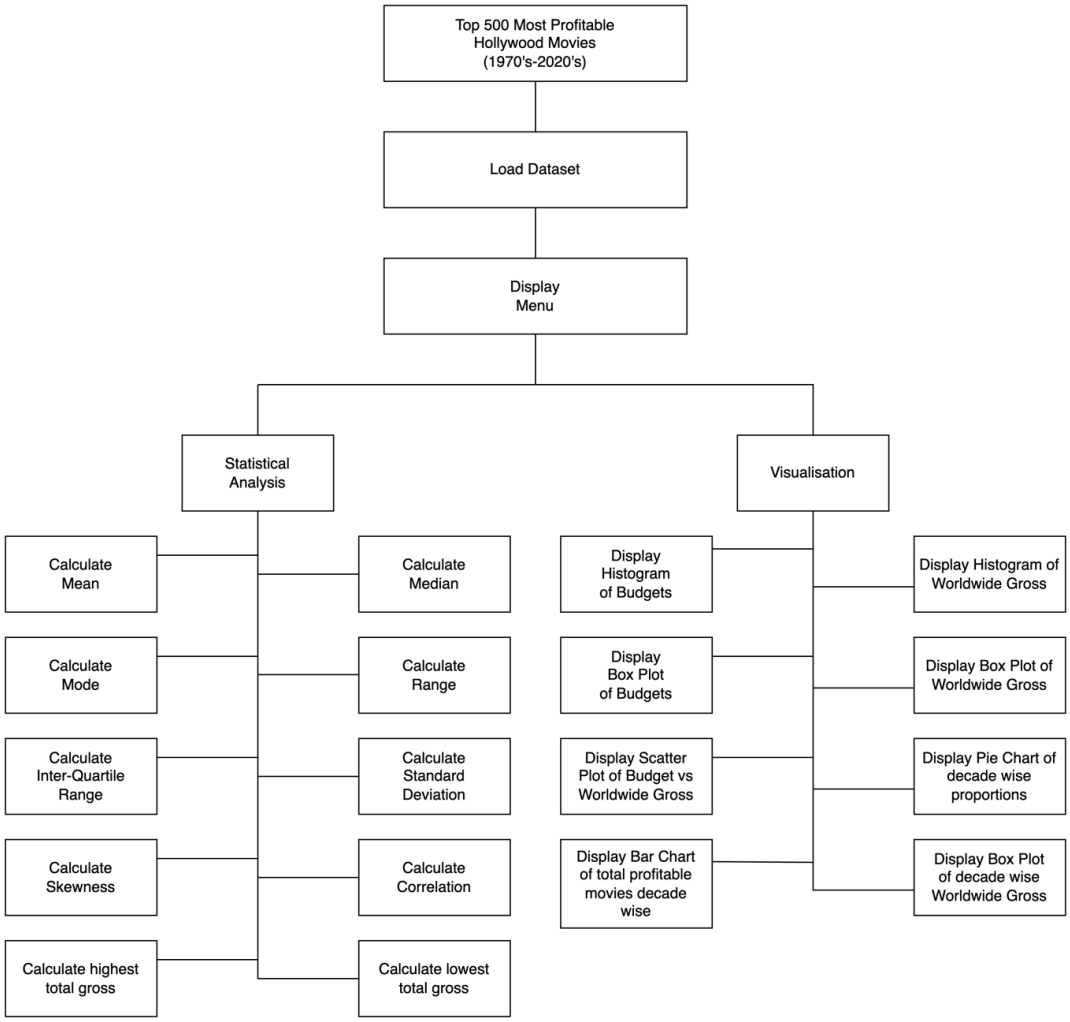
Program Design: Structure Chart

A Structure Chart showing the Top-Down Modular Design of your program.



Unit Testing

Include a Test Case for each User-Defined Statistics Function

Test Case ID	101
Function Tested	calculate_mean
Test Case Description	Tests the calculate_mean function by verifying if the calculated Mean of a numerical list matches the expected result.
Test Data (Inputs)	[11, 12, 13, 14, 15]
Expected Results	13
Actual Results	13
Pass/Fail	Pass

Test Case ID	102
Function Tested	calculate_median
Test Case Description	Tests the calculate_median function by verifying if the calculated Median of a numerical list matches the expected result.
Test Data (Inputs)	[1, 2, 3, 4, 5]
Expected Results	3
Actual Results	3
Pass/Fail	Pass

Test Case ID	103
Function Tested	calculate_mode
Test Case Description	Tests the calculate_mode function by verifying if the calculated Mode of a numerical list matches the expected result.
Test Data (Inputs)	[1, 2, 2, 3, 4, 4, 4, 5, 5, 5, 5]
Expected Results	5
Actual Results	5
Pass/Fail	Pass

Test Case ID	104
Function Tested	calculate_range
Test Case Description	Tests the calculate_range function by verifying if the calculated Range of a numerical list matches the expected result.
Test Data (Inputs)	[1, 5, 9, 3, 7]
Expected Results	8
Actual Results	8
Pass/Fail	Pass

