

# MASTER OF TECHNOLOGY (IS) PROJECT REPORT

---

## HDB RESALE PREDICTION AND RECOMMENDER

---

### **TEAM MEMBERS**

(A0045422R) YEE WEI LIANG

(A0229968E) TOH KAH KHEK

(A0133374J) JEON SUNG MIN

(A0229978A) AHMED SYALABI SEET

# Table of Contents

1. EXECUTIVE SUMMARY .....	3
2. BUSINESS PROBLEM DESCRIPTION .....	4
2.1 PROJECT OBJECTIVE .....	4
2.2 MARKET RESEARCH .....	4
3. KNOWLEDGE MODELING .....	6
3.1 KNOWLEDGE IDENTIFICATION .....	6
3.2 KNOWLEDGE SPECIFICATION .....	8
3.3 KNOWLEDGE REFINEMENT .....	9
4. SOLUTION .....	17
4.1. SYSTEM IMPLEMENTATION .....	17
4.2 SYSTEM PERFORMANCE & VALIDATION .....	20
5. CONCLUSION .....	23
5.1 PROJECT CONCLUSION .....	23
5.2 IMPROVEMENTS .....	23
6. APPENDIX .....	24
6.1 PROJECT PROPOSAL .....	24
6.2 MAPPING OF DOMAIN/TECHNIQUE GROUP .....	29
6.3 INSTALLATION AND USER GUIDE .....	30
6.4 INDIVIDUAL PROJECT REPORTS .....	41
6.5 LIST OF ABBREVIATIONS .....	45
6.6 REFERENCES .....	45

---

## 1. EXECUTIVE SUMMARY

---

*“HDB resale prices climb for 4th consecutive quarter in Q1”, “HDB resale prices rise for 9th straight month in March with cash over valuation back in play”, “HDB resale transactions spike 127.3% in Q3 after end of COVID-19 circuit breaker”.* These are headlines that are getting more common in recent days regarding resale HDB flats in Singapore.

Despite a circuit breaker introduced by government on Apr 7 to curb COVID-19, resale flat prices seemingly were spared the adverse gloom that impacted the economic environment in Singapore. This “stunning rebound” could be attributed to various stimulus packages that were introduced by the government prior to the start of the pandemic.

However, the incessant increase of resale HDB prices certainly make it challenging for prospective homebuyers. While BTO flats are an alternative, the long wait for getting the keys, coupled with construction delays, often are out of consideration for some buyers.

Hence, using the valuable knowledge we have learnt during our lectures in Intelligent Reasoning Systems, our group would like to implement a **HDB Resale Prediction and Recommender (HRPR)** tool to assist prospective buyers in their resale flat search.

For price prediction, we have used past resale flat transactions since 2015 (corrected with Resale Price Index). On the other hand, we also collected survey results from friends, families, and coursemates from various walks of life to get a general sentiment of important features to a flat buyer. The survey results were then used to implement a scoring system for our recommender system.

In summary, we hope that with HRPR, prospective buyers can estimate the suitable pricing to negotiate with sellers. Besides that, the recommender system of the tool can suggest potential listings that the users might overlook, especially swarmed by pages of listings that were refreshed daily.

## 2. BUSINESS PROBLEM DESCRIPTION

### 2.1 PROJECT OBJECTIVE

According to Straits Times article published on 8 Apr 2021, delays in Build-to-Order (BTO) flats, rising private property prices and improving market sentiment drove demand in the Housing Board resale market, sending prices climbing for the ninth straight month.<sup>10</sup>

Armed with limited knowledge, aspiring homeowners often must second guess the suitable price to pay for a resale flat. For example, when a listing is listed at a lower price than what the past transactions or neighbouring listings are showing, it could be due to undesirable conditions. On the other hand, if the pricing is higher than valuation, prospective buyer ends up paying more.

Hence, HRPR would like to provide prospective resale flat buyers with a tool to predict the pricing to negotiate with resale flat sellers, with existing domain knowledge obtained through Data.gov. On top of that, HRPR also comes with Recommender System which recommend listings that they may consider in their purchase journey.

### 2.2 MARKET RESEARCH

Prior to kickstart this project, we have done a survey of similar solutions related to housing price prediction, particularly for Singapore HDB flats.

From similar projects in ISS, there is an existing HDB BTO Recommender System. While the system has its strength derived from rules-based reasoning, the system is limited to 2017 BTO launch and require significant effort to scale to resale flats and other launches.

On the other hand, commercial solutions include the one provided by PropertyGuru, 99.co, UrbanZoom etc. However, limitations with current system include lack of Price Prediction (most of them only provide past transaction price trends) and non-functional search (Urbanzoom).

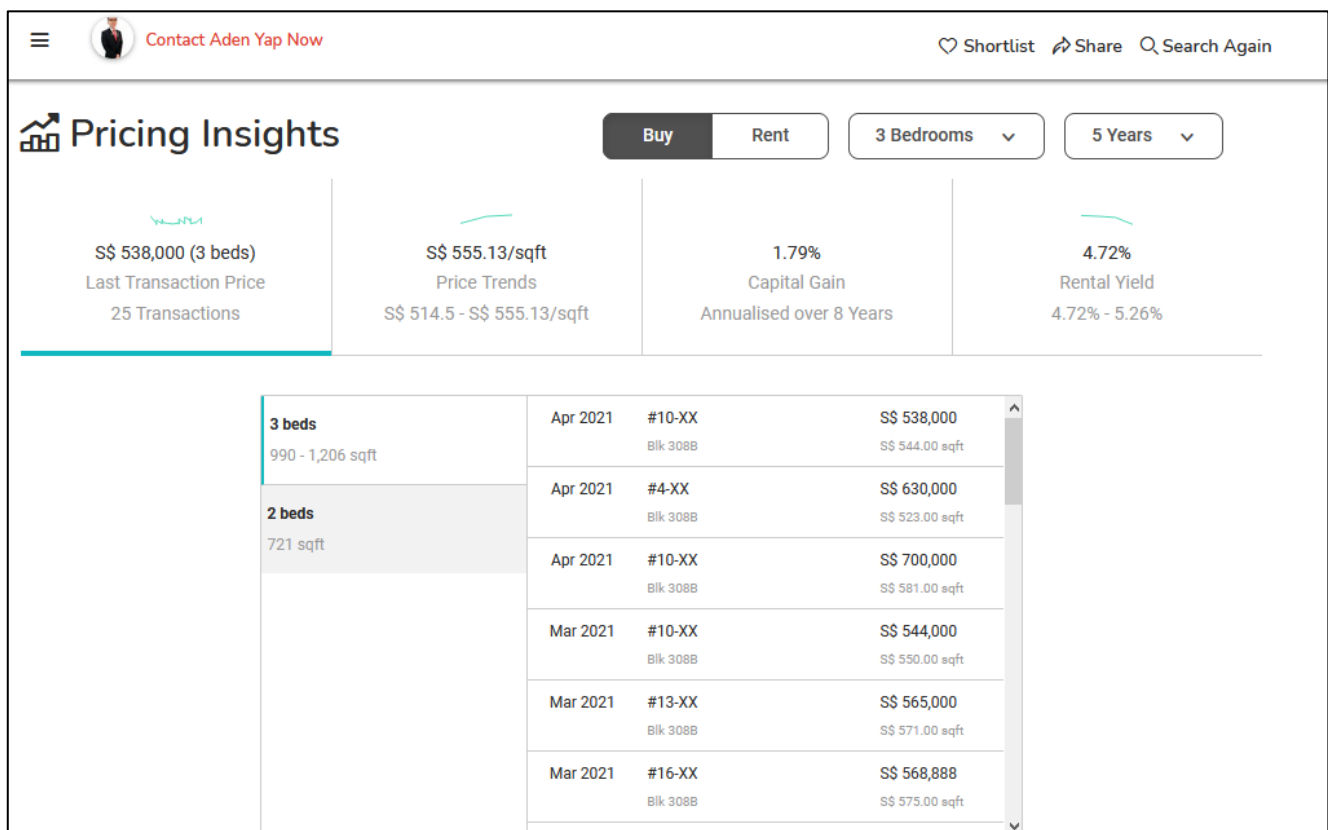


Figure 1. PropertyGuru Pricing Insights based on past transactions

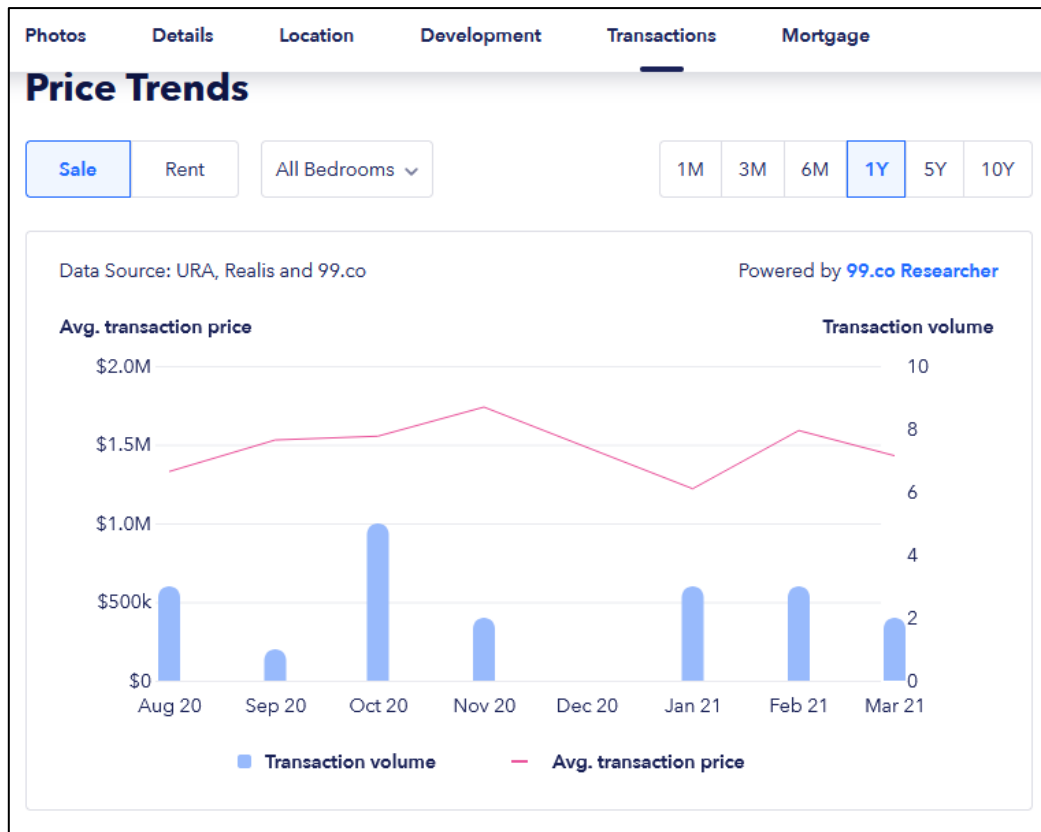


Figure 2. 99.co Price Trends based on past transactions

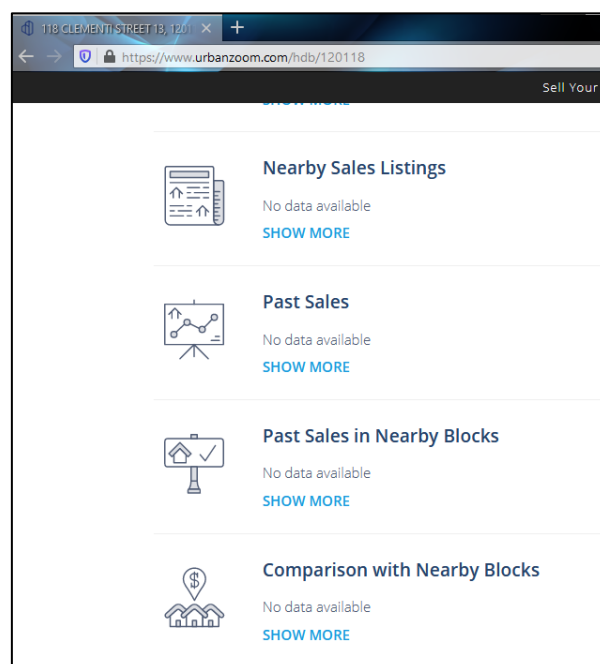


Figure 3. Urban Zoom search is missing data

### 3. KNOWLEDGE MODELING

Knowledge modelling can be decomposed into three main stages, namely:

- (i) Knowledge identification
- (ii) Knowledge specification
- (iii) Knowledge refinement

Various activities are carried out during each of these stages and the crux of the model construction lies in stage (ii), Knowledge specification.

#### 3.1 KNOWLEDGE IDENTIFICATION

Knowledge Identification is a preparatory phase, where information sources are identified and its extraction techniques are explored and validated, for Knowledge Specification in stage 2.

S/N	Source of information	Insights from information source	Knowledge acquisition technique
1	Data.gov.sg	<ul style="list-style-type: none"> <li>Historical resale flat prices from 1990 to current.</li> <li>Acts as training data for the prediction model.</li> </ul>	Extracted using Data.gov.sg API
2	PropertyGuru.com.sg, Real estate agent	<ul style="list-style-type: none"> <li>Individual property listings posted by property agents, inclusive of features such as listing price, type of flat model, location, etc</li> <li>Provides listing prices available for comparison with prediction values</li> </ul>	Web scrapping to obtain actual listings of HDB resale units available to the public
3	Generic population	<ul style="list-style-type: none"> <li>Obtain a ranking of the features that are most important to users for the recommender system.</li> </ul>	Elicitation of tacit knowledge extraction from anonymous online survey

Table 1: Knowledge source and acquisition techniques

We designed an online survey where 74 respondents took part to determine the ranking of 5 of the most relevant features when looking for a resale flat.

**Rank of Importance when Purchasing Resale Flat**

Please rank the following considerations when you look to purchase a resale HDB flat  
(1st choice : Most Important; 5th: Least Important)

\* Required

Rank (1st: Most Important ; 5th: Least Important) \*

	Ease of Access to LRT/MRT Station	Distance to Mall	Distance to Hawker Centre	Distance to City Centre (Orchard)	Age of Flat
1st Choice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2nd Choice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3rd Choice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4th Choice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5th Choice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Submit**

Figure 4 Google Form survey on feature importance

The results were then analysed and used to obtain feature weights for our scoring metric below.

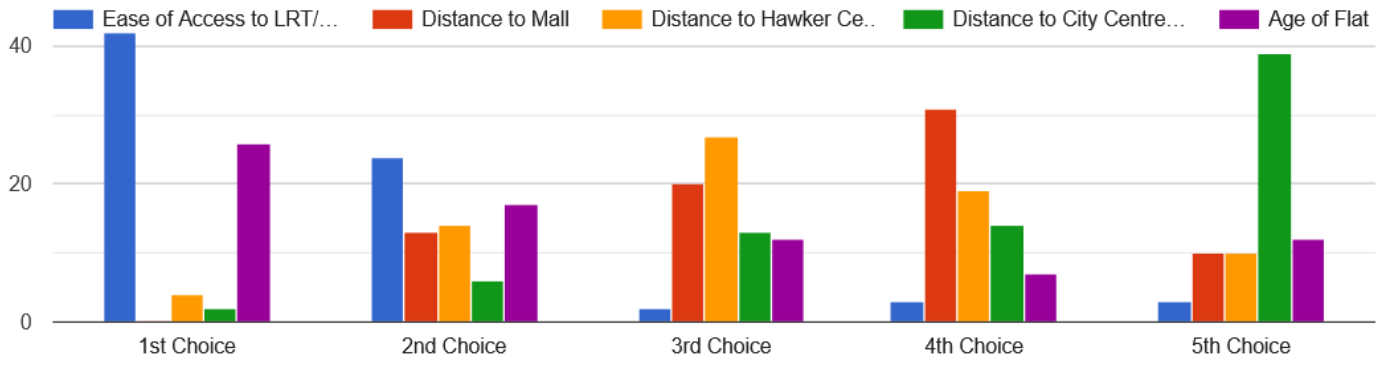


Figure 5: Survey responses on feature importance when scouting for a resale flat

### Recommender scoring metric

$$\sum_{i=1}^{n=5} m_i x_i = m_1 x_1 + \dots + m_n x_n$$

where  $m_i$  is the feature weights derived from the survey and  $x_i$  is the normalized feature value.

### 3.2 KNOWLEDGE SPECIFICATION

After obtaining the knowledge from the first stage, we would require to map out the inference approach to achieve the intent of the knowledge model. Below diagram documents the data flow from knowledge base to inference and the data specified for each task.

