Capstone Project - The Battle of the Neighbourhoods

(Week 2)

by IBM/Coursera

Alex Berschauer

Germany

2019

Table of contents:

1. Introduction: Business Problem

1. Introduction: Business Problem

In this part we shall discuss the business problem and the audience who could be interested in the results of this project.

1.1 Scenario and Background

I am trying to get a data scientist, adding to my MBA degree, and currently living in Frankfurt-am-Main, Germany. I enjoy many amenities in the neighbourhood, such as international cousin restaurants, cafes, food shops and some entertainment options. Recently, I have been offered a opportunity to work in New York, US.

Well, I am really very excited about the future job, I am also a little stressed about to secure a comparable place to live in New York. Therefore, I decided to apply the learned skills during the IBM-Coursera course to explore possibilities to make sure my decision about changing place of living is factual and rewarding. I am pretty sure that there are alternatives to achieve the answer to my question using available media tools but doing it myself with learned tools.

1.2 Problem to be resolved:

The challenge to resolve is being able to find a rental apartment in New York that offers similar characteristics and benefits to my current situation in Frankfurt. Therefore, in order to set a basis for comparison, I want to find a rental unit subject to the following conditions:

Apartment with min 3 bedrooms (we have 2 kids in our family) Monthly rent of the apartment not to exceed US$3000/month Flat should be located within walking distance (up to 2.0 km) from a subway metro station in New York Area of the flat with amenities and venues some similar to the ones described for current location

1.3 Interested Audience

I think this is a relevant project for a person or family considering moving to the Big Apple, since the approach and methodologies used here are supposed to be applicable in all cases. The use of Foursquare data and mapping techniques combined with data analysis will help resolve the key questions. Lastly, this project is a good practical case to practice Data Science skills.

2. Data

In this part we shall discuss the data and its sources that will be used to solve the problem

2.1 Description of the Data: The following data is required to answer the issues of the problem:

List of Boroughs and neighborhoods of Frankfurt with their geodata (including latitude and longitude). List of Boroughs and neighborhoods of Manhattan, NY with their geodata (including latitude and longitude). List of metro stations in Manhattan, NY with their address. List of apartments for rent in Manhattan, NY area with their addresses and price range with additional information, such number of beds, etc. List of Venues for each Manhattan, NY neighborhood List of Venues for metro stations.

I will be able to quickly point out the popups to know the relative price per metro station area. Addresses from rental locations will b converted to geodata(latitude , longitude) using Geopy-distance.

Data will be searched in open data sources if available, for example, - the list of Manhattan neighborhoods has been already worke during the LAb exercise in this course. A .csv file was created which will be read in order to create a dataframe and its mapping. So data will be used as well.

2.2 How the data will be used to solve the problem

The data will be used as follows:

Use Foursquare and geopy data to map top 10 venues for all Manhattan neighbourhoods and clustered in groups (as per Course LAB) Use foursquare and geopy data to map the location of subway metro stations , separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations. create a map that depicts, for instance, the average rental price per square ft, around a radius of 2.0 km around each subway station - or a similar metrics.

A list of places for rent was collected by web-browsing real estate companies in Manhattan:

http://www.rentmanhattan.com/index.cfm?page=search&state=results
https://www.codecademy.com/articles/streeteasy-dataset https://streeteasy.com/blog/data-dashboard/%7B%22topic%22:%2211e3d9ab-a69e-4111-a63d4ff8deddd423%22,%22message%22:%7B%22uid%22:%222333e1da-c940-4eb4-a289b204b5179ad9%22,%22status%22:%22done%22%7D%7D?agg=Total&metric=Inventory&type=Sales&bedrooms=Any%20Bedrooms&property=Any%20Property%20Type&minDate=2010 01-01&maxDate=2019-07-01&area=Flatiron,Brooklyn%20Heights

A .csv file (/resources/data/rentalIndex_All.csv) was locally stored and shows the rental place that indicated: areas of Manhattan, address, number of beds, area and monthly rental price. An algorithm was used to create all the geodata using Nominatim With the of geolocator = Nominatim(http://web.mta.info/developers/data/nyct/subway/Stations.csv) , it was possible to determine the latitude longitude for the subway metro locations as well as for the geodata for each rental place listed. The loop algorithms used are show the execution of data in section 3.0 "Great circle" function from geolocator was used to calculate distances between two points , as the case to calculate average rent price for units around each subway station and at 2.0 km radius. Foursquare is used to find the avenues at Manhattan neighbourhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

3. Methodology

I am currently living in Frankfurt-am-Main, Germany. I enjoy many amenities in the neighbourhood, such as international cousin restaurants, cafes, food shops and some entertainment options.

