

Assignment - Statistics

[Major]

Name – Amit Kumar

E-mail- amitkumar209727@gmail.com

M0- 7881178181

Batch No- 7

1. According to a study, the daily average time spent by a user on a social media website is 50 minutes. To test the claim of this study, Ramesh, a researcher, takes a sample of 25 website users and finds out that the mean time spent by the sample users is 60 minutes and the sample standard deviation is 30 minutes. Based on this information, the null and the alternative hypotheses will be: H_0 = The average time spent by the users is 50 minutes H_1 = The average time spent by the users is not 50 minutes Use a 5% significance level to test this hypothesis.

```
import scipy.stats as stats
import numpy as np
sample_mean = 60 # Sample mean
sample_std = 30 # Sample standard deviation
n = 25 # Sample size
population_mean = 50 # Population mean

# Calculate the t-statistic and p-value
```

```

t_statistic = (sample_mean - population_mean) / (sample_std / np.sqrt(n))
p_value = 2 * (1 - stats.t.cdf(abs(t_statistic), df=n-1)) # Two-tailed test

alpha = 0.05 # Significance level

if p_value < alpha:
    print("Reject the null hypothesis.")
else:
    print("Fail to reject the null hypothesis.")

```

The screenshot shows the Visual Studio Code interface with a file named `a.py` open. The code in the editor is as follows:

```

1 import scipy.stats as stats
2 import numpy as np
3 sample_mean = 60 # Sample mean
4 sample_std = 30 # Sample standard deviation
5 n = 25 # Sample size
6 population_mean = 50 # Population mean
7
8 # Calculate the t-statistic and p-value
9 t_statistic = (sample_mean - population_mean) / (sample_std / np.sqrt(n))
10 p_value = 2 * (1 - stats.t.cdf(abs(t_statistic), df=n-1)) # Two-tailed test
11
12 alpha = 0.05 # Significance level
13
14 if p_value < alpha:
15     print("Reject the null hypothesis.")
16 else:
17     print("Fail to reject the null hypothesis.")
18

```

The terminal window at the bottom shows the command prompt output:

```

PS C:\Users\Amit Kumar\Desktop> c:: cd 'c:\Users\Amit Kumar\Desktop'; & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit K
umar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '65278' '---' 'c:\Users\Amit Kumar\Desktop\a.py'
Traceback (most recent call last):
  File "c:\Users\Amit Kumar\Desktop\a.py", line 9, in <module>
    t_statistic = (sample_mean - population_mean) / (sample_std / np.sqrt(n))
NameError: name 'np' is not defined. Did you mean: 'n'?
PS C:\Users\Amit Kumar\Desktop> c:: cd 'c:\Users\Amit Kumar\Desktop'; & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit K
umar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '65290' '---' 'c:\Users\Amit Kumar\Desktop\a.py'
Fail to reject the null hypothesis.
PS C:\Users\Amit Kumar\Desktop>

```

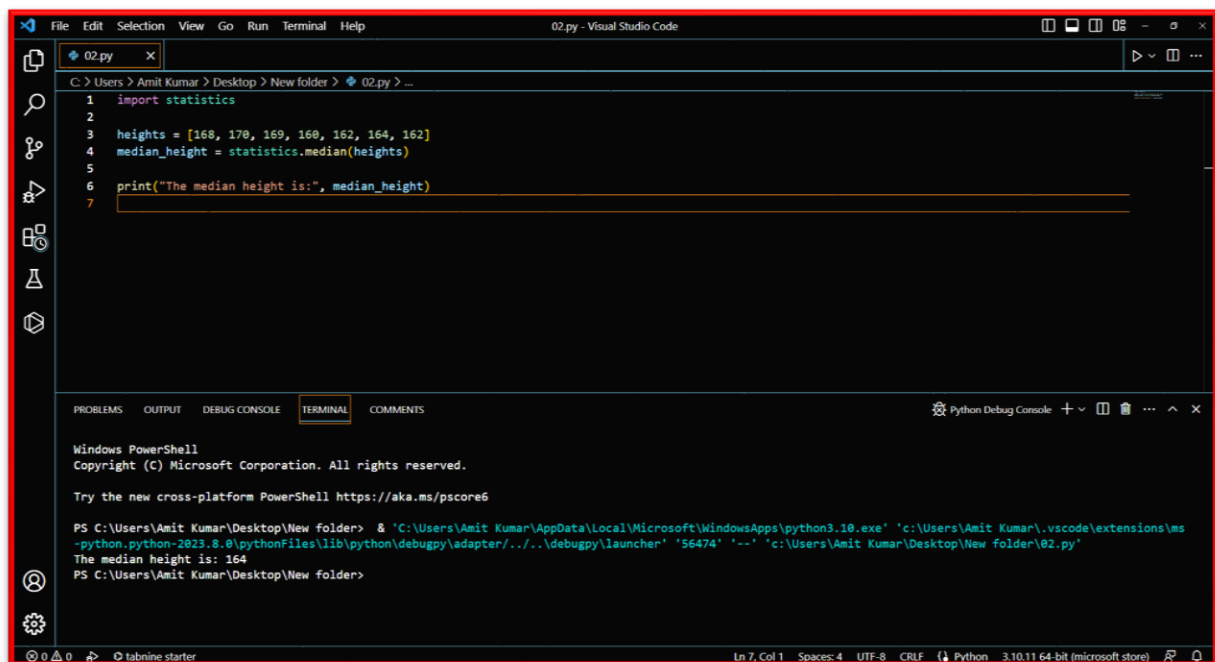
The error message indicates that the variable `np` is not defined, which is a common mistake when importing `numpy` as `np` but not using it correctly in the code.

2. Height of 7 students (in cm) is given below. What is the median? 168 170 169 160 162 164 162.

```
import statistics

heights = [168, 170, 169, 160, 162, 164, 162]
median_height = statistics.median(heights)

print("The median height is:", median_height)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python file named '02.py' with the following code:

```
1 import statistics
2
3 heights = [168, 170, 169, 160, 162, 164, 162]
4 median_height = statistics.median(heights)
5
6 print("The median height is:", median_height)
7
```

The bottom panel shows the 'TERMINAL' tab with the following output:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

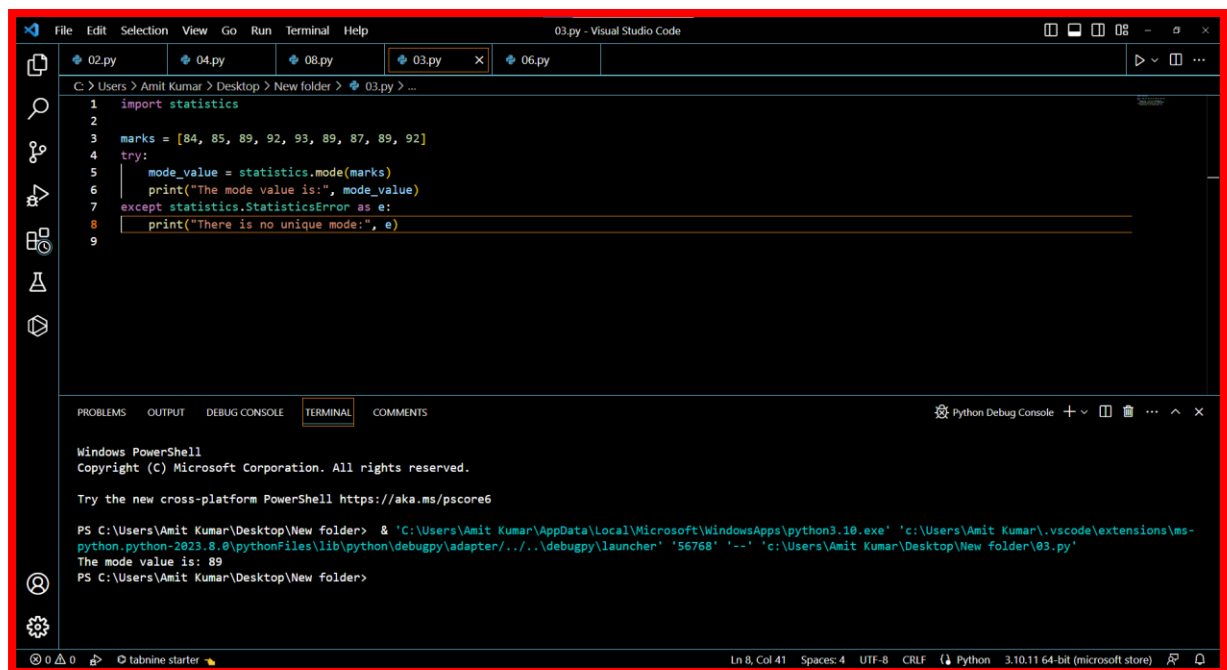
PS C:\Users\Amit Kumar\Desktop\New folder> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '56474' '--' 'c:\Users\Amit Kumar\Desktop\New folder\02.py'
The median height is: 164
PS C:\Users\Amit Kumar\Desktop\New folder>
```

The status bar at the bottom indicates the file is at line 7, column 1, using UTF-8 encoding, and the Python interpreter is 3.10.11 64-bit (microsoft store).