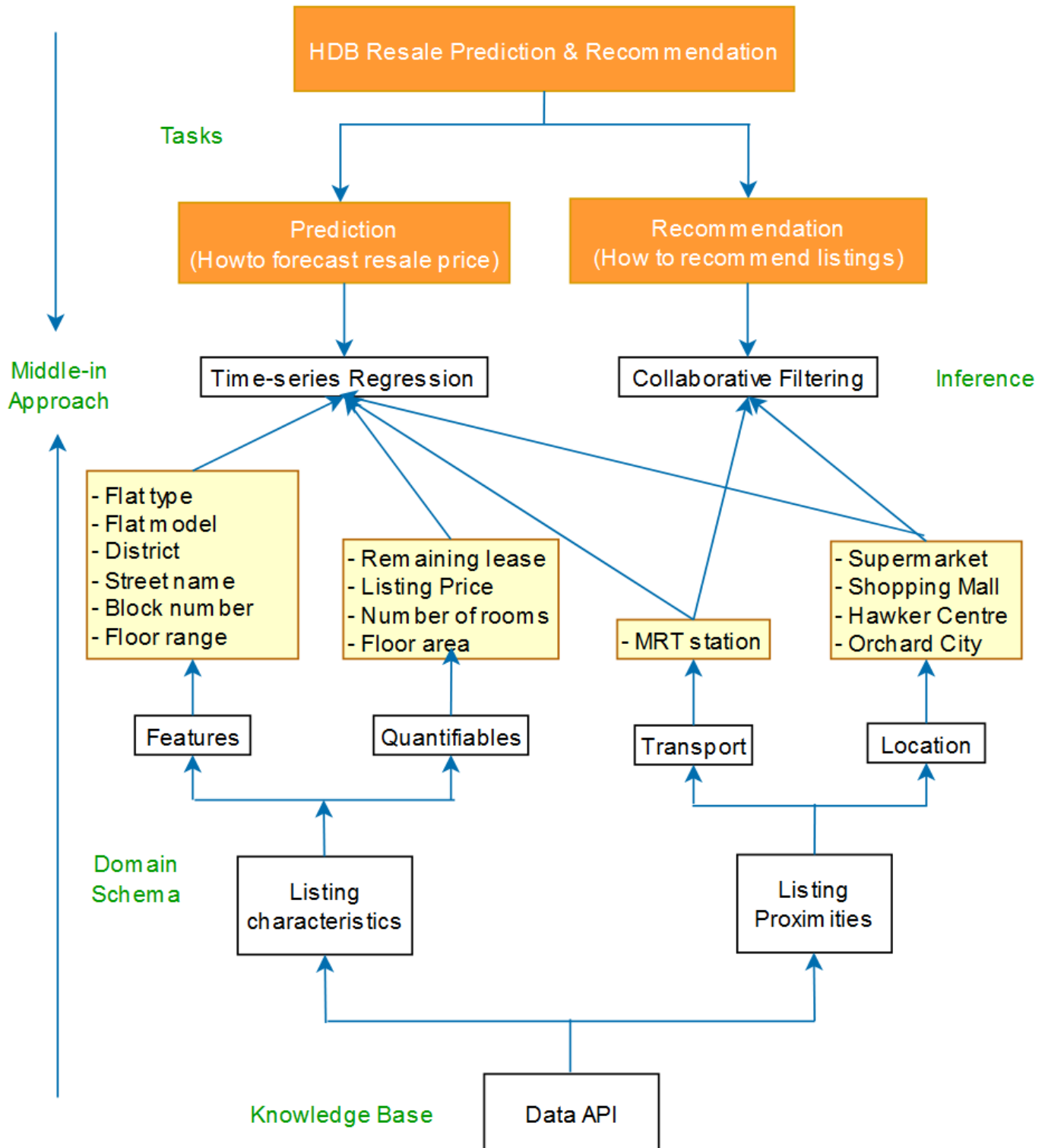


Figure 6: Inference structure



### 3.3 KNOWLEDGE REFINEMENT

In the third and final stage, a thorough refinement of the information is required to be done before ingesting into our models. The domain schema is crucial for our inference models as it must be perfectly scaled and aligned together.

Below are snippets from the notebook used for exploratory data analysis.

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price	remaining_lease
0	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	10 TO 12	31.0	IMPROVED	1977	9000.0	NaN
1	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	04 TO 06	31.0	IMPROVED	1977	6000.0	NaN
2	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	10 TO 12	31.0	IMPROVED	1977	8000.0	NaN
3	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	07 TO 09	31.0	IMPROVED	1977	6000.0	NaN
4	1990-01	ANG MO KIO	3 ROOM	216	ANG MO KIO AVE 1	04 TO 06	73.0	NEW GENERATION	1976	47200.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...
839444	2014-12	YISHUN	5 ROOM	816	YISHUN ST 81	10 TO 12	122.0	Improved	1988	580000.0	NaN
839445	2014-12	YISHUN	EXECUTIVE	325	YISHUN CTRL	10 TO 12	146.0	Maisonette	1988	540000.0	NaN
839446	2014-12	YISHUN	EXECUTIVE	618	YISHUN RING RD	07 TO 09	164.0	Apartment	1992	738000.0	NaN
839447	2014-12	YISHUN	EXECUTIVE	277	YISHUN ST 22	07 TO 09	152.0	Maisonette	1985	592000.0	NaN
839448	2014-12	YISHUN	EXECUTIVE	277	YISHUN ST 22	04 TO 06	146.0	Maisonette	1985	545000.0	NaN

839449 rows × 11 columns

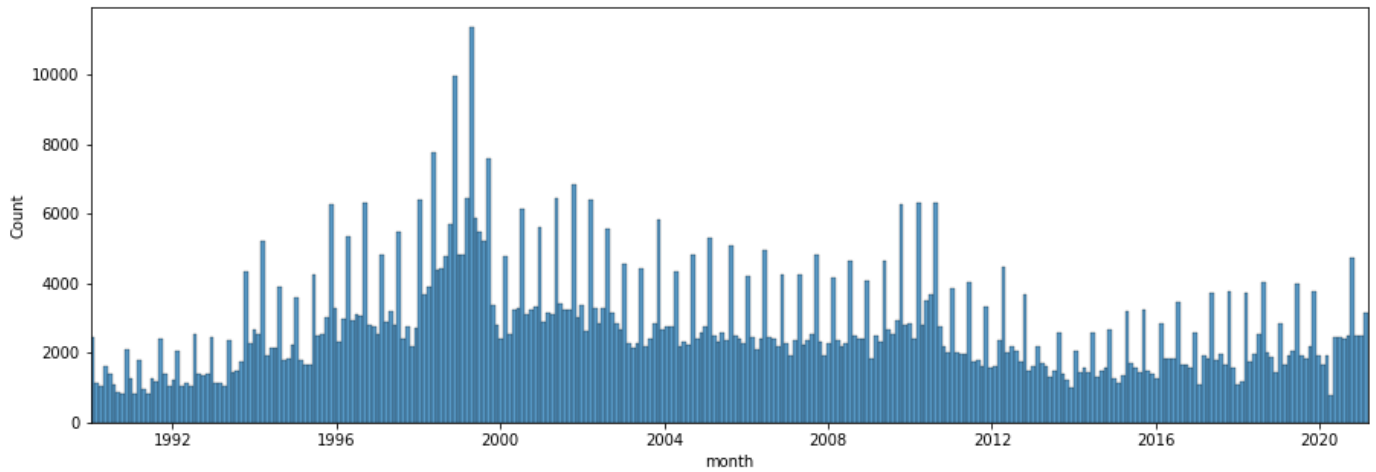
Figure 7: Raw data from data.gov.sg

#### Month & Yearly Quarters

Converting `month` into datetime format. `quarters` required for resale price RPI deflation.

```
data['month'] = pd.to_datetime(data['month'])
data = data.sort_values(by='month').reset_index(drop=True)

# Quarters
def quarters(x):
    if x.month in [1,2,3]:
        return "Q1"
    elif x.month in [4,5,6]:
        return "Q2"
    elif x.month in [7,8,9]:
        return "Q3"
    elif x.month in [10,11,12]:
        return "Q4"
```

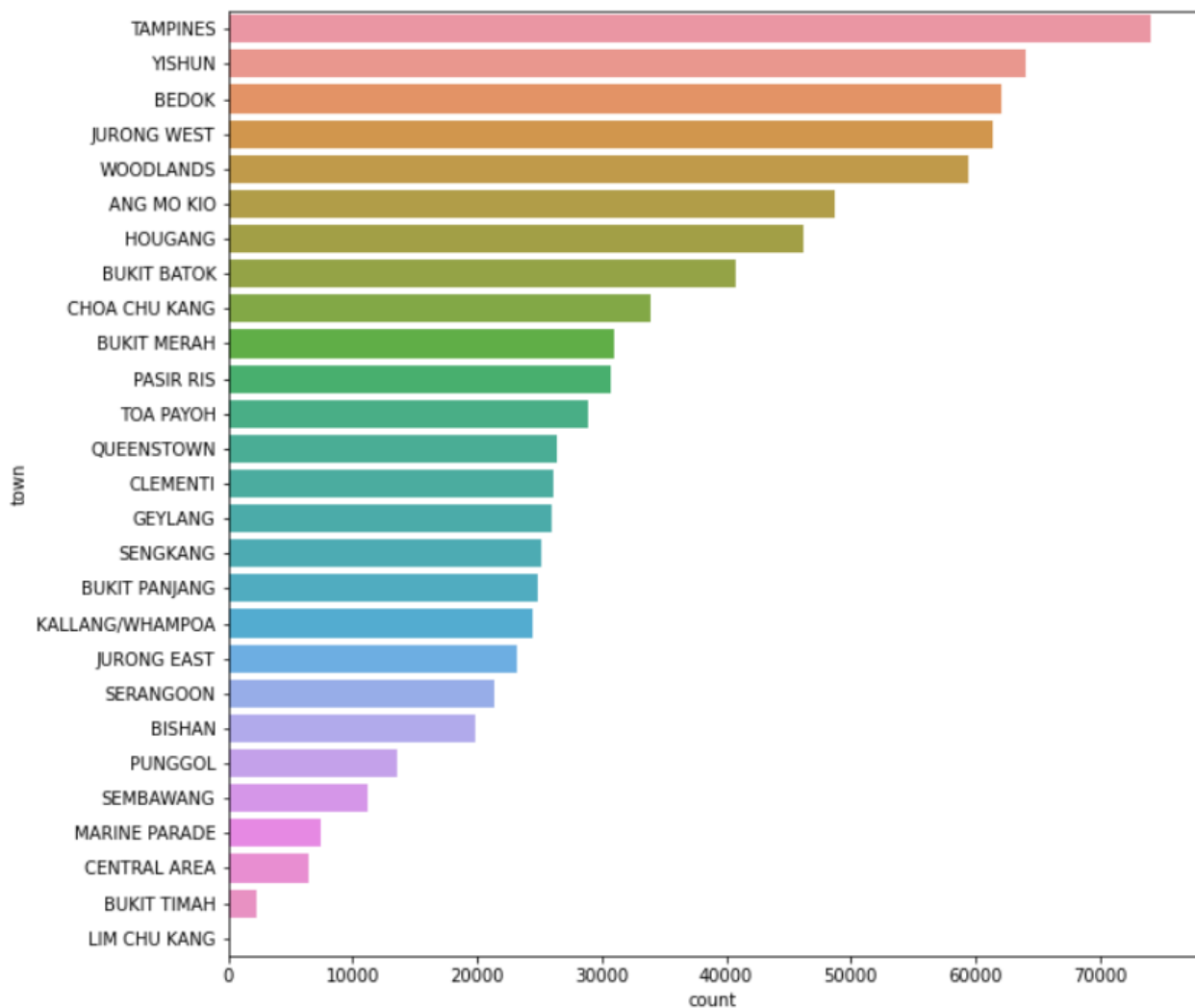


### Town

Town feature will be mapped into `postal_districts` using the link below

This is to align with data coming from PropertyGuru.

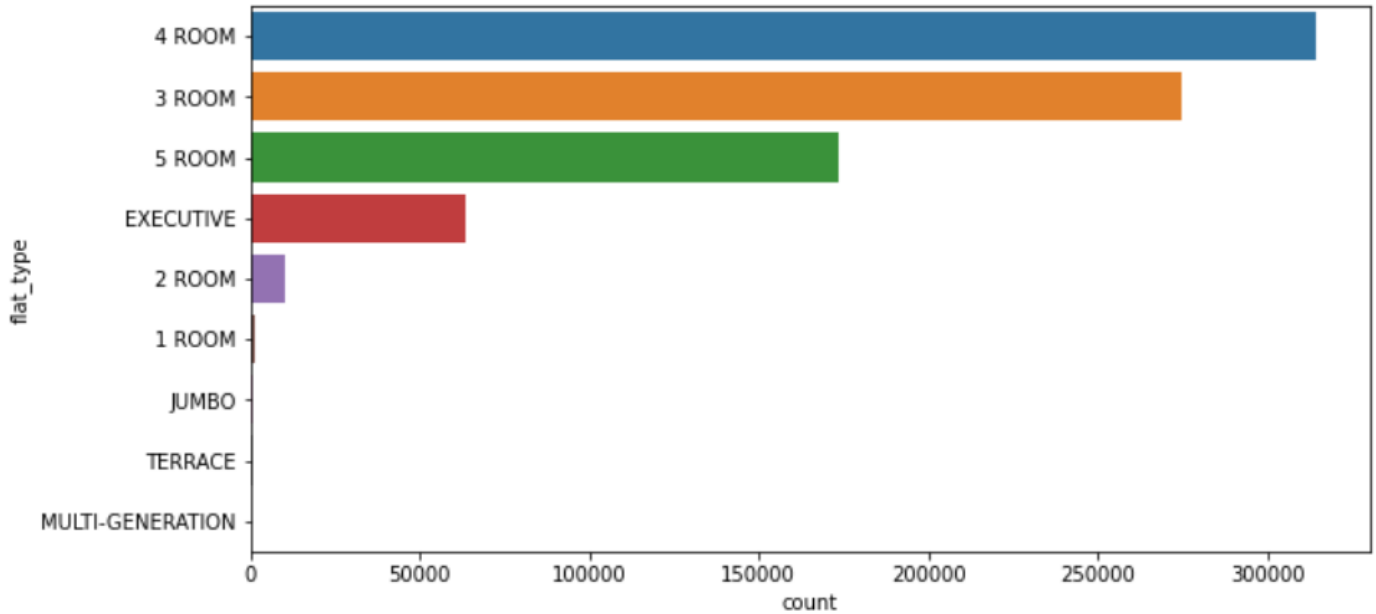
[https://www.ura.gov.sg/realEstateIIWeb/resources/misc/list\\_of\\_postal\\_districts.htm](https://www.ura.gov.sg/realEstateIIWeb/resources/misc/list_of_postal_districts.htm)



**Flat type & model**

`flat_model` was feature engineered together with `flat_type` to extract Jumbo and terrace flats.

```
data['flat_model'] = data['flat_model'].str.upper()
data['flat_type'] = data['flat_type'].str.replace(
    "MULTI GENERATION", "MULTI-GENERATION")
data['flat_type'] = np.where(
    data['flat_model']=='TERRACE', data['flat_model'], data['flat_type'])
data['flat_type'] = np.where(
    data['flat_model']=='ADJOINED FLAT', 'JUMBO', data['flat_type'])
```

**Full Address**

Sample: "837 Hougang Central, Singapore 530837"

`block` and `street_name` were merged to get `address`.

It will later be used for reverse geocoding via OneMapSG API.

```
data['address'] = data['block'] + " " + data['street_name']
data = data.reset_index(drop=True)
```

	block	street_name	address
0	309	ANG MO KIO AVE 1	309 ANG MO KIO AVE 1
1	44	BENDEMEER RD	44 BENDEMEER RD
2	20	ST. GEORGE'S RD	20 ST. GEORGE'S RD
3	14	KG ARANG RD	14 KG ARANG RD
4	46	OWEN RD	46 OWEN RD
...	...	...	...
839444	476A	UPP SERANGOON VIEW	476A UPP SERANGOON VIEW
839445	987C	BUANGKOK GREEN	987C BUANGKOK GREEN
839446	984A	BUANGKOK LINK	984A BUANGKOK LINK
839447	451	HOUGANG AVE 10	451 HOUGANG AVE 10
839448	666B	PUNGGOL DR	666B PUNGGOL DR

