Applied Data Science Capstone

Zeiad Wael Sabra 4/5/2019

Analysis of New York Airbnb Prices

Week 4 Report

Table of Content

1. Introduction
2. Business Plan
3. Data Selection
   1. Geospatial Data
   2. Airbnb Rental data
   3. Foursquare API
4. Next
5. Sources

Introduction

New York City is a huge tourist attraction visited by millions each year, making finding a place to stay a very difficult and pricey endeavour. IsPriceRight is a website that offers recommendations for prices of Airbnb. Is the place you plan on staying at overpriced? Head to RightPrice to check if the price is right or if you are being scammed.

Business Plan

RightPrice wants us to make a model to advise tourists visiting New York City on the optimal price for a place to stay in New York city. A tourist provides us with Information about the place and we provid him/her with the optimal price using our model. Our goal is to build a model that give an estimate of the rent of a place in New York City using available data.

The Desired outcomes are:

- A model for calculating rental prices.
- A description of the most relivant features of the model.

- Cluster the Neigbourhoods based on the Rent, Venues, and location.

Data Selection

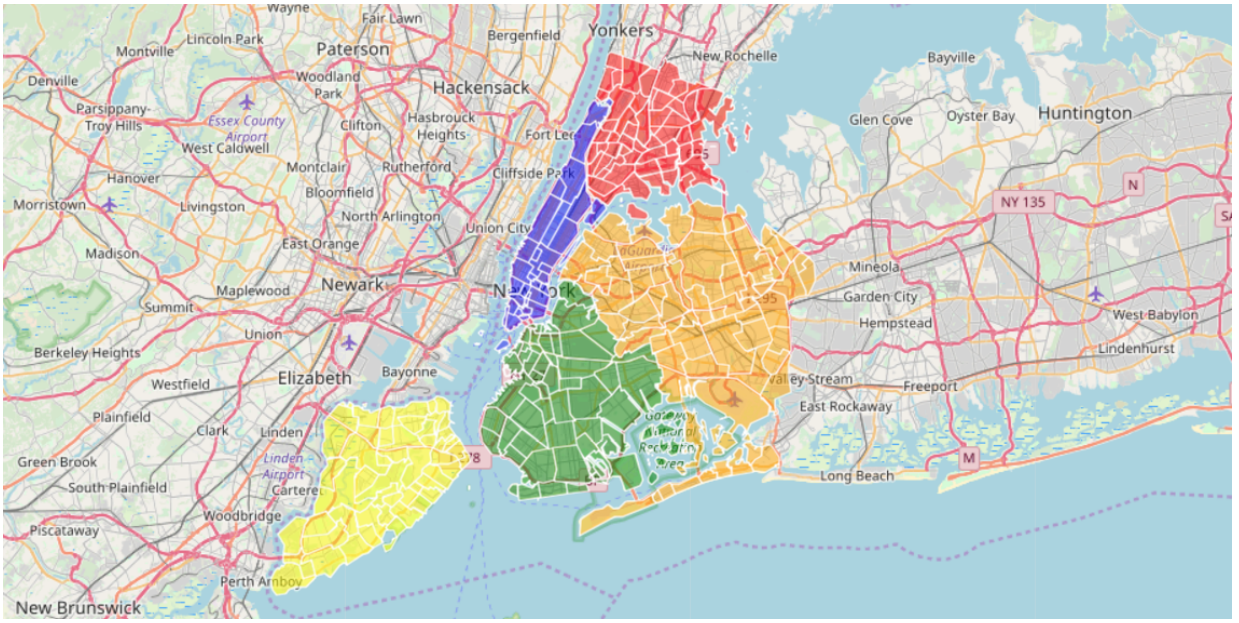
1. Geospatial Data

- The data consists of Neigbourhoods names, Boroughs names, Neigbourhood bountries and some other columns.

Here are the first 5 rows of the geospatial data

	neighborhood	boroughCode	borough	@id	geometry
0	Allerton	2	Bronx	http://nyc.pediacities.com/Resource/Neighborho...	POLYGON ((-73.84859700000018 40.87167000000012...
1	Alley Pond Park	4	Queens	http://nyc.pediacities.com/Resource/Neighborho...	POLYGON ((-73.74333268196389 40.7388830992604,...
2	Arden Heights	5	Staten Island	http://nyc.pediacities.com/Resource/Neighborho...	POLYGON ((-74.169827 40.56107800000017, -74.16...
3	Arlington	5	Staten Island	http://nyc.pediacities.com/Resource/Neighborho...	POLYGON ((-74.15974815874296 40.64141652579018...
4	Arrochar	5	Staten Island	http://nyc.pediacities.com/Resource/Neighborho...	POLYGON ((-74.06077989345394 40.59318800468343...

