Intro to DS Capstone Project - Project Proposal

Vidhyapriya Ramaswamy

October 27, 2017

## 1. Problem:

Using the existing sold homes data for King County in Washington state which includes the homes sold between May 2014 and May 2015, I will be doing an analysis on the following.

* How much does the renovation project affect the price of the homes?
* How much does renovation on a house along with the year built affect the pricing?
* How much does the condition of the house along with the year built affect the pricing?
* Though the data is for only one county, we all know the pricing differs widely for different zip codes. What are the top 10 zip codes in the highest and least priced homes based on average price of homes in each zip codes.
* Predict the pricing of homes based on above factors.

## 2. Data

The data is in excel sheet format, which has been retrieved from Kaggle.com

* Data - King County Homes sold between May 2014 and May 2015
* Provider - King County
* Link - <https://www.kaggle.com/chaitanya94/house-sales-in-king-county/data>

## 3. Project Outline

The first step of the project is to prepare the dataset. Though the data is in excel sheet and looks relatively straightforward, I will apply the data wrangling techniques to clean /modify data as needed.  
The next step would be to apply the exploratory analytical skills learnt to see how the price of the house is changing based on the year built, based on the renovation made, or based on the condition of the house.

Finally, based on how these factors affect the prices of the houses, predict the housing price.

* Find out the price per sq. ft of houses based on the year built, (1900-1925,1926-1950,1951-1975,1976-2000 ,2001- 2010, 2011 - current) without renovation and with renovation, how much percentage of houses in each group (based on the year built) is renovated and predict how the renovation is helping the sellers and how much more price on average are the sellers getting compared to the houses without renovation.
* Find out how the Condition of houses based on the year built (1900-1925,1926-1950,1951-1975,1976-2000 ,2001- 2010, 2011- current) is affecting the price per square feet of the house.
* Find out average price of houses for each zip code and top 10 zip codes with higher average price and top 10 zip codes with lower average price

## 4. Deliverables:

The deliverables for this project are as follows-

* Pricing (Price/sq.ft) comparison report based on the year built (1900-1925,1926-1950,1951-1975,1976-2000 ,2001- 2010, 2011- current) with and without renovation.
* Pricing (Price/sq.ft) comparison report based on the year built (1900-1925,1926-1950,1951-1975,1976-2000 ,2001- 2010, 2011- current) and Condition of house.
* Predict the price of the house based on the above analysis.

This will include the code and the reports generated.