

ICP3 REPORT

+ Code + Text

```
import pandas as pd
import numpy as np
data = {
    'ID': np.arange(1, 1000001),
    'Value': np.random.rand(1000000),
    'Category': np.random.choice(['A', 'B', 'C', 'D'], size=1000000)
}
df = pd.DataFrame(data)
print("First 10 rows of the DataFrame:")
print(df.head(10))
print("\ninformation of 'Value' column:")
print(df['Value'].describe())
```

First 10 rows of the DataFrame:

	ID	Value	Category
0	1	0.937089	D
1	2	0.617650	C
2	3	0.937451	A
3	4	0.482242	C
4	5	0.932629	B
5	6	0.121914	A
6	7	0.542958	A
7	8	0.943599	C
8	9	0.043794	C
9	10	0.655863	D

information of 'Value' column:

count	1000000.000000
mean	0.499845
std	0.288605
min	0.000001
25%	0.249888
50%	0.499710
75%	0.749561
max	0.999999

Name: Value, dtype: float64

```
import pandas as pd
import numpy as np

data = {
    'ID': np.arange(1, 1000001),
    'Value': np.random.rand(1000000),
    'Category': np.random.choice(['A', 'B', 'C', 'D'], size=1000000)
}
df = pd.DataFrame(data)
df.columns = ['ID number', 'Random value', 'Choice']
print("First 5 rows of the modified DataFrame:")
print(df.head())
```

⇒ First 5 rows of the modified DataFrame:

	ID number	Random value	Choice
0	1	0.050753	A
1	2	0.523959	D
2	3	0.847530	A
3	4	0.311173	B
4	5	0.977262	C

+ Code + Text

0s



```
import pandas as pd
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
student_data = pd.DataFrame({
    'school_code': ['s001', 's002', 's003', 's004', 's002', 's005'],
    'class': ['V', 'V', 'VI', 'VI', 'V', 'VI'],
    'name': ['Alberto Franco', 'Gino Mcneill', 'Ryan Parkes', 'Eesha Hinton', 'Gino Mcneill', 'David Parkes'],
    'date_Of_Birth': ['15/05/2002', '17/05/2002', '16/02/1999', '25/09/1998', '11/05/2002', '15/09/1997'],
    'age': [12, 12, 13, 13, 14, 12],
    'height': [173, 192, 186, 167, 151, 159],
    'weight': [35, 32, 33, 30, 31, 32],
    'address': ['street1', 'street2', 'street3', 'street4', 'street2', 'street5']
}, index=['S1', 'S2', 'S3', 'S4', 'S5', 'S6'])
print("Original DataFrame:")
print(student_data)
print("\nSplit the data on 'school_code' and 'class' wise:")
result = student_data.groupby(['school_code', 'class'])
for name, group in result:
    print("\nGroup:")
    print(name)
    print(group)
```



Original DataFrame:

	school_code	class	name	date_Of_Birth	age	height	weight	\
S1	s001	V	Alberto Franco	15/05/2002	12	173	35	
S2	s002	V	Gino Mcneill	17/05/2002	12	192	32	
S3	s003	VI	Ryan Parkes	16/02/1999	13	186	33	
S4	s004	VI	Eesha Hinton	25/09/1998	13	167	30	
S5	s002	V	Gino Mcneill	11/05/2002	14	151	31	
S6	s005	VI	David Parkes	15/09/1997	12	159	32	

	address
S1	street1
S2	street2
S3	street3
S4	street4

✓
0s

[3] S4 street4
S5 street2
S6 street5

Split the data on 'school_code' and 'class' wise:

Group:

('s001', 'V')

	school_code	class	name	date_Of_Birth	age	height	weight	\
S1	s001	V	Alberto Franco	15/05/2002	12	173	35	
			address					
S1			street1					

Group:

('s002', 'V')

	school_code	class	name	date_Of_Birth	age	height	weight	address
S2	s002	V	Gino Mcneill	17/05/2002	12	192	32	street2
S5	s002	V	Gino Mcneill	11/05/2002	14	151	31	street2

Group:

('s003', 'VI')

	school_code	class	name	date_Of_Birth	age	height	weight	address
S3	s003	VI	Ryan Parkes	16/02/1999	13	186	33	street3

