

"""

## Pandas Data Analysis Project

Author: Your Name

Description:

This script demonstrates loading, cleaning,  
filtering, grouping, and analyzing CSV data.

"""

```
import pandas as pd
```

```
# 1. Load Dataset
```

```
try:
```

```
    df = pd.read_csv("students.csv")
```

```
    print("Dataset Loaded Successfully!\n")
```

```
except FileNotFoundError:
```

```
    print("CSV file not found.")
```

```
    exit()
```

```
# 2. Inspect Dataset
```

```
print("First 5 Rows:\n", df.head())
```

```
print("\nDataset Info:")
```

```
print(df.info())
```

```
print("\nSummary Statistics:\n", df.describe())
```

```
# 3. Data Cleaning
```

```
# Convert Marks column to numeric (handles NaN properly)
```

```
df["Marks"] = pd.to_numeric(df["Marks"], errors='coerce')
```

```
# Fill missing marks with average marks
```

```
average_marks = df["Marks"].mean()
```

```
df["Marks"].fillna(average_marks, inplace=True)
```

```
print("\nMissing values handled successfully.")
```

```
# 4. Filtering
```

```
# Students with Marks > 85
```

```
high_scorers = df[df["Marks"] > 85]
```

```
print("\nStudents with Marks > 85:")
```

```
print(high_scorers)
```

```
# 5. Grouping & Aggregation
```

```
department_avg = df.groupby("Department")["Marks"].mean()
```

```
print("\nAverage Marks by Department:")
```

```
print(department_avg)
```

```
print("\nData Analysis Completed Successfully!")
```

## Github repository

<https://github.com/Karan6165/Alfido-Tech-Internship.git>

The screenshot shows a Jupyter Notebook interface with a dark theme. The top bar includes File, Edit, Selection, View, Go, Run, Terminal, Help, and a Search field. The left sidebar contains icons for file operations like New, Open, Save, and a search bar. The main area is titled "Untitled-1" and displays the following Python code:

```
1 """
2 Pandas Data Analysis Project
3 Author: Your Name
4 Description:
5 This script demonstrates loading, cleaning,
6 filtering, grouping, and analyzing CSV data.
7 """
8
9 import pandas as pd
10
11 # 1. Load Dataset
12 try:
13     df = pd.read_csv("students.csv")
14     print("Dataset Loaded Successfully!\n")
15
16 except FileNotFoundError:
17     print("CSV file not found.")
18     exit()
19
20 # 2. Inspect Dataset
21
22 print("First 5 Rows:\n", df.head())
23 print("\nDataset Info:")
24 print(df.info())
25
26 print("\nSummary Statistics:\n", df.describe())
27
28 # 3. Data Cleaning
29
30 # Convert Marks column to numeric (handles NaN properly)
31 df["Marks"] = pd.to_numeric(df["Marks"], errors='coerce')
32
33 # Fill missing marks with average marks
34 average_marks = df["Marks"].mean()
35 df["Marks"].fillna(average_marks, inplace=True)
36
37 print("\nMissing values handled successfully.")
38
39 # 4. Filtering
40
41 # Students with Marks > 85
42 high_scorers = df[df["Marks"] > 85]
43
44 print("\nStudents with Marks > 85:")
45 print(high_scorers)
46
47 # 5. Grouping & Aggregation
48 |
49 department_avg = df.groupby("Department")["Marks"].mean()
50
51 print("\nAverage Marks by Department:")
52 print(department_avg)
53
54 print("\nData Analysis Completed Successfully!")
```

File Edit Selection View Go Run Terminal Help

analysis\_with\_pandas.py students.csv X

D: > project > alfido tasks > task 3 > students.csv

	ID	Name	Department	Marks
1	1	Karan	CSE	85
2	2	Rahul	IT	90
3	3	Sneha	CSE	
4	4	Priya	ECE	78
5	5	Amit	CSE	88
6	6	Neha	IT	NaN
7	7	Rohit	ECE	91

PROBLEMS OUTPUT DEBUG CONSOLE PORTS TERMINAL

PS D:\project\alfido tasks\task 3> & C:\Users\karan\AppData\Local\Programs\Python\Python313\python.exe "d:/project/alfido tasks/task 3/analysis.py"

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7 entries, 0 to 6
Data columns (total 4 columns):
 #   Column   Non-Null Count  Dtype  
--- 
 0   ID       7 non-null      int64  
 1   Name     7 non-null      object 
 2   Department 7 non-null    object  
 3   Marks    5 non-null      float64 
dtypes: float64(1), int64(1), object(2)
memory usage: 356.0+ bytes
None

Summary Statistics:
   ID      Marks
count 7.000000 5.000000
mean 4.000000 86.000000
std 2.160247 5.224940
min 1.000000 78.000000
25% 2.500000 85.000000
50% 4.000000 88.000000
75% 5.500000 90.000000
max 7.000000 91.000000
d:\project\alfido tasks\task 3\analysis.py:40: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

df["Marks"].fillna(average_marks, inplace=True)

Missing values handled successfully.

Students with Marks > 85:
   ID  Name Department  Marks
1   2  Rahul      IT    90.0
2   3  Sneha      CSE   86.4
4   5  Amit      CSE   88.0
5   6  Neha      IT    86.4
6   7  Rohit     ECE   91.0

Average Marks by Department:
Department
CSE    86.466667
ECE    84.500000
IT     88.200000
Name: Marks, dtype: float64

Data Analysis Completed Successfully!
PS D:\project\alfido tasks\task 3>
```

Do you want to install the recommended 'Rainbow CSV' extension from mechatroner for students.csv?

Install Show Recommendations

PROBLEMS OUTPUT DEBUG CONSOLE PORTS TERMINAL

```
PS D:\project\alfido tasks\task 3> & C:\Users\karan\AppData\Local\Programs\Python\Python313\python.exe "d:/project/alfido_tasks/task 3/analysis.py"
2 3 Sneha      CSE  86.4
4 5 Amit       CSE  88.0
5 6 Neha       IT   86.4
6 7 Rohit      ECE  91.0

Average Marks by Department:
Department
CSE    86.466667
ECE    84.500000
IT     88.200000
Name: Marks, dtype: float64

Data Analysis Completed Successfully!
PS D:\project\alfido tasks\task 3>
```