

School of InfoComm Technology

**Business Intelligence Module**

Diploma in IT/FI

Apr 2017

**TEAM/INDIVIDUAL ASSIGNMENT**

(40%)

**Deadline for Final Submission:**

**10 Aug 2017**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tutorial Group | : | P04 | | |
| Team Number | : | 2 | | |
| Tutor | : | Mr Mike Wee | | |
| Members | : | Case Study | Student No. | Student Name |
|  |  | 1 | S10171069 | Lim Seow Chong |
|  |  | 2 | S10170594 | Lean Jun Liang |
|  |  | 3 | S10173220 | Han Weixiang |
|  |  | 4 | S10170920 | Toh Leong Seng |
|  |  | 5 | S10170568 | Nicholas Fang |

**Penalty for late submission:**

10% of the marks will be deducted every day after the deadline.

**NO** submission will be accepted after 14 August 2017, 9:00 AM.

Table of Contents

[1. Introduction 5](#_Toc490146697)

[1.1 Background Information 5](#_Toc490146698)

[1.2 Business Question 6](#_Toc490146699)

[2. Individual Report (Seow Chong) 7](#_Toc490146700)

[2.1 Data Cleansing 7](#_Toc490146701)

[2.1.1 Sources 7](#_Toc490146702)

[2.1.2 Extraction 7](#_Toc490146703)

[2.1.3 Cleansing 8](#_Toc490146704)

[2.2 Analysis 8](#_Toc490146705)

[2.2.1 Data Description 8](#_Toc490146706)

[2.2.2 Graph Creation 9](#_Toc490146707)

[2.3 Dashboard Description 9](#_Toc490146708)

[2.3.1 Home-ownership 10](#_Toc490146709)

[2.3.2 House Pricing 10](#_Toc490146710)

[2.3.3 CPF Housing Scheme 11](#_Toc490146711)

[2.3.4 Analysis 11](#_Toc490146712)

[2.4 Reflection 11](#_Toc490146713)

[3. Individual Report (Nicholas) 12](#_Toc490146714)

[3.1 Data Gathering and Preparations (Before Presentation) 12](#_Toc490146715)

[3.2 Data Gathering and Preparations (After Presentation) 13](#_Toc490146716)

[3.3 Analysis and/or Model Building 14](#_Toc490146717)

[3.4 Final Tableau Story/Dashboard Description 16](#_Toc490146718)

[3.4.1 Observation 17](#_Toc490146719)

[3.4.2 Report Answer 17](#_Toc490146720)

[3.5 Reflection 18](#_Toc490146721)

[4. Individual Report (Jun Liang) 19](#_Toc490146722)

[4.1 Data Gathering 19](#_Toc490146723)

[4.1.1 Data Preparation 20](#_Toc490146724)

[4.2 Analysis 21](#_Toc490146725)

[4.2.1 Singapore Resident by Type of Dwellings 21](#_Toc490146726)

[4.2.2 Total Population 21](#_Toc490146727)

[4.2.3 Population Ethnic 21](#_Toc490146728)

[4.2.4 HDB Sold and Rent 21](#_Toc490146729)

[4.2.5 Data Limitation 22](#_Toc490146730)

[4.3 Graphs 22](#_Toc490146731)

[4.3.1 Trends of Resident Forecast 22](#_Toc490146732)

[4.3.2 Residents by planning area and sub zones 22](#_Toc490146733)

[4.3.3 Average resident by unit type 23](#_Toc490146734)

[4.3.4 Average units sold in the past years 23](#_Toc490146735)

[4.3.5 Ethnic Distribution 23](#_Toc490146736)

[4.4 Dashboard Description 23](#_Toc490146737)

[4.4.1 Advance and non-advance Dashboard 24](#_Toc490146738)

[4.4.2 Observation 24](#_Toc490146739)

[4.5 Reflection 24](#_Toc490146740)

[5. Individual Report (Leong Seng) 26](#_Toc490146741)

[5.1 Data Cleansing 26](#_Toc490146742)

[5.1.1 Sources 26](#_Toc490146743)

[5.1.2 Extraction 26](#_Toc490146744)

[5.2 Analysis 29](#_Toc490146745)

[5.2.1 Data Description 29](#_Toc490146746)

[5.3 Graph Creation 31](#_Toc490146747)

[5.4 Dashboard Description 33](#_Toc490146748)

[5.4.1 Resale Market Environment 33](#_Toc490146749)

[5.4.2 General information for resales market 34](#_Toc490146750)

[5.4.3 Resale Information Map 34](#_Toc490146751)

[5.4.4 Analysis 34](#_Toc490146752)

[5.5 Reflection 35](#_Toc490146753)

[6. Individual Report (Weixiang) 36](#_Toc490146754)

[6.1 Data Gathering and Preparation 36](#_Toc490146755)

[6.1.1 Background Information 36](#_Toc490146756)

[6.1.2 Sources 36](#_Toc490146757)

[6.1.3 Extracting and Cleansing 36](#_Toc490146758)

[6.2 Analysis and/or Model Building 37](#_Toc490146759)

[6.2.1 Data Description 37](#_Toc490146760)

[6.2.2 Creating of graph 37](#_Toc490146761)

[6.3 Final Tableau Story/Dashboard Description 39](#_Toc490146762)

[6.4 Reflection 40](#_Toc490146763)

[7. Report Answer 41](#_Toc490146764)

[7.1 Findings 41](#_Toc490146765)

[7.2 Solution 42](#_Toc490146766)

[8. References 44](#_Toc490146767)

# **1. Introduction**

# **1.1 Background Information**

Singapore is a small country with a land mass of 719.1 kilometre square, as compared to neighbouring countries Malaysia of 330,830 kilometre square and Indonesia of 1.905 million kilometre square. Despite its small size, Singapore is home to 5.607 million people and have an astonishing population concentration of 7797 people per kilometre square, ranking 3rd in the world behind Macau and Monaco. As Singapore is small country, with her limited land size, it is not surprising that she faces problems in ensuring that the population of hers will be able to find a place to live in. In order to solve this problem, HDB Flats(Housing Development Board Flats) were built. It is an incredible feat which is the result of well planned land allocation and the common residential structure - the flats which allowed the government of Singapore to fit more people in a smaller land area by building upwards. Pinnacle@Duxon is the tallest HDB flat in Singapore, with 50 floors at a height of 156 metres tall.

Singapore government has very different plans when it comes to housing. Singapore government hoped that housing can become a tangible asset for all Singaporeans and also provide a sense of belonging. In order to move towards this target, the government launched a scheme in 1964, also known as the Home Ownership Scheme (HOS). The purpose of this scheme is to bridge the gap between affordable, sustainable housing and home-ownership. This scheme allows the low income singapore citizens to buy the units of the flats at affordable prices from the government on a 99-year lease basis. Later in 1968, Singapore government allowed citizens to pay using their Central Provident Fund (CPF). By doing so, more Singapore citizens are able to own their own home which helps foster a sense of rootedness among Singapore citizens.Today, 82% of the population is housed in HDB estates, and the estates (comprising flats and a comprehensive range of facilities) have become a common point of emotional reference for the vast majority of Singaporeans. This "HDB Experience" has played an important role in bonding Singaporean by having a common experiences among Singaporeans of all races and from all walks of life.This idea was supported by Mr Lee Kuan Yew as he believe that by having high homeownership rate, it will lead to social and political stability.

This scheme gained worldwide recognition and attention and received the United Nations (UN) Public Service Award. But this does not means that it is the end, therefore there is a need to continue to improve the scheme so that they can increase and sustain the homeownership rate and ensure social and political stability in singapore.

As years passed, the HOS have undergone several changes and the population have become more affluent and the expectations for standard of public housing grew. The quality of the HDB Flats has thus increased and so has the homeownership rate of singapore. As more singapore residents start owning their own apartment in HDB Flats, this has become an important asset for their retirement fund. Realising this, the government launched several upgrading scheme for the HDB Flats in order to increase the wealth of households through property price appreciation. The factor that will affect this includes the location of the flats, the type, etc. Making use of this, government can increase the value of the HDB being build and using the HOS, able to sell it cheaper than market price to the singaporeans and allow them to have a it to rely on when they are old.

# **1.2 Business Question**

Since HDB has played a large importance in the lives of many Singapore and Singapore’s history, it is important to continue improving the services and products that HDB provides to their customers. Our business question shall be “How to improve the services and products that the HDB provides to their customer?”.  After analysing the question, we have come out with sub-questions that will help to answer the main business question. The sub-questions are: Is there a need to improve the current home-ownership scheme?, will the ratio of each races that purchase HDB affect the sales?, Is the resale price for HDB affordable?, what components will satisfy customer?, does location affect prices of HDB flats?. By answering these question, the main business question will be answered indirectly. The business questions are targeted mainly for HDB. Housing Development Board(HDB) is a government body that is in-charge of planning and developing Singapore’s housing estate which includes recreational properties such as food centre and shopping malls. One of the main reason the decision on this business question is because of the mission of HDB. Its three main mission is to provide affordable housing, creating sustainable and vibrant towns and promoting building of active and cohesive communities.

# **2. Individual Report (Seow Chong)**

## **2.1 Data Cleansing**

### **2.1.1 Sources**

The dataset used is the combination of data from numerous sources, namely data.gov.sg, singstat.gov and HDB.gov.sg. These sources are credible as these are government data-malls whereby all government organization can upload data. There are numerous types of data that is extracted for example, home-ownership rates, flat resale prices and Singapore GDP. The excel sheets were uploaded by related government sector such as HDB and MAS.

### **2.1.2 Extraction**

Before looking for data, I read the business questions of whether is there any need to make changes to the existing home-ownership schemes. I identify all the information required for authorities to make the decision. The identified information are home-ownership rates, number of units sold and rent, transaction of flat units and CPF home-ownership scheme. There was a data sheet about the number of units sold and rent from Teoalida.com, however, due to the information being from 1989/99, it was not useful as it is rather outdated. I tried to look for more recent data but the more recent ones require to be paid for. After continue searching, I found that the recent information from 2011 to 2016 was available on HDB annual report. The details of flats were extracted from the HDB annual reports. The annual reports was not in PDF and could not be converted into excel sheet therefore, manual extraction has to be done. The data has to be copied manually and it was not only 1 report by a total of 6 reports so that data collected can be visualized so that trends can be found. Due to the manual extraction, it is inevitable for entering errors. The errors have to be checked by using a SUM formula to check the total of the entire column corresponds with that shown in the report. If the values does not correspond, the error has to be located through eyeball-scanning. There was a total of 16 columns that is needed to be extracted so it was a tedious process as continuous checking has to be done and it took hours to complete. For the transaction of flat units, it was extracted from data.gov. Transaction of flat units comes in 2 excel sheet, one from 2000 to 2011 and another from 2011 to 2017. There was actually a excel sheet of flats transaction on Teoalida.com, however, the data sheet was locked and requires password to be edited thus it was not used. For the annual GDP of Singapore, it was extracted from singstat.com. The site provided excel sheet for the information, however, due to large amount of data which were not required, I decided to manually extract the data from the table constructed by singstat.