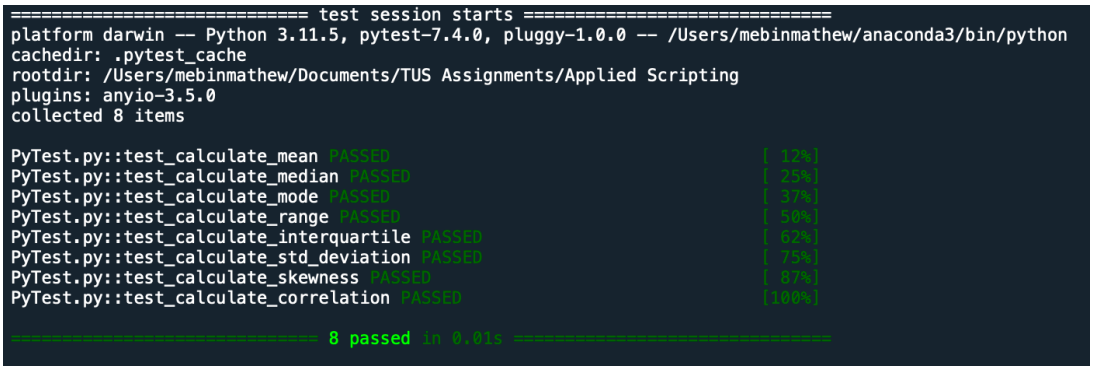
Test Case ID	105
Function Tested	calculate_interquartile
Test Case Description	Tests the calculate_interquartile function by verifying if the calculated Inter-Quartile Range of a numerical list matches the expected result.
Test Data (Inputs)	[1.1, 2, 3, 4, 5]
Expected Results	2.95
Actual Results	2.95
Pass/Fail	Pass

Test Case ID	106
Function Tested	calculate_std_deviation
Test Case Description	Tests the calculate_std_deviation function by verifying if the calculated Standard Deviation of a numerical list matches the expected result.
Test Data (Inputs)	[1, 2, 3, 4, 5.5]
Expected Results	1.74
Actual Results	1.74
Pass/Fail	Pass

Test Case ID	107
Function Tested	calculate_skewness
Test Case Description	Tests the calculate_skewness function by verifying if the calculated Skewness (using Mode and Median) of a numerical list matches the expected result.
Test Data (Inputs)	[1, 2, 2, 3, 3, 3, 4, 4, 4, 4]
Expected Results	-0.95
Actual Results	-0.95
Pass/Fail	Pass

Test Case ID	108
Function Tested	calculate_correlation
Test Case Description	Tests the calculate_correlation function by verifying if the calculated Correlation between two numerical lists matches the expected result.
Test Data (Inputs)	[10, 20, 30, 40, 50], [5, 15, 25, 35, 45]
Expected Results	1.0
Actual Results	1.0
Pass/Fail	Pass

b) Screenshot of the PyTest output in Verbose Mode



User Manual

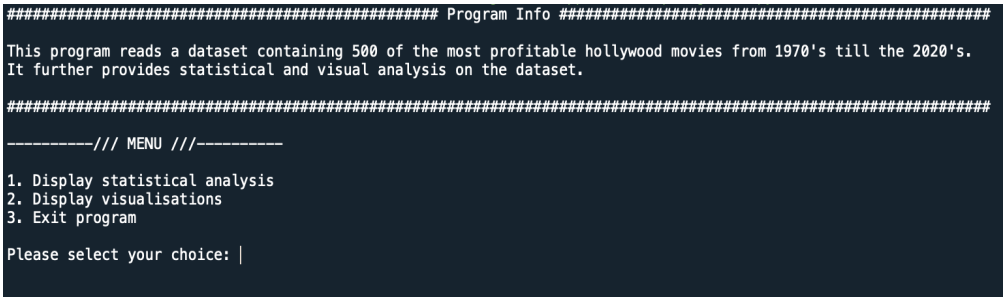
Describe in detail the program's *menu system* using output screenshots with appropriate descriptions.

1. **Running the program**

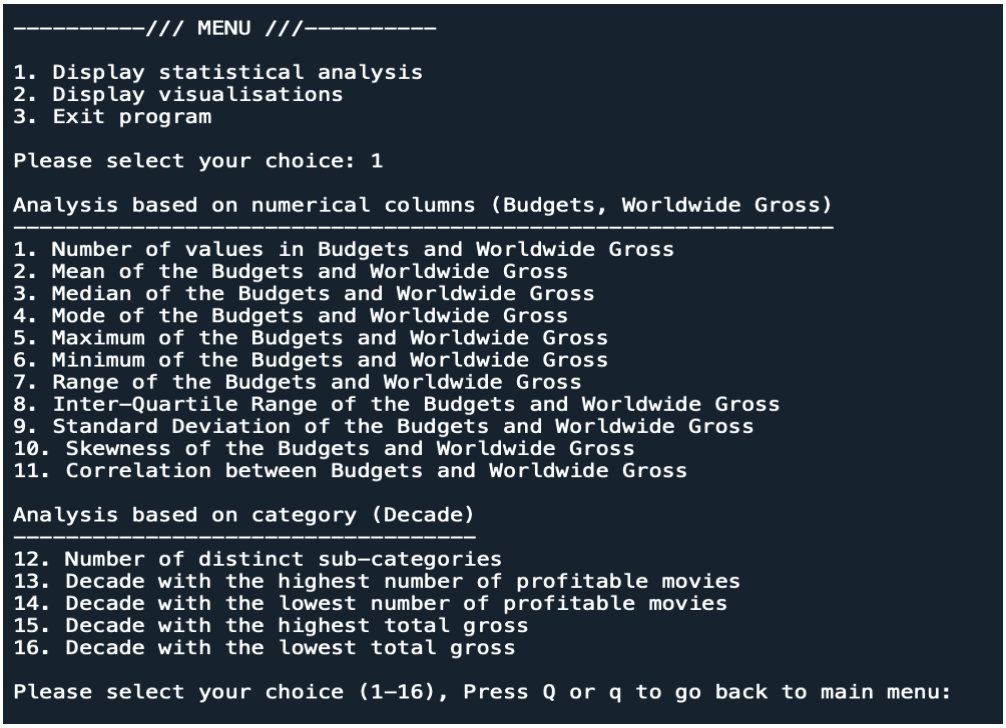
   Open the "main.py" file in PyCharm or Spyder and make sure that the "dataset.csv" file is in the same folder as the "main.py" file. Run the program.

