Capstone Project — The Battle of Neighbourhoods

Title: Assist Mr Edwin to find Best Deal Flat in Singapore (Prepare By Khai Seng)

Introduction
Introduction

Assumingly I am a Real Estate Agent in Singapore assisting my Client Mr Edwin to Find a HDB(Housing Development Board) Flat in Singapore

Problem

My Client Mr Edwin ***Gave me a Stringent Expectation as below :***

1. My client looking for a 4 Room HDB Flat (Can be Model A / ModelA2 / DBSS/ Simplified Etc)
2. Area that My Client looking at (Toa Payoh / Ang Mo Kio / Hougang)
3. Should Be Nearby Train Station (MRT)
4. Should have HyperMart(Fair Price/ Giant / Cold Storage / Sheng Siong) Near By
5. Leasing Left Should be be more that t > 70 Years (In Singapore Government Build HDB Flat have a Leasing Years of 99 Years)
6. Slightly below Market Rate Average Price

Data

For this project we need the following data :

- HDB Resale Flat Historical Transaction Data
  - Data source : Data From Singapore Government
  - Description : This data set contains the required information. And we will use this data set to Extract Flat Resale Price
- One Map API to Extract HDB Flat Coordinate
  Data source :One Map API From Singapore Government
- NearBy Venue Using Foursquare API
  - Data source : Fousquare API
  - Description : By using this api we will get all the venues in each neighborhood.

Approach

->Using data.gov.sg API

->Using One Map API we will find all HDB Flat Coordinates

->Filter out all unwanted Neighbourhood , only left (Ang Mo Kio, Toa Payoh, Hougang)

->Filter Result Base on Client Mr Edwin Requirement

3. Methodology section

The Methodology section will describe the main components of our analysis and prediction system. The Methodology section comprises four stages:

1. **Collect Inspection Data**
2. **Explore and Understand Data**
3. **Data preparation and preprocessing**
4. **Modeling**

3.1 Collect Inspection Data

I collected the data from Data.gov.sg , Convert into CSV and Uploaded to Github

Read CSV Data and Convert it into Data Frame using Pandas

3.2 Explore and Understand the data

Read HDB FLAT Resale Price Data From 2017 till Recent 2021

Out[2]:

month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	remaining_lease	resale_price
2017-01	ANG MO KIO	2 ROOM	406	ANG MO KIO AVE 10	10 TO 12	44.0	Improved	1979	61 years 04 months	232000.0
2017-01	ANG MO KIO	3 ROOM	108	ANG MO KIO AVE 4	01 TO 03	67.0	New Generation	1978	60 years 07 months	250000.0
2017-01	ANG MO KIO	3 ROOM	602	ANG MO KIO AVE 5	01 TO 03	67.0	New Generation	1980	62 years 05 months	262000.0
2017-01	ANG MO KIO	3 ROOM	465	ANG MO KIO AVE 10	04 TO 06	68.0	New Generation	1980	62 years 01 month	265000.0
2017-01	ANG MO KIO	3 ROOM	601	ANG MO KIO AVE 5	01 TO 03	67.0	New Generation	1980	62 years 05 months	265000.0
...	...	...	...	...	...	...	...	...	...	...
2021-04	YISHUN	EXECUTIVE	326	YISHUN RING RD	10 TO 12	146.0	Maisonette	1988	66 years 04 months	650000.0
2021-04	YISHUN	EXECUTIVE	360	YISHUN RING RD	04 TO 06	146.0	Maisonette	1988	66 years 04 months	645000.0
2021-04	YISHUN	EXECUTIVE	326	YISHUN RING RD	10 TO 12	146.0	Maisonette	1988	66 years 04 months	585000.0
2021-04	YISHUN	EXECUTIVE	355	YISHUN RING RD	10 TO 12	146.0	Maisonette	1988	66 years 08 months	675000.0
2021-04	YISHUN	EXECUTIVE	277	YISHUN ST 22	04 TO 06	146.0	Maisonette	1985	63 years 05 months	625000.0

As My Client only looking for Area in (Toa Payoh, Ang MoKio and Hougang) Thus Drop all Area Except the Mentioned Area

```
df_3_town =df_Origin[df_Origin['town'].str.contains('ANG MO KIO|TOA PAYOH|HOUGANG',case=False, regex=True)].reset_index(drop=Fals
df_3_town
```

Out[4]:

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	remaining_lease	resale_price
0	2017-01	ANG MO KIO	2 ROOM	406	ANG MO KIO AVE 10	10 TO 12	44.0	Improved	1979	61 years 04 months	232000.00
1	2017-01	ANG MO KIO	3 ROOM	108	ANG MO KIO AVE 4	01 TO 03	67.0	New Generation	1978	60 years 07 months	250000.00
2	2017-01	ANG MO KIO	3 ROOM	602	ANG MO KIO AVE 5	01 TO 03	67.0	New Generation	1980	62 years 05 months	262000.00
3	2017-01	ANG MO KIO	3 ROOM	465	ANG MO KIO AVE 10	04 TO 06	68.0	New Generation	1980	62 years 01 month	265000.00
4	2017-01	ANG MO KIO	3 ROOM	601	ANG MO KIO AVE 5	01 TO 03	67.0	New Generation	1980	62 years 05 months	265000.00

3.3 Data Preparation and preprocessing

Drop HDB Resale Flat Data before 2021

Filtered out Flat Type only Remain Flat Type == 4 Room

Get the Latest Price at Year 2021

```
In [8]: df_3_town_4R_21 =df_3_town_4R.loc[df_3_town_4R["month"]>="2021-01"]
```

```
In [9]: df_3_town_4R_21.head()
```

Out[9]:

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	remaining_lease	resale_price
3771	2021-01	ANG MO KIO	4 ROOM	305	ANG MO KIO AVE 1	07 TO 09	97.0	New Generation	1977	55 years 06 months	448000.0
3772	2021-01	ANG MO KIO	4 ROOM	413	ANG MO KIO AVE 10	10 TO 12	92.0	New Generation	1979	57 years 08 months	420000.0
3773	2021-01	ANG MO KIO	4 ROOM	414	ANG MO KIO AVE 10	01 TO 03	92.0	New Generation	1979	57 years 09 months	375000.0
3774	2021-01	ANG MO KIO	4 ROOM	415	ANG MO KIO AVE 10	07 TO 09	92.0	New Generation	1979	57 years 08 months	420000.0
3775	2021-01	ANG MO KIO	4 ROOM	547	ANG MO KIO AVE 10	04 TO 06	92.0	New Generation	1981	59 years	370000.0

Filter Out Remaining Lease Year which is Less than 70 Years

Get Lease Remaining More than or equal to 70 Years

```
In [17]: #pd.set_option('display.max_rows', None)
         #pd.set_option('display.max_columns', None)
         #pd.set_option('display.width', None)

         df_filtered = df_3_town_4R_21[df_3_town_4R_21['remain_lease']>=70]
         df_filtered.head()
```

Out[17]:

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	remain_lease	resale_price
3784	2021-01	ANG MO KIO	4 ROOM	619	ANG MO KIO AVE 4	01 TO 03	107.0	Model A	1996	74	508000.0
3786	2021-01	ANG MO KIO	4 ROOM	700B	ANG MO KIO AVE 6	22 TO 24	90.0	Model A	2003	81	690000.0
3788	2021-01	ANG MO KIO	4 ROOM	700C	ANG MO KIO AVE 6	13 TO 15	85.0	Model A	2003	81	665000.0
3790	2021-01	ANG MO KIO	4 ROOM	315A	ANG MO KIO ST 31	04 TO 06	90.0	Model A	2006	84	580000.0
3791	2021-01	ANG MO KIO	4 ROOM	309A	ANG MO KIO ST 31	04 TO 06	90.0	Model A	2006	84	505055.0

```
In [18]: df_filtered.shape
```

Out[18]: (183, 11)

Get Flat Coordinate Using One Map API

Now We have the Address will Use OneMap Api to get Coordinate

```
In [23]:
# Declare eMpty List
latitude = []
longitude = []
blk_no = []
road_name = []
postal_code = []
address = []
count = 0

for row in range(len(address_list)):
    #formulate query string
    query_address = address_list[row]
    # Query the API
    query_string='https://developers.onemap.sg/commonapi/search?searchVal='+str(query_address)+'&returnGeom=Y&getAddrDetails=Y'
    resp = requests.get(query_string)

    #Convert JSON into Python Object
    data_geo_location=json.loads(resp.content)
    if data_geo_location['found'] != 0:
        latitude.append(data_geo_location['results'][0]['LATITUDE'])
```