### **2.1.3 Cleansing**

Cleansing was done for the transactions of flats in different estates. The data of transactions was downloaded from data.gov.sg in an excel format. The data has details such as the month, town, street-name, floor-range, flat-type, floor area and resale price. After looking at the data, I decide which are the columns that are required and came to a conclusion that the month, town, flat-type and resale price is that only details that are useful. Therefore, I hide the columns not needed to ensure that I would not be distracted by the other details. Since the month is in a year-month format and I only wanted the year, I had to create another column named year and use LEFT(B:B,4) to extract only the year. Afterwards, I created a column named mean resale price to contain the average resale price for the different room types, towns and the years. Filter was placed on every column and I tried to filter by year then by town and lastly by flat-type to retrieve the transactions. When using the AVERAGE formula on the filtered resale price column to find the mean resale price, the value appeared, however, when filtered for the next range of results, the value for the previous mean changes. Research on how to convert formula into value was done but no solution was found, therefore I created a new sheet to hold the information of year, town, flat-type and mean resale price. I filter the original sheet to obtain the resulting resale price after that, I copy the resulting rows then paste it in the new excel sheet. I then run an AVERAGE formula to obtain the average and then type it into the cell which has the corresponding year, town and room type. It is time consuming and took around a day to complete. Due to manual typing and filtering, I encountered an error as one of the filter was not changed therefore all the filtered value was not correct. This resulted in a restart for that year. The rest of the data are either from the singstat site which are in a table thus manual entering is done.

## **2.2 Analysis**

### **2.2.1 Data Description**

After cleansing, there are a total of 6 excel sheets containing information of the the number of units sold & rent at each town, home-ownership rates, home-ownership rate details, resales price of flat type by town, Singapore GDP 2011 - 2016 and the CPF Housing Scheme results. The units sold & rent at each town is able to provide the details of total number of sold as compared to total units rent over time. Home-ownership rates and the details are able to provide information about whether how many Singapore citizens are able to own their own flats. Resale price of flat type by town is able to give an in-depth information about the prices of the different types of flat at different places. Singapore GDP is displayed in millions and the data is able to provide information of the economy of Singapore which influences the prices of housing. Lastly, the CPF Housing Scheme is a housing scheme established by HDB to allow Singaporeans to use their CPF to pay for part of their housing.

### **2.2.2 Graph Creation**

Before creation of the graphs, I take at the look at the business question again and start to draft out the information needed to answer the question. I visualized the graphs needed and its objectives before deciding which type of graph to use for example, to show to trends of home-ownership rates I decide to use line graph to show the trend with home ownership rates as y-axis and year as y-axis. As for comparison of prices of housing over the years, I decided to use geological map. Due to the availability of data of the price by town, I wanted to display the data using a map so I tried to change the town into geological town and then set the chart to map. However, Tableau could not identify the name of town to those in Singapore thus, I tried to find the coordinates of the different towns and tried to explore custom geocoding. After coming out with the chart, the dots were small and was not what I wanted. After doing some research and enquiring more from the lecturer, I understood that Tableau required a country to identify the town since Singapore is so small. I created a new excel sheet with the town name and country name. I created a new tableau sheet to test out the filled map and the map of towns in Singapore could be shaded. After that, I tried to link the excel sheet with Singapore details and the transaction, I created a filled map which was shaded mostly, however, due to a missing town name Central Area in the filled map, the central area portion was no shaded. I researched and found out that Central Area consist of Newton, Orchard, Outram and River Valley. I tried to group the four regions, however, I did not know how to link the group to the name Central Area. After much research, I still could not find out how to link the group. Due to time constraints, I decided to use the second-best method of using bar graph to allow clear comparison of prices of units at different towns. I placed the price to the y-axis and town to the x-axis. While doing other graphs, I faced problems of which the measure and dimension were not in the correct format thus, resulting in not being able to create the graphs wanted. Therefore, during the creation of the graphs, there were many occurrences of going back to the dataset to cleanse the data into the form that is needed. For the breakdown of home-ownership, I decided to use the pie chart to effectively show the difference in number of people who owned their unit as compared to tenants. There is also a filter added for user to filter the data by year. Dual-bar graph was used to show the data of CPF Scheme of both residential property scheme and public housing scheme. A filter was created to allow user to change the y-axis from number of member to net amount. During the creation of graphs, a total of 7 graphs of 4 types of graph, namely the line chart, bar chart, dual-bar chart and pie chart, was created.

## **2.3 Dashboard Description**

For the final dashboards, there are a total of 3 dashboards created. Initially, while researching I stumbled across a video called “Conditional Formatting”. The video taught how to create parameters and calculated fields so as to create a single dashboard to have 3 different interfaces which I found was rather cool. After playing around with my dashboard, I found out that in order to have parameters and calculated controls that controls multiple sheets, all the graphs must be using the same data source. Due to the different data sources, I was unable to create the parameters and calculated controls therefore, I ended up with 3 dashboards. The dashboards are categorized by Home-Ownership, House Pricing and CPF Scheme. The dashboards are aimed to aid government make decisions on the current CPF housing scheme and other housing schemes. I established that Home-Ownership is the main purpose for the housing schemes that the government thus having a Home-Ownership dashboard is important to portray the current prospect of home-ownership. A factor that contributes to whether people purchase houses is the pricing. Sales of houses fluctuate according to the prices of housing thus it is important to monitor the prices of houses and so there should be a dashboard showing the pricing of houses. The main purpose of the dashboards is to help user make decision of the CPF housing scheme, therefore, the last dashboard is CPF Scheme.

### **2.3.1 Home-ownership**

Home-Ownership dashboard has a total of 3 graphs, home-ownership rates from 2011 to 2016, home-ownership drilldown which shows the number of tenants and owner-occupied units and the number of units sold/rented over the years. There is a filter for the user to choose which year to drilldown on. By looking at Home-Ownership dashboard, user is able to look at the different aspect of home-ownership including the yearly unit owned and rented count and also the number of units rented and sold yearly. This will allow the user decided whether is there a need to change the scheme if there is a significant drop in home-ownership rate, continuous drop in number of units sold or an significant increase in rented units. Currently, the home-ownership rate is still increasing.

### **2.3.2 House Pricing**

House-ownership has a total of 3 graphs, average price by type by year, Singapore GDP and average price by type by town. After doing some research, I came across a news report stating that Singapore GDP have an influence on the pricing of the units thus I placed the chart in this dashboard. This dashboard has 1 filter which allows the user to select which type of unit to view. By selecting the type of unit, the graph of the average price by year will be changed and the title will state which room type is the graph showing. The average price by town and its title will also change to the selected room type. This dashboard also allow user to do a drill down of the prices for different years. So by clicking on the different bars, the prices by town will change to show the prices of room type by town of the selected year. This dashboard will enable user to compare the prices of unit by year with the GDP so as to effectively step in to come out with a new policy to aid home-ownership. This is evident in 2013, the prices of housing have been on the increase from 2011 to 2013 in relation to the increasing GDP from 2011 to 2013. In order to stop the price from going up, the government have come out with numerous policies to reduce the pricing which can be observed from the graph. This evidently show the dashboard is able to show the result of the new policies based on the pricing of housing in relation to the GDP. The housing prices is still increasing, however, it is still lower as compared to 2013.This is expected as the GDP of Singapore has been increasing.

### **2.3.3 CPF Housing Scheme**

CPF Housing Scheme has only 1 graph comparing the residential property and public housing over the years. There is a filter which allows the user to select what is to be compared; whether is it the amount withdrawn or the number of participants. By looking at this dashboard, the user is able to understand the results of the schemes, whether are they beneficial or is there any changes required to be made. For the CPF Housing Scheme, the net amount withdrawn has been on the increase which shows that the people are benefiting more from the scheme as they are able to withdraw more to pay for housing. The number of participate dropped from 2011 to 2016.

### **2.3.4 Analysis**

In conclusion, home-ownership rate is still on a steady climb since 2011 from 88.6 to 90.9 in 2016. The increase number of units sold as compare to number of units rent is more significant thus corresponding with the Home-Ownership rates. From 2011 to 2016, Singapore’s GDP was on a climb from 342 billion to 402 billion. Prices of houses increased during the period of 2011 to 2013 for all room-types. However, it dropped from 2013 to 2014 and currently, the prices of housing is having a slower climb from 2014 to 2016. This correspond with the GDP rise over the years. As for the CPF Housing Scheme, the total amount of CPF withdrawal has increased from 2011 to 2016 from 5.464 billion to 6.972 billion however the number of people using CPF to withdraw has reduced from 750000 to 740000.

## **2.4 Reflection**

Through this project, I managed to understand many things taught in class. Initially, ETL to me was supposed to be a relatively easy process as the data sheet should be able to be downloaded in excel format and translate just means running excel formula to cleanse the data. However, after this project, I understood that data comes in many format and cleansing does not only include running formulas but also removing unwanted columns and creating new columns to extract required information. This project had required me find a data source, create a business question, carry out ETL, come out with charts and graph and lastly create dashboard to answer the business question. I also experience finding bad dataset which resulted in the inability to create good/many charts for analysis, thus, a good dataset is crucial. As for the creation of graphs, it is important to know the purpose for all the graphs created so that the correct measure, dimension and graph type. Overall, this project have given me a better perspective of the creation of dashboard and an appreciation for the data set given during practical test / tutorials.

# **3. Individual Report (Nicholas)**

## **3.1 Data Gathering and Preparations (Before Presentation)**

The dataset used is a combination of multiple data sources on HDB, taken from government websites, thus ensuring the reliability of the data gathered. Websites include data.gov.sg and hdb.gov.sg. Information such as number of sold and rented units, prices of units, based on location and flat types are collected.