2. **Navigating through the main menu.**

   The main menu displays a brief description of the program along with 3 options. This will be the first screen immediately after the program is run.



Select 1 to display a sub menu of statistical analysis based on numerical columns (budgets and worldwide) and category column (decade) in the dataset.



Select 2 to display a sub menu of visualisations based on numerical columns (budgets and worldwide) and category column (decade) in the dataset.

```
----------/// MENU ///----------

1. Display statistical analysis
2. Display visualisations
3. Exit program

Please select your choice: 2

Visualisations based on numerical columns (Budgets, Worldwide Gross)
-------------------------------------------------------------------
1. Histogram of Budgets
2. Histogram of Worldwide Gross
3. Box plot of Budgets
4. Box plot of Worldwide Gross
5. Scatter plot of Budget vs Worldwide Gross

Visualisations based on category (Decade)
-----------------------------------------
6. Pie chart showing the percentage number of profitable movies in each decade
7. Bar chart showing the total of the profitable movies in each decade
8. Box plots of the worldwide grosses for each decade

Please select your choice (1-8), Press Q or q to go back to main menu:
```

Select 3 to exit program.

```
----------/// MENU ///----------

1. Display statistical analysis
2. Display visualisations
3. Exit program

Please select your choice: 3

Program is exiting... Good Bye!
```

If an **invalid option** is entered, a message will be displayed accordingly, and the main menu will be displayed again.

```
----------/// MENU ///----------

1. Display statistical analysis
2. Display visualisations
3. Exit program

Please select your choice: 4

Invalid Choice, Please try again...



----------/// MENU ///----------

1. Display statistical analysis
2. Display visualisations
3. Exit program

Please select your choice: |
```

3. **Navigating through the sub menus**

   The **sub-menu 1** will be displayed on selecting **option 1** in the main menu. This menu displays 16 options to perform statistical analysis on numerical columns (budgets and worldwide gross) and category column (decade) in the dataset.

   Select any option from **1 - 16** to display each analysis.

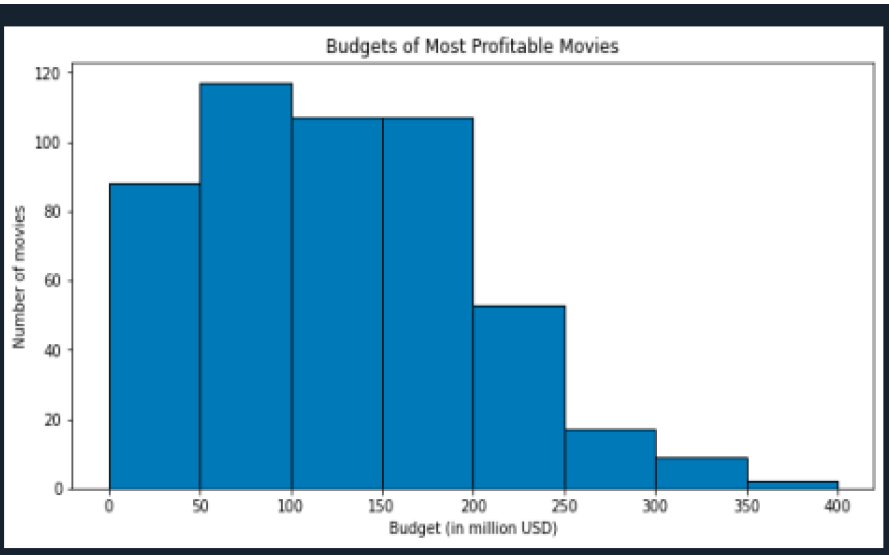
```
Analysis based on numerical columns (Budgets, Worldwide Gross)
--------------------------------------------------------------
1. Number of values in Budgets and Worldwide Gross
2. Mean of the Budgets and Worldwide Gross
3. Median of the Budgets and Worldwide Gross
4. Mode of the Budgets and Worldwide Gross
5. Maximum of the Budgets and Worldwide Gross
6. Minimum of the Budgets and Worldwide Gross
7. Range of the Budgets and Worldwide Gross
8. Inter-Quartile Range of the Budgets and Worldwide Gross
9. Standard Deviation of the Budgets and Worldwide Gross
10. Skewness of the Budgets and Worldwide Gross
11. Correlation between Budgets and Worldwide Gross

Analysis based on category (Decade)
-----------------------------------
12. Number of distinct sub-categories
13. Decade with the highest number of profitable movies
14. Decade with the lowest number of profitable movies
15. Decade with the highest total gross
16. Decade with the lowest total gross

Please select your choice (1-16), Press Q or q to go back to main menu:

1



Number of values in budget: 500
Number of values in Worldwide Gross: 500

Press any key to display sub-menu again...
```