Combine the Coordinates with the Main Data frame

Combine list_of_coordinates (Drop address Column) Join by blk_no with block df_filtered_sorted

[32]:
```
df_hdb_Coor = pd.merge(df_filtered_sorted,df_coordinates, how='left', left_on = 'block', right_on = 'blk_no')
# remove the "address" column
df_hdb_Coor.drop("address_x", axis=1, inplace=True)
df_hdb_Coor.drop("blk_no", axis=1, inplace=True)

df_hdb_Coor.head()
```

t[32]:

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	remain_lease	resale_price	latitude
0	2021-01	ANG MO KIO	4 ROOM	619	ANG MO KIO AVE 4	01 TO 03	107.0	Model A	1996	74	508000.0	1.3785880768512
1	2021-04	ANG MO KIO	4 ROOM	588D	ANG MO KIO ST 52	16 TO 18	91.0	DBSS	2011	89	770000.0	1.37175400663025
2	2021-02	ANG MO KIO	4 ROOM	596A	ANG MO KIO ST 52	13 TO 15	91.0	Model A	2002	80	665000.0	1.37243023974421
3	2021-02	ANG MO KIO	4 ROOM	588B	ANG MO KIO ST 52	22 TO 24	90.0	DBSS	2011	89	768000.0	1.37197571261128
4	2021-02	ANG MO KIO	4 ROOM	596C	ANG MO KIO ST 52	22 TO 24	90.0	Model A	2002	80	680888.0	1.37193571932412

Get the Train Coordinate Information Using One Map API

Convert it into CSV and Upload to GitHub

Call the File From Github

In [34]:
```
#The CSV File copy form data.gov.sg

url_1 = 'https://raw.githubusercontent.com/khaiseng82/mrt/main/mrt_lrt_data.csv'
df_mrt = pd.read_csv(url_1, index_col=0)

df_mrt
```

Out[34]:

station_name	type	lat	lng
Jurong East	MRT	1.333207	103.742308
Bukit Batok	MRT	1.349069	103.749596
Bukit Gombak	MRT	1.359043	103.751863
Choa Chu Kang	MRT	1.385417	103.744316
Yew Tee	MRT	1.397383	103.747523
Kranji	MRT	1.425302	103.762049
Marsiling	MRT	1.432579	103.774150
Woodlands	MRT	1.436984	103.786406
Admiralty	MRT	1.436984	103.786406
Sembawang	MRT	1.449133	103.820060
Yishun	MRT	1.429666	103.835044

After You the MRT Coordinate Use the geodesic to Measure the Distance between the Flat Unit to the Nearest List Of MRT (Train Station)

In [40]:
```
from geopy.distance import geodesic

list_of_dist_mrt = []
min_dist_mrt = []

# list_of_coordinates -> Flat Unit Coordinates
# Loop Theough the Flat Coordinates
for origin in list_of_coordinate:
    # Loop Through All MRT Station in the List in -> "list_of_mrt_coordinates"
    for destination in range(0, len(list_of_mrt_coordinates)):
        #geodesic is a  Function -> is use to Calculate the Distance between The Flat Unit
        list_of_dist_mrt.append(geodesic(origin,list_of_mrt_coordinates[destination]).meters)
        # Append the Result into (list_of_dist_mrt) with every Loop
        #Get the  min () Minimum Distance
    shortest = (min(list_of_dist_mrt))
    #After getting the Shortest Distance append to min_dist_mrt list
    min_dist_mrt.append(shortest)
    # Clear the List  list_of_dist_mrt loop again
    list_of_dist_mrt.clear()
```

In [41]:
```
min_dist_mrt
```

Out[41]:
```
[782.5324575496927,
 514.8306095744726,
 306.0736916342701,
 451.38025278018864,
 220.12880408769516,
 792.0099256205582,
 726.2540809755077,
 720.1050737040086,
 782.5324575496927,
 782.5324575496927.
```