I started out with only data from data.gov.sg, finding what I thought was enough data to help me answer my business questions. What I gathered were the cumulative units completed since 1960, number of sold and rented properties and price range of HDB flats offered. Each of these datasets have the same primary key, Financial Year, given they were from the same website. I then tried multiple ways to combine the datasets into one sheet of data, but failed to do so, thus I resorted to copying the two other sheets into one excel file. From there, I saw multiple na values in all three sheets of data, and thinking that it would affect my graphs, I used the Find & Replace Function to replace all na values with 0.

Seeing as my data was pretty much cleansed for me already, I started to design my graphs on tableau. Once I started designing the graphs, I found problems with the data I had. Firstly, I had many missing data, such as missing towns for certain years. Secondly, the na values which I had replaced with 0 caused my line graph to show weird trend lines. Lastly, the data provided were only for a few areas in Singapore. With this much missing information, I will not be able to design an informative dashboard.

Thus, I went to look for more data. In my struggles to find more data, I found out that my team member, Seow Chong had data that I could use. Even though I did not help in the gathering of the data, he generously gave the dataset to me for me to use. He gave me the data for HDB Annual Report and Resale Prices of Units. From there, I had to further cleanse the data to fit my business questions. I started removing columns that I do not need, such as land areas, population size, targeted dwelling units. I also removed all data for the year 2017, as the data was only up till the second quarter of the year and I am planning on using data which was collected for an entire year.

I also had to manually calculate and input the data for total number of units sold and total number of units rented for each year in order to find out if the total units sold and rented increase or decrease over the years. I decided to put this set of data in a new sheet in order not to mess up the rest of my data. I calculated the total units sold for each year by using the SUM function in Excel and selecting all the cells in the relevant columns such as 1-room (sold) to 5-room (sold) and Executive (sold) for each of the years. The total number of units rented per year was calculated the same way.

## **3.2 Data Gathering and Preparations (After Presentation)**

During the presentation of the dashboard to the tutor, I felt that I could have made my dashboard a lot better because the dashboard then did was not very informative in telling the user why the location could have affected the price of HDB flats. The difference in price could be due to the number of facilities that the people can use in the area, in my case, I chose the number of Mass Rapid Transit(MRT) Stations in each region.

The reason for choosing only the number of MRT Stations in each town was because I could not find data for the number of facilities in each town. I had searched on data.gov.sg but all they had were the cumulative recreational facilities in Singapore, which was not separated by the different town. I tried searching “facilities in <town name>”, “recreational facilities in <town name>” as well as “shopping malls in <town name>” on Google Maps and then manually counting the results shown but I found out that there were many inaccuracies. Firstly, the number of results of each search kept changing whenever the page refresh. For example, when I first searched “facilities in ang mo kio”, I was given a total of 73 results, but when I refreshed the page, it became 84 results. Secondly, the results shown does not actually shows the actual number of facilities in the area. When I changed from searching “facilities in ang mo kio” to “shopping malls in ang mo kio”, I was given 3 results, but on the Ang Mo Kio Wikipedia page, it was stated that there are four shopping malls in Ang Mo Kio. Thus, I decided to stop searching for this set of data as I did not want my dashboard to show inaccurate data.

I started searching for the number of MRT stations in each town. Similarly, such data could not be found on data.gov.sg or on any other government website. Thus, I started searching through the Wikipedia page of each town, and manually counting the number of train stations that were stated to be in each town. However, certain towns such as Central Area and Queenstown did not specify any MRT stations. Once again, I had to find another source.

I managed to find the List of all MRT Stations in Singapore on Wikipedia and found out that each MRT Station’s Wiki page states the area that they are in. Thus, I had to visit each wiki page for every MRT Station on the list and count the number of stations in each area. During the process, I had to skip MRT Stations that were located in an area that I do not have in my previous dataset as well as MRT stations that were still under construction.

After gathering the required data, I entered in to a new data sheet with two columns, Town and No. of MRT Stations.

Additionally, I had to make some edits to my previous 3 data sheets. Firstly, I had to change the column header for similar data to the same name as well as the data name itself, such as Financial Year to Year, Flat Type to Flat-Type and 1 ROOM to 1 Room(same for the other flat-types) instead.

Previously, the HDB Annual Report data sheet had the various flat types, rent and sold, split up into different columns. That would make it hard for me to design a dashboard as the flat-types will be separated in to different Measures, thus, I decided to transfer the data over to a new sheet in a new format. I changed the columns from Year, Town, 1-Room (sold), 2-Room (sold), 3-Room (sold), 4-Room (sold), 5-Room (sold), Executive (sold), 1-Room (rent), 2-Room (rent), 3-Room (rent), 4-Room (rent), 5-Room (rent), Executive (rent) to just Year, Town, Flat-Types, No. Sold and No. Rent. By making these changes, the number of columns decreased but the number of rows increased. I used the Transpose function in Excel to convert the data in each row into a vertical column and then proceed to paste it into the new sheet. I also had to change the name of the flat-types by removing the dash in the middle so that it will be the same as the other data sheets.

I had also decided to remove unneeded data sheets such as the manually calculated total units rented and sold as well as the unedited HDB Annual Report.

With this I can start designing my final dashboard.

## **3.3 Analysis and/or Model Building**

In my revamped data sheet for the HDB Annual Report, it shows me the number of the different flat-types units sold and rented in the various towns from 2011 to 2016. This will be able to tell me if there are any increase in the number of units sold or rented. This will tell me if people are buying or renting HDB flats over the years as well as which location is more popular among the people.

In my second data sheet, I have the Resale prices of the different flat-types by town from transactions that were made from 2011 to 2016. This can tell me how the average prices of each flat-types vary according to the location as well as if the average prices of each flat-types in each area increases over the years.

In the last data sheet, I have the number of MRT stations in each of the towns, which can show me if this is a reason for the difference in prices of HDB flats in the various location. As MRT Stations are one of Singapore’s main public transport systems, an abundance of MRT stations in an area could cause the average prices of HDB flats in that area to increase.

By comparing the second and last data sheet, I can find out which location have the most MRT Stations and if that same location has the highest average price of HDB flats compared to the other locations. Additionally, by looking at the first data sheet, I can check if the high prices in that area affects the number of HDB flats sold.

In order to answer my business question, I had to incorporate the Town column into each of my graphs as my question is mainly focused on the location. To show even more details, I can filter the graphs by Flat-Types to show how the prices of each flat-type varies across the different locations of Singapore.

The first graph I had created is a tree map graph to help visualise the number of MRT Stations in each town. I started out with a simple bar graph but in order to make a more unique graph that can show the same information as a bar graph, I decided to switch to a tree map graph. The Text represents the name of the town and the size and colour represents the number of MRT Stations in that town. The bigger and darker the box, the more number of MRT Stations there are in that area.

The second graph I had created is a horizontal bar graph with one column and two rows, Average Resale Price, Town and Flat-Type respectively. This was created to show the difference between prices of the flat-types in the different towns. I started out with only Resale Prices in the column and Town in rows, using Flat-Types as a filter, but decided to add flat-types into rows as well for better comparison between the flat-types in each area. When designing this graph, I was contemplating between putting which Dimension into the Rows first, Town or Flat-Types. By placing Flat-Types first, it will be easier to compare the difference in prices between locations based on flat-types however, I decided to place Town first instead. This is because with the flat-types filter I created, I can recreate the same thing the can be done as putting flat-types first.

To see a trend in the prices of the HDB flats in the different towns from 2011 to 2016, I created a line graph with Year as the column, Average Price as the row and Towns as the colour. I decided to use a trend line to show this set of data as the user will be able to visualise how the price changes with greater ease. Currently in the sheet, the graph does not have any filters for the different flat-types, thus the average prices of HDB flats in each area is across all the flat-types. Previously, I did not add the Towns measure into colours but instead added Flat-types. However, I realised that that does not help me answer my business question which is based on location and not flat-types, thus I changed it to Towns.

Lastly, to compare the number of sold and rented flats based on town from 2011 to 2016, I created a side-by-side bar graph with Town and Year as the Columns and Number of units sold/rented as the rows. The graph allows me to compare between the different areas and spot those places with significantly more sold flats or rented flats than the other areas. Instead of removing the gridlines like in the practical, I decided to keep it this time to help the user compare between the different bar easily, as some of the changes are insignificant and can be difficult to spot, especially on the dashboard. I did not use any filters yet as I will be using the second graph created to act as the filter for all the other graphs. In addition to creating the graphs, I gave each of them dynamic titles which will change based on the filters selected. Based on all the graphs created, predictions can be made. By looking at the tree map graph, the locations with the higher price range can be predicted as the Towns with a larger and darker box means that there are more MRT Stations in that area, thus giving it a reason for the prices of HDB flats in that area to be higher. Thus, the number of sold flats in that area may be expected to be lower.

By comparing the other three graphs, the rate of flats being sold may be predictable because people tend to avoid high prices, thus when the prices shown on the trend line graph increases, we can expect the number of sold flats to decrease in that year. We can also expect the areas with more pricey units to have lower number of sold flats and possibly more rented flats. On the side-by-side bar graph, if the number of rented flats were to decrease, while the number of sold flats increase, it could mean that people are switching from renting to buying. However, the number of sold flats may also increase as prices of HDB flats increase due to other reasons. People may find the high prices “worth-it” as they will be paying for the number of facilities that they can use in the area compared to other places. Another reason could be the growing population. As population increase, more flats are needed to house this people, thus leading to more flats being sold.

## **3.4 Final Tableau Story/Dashboard Description**

Previously, before the presentation, I had 2 dashboards, with 2 graphs on each dashboard and multiple filters, that did not really show much information. Now, I have only one dashboard but with 4 different graphs and only one filter. At the top left, I had placed the tree map graph showing the Number of MRT Stations by Town. At the top right, it is a line graph that shows the trend of the Average Prices of Units in all Town by Year. At the bottom left, the horizontal bar chart showing the Average Prices of all Units by Town, equipped with a Flat-Type filter, is displayed. At the bottom right, it is the side-by-side bar graph that allows the user to compare the Total No. of All Units in all Towns that are Sold or Rented by Year. In order to make the dashboard easy on the eyes, I avoided 3D graphs and Pie charts to display my data. I also avoided using bright colours in my bar graphs that can be distracting to the users. The line graph on the other hand, have multiple lines with different colours, however, this can be solved when the filter on the dashboard is applied. I have also used dynamic titles in my graphs that changes according to the filters applied. Mentioned previously, I decided to use my horizontal bar graph as a filter for the entire dashboard. Additionally, I set the tree map graph as a filter as well. When an area is clicked on the tree map graph, all the graphs will change accordingly to the area selected. The horizontal bar graph will show the average prices of a HDB unit in the selected area, the line graph will show the trend of the average price of all flat-types in the selected area from 2011 to 2016 and the side-by-side bar chart will show the total units sold and rented in the selected area from 2011 to 2016. As for the horizontal bar chart, by clicking on each Town, it will be filtered the same way the tree map did. However, the horizontal bar chart can be filtered another way, by Flat-Types. The user can click on the flat types of a certain area instead to display its relevant information. This graph also comes with a separate Flat-Type filter that allows the user to select which flat-type to display for all towns instead of manually clicking on each one.