3. Below are the observations of the marks of a student. Find the value of mode. 84 85 89 92 93 89 87 89 92.

```
import statistics

marks = [84, 85, 89, 92, 93, 89, 87, 89, 92]
try:
    mode_value = statistics.mode(marks)
    print("The mode value is:", mode_value)
except statistics.StatisticsError as e:
    print("There is no unique mode:", e)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named 03.py with the following code:

```
1 import statistics
2
3 marks = [84, 85, 89, 92, 93, 89, 87, 89, 92]
4 try:
5     mode_value = statistics.mode(marks)
6     print("The mode value is:", mode_value)
7 except statistics.StatisticsError as e:
8     print("There is no unique mode:", e)
9
```

The terminal window at the bottom shows the output of the script:

```
Windows PowerShell
Copyright (c) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\Amit Kumar\Desktop\New folder> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '56768' '--' 'c:\Users\Amit Kumar\Desktop\New folder\03.py'
The mode value is: 89
PS C:\Users\Amit Kumar\Desktop\New folder>
```

The status bar at the bottom indicates the file is at line 8, column 41, using UTF-8 encoding with CRLF line endings. The Python version is 3.10.11 64-bit (microsoft store).

4. From the table given below, what is the mean of marks obtained by 20 students?

Marks X_i	No. of students f_i
3	1
4	2
5	2
6	4
7	5
8	3
9	2
10	1
Total	20

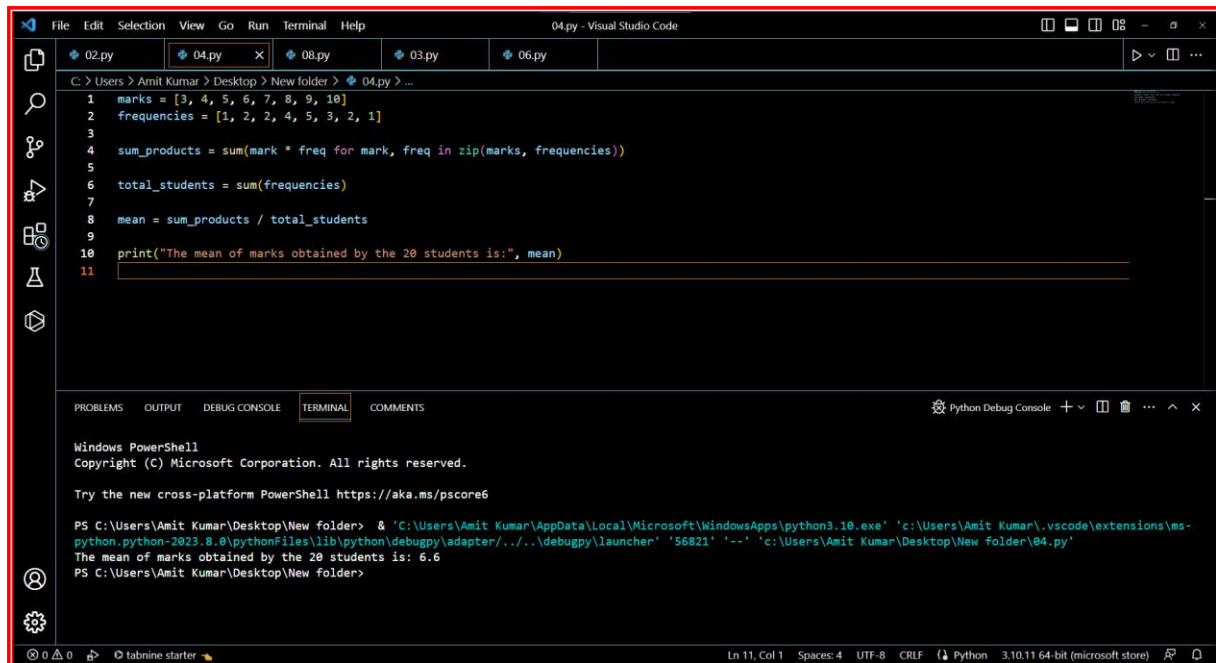
```
marks = [3, 4, 5, 6, 7, 8, 9, 10]
frequencies = [1, 2, 2, 4, 5, 3, 2, 1]

sum_products = sum(mark * freq for mark, freq in zip(marks, frequencies))

total_students = sum(frequencies)

mean = sum_products / total_students

print("The mean of marks obtained by the 20 students is:", mean)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python file named 04.py with the following code:

```
1 marks = [3, 4, 5, 6, 7, 8, 9, 10]
2 frequencies = [1, 2, 2, 4, 5, 3, 2, 1]
3
4 sum_products = sum(mark * freq for mark, freq in zip(marks, frequencies))
5
6 total_students = sum(frequencies)
7
8 mean = sum_products / total_students
9
10 print("The mean of marks obtained by the 20 students is:", mean)
11
```

The terminal window at the bottom shows the execution of the script using PowerShell:

```
PS C:\Users\Amit Kumar\Desktop\New folder> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\launcher' '56821' '--' 'c:\Users\Amit Kumar\Desktop\New folder\04.py'
The mean of marks obtained by the 20 students is: 6.6
PS C:\Users\Amit Kumar\Desktop\New folder>
```

5. For a certain type of computer, the length of time between charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. John owns one of these computers and wants to know the probability that the length of time will be between 50 and 70 hours.

```
import scipy.stats as stats

mean = 50 # Mean of the normal distribution
std_dev = 15 # Standard deviation of the normal distribution

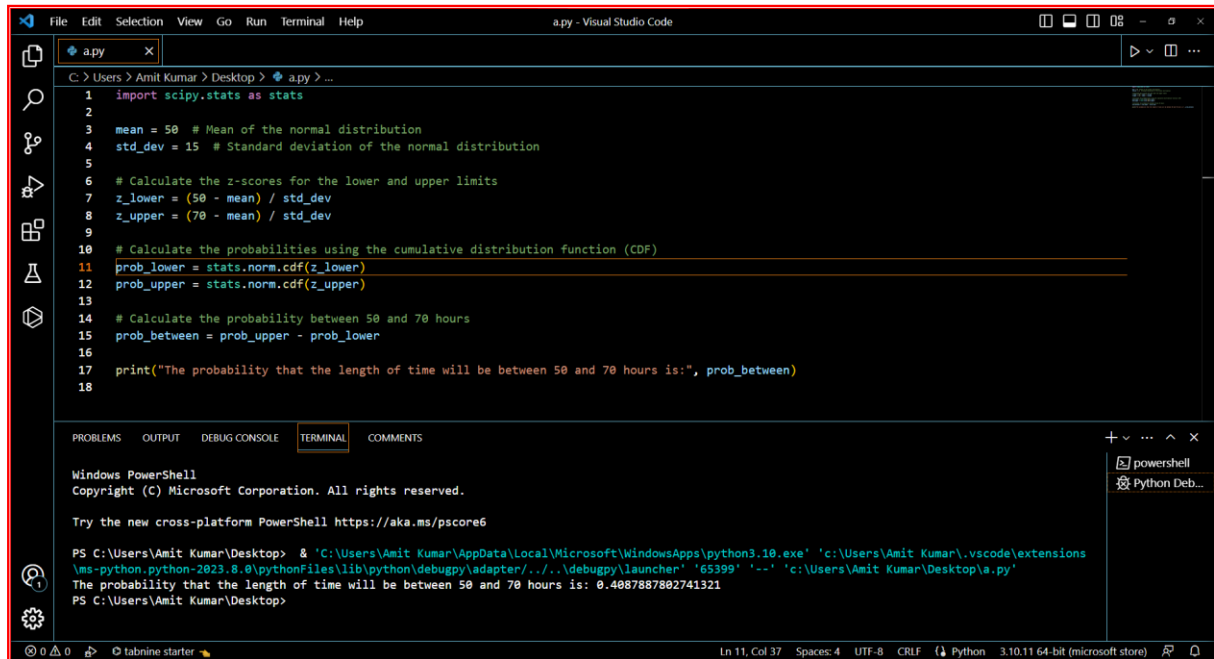
# Calculate the z-scores for the lower and upper limits
z_lower = (50 - mean) / std_dev
z_upper = (70 - mean) / std_dev

# Calculate the probabilities using the cumulative distribution function (CDF)
prob_lower = stats.norm.cdf(z_lower)
prob_upper = stats.norm.cdf(z_upper)

# Calculate the probability between 50 and 70 hours
```

```
prob_between = prob_upper - prob_lower
```

```
print("The probability that the length of time will be between 50 and 70 hours  
is:", prob_between)
```



