CM3010 Mid-Term Project

About the dataset

The dataset that I have chosen covers the field of housing in Singapore. The dataset contains the information of resale flats in Singapore. Since housing is a trending topic in Singapore, especially among young couples, it is interesting to study the dataset and draw conclusions from it. For example, how do the prices in each estate vary from each other? Such queries can help people to identify where the cheapest estates are for them to purchase their home.

Quality

The data is found on data.gov.sg, a Singapore government based website that stores a host of dataset related to anything in Singapore. Furthermore, the source of data is by the Housing and Development Board. Since it is a dataset provided by the government, it definitely has to be accurate and reliable.

Detail

Attributes provided in the dataset are detailed and insightful from the resale price to the specific location of the resale flats. They are more than sufficient for me to study the dataset and draw conclusions from it.

Documentation

Its documentation could be found easily on the website itself which clearly states what the dataset is about.

Interrelation

This dataset can be very useful to other datasets as it can be used to compare the prices of the resale_flats. Other datasets could easily be found on the data.gov.sg website that is related to this dataset.

Use

The usability of this dataset is very good as there are many things one can use it for. I, for one, can use it to identify which types or locations of the resale flats cost more or less. If there is one thing that I could not derive from this dataset is the number of previous owners the flat has to determine whether the number of previous owners affect the resale price of the flat as it is not provided by the dataset.

Discoverability

The data was readily available as it is provided by the Singapore government for people to use on a website dedicated to datasets related to anything in Singapore on data.gov.sg.

The dataset has the Singapore open data licence. This information is found on the documentation of the dataset on the website. Upon further research on the mentioned licence, with this licence, I am able to use, access, download, modify and adapt the datasets which means I am allowed to use the data for this project. The terms of the licence does not restrict the use of the dataset in any significant way however I must acknowledge the use of their dataset. [1]

Database Modelling
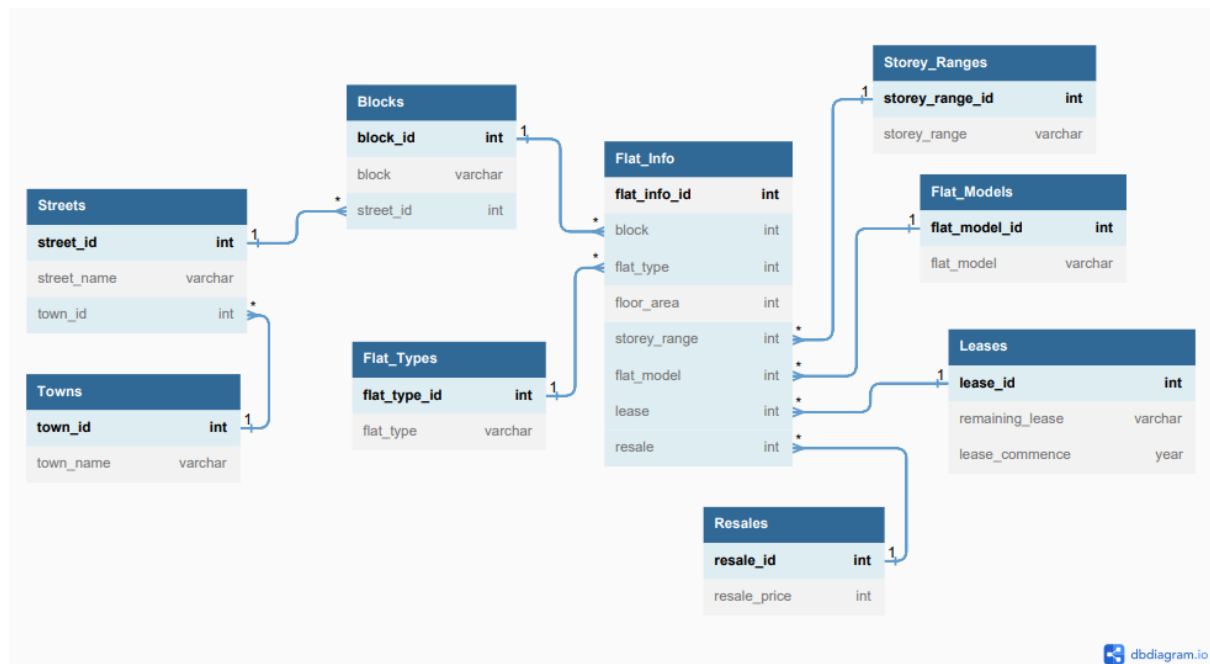
(Fig 1 Database Diagram)
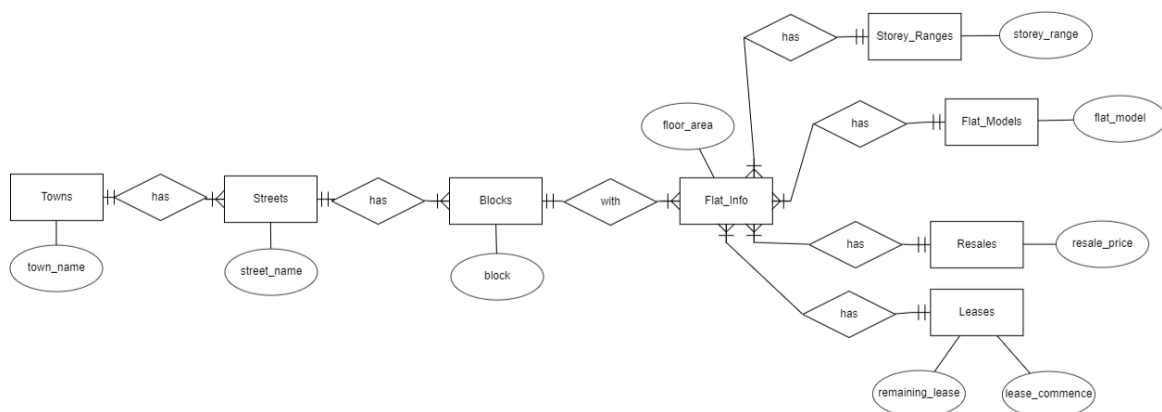
(Fig 2 ERD Diagram)