The program displays each analysis one by one, so **press any key** to display the sub menu again and select other options.

```
Press any key to display sub-menu again...



Analysis based on numerical columns (Budgets, Worldwide Gross)
--------------------------------------------------------------
1. Number of values in Budgets and Worldwide Gross
2. Mean of the Budgets and Worldwide Gross
3. Median of the Budgets and Worldwide Gross
4. Mode of the Budgets and Worldwide Gross
5. Maximum of the Budgets and Worldwide Gross
6. Minimum of the Budgets and Worldwide Gross
7. Range of the Budgets and Worldwide Gross
8. Inter-Quartile Range of the Budgets and Worldwide Gross
9. Standard Deviation of the Budgets and Worldwide Gross
10. Skewness of the Budgets and Worldwide Gross
11. Correlation between Budgets and Worldwide Gross

Analysis based on category (Decade)
-----------------------------------
12. Number of distinct sub-categories
13. Decade with the highest number of profitable movies
14. Decade with the lowest number of profitable movies
15. Decade with the highest total gross
16. Decade with the lowest total gross

Please select your choice (1-16), Press Q or q to go back to main menu:

12



Number of distinct sub-categories: 6 (2000's, 2010's, 1990's, 1970's, 1980's, 2020's)

Press any key to display sub-menu again...
```

The **sub menu 2** will be displayed on selecting **option 2** in the main menu. This menu displays 8 options to display visualisations based on numerical columns (budgets and worldwide gross) and category column (decade) in the dataset.

Select any option from **1 - 8** to display each visualisation.

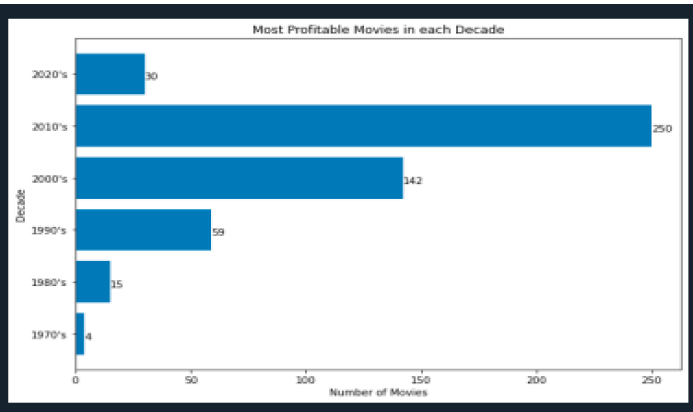


Help | Variable Explorer | Plots | Files

Console 1/A

```
Visualisations based on numerical columns (Budgets, Worldwide Gross)
-------------------------------------------------------------------
1. Histogram of Budgets
2. Histogram of Worldwide Gross
3. Box plot of Budgets
4. Box plot of Worldwide Gross
5. Scatter plot of Budget vs Worldwide Gross

Visualisations based on category (Decade)
-----------------------------------------
6. Pie chart showing the percentage number of profitable movies in each decade
7. Bar chart showing the total of the profitable movies in each decade
8. Box plots of the worldwide grosses for each decade

Please select your choice (1-8), Press Q or q to go back to main menu:

1

Press any key to display sub-menu again...
```

The program displays each visualisation one by one, so **press any key** to display the sub menu again and select other options.



Help | Variable Explorer | Plots | Files

Console 1/A

```
Press any key to display sub-menu again...


Visualisations based on numerical columns (Budgets, Worldwide Gross)
-------------------------------------------------------------------
1. Histogram of Budgets
2. Histogram of Worldwide Gross
3. Box plot of Budgets
4. Box plot of Worldwide Gross
5. Scatter plot of Budget vs Worldwide Gross

Visualisations based on category (Decade)
-----------------------------------------
6. Pie chart showing the percentage number of profitable movies in each decade
7. Bar chart showing the total of the profitable movies in each decade
8. Box plots of the worldwide grosses for each decade

Please select your choice (1-8), Press Q or q to go back to main menu:

7
```

In both sub menus, selecting **Q or q** will go back to main menu.

```
Visualisations based on numerical columns (Budgets, Worldwide Gross)
-------------------------------------------------------------------
1. Histogram of Budgets
2. Histogram of Worldwide Gross
3. Box plot of Budgets
4. Box plot of Worldwide Gross
5. Scatter plot of Budget vs Worldwide Gross

Visualisations based on category (Decade)
-----------------------------------------
6. Pie chart showing the percentage number of profitable movies in each decade
7. Bar chart showing the total of the profitable movies in each decade
8. Box plots of the worldwide grosses for each decade

Please select your choice (1-8), Press Q or q to go back to main menu:

q

----------/// MENU ///----------

1. Display statistical analysis
2. Display visualisations
3. Exit program

Please select your choice:
```