### **3.4.1 Observation**

I have observed that most of the Towns with a relatively higher number of MRT Stations in them have HDB flats that are of higher price, regardless of flat-types. Areas such as Central Area, Bukit Timah and Queenstown have much higher pricing for HDB flats. However, I had also noticed that some areas with only one MRT Station have HDB flats that are higher in price compared to the areas with more stations. For example, even though Central Area have 27 MRT Stations, and Bishan only have 1, most of the flat-types (except Executive) in Bishan are of higher price compared to Central Area.

I also observed that as prices of HDB flats change over the years, the number of sold and rented flats also changes. When prices decrease, the number of units sold will increase. For example, in Yishun, as average prices of a unit decrease from $403,538 to $361,865 in 2013 to 2015, the number of sold units increased drastically from 48,263 to 54,751 in 2013 to 2015. There were abnormalities in some of the areas such as in Yishun, where in 2011 to 2013, average prices of a unit increased from $358,377 to $403,538, the number of sold flats increased from 47,166 to 48,263. Usually, when prices increases, sales will decrease, however in this case, sales increased.

### **3.4.2 Report Answer**

Based on my findings, I can conclude that locations do in fact affect the prices of HDB flats in the area. However, as to how it affects, the reasons can be very broad and not just be due to the number of facilities in the area. Reasons such as nearby schools or recreational facilities can be the cause as well. Reputation of the area can be another reason because an area with a bad reputation may have low flat prices in order to attract customers. Even if the number of MRT Stations in the area is not the only reason for the difference in prices, it is definitely one of the more influential reasons. Another thing that I can conclude is that prices affect the sales of HDB flats, despite seeing number of sold units increased as prices increase, we can look at the data another way and see that people are buying 4 & 5 room flats more than executive flats which are usually priced higher. However, it is still a fact that a drop in price can lead to a drastic increase in sold flats. Furthermore, the Homeownership Scheme could be encouraging people to buy a flat instead of renting one, thus the increase in number of sold flats every year.

In conclusion, locations do affect prices of HDB flats and the price in turn can also affect sales of the flats. However, there are many other factors that contributes to this result as well.

## **3.5 Reflection**

From this project, I learnt that gathering the required data is one of the hardest parts of Data Analytics and then displaying the data in a perfect dashboard is the other hard part. Data gathering takes a lot of time and effort, and the required data may not be available sometimes, so I have to make do with what I can find. It is also important to know what data is needed before testing and experimenting with the data as well as familiarize with the software. This is because I started out not knowing what I needed to really answer my business question and just came up with a half-baked dashboard before the presentation. I was also not very experienced with Excel and Tableau, thus was unable to cleanse properly or create a relatively good dashboard.

I could see that I did not put in enough effort into this project and thus ended up with a sub-par dashboard and I hope that I am able to get better at doing similar projects in the future.

# **4. Individual Report (Jun Liang)**

HOS purpose is to ensure that the people of Singapore will be able to have a housing unit they can call their own and in order to ensure that this is possible to be achieved with the amount of people and its land limitations, HDB Flats are built as the solution. As more of the population owns their own units in the flats, the expectation of the standard of quality of HDB Flats also increased. In order to meet these expectation, there is a need to have some information which will guide them in building a HDB Flat that will be able to satisfy the customer. So, the question that I am answering is what components will improve the customer satisfactory on HDB Flat. This is important as the HDB Flats when it is built and few wants to rent it, it will be wasting the space that has been used to build the flat and this is not good as land space is limited for Singapore and this is not maximising the use of land area. Some information that will have to be known in order to build a HDB flat that people will be interested in is the location, the location of the flat is important as some people wants to their house to be near their workplace, wants to escape from the big crowd or traffic, etc. Another of information that will be good to know will be the type of the unit for the flat, whether it is 4-room or 3-room units. This is also important as this is also something that the customer will be keeping out for. Some other factors like the ethnic group of the people have to be taken into account when building HDB Flats. This is so as there is a certain rules and regulation that they have to adhere to thus, this calls for the need to have the information in order to build a HDB that sticks to the condition and also able to satisfy the customers. The type of users for this will be the HDB representatives, the ones that will decide on the details of the HDB Flats to be build. They will be able interact with the dashboard in order to get the information to build a HDB of good standard. The market size of this question will be people who are interested in finding a new HDB Flat to live in which generally could mean the population of Singapore.

## **4.1 Data Gathering**

In order to create a dashboard for the users, the dashboard must first have worksheets which is made from datasets. The dataset that I have used come from different sources. This is so as the first original data set that I have found is not enough to create enough worksheet for the dashboard. Therefore, I have to go and find more datasets, which caused me to have a dataset that come from different sources. I got the data mainly from data.gov.sg and singstat.gov.sg. A few data came from a report that is in pdf format which have to be copied. I got it from my friend who did the copying, this is so as I realised I need it only when I have started the creation of dashboard.

### **4.1.1 Data Preparation**

In order to prepare my data so that it can be used for analysis in tableau. During the preparation phase of the data, I have to filter some of the values that are not needed. There are some datasets which contains the value ‘na’ or ‘-‘ and these have to be changed to 0 or filtered out. And in order to do so is kind of tedious as I have to filter manually and change it.

The dataset that I have sourced is in a table format too. Which means that I have to convert it into row dataset so that tableau will be able to read it properly. In order for this to be done, I have to first understand the data, identify all of the attributes and measure and link it such that it will be like a star schema table but in rows form. In order to understand the data, sometimes I have to search to understand the relationship of the different attributes. In some cases, the names for the data does not make sense as it is in letters and underscores. In this case I have to read the metadata that I have downloaded along with the dataset and study it.

After understanding the data, I have to see if my understanding of it is correct and to do this, I have to do some calculation of some measures and see if it matches with the total that usually have been given in the dataset. Once all of this is done, then I will start preparing the data into row format. As data cleansing is an iterative process, which I only learn it when I started to do the worksheets for the tableau, I have to go back to the source and retrieve what I need and start the cleansing of that data part and merge it with the cleansed one that I have been working on. The process of changing the data from the table format to row form is a very long process especially if the dataset is big. There are many rows and columns that we have arrange properly before we can start the copy pasting and rearranging of the data.

Another big thing during the preparation of the data is the format of the data. As my source is from different sources, they use different term in order to describe the same thing. For example, data that shows the number of residents, one of my data shows the title of that column to be ‘Population’ and another as ‘Resident Count’. Things like this have to change to a constant name such that it won’t be as confusing and also allowing tableau to be able to detect the relationship easily. In some cases, the attribute for the different dataset is similar but some values from the datasets are different. In order to integrate them, I have to find out what is the missing values and what is it about then I have to come out with a solution to integrate them together such that they will be the same and there wouldn’t be inconsistency in the data. If there are inconsistencies, then when the worksheets have to display the data, it will not have the right value.

During the preparation, I have also realised that I have not gotten some data that I need in order to create the type of graph. This caused me to have to do more research and try to get datasets on what I need and in my case, I cannot find it as there is a lack of time therefore, I can only use what I have got and make the best out of it. For example, me and one of my teammate wants to create a map but then we realised that we will need the postal code or coordinates of the region which we can’t find easily. Therefore, we can’t create a map graph.

## **4.2 Analysis**

My data contains mainly about HDB and the residents. There are 4 datasets namely, Singapore Resident by Type of Dwellings, Total population, Population ethnic and HDB sold and rent.

### **4.2.1 Singapore Resident by Type of Dwellings**

This data has data on the number of residents by the town and their more specific parts. It also has the number of residents by the type of dwellings. Type of dwellings contains the type of HDB units and HUDC Flats that are non-privatised. The type of HDB Units is generally categorised by the amount of room the units have. The town attribute contains towns in Singapore like Yishun, Queenstown, etc. The subzones are the more specific parts of the town, for example, Yishun have subzones like Yishun Central, Yishun East, Yishun South, Yishun West, etc. All of these is recorded under the span of 17 years, from 2000 to 2016.

### **4.2.2 Total Population**

This dataset contains the data on population related data. There is the Total Population of Singapore, the Resident Population of Singapore, etc. The Total Population is the total number of people in Singapore whereas the resident population includes only Singaporean and Permanent Residents that are not overseas for a period not longer than 12 months. This data is also collected over the years from as early as 1960 till 2016.

### **4.2.3 Population Ethnic**

This dataset has the number of residents that are of the different ethnic group. The different ethnic group consist of Malaya, Indians, Chinese and others. It also contains data on the different gender of the ethnic group. This data is collected from 1960 onwards to 2016.

### **4.2.4 HDB Sold and Rent**

This dataset contains data that shows the Type of HDB units that have been rented or sold at 2011 to 2016. They can be further be categorised by the town that the flats are at. For example, there are 4947 Units of 1 to 2 room flats sold in Ang Mo Kio in year 2011.

### **4.2.5 Data Limitation**

The data that I have collected and filtered, have some limitations. Firstly, is that the value from the different sources may not match properly. This is so as when I tried to merge the data together, I checked if the data is gives the same result for the similar attributes, it returns values that are close to each other but not exactly the same. For example, the amount of Singapore residents total added from dataset ‘Singapore residents by type of dwellings’ and dataset ‘Total Population’ doesn’t show the same amount but both of them is about 3 million. This is so as when I did some filtering of data in the dataset ‘Singapore residents by type of dwellings’, I filtered out some values that I feel that it is insignificant and will not serve its purpose with the graph that I am going create with it. That is one of the reason why it doesn’t tally with each other. But then from the start, it is already different from the other sources that also have the resident population. Another reason is that some datasets have different categorisation of the towns and when the resident population is added up, there is also a difference between them.