The original data from this database contains over 140,000 flats. Since this will be too much data for the environment to process, I have decided to reduce the data to about 1000 flats. I have also decided to remove a column which contains dates as the dates provided are dates

of when the flat information was added into the dataset which is irrelevant to the analysis of the dataset. Hence, the column is removed from the dataset.

Referring to figure 2 above which shows the ERD Diagram, the tables have a many to one relationship to one another for example, the table Towns, has a mandatory many to mandatory 1 relationship with Streets. Figure 1 which depicts the Database diagram elaborates the relationship of the tables with one another showing the foreign keys connecting one table to the other. The database is normalised into its 3rd Normal Form.

List of tables and fields

Table name	Fields
sample_data	Town varchar(50), Flat_Type varchar(50), Block varchar(10), Street varchar(50), Storey_Ranges varchar(50), Floor_Area int, Flat_Model varchar(50), Lease_Commence year, Remaining_Lease varchar(50), Resale_Price int
Towns	town_id int, town_name varchar (50)
Streets	street_id int , street_name varchar (50) not null, town_id int,
Flat_Models	flat_model_id int, flat_model varchar (50) not null
Storey_Ranges	storey_range_id int, storey_range varchar (30) not null
Blocks	block_id int, Block varchar (10) not null, street_id int,
Flat_Types	flat_type_id int, flat_type varchar (50) not null
Flat_Info	flat_info_id int,

	block int, flat_type int, floor_area int, storey_range int, flat_model int, lease int, resale int,
Leases	lease_id int, remaining_lease varchar(30) not null, lease_commence year not null
Resales	resale_id int, resale_price int

Database Creation

The create commands are input in the terminal of the coursera environment of mysql.

The database is created by the following command, CREATE DATABASE resale_flats;

Next, a user is created, CREATE USER 'user1' IDENTIFIED BY 'password';
Followed by granting it the permission to access the database, GRANT ALL ON resale_flats.* TO 'user1'@'%';

The table below shows the creation of the rest of the tables in the database.

