df.head())
```

b. Desde un Archivo Excel

```
# Cargar datos desde un archivo Excel
df = pd.read_excel('archivo.xlsx')

# Mostrar las primeras filas del DataFrame
print(df.head())
```

c. Desde una Base de Datos SQL

```
import sqlite3

# Conectar a la base de datos SQLite
conn = sqlite3.connect('base_de_datos.db')
```

```
# Cargar datos desde una consulta SQL
query = 'SELECT * FROM tabla'
df = pd.read_sql_query(query, conn)

# Mostrar las primeras filas del DataFrame
print(df.head())
```

2. Explorar los Datos

a. Verificar Información Básica

```
# Verificar las primeras y últimas filas
print(df.head())
print(df.tail())

# Verificar el tamaño del DataFrame
print(df.shape)

# Obtener información sobre los tipos de datos y valores nulos
print(df.info())

# Ver estadísticas descriptivas
print(df.describe())
```

b. Seleccionar y Filtrar Datos

```
# Seleccionar una columna
column = df['nombre_columna']

# Seleccionar múltiples columnas
subset = df[['columna1', 'columna2']]

# Filtrar filas con una condición
filtered_df = df[df['columna'] > valor]
```

```
# Filtrar filas con múltiples condiciones
filtered_df = df[(df['columna1'] > valor1) & (df['columna2'] < valor2)]
```

c. Agrupar Datos y Aplicar Agregaciones

```
# Agrupar datos por una columna y calcular estadísticas agregadas
grouped = df.groupby('columna').agg({
    'columna_numerica': ['mean', 'sum'],
    'otra_columna': 'count'
})

print(grouped)
```

3. Manipular Datos

a. Añadir o Eliminar Columnas

```
# Añadir una nueva columna
df['nueva_columna'] = valor

# Eliminar una columna
df = df.drop(columns=['columna_a_eliminar'])
```

b. Modificar Valores

```
# Modificar valores en una columna
df.loc[df['columna'] == valor_antiguo, 'columna'] = nuevo_valor
```

c. Lidar con Valores Nulos

```
# Verificar valores nulos
print(df.isnull().sum())

# Eliminar filas con valores nulos
df = df.dropna()
```

```
# Rellenar valores nulos con un valor específico
df = df.fillna(valor)
```

4. Transformar Datos

a. Aplicar Funciones a las Columnas

```
# Aplicar una función a una columna
df['columna'] = df['columna'].apply(funcion)

# Aplicar una función a varias columnas
df[['columna1', 'columna2']] = df[['columna1', 'columna2']].apply(funcion)
```

b. Reordenar Filas y Columnas

```
# Reordenar columnas
df = df[['columna2', 'columna1', 'columna3']]

# Ordenar filas por una columna
df = df.sort_values(by='columna', ascending=False)
```

5. Visualizar Datos

a. Visualización Básica con Pandas

```
import matplotlib.pyplot as plt

# Histograma
df['columna'].hist()
plt.show()

# Gráfico de dispersión
df.plot.scatter(x='columna1', y='columna2')
```

```
plt.show()
```

b. Gráficos Avanzados con Seaborn

```
import seaborn as sns

# Gráfico de caja
sns.boxplot(x='columna1', y='columna2', data=df)
plt.show()

# Gráfico de pares
sns.pairplot(df)
plt.show()
```

6. Guardar Datos

a. Guardar en un Archivo CSV

```
df.to_csv('archivo_modificado.csv', index=False)
```

b. Guardar en un Archivo Excel

```
df.to_excel('archivo_modificado.xlsx', index=False)
```