839449 rows × 3 columns

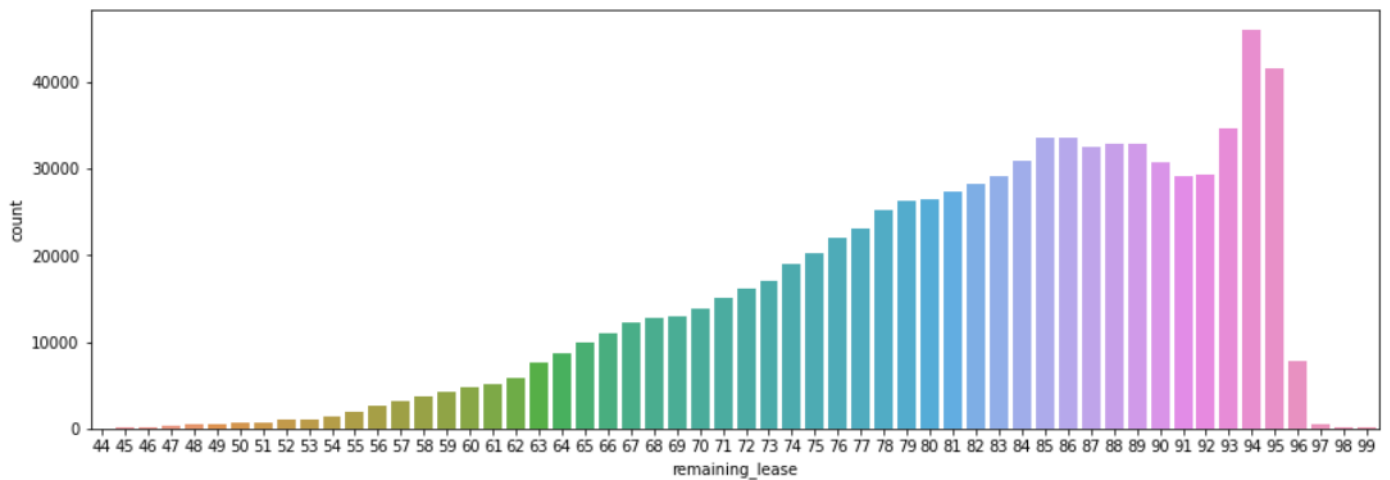
## Remaining Lease

Due to missing values from mainly the older flats, imputation will be required.

```
data['remaining_lease'] = data['remaining_lease'].fillna(0)

def remaining_lease(row):
    """Take in row index and outputs remaining_lease"""
    lease_year = data['lease_commence_date'][row]
    transaction_year = data['month'][row].year
    if data['remaining_lease'][row] == 0:
        remaining_lease = int(99 - (transaction_year - lease_year))
        # some listing date before lease commence date,
        # remaining_lease was showing more than 99 years
        if remaining_lease > 99:
            return 99
        else:
            return remaining_lease
    else:
        remaining_lease = int(data['remaining_lease'][row].split(' ')[0])
        if remaining_lease > 99:
            return 99
        else:
            return remaining_lease

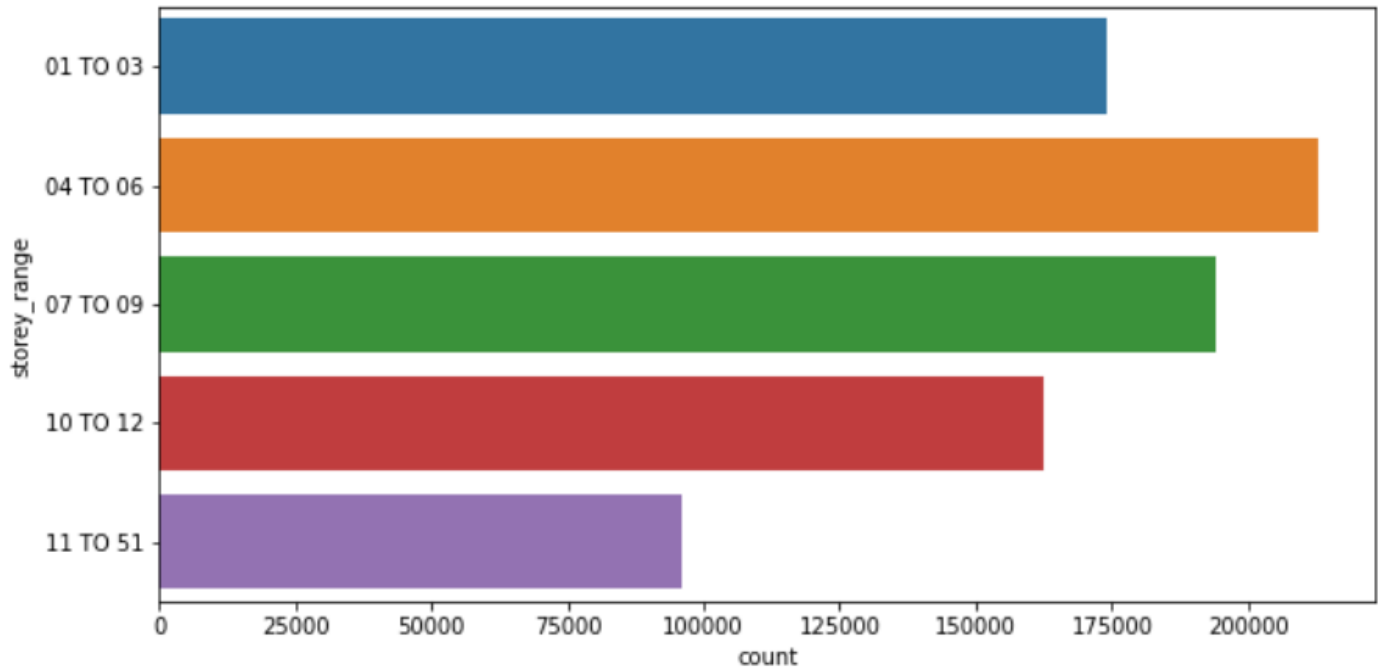
data['remaining_lease'] = data.reset_index()['index'].apply(remaining_lease)
```



**Storey Range**

Further discretize into 5 bins.

```
def storey_map(x):  
    if x == '01 TO 05': return '01 TO 03'  
    if x == '06 TO 10': return '07 TO 09'  
    if x in ['01 TO 03', '04 TO 06', '07 TO 09', '10 TO 12']:  
        return x  
    else:  
        return '11 TO 51'
```



**Resale Price**

Resale Price Index: <https://data.gov.sg/dataset/hdb-resale-price-index>

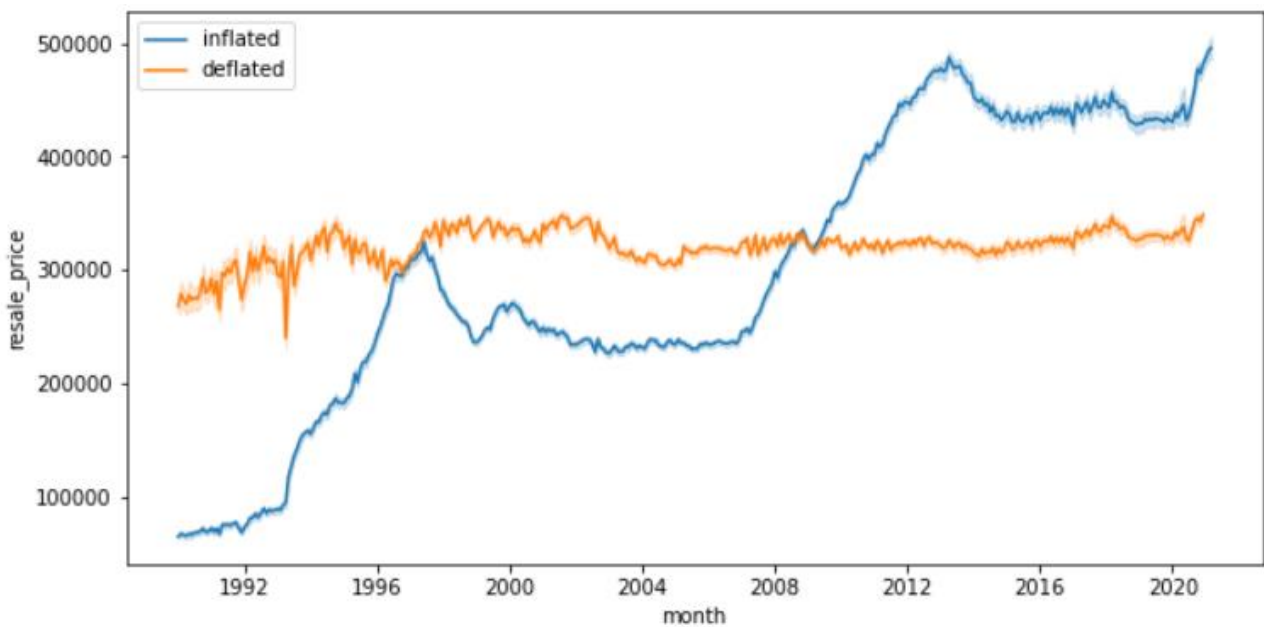
As the resale price spans from 1990 to 2021, it is affected by time therefore it has to be deflated to maintain linearity.

- $x_o$  - resale price at base year, 2019
- $x_i$  - resale price at year of sale
- $C_i$  - RPI at year of sale
- $C_o$  - RPI at year at base year, 2019

$$x_o = x_i \cdot \frac{C_o}{C_i}$$

```
def deflation_fn(row):
    while True:
        try:
            x_i = data['resale_price'][row]
            c_o = 100.0
            c_i = cpi[
                (cpi['year'] == data['month'][row].year) &
                (cpi['quarter'] == data['quarter'][row])
            ]['index'].item()
            return x_i * (c_o/c_i)
        except ValueError:
            return np.nan

data['corrected_resale_price'] = data['resale_price'] \
    .reset_index()['index'].apply(deflation_fn)
```



**Reverse Geocoding**

Obtain coordinates for all flats and amenities.

```
def regex_func(x):
    regex = r"[A-Z]{1,}[^A-Z]{2,}"
    regex1 = r".+?(Mall|mall){1}"
    if re.match(regex, x):
        return " ".join(re.findall(regex, x))
    else:
        return x

def latitude(x):
    try:
        return (Client.search(x) ['results']).pop() ['LATITUDE']
    except:
        return np.nan

def longitude(x):
    try:
        return (Client.search(x) ['results']).pop() ['LONGITUDE']
    except:
        return np.nan

def full_address(x):
    try:
        return (Client.search(x) ['results']).pop() ['ADDRESS']
    except:
        return np.nan
```

**Distance from nearest Amenities**

Obtain great arc distance between flats and nearest amenities using haversine function.

```
class nearest_Mall():
    def location(p1):
        nearest_y = 1000
        p3 = 0
        distances = []
        for i, p2 in zip(malls['Shopping_Malls'], malls['coordinates']):
            y = haversine(p1,p2) # distance
            distances.append(y)
            if y < nearest_y:
                nearest_y = y
                p3 = i
        return str(p3)

    def distance(p1):
        nearest_y = 1000
        p3 = 0
        distances = []
        for i, p2 in zip(malls['Shopping_Malls'], malls['coordinates']):
            y = haversine(p1,p2) # distance
            distances.append(y)
            if y < nearest_y:
                nearest_y = y
                p3 = i
        return float(nearest_y)
```

## Outlier Detection

```
outliers = []

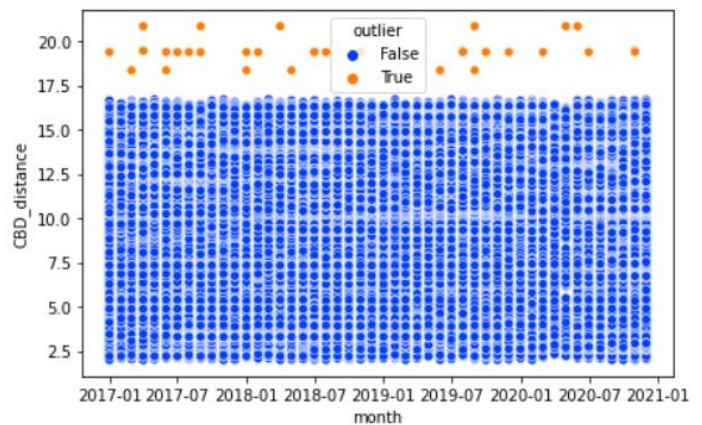
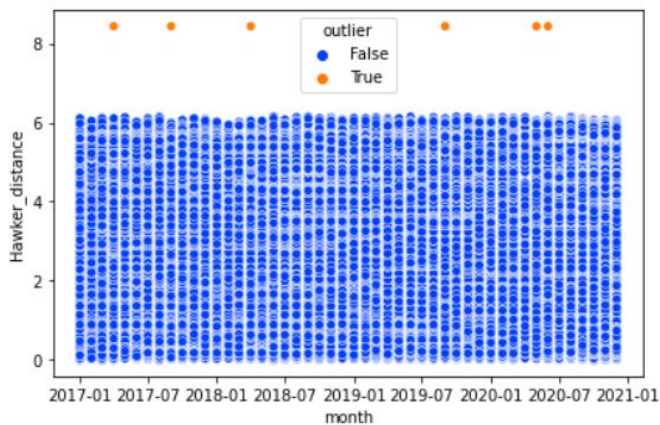
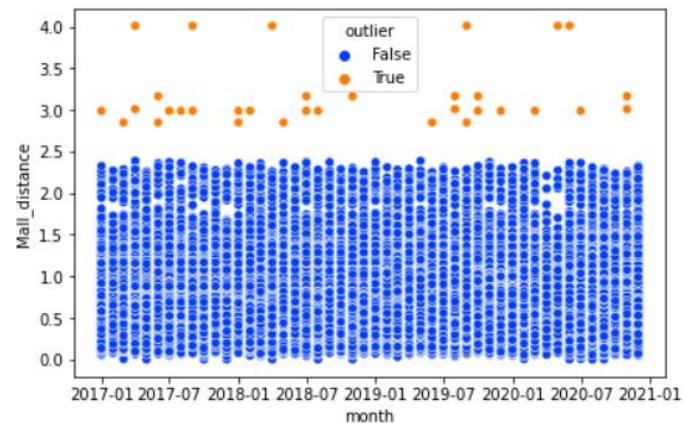
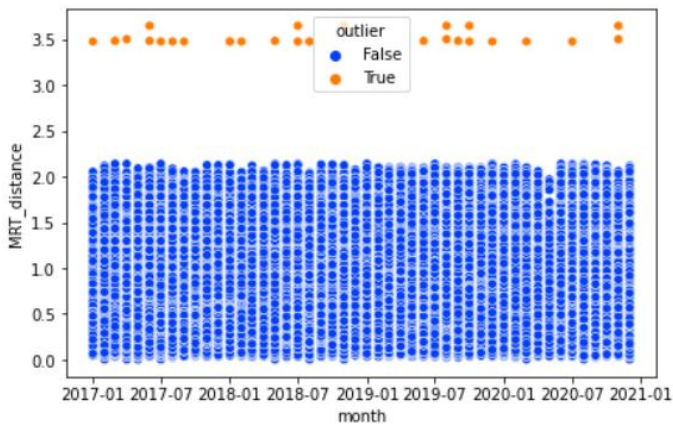
plt.figure(figsize=(15,20))
for i, feature in enumerate([
    'floor_area_sqm', 'remaining_lease', 'MRT_distance',
    'Mall_distance', 'Hawker_distance', 'CBD_distance',
    'target']):

    data = df[['month', feature]]
    if feature=='floor_area_sqm':
        data['outlier'] = data[feature]>200
    elif feature=='MRT_distance':
        data['outlier'] = data[feature]>2.5
    elif feature=='Mall_distance':
        data['outlier'] = data[feature]>2.5
    elif feature=='Hawker_distance':
        data['outlier'] = data[feature]>7
    elif feature=='CBD_distance':
        data['outlier'] = data[feature]>17.5
    else:
        data['outlier'] = False

    for outlier in data[data['outlier']==True].index:
        outliers.append(outlier)
```

```
outliers = set(outliers)
df = df[~df.index.isin(outliers)]
print('Number of outliers removed:', len(outliers))
```

Number of outliers removed: 39





## 4. SOLUTION

In the previous section, the knowledge models that are used by the application for making predictions and recommendations are key functions designed to provide HDB buyers with more information when they are looking to purchase a resale HDB flat. This section will provide further information on the overall design and architecture of the application.

### 4.1. SYSTEM IMPLEMENTATION

The HRPR application backend is developed with Flask, a web framework used to develop web applications in Python. The web frontend is developed using HTML, CSS and Javascript with the help of Bootstrap, a CSS framework. Transaction records from Data.gov.sg, listings of PropertyGuru and information of amenities are stored in the tables of SQLite, a relational database management system. The application's system architecture is illustrated in Figure 8. The list of key functions developed in the backend and its definition can be found in Table 2.