Another limitation is that all of the towns may not be shown. This is so as I integrate them together and in order to do so, I have to remove some of the towns as the other dataset doesn’t have it. Instead that very dataset has a row of data on the attribute town called ‘others’. So, I changed some of it into others in order for the data to be able to be integrated together. This is a limitation as this will result in the data not being able to show all the town name properly and can only be called ‘others’.

## **4.3 Graphs**

I have created 5 graphs that allows the user to gain interact with.

### **4.3.1 Trends of Resident Forecast**

This graph shows the trend of the amount of resident that is forecasted by tableau. The aim of this graph is to show the estimated increase of residents in Singapore. This will allow them to know how many people should the flats to be built be able to house. This is done by using the forecast function in tableau and by using the previous data, forecast the increase of the residents. The forecast is done by years so that the user can use the estimated year to complete and compare with the estimated increase for that year.

### **4.3.2 Residents by planning area and sub zones**

This graph shows the trend of the residents in Singapore as a whole. It has the ability to drill down to more specific area to show the trend at that particular area. The area is called planning area, a term that is used by the government and it is used as the main urban planning and census divisions of Singapore. They are delineated by the Urban Redevelopment Authority. This is further able to drill down even more to the subzones which is the different areas in the planning area itself. The graph generally helps the user to decide where to build the HDB. As this shows the trend, they can use it and decide if there is a need to build more HDB in that area.

### **4.3.3 Average resident by unit type**

This will show the user the average residents that are housed by the type of flats. This will be filtered with the graph ‘Residents by planning area and subzones’. It allows the user to see the information at different granularity, just like the graph just mentioned. This will help them to make the decision to know how many to build by finding the average amount of people per unit by using this information in conjunction with the number of units they have in that area or subzone.

### **4.3.4 Average units sold in the past years**

This will show the average amount of HDB Units sold and rented for the 2011 onwards. This will show them in which area is the HDB more sold in. This information can allow them to see in which area have more HDB flats sold. From here they can see the most sold on easily as it is a heatmap chart which can help them in seeing the ‘famous’ area for the HDB flats and it will also allow them to know the number of units in average that the town have. Which can be used in conjunction with the chart of average Residents by unit type. To find out how many residents should one type of units contains.

### **4.3.5 Ethnic Distribution**

This graph shows the average ethnic distribution throughout Singapore, this will help the users to decide the ethnic ratio for the flat that they are building. This graph shows the percentage to aid them in deciding the ratio for the flats.

## **4.4 Dashboard Description**

The dashboard will include the 5 graphs mentioned above. There will be 4 filters in the dashboard. The first filter will be the planning area filter used for the average residents by type of dwellings, residents by planning area and subzones and average units sold for past years graph. The second filter will be for the subzones for the residents by the planning area and subzones and the residents by the type of dwellings. The last 2 filters will be used for the user to swap the type of view in the dashboard, graph filtering. The dashboard will allow the users to analyse the data presented to them in graphs and interacts with it such that they will be able to answer the question as to which components will increase the customer satisfactory of the HDB.

### **4.4.1 Advance and non-advance Dashboard**

I have placed the most important graph on the top left-hand corner as that will be the graph that should and will be seen first. I have also try to make sure that the dashboard would not look too squeezed by giving them the option to filter some graphs that they may not need at that point of time. For example, the user is supposed to look at the graph of forecasted trend on residents first in order to know about how many residents are they are going to build for. Then they will look at the residents by planning area and subzones to see the trend and decide where the HDB should be built. Then they will have to know the amount of HDB Flats to build which is the 2 graphs below, Average residents by Types of dwellings and Average units sold in past years. Then finally, after that, they will have to see the graphs on the ethnic distribution which is placed on the top right-hand corner, a place where we will usually check last. The dashboard is done in such a way that it is organized and allows the user to find what they need easily as it is not cluttered. I have used the action filter which allows some of the graphs to be filtered by the year. I have the graph filter which will allow them to choose the graphs that they want to view, this will give them a clearer view.

### **4.4.2 Observation**

I have observed that the most common type of units built is the 4-room Flat type followed by the 5-room Flat type. In average for the past 6 years, Jurong West seems to have the most sold HDB Flats for 5-room and 4-room but Ang Mo Kio have the Most for 3-Room type. But the residents in Jurong West don’t seems to be increasing. Whereas, Seng Kang residents is increasing steadily, this shows that Seng Kang can be a good area to build new HDB flats and 5 or 4 room flat would be good, seeing that there is a lot of it sold and that since it is a 4 or 5 room Flat type, it will be able to house more residents.

## **4.5 Reflection**

After doing this assignment, I have learned that the data that one has will define what type of graph can one come out with. Take the example of me and my friend who wanted to do a map graph but unable to do so as our data don’t have the data that allows it to do so. Another thing that I have learned is that the data preparation is an iterative process and that it isn’t easy. The quality of the data preparation will define how well the dashboard can provide the information. If the preparation is done well, and the data contains what we need we will be able to obtain a clearer answer for the question. The clarity of the question is also very important as only with a clear understanding of what we want to answer will we be able to come out with a good dashboard that will answer the question clearly. I have also learned that it is important that the dashboard is not as easy to make as there can’t be too much graph or it will look too cluttered but it can’t have too little graph as it may not be able to give the necessary information needed. Lastly, I have learned that the integration of the data requires understanding of how the business work and how the data should interact with each other.

# **5. Individual Report (Leong Seng)**

## **5.1 Data Cleansing**

### **5.1.1 Sources**

The dataset used is the combination of data from numerous sources, namely data.gov.sg, HDB.gov.sg, goggle map and Wikipedia.

The data source from gov.sg and HDB.gov.sg are creditable as these are government data-mall where government organization can provide reliable, relevant and timely statistics to support Singapore’s social and economic development. Data have been made available and accessible to the public through online portal so anyone can download and analysis them. There are numerous data that is extracted for example, resale-flat price, resale price Index and number of resales flat application. I even extracted spatial data on planning region and towns. The data were uploaded by related government sector such as HDB and URA.

The data from Wikipedia are quite creditable as Wikipedia covers a lot of topic across the entire globe. All the data comes from different source and Wikipedia compile them together and provide useful information. The data I extracted from Wikipedia is the planning area of Singapore.

The data from goggle map are creditable as there covers thousands of roads in hundreds of countries across the entire globe. All geographical data comes from different source and goggle complies everything and provide an accurate map experience. The data I extracted from goggle map is the postal code for the resale flat.

### **5.1.2 Extraction**

**Getting postal code for resale price flat data**

Before looking for data, I read and analyze the business questions of whether resale flat is affordable. I identify all the information required for the authorities to research and analyze. The identified information’s are resale-flat price, resale price index and number of resales flat application, planning area of Singapore. There was a data base about all HDB information (GPS coordinates, unit numbers, postal code, lease year, number of unit’s breakdown by flat type, upgrading program, etc.). However, I need pay to use the database. It was very impractical as I need to be paid to use. After further researching, I found similar data from data.gov.sg. The data is in excel format. However, there is no postal code. Therefore, I had to extract manually using goggle map. The resale-flat data only have block number and street name, therefore I had to use formula CONCATENATE ("blk"," ",blk, " ",street) to combine them into a full address. After having the full address for the different transactions, I had to copy the address one by one to goggle map to obtain the postal code. Due to manual extraction, it is inevitable for having errors. Since postal code are made up of 6digit code, the error has to been checked using IF(LEN(A1)=6,"True","False") to check if there are six postal digit. If the postal code does not correspond, the error should be located through eyeball-scanning. It was a tedious process as continuous checking has to be done and it took hours to complete due to the large amount of record 100332 record. It is not possible to obtain postal code for all of the record due to the short duration. therefore, I have only managed to extract some postal code for some records in order to do a mock-up dashboard for the map feature.

**Getting Resale Application Data**

I found the number of applicants for resales flat in HDB.gov.sg. The details of Resale application were extracted from the website. The columns are quarter, different type of room and the total. However, the data only show the individual year and not complied together. Thankfully I have copied the data from the website. I compile them in excel manually by copying the different year together. For consistency, I copied from year 2012 to 2017 since my data is based on the year.

**Getting Planning Areas of Singapore**

I found the planning areas of Singapore in <https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore>.

The details of the planning area were extracted from the website. The columns are the various name different languages, the region, area, population and density. I can also copied from the website and I compile them in excel together.

**Getting spatial Data for Boundary’s**

The geospatial Data was extracted from data.gov.sg. This data is very important to create an interactive map in tableau. The data that I wanted to obtain is the polygon data. Polygon data is used to represent real world features in a GIS. A vector feature can have a geometry type of a polygon. The polygon data I found is an indicative Polygon of region boundary. To facilitate urban planning, the Urban Redevelopment Authority (URA) divides Singapore into 5 regions, namely Central, West, North, North-East and East Regions. Tableau accepts a variety of spatial file such as ESRI shapefile, MapInfo Tables, KML files and GeoJSON File. The spatial data I obtained is in KML format which can be directly imported and used in tableau.

**Cleansing Resale Price Flat Data**

Cleansing was done for Resale Price Flat. The data of transaction was downloaded from data.gov.sg in an excel format. The data has details such as month, town, flat-type, block, street name, storey range, floor area, flat model, leash commerce date and resale price. After looking at the data, I decided which are the columns required and what I needed is month, town, flat-type, block, street name and floor area, flat model, leash commerce date and resale price. Therefore, I hide the columns that is not required to ensure that I would not be distracted by the information.

Since the month is in year-month and I want the year details, I created a new column called year using LEFT (year-month,4) to obtain the year. Afterwards, I created a column named remaining leash to find out the year leash left for the different transactions. Since every HDB have a leash of 99 years, I use a formula 99-(2017-leash commerce date) to obtain the remaining leash.

Since real estate are measured either using per square foot or per square meter. I decided to include both measurement for convenience of the user as they can choose what format to use to make comparison. Floor area sqm column is given, I decided to create a new column named floor area psf by using formula (floor area sqm\*10.7639) to obtain the floor area psf value.

Price per square foot and price per square meter are helpful when weighing up an HDB Flat’s value and learning the general value of HDB Flat’s value in a particular area. Therefore, I decided to create two column prices per square foot and price per square meter by diving the resale price with the area.

