

### Explore and pre-process Data

```

Exploring Data:
# Get basic information about the DataFrame
df.info()
# Summary statistics for numerical columns
df.describe()
# Number of unique values in each column
df.nunique()
# Count missing values in each column
df.isnull().sum()
# Remove duplicate rows
df = df.drop_duplicates()
# Drop columns with missing values
df = df.dropna(axis=1)
# Fill missing values with a specific value
df['column_name'].fillna(value, inplace=True)
# Drop rows with missing values
df = df.dropna()
# Replace values in a column
df['column_name'].replace({old_value: new_value}, inplace=True)
# Convert data types
df['column_name'] = df['column_name'].astype('new_data_type')
# Rename columns
df.rename(columns={'old_column_name': 'new_column_name'}, inplace=True)
# Filter rows based on a condition
filtered_df = df[df['column_name'] > value]
# Multiple conditions
filtered_df = df[(df['column1'] > value1) & (df['column2'] < value2)]
# Select specific columns
selected_columns_df = df[['column1', 'column2']]
# Sorting the DataFrame
df.sort_values(by='column_name', ascending=False, inplace=True)
# Create a new column based on existing columns
df['new_column'] = df['column1'] + df['column2']
# Apply a function to a column
df['new_column'] = df['existing_column'].apply(lambda x: your_function(x))

```

### Analyse

# Plotting histograms

# Box plot

# Scatter plot

# line graph



By **prettyhatcobb**

Not published yet.

Last updated 5th January, 2024.

Page 1 of 3.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>

### Analyse (cont)

```
df['column_n-ame'].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()
```



By **prettyhatcobb**

[cheatography.com/prettyhatcobb/](https://cheatography.com/prettyhatcobb/)

Not published yet.

Last updated 5th January, 2024.

Page 3 of 3.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>