Function	Commands
Create a table called sample_data that will store data from the csv file	CREATE TABLE sample_data (Town varchar(50), Flat_Type varchar(50), Block varchar(10), Street varchar(50), Storey_Ranges varchar(50), Floor_Area int, Flat_Model varchar(50), Lease_Commence year, Remaining_Lease varchar(50), Resale_Price int);

Create a table called Towns that stores town_name	CREATE TABLE Towns (town_id int PRIMARY KEY AUTO_INCREMENT, town_name varchar (50) not null);
Create a table called Streets storing street_name	CREATE TABLE Streets(street_id int PRIMARY KEY AUTO_INCREMENT, street_name varchar (50) not null, town_id int, FOREIGN KEY(town_id) REFERENCES Towns(town_id));
Create a table called Flat_Models	CREATE TABLE Flat_Models(flat_model_id int PRIMARY KEY AUTO_INCREMENT, flat_model varchar (50) not null);
Create a table called Storey_Ranges	CREATE TABLE Storey_Ranges(storey_range_id int PRIMARY KEY AUTO_INCREMENT, storey_range varchar (30) not null);
Create a table called Blocks	CREATE TABLE Blocks(block_id int PRIMARY KEY AUTO_INCREMENT, Block varchar (10) not null, street_id int, FOREIGN KEY(street_id) REFERENCES Streets(street_id));
Create a table called Flat_Types	CREATE TABLE Flat_Types(flat_type_id int PRIMARY KEY AUTO_INCREMENT, flat_type varchar (50) not null);
Create a table called Flat_Info	CREATE TABLE Flat_Info(

	flat_info_id int PRIMARY KEY AUTO_INCREMENT, block int, flat_type int, floor_area int, storey_range int, flat_model int, lease int, resale int, FOREIGN KEY (block) REFERENCES Blocks(block_id), FOREIGN KEY (flat_type) REFERENCES Flat_Types(flat_type_id), FOREIGN KEY (storey_range) REFERENCES Storey_Ranges(storey_range_id), FOREIGN KEY (flat_model) REFERENCES Flat_Models(flat_model_id), FOREIGN KEY (lease) REFERENCES Leases(lease_id), FOREIGN KEY (resale) REFERENCES Resales(resale_id));
Create a table called Leases	CREATE TABLE Leases(lease_id int PRIMARY KEY AUTO_INCREMENT, remaining_lease varchar(30) not null, lease_commence year not null);
Create a table called Resales	CREATE TABLE Resales(resale_id int PRIMARY KEY AUTO_INCREMENT, resale_price int);

Instance data

Description	Command
Adding data into Towns table	INSERT INTO Towns(town_name) SELECT DISTINCT Town FROM sample_data;
Adding data into Flat_Info table	INSERT INTO Flat_Info(block, flat_type, floor_area, storey_range, flat_model, lease,resale) SELECT b.block_id,ft.flat_type_id,sd.floor_area,sr.storey_range_id,fm.flat_model_id,l.lease_id,rp.resale_id FROM sample_data sd LEFT JOIN (SELECT bl.block_id, bl.block, st.street_name FROM Streets st LEFT JOIN Blocks bl ON st.street_id = bl.street_id) b ON b.street_name = sd.Street AND b.Block = sd.Block LEFT JOIN Flat_Types ft ON ft.flat_type = sd.Flat_Type LEFT JOIN Storey_Ranges sr ON sr.storey_range = sd.Storey_Ranges LEFT JOIN Flat_Models fm ON fm.flat_model = sd.Flat_Model LEFT JOIN Leases l ON l.remaining_lease = sd.Remaining_Lease AND l.lease_commence = sd.Lease_Commence LEFT JOIN Resales rp ON rp.resale_price = sd.Resale_Price;
Adding data into Blocks table	INSERT INTO Blocks(block, street_id) SELECT b.Block,s.street_id FROM (SELECT DISTINCT Block, Street FROM sample_data) b LEFT JOIN Streets s ON b.Street = s.street_name;
Loading data from csv file into sample_data table	LOAD DATA INFILE '/home/coder/project/mid-term/resale-flats/data/edited_sample_data.csv' INTO TABLE resale_flats.sample_data FIELDS TERMINATED BY ',' ENCLOSED BY " LINES TERMINATED BY '\n' IGNORE 1 LINES;

How well the database reflects the data

In the original dataset, there is a date column which reflects on when the specific column was added into the dataset. Since it is unimportant, I have decided to remove the column from the dataset. There is a column called remaining_lease which holds the information of how long left the flat has before its lease expires. That column has been left in varchar as it is input with both the year and month. Even though I am unable to convert it into a usable format for insights, I have decided to keep it in the dataset as it is vital information for users.