Some word about this beautiful city:

Frankfurt is the most international city in Germany, the largest financial centre on the continent, the historical city of coronations, the city of Goethe and the Frankfurt School… In brief, to the smallest metropolis in the world, in which there is a lot to discover at close hand. Whether this glance at the city's home page is the start of a longer, maybe even permanent, stay on the banks of the Main, or if, as a business visitor, you only have a limited amount of time available: you will find that the city has interesting offers for extensive tours of the city, cultural enjoyment and attractive shopping trips waiting for you.

In this section we represent the main components of the report where the data is gathered, prepared for analysis.

Two datasets will be downloaded:

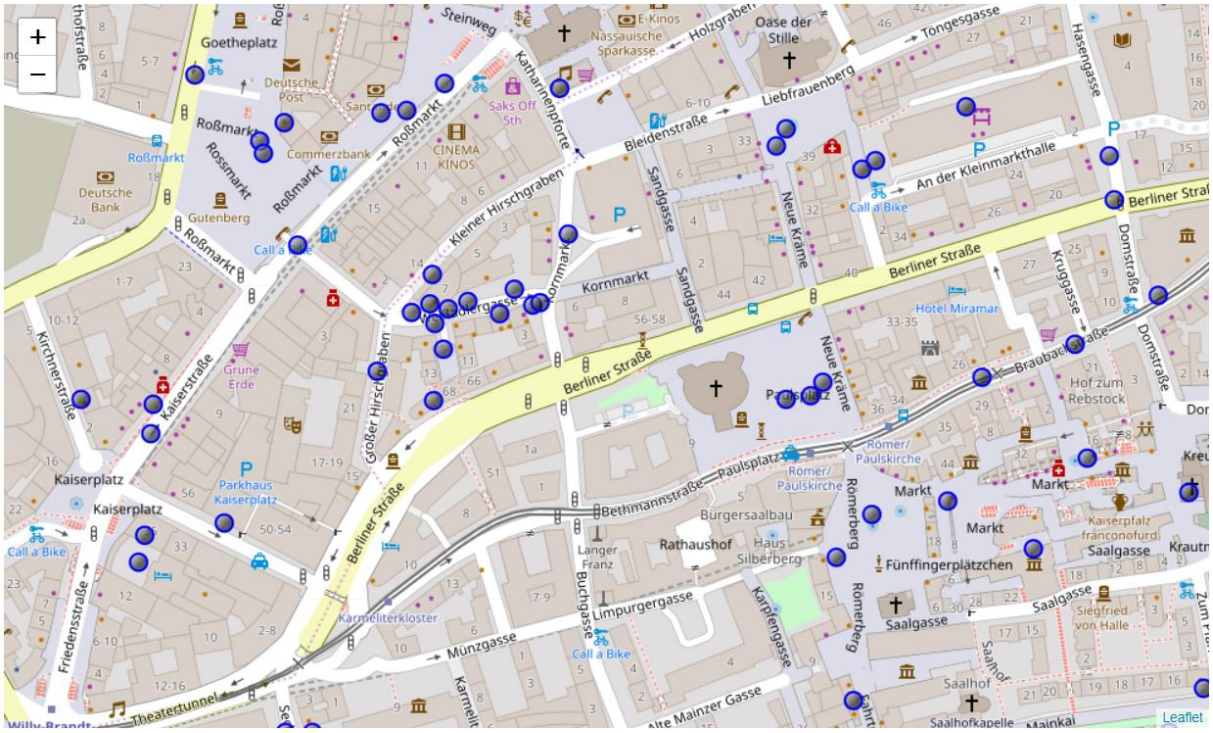
The data of the current location in Frankfurt. Frankfurt Map - Current residence and venues in neighbourhood.

