Assignment

2023004473
G. Keerthi

(3) Fri 14/9/24.

```
pip install pandas
import pandas as pd
df = pd.read_excel('marks_stu.xlsx')
original_df = pd.DataFrame(data)
original_df['Total marks'] = original_df[['Math score', 'English score',
'science score']]
modified_df = original_df.copy()
modified_df.drop(['Name', 'Age'], axis=1, inplace=True)
print("Original Dataset:")
print(original_df)
print("\nModified Dataset:")
print(modified_df)
avg_marks = modified_df[['Math Score', 'English Score', 'Science Score']]
.mean()
median_marks = modified_df[['Math Score', 'English Score', 'Science Score']]
.median()
min_marks = modified_df[['Math Score', 'English Score', 'Science Score']]
.min()
max_marks = modified_df[['Math Score', 'English Score', 'Science Score']]
.max()
mode_age = original_df['Age'].mode()[0]
num_students = len(original_df)
print("\nAverage Marks:")
print(avg_marks)
print("\nMedian Marks:")
print(median_marks)
print("\nMinimum Marks:")
```

```
print(min_marks)
print("\nMaximum Marks:")
print(max_marks)
print("\nMode of Ages:", mode_age)
print("\nCount of students:", num_students)
print("\nTotal Marks:")
print(original_df[['Student ID', 'Name', 'Total Marks']])
```

Original Dataset:

	Student ID	Name	Gender	Age	Math Score	English Score	Science Score
0	1	John Smith	Male	15	85	78	92
1	2	Emily Chen	Female	16	92	89	88
2	3	David Lee	Male	15	78	82	80
3	4	Sarah Wang	Female	16	90	85	94
4	5	Michael Liu	Male	17	88	90	85

Modified Dataset:

	Student ID	Gender	Math Score	English Score	Science Score	Total Marks
0	1	Male	85	78	92	255
1	2	Female	92	89	88	269
2	3	Male	78	82	80	240
3	4	Female	90	85	94	269
4	5	Male	88	90	85	263

Average Marks:

Math score 86.6
English score 86.8
Science score 87.8
dtype: float64

Median Marks:

Math score 88.0
English score 89.0
Science score 85.0
dtype: float64

Minimum Marks:

Math score 78
English score 78
Science score 80
dtype: int64

Maximum Marks

Math score 92
English score 90
Science score 94
dtype: int 64

Mode of Ages: 15

Count of students: 5

Total Marks :

	Student ID	Name	Total Marks
0	1	John Smith	255
1	2	Emily Chen	269
2	3	David Lee	240
3	4	Sarah Wang	269
4	5	Michael Liu	263

1. Add/Drop columns and rows excel files
2. Using Aggregations and applying functions on above example.

Lets assume we have an employee salary dataset in excel file.

Task:1: Make sure to install the pandas library.

Task:2: Read the excel file into a data frame and print the initial data frame.

Task:3: Adding columns / rows

Add a column 'Bonus' by assuming a bonus of twice of the salary.

→ dropping column -

drop the age column by using drop() function

→ dropping rows -

based on the conditions like removing employes with salary less than 50,000.

Now display the modified data frame.

Task: 4: Using Aggregations and applying functions

(i) Average salary: The "mean()" method will calculate the average salary of the employees in the dataframe.

(ii) Median Salary: The "median()" method will calculate the median of the salary of the employees in the Dataframes

(iii) Mode Experience: The "mode()" method will calculate the mode of the experienced years of the employees.

(iv) Minimum salary: The "min()" method finds the minimum salary of an employees in the dataframe.

(v) Maximum salary: The "max()" method finds the maximum salary of an employees in the data frame.

(vi) Sume of the bonus salary: The "sum()" method calculates the sum of bonus salaries of the employees in the Dataframe.

(vii) Count of the ID's: The "count()" method counts the no. of employees in the data frame.