SQL Commands

Q1. What is the ranking of towns from most expensive to cheapest?

Taking the average resale price of all the flats in the specific town and ordering it in descending order, I will get a table of towns together with the resale price listing from the most expensive town to the cheapest.

SELECT tn.town_name as TownName, AVG(rp.resale_price) as ResalePrice from Towns tn join Resales rp join Flat_Info fi join Blocks bl join Streets st WHERE tn.town_id = st.town_id AND st.street_id = bl.street_id AND bl.block_id = fi.block AND rp.resale_id = fi.resale GROUP BY TownName ORDER BY ResalePrice DESC;

```
+-----------------+-------------+
| TownName        | ResalePrice |
+-----------------+-------------+
| BUKIT TIMAH     | 812296.0000 |
| BISHAN          | 604892.1818 |
| CENTRAL AREA    | 594583.3333 |
| BUKIT MERAH     | 512267.8571 |
| QUEENSTOWN      | 479381.7143 |
| MARINE PARADE   | 474500.0000 |
| KALLANG/WHAMPOA | 473707.6216 |
| PUNGGOL         | 469814.9474 |
| SERANGOON       | 466450.0000 |
| PASIR RIS       | 453576.0000 |
| TAMPINES        | 449650.7805 |
| TOA PAYOH       | 446057.2121 |
| SENGKANG        | 421884.7407 |
| CLEMENTI        | 420814.2857 |
| HOUGANG         | 416688.8485 |
| ANG MO KIO      | 413971.2143 |
| SEMBAWANG       | 410043.4783 |
| BEDOK           | 408598.4000 |
| BUKIT PANJANG   | 408351.1111 |
| JURONG WEST     | 402120.2143 |
| CHOA CHU KANG   | 392424.7273 |
| WOODLANDS       | 390294.7349 |
| JURONG EAST     | 389646.5556 |
| BUKIT BATOK     | 373450.8085 |
| GEYLANG         | 370543.3750 |
+-----------------+-------------+
25 rows in set (0.09 sec)
```

Q2. Does the storey range affect the pricing of the flat?

By calculating the average price of each storey range and ordering them by the highest resale price to the lowest, I can deduce that the higher the level of the flat is, the more expensive it is.

SELECT sr.storey_range as StoreyRange, AVG(rp.resale_price) as ResalePrice from Storey_Ranges sr join Resales rp join Flat_Info fi WHERE sr.storey_range_id = fi.storey_range AND rp.resale_id = fi.resale GROUP BY StoreyRange ORDER BY ResalePrice DESC;

```
+-------------+---------------+
| StoreyRange | ResalePrice   |
+-------------+---------------+
| 49 TO 51    | 1108000.0000  |
| 34 TO 36    |  808000.0000  |
| 37 TO 39    |  801500.0000  |
| 28 TO 30    |  733600.0000  |
| 22 TO 24    |  663722.0000  |
| 25 TO 27    |  640000.0000  |
| 19 TO 21    |  549437.5000  |
| 16 TO 18    |  489799.0000  |
| 13 TO 15    |  476320.5833  |
| 10 TO 12    |  440088.0370  |
| 07 TO 09    |  414696.2888  |
| 04 TO 06    |  408646.5034  |
| 01 TO 03    |  397525.4010  |
+-------------+---------------+
13 rows in set (0.00 sec)
```

Q3. Which town has the oldest flats?

By calculating the average years of the lease commence of every flat in a town, I am able to derive a table with the town name and average lease commence year in ascending order which tells me that Marine Parade has the oldest flats. From this table I can also come to the conclusion that Marine Parade is an old town with Punggol being the youngest.