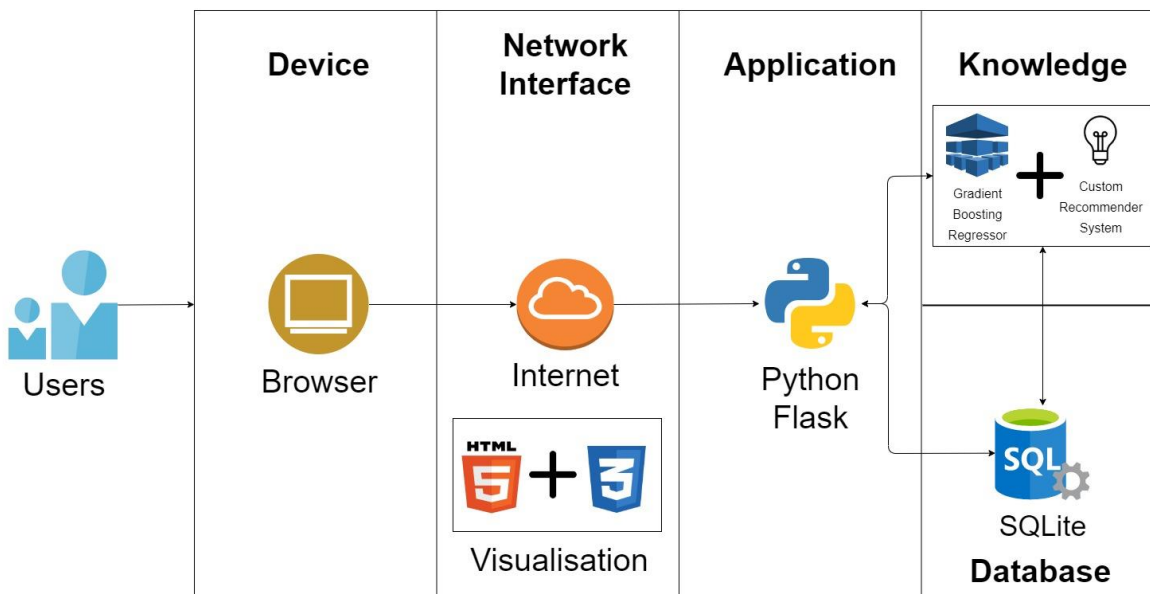


Figure 8: HRPR System

Function Name	Function URL	Definition
update_datagov	/update/datagov	<p>If the response from the form to update Data Gov Table is “Yes”, the update_datagov_table function in the functions.py file will be executed.</p> <p>If the response from the form to update Data Gov Table is “No”, the user will be redirected back to the homepage.</p>
update_propguru	/update/propguru	<p>If the response from the form to update Propertyguru Table is “Yes”, the scrapeType function in the PropertyGuruRetriever.py file will be executed.</p> <p>If the response from the form to update Propertyguru Table is “No”, the user will be redirected back to the homepage.</p>
train_model	/update/trainmodel	<p>If the response from the form to update Training Model is “Yes”, the train_regression_model function in the functions.py file will be executed.</p> <p>If the response from the form to update Training Model is “No”, the user will be redirected back to the homepage.</p>
update_amenities	/update/amenities	<p>If the response from the form to update Amenities Table is “Yes”, the following functions in the functions.py file will be executed.</p> <ul style="list-style-type: none"> <li>• insert_railtransit_data</li> <li>• insert_shoppingmalls_data</li> <li>• insert_hawkercentre_data</li> <li>• insert_supermarket_data</li> </ul> <p>If the response from the form to update Amenities Table is “No”, the user will be redirected back to the homepage.</p>
update_datagov_table	Executed from update_datagov	<ol style="list-style-type: none"> <li>1. Retrieve the HDB resale transaction records and the HDB Resale Price Index (RPI) using data.gov API.</li> <li>2. Check for the latest quarter of the RPI available and the month of the last record in the database. If database is updated with records until the latest quarter, no update is required.</li> <li>3. If there are records to be updated, get the features required for model learning from the API. Additional features such as latitude, longitude, distances to amenities are created from the following functions in addfeatureslib.py using the address from individual resale transaction record. <ul style="list-style-type: none"> <li>• geographic_position</li> <li>• get_nearest_railtransit</li> <li>• get_nearest_shoppingmall</li> <li>• get_orchard_distance</li> <li>• get_nearest_hawkercentre</li> <li>• get_nearest_supermarket</li> </ul> </li> <li>4. The postal district, storey range and flat type are then mapped according to each transaction record.</li> <li>5. Records appended to the table in the database.</li> </ol>

scrapeType	Executed from update_propguru	<ol style="list-style-type: none"> <li>1. Scrape Propertyguru for features required for recommender system from listings of different HDB resale properties using Selenium.</li> <li>2. Validate information of listings before adding to the list of records to be added to the database.</li> <li>3. Additional features such as latitude, longitude, distances to amenities are created from the following functions in addfeatureslib.py using the address from individual listing. <ul style="list-style-type: none"> <li>• geographic_position</li> <li>• get_nearest_railtransit</li> <li>• get_nearest_shoppingmall</li> <li>• get_orchard_distance</li> <li>• get_nearest_hawkercentre</li> <li>• get_nearest_supermarket</li> </ul> </li> <li>4. Remove duplicate listings and listing without valid postal district.</li> <li>5. Delete existing listings in the table and update table with new listings.</li> </ol>
train_regression_model	Executed from train_model	<ol style="list-style-type: none"> <li>1. Query from table in database for HDB resale transaction records with created features.</li> <li>2. Remove outliers and data without valid postal district from the dataset.</li> <li>3. Preprocess the dataset by normalising continuous data and encoding categorical data.</li> <li>4. Split the dataset into training data (80%) and test data (20%).</li> <li>5. Use scaler to fit the training and use Gradient Boosting Regressor as the model.</li> <li>6. Use scaler to transform the test data and perform predictions on the test data.</li> <li>7. Evaluate accuracy of the model using Root Mean Square Error (RMSE).</li> <li>8. Save the scaler and model using joblib.</li> </ol>
<ol style="list-style-type: none"> <li>1. scrapeSearchListing</li> <li>2. load_regression_model</li> <li>3. get_history_transactions</li> <li>4. recommender_system</li> </ol>	Executed when user provides valid Propertyguru URL and clicks on "Search"	<ol style="list-style-type: none"> <li>1. Validate and scrape Propertyguru URL provided by user for features required for prediction and recommender system</li> <li>2. Additional features such as latitude, longitude, distances to amenities are created from the following functions in addfeatureslib.py using the address from individual listing. <ul style="list-style-type: none"> <li>• geographic_position</li> <li>• get_nearest_railtransit</li> <li>• get_nearest_shoppingmall</li> <li>• get_orchard_distance</li> <li>• get_nearest_hawkercentre</li> <li>• get_nearest_supermarket</li> </ul> </li> <li>3. Predict price of Propertyguru listing provided by user using scaler and model saved by train_regression_model function</li> <li>4. Query database for historical transactions</li> <li>5. Recommend other Propertyguru listings to users such as listings with similar scores, similar listing price and similar flat size</li> </ol>

Table 2: List of functions developed in the backend

## 4.2 SYSTEM PERFORMANCE & VALIDATION

A detailed list of functional test cases was performed after the development of the application was completed. The test scope included validation of functions developed, the look and feel of the user interface and performance of the application. The lists of functional tests along with expected and actual results are detailed in Table 3.

S/N	Test Case Description	Location	Test type	Expected Result	Actual Result
1	Start HRPR website using Ubuntu OS	Home	Happy path	HRPR website loads up successfully without any format/layout breaking.	HRPR website loads up successfully without any format/design breaking
2	Start HRPR website using Mac OS	Home	Exceptional Handling		
3	Start HRPR website using Windows OS	Home	Exceptional Handling	HRPR website loads up successfully but CSS layout breaks.	HRPR website loads up successfully but CSS layout breaks.
4	Click “HDB App” on top left corner	Anywhere	Happy path	HRPR website loads up successfully format/layout breaking.	HRPR website loads up successfully without any format/design breaking
5	Click “Home” tab on top menu bar	Anywhere	Happy path	Goes back to Home page	Goes back to Home page
6	Click “Update” tab on top menu bar	Home	Happy path	Pull down pop up	Pull down pop up
7	Click “Update” tab on top menu bar and click “Update Data Gov Table”	Home	Happy path	Goes to Update Data Gov page	Goes to Update Data Gov page
8	Click “Update” tab on top menu bar and click “Update PropertyGuru Table”	Home	Happy path	Goes to Update PropertyGuru page	Goes to Update PropertyGuru page
9	Click “Update” tab on top menu bar and click “Update Amenities Table”	Home	Happy path	Goes to Update Amenities page	Goes to Update Amenities page
10	Click “Update” tab on top menu bar and click “Update Training Table”	Home	Happy path	Goes to Update Training page	Goes to Update Training page
11	Click one of the Resale HDB Article	Home	Happy path	Redirects to article webpage	Redirects to article webpage
12	Click on Pull down menu	Update Data Gov Table	Happy path	Pull-down shows “Yes” and “No”	Pull-down shows “Yes” and “No”
13	Leave pull-down to “Yes” and click “Update”	Update Data Gov Table	Happy path	If there are no new records: “There are no new records to be updated into the database.”  If there are new records: “Updated Data Gov tables in database.”	If there are no new records: “There are no new records to be updated into the database.”  If there are new records: “Updated Data Gov tables in database.”
14	Leave pull-down to “No” and click “Update”	Update Data Gov Table	Happy path	Return to Home	Return to Home

15	Click on pull-down menu	Update PropertyGuru Table	Happy path	Pull-down shows “Yes” and “No”	Pull-down shows “Yes” and “No”
16	Leave pull-down to “Yes” and click “Update”	Update PropertyGuru Table	Happy path	If there are no new records: “There are no new records to be updated into the database.”  If there are new records: “Updated PropertyGuru tables in database.”	If there are no new records: “There are no new records to be updated into the database.”  If there are new records: “Updated PropertyGuru tables in database.”
17	Leave pull-down to “No” and click “Update”	Update PropertyGuru Table	Happy path	Return to Home	Return to Home
18	Click on pull-down menu	Update Amenities Table	Happy path	Pull-down shows “Yes” and “No”	Pull-down shows “Yes” and “No”
19	Leave pull-down to “Yes” and click “Update”	Update Amenities Table	Happy path	If there are no new records: “There are no new records to be updated into the database.”  If there are new records: “Updated amenities tables in database.”	If there are no new records: “There are no new records to be updated into the database.”  If there are new records: “Updated amenities tables in database.”
20	Leave pull-down to “No” and click “Update”	Update Amenities Table	Happy path	Return to Home	Return to Home
21	Click on pull-down menu	Update Training Models	Happy path	Pull-down shows “Yes” and “No”	Pull-down shows “Yes” and “No”
22	Leave pull-down to “Yes” and click “Update”	Update Training Models	Happy path	“Trained regression model.”	“Trained regression model.”
23	Leave pull-down to “No” and click “Update”	Update Training Models	Happy path	Return to Home	Return to Home
24	Input valid PropertyGuru resale HDB URL input box and click “Search”	Home	Happy path	Successfully give user the resale price recommendation relative to the floor, latest 5 transactions and 9 other recommended resale HDB listings	Successfully give user the resale price recommendation relative to the floor, latest 5 transactions and 9 other recommended resale HDB listings

25	Input invalid PropertyGuru resale HDB URL input box and click “Search”	Home	Exceptional handling	“Please use a valid PropertGuru HDB listing URL”	“Please use a valid PropertGuru HDB listing URL”
26	Input PropertyGuru Condominium listing URL input box and click “Search”	Home	Exceptional handling	“Please use a valid PropertGuru HDB listing URL”	“Please use a valid PropertGuru HDB listing URL”
27	Input PropertyGuru Landed Property listing URL input box and click “Search”	Home	Exceptional handling	“Please use a valid PropertGuru HDB listing URL”	“Please use a valid PropertGuru HDB listing URL”
	Click listing information	Resale price recommendation page	Happy path	Redirect to the PropertyGuru listing webpage	Redirect to the PropertyGuru listing webpage
28	Click one of the recommended listing from “You Might Also Consider?”	Resale price recommendation page	Happy path	Redirect to the PropertyGuru listing webpage	Redirect to the PropertyGuru listing webpage
29	Click one of the recommended listing from “Similar size, better price!”	Resale price recommendation page	Happy path	Redirect to the PropertyGuru listing webpage	Redirect to the PropertyGuru listing webpage
30	Click one of the recommended listing from “Similar price, bigger house!”	Resale price recommendation page	Happy path	Redirect to the PropertyGuru listing webpage	Redirect to the PropertyGuru listing webpage

Table 3: Functional Test Cases

---

## 5. CONCLUSION

---

### 5.1 PROJECT CONCLUSION

In a nutshell, we enjoyed our time getting our hands dirty in doing this Practice Module. Albeit it is challenging to juggle between full-time work and part-time studies, the project allows us to appreciate Intelligent Reasoning Systems better. For instance, this project taught us the importance of ETL, building a model and using the model to predict and recommend.

We also appreciate the feedback that we have received from our lecturers, friends and family. Their input is valuable to us as they could help us to see from another point of view.

We hope that HPRP can be a useful tool that can be further scaled as we learnt more during our part time M.Tech course.

### 5.2 IMPROVEMENTS

We acknowledge that the current HPRP still has multiple rooms of improvements. Following are some of the areas that we identify:

1. **Ability to work with other property listings website** such as 99.co, SRX etc. Currently, our scraped data is only from Property Guru and users can only provide search listing from there. Reason being there is constraint on the speed of web scraping (which currently takes about 4 hours for complete listing). Aside, duplicate listings of same unit might be posted on various property listings websites.
2. **User-based Collaborative filtering** user recommender system. In such system, we could obtain the profile of prospective homebuyer and request them to rank few listings, which could assist in providing a more accurate recommender system.
3. **Model hyperparameter tuning using Genetic Evolutionary algorithms**, to obtain best model hyperparameters with no human supervision. With periodic database updates, model heuristics would change meaning that hyperparameters that were used in the previous update might not give the best outcome anymore. With genetic algorithms implemented, model accuracy could improve with faster convergence time, all with a single command.
4. **Duplicate listings posted by different agents**. Currently, we are still unable to filter out such rare occurrences. However, a possible idea is to compare the listing images and other features and if the similarity score is above a certain threshold, duplicate listing should be removed.

## 6. APPENDIX

### 6.1 PROJECT PROPOSAL

#### GRADUATE CERTIFICATE: Intelligent Reasoning Systems (IRS)

#### PRACTICE MODULE: Project Proposal

##### Date of proposal:

3 March 2021

##### Project Title:

IRS-PM Project: HDB RESALE PREDICTION AND RECOMMENDER

##### Sponsor/Client: (Name, Address, Telephone No. and Contact Name)

Institute of Systems Science (ISS) at 25 Heng Mui Keng Terrace, Singapore  
 NATIONAL UNIVERSITY OF SINGAPORE (NUS)  
 Contact: Mr. GU ZHAN / Lecturer & Consultant  
 Telephone No.: 65-6516 8021  
 Email: [zhan.gu@nus.edu.sg](mailto:zhan.gu@nus.edu.sg)

##### Background/Aims/Objectives:

Given a Property Guru listing, the proposed HDB Resale Prediction and Recommender (HRPR) assists prospective buyers in their resale flat search by predicting a suitable price. At the same time, the system also recommends other listings that are similar to what the user has provided, using matrix obtained from survey results and other suitable filters.

##### Requirements Overview:

- Webscraping ability
- Database ability
- Feature Engineering ability
- Data Analysis ability
- System Integration ability
- Video Editing ability

##### Resource Requirements (please list Hardware, Software and any other resources)

Hardware proposed for consideration:

- PC (Ubuntu 20.04-based, can be hosted on VM)

Software proposed for consideration:

- Firefox (for Selenium webscraping purpose)
- Web Browser for navigating the system
- Flask
- Various Python libraries (as indicated in requirements.txt)



**Number of Learner Interns required: (Please specify their tasks if possible)**

4 Team Members

Full Name	Student ID	Work Items	Email
Yee Wei Liang	A0045422R	Web Scraping, Data Cleaning, Feature Engineering	<a href="mailto:E0258287@u.nus.edu">E0258287@u.nus.edu</a>
Toh Kah Khek	A0229968E	Database Setup, Front-end Development, Back-end Integration	<a href="mailto:E0687376@u.nus.edu">E0687376@u.nus.edu</a>
Jeon Sungmin	A0133374J	Collaborative Filtering Recommendation System, System Performance & Testing, Video Editing	<a href="mailto:E0689806@u.nus.edu">E0689806@u.nus.edu</a>
Ahmed Syalabi Seet	A0229978A	Data Analysis, Data Cleaning, Prediction Tree Modelling	<a href="mailto:E0687386@u.nus.edu">E0687386@u.nus.edu</a>

**Methods and Standards:**

Procedures	Objective	Key Activities
<b>Requirement Gathering and Analysis</b>	<p>The team should meet with ISS to scope the details of project and ensure the achievement of business objectives.</p> <p>After brainstorming of few ideas, the team met with Sam Gu to discuss on the ideas and eventually decided on current idea.</p>	<ol style="list-style-type: none"> <li>Gather &amp; Analyze Requirements <ul style="list-style-type: none"> <li>Google Sheet to discuss ideas</li> <li>Explore existing solutions</li> <li>Decide on HPRP</li> </ul> </li> <li>Define internal and external design <ul style="list-style-type: none"> <li>System design</li> <li>Roles discussion</li> </ul> </li> <li>Prioritize &amp; Consolidate Requirements <ul style="list-style-type: none"> <li>Draft system requirements</li> </ul> </li> <li>Establish Functional Baseline <ul style="list-style-type: none"> <li>User inputs</li> <li>Data model design concept</li> <li>Output</li> </ul> </li> </ol>
<b>Technical Construction</b>	<p>To develop the source code in accordance to the design.</p> <p>To perform unit testing to ensure the quality before the components are integrated as a whole project</p>	<ol style="list-style-type: none"> <li>Setup Development Environment (Ubuntu-VM)</li> <li>Understand the System Context, Design</li> <li>Perform Coding</li> <li>Conduct Unit Testing</li> </ol>

<b>Integration Testing and acceptance testing</b>	To ensure interface compatibility and confirm that the integrated system hardware and system software meets requirements and is ready for acceptance testing.	<ol style="list-style-type: none"> <li>1. Prepare System Test Specifications</li> <li>2. Prepare for Test Execution</li> <li>3. Conduct System Integration Testing between different modules developed by team members</li> <li>4. Evaluate Testing</li> </ol>
<b>Acceptance Testing</b>	To obtain ISS user acceptance that the system meets the requirements.	<ol style="list-style-type: none"> <li>1. ISS Lecturers Evaluate Testing</li> </ol>
<b>Delivery</b>	To deploy the system into production (ISS standalone server) environment.	<ol style="list-style-type: none"> <li>1. Deployment using Flask framework served in supported Web Browser</li> </ol>

### Team Formation & Registration

Team Name: JSSW
Project Title (repeated): HDB RESALE PREDICTION AND RECOMMENDER
Team Member 1 Name: YEE WEI LIANG
Team Member 1 Matriculation Number: A0045422R
Team Member 1 Contact (Mobile/Email): <a href="mailto:E0258287@u.nus.edu">E0258287@u.nus.edu</a>
Team Member 2 Name: TOH KAH KHEK
Team Member 2 Matriculation Number: A0229968E
Team Member 2 Contact (Mobile/Email): <a href="mailto:E0687376@u.nus.edu">E0687376@u.nus.edu</a>
Team Member 3 Name: JEON SUNGMIN
Team Member 3 Matriculation Number: A0133374J
Team Member 3 Contact (Mobile/Email): <a href="mailto:E0689806@u.nus.edu">E0689806@u.nus.edu</a>
Team Member 4 Name: AHMED SYALABI SEET
Team Member 4 Matriculation Number: A0229978A
Team Member 4 Contact (Mobile/Email): <a href="mailto:E0687386@u.nus.edu">E0687386@u.nus.edu</a>

