

Buying A House In Pittsburgh

—Analyzing Property Prices in Pittsburgh

Congyu Wang

July 16, 2020

1 Introduction

There are many factors to think over before purchasing a property. The areas of houses are surely an important factor to take into consideration. A family with children would need a bigger house than a single person. The price of a property is largely determined by the size of the property, but there are other factors affecting the total price as well. For instance, houses of a same size located in downtown are usually more expensive than those located near the periphery of downtown area. Except for the proximity to city center, different access to public transportation, schools, and hospitals may also affect the house price per unit. The first task of this project is to analyze how different factors affect property prices.

Buying a property can be an exhausting task in real life. When people are buying houses, it is often difficult for them to obtain important information all at once. People spend a lot of time visiting different houses and apartments, which is exhausting and time-consuming. *Foursquare* as a geo-location provides therefore provides very needy information for choosing a house, as it has information about different venues near a house with location information associated to them. By utilizing *Foursquare API*, it would save house-seekers a great amount of time and effort to pin down a few number of properties that satisfy the specific demand of this person. The second task of this project is thus to assume the role of a property buyer, and attempt to find a small number of target properties based on the many useful information provided by *Foursquare*.

The city chosen for this task is Pittsburgh. The local government of Pittsburgh has many useful public datasets provided for completing our task. The specific data used will be discussed more carefully in the data section.

2 Data

The Allegheny County Information Portal provides all house transaction data from 2016 to 2019, which can be used for analyzing house prices associated with each neighborhood. The data include transaction date, address of the property, municipality, price, and types of transactions. Some of the transactions will not be used for the estimation of

house prices. For example, there are government transfer, love and affection sale. Such transactions do not reflect a valid market price.

From another public data source—Burgh’s eye, I obtained the crime records in Pittsburgh from year 2016 to 2019. Using this data, it is possible to give an estimation of the safety and security of a neighborhood, and we shall see whether this is related to the house price of the specific neighborhood.

The most important data source is *Foursquare*. This API provides information such as the location of Arts & Entertainment venues, College & Universities, Food & Restaurants, Nightlife Spots, Outdoors, Shops, Transportation, and so on. Using such data, one can estimate the number of different types of venues near the properties of interest. Use clustering algorithms, properties can be grouped into different groups with various prices and characteristics.