	name	categories	lat	lng
0	Weinterasse Rollanderhof	Wine Bar	50.112473	8.682164
1	Römerberg	Plaza	50.110489	8.682131
2	SCHIRN Kunsthalle	Art Museum	50.110291	8.683542
3	Dom Aussichtsplattform	Scenic Lookout	50.110609	8.684908
4	Kleinmarkthalle	Market	50.112778	8.682958
5	Bitter & Zart Chocolaterie	Chocolate Shop	50.111444	8.683904
6	Jamy's Burger	Burger Joint	50.111226	8.681699
7	Wackers Kaffee	Coffee Shop	50.112064	8.679467
8	Main	River	50.108390	8.682631
9	Superkato	Sushi Restaurant	50.111664	8.679153

Here we can see some example venues in neighbourhood of the current residence in Frankfurt-am-Main. Germany.

But, what about to have a look at them on a map?

Here we go – the map of Frankfurt residence place with venues in Neighbourhood:



But now I must think about the future. I Know, New York is the Capital of the World and also called the Big Apple.

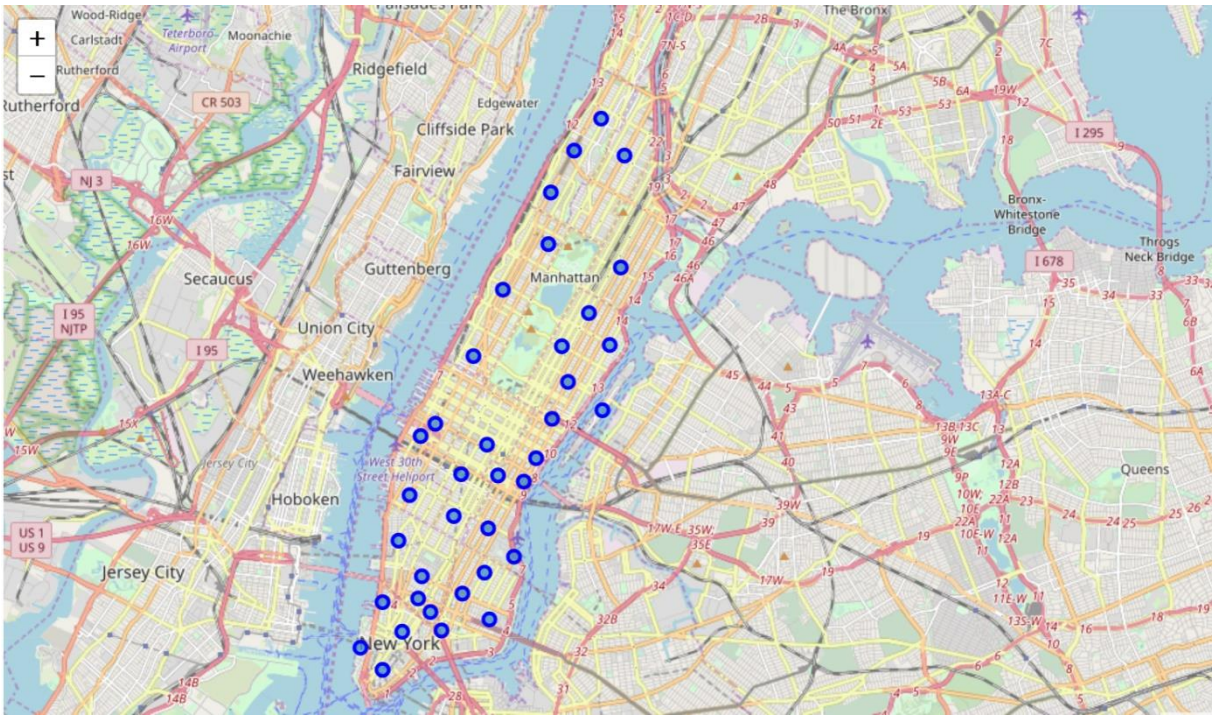
Here are some words about this exciting metropole:

Hustle. Heart. Glamour. Grit. This city is an ever-changing blend of these elements. Hustle powers Michelin-starred kitchens. Heart pulses under Broadway lights. Glamour shines in art deco palaces. And our New York City grit lends indelible character to every experience.

Well, I am really very excited about what I am expected of in the Big Apple!!!