Append the Distance Difference Result Back to the Main Data Frame

Append Distance between MRT and Flat Unit List to df_hdb_Coor Data Frame

In [42]:
```
len(min_dist_mrt)
```

Out[42]: 189

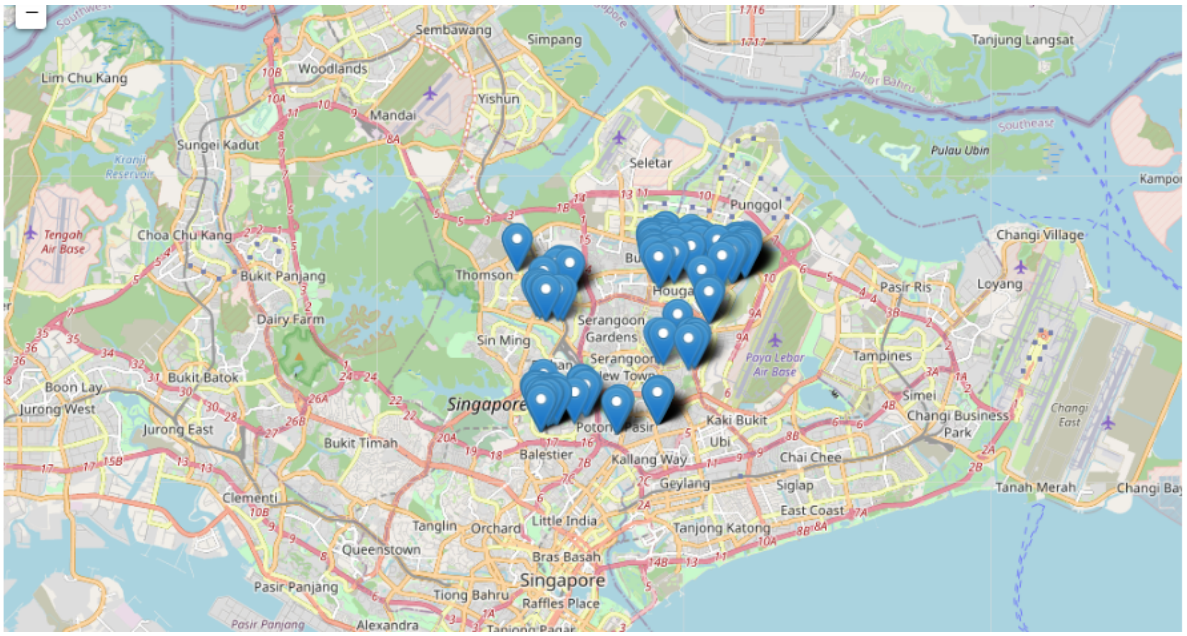
Can Only Merge if Same length

In [43]:
```
df_hdb_Coor.shape
```

Out[43]: (189, 16)

In [44]:
```
df_hdb_Coor['min_dist_mrt'] = min_dist_mrt
```

Use FourSquare to Create Flat Unit clusters



3.4 Reduction Modeling

Now We have Scan through the Data Frame and Got 189 Potential Properties to be recommended to My Client .

My Job now is to reduce the Number and Provide my client the Best Recommendation, One Flat for each Single Area(Toa Payoh / Ang Mo kio / Hougang)

Frame Reduction Base On Distance FLat unit to Nearest HyperMart

In [55]:
```
hdb_venues_tpy.head()
```

Out[55]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	t	1.33626413671063	103.844907413472	Creamier	1.337754	103.843380	Ice Cream Shop
1	t	1.33626413671063	103.844907413472	Uggli Muffins	1.338179	103.844616	Bakery
2	t	1.33626413671063	103.844907413472	Toa Payoh Lor 1 Market & Food Centre	1.337999	103.844623	Food Court
3	t	1.33626413671063	103.844907413472	Hup Chong Hakka Yong Dou Foo	1.338293	103.844422	Hakka Restaurant
4	t	1.33626413671063	103.844907413472	Hua Fong Kee Roasted Duck	1.338283	103.844420	BBQ Joint

We Are Only Interested in the Toa Payoh SuperMarket Cateogory

In [56]:
```
hdb_venues_tpy_sp = hdb_venues_tpy[hdb_venues_tpy['Venue Category'].str.contains("Supermarket")].reset_index(drop=True)
hdb_venues_tpy_sp
```

Out[56]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	t	1.33626413671063	103.844907413472	NTUC FairPrice	1.332965	103.847317	Supermarket
1	p	1.33279631135029	103.845588638625	NTUC FairPrice	1.332965	103.847317	Supermarket
2	a	1.33270900444626	103.867989929028	NTUC FairPrice	1.331637	103.868364	Supermarket
3	y	1.33180078265389	103.867242849282	NTUC FairPrice	1.331637	103.868364	Supermarket
4	o	1.33466409464552	103.846451369821	NTUC FairPrice	1.332965	103.847317	Supermarket