The screenshot shows a Visual Studio Code window with a file named 'a.py' open. The editor displays a Python script that calculates the probability of a normal distribution between two values. The script uses the 'scipy.stats' module to calculate z-scores and cumulative distribution functions. Below the editor, the 'TERMINAL' tab is active, showing the command prompt output. The output confirms the calculation of the probability between 50 and 70 hours, resulting in 0.4087887802741321.

```
File Edit Selection View Go Run Terminal Help
a.py - Visual Studio Code

C > Users > Amit Kumar > Desktop > a.py > ...
1 import scipy.stats as stats
2
3 mean = 50 # Mean of the normal distribution
4 std_dev = 15 # Standard deviation of the normal distribution
5
6 # Calculate the z-scores for the lower and upper limits
7 z_lower = (50 - mean) / std_dev
8 z_upper = (70 - mean) / std_dev
9
10 # Calculate the probabilities using the cumulative distribution function (CDF)
11 prob_lower = stats.norm.cdf(z_lower)
12 prob_upper = stats.norm.cdf(z_upper)
13
14 # Calculate the probability between 50 and 70 hours
15 prob_between = prob_upper - prob_lower
16
17 print("The probability that the length of time will be between 50 and 70 hours is:", prob_between)
18

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL COMMENTS

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

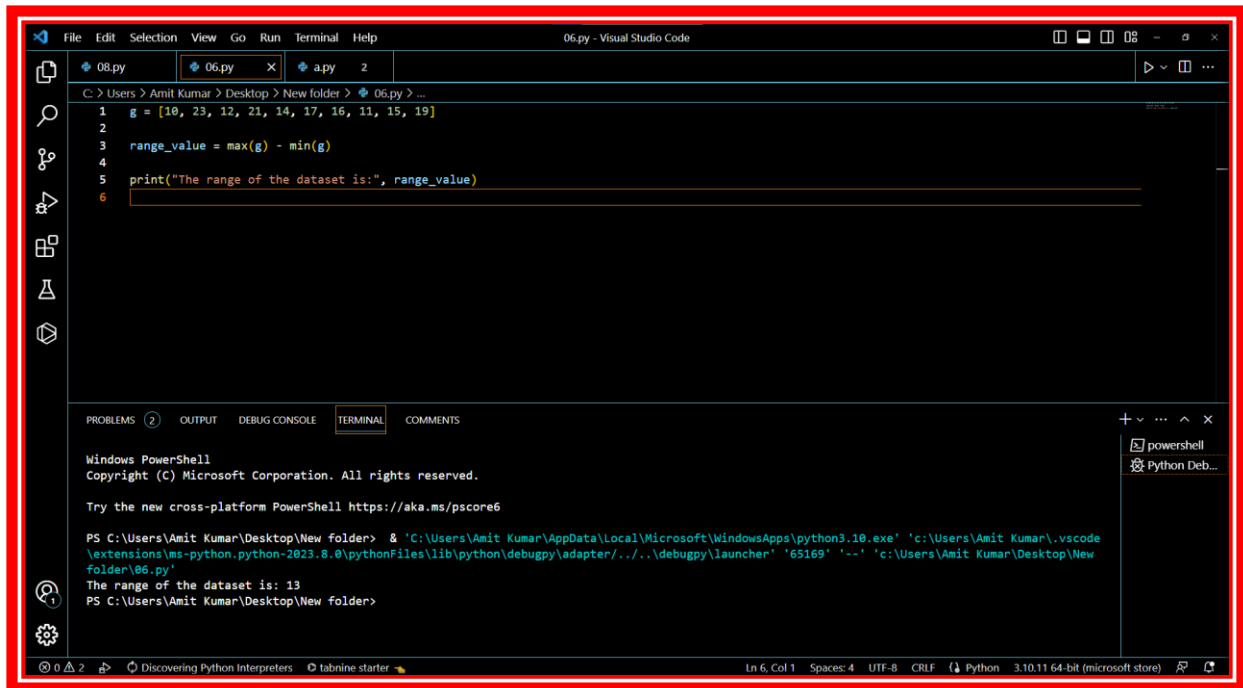
PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '65399' '-.' 'c:\Users\Amit Kumar\Desktop\a.py'
The probability that the length of time will be between 50 and 70 hours is: 0.4087887802741321
PS C:\Users\Amit Kumar\Desktop>
```

6. Find the range of the following. $g = [10, 23, 12, 21, 14, 17, 16, 11, 15, 19]$

```
g = [10, 23, 12, 21, 14, 17, 16, 11, 15, 19]
```

```
range_value = max(g) - min(g)
```

```
print("The range of the dataset is:", range_value)
```

A screenshot of the Visual Studio Code interface. The editor window shows a Python file named '06.py' with the following code:

```
1 g = [10, 23, 12, 21, 14, 17, 16, 11, 15, 19]
2
3 range_value = max(g) - min(g)
4
5 print("The range of the dataset is:", range_value)
6
```

The terminal at the bottom shows the command prompt output:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\Amit Kumar\Desktop\New folder> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '65169' '--' 'c:\Users\Amit Kumar\Desktop\New folder\06.py'
The range of the dataset is: 13
PS C:\Users\Amit Kumar\Desktop\New folder>
```

7. It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam 2 email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

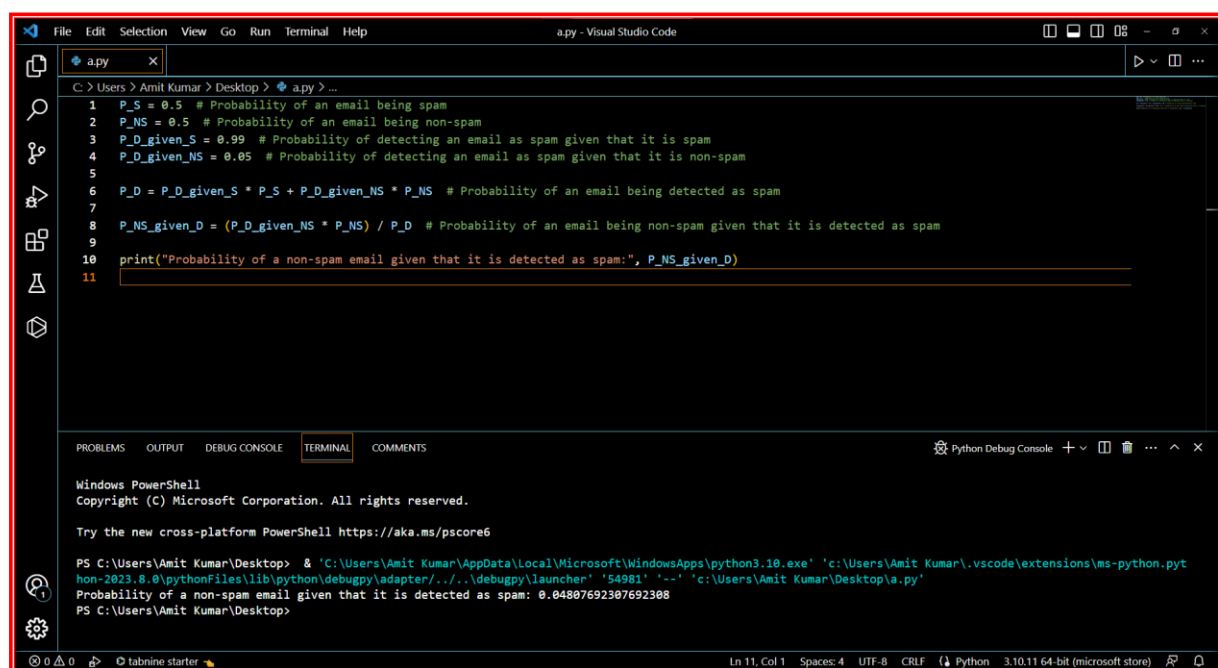
```
P_S = 0.5 # Probability of an email being spam
P_NS = 0.5 # Probability of an email being non-spam
P_D_given_S = 0.99 # Probability of detecting an email as spam given that it is spam
P_D_given_NS = 0.05 # Probability of detecting an email as spam given that it is non-spam
```



```
P_D = P_D_given_S * P_S + P_D_given_NS * P_NS # Probability of an email being
detected as spam
```

```
P_NS_given_D = (P_D_given_NS * P_NS) / P_D # Probability of an email being
non-spam given that it is detected as spam
```

```
print("Probability of a non-spam email given that it is detected as spam:",
P_NS_given_D)
```



The screenshot shows a Visual Studio Code window with a file named 'apy'. The code in the editor is a Python script for calculating probabilities. Below the editor, the 'TERMINAL' tab is active, showing the execution of the script. The output of the script is 'Probability of a non-spam email given that it is detected as spam: 0.04807692307692308'.

```
File Edit Selection View Go Run Terminal Help
apy - Visual Studio Code

C:\Users\Amit Kumar\Desktop> apy > ...
1 P_S = 0.5 # Probability of an email being spam
2 P_NS = 0.5 # Probability of an email being non-spam
3 P_D_given_S = 0.99 # Probability of detecting an email as spam given that it is spam
4 P_D_given_NS = 0.05 # Probability of detecting an email as spam given that it is non-spam
5
6 P_D = P_D_given_S * P_S + P_D_given_NS * P_NS # Probability of an email being detected as spam
7
8 P_NS_given_D = (P_D_given_NS * P_NS) / P_D # Probability of an email being non-spam given that it is detected as spam
9
10 print("Probability of a non-spam email given that it is detected as spam:", P_NS_given_D)
11

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL COMMENTS
Python Debug Console + - [ ] ... ^ X

