Explore and pre-process Data

```
Exploring Data:
# Get basic inform ation about the DataFrame
df.info()
# Summary statistics for numerical columns
df.des cribe()
# Number of unique values in each column
df.nun ique()
# Count missing values in each column
df.isn ull ().s um()
# Remove duplicate rows
df = df.dro p_d upl ica tes()
# Drop columns with missing values
df = df.dro pna (ax is=1)
# Fill missing values with a specific value
df['co lum n_n ame '].f il lna (value, inplac e=True)
# Drop rows with missing values
df = df.dro pna()
# Replace values in a column
df['co lum n_n ame '].r ep lac e({ old _value: new_va lue}, inplac e=True)
# Convert data types
df['co lum n_n ame'] = df['co lum n_n ame '].a st ype ('n ew_ dat a_t ype')
# Rename columns
df.ren ame (co lum ns= {'o ld_ col umn _name': 'new_c olu mn_ name'}, inplac e=True)
# Filter rows based on a condition
filter ed_df = df[df[ 'co lum n_n ame'] > value]
# Multiple conditions
filter ed_df = df[(df ['c olu mn1'] > value1) & (df['c olu mn2'] < value2)]
# Select specific columns
select ed_ col umns_df = df[['c olu mn1', 'colum n2']]
# Sorting the DataFrame
df.sor t_v alu es( by= 'co lum n_n ame', ascend ing =False, inplac e=True)
# Create a new column based on existing columns
df['ne w_c olumn'] = df['co lumn1'] + df['co lumn2']
# Apply a function to a column
df['ne w_c olumn'] = df['ex ist ing _co lum n'].ap ply (lambda x: your_f unc tio n(x))
```

Analyse

# Plotting histograms	# Box plot	# Scatter plot	# line graph

Sponsored by **Readable.com**
Measure your website readability!
https://readable.com

Analyse (cont)

df['column_name'].hist()	sns.boxplot(x='column1', y='column2', data=df)	plt.scatter(df['column1'], df['column2']) plt.xlabel('Column1') plt.ylabel('Column2') plt.title('Scatter Plot') plt.show()	plt.plot(df['x'], df['y']) plt.title('Sample Line Graph') plt.xlabel('X-axis label') plt.ylabel('Y-axis label') plt.show()

First

# import all needed libraries	# Load data from a CSV file	# Display the first few rows of the DataFrame	# Get basic information about the DataFrame	# Summary statistics for numerical columns
import pandas as pd import matplotlib.pyplot as plt import seaborn as sns	df = pd.read_csv('your_file.csv')	df.head()	df.info()	df.describe()

Sponsored by **Readable.com**
Measure your website readability!
https://readable.com