**Cleansing HDB resale house index source data**

Another cleansing was done for house resale index. The data of transaction was download from data.gov.sg in an excel format. There data has details such as quarter and index. I decided to use all the data and include a new column called year. Since the month is in year-month format and I only wanted the year, I created a new Colum named year using left LEFT(B:B,4) to extract only the year. The year start from 1992 to 2017. For Data consistency, I only extract from year 2012 to 2017. This allow user to see a consistent view of the data, as all my data start from the year 2012. This also prevent user being misled by the information from 1992 to 2012 affecting their decision. I remove the resale index for 2017-Q1 as it is not complete as a year. I am comparing based on year by adding Q1-Q4, since 2017 data is incomplete there no pointing keeping the data.

**Cleansing resale application data**

Another cleansing was done for resale application data. There data has details such as year quarter, various room and total. I decided to use all the data and include a new column called year. Since the month is in year-month format and I only wanted the year, I created a new Colum named year using left LEFT(B:B,4) to extract only the year. I hide the year quarter as it is not useful for me.

**Cleansing planning area of Singapore**

The details of planning area of Singapore were extracted from Wikipedia planning area of Singapore. The data of planning area was in a table format. I am able to convert into excel format by copying the table into excel. There data has details such as Name, Malay Name, Chinese Name, Pin yin Name, Tamil name for locations and other data such as region, area, population and density(/km2). After looking at the data, I decide to remove Malay name, Chinese Name, Pin yin Name and Tamil name as there are not relevant. I hide population and density(km2) as they are not relevant to the topic.

## **5.2 Analysis**

### **5.2.1 Data Description**

After cleansing, there are a total of 5 excel sheets containing information of the resale price flat, housing resale index, number of resale application and region information and map information’s. There are also two spatial file containing information of planning region boundary and town region boundary.

**The resale price flat data**

The resale price flat data can give an in-depth information about the resale price transaction. The transaction information is managed by HDB who will regularly update the data. The resale price flat data consist of both qualitative data and quantitative data.

**Qualitative (text data) data column:** town, flat type, block, street name, flat model, leash commence date, Address, planning region

**Reason for categorizing them in such way:** Each of them describes the quality or characteristics of HDB. Although block number is in numerical form but it is group under qualitative data because it acts as a tag to identify the HDB together with the street name. They are ordinal values, it is possible to sort them in order but not use them as measure.

**Quantitative (number data) data column:** year, floor area sqm, resale price, remaining leash, floor area per psf, floor area per sqm, unit price per psf

**Reason for categorizing them in such way:** These data have meaning as a measurement for HDB. E.G remaining leash measure the remaining life span of the HDB and floor area per sqm measure the price of the HDB. We can also apply mathematical data operation on the data in tableau. We can also sort the data in ascending and descending order.

**Combination of qualitative and quantitative**

The resale price flat data use a combination of qualitative data and quantitative data. This can improve the analysis because by ensuring that the limitations of one type of data are balanced by the strength of another. This can improve the understanding of HDB resale flat by integrating different perspective of HDB resale flat data.

This also allow us to generate hypotheses from qualitative data to be tested through the quantitative approach. E.G qualitative data(town) and quantitative data (resale price). Will the location of resale flat affects the resale price?? This allow us to come out with trends using hypotheses to test allow us to answer the business question.

**Resale Price Index Data**

The resale price index can track the overall price movement of the public residential market. The index is computed using both stratification method and stratified hedonic regression method. The housing index consist of index and year column both are numeric number.

The resale price index data is considered a quantitative data because the data is in numeric form. It can be further categorized into discrete data because they can be counted and even be listed out.

We can even apply statistical test in making statement about the data in tableau. These include descriptive statistics like the mean, median and standard deviation. We can even use advance feature in tableau to forecast using inferential statistics like t-test or multiple regression correlation(MRC). This allow us to derive important facts from this data and even forecast for the future.

**Resale Application Data**

The resale application data track the number of applicant for resales flat based on the different type of room flat. The resale application consists of year, different type of room with values and total.

The resale application data is considered a quantitative data because the data is in numeric form.

**Planning area Data**

Planning areas are the main urban planning and census division of Singapore delineated by the urban redevelopment authority. There is a total of 55 areas divided into 5 different sub zones. The columns are Name, Region and areas.

The planning areas is considered a qualitative data because name, region and areas describe the characteristic of planning area.

**Spatial Data**

The spatial Data that I collected is also known as geospatial Data. It is information about a physical object that can be represented by numerical values in a geographic coordinate system. After I import the data into the tableau, I was given three column, Name, Description and Geometry. The most unique part for this data is the geometry. It uses MULTIPOLYGON. A MULTIPOLYGON is a Multi Surface whose elements are Polygons. Using them we will be able to divide the Singapore map into desired areas.

## **5.3 Graph Creation**

**1)Bubble Graph using bar graph**

To find out if the location of the resale flat affects the price of the resale flat. I decided to use bar graph to compare. My X-axis is Town which is location and My Y-axis is unit price per psf. The reason why I choose unit price per psf instead of resale price because unit price per psf is a more useful indicator when comparing flat in various location.

The reason why I choose bar graph because bar graph are an effective way to compare items between different groups. After seeing the bar graph, I am unhappy with the results produces. This is because the bar sticks too closely and their height is very similar. Therefore, I decided to play around with the bar graph. Instead of using the traditional bar, I decide to play with the various option e.g. areas and shapes. Circle Produce the most satisfying results as It allow user to know the difference immediately just by having a look.

To provide a more visual impactful, I decided to play with the sizes of the circle. I made the circle be proportional to the price per psf. This allow the user to see just base on the size. I also created a minimum and maximum measures to display the minimum and maximum value using different color.

**2)Tree Map**

To find out the difference in resale flat price in different areas. I decided to use tree map to represent the data. I will be using Average resale price to compare with the town. Using tree map allows me to display large number of items with the limited space.

To provide a more visual impactful graph, I decided to play with the sizes of the boxes. In this tree map, both the size of the rectangles and their color are determined by the value of resale price—the greater the resale price for each category, the darker and larger its box.

**3)Line graph**

To show the trends of housing resale index, I decided to use line graph to show the trend with index as y axis and year as x-axis. Since the housing index is a time series data. Using line graph are particularly useful for identifying patterns and trend in the data.

Line graph are also useful as they can help to make predictions about the results of data not yet recorded. Tableau allow forecast future values of a regular time series of values from weighted average of past values of the data. I manage to forecast the year 2017 index.

**4)Bubble chart**

To show the number of application for resale application, I decided to use bubble chart to interpret the data. I use sum of total and filter it by years. Bubble chart allow comparison of entities in terms of the position and their size.

Color-coding bubble further categorizes them and add clarification for the user. I color-code the bubble base the year. I also add label, it is important because it increase the readability of the graph. This allow user to find insights without any confusion.

**5)Line chart for resale application**

To show the trends of resale application, I decided to use line graph to show the trend with application as y axis and year as x-axis. Since the housing index is a time series data. Using line graph are particularly useful for identifying patterns and trend in the data.

**6) Bar Graph**

To find out the number of transaction for resale flat for different flat type. I decided to use bar graph to interpret the data. I use flat type as my x-axis and the number of transaction as my y-axis.

This type of chart is one of the more familiar options as it is easy to interpret. It is also very useful for displaying data that is classified into nominal or ordinal categories.

It is also very easy for the user to compare between the categories of data. I also color code the categorizes to add clarification for the user.

**7) Choropleth Map for regional**

To find out the resale price for each region I decide to use a map to represent the data. A map is a representation of the earth’s surface or a part of it drawn on a flat surface according to a scale.

Choropleth maps are one of the most frequently used maps in infographic style visualizations. As the name suggests, color is the important part to these maps. A color scale is assigned to categorical and the value for each region is used to color the region. I manage to create the region by using polygon data. The entire Singapore map is divided into few regions after inserting the polygon data.

I insert information of average resale price into the map. To do this, I must use inner join to join both data together, spatial data and resale flat price data. I use the planning region in result flat price and join the spatial data. At first, I am unable to do as the word region is missing from flat price data and the spatial data is all caps. Therefore, I have to re cleanse my flat price data, by adding the word region and caps every word to every transaction.

It is very easy for the user to visualize areas that is more expensive just by looking at the map.

**8)Choropleth Map for areas**

Due to the availability of data of the price by town and polygon data. I decided to use choropleth map. I tried to change town into geological town and apply to the map. However, Tableau is not able to recognize the different towns in Singapore.

I tried to use polygon data instead. I manage to create the town by using the polygon data. The entire Singapore is divided into all the different towns.

To use both data, resale price with postal code sample data and the spatial file. I must inner join them together. I join them by using town from resale price data and name from spatial file since there are the same.

When I insert polygon data, the map is not fully filled with towns. This is because due to time constraint, I am unable to obtain all the postal code and I use sample data that covers only a few locations.

To show different data together in a same map, I must create a multiple layer maps. To create multiple layer maps, I just drag my longitude and latitude to the column and rows to duplicate two maps. Both maps are independent from one another before I merge them.

In Map 1, I insert all the information about the resale flat such as address, postal code, price per psf, resale price and remaining leash.

In Map 2, I insert all the polygon data to divide the map into different town.

To merge them I click on either on longitude or latitude, by clicking dual axis I am able to merge the map together.

User can click and interact with the map to find useful data.

## **5.4 Dashboard Description**

For the final dashboards, there are 3 dashboards created. The dashboards are categorized by resale market environment, general trend and map information. The dashboards are aimed to aid HDB analyst to monitor the current resale market. Thus, having a resale market environment dashboard is important to portray the current prospect of the resale market. There should be a dashboard showing the general trend of the market. The main purpose of this dashboard is to analyze the pattern of the market. The Map dashboard shows the information of the resale flat. The main purpose is to analyze the information of the resale flat.

### **5.4.1 Resale Market Environment**

Resale market environment dashboard has a total of 4 graph, Map that is divided into various region. A flat type graph, bubble graph and a tree map. User can click on the Map to drilldown onto each individual region. If user want to compare with different region, there is a filter for the user to choose which region to compare and drilldown on. There is also a filter for user to drill down for the different year. By looking at resale market environment dashboard, user is able to look at different aspect of the resale market including the price of the resale market, the trend and even the type of flat-sold. This will allow user to decides if there is a need to monitor the market due to the significant increase in resale price. Currently, the price of the resale flat is increasing steadily.