Group:

('s004', 'VI')

	school_code	class	name	date_Of_Birth	age	height	weight	address
S4	s004	VI	Eesha Hinton	25/09/1998	13	167	30	street4

Group:

('s005', 'VI')

	school_code	class	name	date_Of_Birth	age	height	weight	address
S6	s005	VI	David Parkes	15/09/1997	12	159	32	street5

5s

pip install pandas

Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: numpy<2,>=1.22.4 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages

2s

```
[5] from google.colab import drive
    drive.mount('/content/drive')
```

Mounted at /content/drive

0s

```
[10] filepath = ('/content/drive/MyDrive/data.csv')
      with open(filepath, 'r') as file:
          content = file.read()
          print(content)
```

Duration,Pulse,Maxpulse,Calories

60,110,130,409.1
60,117,145,479.0
60,103,135,340.0
45,109,175,282.4
45,117,148,406.0
60,102,127,300.0
60,110,136,374.0
45,104,134,253.3
30,109,133,195.1
60,98,124,269.0
60,103,147,329.3
60,100,120,250.7
60,106,128,345.3
60,104,132,379.3
60,98,123,275.0
60,98,120,215.2
60,100,120,300.0

✓
0s



```
60,103,147,329.3
60,100,120,250.7
60,106,128,345.3
60,104,132,379.3
60,98,123,275.0
60,98,120,215.2
60,100,120,300.0
45,90,112,
60,103,123,323.0
45,97,125,243.0
60,108,131,364.2
45,100,119,282.0
60,130,101,300.0
45,105,132,246.0
60,102,126,334.5
60,100,120,250.0
60,92,118,241.0
60,103,132,
60,100,132,280.0
60,102,129,380.3
60,92,115,243.0
45,90,112,180.1
60,101,124,299.0
60,93,113,223.0
60,107,136,361.0
60,114,140,415.0
60,102,127,300.0
60,100,120,300.0
60,100,120,300.0
45,104,129,266.0
45,90,112,180.1
60,98,126,286.0
60,100,122,329.4
60,111,138,400.0
60,111,131,397.0
60,99,119,273.0
60,109,153,387.6
45,111,136,300.0
45,108,129,298.0
60.111.139.397.6
```

✓
0s



```
100,109,135,853.0  
45,118,141,341.0  
20,110,130,131.4  
180,90,130,800.4  
150,105,135,873.4  
150,107,130,816.0  
20,106,136,110.4  
300,108,143,1500.2  
150,97,129,1115.0  
60,109,153,387.6  
90,100,127,700.0  
150,97,127,953.2  
45,114,146,304.0  
90,98,125,563.2  
45,105,134,251.0  
45,110,141,300.0  
120,100,130,500.4  
270,100,131,1729.0  
30,159,182,319.2  
45,149,169,344.0  
30,103,139,151.1  
120,100,130,500.0  
45,100,120,225.3  
30,151,170,300.0  
45,102,136,234.0  
120,100,157,1000.1  
45,129,103,242.0  
20,83,107,50.3  
180,101,127,600.1  
45,107,137,  
30,90,107,105.3  
15,80,100,50.5  
20,150,171,127.4  
20,151,168,229.4  
30,95,128,128.2  
25,152,168,244.2  
30,109,131,188.2  
90,93,124,604.1  
20,95,112,77.7  
90,90,110,500.0
```


✓
0s



60,102,124,325.2
45,107,124,275.0
15,124,139,124.2
45,100,120,225.3
60,108,131,367.6
60,108,151,351.7
60,116,141,443.0
60,97,122,277.4
60,105,125,
60,103,124,332.7
30,112,137,193.9
45,100,120,100.7
60,119,169,336.7
60,107,127,344.9
60,111,151,368.5
60,98,122,271.0
60,97,124,275.3
60,109,127,382.0
90,99,125,466.4
60,114,151,384.0
60,104,134,342.5
60,107,138,357.5
60,103,133,335.0
60,106,132,327.5
60,103,136,339.0
20,136,156,189.0
45,117,143,317.7
45,115,137,318.0
45,113,138,308.0
20,141,162,222.4
60,108,135,390.0
60,97,127,
45,100,120,250.4
45,122,149,335.4
60,136,170,470.2
45,106,126,270.8
60,107,136,400.0
60,112,146,361.9
30,103,127,185.0
60,110,150,409.4
50,150,134,545.0

```
✓ [14] filepath = '/content/drive/My Drive/data.csv'
0s data_frame = pd.read_csv(filepath)

stats_summary = data_frame.describe()
print(stats_summary)
```

```
↗
```