SELECT tn.town_name as TownName,CEILING(AVG(l.lease_commence)) as LeaseCommence from Towns tn join Leases l join Flat_Info fi join Blocks bl join Streets st WHERE tn.town_id = st.town_id AND st.street_id = bl.street_id AND bl.block_id = fi.block AND l.lease_id = fi.lease GROUP BY TownName ORDER BY LeaseCommence ASC;

```
+------------------+----------------+
| TownName         | LeaseCommence  |
+------------------+----------------+
| MARINE PARADE    |          1976  |
| QUEENSTOWN       |          1979  |
| GEYLANG          |          1979  |
| TOA PAYOH        |          1983  |
| CLEMENTI         |          1983  |
| ANG MO KIO       |          1983  |
| BEDOK            |          1984  |
| BUKIT TIMAH      |          1984  |
| BUKIT MERAH      |          1986  |
| SERANGOON        |          1987  |
| KALLANG/WHAMPOA  |          1987  |
| JURONG EAST      |          1988  |
| BUKIT BATOK      |          1989  |
| BISHAN           |          1989  |
| HOUGANG          |          1991  |
| TAMPINES         |          1991  |
| CENTRAL AREA     |          1992  |
| PASIR RIS        |          1993  |
| JURONG WEST      |          1995  |
| CHOA CHU KANG    |          1996  |
| WOODLANDS        |          1996  |
| BUKIT PANJANG    |          1997  |
| SEMBAWANG        |          2001  |
| SENGKANG         |          2006  |
| PUNGGOL          |          2009  |
+------------------+----------------+
25 rows in set (0.00 sec)
```

Q4. What are the top 10 most expensive flats in Singapore?

Select the top 10 flats with the highest resale price in sample_data which displays a table of 10 flats with the highest resale prices.

SELECT * FROM sample_data ORDER BY Resale_Price DESC LIMIT 10;

```
+----------------+-----------+-------+--------------------+---------------+------------+------------+----------------+-------------------+--------------+
| Town           | Flat_Type | Block | Street             | Storey_Ranges | Floor_Area | Flat_Model | Lease_Commence | Remaining_Lease   | Resale_Price |
+----------------+-----------+-------+--------------------+---------------+------------+------------+----------------+-------------------+--------------+
| CENTRAL AREA   | 5 ROOM    | 1B    | CANTONMENT RD      | 49 TO 51      |        105 | Type S2    |           2011 | 93 years          |      1108000 |
| KALLANG/WHAMPOA| 5 ROOM    | 7     | BOON KENG RD       | 25 TO 27      |        119 | DBSS       |           2011 | 93 years          |      1005000 |
| TOA PAYOH      | EXECUTIVE | 129   | POTONG PASIR AVE 1 | 01 TO 03      |        161 | Maisonette |           1984 | 66 years 02 months|       905000 |
| CENTRAL AREA   | 4 ROOM    | 1F    | CANTONMENT RD      | 37 TO 39      |         95 | Type S1    |           2011 | 93 years          |       903000 |
| BISHAN         | EXECUTIVE | 134   | BISHAN ST 12       | 04 TO 06      |        143 | Maisonette |           1986 | 68 years 10 months|       895000 |
| BUKIT TIMAH    | 5 ROOM    | 5     | FARRER RD          | 10 TO 12      |        120 | Standard   |           1974 | 56 years 03 months|       890000 |
| BISHAN         | EXECUTIVE | 231   | BISHAN ST 23       | 04 TO 06      |        146 | Maisonette |           1992 | 74 years 04 months|       888000 |
| TOA PAYOH      | 5 ROOM    | 81    | LOR 4 TOA PAYOH    | 22 TO 24      |        122 | Improved   |           1997 | 79 years 02 months|       888000 |
| ANG MO KIO     | EXECUTIVE | 613   | ANG MO KIO AVE 4   | 07 TO 09      |        147 | Apartment  |           1996 | 78 years 04 months|       888000 |
| BUKIT TIMAH    | EXECUTIVE | 15    | TOH YI DR          | 01 TO 03      |        146 | Maisonette |           1988 | 70 years 06 months|       876888 |
+----------------+-----------+-------+--------------------+---------------+------------+------------+----------------+-------------------+--------------+
10 rows in set (0.00 sec)
```

Q.5 Find out which flat types are the most expensive

Select the different flat types and calculate the average of each flat types which displays a table ranking the flat types from most expensive to cheapest

SELECT ft.flat_type as flatType, AVG(rp.resale_price) as averagePrice from Flat_Types ft join Resales rp join Flat_Info fi WHERE ft.flat_type_id = fi.flat_type AND rp.resale_id = fi.resale GROUP BY flatType ORDER BY averagePrice DESC;

```
+-----------+---------------+
| flatType  | averagePrice  |
+-----------+---------------+
| EXECUTIVE |    622018.5205 |
| 5 ROOM    |    519163.9472 |
| 4 ROOM    |    434536.9277 |
| 3 ROOM    |    320546.6343 |
| 2 ROOM    |    242970.5882 |
+-----------+---------------+
5 rows in set (0.00 sec)
```

