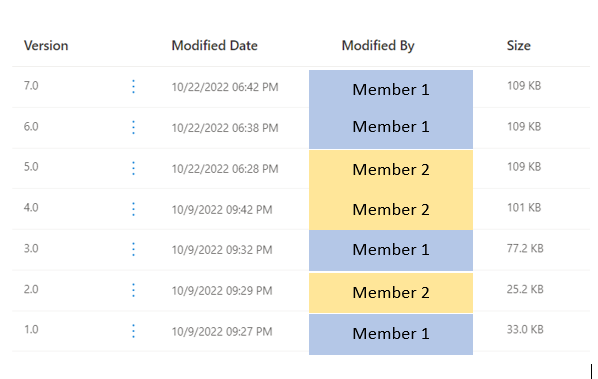
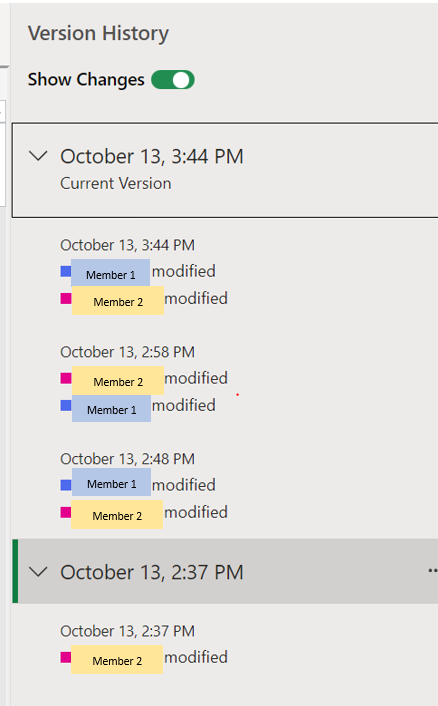
**Final Assignment instructions:**

**Form a team of 2 to work on this final assignment together online using either Zoom or MS Teams. You may choose to use either MS Excel OR Python to complete this assignment.**

**OneDrive:**

Create a OneDrive folder and share it with the team. Edit and save the work in only 1 document in OneDrive. Take a screenshot showing the edit history by all members within the team and versioning in OneDrive and include it in the submission template. Please see example below:



**Documents to be submitted in Dropbox:**

1. Assignment completed using MS Excel **OR** Python

For assignment completed using **MS Excel:**

Please submit the merged Excel dataset including the visualization charts within the MS Excel file (1 MS Excel file per team)

For assignment completed using **Python**:

please submit the Python notebook (1 Python notebook per team) with codes used to generate the visualization charts. Please include the team members’ names under comments of the charts.

1. Assignment Submission Template (MS Word document)

Completed this assignment submission template as per the instructions given.

**Assignment Answers:**

**Part 1a**

**IDENTIFY data columns**

Before starting on the assignment, each team member is required to identify the data columns from the merged dataset (at least 2 data columns per member) that you will be working on for this assignment in the table below.

|  |  |  |
| --- | --- | --- |
|  | **Team Member 1** | **Team Member 2** |
| **Name** | Justine Carlo Villa Ilao | Muhammad Farhan Bin Ahmad |
| **Student ID** | **2200571** | **2200544** |
| **Data Column 1** *(required)* | Flat Type | Flat type |
| **Data Column 2** *(required)* | Flat Model | Flat Model |
| **Data Column 3** *(optional)* | Floor area sqm | Resale value |
| **Data Column 4** *(optional)* |  | Town |

**Part 1b**

**Team member 1: Muhammad Farhan Bin Ahmad / 2200544**

1. Using the data columns you have selected, create a visualisation chart.Paste a screenshot of your visualization chart and describe what conclusion can be drawn from the chart

*(write approx. 30–50 words).*

A graph of blue rectangular bars

Description automatically generated

Based on the graph, and filtering the graph to only count a 2 room Model A flat, we find that depending on which town you reside in, the resale price of your flat may be more/lesser than others.

1. Using the data columns selected by your team member, create another visualisation chart. Paste a screenshot of your visualization chart here and describe what conclusion can be drawn from the chart *(write approx. 30–50 words)*

A graph with blue lines

Description automatically generated

The graph tell what type of flats are available in each town in Singapore and it also shows the total sum of the resale price of each type of flat for each of the town.

***Part 1c***

**Team member 2: Justine Carlo Villa llao/ 2200571**

1. Using the data columns you have selected, create a visualisation chart. Paste a screenshot of your visualization chart and describe what conclusion can be drawn from the chart

*(write approx. 30–50 words).*

A graph with blue lines

Description automatically generated

The person’s flat floor area may differ regardless as, although they may share the same flat type, the amount of the floor area available will differ as some have more than others

1. Using the data columns selected by your team member, create another visualisation chart. Paste a screenshot of your visualization chart here and describe what conclusion can be drawn from the chart *(write approx. 30–50 words).*

A graph with blue lines

Description automatically generated

Depending on the flat model you are living, the range of area you have in flat may differ as flat like the Improved area may range from 45 – 131 sqm while the DBSS have a smaller range. Ranging from 67 – 111 sqm.

**Part 2**

**Group Discussion**

As a team, select a visualisation chart that best represents the conclusion the team wishes to draw from the given data set. Support the conclusion with one statistical calculation with description and provide a short paragraph detailing the conclusion drawn. The report should include at least (i) objective, (ii) reason for using the selected statistical calculation, and (iii) the drawn conclusion.