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.pyt
hon-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '54981' '-.' 'c:\Users\Amit Kumar\Desktop\apy'
Probability of a non-spam email given that it is detected as spam: 0.04807692307692308
PS C:\Users\Amit Kumar\Desktop>
```

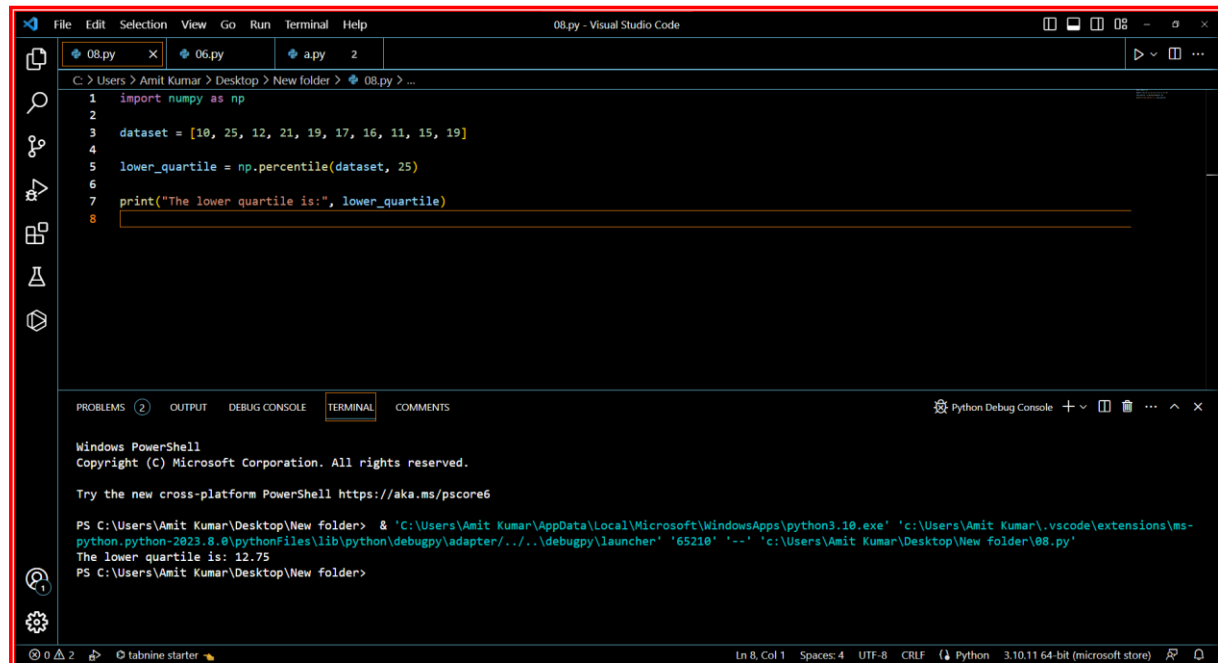
8. Given the following distribution of returns,
determine the lower quartile: {10 25 12 21 19 17 16
11 15 19}

```
import numpy as np

dataset = [10, 25, 12, 21, 19, 17, 16, 11, 15, 19]

lower_quartile = np.percentile(dataset, 25)

print("The lower quartile is:", lower_quartile)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python file named 08.py with the following code:

```
1 import numpy as np
2
3 dataset = [10, 25, 12, 21, 19, 17, 16, 11, 15, 19]
4
5 lower_quartile = np.percentile(dataset, 25)
6
7 print("The lower quartile is:", lower_quartile)
8
```

The terminal window at the bottom shows the execution of the script using PowerShell:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

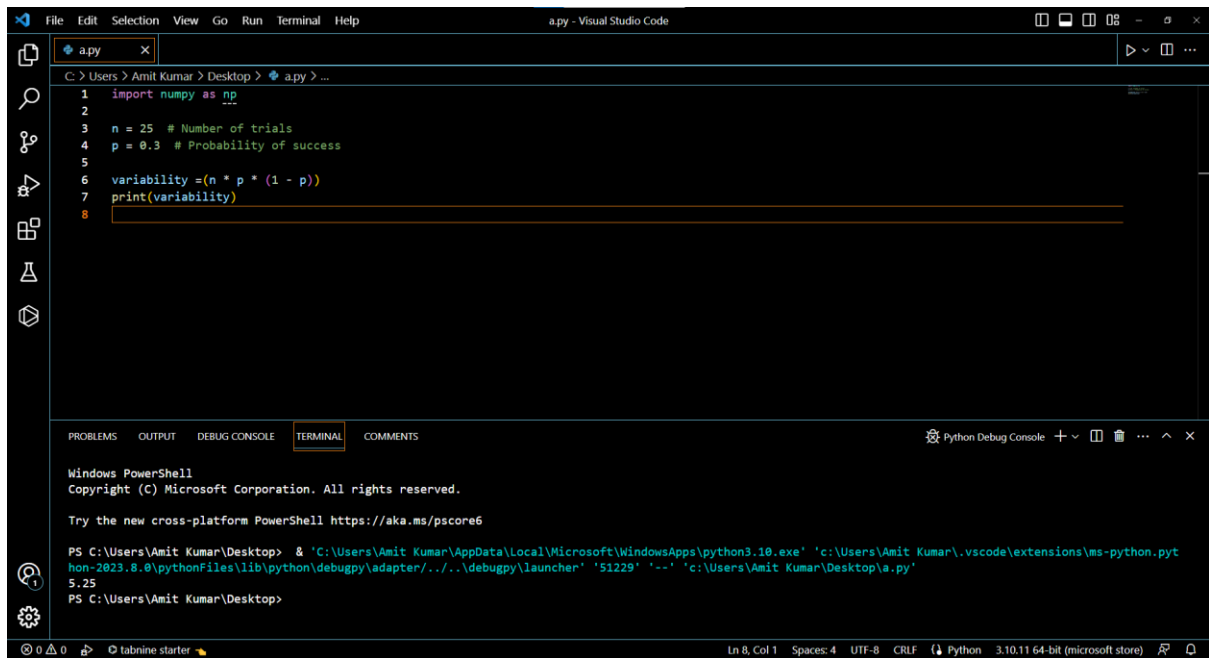
PS C:\Users\Amit Kumar\Desktop\New folder> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '65210' '--' 'c:\Users\Amit Kumar\Desktop\New folder\08.py'
The lower quartile is: 12.75
PS C:\Users\Amit Kumar\Desktop\New folder>
```

9. For a Binomial distribution, the number of trials(n) is 25, and the probability of success is 0.3. What's the variability of the distribution?

```
import numpy as np

n = 25 # Number of trials
p = 0.3 # Probability of success

variability=(n * p * (1 - p))
print(variability)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named `a.py` with the following code:

```
1 import numpy as np
2
3 n = 25 # Number of trials
4 p = 0.3 # Probability of success
5
6 variability = (n * p * (1 - p))
7 print(variability)
8
```

The terminal window at the bottom shows the command prompt output:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.pyth
hon-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '51229' '---' 'c:\Users\Amit Kumar\Desktop\la.py'
5.25
PS C:\Users\Amit Kumar\Desktop>
```

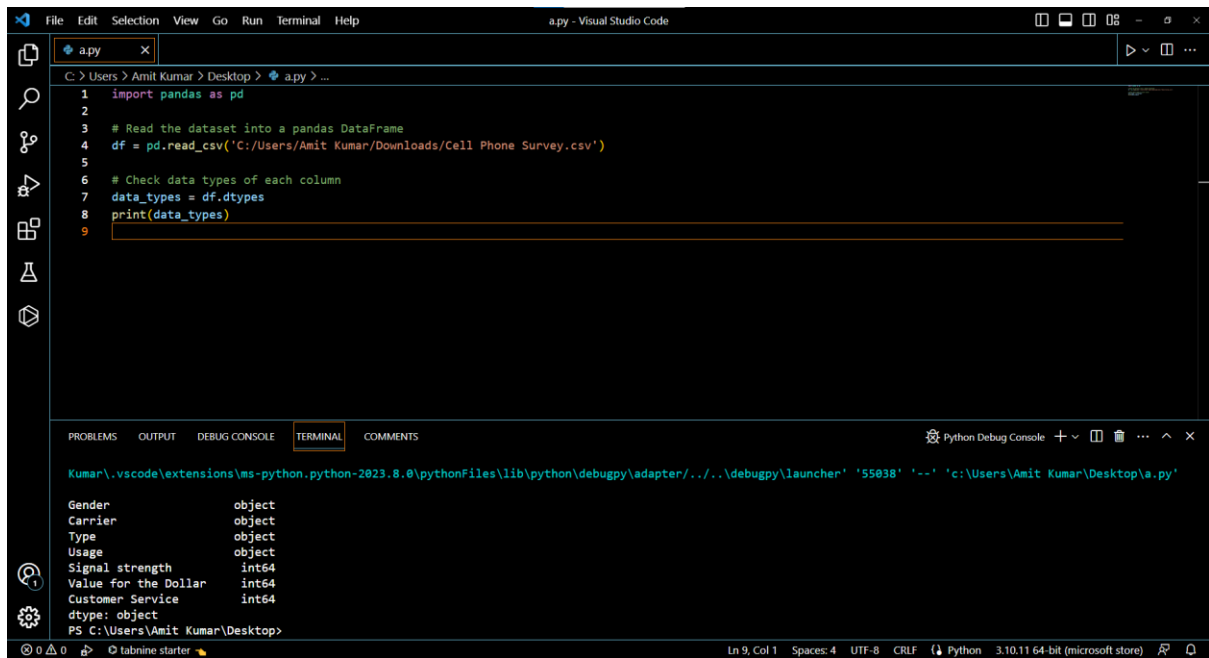
10. Download the Cell Phone Survey Dataset and perform the below mentioned operations on the dataset:-

- **Checking datatypes of each column in the dataset.**

```
import pandas as pd

# Read the dataset into a pandas DataFrame
df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

# Check data types of each column
data_types = df.dtypes
print(data_types)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named 'a.py' with the following code:

```
1 import pandas as pd
2
3 # Read the dataset into a pandas DataFrame
4 df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
5
6 # Check data types of each column
7 data_types = df.dtypes
8 print(data_types)
9
```

The bottom panel shows the 'TERMINAL' tab with the output of the script:

```
Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher '55038' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
Gender                object
Carrier               object
Type                 object
Usage                object
Signal strength       int64
Value for the Dollar  int64
Customer Service      int64
dtype: object
PS C:\Users\Amit Kumar\Desktop>
```

