

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


# Step 1: Extract - Load the dataset from Excel

df = pd.read_excel(Google Apps data.xlsx')


# Step 2: Transform - Clean and preprocess the data
# Example: Handle missing values, convert data types

df.dropna(inplace=True) # Drop rows with missing values

df['Date'] = pd.to_datetime(df['Date']) # Convert 'Date' to datetime format


# Step 3: Load - Save the cleaned dataset to Excel

# Replace 'cleaned_dataset.xlsx' with the desired output file name and path

df.to_excel('cleaned_dataset.xlsx', index=False)


# Step 4: Exploratory Data Analysis (EDA)
# Example: Basic statistics and visualizations

print(df.info())

print(df.describe())


# Example: Visualize the distribution of Android versions

plt.figure(figsize=(10, 6))

sns.countplot(x='Android_Version', data=df, order=df['Android_Version'].value_counts().index)

plt.title('Distribution of Android Versions')

plt.xticks(rotation=45)

plt.show()


# Step 5: Key Metrics
# Example: Calculate average daily app usage

average_daily_usage = df['App_Usage'].mean()
```

```
print(f'Average Daily App Usage: {average_daily_usage:.2f} minutes')
```

```
# Example: Calculate retention rate
```

```
total_users = df['User_ID'].nunique()
```

```
users_retained = df[df['Retention'] == 1]['User_ID'].nunique()
```

```
retention_rate = (users_retained / total_users) * 100
```

```
print(f'Retention Rate: {retention_rate:.2f}%')
```