For ISS Use Only		
<b>Programme Name:</b>	<b>Project No:</b>	<b>Learner Batch:</b>
<b>Accepted/Rejected/KIV:</b>		
<b>Learners Assigned:</b>		
<b>Advisor Assigned:</b>  Contact: Mr. GU ZHAN / Lecturer & Consultant Telephone No.: 65-6516 8021 Email: <a href="mailto:zhan.gu@nus.edu.sg">zhan.gu@nus.edu.sg</a>		

## 6.2 MAPPING OF DOMAIN/TECHNIQUE GROUP

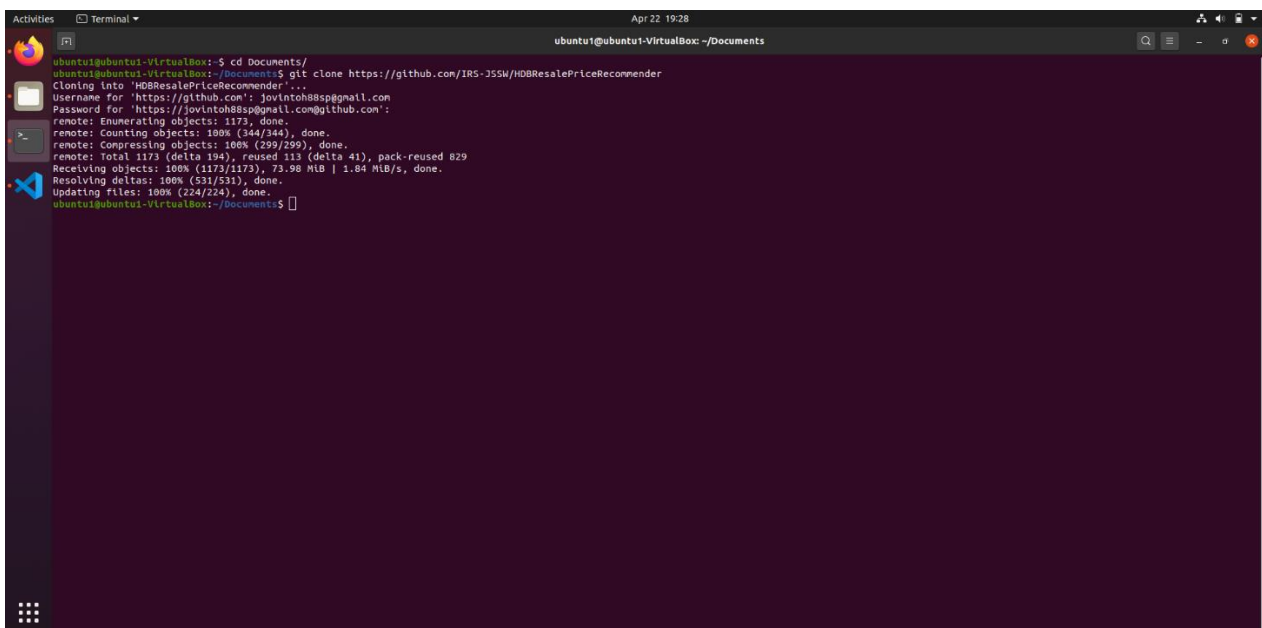
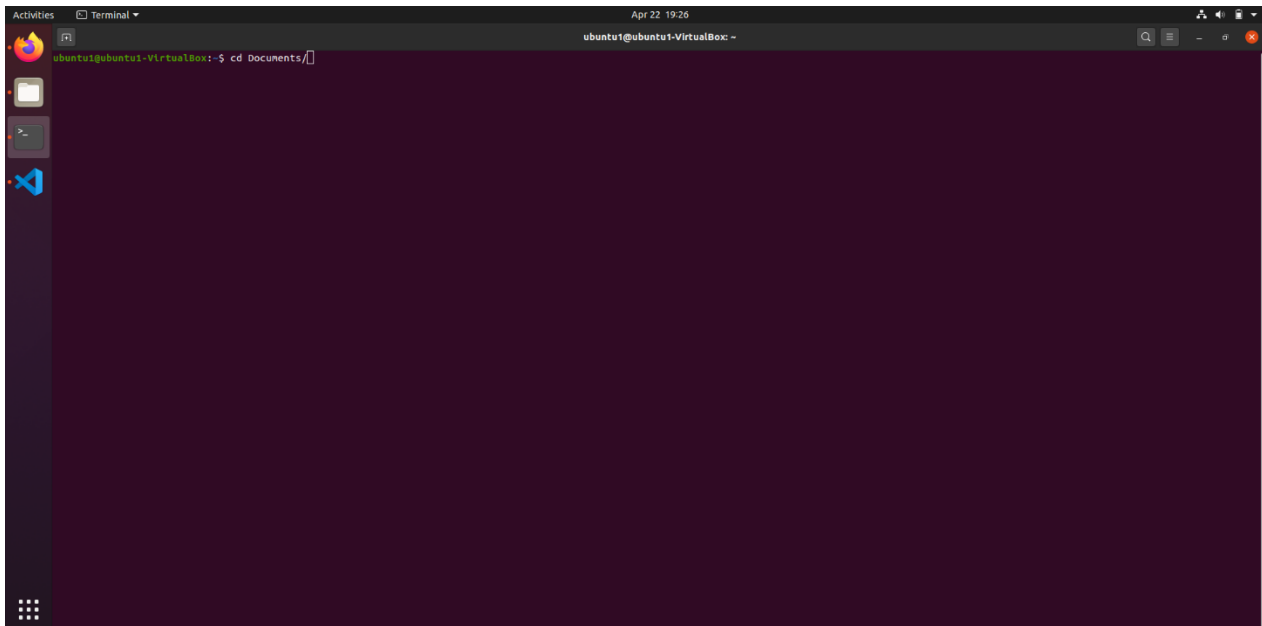
	Domain/Technique Group	Shown in:
1	<b>Knowledge Based Reasoning</b>	a. Price Prediction Tree Modelling
2	<b>Knowledge Discovery using Data Mining</b>	a. Web Scraping of listings details from Property Guru b. Data Retrieval of Past Resale Transactions using DataGov API c. Data Retrieval of Location Information using OneMap API/Google Map API
3	<b>User-based Recommender System</b>	a. Scoring of Recommender System of “You Might Also Consider?” obtained through User Survey

## 6.3 INSTALLATION AND USER GUIDE

### Installation guide for Ubuntu 20.04

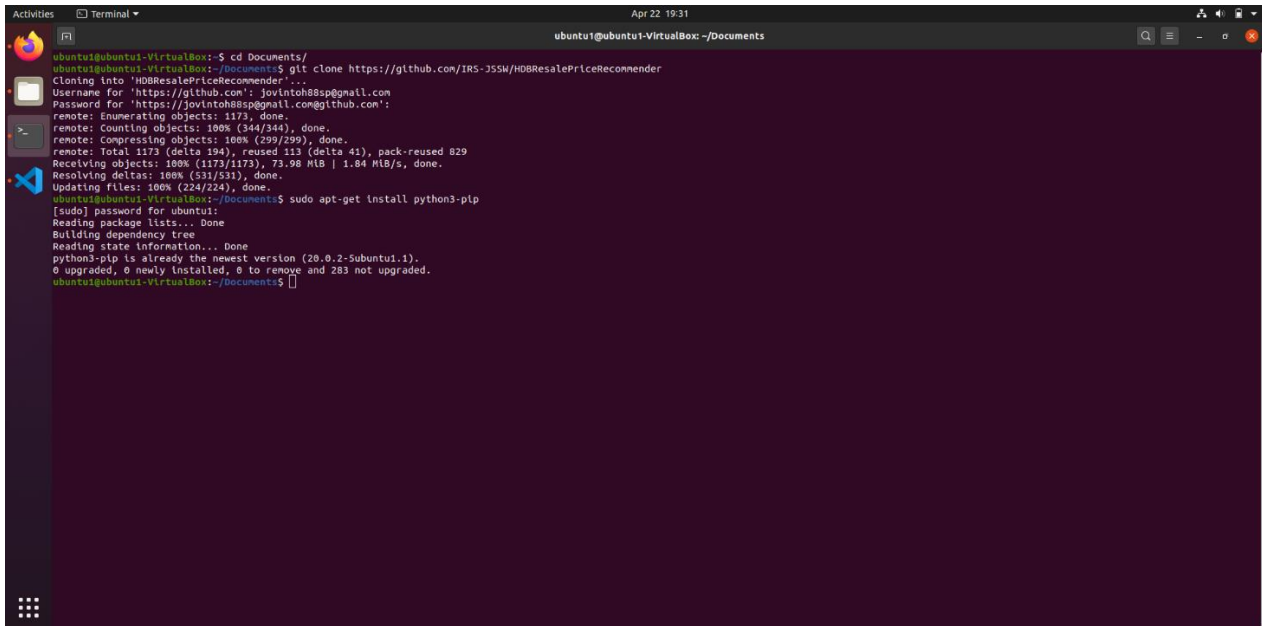
1. Navigate to folder of your choice and download the github repository

command: `git clone https://github.com/IRS-JSSW/HDBResalePriceRecommender`



## 2. Install pip3 package

command: `sudo apt-get install python3-pip`



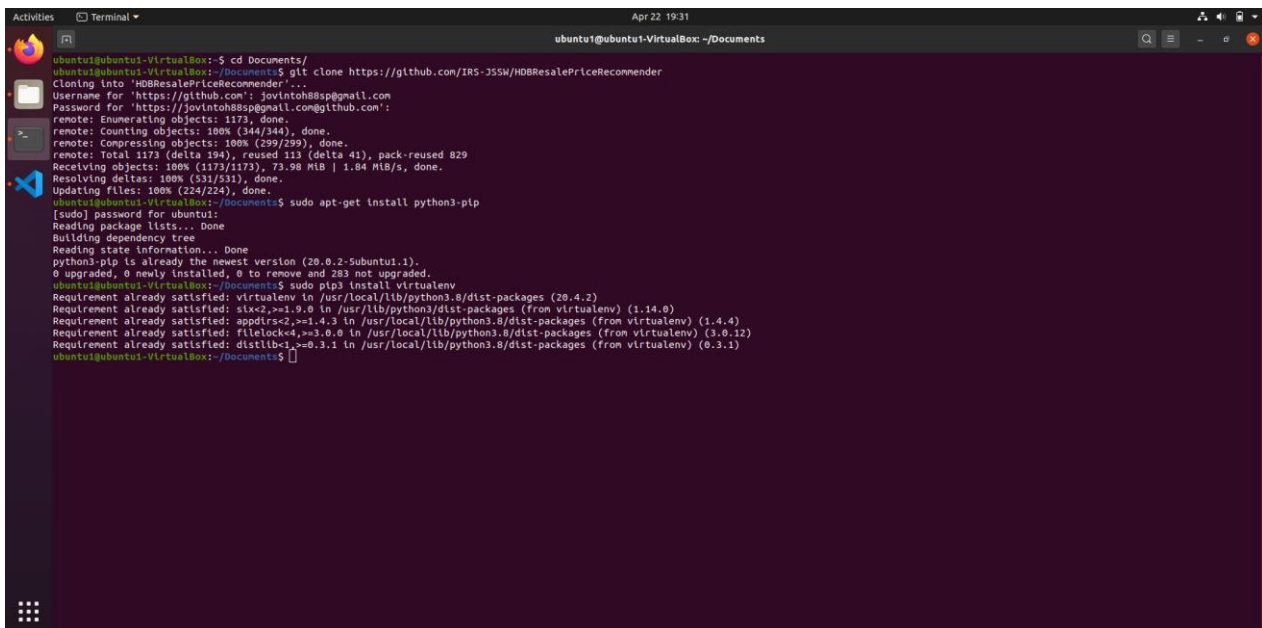
```

ubuntu@ubuntu-VirtualBox:~$ cd Documents/
ubuntu@ubuntu-VirtualBox:~/Documents$ git clone https://github.com/IRS-355W/HDBResalePriceRecommender
Cloning into 'HDBResalePriceRecommender'...
Username for 'https://github.com': jovintoh88sp@gmail.com
Password for 'https://github.com': 
remote: Enumerating objects: 1173, done.
remote: Counting objects: 100% (344/344), done.
remote: Compressing objects: 100% (299/299), done.
remote: Total 1173 (delta 194), reused 113 (delta 41), pack-reused 629
Receiving objects: 100% (1173/1173), 73.98 MiB | 1.84 MiB/s, done.
Resolving deltas: 100% (531/531), done.
Updating files: 100% (224/224), done.
ubuntu@ubuntu-VirtualBox:~/Documents$ sudo apt-get install python3-pip
[sudo] password for ubuntu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
python3-pip is already the newest version (20.0.2-Subuntui.1).
0 upgraded, 0 newly installed, 0 to remove and 283 not upgraded.
ubuntu@ubuntu-VirtualBox:~/Documents$

```

## 3. Install virtualenv package

command: `sudo pip3 install virtualenv`



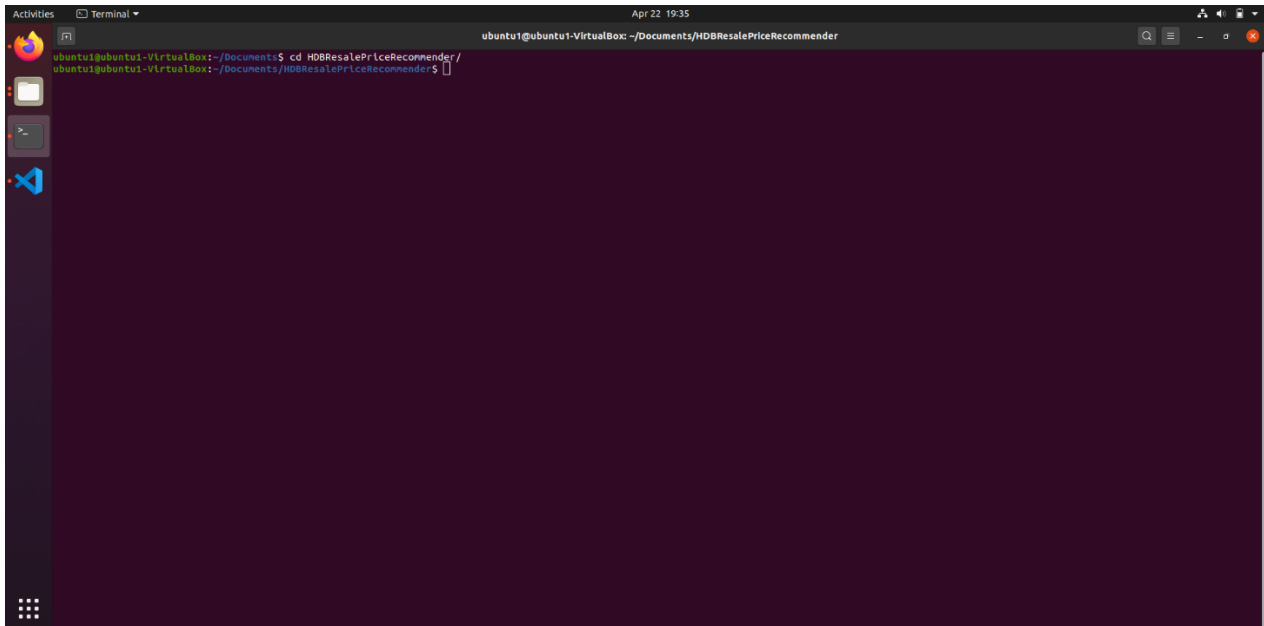
```

ubuntu@ubuntu-VirtualBox:~$ cd Documents/
ubuntu@ubuntu-VirtualBox:~/Documents$ git clone https://github.com/IRS-355W/HDBResalePriceRecommender
Cloning into 'HDBResalePriceRecommender'...
Username for 'https://github.com': jovintoh88sp@gmail.com
Password for 'https://github.com': 
remote: Enumerating objects: 1173, done.
remote: Counting objects: 100% (344/344), done.
remote: Compressing objects: 100% (299/299), done.
remote: Total 1173 (delta 194), reused 113 (delta 41), pack-reused 629
Receiving objects: 100% (1173/1173), 73.98 MiB | 1.84 MiB/s, done.
Resolving deltas: 100% (531/531), done.
Updating files: 100% (224/224), done.
ubuntu@ubuntu-VirtualBox:~/Documents$ sudo apt-get install python3-pip
[sudo] password for ubuntu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
python3-pip is already the newest version (20.0.2-Subuntui.1).
0 upgraded, 0 newly installed, 0 to remove and 283 not upgraded.
ubuntu@ubuntu-VirtualBox:~/Documents$ sudo pip3 install virtualenv
Requirement already satisfied: virtualenv in /usr/local/lib/python3.8/dist-packages (20.4.2)
Requirement already satisfied: six<2,=>1.9.0 in /usr/local/lib/python3.8/dist-packages (from virtualenv) (1.14.0)
Requirement already satisfied: appdirs<2,=>1.4.3 in /usr/local/lib/python3.8/dist-packages (from virtualenv) (1.4.4)
Requirement already satisfied: filelock<4,=>3.0.0 in /usr/local/lib/python3.8/dist-packages (from virtualenv) (3.0.12)
Requirement already satisfied: distlib<1,=>0.3.1 in /usr/local/lib/python3.8/dist-packages (from virtualenv) (0.3.1)
ubuntu@ubuntu-VirtualBox:~/Documents$

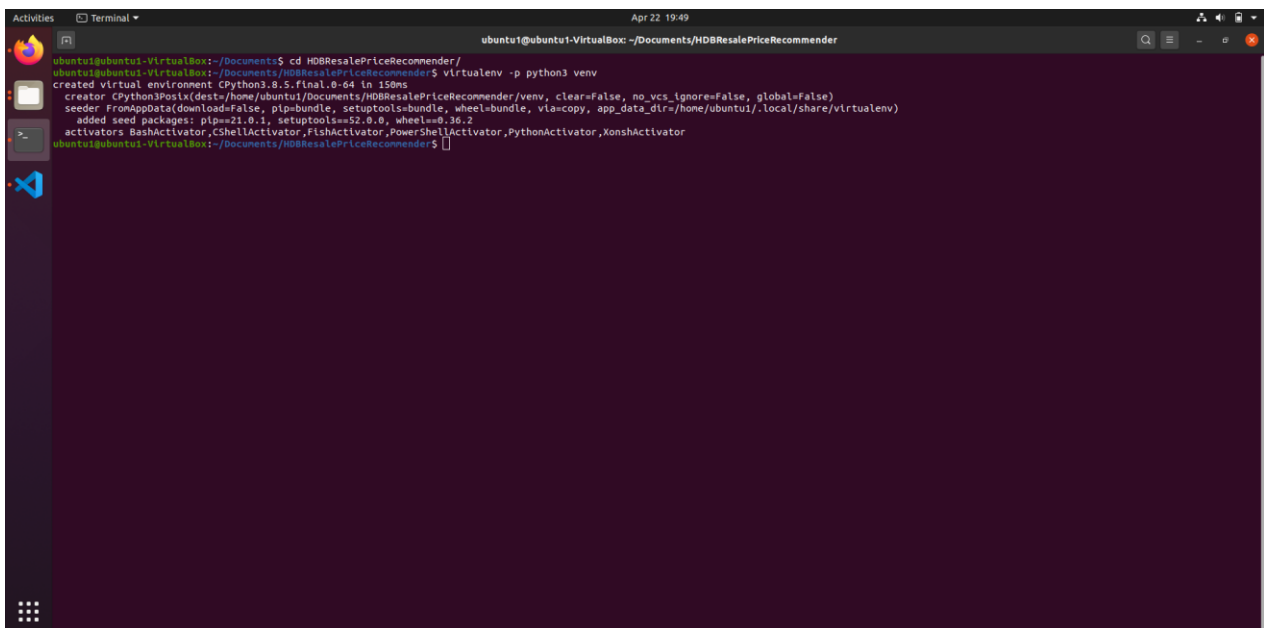
```

