**Analysis HDB Resale Price vs Venues Data in Singapore**

**Applied Data Science Capstone by Sazili Muhammad**

**Table of contents**

* [1. Introduction: Business Problem](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#introduction)
* [2. Data](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#data)
* [3. Methodology](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#methodology)
* [4. Analysis](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#analysis)
* [5. Results and Discussion](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#results)
* [6. Conclusion](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#conclusion)
* [7. References](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#references)

**1. Introduction**

**1.1 Background**

Majority of Singapore population are living in the public housing, in which the developments are publicly governed and developed by the state **Housing and Development Board (HDB)** under a 99-year lease. These flats are located in housing estates, which are self-contained satellite towns with well-maintained schools, supermarkets, malls, community hospitals, clinics, hawker centres (food court) and sports and recreational facilities. Every housing estate includes **MRT stations** and bus stops that link residents to other parts of the city-state.

As compared to most parts of the world, public housing in Singapore is not ostracised by a wide majority of the population and its government, and acts as a necessary and vital measure to provide immaculate and safe housing surrounded by public amenities at affordable prices, especially during its rapid development and industrialisation in the early years of independence. It is also meant to foster social cohesion between social classes and races of Singapore, and prevent neglected areas or districts and ethnic enclaves from developing. As such, **it is considered a unique part of Singaporean culture and identity**, being commonly associated with the country.

There are a large variety of flat types and layouts which cater to various housing budgets. HDB flats were built in mind to primarily provide affordable housing for **Singaporeans/Permanent Residents** and their purchase can be financially aided by the **Central Provident Fund (CPF)** in addition to various grants. Due to changing demands, HDB introduced the Design, Build and Sell Scheme to produce up-market public housing developments.

New public housing flats are strictly only eligible for purchase towards Singaporean citizens. The housing schemes and grants available to finance the purchase of a flat are also only extended to households owned by Singaporeans, while Permanent Residents do not get any housing grants or subsidies from the Singaporean government and could only purchase **resale flats from the secondary market at a market price**. Such policies have helped **Singapore reach a home-ownership rate of 91%, one of the highest in the world**. In 2008, Singapore was lauded by the United Nations Habitat's State of the World's Cities report as the only slum-free city in the world. [[1]](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#references)

**1.2 Problem**

HDB flats surrounded by various amenities such as hawker centers, coffee shop, school, carpark, community centers, shopping center/malls, playground, sport center. Those amenities impact to housing resale price fluctuation and hence resale prices among towns in Singapore are not balanced. Some town are highly popular and more matured, makes increasing of resale price faster than other town.

This situation makes house buyer has difficulties and limitation to purchase flats as they need. The challenges also facing to business owner to invest which location has more profitable and more promising. For government and regulation makers will have difficulties to make fair decision to build facilities in certain location.

**1.3 Purpose**

This project will try to analyze the housing resale price vs various amenities (venues) in different angle. This project will give better insight to house buyer, business owner, or government/regulator to have better decision to invest and improve in various location across the island.

**2. Data aquisition and data cleaning**

**2.1 Data Sources**

The dataset for this project are captured from:

1. Data.gov.sg [[2]](https://dataplatform.cloud.ibm.com/data/jupyter2/runtimeenv2/v1/wdpx/service/notebook/conda2py3783032163a51e4806a12333ad0b448acb/dsxjpy/VdC8FFUw4cGfOh4NJOx27w:OnxLWmSZdJzmdAptA4cqADNJtOT_MgOP28PV2Jl73gnwORDDyrg3d3WamySQOL0q7vlzZVY/container/notebooks/679d1ae5-2951-4cc7-ada4-bce841fb559a?api=v2&project=83032163-a51e-4806-a123-33ad0b448acb#references). This is Singapore government body which release data for public purpose.
2. Google Maps to capture latitude and longitude.

**2.2 Data cleaning and data selection**

Originally this data is very detail which comprises resale price from all towns in Singapore from Jan-2017 until Oct-2020 which has in detail block number, flat type, and flat floor. Since this project is not purpose into detail flat address but instead only interested on towns. So that, I clean up and summarize the dataset into average of housing resale price across all Singapore towns for year 2020 (since Jan-Oct).

Note that, all the prices is in Singapore dollars (SGD).

The total Singapore towns are 26 towns, as follows:



In order to use Foursquare to find venues data, I capture latitude and longitude for each Singapore town from Google Maps, as shown below:



Combine original data from Data.gov.sg and Google maps (latitude/longitude), hence getting final dataset as follows:



Hence, I will use this dataset above for further analysis for this project

**3. Methodology**

The methodology in this project, I am using clustering approach based on the most common venues in various towns in Singapore. The venues data is getting from Foursquare website.