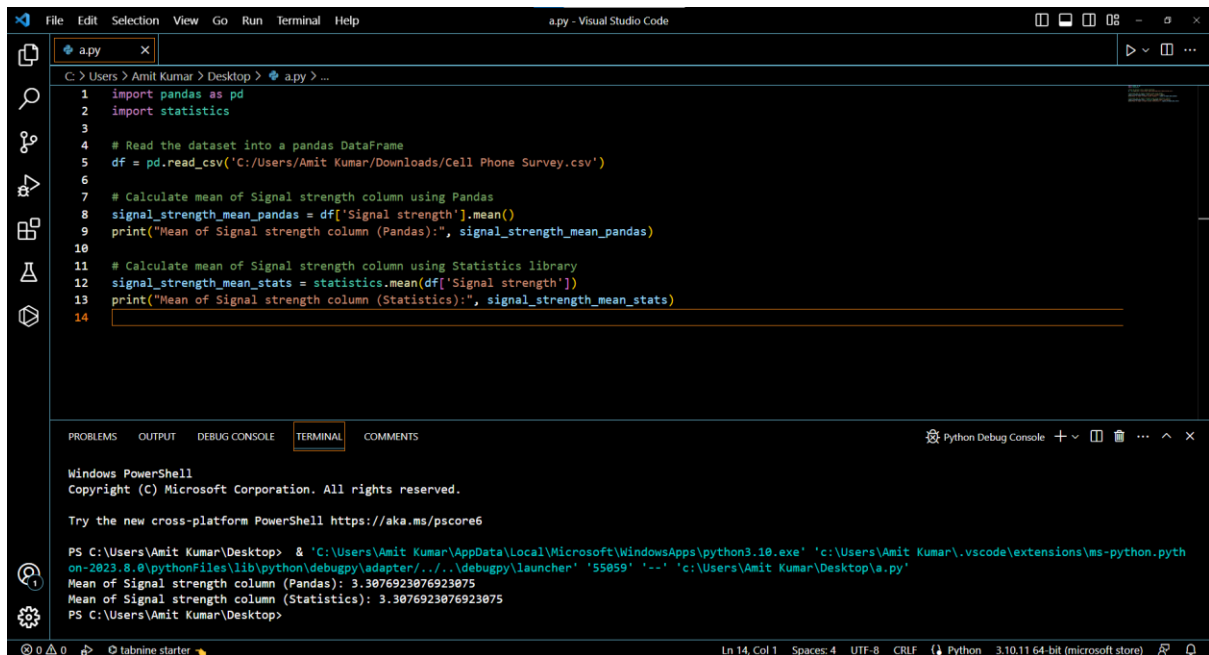
● Find Mean of Signal strength column using Pandas and Statistics library.

```
import pandas as pd
import statistics

# Read the dataset into a pandas DataFrame
df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

# Calculate mean of Signal strength column using Pandas
signal_strength_mean_pandas = df['Signal strength'].mean()
print("Mean of Signal strength column (Pandas):", signal_strength_mean_pandas)

# Calculate mean of Signal strength column using Statistics library
signal_strength_mean_stats = statistics.mean(df['Signal strength'])
print("Mean of Signal strength column (Statistics):",
signal_strength_mean_stats)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named 'a.py' with the following code:

```
1 import pandas as pd
2 import statistics
3
4 # Read the dataset into a pandas DataFrame
5 df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
6
7 # Calculate mean of Signal strength column using Pandas
8 signal_strength_mean_pandas = df['Signal strength'].mean()
9 print("Mean of Signal strength column (Pandas):", signal_strength_mean_pandas)
10
11 # Calculate mean of Signal strength column using Statistics library
12 signal_strength_mean_stats = statistics.mean(df['Signal strength'])
13 print("Mean of Signal strength column (Statistics):", signal_strength_mean_stats)
14
```

The bottom panel shows the 'TERMINAL' tab with the following output:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/powershell

PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '55059' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
Mean of Signal strength column (Pandas): 3.3076923076923075
Mean of Signal strength column (Statistics): 3.3076923076923075
PS C:\Users\Amit Kumar\Desktop>
```

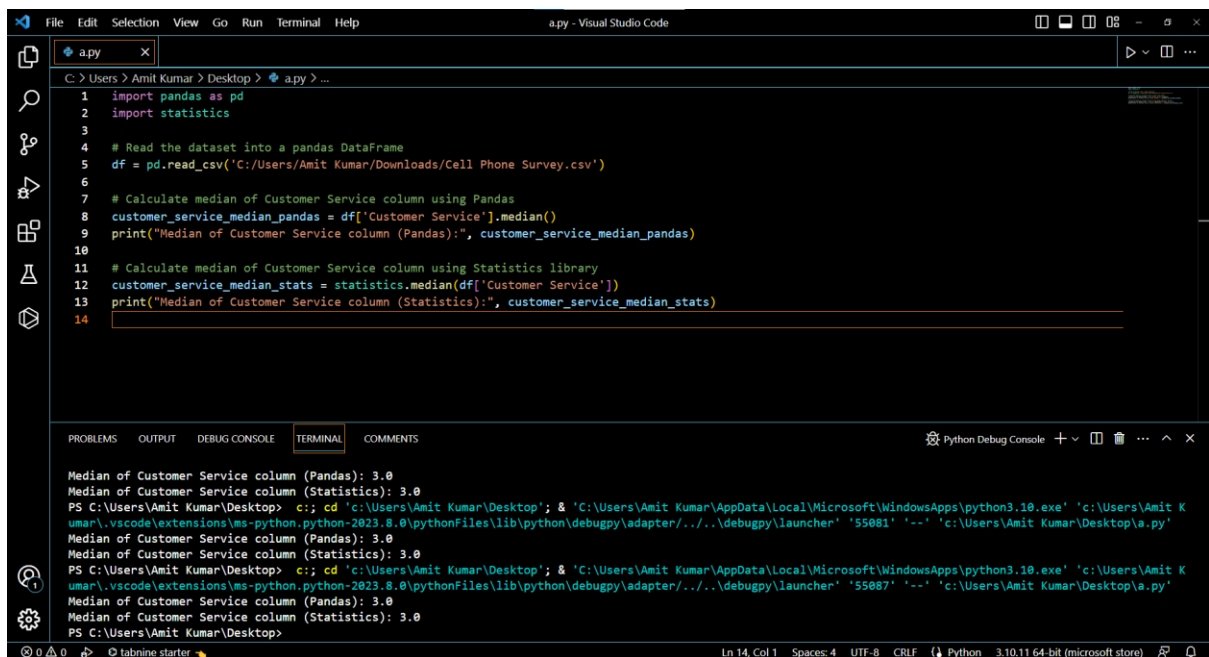
● Find the Median of Customer Service column using Pandas and Statistics library.

```
import pandas as pd
import statistics

# Read the dataset into a pandas DataFrame
df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

# Calculate median of Customer Service column using Pandas
customer_service_median_pandas = df['Customer Service'].median()
print("Median of Customer Service column (Pandas):",
customer_service_median_pandas)

# Calculate median of Customer Service column using Statistics library
customer_service_median_stats = statistics.median(df['Customer Service'])
print("Median of Customer Service column (Statistics):",
customer_service_median_stats)
```



The screenshot shows the Visual Studio Code interface with a Python file named 'a.py' open. The script calculates the median of the 'Customer Service' column using both Pandas and the Statistics library. The terminal output shows the execution results, including the median value of 3.0 for both methods. The interface includes a sidebar with icons for Explorer, Search, Source Control, Run and Debug, and Extensions. The bottom status bar indicates the file is at line 14, column 1, with 4 spaces, UTF-8 encoding, CRLF line endings, and is a Python file.

```
File Edit Selection View Go Run Terminal Help
a.py - Visual Studio Code

C:\Users\Amit Kumar\Desktop> a.py ...
1 import pandas as pd
2 import statistics
3
4 # Read the dataset into a pandas DataFrame
5 df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
6
7 # Calculate median of Customer Service column using Pandas
8 customer_service_median_pandas = df['Customer Service'].median()
9 print("Median of Customer Service column (Pandas):", customer_service_median_pandas)
10
11 # Calculate median of Customer Service column using Statistics library
12 customer_service_median_stats = statistics.median(df['Customer Service'])
13 print("Median of Customer Service column (Statistics):", customer_service_median_stats)
14

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL COMMENTS
Python Debug Console + - [ ] ... ^ x

Median of Customer Service column (Pandas): 3.0
Median of Customer Service column (Statistics): 3.0
PS C:\Users\Amit Kumar\Desktop> cd 'c:\Users\Amit Kumar\Desktop'; & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '55081' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
Median of Customer Service column (Pandas): 3.0
Median of Customer Service column (Statistics): 3.0
PS C:\Users\Amit Kumar\Desktop> cd 'c:\Users\Amit Kumar\Desktop'; & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '55087' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
Median of Customer Service column (Pandas): 3.0
Median of Customer Service column (Statistics): 3.0
PS C:\Users\Amit Kumar\Desktop>

Ln 14, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.11 64-bit (microsoft store)
```