But what is my future neighbourhood?

Let us have a look at the map:



And at the list of my neighbourhood:

Marble Hill	Roosevelt Island	Little Italy
Chinatown	Upper West Side	Soho
Washington Heights	Lincoln Square	West Village
Inwood	Clinton	Manhattan Valley
Hamilton Heights	Midtown	Morningside Heights
Manhattanville	Murray Hill	Gramercy
Central Harlem	Chelsea	Battery Park City
East Harlem	Greenwich Village	Financial District
Upper East Side	East Village	Carnegie Hill
Yorkville	Lower East Side	Noho
Lenox Hill	Tribeca	Civic Center

What can I say? - Wow, that is cool!!

Here are some examples of amenities in the neighbourhood:

Battery Park City

Park
Coffee Shop
Hotel
Memorial Site
Gym

Central Harlem

Seafood Restaurant
Bar
Cosmetics Shop
French Restaurant
Public Art

Carnegie Hill

Pizza Place
Coffee Shop
Café
Cosmetics Shop
Bakery

Chelsea

Coffee Shop
Italian Restaurant
Ice Cream Shop
Bakery
Nightclub

Well, not bad!!!

4. Analysis

The analysis and the strategy:

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed: location near a subway, rental price and similar venues to Singapore. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

The processing of these DATA and its mapping will allow to answer the key questions to make a decision:

what is the cost of available rental places that meet the demands?

what is the area of Manhattan with best rental pricing that meets criteria established?

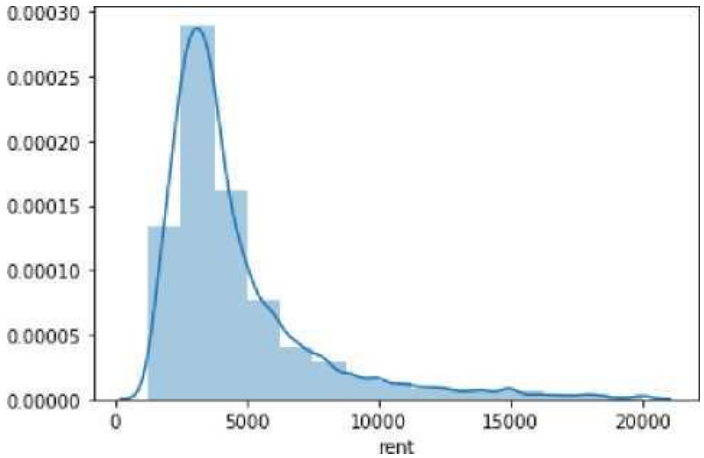
What are the venues of the two best places to live?

How the prices compare?

How venues distribute among Manhattan neighborhoods and around metro stations?

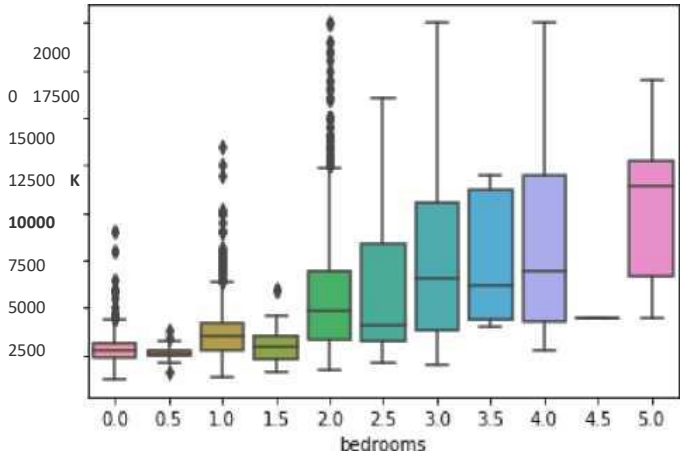
Here we see some results of our analysis:

MANHATTAN APARTMENT RENT PRICE STATISTICS¶



As we can see the rental prices are mainly set at actually round $3000 per month, which is satisfactory!