	Duration	Pulse	Maxpulse	Calories
count	169.000000	169.000000	169.000000	164.000000
mean	63.846154	107.461538	134.047337	375.790244
std	42.299949	14.510259	16.450434	266.379919
min	15.000000	80.000000	100.000000	50.300000
25%	45.000000	100.000000	124.000000	250.925000
50%	60.000000	105.000000	131.000000	318.600000
75%	60.000000	111.000000	141.000000	387.600000
max	300.000000	159.000000	184.000000	1860.400000

```
✓ 1 filepaths = '/content/drive/My Drive/data.csv'
0s df = pd.read_csv(filepath)
print("Null values before replacement:")
print(df.isnull().sum())
df.fillna(df.mean(), inplace=True)
print("\nNull values after replacement:")
print(df.isnull().sum())
df.to_csv('data_cleaned.csv', index=False)
```

```
↗ Null values before replacement:
Duration    0
Pulse       0
Maxpulse    0
Calories    5
dtype: int64

Null values after replacement:
Duration    0
Pulse       0
Maxpulse    0
Calories    0
```

```
[25] filepath = '/content/drive/My Drive/data.csv'
df = pd.read_csv(filepath)
columns_to_aggregate = ['Duration', 'Pulse']
a = df[columns_to_aggregate].agg(['min', 'max', 'count', 'mean'])
print("Aggregated Data:")
print(a)
```

⇒ Aggregated Data:

	Duration	Pulse
min	15.000000	80.000000
max	300.000000	159.000000
count	169.000000	169.000000
mean	63.846154	107.461538

✓
0s

```
▶ print("\nRows with Calories between 500 and 1000:")
filtered_df_2= df[(df['Calories'] > 500) & (df['Pulse'] < 100)]
print (filtered_df_2)
```

⇒ Rows with Calories between 500 and 1000:

	Duration	Pulse	Maxpulse	Calories
65	180	90	130	800.4
70	150	97	129	1115.0
73	150	97	127	953.2
75	90	98	125	563.2
99	90	93	124	604.1
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

```
✓ [36] print("\nRows with Calories between 500 and 1000:")
0s filtered_df = df[(df['Calories'] >= 500) & (df['Calories'] <= 1000)]
print (filtered_df)
```



```
Rows with Calories between 500 and 1000:
   Duration  Pulse  Maxpulse  Calories
51        80    123      146    643.1
62       160    109      135    853.0
65       180     90      130    800.4
66       150    105      135    873.4
67       150    107      130    816.0
72        90    100      127    700.0
73       150     97      127    953.2
75        90     98      125    563.2
78       120    100      130    500.4
83       120    100      130    500.0
90       180    101      127    600.1
99        90     93      124    604.1
101       90     90      110    500.0
102       90     90      100    500.0
103       90     90      100    500.4
106       180     90      120    800.3
108       90     90      120    500.3
```

```
✓ [37] df_modified = df.drop(columns=['Maxpulse'])
0s print("\nDataframe 'df_modified' without 'Maxpulse' column:")
print(df_modified.head())
```



```
Dataframe 'df_modified' without 'Maxpulse' column:
   Duration  Pulse  Calories
0        60    110    409.1
1        60    117    479.0
2        60    103    340.0
3        45    109    282.4
4        45    117    406.0
```

```
✓ [39] df.drop(columns=['Maxpulse'], inplace=True)
0s print("\nOriginal dataframe after deleting 'Maxpulse':")
print(df.head())
```



```
Original dataframe after deleting 'Maxpulse':
   Duration  Pulse  Calories
0        60    110    409.1
1        60    117    479.0
2        60    103    340.0
3        45    109    282.4
4        45    117    406.0
```



MY GITHUB REPOSITORY:

<https://github.com/nithin1086/BDA>