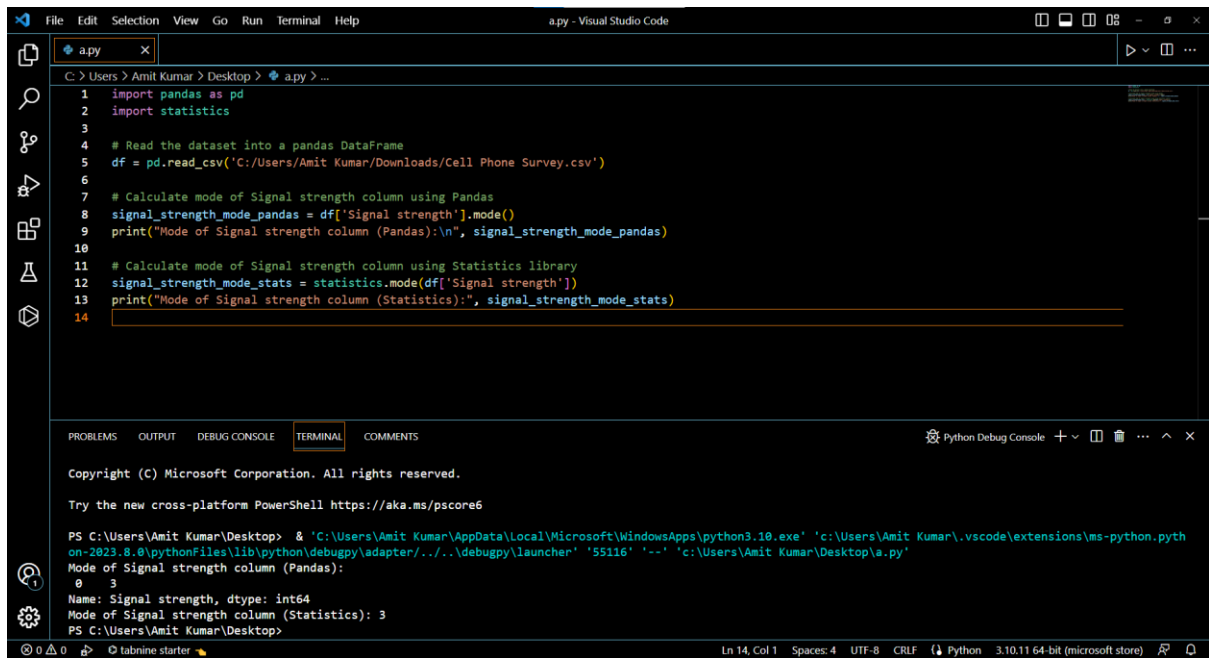
● Find Mode of Signal strength column using Pandas and Statistics library.

```
import pandas as pd
import statistics

# Read the dataset into a pandas DataFrame
df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

# Calculate mode of Signal strength column using Pandas
signal_strength_mode_pandas = df['Signal strength'].mode()
print("Mode of Signal strength column (Pandas):\n",
signal_strength_mode_pandas)

# Calculate mode of Signal strength column using Statistics library
signal_strength_mode_stats = statistics.mode(df['Signal strength'])
print("Mode of Signal strength column (Statistics):",
signal_strength_mode_stats)
```



The screenshot shows the Visual Studio Code interface with a Python file named 'a.py' open. The code in the editor calculates the mode of the 'Signal strength' column using both Pandas and the Statistics library. The terminal at the bottom shows the execution output, including the mode value of 3 for the Signal strength column.

```
File Edit Selection View Go Run Terminal Help
a.py - Visual Studio Code

C:\Users\Amit Kumar\Desktop> a.py ...
1 import pandas as pd
2 import statistics
3
4 # Read the dataset into a pandas DataFrame
5 df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
6
7 # Calculate mode of Signal strength column using Pandas
8 signal_strength_mode_pandas = df['Signal strength'].mode()
9 print("Mode of Signal strength column (Pandas):\n", signal_strength_mode_pandas)
10
11 # Calculate mode of Signal strength column using Statistics library
12 signal_strength_mode_stats = statistics.mode(df['Signal strength'])
13 print("Mode of Signal strength column (Statistics):", signal_strength_mode_stats)
14
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL COMMENTS

Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell <https://aka.ms/powershell>

PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '55116' '--' 'c:\Users\Amit Kumar\Desktop\a.py'

Mode of Signal strength column (Pandas):
0 3
Name: Signal strength, dtype: int64
Mode of Signal strength column (Statistics): 3
PS C:\Users\Amit Kumar\Desktop>

Ln 14, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.11 64-bit (microsoft store)

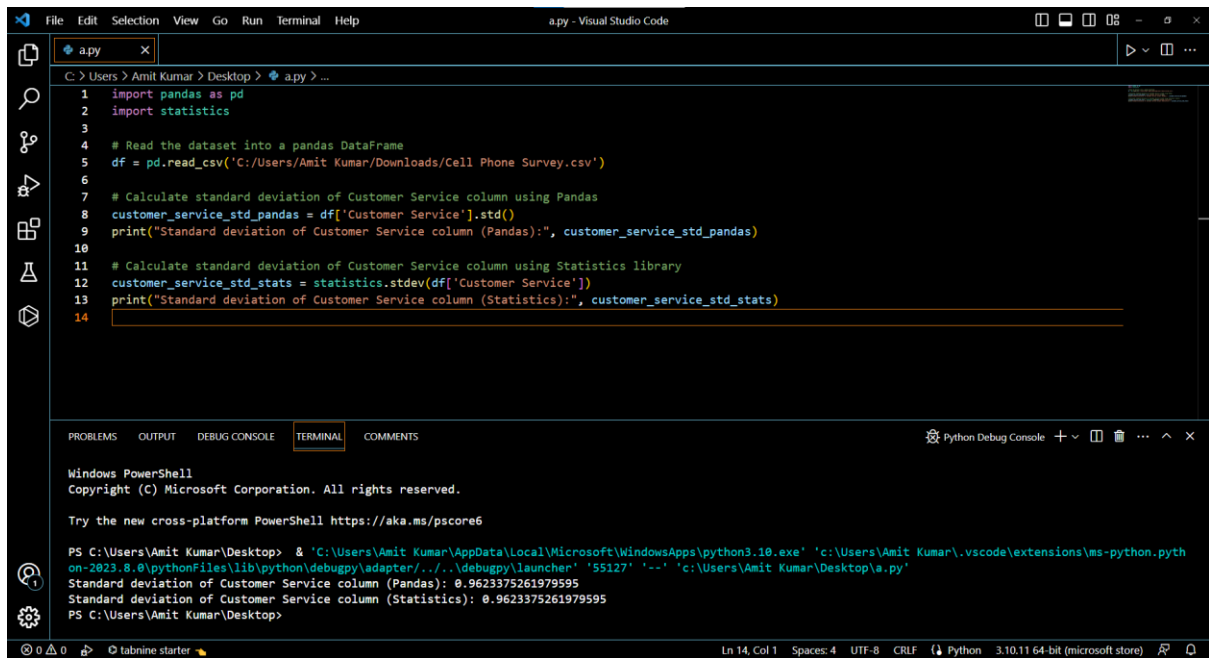
● Find Standard deviation of Customer Service column using Pandas and Statistics library.

```
import pandas as pd
import statistics

# Read the dataset into a pandas DataFrame
df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

# Calculate standard deviation of Customer Service column using Pandas
customer_service_std_pandas = df['Customer Service'].std()
print("Standard deviation of Customer Service column (Pandas):",
customer_service_std_pandas)

# Calculate standard deviation of Customer Service column using Statistics
library
customer_service_std_stats = statistics.stdev(df['Customer Service'])
print("Standard deviation of Customer Service column (Statistics):",
customer_service_std_stats)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named 'a.py' with the following code:

```
1 import pandas as pd
2 import statistics
3
4 # Read the dataset into a pandas DataFrame
5 df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
6
7 # Calculate standard deviation of Customer Service column using Pandas
8 customer_service_std_pandas = df['Customer Service'].std()
9 print("Standard deviation of Customer Service column (Pandas):", customer_service_std_pandas)
10
11 # Calculate standard deviation of Customer Service column using Statistics library
12 customer_service_std_stats = statistics.stdev(df['Customer Service'])
13 print("Standard deviation of Customer Service column (Statistics):", customer_service_std_stats)
14
```

The bottom panel shows the 'TERMINAL' tab with the following output:

```
Windows PowerShell
Copyright (c) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/powershell

PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '55127' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
Standard deviation of Customer Service column (Pandas): 0.9623375261979595
Standard deviation of Customer Service column (Statistics): 0.9623375261979595
PS C:\Users\Amit Kumar\Desktop>
```