The web application

Homepage of the web application listing all the flats in the database

Resale Flats in Singapore

- Home
- Most Expensive Flat
- Flat Type Price
- Most Expensive Town
- Oldest to Newest Towns
- Storey Range Price

Town	Flat Type	Block	Street	Storey Ranges	Floor Area	Flat Model	Lease Commence	Remaining Lease	Resale Price
ANG MO KIO	2 ROOM	406	ANG MO KIO AVE 10	10 TO 12	44	Improved	1979	61 years 04 months	232000
ANG MO KIO	3 ROOM	108	ANG MO KIO AVE 4	01 TO 03	67	New Generation	1978	60 years 07 months	250000
ANG MO KIO	3 ROOM	602	ANG MO KIO AVE 5	01 TO 03	67	New Generation	1980	62 years 05 months	262000
ANG MO KIO	3 ROOM	465	ANG MO KIO AVE 10	04 TO 06	68	New Generation	1980	62 years 01 month	265000
ANG MO KIO	3 ROOM	601	ANG MO KIO AVE 5	01 TO 03	67	New Generation	1980	62 years 05 months	265000
ANG MO KIO	3 ROOM	150	ANG MO KIO AVE 5	01 TO 03	68	New Generation	1981	63 years	275000
ANG MO KIO	3 ROOM	447	ANG MO KIO AVE 10	04 TO 06	68	New Generation	1979	61 years 06 months	280000
ANG MO KIO	3 ROOM	218	ANG MO KIO AVE 1	04 TO 06	67	New Generation	1976	58 years 04 months	285000
ANG MO KIO	3 ROOM	447	ANG MO KIO AVE 10	04 TO 06	68	New Generation	1979	61 years 06 months	285000
ANG MO KIO	3 ROOM	571	ANG MO KIO AVE 3	01 TO 03	67	New Generation	1979	61 years 04 months	285000
ANG MO KIO	3 ROOM	534	ANG MO KIO AVE 10	01 TO 03	68	New Generation	1980	62 years 01 month	288500
ANG MO KIO	3 ROOM	233	ANG MO KIO AVE 3	10 TO 12	67	New Generation	1977	59 years 08 months	295000
ANG MO KIO	3 ROOM	235	ANG MO KIO AVE 3	04 TO 06	67	New Generation	1977	59 years 08 months	295000
ANG MO KIO	3 ROOM	219	ANG MO KIO AVE 1	07 TO 09	67	New Generation	1977	59 years 06 months	297000
ANG MO KIO	3 ROOM	536	ANG MO KIO AVE 10	07 TO 09	68	New Generation	1980	62 years 01 month	298000
ANG MO KIO	3 ROOM	230	ANG MO KIO AVE 3	04 TO 06	67	New Generation	1978	60 years	298000
ANG MO KIO	3 ROOM	570	ANG MO KIO AVE 3	10 TO 12	67	New Generation	1979	61 years 04 months	300000
ANG MO KIO	3 ROOM	624	ANG MO KIO AVE 4	04 TO 06	68	New Generation	1980	62 years 08 months	301000
ANG MO KIO	3 ROOM	441	ANG MO KIO AVE 10	07 TO 09	67	New Generation	1979	61 years	306000

Webpage of the 10 most expensive flats showing the top 10 most expensive flats in Singapore

Top 10 Most Expensive Resale Flats in Singapore

- Home
- Most Expensive Flat
- Flat Type Price
- Most Expensive Town
- Oldest to Newest Towns
- Storey Range Price

