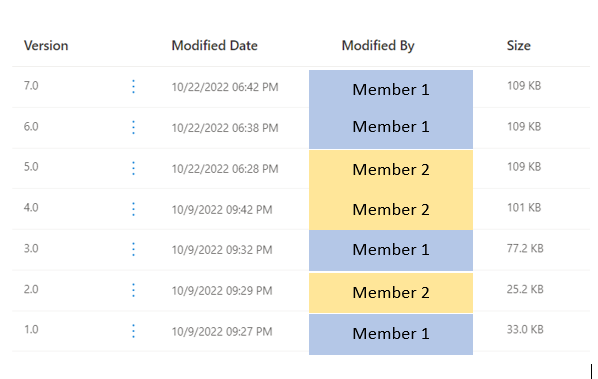
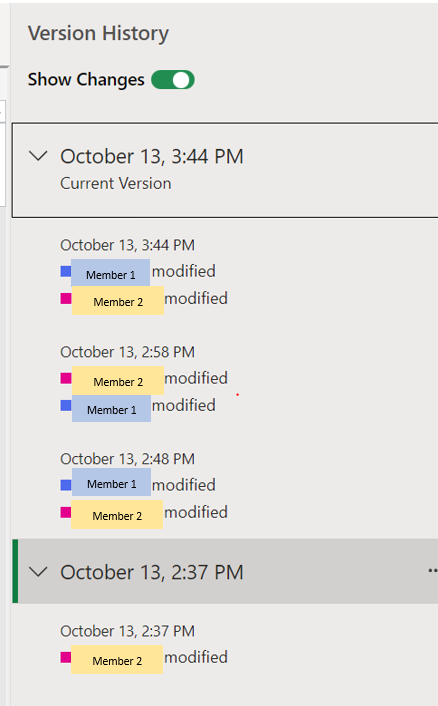
**Final Assignment instructions:**

**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 with blue lines

Description automatically generated

Based on the graph, depending on your flat model, flat type and the town you currently reside in, the resale value of your house may differ from other. Even if the type and model are the same.

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 bars

Description automatically generated

The amount of floor area per square meter the person has on their flat is dependent on the flat type of the and the flat model the person is currently residing in.

***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 orange lines

Description automatically generated

Where the seller reside plays a part in dictating the resell price of the flat as although the flat type and flat model are the same, their cost differ depending on the location the flat is at.

**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 with blue lines

Description automatically generated

Based on the graph, depending on your flat model, flat type and the town you currently reside in, the resale value of your house may differ from other. Even if the type and model are the same.

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.)*

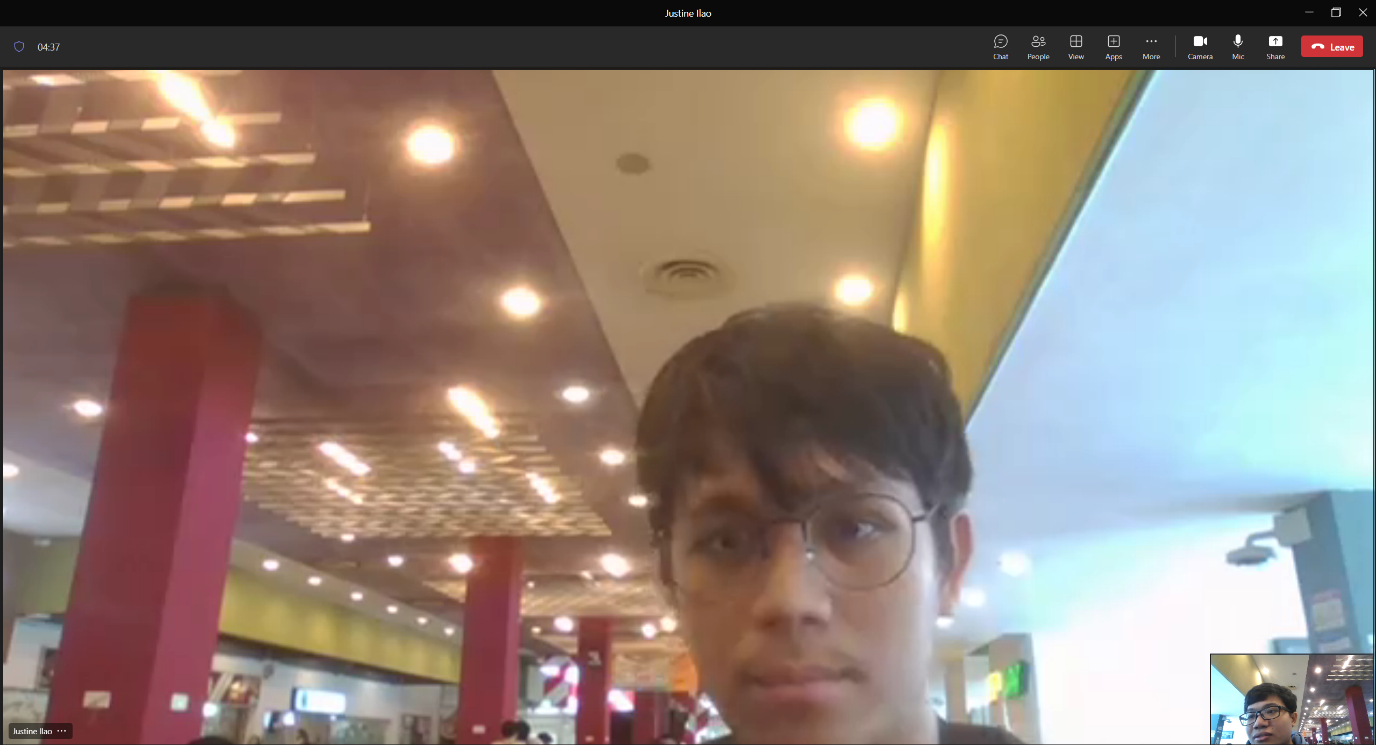
Function: =AVERAGE (D4:D149)

Result: 2084948.38

Description:

We aim to find the average resell price of a house in Singapore. And upon using the AVERAGE function (Mean), we were able to deduce that the average reselling price of a house is 2084948.38 if we take in consideration of the seller’s flat model, the flat type, which town the seller’s flat is resided in as all this does affect the selling power of the flat

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>