● Find Variance of Customer Service column using Pandas and Statistics library.

```
import pandas as pd
import statistics

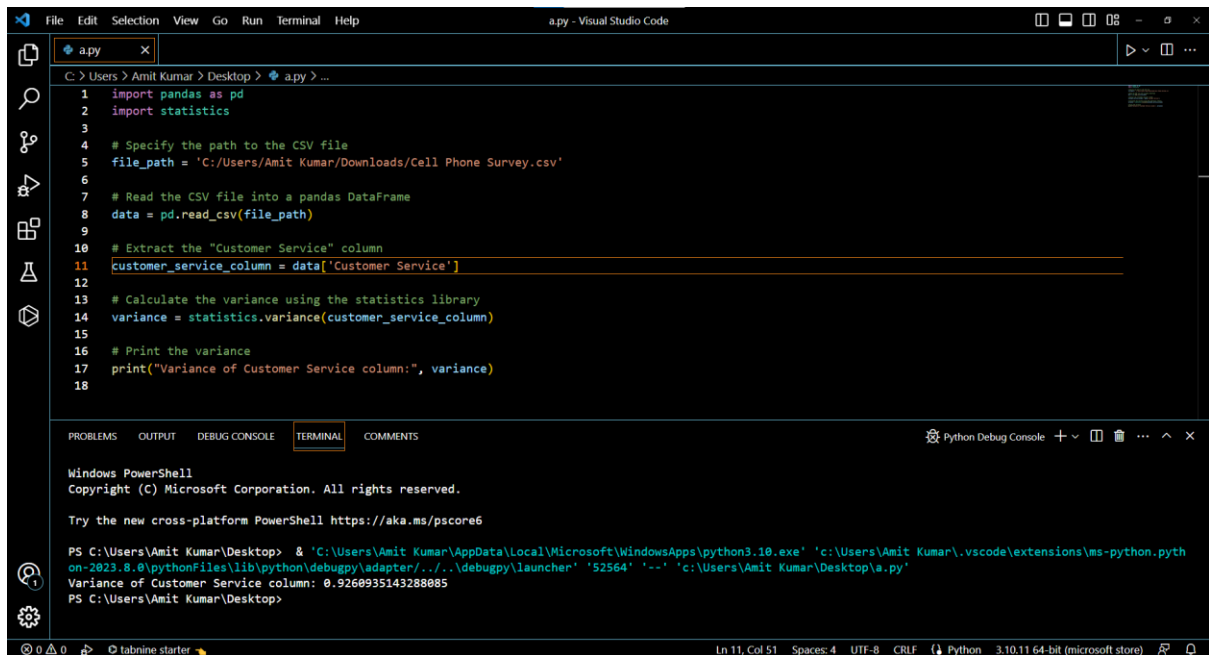
# Specify the path to the CSV file
file_path = 'C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv'

# Read the CSV file into a pandas DataFrame
data = pd.read_csv(file_path)

# Extract the "Customer Service" column
customer_service_column = data['Customer Service']

# Calculate the variance using the statistics library
variance = statistics.variance(customer_service_column)

# Print the variance
print("Variance of Customer Service column:", variance)
```

The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named `a.py` with the following code:

```
1 import pandas as pd
2 import statistics
3
4 # Specify the path to the CSV file
5 file_path = 'C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv'
6
7 # Read the CSV file into a pandas DataFrame
8 data = pd.read_csv(file_path)
9
10 # Extract the "Customer Service" column
11 customer_service_column = data['Customer Service']
12
13 # Calculate the variance using the statistics library
14 variance = statistics.variance(customer_service_column)
15
16 # Print the variance
17 print("Variance of Customer Service column:", variance)
18
```

The terminal window at the bottom shows the execution of the script using the Python Debug Console:

```
PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '52564' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
Variance of Customer Service column: 0.9260935143288885
PS C:\Users\Amit Kumar\Desktop>
```

- Calculate Percentiles of Value for the Dollar column using Numpy.

```
import pandas as pd
import numpy as np

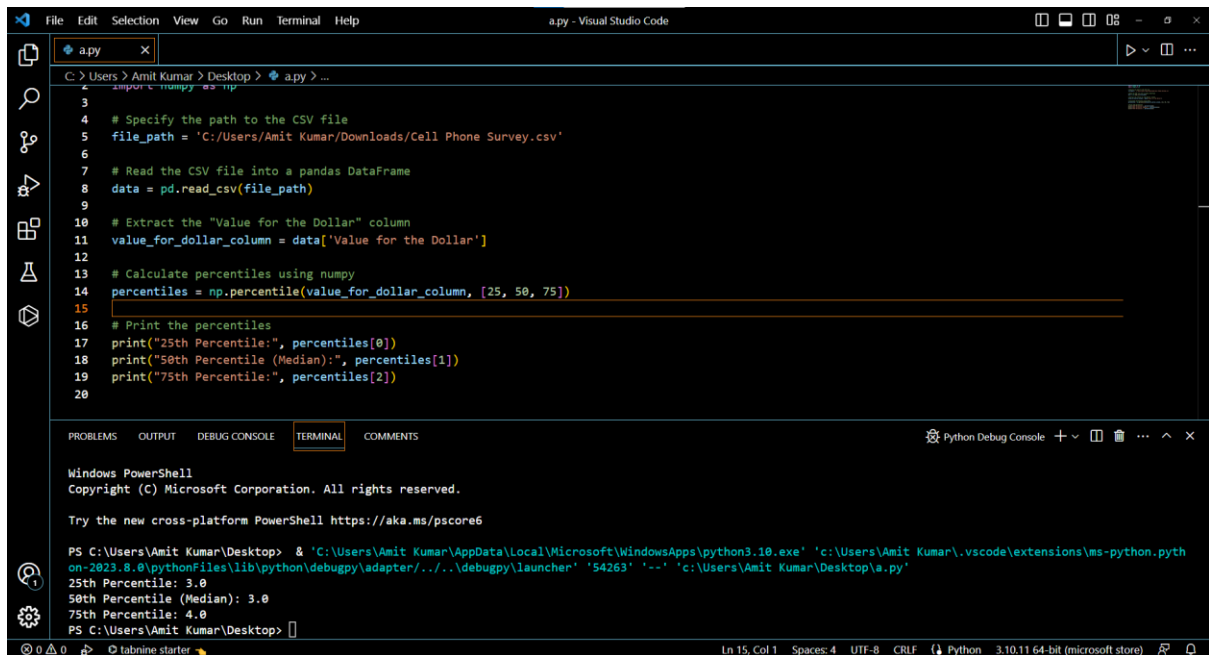
# Specify the path to the CSV file
file_path = 'C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv'

# Read the CSV file into a pandas DataFrame
data = pd.read_csv(file_path)

# Extract the "Value for the Dollar" column
value_for_dollar_column = data['Value for the Dollar']

# Calculate percentiles using numpy
percentiles = np.percentile(value_for_dollar_column, [25, 50, 75])

# Print the percentiles
print("25th Percentile:", percentiles[0])
print("50th Percentile (Median):", percentiles[1])
print("75th Percentile:", percentiles[2])
```



The screenshot shows the Visual Studio Code interface with a Python file named 'a.py' open. The code in the editor calculates percentiles for a CSV file. The terminal at the bottom shows the command prompt output, which includes the execution of the script and the resulting percentiles.

```
File Edit Selection View Go Run Terminal Help
a.py - Visual Studio Code

C:\Users\Amit Kumar\Desktop> a.py > ...
3
4 # Specify the path to the CSV file
5 file_path = 'C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv'
6
7 # Read the CSV file into a pandas DataFrame
8 data = pd.read_csv(file_path)
9
10 # Extract the "Value for the Dollar" column
11 value_for_dollar_column = data['Value for the Dollar']
12
13 # Calculate percentiles using numpy
14 percentiles = np.percentile(value_for_dollar_column, [25, 50, 75])
15
16 # Print the percentiles
17 print("25th Percentile:", percentiles[0])
18 print("50th Percentile (Median):", percentiles[1])
19 print("75th Percentile:", percentiles[2])
20

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL COMMENTS
Python Debug Console + - [ ] ... ^ x

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

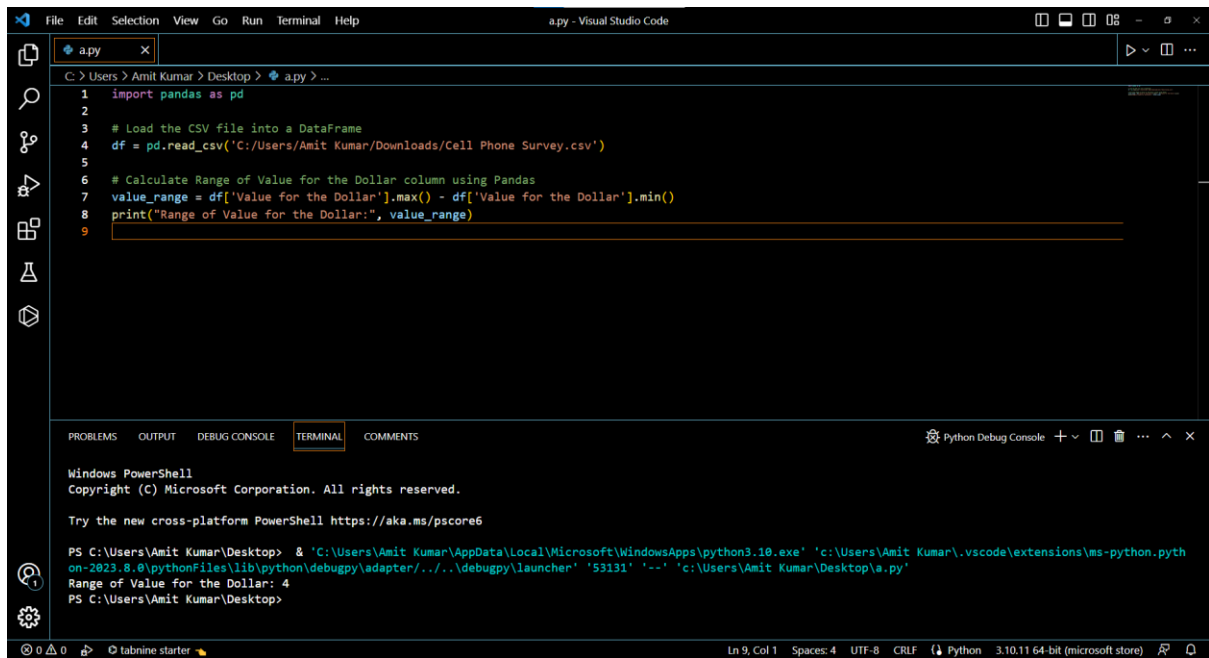
PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '54263' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
25th Percentile: 3.0
50th Percentile (Median): 3.0
75th Percentile: 4.0
PS C:\Users\Amit Kumar\Desktop>
```