### **5.4.2 General information for resales market**

General information for resale market has only 3 graph bubble chart of resale application, line graph of resale application and line graph of resale index. By looking at the dashboard, the user is able to find trends from the resale house index. They will also be able to compare the number of applicant for the different year.

Currently the number of applicant and housing index is increasing steadily. Resales flats are now attractive for people as transactions can be completed in few months compare to BTO. The number of applicant has been increasing 2014 to 2016. This also means resale market Is more popular compare to BTOs.

### **5.4.3 Resale Information Map**

The resale information map has only 1 map that provide all the information about the transaction records. There is a filter which allow user to select regions to be compared, towns to be compared. By looking at the dashboard, the user will be able to get useful information on resale flat such as location, price, leash remaining. So that they can come up with measures to control the resale market if one area is too popular. To customize resale information map to suit the theme, I create custom shape in tableau to import shapes to adjust the shape. I replace the circle with house icon. So, that user can see a more meaningful map instead of some symbol that they have no idea what is it about.

### **5.4.4 Analysis**

In conclusion, the central region has the highest resale price by comparing with other region they have highest resale price. Also by looking at other indicators like price per psf, central region is the most expensive per psf compare to other region. 4 room flat is also a very popular choice based on transaction record as they have the number of transaction every year. From 2014 to 2016, the number of resale application increase sharply, this correspond with the housing index. The housing index is risking over the years.

## **5.5 Reflection**

Through this project, I managed to apply and have a better understanding the theory I learnt in class. At first, the theory such as ETL seem very easy, but I was wrong. ETL process requires a lot of effort and skills. I remembered what my lecturer told me is true, getting data is the hardest part while coming up with graph is the easy part. I read some article from professional data analyst, they spent 80% of their time on data prepping and ¼ of that time analyzing data and visualizing the results. I was convinced by them because I spent hours cleaning and verifying my data, although it was very tiring but I really learnt a lot through cleansing. Data cleansing is also very important process. I have to identify inaccurate data and unnecessary data and modifying them or even remove them. I have to apply a lot of method in excel to obtain a clean data. By removing columns and creating new columns to extract the right information. I am also required to use formula to validate if the data is accurate or not. Data quality is very important if I want to take advantage of the obtained data in a meaningful way and productive way to give me that competitive advantage. I really appreciate the importance of keeping data clean and it is an important aspect in every industry. Data integration is another thing I learnt. My data involves combining data from several disparate sources, which are stored in different format. Data integration allows me to make the different source into meaningful and valuable information. I also learnt that data does not only come with excel format, it can come in many different sources such as text file, JSON file, spatial file and many more. Thus, by mastering data integration, it allows me to merge them into useful data. One of the hardest part for this project is to determine where should we get the data from. This project requires me to source my own data source. I have to determine which one is a “good” or “bad” dataset. It takes times to learn and determine. I have learnt that a good dataset will allow easy manipulation and calculation for meaningful statistical analysis. A bad dataset is poorly structured and is very hard for me to manipulate the data. A good dataset is very important for analyzing data.

# **6. Individual Report (Weixiang)**

## **6.1 Data Gathering and Preparation**

### **6.1.1 Background Information**

In Singapore, ethnic quota will affect the residents to buy HDB flats and the chance to succeed in balloting of HDB Build-To-Order (BTO). The ethnic quota when it was first introduced was a novel idea meant to keep the peace among the races by having a percentage distribution of the races within the blocks of flats in an estate to maintain not just a kampong culture but also to help foster racial tolerance in a burgeoning Singapore. The racial allocation are as follows; 84% for Chinese, 22% for Malays and 10% for Indians and other minorities. The percentages for the races in a block and the estate are close as well. And hence, the business question that I am answering is with ethnic quota, will the ratio of each raves that purchase HDB flats affect the sales? Hence, I will be making this dashboard, whose end users will be the HDB sales analyst, for them to view the trend of this question.

### **6.1.2 Sources**

The dataset used is collected from data.gov.sg. It is a site that has Singapore’s public data. As the data is by the government of Singapore, the data collected from here is reliable. By searching HDB data, I am able to retrieve relevant data about my Case study question: To find out will the ratio of each races that purchase HDB affect the sales. I have retrieved several data namely HDB resident population by ethnic group and flat type, estimated resident population living in HDB flats, economically active HDB population, area population and the median size per household.

### **6.1.3 Extracting and Cleansing**

Before extracting data from the site, I read the background information regarding to my business questions to further understand what data set I would need to use to design my dashboard. By searching “HDB” in the site, 59 datasets were shown. As not all datasets were relevant to my business question, I had to look through the datasets one by one and choose those that I would need. I had took 11 datasets among the 59 datasets, afterwards I took out more datasets as some excel sheets had the data sort by both flat type and ethnic group. Datasets with ethnic group was what I need, hence those datasets sorted by flat type were filtered out.

After extracting the datasheets needed, I copied the original dataset to a new excel file with only the columns and rows that is related to the business question. I proceed to cleanse data as there were irrelevant data columns in the dataset that need to be removed. Example will be for the data of individual ethnic group that bought HDB flats from 1960 to 2016. There were columns that are not needed when making dashboard; Gender and total resident. As my main topic was about ethnic groups, gender was not needed in this case. After removing, I merge all the excel files together to produce a final excel document with the datasets I would use for dashboard. Final names for the excel sheets will be Ethnic Group Population, Employment Status, Area Population and Median Size per Household.

Afterwards I sort the data by adding headings and filtering. This made me easier to view the datasets instead of the original messy datasets. By having common headings, it was possible for me to do filtering in the dashboard in Tableau. By using filter function, I was able to filter out the data that was not needed and delete them, and also to sort the data alphabetically. With the sort function, I was able to view individual data. Example will be for “Ethnic Group”, it would be tedious for me to look for Chinese data if the sort function was not implemented. With the function implemented, I was able to view the rows for ethnic group “Chinese” with just a click. For the Ethnic Group Population sheet, as ethnic group names were “Total Chinese”, “Total Indian”, “Total Malay” and “Other races(Total)”, I need to replace them with “Chinese”, “Indian”, “Malay” and “Others” for consistency with other datasets. As there were hundreds of rows, I used Ctrl F command to use find and replace function. This reduced the time needed for me to replace all the rows, instead of replacing row by row. For consistency in all excel sheets, columns heading were renamed. “Financial year” was indicated for years, “Ethnic group” for different races.

## **6.2 Analysis and/or Model Building**

### **6.2.1 Data Description**

After cleansing the datasets, I am left with 4 excel sheets, which would be Ethnic Group Population, Area Population, Employment Status and Median Size per Household. Ethnic group is able to provide the details of the number of individual ethnic groups from 1960 to 2016; Employment Status is able to provide the details of the number of employed residents of individual ethnic groups; Area Population is able to provide the details of the number of residents living in HDB in different town or estates in Singapore and lastly, Median Size per Household is able to provide the details of the median size of individual ethnic groups.

### **6.2.2 Creating of graph**

I have created 4 graphs based on the dataset collected. Mainly pie charts and bar graphs to describe the business question. Before creating graphs, I had to visualize how the graphs would look like, with what kind of filters and which dimensions to fill up the columns and rows of the graphs. As the large-scale Sample Household Surveys (SHS) is conducted every 5 years, the financial year for Employment Status and Median Size per Household have only 2008 and 2013.

Ethnic Group Population-

Shows the total of the individual ethnic group that bought HDB flats from 1960 to 2016.

From the pie chart, it shows that the most number of ethnic group that bought HDB flats during that period of time was Chinese (74.47%), followed by Malay (13.46%), Indian (9.02

%). And lastly other ethnic groups (3.06%). The reason would be that the biggest ethnic group in Singapore is Chinese, and with the ethnic quota for each HDB flat, hence Chinese occupied the most portion of the residents living in HDB flats. Although a ratio of people may purchase other types of housing like condominium and terrace house, it will not affect much of the data collected. Filters for Financial Year and Ethnic Group are added, to allow users to change the data shown according to year or ethnic group, which will be used in the dashboard to filter out data. Color of each ethnic group are identified to differentiate the different ethnic groups more effectively, which allows the user to spot easily.

Area Population-

Initially, I wanted to create filled map for Area Population, however I was unable to find datasets that contain specific geographical locations of individual town and estate to use as geographical dimension in Tableau. Hence I resort in using bar graph to show the area population in Singapore.

From the bar graph, it shows that Jurong West (502,200) has the largest population that lived in HDB flats, followed by Tampines (473,300) and Woodland (468,800). After doing background information checking, I am aware that Bedok, which is ranked 4 in the graph, is the top 3 areas with the largest population. The reason of being rank 4 as it has the most amount of condominiums, private flats and landed properties, which is not the field I am researching at. The least population that lived in HDB flats is located at Bukit Timah (17,000), the reason will be that Bukit Timah has one of the highest densities of private housing compared to other town and estate. Due to the low amount of HDB flats built, the data will show that Bukit Timah has the least population living in HDB flats. For Jurong West, Woodlands and Yishun, there are several reasons for higher population of residents. Firstly, these areas are considered as older areas, as they were developed earlier compared to other areas like Sengkang and Punggol. As a result there are many elderly and mid age residents continue to live at those areas. As I had been living in Yishun for more than 10 years, I had observed that not much changes had been made. Secondly, the distance between these areas to Malaysia checkpoint is short compared to other areas. This advantage attracts many residents to buy HDB flats at those areas, mainly Malaysians that need to travel back and forth to Johor Bahru or visit relatives during festivals. Hence the results are shown as this.

Filter for Financial Year is added, to allow users to change the data shown according to year or ethnic group, which will be used in the dashboard to filter out data. Sorting is added to the column to display the graph is descending order, which can be changed by clicking to display in ascending or original order.

Employment Status-

Bar graph is used to show the differences of the employment status between the ethnic groups. From the graph, it shows that Chinese has the most employment status, followed by Malay, Indian and other ethnic groups. Being pointed out to have percentage values are better for individual bar in the graph during presentation, I had changed the mark labels with percentage of the values, which is easier for the user to view the data too.

The reason for the increased of employment status of the residents is due to the increase in the rate of home-ownership from 2008 to 2013, which is researched by my team mate as his business question. Filters for Financial Year and Ethnic Group are added, to allow users to change the data shown according to year or ethnic group, which will be used in the dashboard to filter out data.