## 4. Navigate to HDBResalePriceRecommender and create virtualenv

command: `virtualenv -p python3 venv`



A terminal window titled 'Terminal' with a dark background. The prompt is `ubuntu1@ubuntu1-VirtualBox: ~/Documents/HDBResalePriceRecommender`. The user has entered the command `cd HDBResalePriceRecommender/` and the prompt is now `ubuntu1@ubuntu1-VirtualBox: ~/Documents/HDBResalePriceRecommender$`.



The same terminal window showing the output of the `virtualenv -p python3 venv` command. The output is as follows:

```
ubuntu1@ubuntu1-VirtualBox: ~/Documents/HDBResalePriceRecommender$ virtualenv -p python3 venv
created virtual environment CPython3.8.5.final.0-64 in 150ms
creator CPython3Posix(dest=/home/ubuntu1/Documents/HDBResalePriceRecommender/venv, clear=False, no_vcs_ignore=False, global=False)
seeder FromAppData(download=False, pipbundle=True, setuptoolsbundle=True, wheelbundle=True, via=copy, app_data_dir=/home/ubuntu1/.local/share/virtualenv)
added seed packages: pip==21.0.1, setuptools==52.0.0, wheel==0.36.2
activators BashActivator,CShellActivator,FishActivator,PowerShellActivator,PythonActivator,KonshActivator
ubuntu1@ubuntu1-VirtualBox: ~/Documents/HDBResalePriceRecommender$
```



## 5. Activate virtualenv

command: `source venv/bin/activate` (Note: Ensure that you are seeing (venv) in the terminal)

```

ubuntu1@ubuntu1-VirtualBox: ~/Documents/HDBResalePriceRecommender
ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$ venv/bin/python3 -m venv venv
created virtual environment CPython3.8.5.final.0-64 in 150ms
creator CPython3Posix(dest=/home/ubuntu1/Documents/HDBResalePriceRecommender/venv, clear=False, no_vcs_ignore=False, global=False)
seeder FromAppData(download=False, pip=bundle, setuptools=bundle, wheel=bundle, via=copy, app_data_dir=/home/ubuntu1/.local/share/virtualenv)
added seed packages: pip==21.0.1, setuptools==52.0.0, wheel==0.36.2
activators BashActivator,CShellActivator,FishActivator,PowershellActivator,PythonActivator,KonshActivator
ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$ source venv/bin/activate
(venv) ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$
  
```

## 6. Install project dependencies from requirements.txt

command: `pip install -r requirements.txt`

No	Package	Version
1	flask	1.1.2
2	flask-wtf	0.14.3
3	haversine	2.3.0
4	requests	2.25.1
5	selenium	3.141.0
6	sklearn	-----
7	sqlalchemy	1.3.23

```

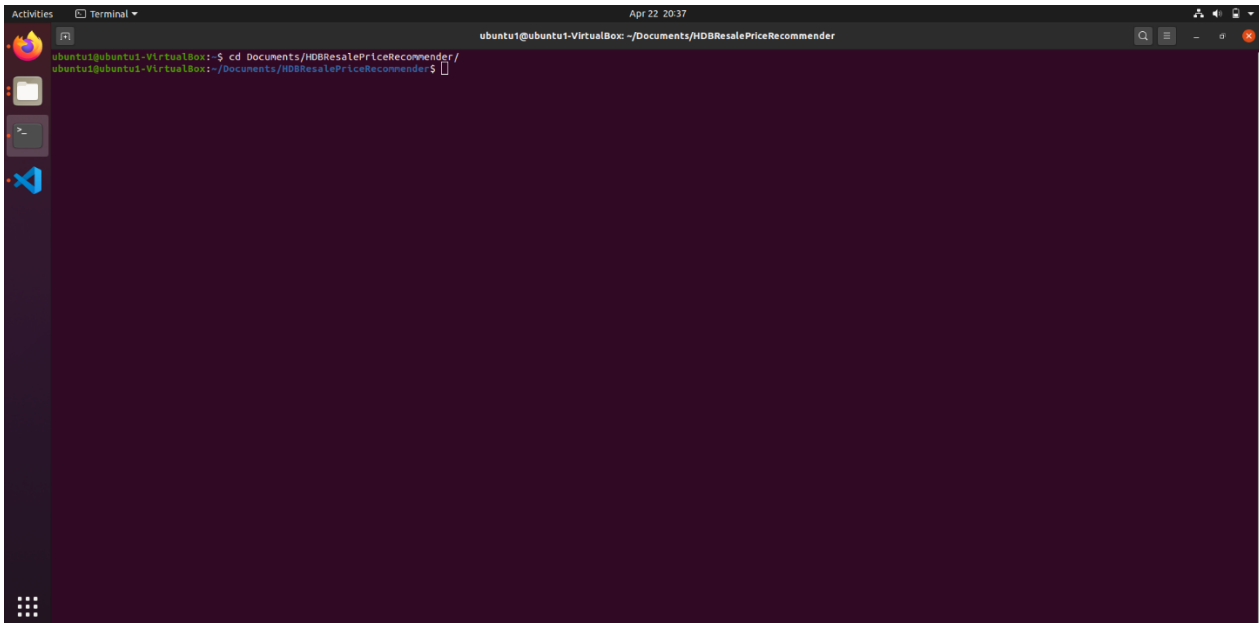
(venv) ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$ pip install -r requirements.txt
Collecting idna==2.10
Using cached idna-2.10-py2.py3-none-any.whl (58 kB)
Collecting itsdangerous==1.1.0
Using cached itsdangerous-1.1.0-py2.py3-none-any.whl (16 kB)
Collecting Jinja2==2.11.3
Using cached Jinja2-2.11.3-py2.py3-none-any.whl (125 kB)
Collecting joblib==1.0.1
Using cached joblib-1.0.1-py3-none-any.whl (303 kB)
Collecting MarkupSafe==1.1.1
Using cached MarkupSafe-1.1.1-cp38-cp38-manylinux2010_x86_64.whl (32 kB)
Collecting numpy==1.20.1
Using cached numpy-1.20.1-cp38-cp38-manylinux2010_x86_64.whl (15.4 MB)
Collecting pandas==1.1.5
Using cached pandas-1.1.5-cp38-cp38-manylinux1_x86_64.whl (9.3 MB)
Collecting python-dateutil==2.8.1
Using cached python_dateutil-2.8.1-py2.py3-none-any.whl (227 kB)
Collecting pytz==2021.1
Using cached pytz-2021.1-py2.py3-none-any.whl (510 kB)
Collecting requests==2.25.1
Using cached requests-2.25.1-py2.py3-none-any.whl (61 kB)
Collecting requests-toolbelt==0.9.1
Using cached requests_toolbelt-0.9.1-py2.py3-none-any.whl (54 kB)
Collecting scikit-learn==0.24.1
Using cached scikit_learn-0.24.1-cp38-cp38-manylinux2010_x86_64.whl (24.9 MB)
Collecting scipy==1.6.2
Using cached scipy-1.6.2-cp38-cp38-manylinux1_x86_64.whl (27.2 MB)
Collecting selenium==3.141.0
Using cached selenium-3.141.0-py2.py3-none-any.whl (904 kB)
Collecting six==1.15.0
Using cached six-1.15.0-py2.py3-none-any.whl (10 kB)
Collecting sklearn==0.0
Using cached sklearn-0.0-py2.py3-none-any.whl
Collecting SQLAlchemy==1.3.23
Using cached SQLAlchemy-1.3.23-cp38-cp38-manylinux2010_x86_64.whl (1.3 MB)
Collecting threadpoolctl==2.1.0
Using cached threadpoolctl-2.1.0-py3-none-any.whl (12 kB)
Collecting urllib3==1.26.4
Using cached urllib3-1.26.4-py2.py3-none-any.whl (153 kB)
Collecting Werkzeug==1.0.1
Using cached Werkzeug-1.0.1-py2.py3-none-any.whl (298 kB)
Collecting WTForms==2.3.3
Using cached WTForms-2.3.3-py2.py3-none-any.whl (169 kB)
Installing collected packages: numpy, MarkupSafe, Werkzeug, urllib3, threadpoolctl, six, scipy, joblib, Jinja2, itsdangerous, idna, click, chardet, certifi, WTForms, SQLAlchemy, scikit-learn, requests, pytz, python-dateutil, Flask, sklearn, selenium, requests-toolbelt, pandas, haversine, Flask-WTF, Flask-SQLAlchemy
Successfully installed Flask-1.1.2 Flask-SQLAlchemy-2.4.4 Flask-WTF-0.14.3 Jinja2-2.11.3 MarkupSafe-1.1.1 SQLAlchemy-1.3.23 WTForms-2.3.3 Werkzeug-1.0.1 certifi-2020.12.5 chardet-4.0.0 click-7.1.2 haversine-2.3.0 idna-2.10 itsdangerous-1.1.0 joblib-1.0.1 numpy-1.20.1 pandas-1.1.5 python-dateutil-2.8.1 pytz-2021.1 requests-2.25.1 requests-toolbelt-0.9.1 scikit-learn-0.24.1 scipy-1.6.2 selenium-3.141.0 six-1.15.0 sklearn-0.0 threadpoolctl-2.1.0 urllib3-1.26.4
(venv) ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$
  
```

## User Guide

### Part 1 – To start the app

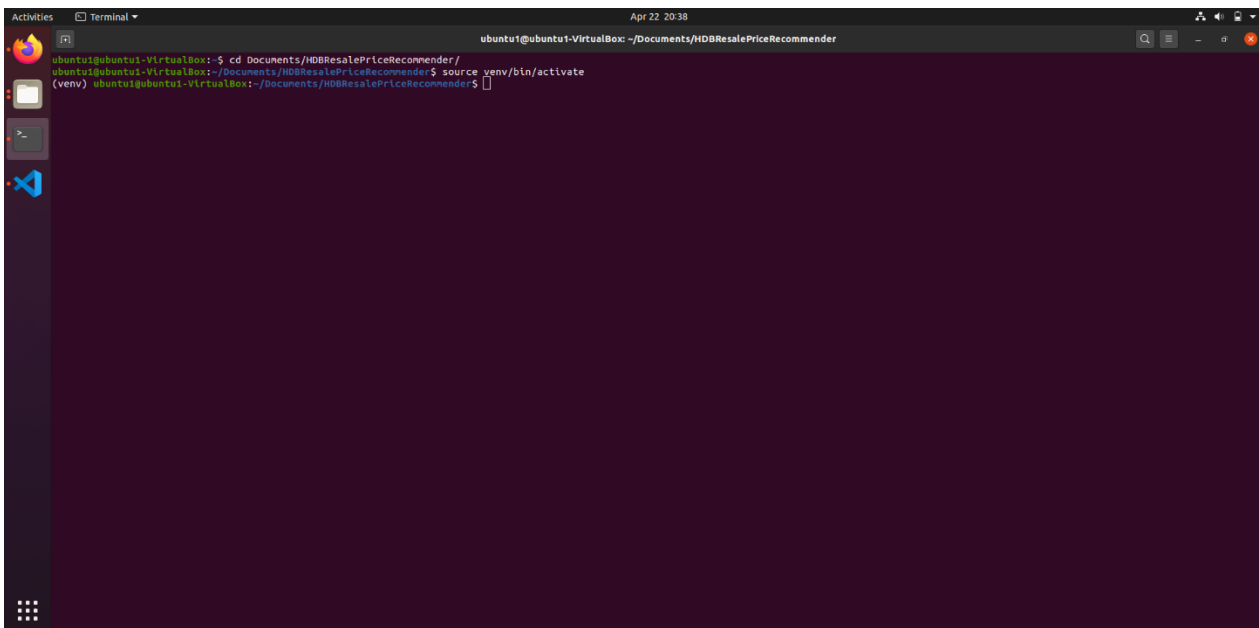
1. Navigate to HDBResalePriceRecommender and activate virtualenv

command: `source venv/bin/activate` (Note: Ensure that you are seeing `(venv)` in the terminal)



A terminal window titled 'Terminal' with a dark background. The prompt is 'ubuntu1@ubuntu1-VirtualBox: ~/Documents/HDBResalePriceRecommender'. The user has entered the command `cd Documents/HDBResalePriceRecommender/` and the prompt has moved to the subdirectory. The terminal shows the following commands and output:

```
ubuntu1@ubuntu1-VirtualBox:~$ cd Documents/HDBResalePriceRecommender/
ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$
```

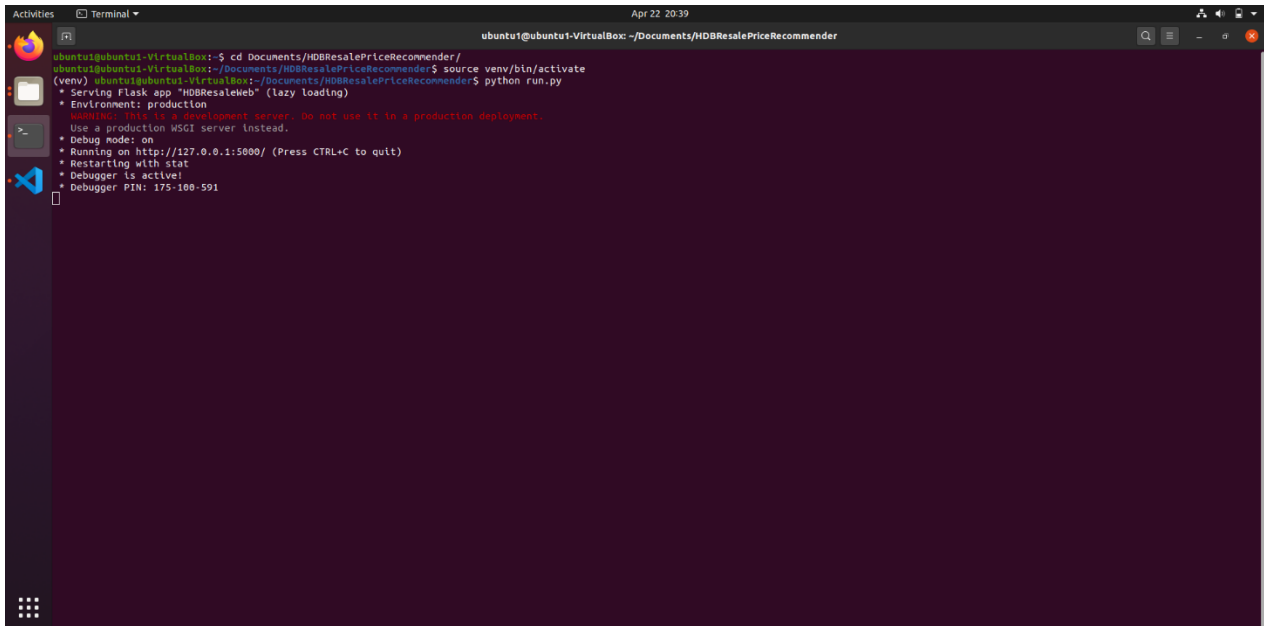


A terminal window titled 'Terminal' with a dark background. The prompt is 'ubuntu1@ubuntu1-VirtualBox: ~/Documents/HDBResalePriceRecommender'. The user has entered the command `source venv/bin/activate` and the prompt has changed to `(venv) ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$`. The terminal shows the following commands and output:

```
ubuntu1@ubuntu1-VirtualBox:~$ cd Documents/HDBResalePriceRecommender/
ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$ source venv/bin/activate
(venv) ubuntu1@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$
```