● Calculate Range of Value for the Dollar column using Pandas.

```
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

# Calculate Range of Value for the Dollar column using Pandas
value_range = df['Value for the Dollar'].max() - df['Value for the Dollar'].min()
print("Range of Value for the Dollar:", value_range)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named `a.py` with the following code:

```
1 import pandas as pd
2
3 # Load the CSV file into a DataFrame
4 df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
5
6 # Calculate Range of Value for the Dollar column using Pandas
7 value_range = df['Value for the Dollar'].max() - df['Value for the Dollar'].min()
8 print("Range of Value for the Dollar:", value_range)
9
```

The bottom panel shows the TERMINAL tab with the following output:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/powershell

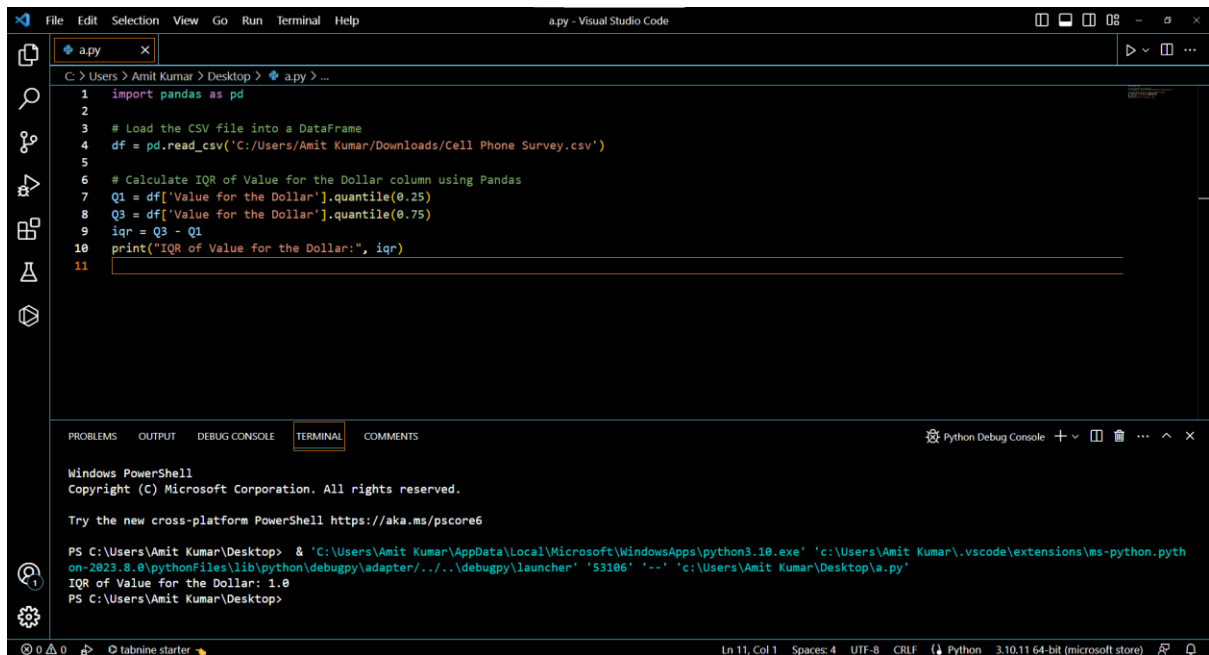
PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '53131' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
Range of Value for the Dollar: 4
PS C:\Users\Amit Kumar\Desktop>
```

● Calculate IQR of Value for the Dollar column using Pandas.

```
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

# Calculate IQR of Value for the Dollar column using Pandas
Q1 = df['Value for the Dollar'].quantile(0.25)
Q3 = df['Value for the Dollar'].quantile(0.75)
iqr = Q3 - Q1
print("IQR of Value for the Dollar:", iqr)
```



The screenshot shows the Visual Studio Code interface. The editor window displays a Python script named `a.py` with the following code:

```
1 import pandas as pd
2
3 # Load the CSV file into a DataFrame
4 df = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
5
6 # Calculate IQR of Value for the Dollar column using Pandas
7 Q1 = df['Value for the Dollar'].quantile(0.25)
8 Q3 = df['Value for the Dollar'].quantile(0.75)
9 iqr = Q3 - Q1
10 print("IQR of Value for the Dollar:", iqr)
11
```

The terminal window at the bottom shows the execution of the script using PowerShell:

```
PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.python-2023.8.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '53106' '--' 'c:\Users\Amit Kumar\Desktop\a.py'
IQR of Value for the Dollar: 1.0
PS C:\Users\Amit Kumar\Desktop>
```

● Hypothesis Testing - Using the data in the Cell Phone Survey dataset, apply ANOVA to determine if the mean response for Value for dollar is the same for different types of cell phones.

```
import pandas as pd
from scipy.stats import f_oneway

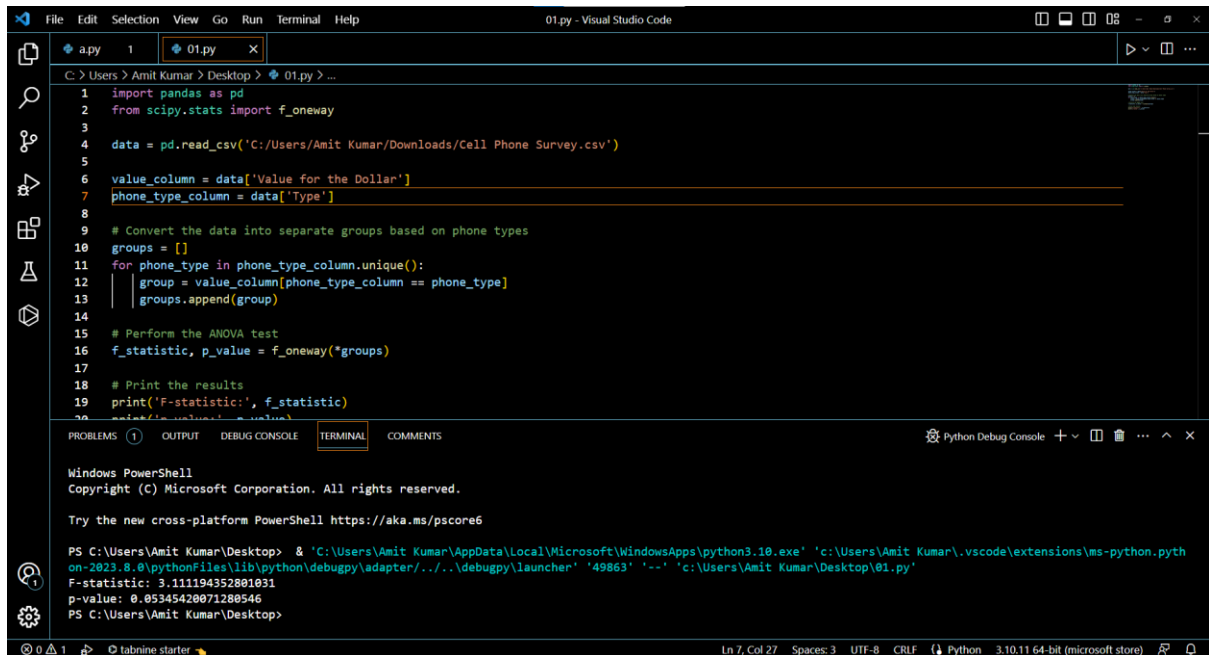
data = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')

value_column = data['Value for the Dollar']
phone_type_column = data['Type']

# Convert the data into separate groups based on phone types
groups = []
for phone_type in phone_type_column.unique():
    group = value_column[phone_type_column == phone_type]
    groups.append(group)

# Perform the ANOVA test
f_statistic, p_value = f_oneway(*groups)

# Print the results
print('F-statistic:', f_statistic)
print('p-value:', p_value)
```



The image shows a Visual Studio Code window with a Python file named `01.py` open. The script performs an ANOVA test using pandas and scipy. The terminal output shows the execution of the script, displaying the F-statistic and p-value.

```
File Edit Selection View Go Run Terminal Help
01.py - Visual Studio Code

C:\Users\Amit Kumar\Desktop> 01.py > ...
1 import pandas as pd
2 from scipy.stats import f_oneway
3
4 data = pd.read_csv('C:/Users/Amit Kumar/Downloads/Cell Phone Survey.csv')
5
6 value_column = data['Value for the Dollar']
7 phone_type_column = data['Type']
8
9 # Convert the data into separate groups based on phone types
10 groups = []
11 for phone_type in phone_type_column.unique():
12     group = value_column[phone_type_column == phone_type]
13     groups.append(group)
14
15 # Perform the ANOVA test
16 f_statistic, p_value = f_oneway(*groups)
17
18 # Print the results
19 print('F-statistic:', f_statistic)
20 print('p-value:', p_value)
```

PROBLEMS (1) OUTPUT DEBUG CONSOLE TERMINAL COMMENTS Python Debug Console + - [] ... ^ x

Windows PowerShell
Copyright (c) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell <https://aka.ms/pscore6>

```
PS C:\Users\Amit Kumar\Desktop> & 'C:\Users\Amit Kumar\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Amit Kumar\.vscode\extensions\ms-python.pyth
on-2023.8.0\pythonFiles\lib\python\debugpy\adapter\...\debugpy\launcher' '49863' '--' 'c:\Users\Amit Kumar\Desktop\01.py'
F-statistic: 3.111194352801031
p-value: 0.05345420071280546
PS C:\Users\Amit Kumar\Desktop>
```

Ln 7, Col 27 Spaces: 3 UTF-8 CRLF Python 3.10.11 64-bit (microsoft store)