If an **invalid option** is entered in the sub menu, a message will be displayed accordingly. **Press any key** to display sub menu again.

```
Visualisations based on numerical columns (Budgets, Worldwide Gross)
-------------------------------------------------------------------
1. Histogram of Budgets
2. Histogram of Worldwide Gross
3. Box plot of Budgets
4. Box plot of Worldwide Gross
5. Scatter plot of Budget vs Worldwide Gross

Visualisations based on category (Decade)
-----------------------------------------
6. Pie chart showing the percentage number of profitable movies in each decade
7. Bar chart showing the total of the profitable movies in each decade
8. Box plots of the worldwide grosses for each decade

Please select your choice (1-8), Press Q or q to go back to main menu:

33

Invalid Choice!

Press any key to display sub-menu again...
```

Analysis, Visualisation, Results and Conclusions

a): Analyse the Two Numeric Columns using a List

	Budget **(in million USD)**	***Worldwide Gross*** **(in million USD)**
Number of values	500	500
Total	NA	NA
Mean	121.04	585.49
Median	120.00	468.50
Mode	200.00	352.00
Maximum	379.00	2923.00
Minimum	0.01	4.00
Range	378.99	2919.0
Inter-Quartile Range	100.00	355.5
Standard Deviation	71.11	343.87
Median Skewness	0.04	1.02
Mode Skewness	-1.11	0.68
Correlation	0.53	

Output Screenshot(s) showing the above results.

```
Please select your choice (1-16), Press Q or q to go back to main menu:

1



Number of values in budget: 500
Number of values in Worldwide Gross: 500
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

2



Mean of the budgets of the most profitable movies: 121.04 million USD
Mean of the worldwide grosses of the most profitable movies: 585.49 million USD
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

3



Median of the budgets of the most profitable movies: 120.00 million USD
Median of the worldwide grosses of the most profitable movies: 468.50 million USD
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

4



Mode of the budgets of the most profitable movies: 200.00 million USD
Mode of the worldwide grosses of the most profitable movies: 352.00 million USD
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

5



Maximum budget of the most profitable movie: 379.00 million USD (Pirates of the Caribbean: On Stranger Tides)
Maximum worldwide gross of the most profitable movie: 2923.00 million USD (Avatar)
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

6



Minimum budget of the most profitable movie: 0.01 million USD (Paranormal Activity )
Minimum worldwide gross of the most profitable movie: 4.00 million USD (Sleight)
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

7



Range of the budgets of the most profitable movies: 378.99 million USD
Range of the worldwide grosses of the most profitable movies: 2919.0 million USD
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

8



Inter-quartile range of budgets of the most profitable movies: 100.00 million USD
Inter-quartile range of worldwide grosses of the most profitable movies: 355.5 million USD
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

9



Standard Deviation of the budgets of the most profitable movies: 71.11 million USD
Standard Deviation of the worldwide grosses of the most profitable movies: 343.87 million USD
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

10



Pearson Mode Skewness of the budgets of the most profitable movies: -1.11
Alternative Pearson Mode Skewness of the budgets of the most profitable movies: 0.04
Pearson Mode Skewness of the worldwide grosses of the most profitable movies: 0.68
Alternative Pearson Mode Skewness of the worldwide grosses of the most profitable movies: 1.02
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:

11



Correlation value of budget with worldwide gross: 0.53
```