We also use Folim to plot the boundries of each Neighbourhood assigning then colors by their borough



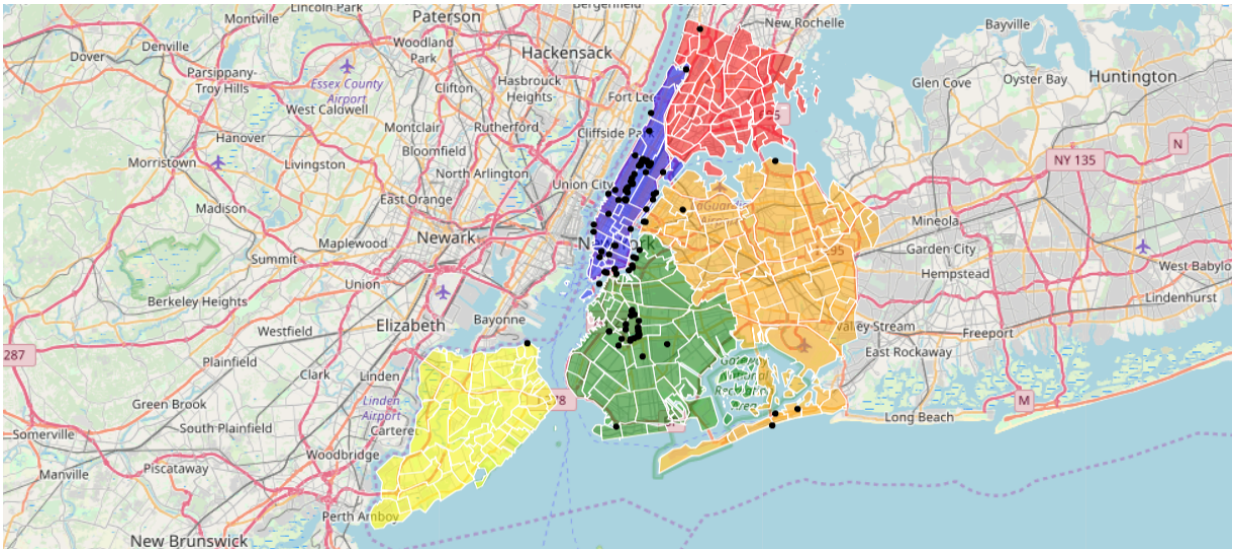
2. Airbnb Rental data

Here are the first 5 rows of the data

	id	listing_url	...	calculated_host_listings_count	reviews_per_month
0	1533652	https://www.airbnb.com/rooms/1533652	...	1	1.1
1	3423077	https://www.airbnb.com/rooms/3423077	...	1	1.5
2	326908	https://www.airbnb.com/rooms/326908	...	1	2.4
3	4625178	https://www.airbnb.com/rooms/4625178	...	1	0.5
4	3614041	https://www.airbnb.com/rooms/3614041	...	1	0.2

The data also contains price per night, latitude, longitude and many other features

Here are the first 100 Places in the data



3. Foursquare API

- We are going to use the Foursquare API to explore the nearby venues available around each listing of the Airbnb dataset and see how they affect the price of the listing.

Here are the list of venues near the first Place listed in the data

	Place	Place Latitude	Place Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Charming Studio - Central Park	40.781561	-73.971238	American Museum of Natural History	40.781282	-73.973238	Science Museum
1	Charming Studio - Central Park	40.781561	-73.971238	Hayden Planetarium	40.781718	-73.973239	Planetarium
2	Charming Studio - Central Park	40.781561	-73.971238	American Museum of Natural History Museum Shop	40.780973	-73.973028	Souvenir Shop
3	Charming Studio - Central Park	40.781561	-73.971238	Rose Center for Earth and Space	40.781741	-73.973127	Planetarium
4	Charming Studio - Central Park	40.781561	-73.971238	Shakespeare Garden	40.779755	-73.969976	Garden

Next

Since there are so many listings and some of them are very close togther requesting nearby venues from Foursquare API would be difficult. Instead, for each neighbourhood, we will request the available venues, then place each place listed in its neighbourhood. We then create a model that will predict the optimal price for that neighbourhood.

Sources

We used http://data.beta.nyc (http://data.beta.nyc) as the source of our data.

The data for the Rental Data was download from http://data.insideairbnb.com/united-states/ny/new-york-city/2015-05-01/data/listings.csv.gz (http://data.insideairbnb.com/united-states/ny/new-york-city/2015-05-01/data/listings.csv.gz)

The data for the geolocations and boundries of New York's neighbourhoods was downloaded from: http://data.beta.nyc/dataset/pediacities-nyc-neighborhoods (http://data.beta.nyc/dataset/pediacities-nyc-neighborhoods)

The venues data was aquired using Foursquare API at https://foursquare.com (https://foursquare.com)