**King County Housing Dataset**

City University of Seattle

CS612 Summer 2020

Group Members:

Anil Erturk [erturkanil@cityuniversity.edu](mailto:erturkanil@cityuniversity.edu)Minh Nguyen [minh.truong@cityuniversity.edu](mailto:erturkanil@cityuniversity.edu)

Faculty Advisor: Professor Brian Maeng

**Project Summary (Anil)**

The project aims to answer the question what features influence house prices in King County the most, and whether these features affects the price based on different classification

**Project Description (Minh)**

The King County Housing data set was collected from May 2014 to May 2015 which specifies relevant information of each listing such as price, size, views, selling date, posting date and geographical figures. The project aims to find a correlation between each of these features to the price of the house and try to determine which features have a significant impact on a house price, and which doesn’t. We will also examine the distribution of each of these features to see how normally distributed they are.

**Usefulness (Minh)**

This project mostly serves the purpose of exploratory data analysis whose main goal is to find meaningful information about the dataset. For this project in particular, it will help the audience have a better understanding of how to evaluate a house in King County. This will especially benefit real estate agents or people with the goal to sell houses. This insight will also help data scientist make proper preparation for the dataset to create a house price prediction model

**Dataset (Minh)**

The dataset is downloaded from Kaggle. It’s the property listing dataset collected from May 2014 to May 2015. The data is extremely biased due to being collected within only 1 year. It contains mostly numeric features and is used mostly for regression problems