1. Select 1 out of the 4 visualisation charts from Part 1b/1c for this group discussion. Provide the same visualisation chart and the description here.

A graph of blue rectangular bars

Description automatically generated

Based on the graph, and filtering the graph to only count a 2 room Model A flat, we find that depending on which town you reside in, the resale price of your flat may be more/lesser than others.

1. Using the selected visualisation chart data above, choose **one** of the following corresponding statistical calculations to work on, state the result obtained and provide a brief description *(refer to Appendix A for example of answer format).*

(1) Mean (2) Median (3) Mode (4) Standard Deviation (5) Variance (6) Correlation

This calculation/formula should be included in the MS Excel merged data file / Python notebook that is to be submitted.

*(You may refer to Topic 2 - Data Exploration for additional information on the statistical calculations.)*

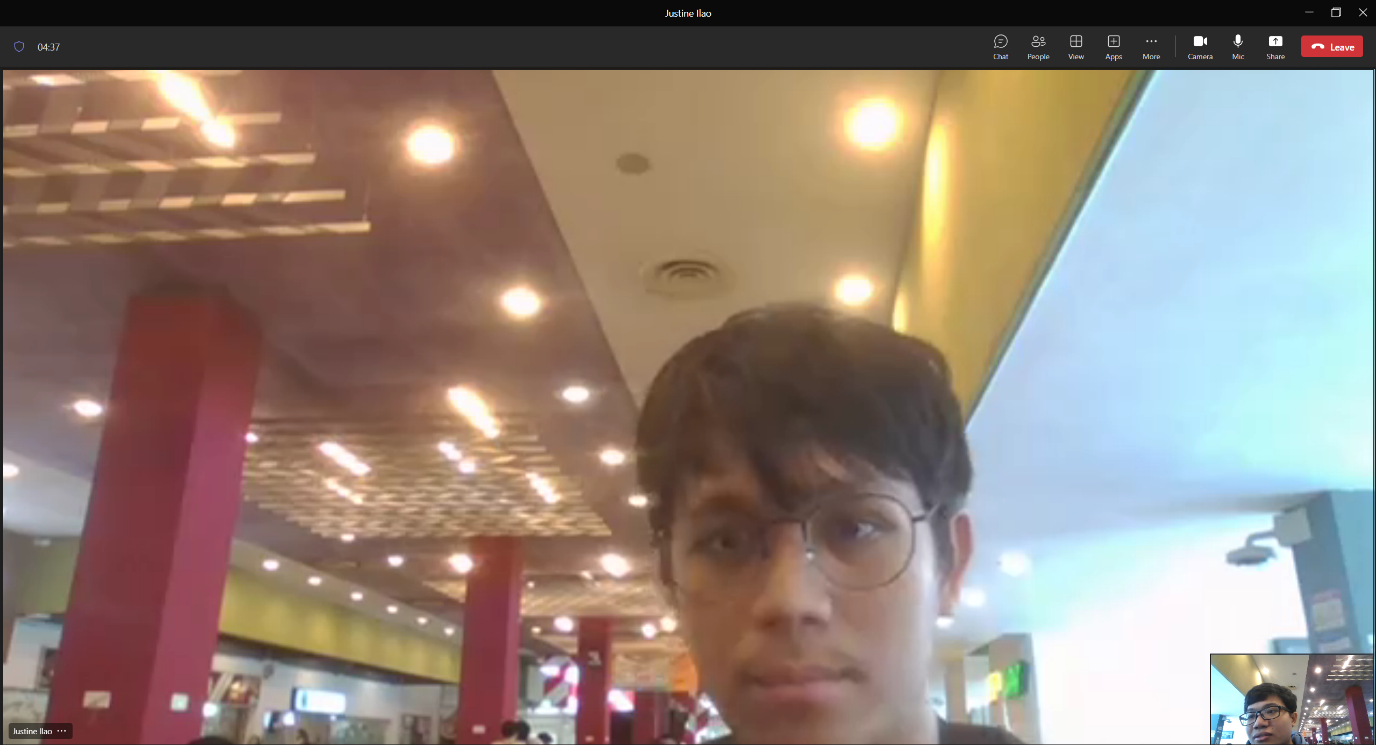
Function: =AVERAGE (D4:D51)

Result: 2084948.38

Description:

We aim to find the average resell price of a 2 room Model A flat in Singapore. And upon using the AVERAGE function (Mean), we were able to deduce that the average reselling price of a house is $358436.3636. It seems that location of where the house reside in matters as that can affect the resell cost of the flat as those in certain area cost more than those in other areas in Singapore and some area may cause the resell price to be lower than the average selling price.

1. Zoom / MS Teams screenshot (with your camera turned on):



1. Paste a screenshot of your OneDrive showing the edit history and versioning for all team members here.

A screenshot of a computer

Description automatically generated

**Appendix A**

**Example Answer Format**

Microsoft Excel answer:

Function: =MODE(A1:A100)

Result: Bedok

Description: We aim to find out which town has the most transactions during the past 5 years. By using MODE function, we find out that Bedok has the most occurrence in our merged data.

Python answer:

Function: dataMerge[“Town”].mode(1)

Result: Bedok

Description: We aim to find out which town has the most transactions during the past 5 years. By using MODE function, we find out that Bedok has the most occurrence in our merged data.

**Appendix B Additional Statistical functions in Panda Not Covered in Slides:**

* Standard Deviation:
* <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.std.html>
* Variance:
* <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.var.html>
* Correlation:
* <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.corr.html>