Introduction

The main objective of tis study will be to explore the neighbourhoods of New York city to determine the correlation between real estate value and the venues surrounding each parcel of land. This study aims to help people moving into the city to find a suitable place to live.

Property owners and real estate agents usually advertise the fact that their properties are located close to supermarkets, landmarks, restaurants etc. In this manner they raise awareness of the attractiveness of their property and as a result, increase the property's price.

The main question is: Do surrounding venues influence the price of a property? If this is the case, then what are the types of venues which impact the property's value the most?

The intended audience for this study includes real estate agents, house vendors and customers interested in acquiring a real estate property in New York City.

Data overview

Real estate prices in neighbourhoods of New York City are readily available and there is also a wide range of them. For instance, they might range between $500k and $5m. In addition, geospatial data is readily available and this aids in visualising the results on a map of New York.

The dataset comprises of data collected by calling FourSquare API and from https://www.cityrealty.com/nyc/market-insight/features/get-to-know/average-nyc-condo-prices-neighborhood-august-2020/18804.

The CityRealty webpage is scraped for the list of New York City neighbourhoods. In addition, the geographic data of the neighbourhoods is identified including their centre and borders. This data is passed to the FourSquare API and the 'explore' endpoint is used to return a list of surrounding venues. The occurrence of each venue is measured, and one hot encoding is then applied to convert each venue type into a column. The venue frequency is catalogued as a typefloat. The average price is then standardised by subtracting the mean of the column and dividing by σ^2. Every row in the data frame represents a neighbourhood and every column the frequency of a venue. The last column represents the standardised real estate property price.

A sample of 5 neighbourhoods is depicted in figure 1. The dataframe consists of 50 samples and 300 features in total

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Animal Shelter	Antiq Shop		Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio	StandardizedAvgPrice
0	Battery Park City	0	0	0	3	0	0		0	1	4	0	1	0	-1.303912
1	Bedford-Stuyvesant	0	0	0	0	0	0	•••	0	1	6	0	0	1	-0.418350
2	Boerum Hill	0	0	0	1	0	0		0	0	2	0	0	2	0.015011
3	Brooklyn Heights	0	0	0	2	0	0		0	1	4	0	0	5	-1.099479
4	Bushwick	0	0	0	1	0	0		0	0	1	0	0	2	-0.587926

Figure 1: Sample from data frame used to conduct study