First, I will Singapore town data including the latitude and longitude.

Then, I will get venues data from Foursquare for each town within 1 km and limit to 1000 venues. This data collection based on the latitude and longitude from each town.

I create bar chart to show total venues for each town and bar chart to show housing resale price for each town, side-by-side, to visualize the correlation between total venues effected the housing resale price.

Another analysis, I am using clustering which finding the best cluster number using Elbow Method to find optimal k.

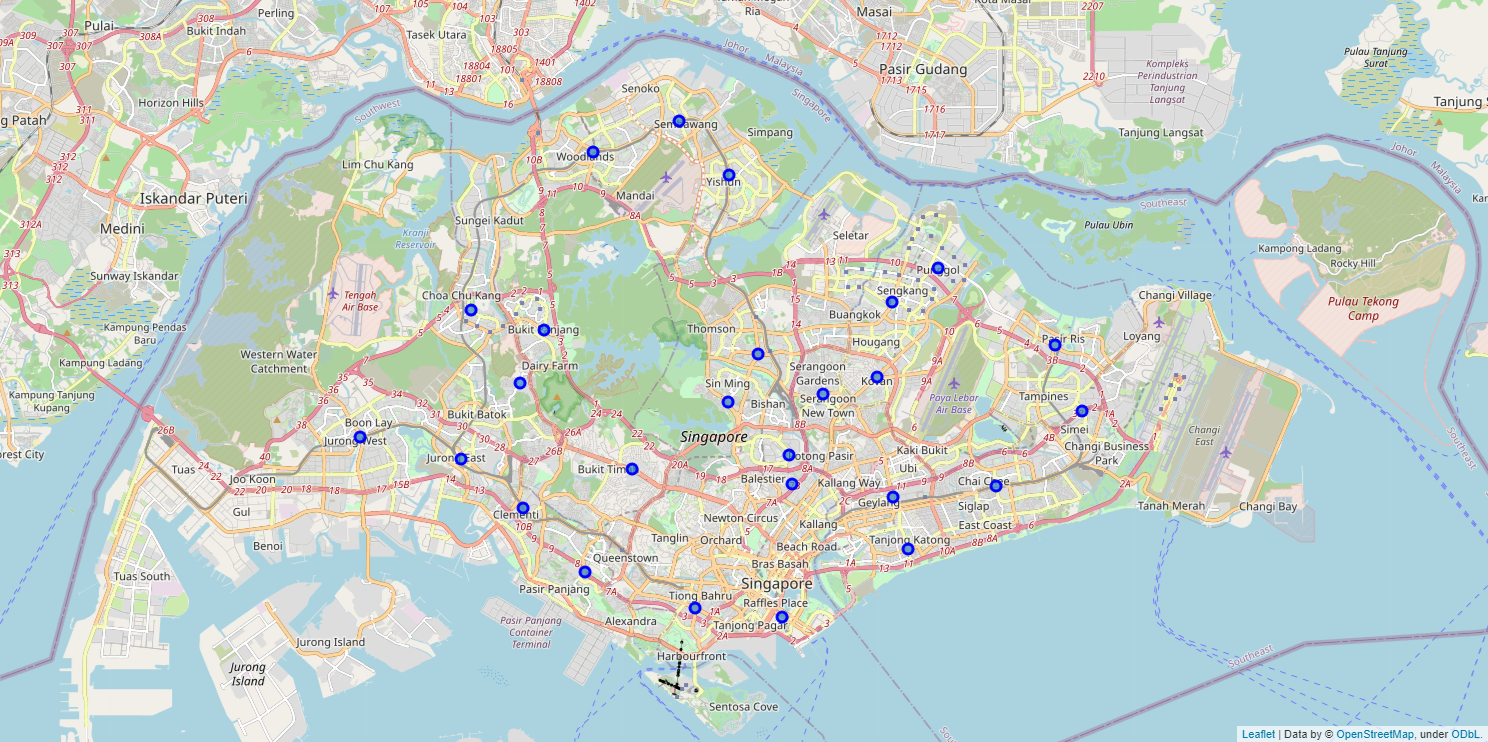
This k-cluster will be used to defined clustering for all towns and visualize in map.

The detail analysis will show as below.

**4. Analysis**

In section Data, I grabbed the Singapore town data including latitude and longitude.

Now, I will visualize all the Singapore town in the map as per starting analysis.



The venues data collection from Foursquare is processed using this steps below. For each town will search venues in radius 1km which limited maximum up to 1000 venues.

Below is the result of venues data for each town.

