****

**C355 Business Intelligence**

**AY2021 Sem 2 – ASSIGNMENT**

|  |  |
| --- | --- |
| **Personal Details** | |
| **Name** | **SAMIA AFRIN** |
| **Student ID** | **20008933** |
| **Class** | **W66P** |

|  |  |  |
| --- | --- | --- |
| **Compliance Statement** | | |
| **Plagiarism**  I declare that this report is my original work. I understand that if I am suspected of plagiarism either through sharing your work or copying other’s work, I may be given an ‘**F’** grade. | | **🗸** |
| **Retention of Backup Copy**  I declare that I have a back-up electronic copy of this report for immediate submission. | | **🗸** |
| **Signature:** | **Date: 07/02/2022** | |
| **IMPORTANT:** Non-compliance to these clauses will result in unconditional rejection of your submission, resulting in failure in the assignment. | | |

**ESE Assignment**

**Worksheet**

As part of the assignment requirements, you have to complete:

|  |  |  |
| --- | --- | --- |
| **Task#** | **Task** | **Completed by** |
| Task 1 | Understand Business Requirement   * + Examine the dataset carefully to identify the Measures (what to analyze) and Dimensions (how  to analyse the identified measures) in the data set   + Define business questions | End of Lesson 10 |
| Task 2 | Data Preparation   * Clean and prepare data file using Power Query | End of Lesson 11 |
| Task 3 | Analysis and Data Visualisation   * Charts and Dashboard creation * Discovering insights and answer Business questions | End of Lesson 12 |
| Task 4 | Individual oral presentation   * Demo final dashboard in Tableau | End of Lesson 13 |

## Understand the business requirements (Complete by Lesson 10)

1. Derive SIX business questions to provide insights on the resale transactions. For each business question, identify the dimension attribute(s) and measure(s).

You are to come up with 1 to 3 dimensional business questions basing on different measures and have in mind varieties of charts to present the answers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Business Question** | **Dimension** | **Measure** | **Rationale of Question** |
| **BQ1** | **What is the most popular purchased flat type** | **- Flat Type**  **(Rows)** | **- TrnID (Count)** | **By knowing which flat type has been the most popular over the past years. We can focus on that particular flat type as it is mostly the type of flat purchased by Singaporeans** |
| **BQ2** | **Which top 10 estates has the most resale flats that has a resale price equivalent to its valuation price based on the most popular flat type** | **- Estate**  **(Filter top 10 & Rows)**  **- Flat Type (Filter)** | **- Cash Over Valuation**  **(Filter value = 0)**  **- TrnID (Count) (Column)** | **When buying a resale flat , buyers would try to buy a house closer to its valuation price instead of a price that is way above the valuation price. This allows a buyer to get a valued house at a decent pricing**  **By knowing which estate (narrowed down to top 10 choices) sells their most popular flats nearer to its valuation price , buyers can choose to buy a flat in that estate** |
| **BQ3** | **Which top 10 estate has a longer remaining lease duration on average based on the most popular flat type** | **- Estate**  **(Filter top 10 & Rows**  **- Flat Type(Filter)** | **- Remaining Lease Duration (Average)(Column)** | **The shorter the lease duration , the lower the value of the flat, the lesser amount of time lived in the flat.**  **The current buyer would not be able to spend much time in the flat as once the lease is over , they would need to evacuate the house. Hence , buyers would want a house with longer remaining lease so they can stay in the house longer.**  **Reselling the house with a short lease would be difficult as well** |
| **BQ4** | **What are the Top 5 estates that sold the most popular flat type** | **- Estate**  **(Filter)**  **- Flat Type**  **(Filter)** | **- Frequency Count (Tenant ID)**  **(Label)** | **The buyers would be able to find out which estate has the high probability of buying a 4 room flat. (narrowed down to top 5 )** |
| BQ | **What is the minimum, average and maximum resale price for the most popular flat type in the Top 5 estates that has the most popular flat type** | **- Flat Type**  **(Filter)**  **- Estate Set from BQ4 (Top 5) (Column)** | **- Resale Price (Max)**  **- Resale Price (Min)**  **- Resale Price (Average)**  **(All in Rows)** | **The buyers would be able to know what is the maximum and minimum cost that a particular flat type can go in the top 5 estate popular for selling 4 Room flats.**  **They would also know how much a particular flat type cost on average in in the top 5 estate popular for selling 4 Room flats.** |
| BQ | **What is the average floor area for the most popular flat type in the top 5 estates that has the most popular flat type** | **- Flat Type**  **- Estate**  **(All in Column)** | **- Floor Area (Average)**  **(Rows)** | **Buyers would be able to estimate how big different flat types are. This can help them plan their housing better. (E.g Buyers might want a bigger bedroom as they want to make it their work/study area as well )** |