Town	Flat Type	Block	Street	Storey Ranges	Floor Area	Flat Model	Lease Commence	Remaining Lease	Resale Price
CENTRAL AREA	5 ROOM	1B	CANTONMENT RD	49 TO 51	105	Type S2	2011	93 years	1108000
KALLANG/WHAMPOA	5 ROOM	7	BOON KENG RD	25 TO 27	119	DBSS	2011	93 years	1005000
TOA PAYOH	EXECUTIVE	129	POTONG PASIR AVE 1	01 TO 03	161	Maisonette	1984	66 years 02 months	905000
CENTRAL AREA	4 ROOM	1F	CANTONMENT RD	37 TO 39	95	Type S1	2011	93 years	903000
BISHAN	EXECUTIVE	134	BISHAN ST 12	04 TO 06	143	Maisonette	1986	68 years 10 months	895000
BUKIT TIMAH	5 ROOM	5	FARRER RD	10 TO 12	120	Standard	1974	56 years 03 months	890000
BISHAN	EXECUTIVE	231	BISHAN ST 23	04 TO 06	146	Maisonette	1992	74 years 04 months	888000
TOA PAYOH	5 ROOM	81	LOR 4 TOA PAYOH	22 TO 24	122	Improved	1997	79 years 02 months	888000
ANG MO KIO	EXECUTIVE	613	ANG MO KIO AVE 4	07 TO 09	147	Apartment	1996	78 years 04 months	888000
BUKIT TIMAH	EXECUTIVE	15	TOH YI DR	01 TO 03	146	Maisonette	1988	70 years 06 months	876888

Webpage of the cost of every flat types

Average Prices of Flat Types

Flat Type	Average Flat Type Price
EXECUTIVE	622018.5205
5 ROOM	519163.9472
4 ROOM	434536.9277
3 ROOM	320546.6343
2 ROOM	242970.5882

Webpage listing the price of flats in towns from most expensive to cheapest

Expensive Towns in Singapore

Town Name	Average Flat Price
BUKIT TIMAH	812296
BISHAN	604892.1818
CENTRAL AREA	594583.3333
BUKIT MERAH	512267.8571
QUEENSTOWN	479381.7143
MARINE PARADE	474500
KALLANG/WHAMPOA	473707.6216
PUNGGOL	469814.9474
SERANGOON	466450
PASIR RIS	453576
TAMPINES	449650.7805
TOA PAYOH	446057.2121
SENGKANG	421884.7407
CLEMENTI	420814.2857
HOUGANG	416688.8485
ANG MO KIO	413971.2143
SEMBAWANG	410043.4783
BEDOK	408598.4
BUKIT PANJANG	408351.1111
JURONG WEST	402120.2143
CHOA CHU KANG	392424.7273
WOODLANDS	390294.7349
JURONG EAST	389646.5556
BUKIT BATOK	373450.8085
GEYLANG	370543.375

Webpage listing the towns with the oldest and newest flats

Oldest to Newest Town in Singapore

Town Name	Average Lease Commence Year
MARINE PARADE	1976
QUEENSTOWN	1979
GEYLANG	1979
TOA PAYOH	1983
CLEMENTI	1983
ANG MO KIO	1983
BEDOK	1984
BUKIT TIMAH	1984
BUKIT MERAH	1986
SERANGOON	1987
KALLANG/WHAMPOA	1987
JURONG EAST	1988
BUKIT BATOK	1989
BISHAN	1989
HOUGANG	1991
TAMPINES	1991
CENTRAL AREA	1992
PASIR RIS	1993
JURONG WEST	1995
CHOA CHU KANG	1996
WOODLANDS	1996
BUKIT PANJANG	1997
SEMBAWANG	2001
SENGKANG	2006
PUNGGOL	2009

Webpage of the prices of the storey ranges.

Storey Range Prices in Singapore From most expensive to cheapest

Storey Range	Price
49 TO 51	1108000
34 TO 36	808000
37 TO 39	801500
28 TO 30	733600
22 TO 24	663722
25 TO 27	640000
19 TO 21	549437.5
16 TO 18	489799
13 TO 15	476320.5833
10 TO 12	440088.037
07 TO 09	414696.2888
04 TO 06	408646.5034
01 TO 03	397525.401

References and Acknowledgement

[1] https://data.gov.sg/dataset/resale-flat-prices, https://data.gov.sg/open-data-licence

I acknowledge the data I used is made available under the terms of the Singapore Open Data Licence version 1.0 {https://data.gov.sg/open-data-licence}

Link to coursera web application page

https://hub.labs.coursera.org:443/connect/sharedvpghxxlo?forceRefresh=false&path=%2F%3Ffolder%3D%2Fhome%2Fcoder%2Fproject