Venues Popularity Data Analysis of New York

Wilton Beltré

Septiembre 20, 2019

1. Introduction

1.1 Background

The City of New York, is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km2), New York is also the most densely populated major city in the United States. The New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

1.2 Interest

New York City holds one of the greatest dynamic economy and high foot traffic that allows me to study three basic points on the subject of this study to answer the following questions:

- What are the venues preferred by people?
- Where do those people come from?
- What are the favorite places of the cities that visit the venue?

I am trying to find the answers to these questions focused on customer experience as retailers seek to seamlessly integrate the best of the digital and physical world.

US E-Commerce Sales as a Percent of Retail Sales represent 10.70% for Q2 2019
(source https://ycharts.com/indicators/us_ecommerce_sales_as_percent_retail_sales)

2. Data Acquisition

I have considered supporting the solution with the following data sources listed below

- I used Forsquare API to get the most common venues of given Borough of New York.

- Looking at kaggle.com I could find the SafeGraph data that contains foot traffic data for points of interest in the U.S. containing the most granular level the US Census Bureau reports.
- SaveGraph has another data source with the foot traffic coordinates that are necessary to determine the origin of the visitors.

- In order to group each identified coordinate, that is: each coordinate corresponds to a parent neighbor, I will be implementing reverse geocoding, therefore I need the GeoNames data source that helps this task.

Below I summarize what is intended to be achieved with the data.

Solution # 1 - We see 1000 visits to the "Accessory Store" in the Chelsea neighborhood.

	Accesorie Store	Arcade	Restaurant Mexico	Restaurant Italian
Neighborhood				
Chelsea	1000	0	560	1560
Bronx	20	100	1000	100
Chinatown	1200	300	0	10

Solution # 2 - We see 300 visits from people from New Jersey who visited the Chelsea neighborhood.

	New Jersey	Bronx	Pensilvania	Connecticut
Neighborhood				
Chelsea	300	1000	300	221
Bronx	600	200	645	100
Chinatown	1200	300	44	40

Solution # 3 - 600 people from New Jersey visited or prefer "Accessory Store".

	New Jersey	Bronx	Pensilvania	Connecticut
Venues				
Accesorie Store	600	500	740	630
Arcade	100	1000	1200	144
Restaurant Mexico	230	630	200	23
Restaurant Italian	611	100	10	55