## 2. Start the flask application on local machine

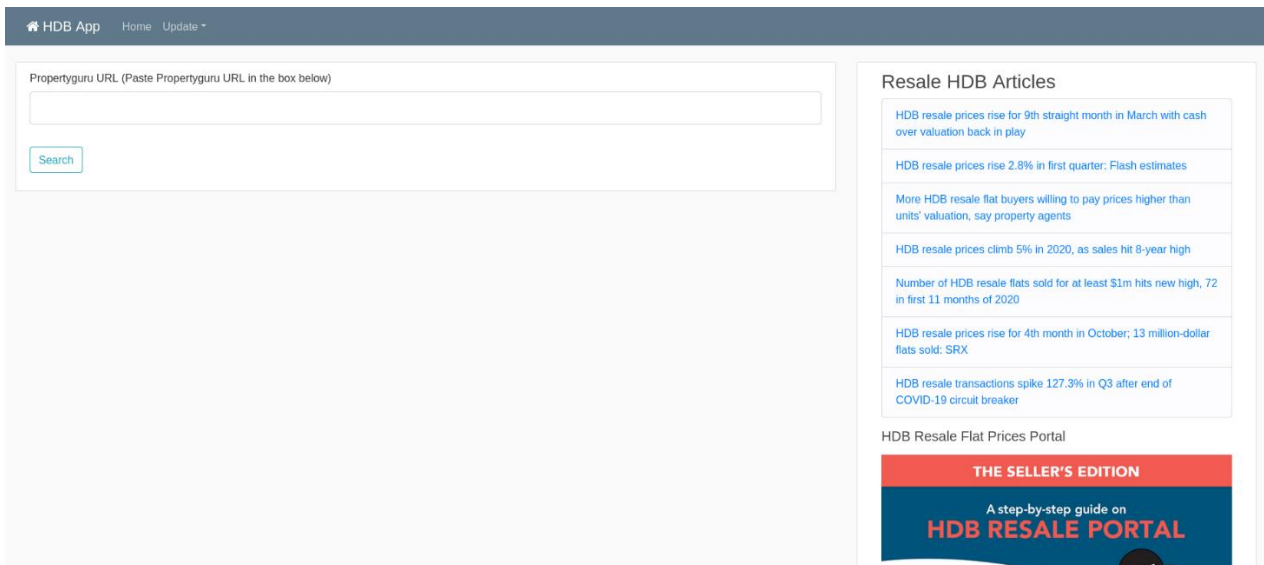
command: `project run.py`



```

ubuntu@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$ cd Documents/HDBResalePriceRecommender/
ubuntu@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$ source venv/bin/activate
(venv) ubuntu@ubuntu1-VirtualBox:~/Documents/HDBResalePriceRecommender$ python run.py
* Serving Flask app "HDBResaleWeb" (lazy loading)
* Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
* Debug mode: on
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
* Restarting with stat
* Debugger is active!
* Debugger PIN: 175-188-591
  
```

## 3. Open browser (Firefox or Google Chrome) and navigate to URL <http://127.0.0.1:5000/home>



## Part 2 - To get predicted price of HDB Resale Flats

1. Input valid Propertyguru Resale HDB listing in textbox and click on the Search button

HDB App
Home
Update

Propertyguru URL (Paste Propertyguru URL in the box below)

### Resale HDB Articles

- HDB resale prices rise for 9th straight month in March with cash over valuation back in play
- HDB resale prices rise 2.8% in first quarter: Flash estimates
- More HDB resale flat buyers willing to pay prices higher than units' valuation, say property agents
- HDB resale prices climb 5% in 2020, as sales hit 8-year high
- Number of HDB resale flats sold for at least \$1m hits new high, 72 in first 11 months of 2020
- HDB resale prices rise for 4th month in October; 13 million-dollar flats sold: SRX
- HDB resale transactions spike 127.3% in Q3 after end of COVID-19 circuit breaker

### HDB Resale Flat Prices Portal

THE SELLER'S EDITION  
A step-by-step guide on  
**HDB RESALE PORTAL**

2. Results of predicted price and other recommended Propertyguru listings will be generated

HDB App
Home
Update

**258B Punggol Field**  
Flat type: 4 ROOM  
Floor Area: 93.5 sqm  
Year built: 2018  
Listed Price: \$480,000  
Predicted Price (By Storey Range):  
Level 1 to 3: \$462,700  
Level 4 to 6: \$487,700  
Level 7 to 9: \$508,700  
Level 10 to 12: \$512,800  
Level 13 and above: \$516,740

Transaction month	Flat type	Storey range	Floor area (sqm)	Resale Price
2020-12	3 ROOM	13 and above	68	\$390,000
2020-10	4 ROOM	04 TO 06	93	\$445,000
2020-09	4 ROOM	04 TO 06	93	\$433,000
2020-09	4 ROOM	04 TO 06	93	\$423,000
2020-07	3 ROOM	04 TO 06	68	\$375,000

You May Consider?

Similar size, better price!

Similar price, bigger house!

**161 Mei Ling Street**  
Flat type: 4 ROOM  
Floor Area: 85  
Year built: 2015  
Listed Price: \$450,000

**708 Hougang Avenue 2**  
Flat type: 4 ROOM  
Floor Area: 95  
Year built: 1984  
Listed Price: \$388,000

**206B Compassvale Lane**  
Flat type: 5 ROOM  
Floor Area: 123  
Year built: 1997  
Listed Price: \$495,000

**209 Bukit Batok Street 21**  
Flat type: 4 ROOM  
Floor Area: 110  
Year built: 2014  
Listed Price: \$450,000

**251 Hougang Avenue 3**  
Flat type: 4 ROOM  
Floor Area: 95  
Year built: 1984  
Listed Price: \$389,000

**939 Hougang Street 92**  
Flat type: 5 ROOM  
Floor Area: 122  
Year built: 1998  
Listed Price: \$500,000

**442 Ang Mo Kio Avenue 10**  
Flat type: 4 ROOM  
Floor Area: 92  
Year built: 2015  
Listed Price: \$408,000

**227A Compassvale Drive**  
Flat type: 4 ROOM  
Floor Area: 101  
Year built: 1998  
Listed Price: \$498,000

**206A Compassvale Lane**  
Flat type: 5 ROOM  
Floor Area: 122  
Year built: 1997  
Listed Price: \$480,000

### Resale HDB Articles

- HDB resale prices climb for 4th consecutive quarter in Q1
- HDB resale prices rise for 9th straight month in March with cash over valuation back in play
- HDB resale prices rise 2.8% in first quarter: Flash estimates
- More HDB resale flat buyers willing to pay prices higher than units' valuation, say property agents
- HDB resale prices climb 5% in 2020, as sales hit 8-year high
- Number of HDB resale flats sold for at least \$1m hits new high, 72 in first 11 months of 2020
- HDB resale prices rise for 4th month in October; 13 million-dollar flats sold: SRX
- HDB resale transactions spike 127.3% in Q3 after end of COVID-19 circuit breaker

### HDB Resale Flat Prices Portal

THE SELLER'S EDITION  
A step-by-step guide on  
**HDB RESALE PORTAL**

We plan to sell our HDB flat and have on view of:  

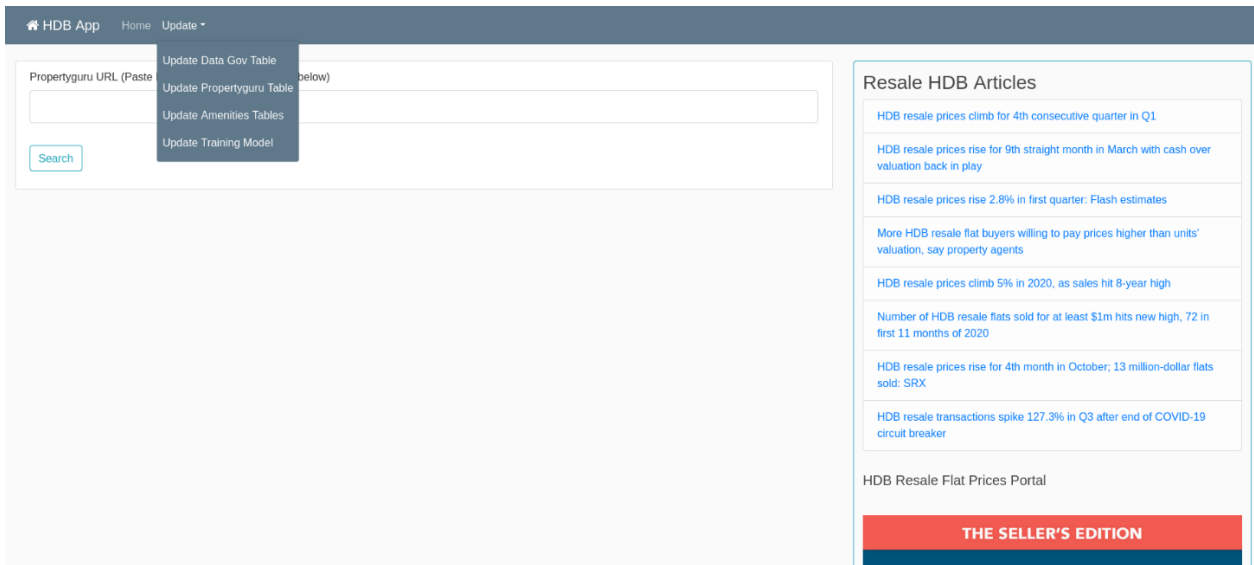
- The potential benefits of our flat
- The selling price that we are willing to accept
- The sale proceeds that we may get after selling
- The budget and options for our next home

#### USING THE HDB RESALE PORTAL

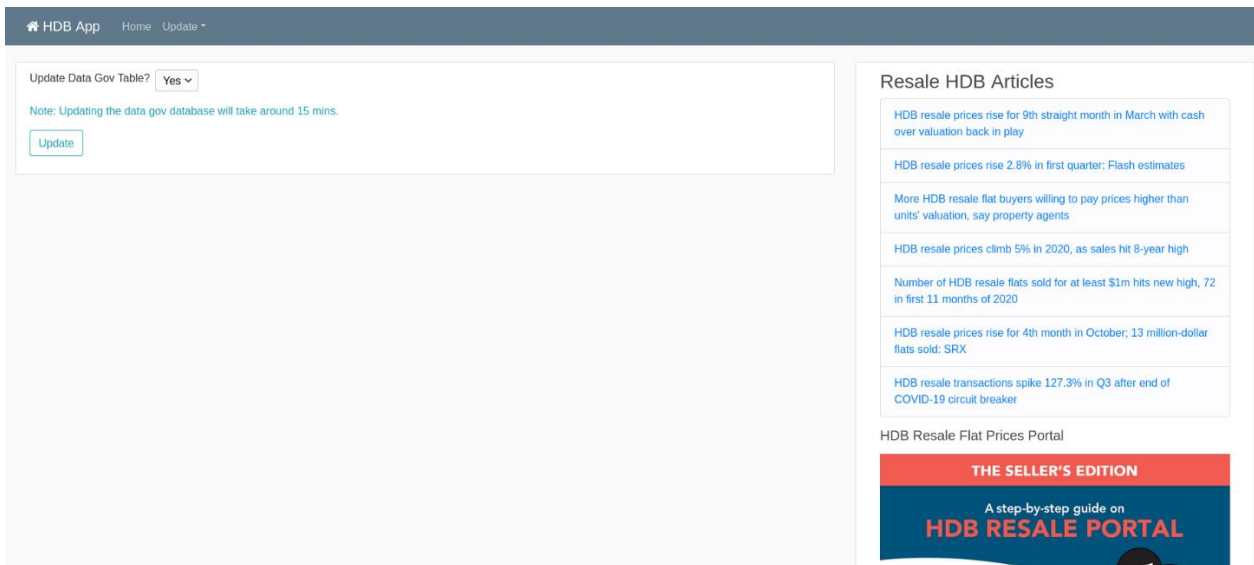
- REGISTER**  
The Internet is all you need! Register online and submit your details to the relevant authorities. You will receive a confirmation email and a QR code to use for registration.
- SUBMIT**  
Resale Application  
HDB will verify the information provided and notify you of the outcome of the application.
- ENDORSE**  
Resale Application  
We will be notified via SMS to return the documents on the resale application to HDB.
- PAY**  
Resale Fee Online  
You can pay the fee by Credit Card.
- RECEIVE**  
Approval of Resale Application  
We will be notified via SMS when HDB's approval letter is posted in the portal.
- ATTEND**  
Resale Completion Appointment  
We will be informed via SMS on the appointment to complete the transaction.

### Part 3 - To update HDB Resale Transactions Records from Data.gov

1. Click on “Update” on the navigation bar and click on “Update Data Gov Table”

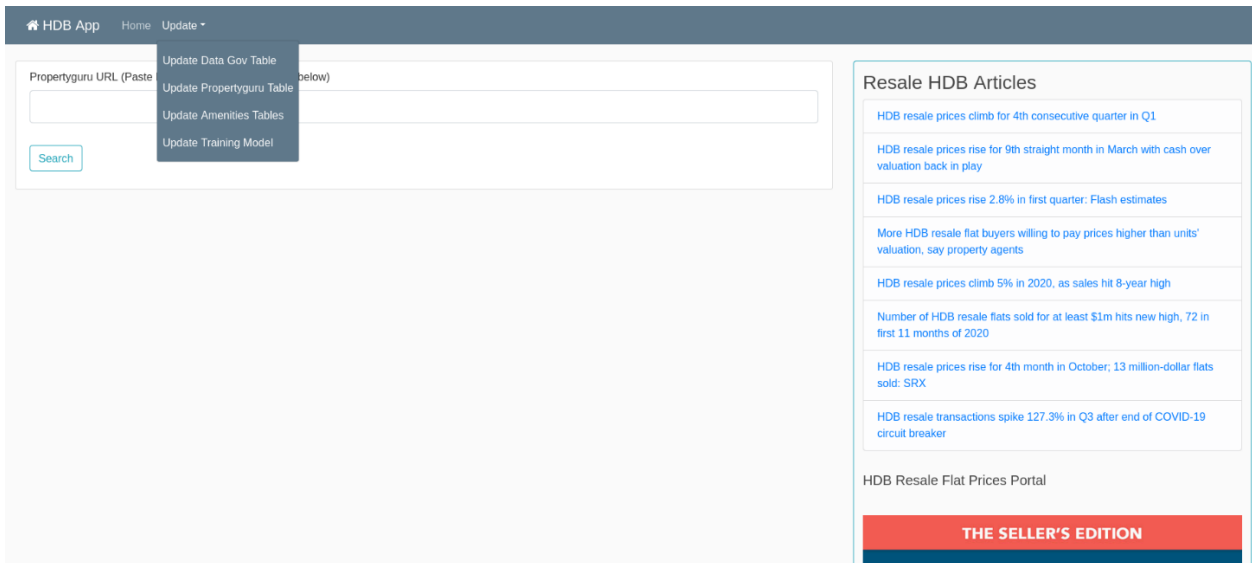


2. Next, select “Yes” in the dropdown menu and click on “Update” button. Selecting “No” will not update the database and redirects back to the homepage