Completed on (Date) : **20/01/2021**

**Data Preparation (Complete by Lesson 11)**

1. Examine the C355\_Resale\_Dataset given. Identify the data issues.

|  |  |
| --- | --- |
| **Data Issue (which column)** | **Propose Cleansing/Transformation Rules** |
| **Block** | **Format 🡪 Clean** |
| **Flat Type** | **Replace Values:**  **3 ROOMS 🡪 3 ROOM**  **4 ROOMS 🡪 4 ROOM** |
| **Floor Area** | **Remove Null Values** |
| **Resale Price** | **Remove Null Values** |
| **Valuation Price** | **Remove Null Values** |
| **Column 16 & Column 17** | **Remove Invalid Columns** |
| **Extract the year out from the column named “Lease Start Date” into a new column and rename it “Lease Start Year”** | **Add Column 🡪 Date 🡪 Year** |
| **Extract the year out from the column named “Lease Start Date” into a new column and rename it “Lease Start Year”** | **Add Column 🡪 Date 🡪 Year** |
| **Remove the column named “Remaining Lease Duration”** | **Home 🡪 Remove Column** |
| **Add a Custom Column , Name the column as “ Remaining Lease Duration”** | **Add Column 🡪 Custom Column**  **🡪 99 – (Resale Year - Lease Start Year)** |
| **Add a Custom Column , Name the column as “ Cash Over Valuation”** | **Add Column 🡪 Custom Column**  **🡪 Resale Price – Valuation Price)** |
| **Cash Over Valuation** | **Remove Negative Values as no one would sell their house lower than its value** |

1. Carry out the prosposed data cleaning and transformed identified in step 2 above.

Completed on (Date) : **08/01/2022**

**Analysis and Data Visualisation (Complete by Lesson 12)**

This will be done in Tableau. You should begin visualisation as soon as you complete your data cleansing. You can continue the visualization and report in Lesson 12 class.

1. Import the cleansed excel file into Tableau.
2. Do any join(s) where appropriate between the 2 datasets.
3. Ensure that the columns are of correct data types (string, numeric, date, geographic).
4. For each of the chart, consider :
5. What appropriate charts to use : Table, Bar, Pie, Line, GeoMap, TreeMap, Boxplot, Scatterplot, Bubble, Dual Axis, Area Chart, etc ?
6. Which attribute to add to colors?
7. What to add for labels?
8. What to add for tooltips?
9. What filters to use?
10. What sorting to use?
11. To achieve **aesthetic** effects, refer to resources on how to create different types of charts:

* Donut, Waterfall, Filled Map, ScatterPlots, Butterfly, etc

<https://help.tableau.com/current/pro/desktop/en-us/design_and_analyze.htm>

<https://help.tableau.com/current/pro/desktop/en-us/maps.htm>

1. Refer to some useful websites on dashboards for design ideas :

<https://www.klipfolio.com/resources/dashboard>

<https://www.geckoboard.com/dashboard-examples/>

<https://www.tableau.com/learn/articles/business-intelligence-dashboards-examples>

1. Refer to **Tableau Best Practice** and **Lesson 09** on Tips on dashboard design:

<https://help.tableau.com/current/pro/desktop/en-us/dashboards_best_practices.htm>

1. Create your dashboard.
   1. Screen capture the dashboard including your username and insert below and provide the insights and how the insights help to answer the business questions