Visualisations

Histograms

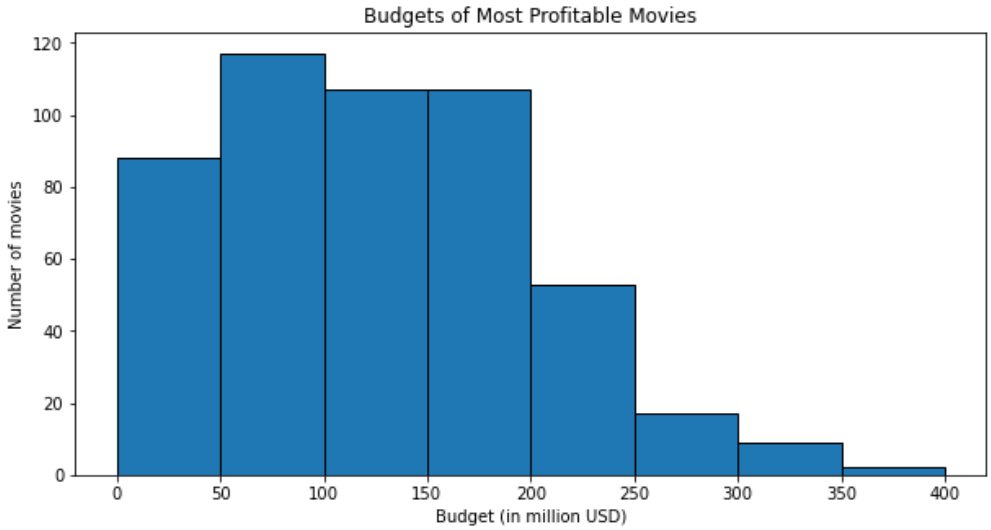


Figure 1: Histogram of Budgets

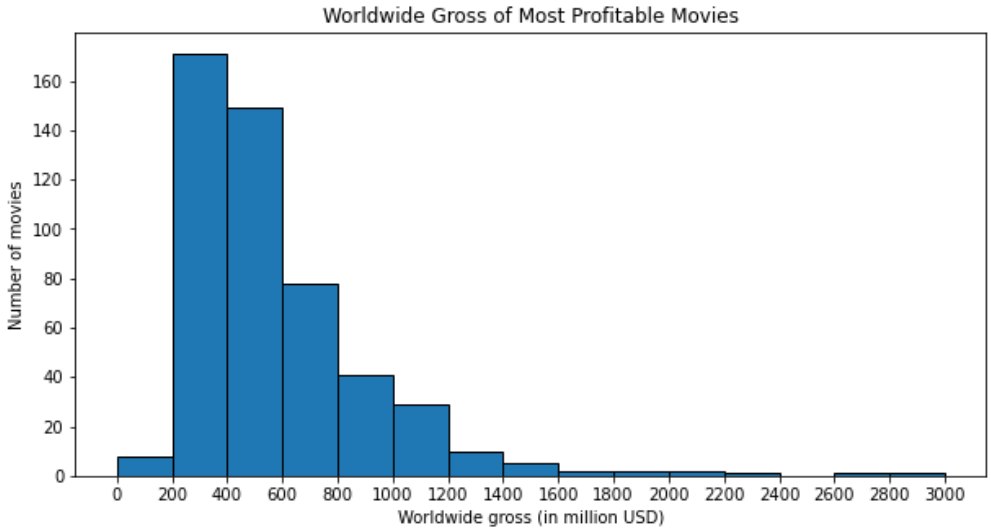


Figure 2: Histogram of Worldwide Gross

Box Plots

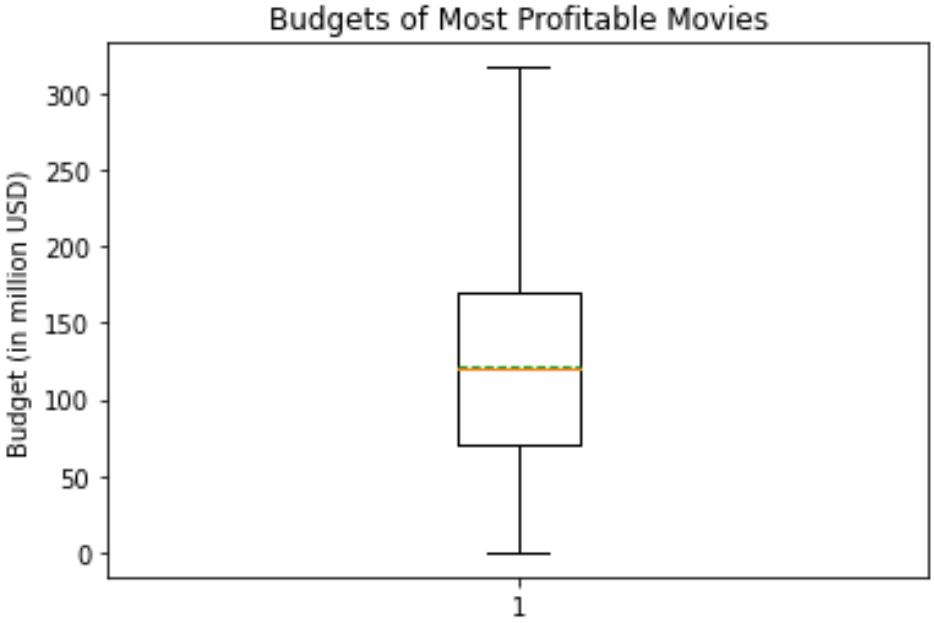


Figure 3: Box Plot of Budgets

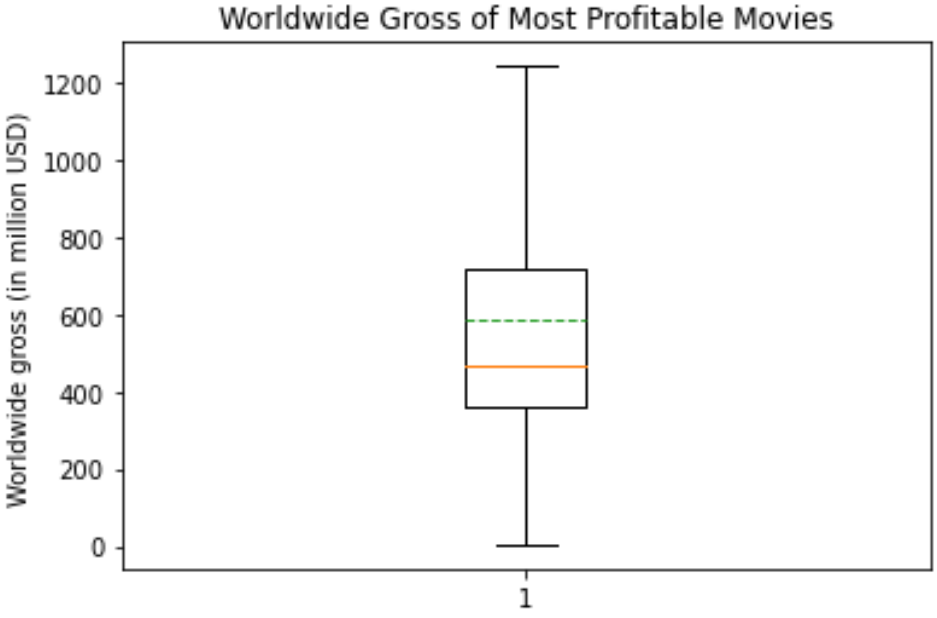


Figure 4: Box plot of Worldwide Gross

Scatter Plot

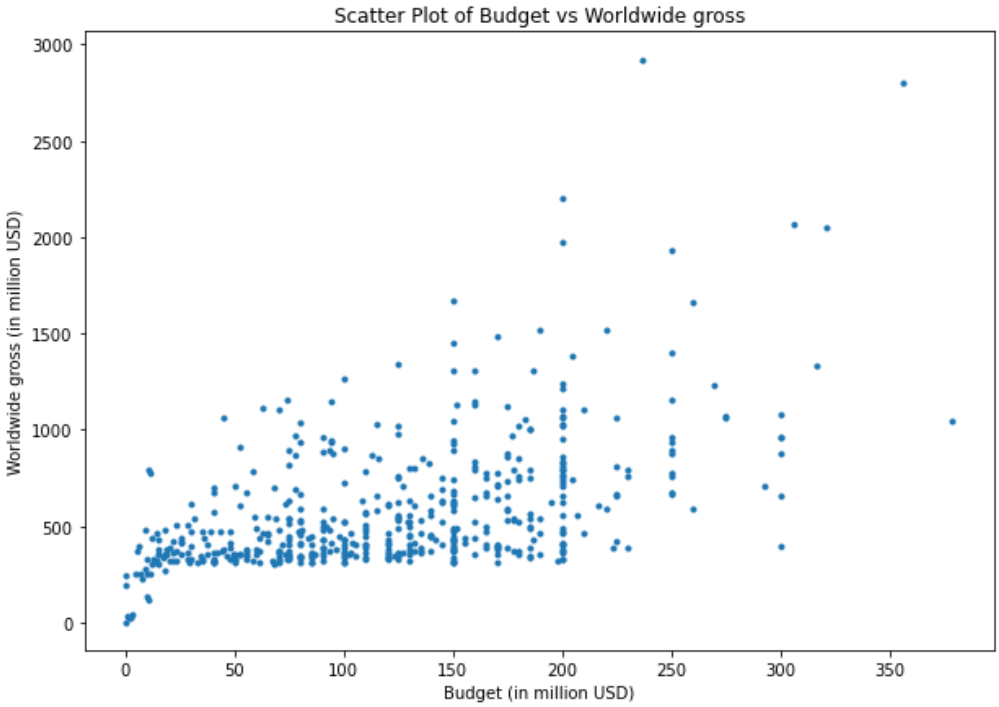


Figure 5: Scatter Plot of Budgets vs Worldwide Gross

b) Analyse by Category using a Dictionary

Category Name	***Decade***
Number of Subcategories	*6*
Subcategory with highest frequency	*2010's* (*250*)
Subcategory with lowest frequency	*1970's* (*4*)

Analysis by Category	***Total - Worldwide Gross***
Subcategory with highest total	*2010's* (*161.18*)
Subcategory with lowest total	*1970's* (*2.09*)

Output Screenshot(s) showing the above results.

```
Please select your choice (1-16), Press Q or q to go back to main menu:
12


Number of distinct sub-categories: 6 (2000's, 2010's, 1990's, 1970's, 1980's, 2020's)
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:
13


The decade with the highest number of profitable movies is the 2010's with 250 movies
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:
14


The decade with the lowest number of profitable movies is the 1970's with 4 movies
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:
15


The highest grossing decade is the 2010's with a total of 161.18 Billion USD
```

```
Please select your choice (1-16), Press Q or q to go back to main menu:
16


The lowest grossing decade is the 1970's with a total of 2.09 Billion USD
```

Visualisations

Pie chart showing the percentage number of values in each category.

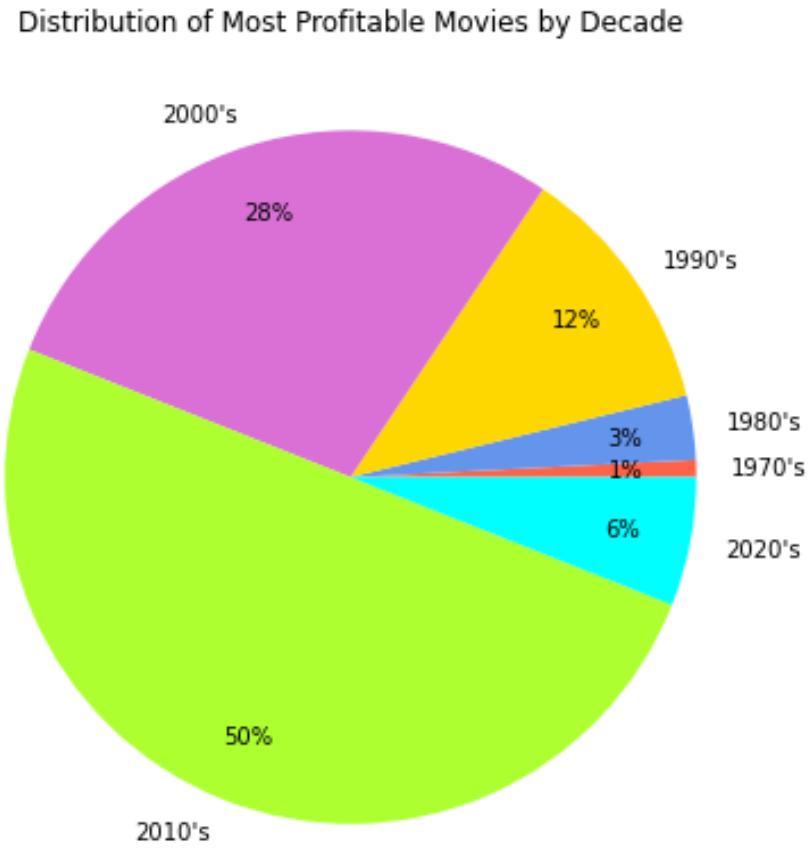


Figure 6: Pie Chart

Bar chart showing the total/average of the values in each sub-category.