Out[66]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	h	1.38198270653355	103.881217675808	Kampung Lorong Buangkok	1.383242	103.878222	Historic Site
1	h	1.38198270653355	103.881217675808	McDonald's	1.384216	103.880951	Fast Food Restaurant
2	h	1.38198270653355	103.881217675808	Bus Stop 66489 (Blk 969)	1.379076	103.879003	Bus Stop
3	h	1.38198270653355	103.881217675808	Fitness Corner	1.378371	103.880273	Gym / Fitness Center
4	h	1.38198270653355	103.881217675808	Bus Stop 67079	1.382573	103.876795	Bus Station

We Are Only Interested in the Hougang SuperMarket Cateogory

In [67]:
```
hdb_venues_hg_sp = hdb_venues_hg[hdb_venues_hg['Venue Category'].str.contains("Supermarket")].reset_index(drop=True)
hdb_venues_hg_sp
```

Out[67]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	n	1.37499383603396	103.883486736367	FairPrice Xtra	1.375279	103.879216	Supermarket
1	g	1.36987745772011	103.891412815739	NTUC FairPrice	1.371776	103.894068	Supermarket
2	g	1.36987745772011	103.891412815739	NTUC FairPrice	1.372699	103.893807	Supermarket

Hougang : Calculate Flat Distance to Supermarket (Get the Minimum Distance)

In [68]:
```
hdb_venues_hg_la = hdb_venues_hg ["Neighborhood Latitude"].tolist()
hdb_venues_hg_lo = hdb_venues_hg ["Neighborhood Longitude"].tolist()
```

In [77]:
```
hdb_venues_amk.head()
```

Out[77]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	a	1.3785880768512	103.838639699065	香记粥 (Xiang Ji Porridge)	1.381202	103.840782	Chinese Restaurant
1	a	1.3785880768512	103.838639699065	Ho-Ree Roasted Food 好味烧腊	1.381160	103.840977	Chinese Restaurant
2	a	1.3785880768512	103.838639699065	Mr Teh Tarik Eating House	1.378301	103.836725	Food Court
3	a	1.3785880768512	103.838639699065	Shanghai Ren Jia 上海人家	1.377068	103.841329	Shanghai Restaurant
4	a	1.3785880768512	103.838639699065	Charcos The Flaming Chicken	1.378220	103.836871	Australian Restaurant

We Are Only Interested in Ang Mo Kio SuperMarket Cateogory

In [78]:
```
hdb_venues_amk_sp = hdb_venues_amk[hdb_venues_amk['Venue Category'].str.contains("Supermarket")].reset_index(drop=True)
hdb_venues_amk_sp
```

Out[78]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	g	1.37243023974421	103.850949173949	FairPrice Xtra	1.369279	103.848886	Supermarket
1	g	1.37243023974421	103.850949173949	NTUC FairPrice	1.371507	103.847082	Supermarket

3.5 Provide Recommendation to Client

Now We Have reduce the Data Frame Size ,

Final Steps is to Provide Client My Recommendation

Toa Payoh Summary : Base on the Result I would Recommend Block 79 A to My Client
Reason : (1)Cheapest Resale Price

(2)Distance to MRT Not the Nearest but within the Top 5

(3)Distance to Market Not the Nearest but within the Top 5

(4) LeaseCOst per Year is Cheaper

Ang Mo Kio

AMK Summary : Base on the Result I would Recommend Block 596 A to My Client
Reason : (1)Cheapest Resale Price

(2)Distance to MRT Not the Nearest but within the Top 5

(3)Distance to Market Not the Nearest but within the Top 5

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	remain_lease	resale_price	la
75	2021-01	HOUGANG	4 ROOM	220	HOUGANG ST 21	13 TO 15	104.0	Model A	1992	70	590000.0	1.356934396

Hougang Summary : Base on Observation to the Table I would Recommend Block 220 A to My Client
Reason : (1)Cheapest Resale Price

(2)Distance to MRT Not the Nearest but within the Top 5

(3)Distance to Market Not the Nearest but within the Top 5

For Detail Explanation please refer my Source Code at Git Hub

Conclusion and Result:

For Toa Payoh Area -> Block 79A will be the Best Deal

For Ang Mo Kio Area -> Block 596A will be the Best Deal

For Hougang Area -> Block 220A will be the Best Deal