|  |
| --- |
| **Chart - Dashboard** |
|  |
| **Insights** |
| The most popular flat type that has been sold for the past few years is the **4 ROOM Flats.** This can be seen in the “Popular Flat Type” Chart. This means that people in Singapore most likely purchases a 4 Room Flat as compared to the other flat types. The top 5 estates that sold the most number of 4 room flats is Woodlands , Sengkang, Yishun, Jurong West and Tampines. Since 4 room flat is the most popular flat type that Singaporeans purchases , the probability of finding and buying 4 room flats in those estates is high.  When buying a house , customer would try to find a house that is worth its resale price. Customers who might be cost saving do not want to purchase a flat that cost more than the flat’s worth. Since the popular estates to purchase a 4 room flat has been identified (found in “Top 5 Estates that sold the most number of 4 Room Flat” Chart) , it wise to check those estates first. In the chart “Number of 4 Room Resale Flats With Same Resale & Valuation Price (By Estate - Top 10)” , it displays the top 10 flats where the resale price is exactly equivalent to its valuation price. In that chart , the popular estates identified in Top 5 Estates that sold the most number of 4 Room Flat” Chart is present and is also the top 5 estates where the resale price is exactly equivalent to its valuation price. This makes the decision to where one should buy their 4 room flat strong.  Another important factor when buying a house is the remaining lease duration (RLD), as this factor determines how long one gets to stay in the flat before the lease ends. The top 10 estates that have a high remaining lease duration is displayed in the chart named “Average Remaining Lease for 4 Room Flat (By Estate - Top 10)”. In the chart , it is shown that Punggol has the highest RLD. However , if we compare with the estates that sold the most 4 room flat over the years , Sengkang has a high RLD. Estates such as Yishun and Tampines did not make the top 10. Hence, the best estates that has resale price closer to its valuation price and has a relatively high RLD is Sengkang , Jurong West and Woodlands.  In order to purchase a flat, one needs to know how much a flat cost in average. Since the most popular flat type is a 4 Room flat , the minimum , maximum and average cost of a 4 room flat in top 5 estates where 4 Room flats are sold has been identified and displayed in the “Min, Avg , Max Resale Price of 4 Room Flats (By Popular Estates)” Chart. From the chart , the highest resale price that a 4 room flat can be is $650,000 at SengKang while the lowest resale price that a 4 room flat can be is $225,000 at Woodlands. The average cost of a 4 room flat in the popular estates (that known for selling 4 room flats) cost around $366,954 - $445,695.  In my experience, some people would like to purchase a new flat to due lack of space in their previous flat. Some customers would prefer a slightly bigger living area rather than buying an extra room which can be very costly. ( E.g Buying a bigger 4 Room Flat instead of a 5 Room flat is a cost saving option). Hence , in the Average Floor Area for 4 Room Flat( By Estate) Chart. All the estate’s floor area is displayed. Different estates have different floor areas for a flat. The biggest 4 room flat is sold at Pasir Ris while the smallest 4 room flat sold is Marine Parade. In the Top 5 estates, where the probability of purchasing a 4 room flat is high, Tampines is the biggest while Sengkang is the smallest. This helps customers who are finding for a slightly bigger 4 room flat for different purposes.  **In conclusion , the best option for prospective homeowners who would like to buy a 4-room flat is at Woodlands. This is because, the probability of finding a 4-room flat is the highest, most of the 4 room flats are resold at the same price as its valuation price , the average remaining lease duration is 79.78 out of 99 years which is still relatively high , the cost of the flats is cheapest, and the floor area aligns closely with the higher average.** |

* 1. What are the **visualization principles** you adopt on your dashboard charts ? Give at least ONE examples of each visualization principles covered in P08 and P09.

|  |
| --- |
| **Chart (screen capture)** |
|  |
| **Visualization Principle(s) applied** |
| **Continuity**: The bar charts is used in the dashboard , follow the law of continuity. This is because the chart is ordered in descending by cash over valuation or average remaining Lease.  **Similarity**: A multi-colored graph is harder to read which can cause cognitive overload. Since coloring the bars individually in different colors do not have any special meaning. I have only used 2 colors. The dark blue to group the estates with high probability of selling 4 room flats and the light blue to group the others.  **Data ink ratio**: I do not have any redundant data ink and have only used in necessary bars. For example , I have used ink to group by data to help in group estates.  **Chart Junk**: There is no chart junk used in my graphs. For example , I do not have any irrelevant graphics such as images of flats on my charts or using 3D bars.  **Graphical Excellence**: All my charts starts from 0 in the y and x axis.    **Data Density**: The portion of the total size of my graph is fully covered with the amount of data I have. |

# Reflections

Complete the reflections on the 2 questions below:

1. What are the challenges or issues faced in the assignment?

**The challenges I have faced with my assignment would be cleansing the data , I often find myself cleansing the data many times as there would be errors such as the datatype or sometimes I would have missing values. After organising what data needs to be cleaned , I was able to clean my data properly. Another problem I faced was coming up with business questions that link together to create a story. After some guidance from my facilitator who explained how a story should flow and link. I was able to create business questions that link together to create a story.**

1. Which part of the assignment you have done well and which part could be better and how?

**I think the part that I have done well would be creating a story with insights and I was able to give my final conclusions. For example, after analyzing my charts, I was able to conclude that Woodlands was a great estate to buy a 4 ROOM Flat.**

--- End of Document ---