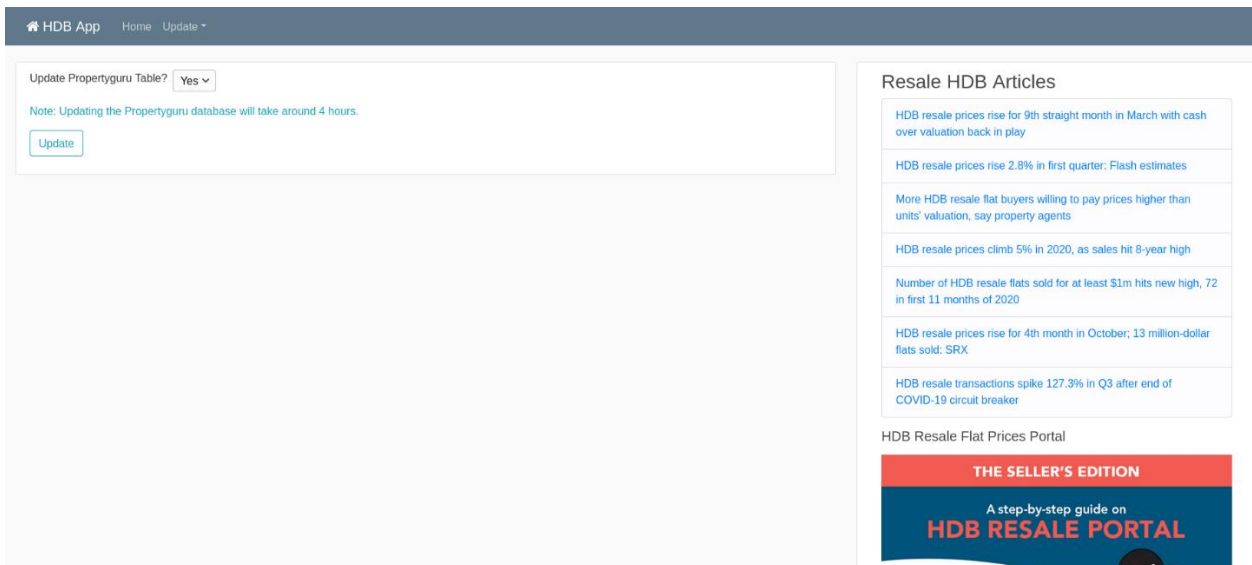


## Part 4 - To update Propertyguru listings

1. Click on “Update” on the navigation bar and click on “Update Propertyguru Table”

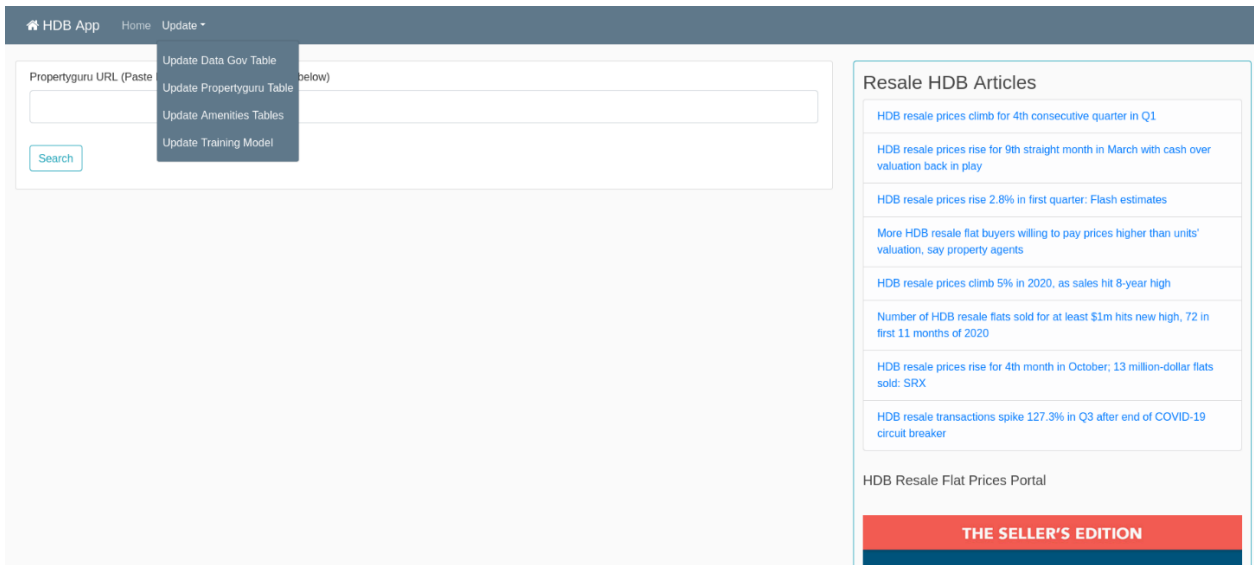


2. Next, select “Yes” in the dropdown menu and click on “Update” button. Selecting “No” will not update the database and redirects back to the homepage

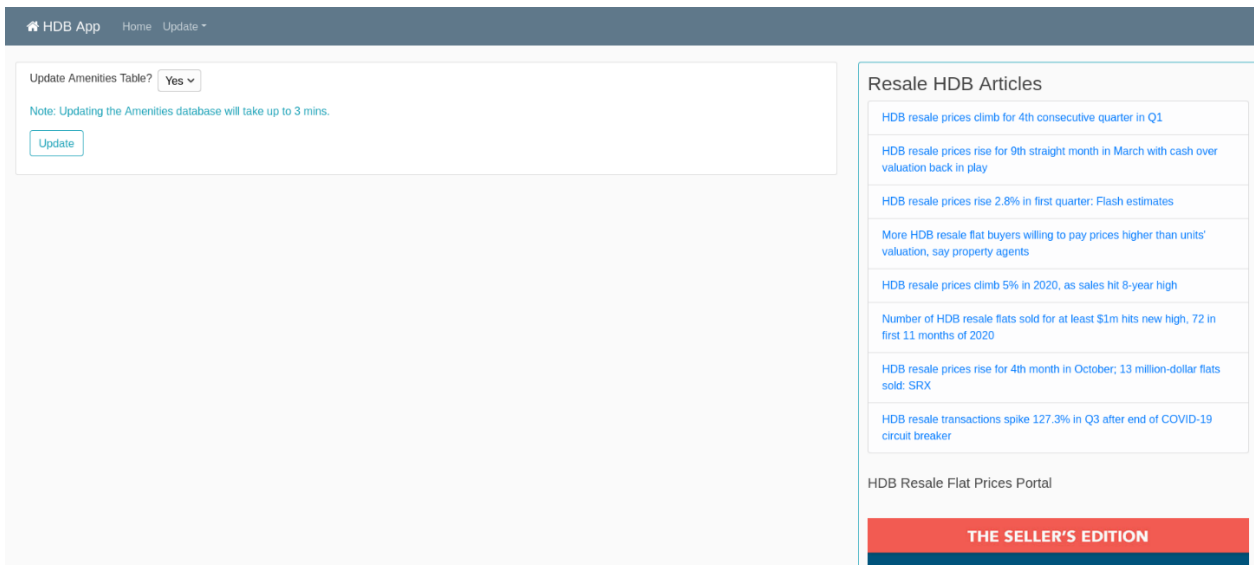


## Part 5 – To update Amenities Table

1. Click on “Update” on the navigation bar and click on “Update Amenities Tables”

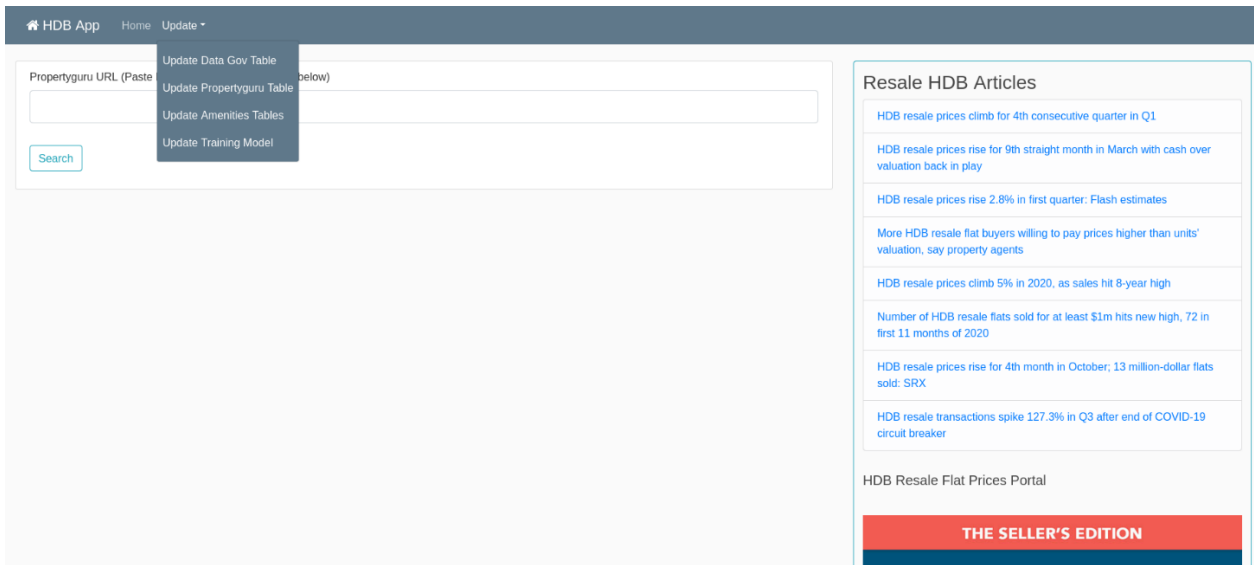


2. Next select “Yes” in the dropdown menu and click on “Update” button. Selecting “No” will not update the database and redirects back to the homepage

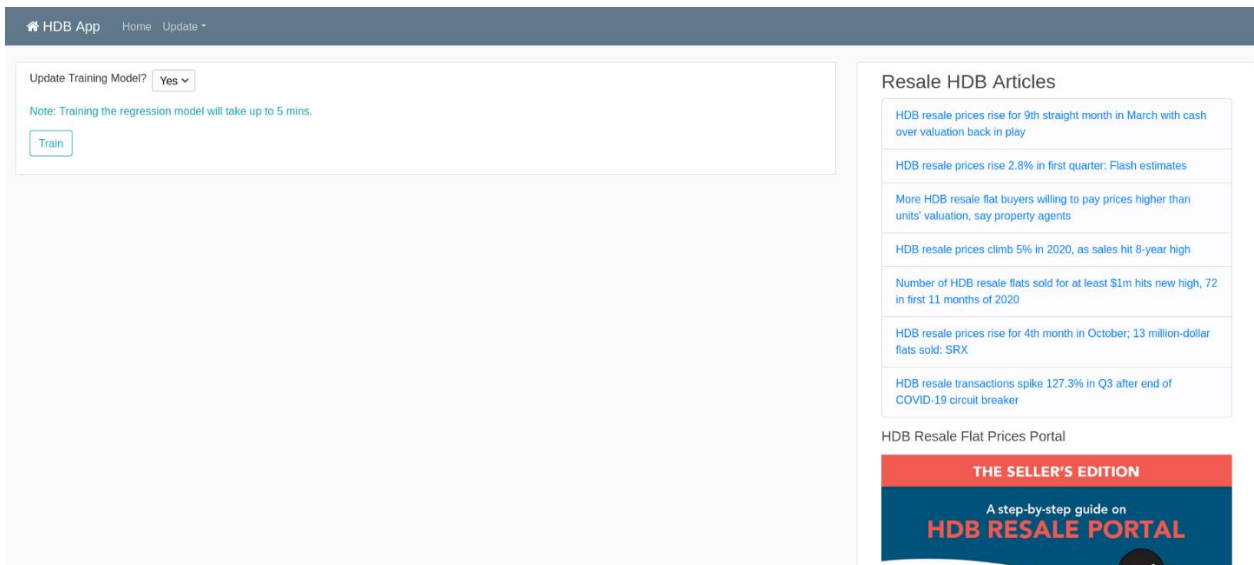


## Part 6 - To train Regression Model

1. Click on “Update” on the navigation bar and click on “Update Training Model”



2. Next, select “Yes” in the dropdown menu and click on “Train” button. Selecting “No” will not update the regression model and redirects back to the homepage







<b>Name:</b>	TOH KAH KHEK
--------------	--------------

### **1. Your personal contribution to the project.**

In this project, I helped to develop the database and the structure of the data flow. This requires me to understand the data required by each function and coordinate with my teammates. To keep the database clean and compact, I am required to review the data to ensure only essential data are kept and null values (if any) are reasonable.

I am also responsible over the frontend development of the project and the integration of the functions developed. As none of the members in the team had frontend development experience, I took up the role by looking into resources for developing frontend languages that works with Python, mainly the Flask and Django library. After discussion with the other members, we decided on Flask as our project are relatively simple and scales well with the resources available with using Flask. From an Udemmy course that I was learning previously, I made use of the Bootstrap (CSS framework) to minimise the effort required for designing the webpages.

### **2. What you have learnt from the project.**

Having acquired knowledge of Python programming before this course, my challenge has always been developing an application with an interface. Through this project, I have acquired the knowledge and skills required to develop an interface for a working application by using available resources on the net such as Bootstrap, Font Awesome and Flask.

To integrate the functions developed by the rest in the team with the Flask library, I have to understand when to execute each function and the inputs required and outputs generated by the functions.

Time management and coordination with other members in the team are also skills that are essential to ensure project milestones are met so that the project can be completed in time and issues along the way are resolved timely.

### **3. How you can apply this in future work-related projects.**

The experience for building this application will enable me to develop meaningful application in my workplace using Python with the help of the Flask library and resources such as Bootstrap. This will enable me to develop application for my team in addition to using Python for data analytics.

In addition, I will be able to apply my experience in future projects as I have a better understanding over how to break down the development of an application into modular functions such that each member in the project understand their role and the manner which the application should be developed e.g the inputs, outputs and data structure.

<b>Name:</b>	JEON SUNGMIN
--------------	--------------

### 1. Your personal contribution to the project.

In this project, I have designed and wrote Python based property collaborative filtering recommender system. The recommender system takes in the user input data, filter property database based on the selected important features which was selected based on the feature importance testing. I have designed the survey which was used to get the ranking of the data features which was used in the score calculation.

I have also created test cases for the website by thoroughly going through the features in the H.R.P.R website. Went through test cases to test out the website's functionality.

Finally, I have edited the videos for the project submission.

### 2. What you have learnt from the project.

Prior to this course, I have very limited knowledge in python coding. With the experience to write the recommender system, I have gained more knowledge and confidence in python code writing.

I have learnt various technics in reasoning system, how and where to apply these technics. With collaboration with the team members, I have learnt all the works that they have done. Web scraping technics and the intelligent price prediction modelling.

I have also learnt how to collaborate effectively online. Due to Covid-19 and the busy work schedule, we were not able to meet face-to-face regularly. We have overcome this through regular Zoom session. Every team member has worked collaboratively and individually to complete the tasks that they were responsible for.

### 3. How you can apply this in future work-related projects.

In my workplace, there exist past historical data on the lot run time in every machines and test types. This data also shares similar characteristics to the data we have used in this project. Upon carefully selecting the important features which will be used as hard filter and applying selective filtering, we will be able to accurately predict the future lot run time. This will then integrate with the auto trigger for the transport. This project's method can help in my workplace to integrate just-in-time lot arrival to avoid loss in OEE and thus, improving the overall efficiency and effectiveness in the productivity.

<b>Name:</b>	AHMED SYALABI SEET
--------------	--------------------

### **1. Your personal contribution to the project.**

I performed exploratory data analysis on the training data and aligned the schema to fit the test data which was web scrapped information from PropertyGuru.com.sg website. The alignment of the data took majority of our time as it was an iterative process that involved all of the members. This was due to some data that was present in the training data was not attainable after web scrapping online data.

Once the data was cleaned and prepared, I delved into model training, obtaining the best possible RMSE scores for our model through random search.

### **2. What you have learnt from the project.**

The main thing I learnt in this project was how to work with others in an IT systems project. It was an eye-opener for someone with a non-IT background.

Another thing was that, it is important to draw out your ideas so that everyone in the team knows what you are talking about or what is your vision for the project. This is important as other people might be seeing things differently from you therefore, we have to benchmark our expectations firsthand before commencing with the project.

Version control is an important tool for collaboration in IT projects and is a must-have skill that I must learn in the near-future.

Lastly, the knowledge learnt in the course provided exposure to the many possibilities and the myriad of data science tools available for system solutions. Data science was not just decision trees and vector machines models after all.

### **3. How you can apply this in future work-related projects.**

After experiencing the basic framework for IT projects, I believe subsequent attempts in projects will get easier. I would like to try building a more complex recommender system next time.

## 6.5 LIST OF ABBREVIATIONS

Abbreviation	Full Representation
API	Application Programming Interface
CSS	Cascading Style Sheets
HDB	Housing & Development Board
HPRP	HDB Resale Prediction and Recommender
HTML	HyperText Markup Language
RMSE	Root Mean Square Error
RPI	Resale Price Index

## 6.6 REFERENCES

1. Anwla, P. K. (2020, August 30). Web Scraping using Selenium with Python! Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/08/web-scraping-selenium-with-python/>
2. Brownlee, J. (2020, August 28). Save and Load Machine Learning Models in Python with scikit-learn. Machine Learning Mastery. <https://machinelearningmastery.com/save-load-machine-learning-models-python-scikit-learn/>
3. Fernando, J. (2020, December 15). Consumer Price Index (CPI) Definition. Investopedia. <https://www.investopedia.com/terms/c/consumerpriceindex.asp>
4. GovTech. (2011, March 29). SDCP MRT Station Point. Data.Gov.Sg. <https://data.gov.sg/dataset/sdcp-mrt-station-point>
5. GovTech. (2021, April 27). HDB Resale Price Index. Data.Gov.Sg. <https://data.gov.sg/dataset/hdb-resale-price-index>
6. Housing & Development Board (HDB). (n.d.). Procedures. <https://www.hdb.gov.sg/business/estate-agents-and-salespersons/selling-a-flat/procedures>
7. J. (2017, June 29). Selecting the best regression model. Kaggle. <https://www.kaggle.com/junkal/selecting-the-best-regression-model>
8. Lee, Y. X. (2019, August 3). Singapore Train Station Coordinates. Kaggle. <https://www.kaggle.com/yxlee245/singapore-train-station-coordinates>
9. Ng, T. (2021, January 12). Exploring Data Gov SG API. Towards Data Science. <https://towardsdatascience.com/exploring-data-gov-sg-api-725e344048dc>
10. Sherlyn, S. (2021, April 8) Straits Times. HDB resale prices rise for 9th straight month in March with cash over valuation back in play <https://www.straitstimes.com/business/property/hdb-resale-prices-rise-for-9th-straight-month-in-march-with-cash-over-valuation>
11. T. (n.d.). HDB flat types and their sizes. The World of Teoalida. <https://www.teoalida.com/singapore/hdbflattypes/>
12. URA. (n.d.). List of Postal Districts. [https://www.ura.gov.sg/realEstateIIWeb/resources/misc/list\\_of\\_postal\\_districts.htm](https://www.ura.gov.sg/realEstateIIWeb/resources/misc/list_of_postal_districts.htm)
13. Wikipedia. (n.d.). List of shopping malls in Singapore. [https://en.wikipedia.org/wiki/List\\_of\\_shopping\\_malls\\_in\\_Singapore](https://en.wikipedia.org/wiki/List_of_shopping_malls_in_Singapore)