Median Size per Household-

Pie chart is used to show the median size per household. From the graph, it shows that Indian and Malay (8) has the biggest median size, followed by other ethnic groups (7) and Chinese (6). This shows that ethnic groups except Chinese tends to have a bigger family, which cause the monthly income per household member to decrease, and hence the breadwinner would need to earn more in order to buy HDB flats. Filters for Financial Year and Ethnic Group are added, to allow users to change the data shown according to year or ethnic group, which will be used in the dashboard to filter out data. Color of each ethnic group are identified to differentiate the different ethnic groups more effectively, which allows the user to spot easily.

## **6.3 Final Tableau Story/Dashboard Description**

Before making the final dashboard, I had planned how will it looks like to prevent messy placing of the graphs and filter functions. All the graphs are using the same data source, hence it is possible to implement filter and drill down functions with the same heading name, which I had changed when cleansing the datasets. As every graphs have individual filters, I had to remove 3 the financial year filter, 2 Ethnic Group filter and 1 Ethnic Group color chart. The main purpose of the dashboard is to help user to identify how the ratio of the ethnic groups affect the sales of HDB flats, and the user can plan measures on how to improve the sales of HDB flats, by targeting at the specific ethnic groups.

There are 2 filters applied, financial year and ethnic group. User can choose to show all the years or only 2008 and 2013. For example, user can select year 2008 and only dataset in the year of 2008 will be shown in the dashboard. User can choose to view data of the specific ethnic groups by choosing the specific ethnic group at the filter button. For example, user can select Malay and only dataset that with Malay will be shown in the dashboard. Color chart is implemented to view each ethnic group without drilling down. For example, user can press blue for Chinese, and all the graphs that contains Chinese will be highlighted. This helps the user to look at specific ethnic group more easily.

In conclusion, ratio of ethnic groups will affect the sales of HDB flats. From 2008 to 2013, with the increase of employment status of the residents, more HDB flats are sold, which can be shown by filtering the years individually. Data shown from both the bar graphs support the fact.

## **6.4 Reflection**

After this project, I am able to use all the skills that I had learnt from the practical through the semester, and also other features of Tableau that was not taught in the lesson. I had to look for tutorials online in order to carry out the functions. This is the first time that I had to look for raw data and cleanse it in order to make a clean dashboard. As usually datasets were given and were cleansed for us to use, it is a fresh experience to do the things myself.

Build-To-Order (BTO) - Housing and Development Board (HDB) flat allocation system that offers flexibility in timing and location for owners buying a New HDB flats in Singapore.

Ethnic quota – Implemented to ensure a better racial mix in HDB estates, to promote racial integration and harmony.

The goal of this research is to find out that will the ratio of the ethnic groups affect the sales of HDB flats, goal was made by looking for dataset from government data sites and use them to create a dashboard to see the result and trends. By following the timelines closely, I am able to finish the research and presentations by the deadlines.

In conclusion, this research will address that ratio of ethnic groups do affect the sales of the HDB flats, as Chinese occupied the largest ratio, if there are lesser Chinese to purchase HDB flats, the sales will be heavily affected.

# **7. Report Answer**

## **7.1 Findings**

In conclusion, there is a need to a certain extend to improve the services and products that the HDB is providing with the residents. We can use our findings from the dashboard to come out with a solution in order to improve it. The findings from the dashboard will help to improve the services and products provided by HDB.

From Jun Liang’s dashboard, we can see that Seng Kang seems to be a good place to build flats as there is a increase in the residents. The type of flats that should be build would ideally be either 4 room or 5 room as it seems to be the type most commonly seen in Seng Kang. The ethnic ratio can be about 75% chinese, 9% indians, 13% malays and the remaining for the other ethnic group. The amount of people the flats should be able to house is about 130000.

From leong seng’s dashboard, We can see that town centre area is the most expensive places to get HDB resale flat. Every year, it top the most expensive per psf. We can see that choa chu kang is the cheapest places to get HDB resale flat. Every year, it top from being the cheapest per psf.We can also see the housing index is increasing steadily from 2016-2017.In 2016, there is a huge demand for resale flat application. Centre region is the most costly for resale flat price.

From Seow Chong’s dashboard, Home-Ownership rate is still on a steady climb since 2011 from 88.6 to 90.9 in 2016. The increase number of units sold as compare to number of units rent is more significant thus corresponding with the Home-Ownership rates. From 2011 to 2016, Singapore’s GDP was on a climb from 342 billion to 402 billion. Prices of houses increased during the period of 2011 to 2013.However, it dropped from 2013 to 2014 and currently, the prices of housing is having a slower climb from 2014 to 2016.The total amount of CPF withdrawal has increased from 2011 to 2016 however the number of people using CPF to withdraw has lessen.

From Weixiang’s dashboard, we can see that Jurong West has the most number of residents that live in HDB flats, while Bukit Timah has the least number of residents that live in HDB flats. Ethnic group population within each HDB flat follow closely to the Ethnic quota set by the government. In conclusion, ratio of ethnic groups will affect the sales of HDB flats. From 2008 to 2013, with the increase of employment status of the residents, more HDB flats are sold, which can be shown by filtering the years individually. Data shown from both the bar graphs support the fact.

From Nicholas’ dashboard, we can see that locations do in fact affect the prices of HDB flats in the area. However, as to how it affects, the reasons can be very broad and not just be due to the number of facilities in the area. Reasons such as nearby schools or recreational facilities can be the cause as well. Reputation of the area can be another reason because an area with a bad reputation may have low flat prices in order to attract customers. Even if the number of MRT Stations in the area is not the only reason for the difference in prices, it is definitely one of the more influential reasons.

Another thing that can be seen is that prices affects the sales of HDB flats, despite seeing number of sold units increased as prices increase, we can look at the data another way and see that people are buying 4 & 5 room flats more than executive flats which are usually priced higher. However, it is still a fact that a drop in price can lead to a drastic increase in sold flats. Furthermore, the Homeownership Scheme could be encouraging people to buy a flat instead of renting one, thus the increase in number of sold flats every year.

In conclusion, locations do affect prices of HDB flats and the price in turn can also affect sales of the flats. However, there are many other factors that contributes to this result as well.

## **7.2 Solution**

From these findings, we have came out with some solution that will make use of the findings and improve the services and products that is provided by the HDB.

Firstly, there is need to improve the HDB in order to meet the expectation of the customers. The  solution to improve the the standard of quality of HDB Flats is to make use of the information gotten from the dashboard and help in the decision of the details of the HDB flats to be built. From the findings, they should build new HDB Flats in Seng Kang as there seems to be a well recieved place, seeing that there is a high trend that more residents is living there. And the type of flats build should be 4 to 5 room type of flats as it is common and also at the same time able to house more residents. The time of completion for the HDB Flats should be at around year 2019 as the estimated increase is for the year of 2019.

Secondly, there is need to improve and monitor the HDB resale market to meet the expectation of the buyer. Firstly,more grants can be provided for buyers. The Grant provided will be able to soak up the increasing cost of HDB resale price. This will help some families to relieve the expensive cost of HDB. HDB also can implement measures to cool the red-hot real estate market.They can increase the tax to be imposed on buyers in the who are not citizens, permanent residents.This ensure that the HDB resale flat is for those who really need a house which is singaporean. This will cool down the market making resale price flat affordable. They can also lower the seller stamp duty.  SSD is currently payable by those who sell a residential property within 4 years of purchase, at rates of between 4 per cent and 16 per cent of the property's value. So that they will be more supply for resale flat thus lower the resale price. Lastly they can build more BTOs so that they will be more supply in the market making resale price HDB flat affordable.

From Seow Chong’s findings, we have realised that there is no need to make any changes to policies and schemes regarding the prices of housing as the homeownership rates continue to increase and prices of housing does not have any sharp increases

Improvements can be made to the CPF Housing Scheme so that more people are able to benefit from this scheme as the number of people has dropped since 2011.

From Nicholas’ findings, there may be a need for the HDB representatives to work with the government to plan out the area in Singapore, such as planning which area to set up more MRT Stations or other facilities to make the prices of HDB Flats more worth. They can also work to put a limit on the resale prices based on location so that the flats will not be overpriced for other buyers. Thus, encourages homeownership in Singapore, as people will be able to afford to buy a flat instead of renting it.

From Weixiang’s finding, there may be a need to make changes as by viewing the data of 2008 and 2013, there is an increase of area population, which shows that the sales of HDB flats are increasing steadily. However, HDB analysts can further improve the sales of HDB flats in other areas more by building more affordable flats and provide schemes and subsidies for the residents in those areas that hope to purchase a HDB flat apartment.

# **8. References**

<https://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/geo_spatial_data/ssnsep09-pg8-12.pdf>

<https://en.wikipedia.org/wiki/Bukit_Timah>

<http://www.propertyguru.com.sg/singapore-property-resources/singapore-hdb/how-the-ethnic-quota-can-affect-your-selling-buying-ability>

<https://data.gov.sg/dataset/number-of-members-net-amount-withdrawn-under-housing-schemes-annual?view_id=dd81f418-34f2-401f-b089-b854ec3e6556&resource_id=c10b6660-2926-4f0b-b997-f1d2d0d26cf3>

<http://www.singstat.gov.sg/statistics/visualising-data/charts/home-ownership-rate-of-resident-households>

<https://data.gov.sg/dataset/resale-flat-prices>

<http://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=12359>

<http://www10.hdb.gov.sg/eBook/AR2016/key-statistics.html>

<http://www.todayonline.com/commentary/rethinking-singapores-housing-policies>

<https://data.gov.sg/dataset/cumulative-units-completed-since-1960>

<https://data.gov.sg/dataset/estimated-resident-population-living-in-hdb-flats>

<https://data.gov.sg/dataset/hdb-properties-facilities-under-management>

<https://data.gov.sg/dataset/number-of-units-of-hdb-developments-by-status>

<https://data.gov.sg/dataset/number-of-sold-and-rented-hdb-properties-facilities>

<https://data.gov.sg/dataset/age-of-hdb-population-by-ethnic-group-and-flat-type-in-age-groups>

<https://data.gov.sg/dataset/estimated-resident-population-living-in-hdb-flats>

<https://data.gov.sg/dataset/resident-population-by-ethnicity-gender-and-age-group>

<http://www.propertyguru.com.sg/lifestyle/article/3/9-things-to-consider-before-buying-a-hdb-flat>

<http://www.hdb.gov.sg/cs/infoweb/business/estate-agents--salespersons/selling-a-flat/ethnic-integration-policy--spr-quota>