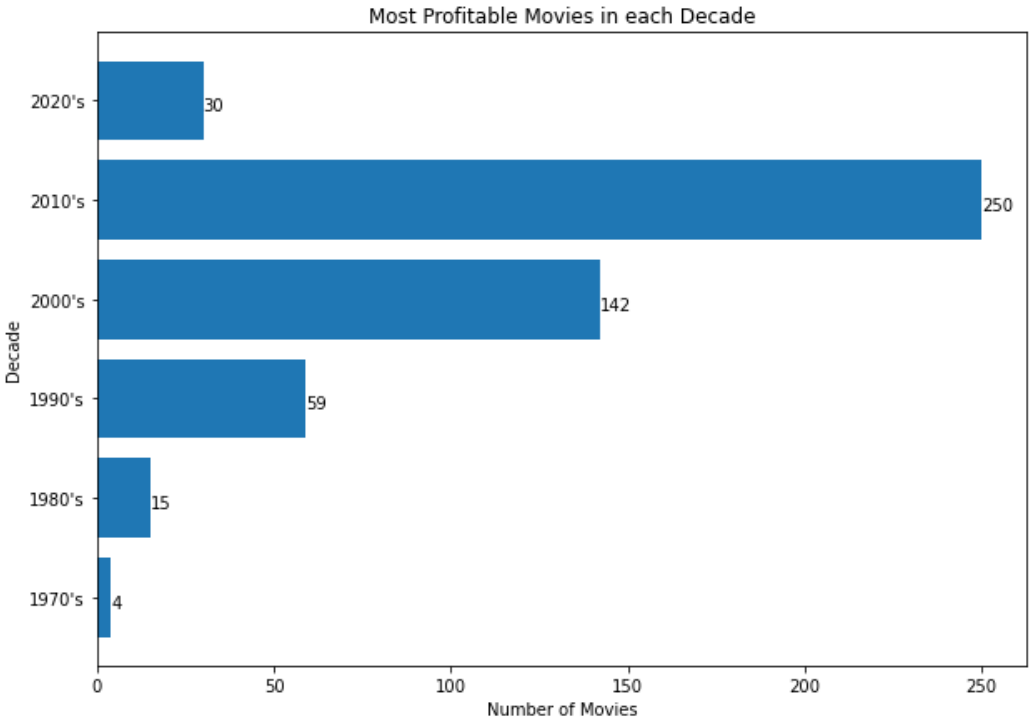


Figure 7: Horizontal Bar Chart

Box plots of the values for each sub-category, done in a single visualisation

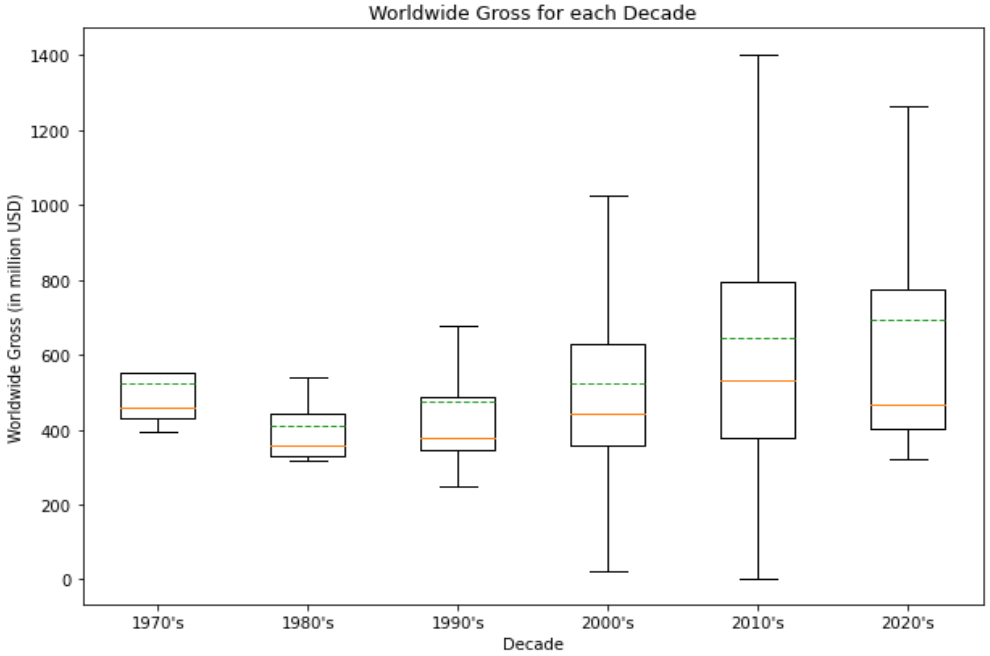


Figure 8: Box Plots