Manhattan locations analysis for targeted advertisement

Clustering of advertising kiosks in Manhattan based on data from FourSquare API

Introduction/Business Problem

Consider you as owner of some kiosks network in Manhattan. Let's say you have the some small network of 100 locations similar to LinkNYC https://www.link.nyc/

Similarly to LinkNYC your spots provide free Wi-Fi, device charging etc.

Your network is completely free because it’s funded through advertising. Its groundbreaking advertising network not only provides brands with a rich, context-aware platform to reach New Yorkers and visitors, but will generate big revenue for New York City.

You have scope of customers for your advertising platform. Most of them sell products for women. Such as Fashion, Jewelry, Shoes etc.

Your goal is to make some analysis of your spots to do better offers segmentation and price diversification for different locations. Basing on people testimonials from places nearby.

With data based on recommended and popular venues from FourSquare. With their API.

Depending ont the analysis your company will build strategy for different types of customers and their products. Trying to reach the best price options and best location for their ads.

Data

Data sources and its exploratory analysis.

Data Sources

In our work we are going to use two data sources:

1) **LinkNYC kiosks locations dataset from https://data.cityofnewyork.us**

https://data.cityofnewyork.us/Social-Services/LinkNYC-Locations/s4kf-3yrf

LinkNYC's guys are generous and smart. And posts their locations dataset to NYC Open Data project.

In our work we will gently use their locations as our sample kiosks dataset.

2) **Foursquare Places API**

https://developer.foursquare.com/places-api

The main data source giving us information on popular places, people testimonials and other useful data neaarby our kiosks.

In the API the method we will use is "explore" https://developer.foursquare.com/docs/api/venues/explore. That "Returns a list of recommended venues near the current location".

For different purposes in our analysis we may use various endpoints such as:

https://developer.foursquare.com/docs/api/venues/trending Returns a list of venues near the current location with the most people currently checked in. May be useful for long-term time based analysis for your strategies

https://developer.foursquare.com/docs/api/venues/suggestcompletion Returns a list of mini-venues partially matching the search term, near the location.

We could use various methods to get venues data. But our goal is to target products to popular locations that could be provided with "explore" method. That's how we can explain our choice of API endpoint.

Additional API using

Nearby Places categories taken from FourSquare categories https://developer.foursquare.com/docs/resources/categories

Categories

Assigning categories to our location.

Based on FourSquare categories https://developer.foursquare.com/docs/resources/categories

As we remember our clients are producers of products for women audience. Such as Jewelry, Fashion, Cosmetics etc. That's why we are trying to do some filtering for target advertising.

Methodology

Basing on exploratory data analysis we did before we can use some unsupervised learning algorithm. As a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses.

What we have

Unlabeled data. We don't know how to split our locations to get the best segments for targeted avertising and customers and price diversification.

Our goal

Get grouped information of various kiosks locations based on such features as counts of popular venues categories nearby.

Machine Learning Algorithm

Due to unlabeled nature of our source data we will use some of of **Unsupervised Learning** algorithms. Without defined target labels our plan is to use **K-means Clustering**.

We can say places popularity is formed organically. As features for our data we will use this popularity rates (grouped counts of Foursquare recommended places nearby). K-mean algoritghm will help us to get:

- The centroids of the K clusters, which can be used to label our kiosks data
- Labels for the training data (each kiosk location will assigned to a single cluster)

Analysis of our clusters

	Woman_common_count	Cosmetics_count	Shoes_count	Jewelry_count	Fashion_count
0	4.037037	2.222222	0.611111	0.203704	1.444444
1	36.500000	2.000000	1.250000	1.000000	34.250000
2	19.428571	8.571429	4.857143	2.000000	9.428571
3	10.225806	3.709677	1.290323	0.806452	5.677419
4	36.000000	15.000000	14.250000	16.500000	10.000000

Visualization

Let's see what kind of clusters we got.

See screenshots of rendered maps below the cell. Obvious way for maps presentation for GitHub problems with Folium map rendering :(

Map screenshots

Adding screenshots of Folium maps. Due to problems of their rendering on github side :(
Map with markered locations

Map with popup

The map shows groups of clustered locations for our diversified advertising strategy

Results and Conclusion

Goals

K-mean algorithm helped us to detect groups for our goals:

- Set diversified prices for different kiosks. Basing on its cluster (group, type of popularity, value)
- Targeting types of advertisement for different products. Fashion products ads can be set on kiosks from our "Fashion oriented" Cluster, Jewelry products can be targeted to "Best location" Cluster with the most valueble places, medium prices can be set for "Medium" cluster, and so on. "Poor" clusters can be targeted to differen products after additional analysis with differen target categories. If it is not popular place for women, so let's analyze it different way and send offers to new interested customers.

Observations on Foursquare API data

As we mentioned above during our datasets analysis we researched different methods the API provides.

According to our purposes and goals we may implement and may build different models with ideas:

- https://developer.foursquare.com/docs/api/venues/trending Returns a list of venues near the current location with the most people currently checked in. May be useful for long-term time based analysis for your strategies

https://developer.foursquare.com/docs/api/venues/suggestcompletion Returns a list of mini-venues partially matching the search term, near the location.

We could use various methods to get venues data. But our goal is to target products to popular locations that could be provided with "explore" method. That's how we can explain our choice of API endpoint.