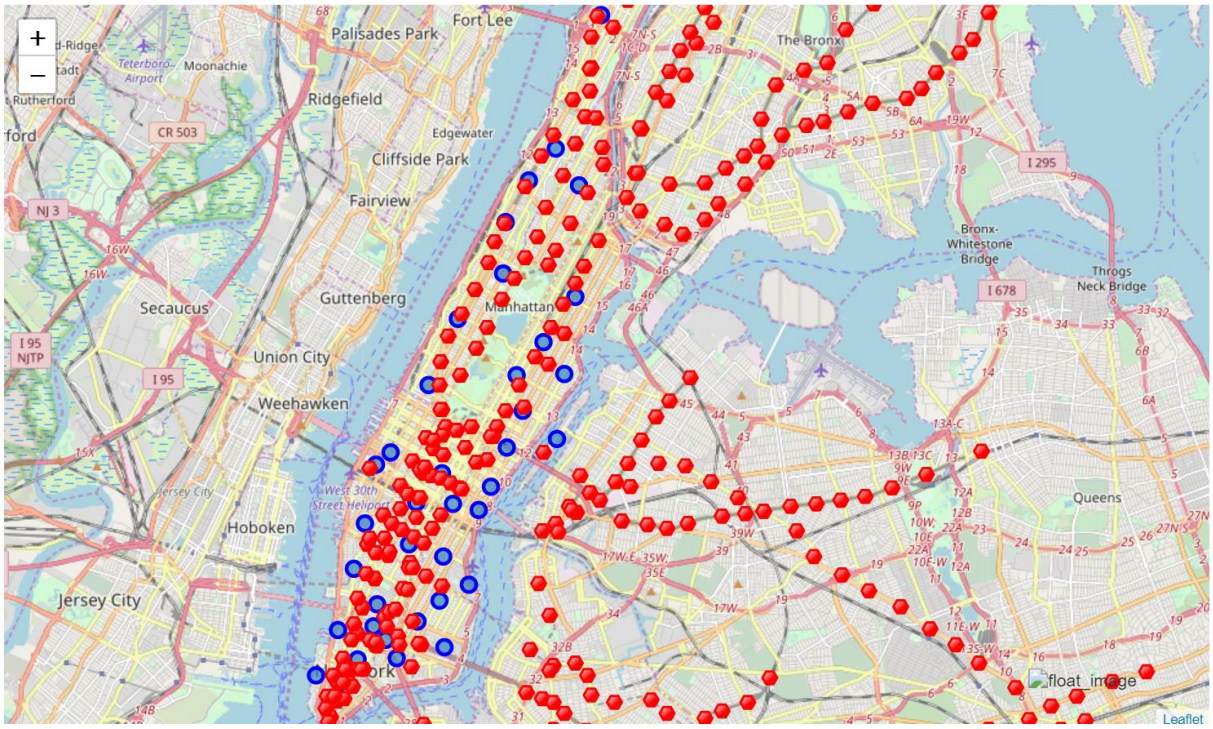
Remember, we need a 3 bedrooms apartment, for we have two children.



The boxplot shows us that there are not very many offers to the desired price about $3000 per month. But we still have a choice!

5. Results and Discussion

Consolidate Map of Manhattan showing places for rent and the subway locations nearby



To remember of the main question asked in the 1. Section of our survey:

Question: *I would like to have a choice of locations with the following criteria:*

Apartment with min 3 bedrooms (we have 2 kids in our family)

Monthly rent of the apartment not to exceed US$3000/month

Flat should be located within walking distance (up to 2.0 km) from a subway metro station in New York

Area of the flat with amenities and venues some like the ones described for current location

Apartment Selection

I can now choose the apartment which correlates with my idea:

- of price (US$3000/month)
- of flat size (with min 3 bedrooms)
- up to 2.0 km from a subway metro station
- amenities and venues some like the ones described for current location in Frankfurt

Discussion

In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course

As to me, this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned.

I hope I have created a good project that I can present to potential employers as an example to show my potential.

With this work, I believe I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.

6. Conclusion

I feel rewarded with the efforts and time spent to the project.

I believe this course with all the topics covered is well worthy of appreciation is to be recommended to other people willing to get theoretical and practical tools when learning Data Science.

This project has shown me a practical application to resolve a real-world situation that has impacting personal and financial impact using diverse Data Science tools.

The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend it for use in similar situations.

One must keep abreast of new tools for Data Science that continue to appear for application in several business fields.