Understanding the effect of Housing Prices based on Clustered Neighborhoods

in King County, USA

# Introduction

Having one’s own home is probably the desire of every individual on this planet. However, buying a house is not an easy feat and one needs to put a lot of considerations before the deciding to buy one, the price of a house being the most important.

Housing prices can vary on a lot of aspects, such as the size of the house, the number of bedrooms or bathrooms, etc. Also, the neighborhood where the house is located can count heavily in determining the price of a house as well.

In this Capstone Project, we would try to understand what impact the locality has in determining the housing prices in King County, USA. We would use Foursquare APIs to fetch the local venues for neighborhoods and then cluster the neighborhoods based on few top localities for each neighborhood and then try to understand if there is any correlation between these clusters and housing prices.

This estimation will certainly help people trying to find the best value for their money they would be investing in buying a house. Also, it will help understand how the housing prices vary based on the clusters of neighborhoods based on top venues in a locality.

# Data

* We will be using the following dataset from Kaggle to get details regarding housing prices for King County, USA

<https://www.kaggle.com/harlfoxem/housesalesprediction>

Below is the list of columns that the dataset has:

id: a notation for a house

date: Date house was sold

price: Price is prediction target

bedrooms: Number of Bedrooms/House

bathrooms: Number of bathrooms/House

sqft\_living: square footage of the home

sqft\_lot: square footage of the lot

floors: Total floors (levels) in house

waterfront: House which has a view to a waterfront

view: Has been viewed

condition: How good the condition is (overall)

grade: overall grade given to the housing unit, based on King County grading system

sqft\_above: square footage of house apart from basement

sqft\_basement: square footage of the basement

yr\_built: Built Year

yr\_renovated: Year when house was renovated

zipcode: zip

lat: Latitude coordinate

long: Longitude coordinate

sqft\_living15: Living room area in 2015(implies-- some renovations) This might or might not have affected the lotsize area

sqft\_lot15: lotSize area in 2015(implies-- some renovations)

* It is important to note the column zipcode in the above data set. A zipcode can be considered as the representation of sub-areas within King County. For each of the zipcodes, we can easily find the coordinates using geolocator. Later, we can feed it to Foursquare API to fetch the nearby venues.
* Hence, we will also be using Foursquare APIs to retrieve the JSON output for each zipcode and then use it to cluster our datasets.