**E**xtreme **E**xperience Trip in New York City

-IBM Data Science Capstone Project

Maojie Gong

Email: [gongmj2000@gmail.com](mailto:gongmj2000@gmail.com)

1/31/2021

**Major Contents:**

1.Business problems

2.Solution proposal

3.Data to be used

4.Methodology and process

5.Results

6.Discussion

7.Conclusion

8.Appreciation

**1.Business problems**

Business Problem to Solve: **E**xtreme **E**xperience Trip in New York City (Optimal Travel Experience) which will be interested by Travel Corporation or Travel Agency.

New York is the dream city to visit for lots of people around the world and there are lots of places to see, various food to taste, lots of places for shopping, different cultures to experience and lots of entertainment activities to engage. But there are some issues for current tourism practice for both traditional tourist groups and individual tourists.

For traditional tourist group, the trip schedule and arrangement are normally fixed and there is little flexibility for personal activity arrangement and the places to visit are normally very famous but expensive with too many tourists which caused bad experience due to crowding and long waiting time etc. For individual tourists, they are free for their own arrangement, but they are normally from other countries or states and might not know New York city very well and they don’t have good idea about which the good places are to visit just normally fascinated by famous places like “Statue of Liberty” and “Empire State Building” etc. while there are much more and better places to visit in New York city. In that sense, they either are not satisfied with current trip arrangement or need help and guidance to select places to visit.

There is good business opportunity to offer kind of tourist solution in which more flexibility can be offered, and Extreme Experience can be achieved by offering top places to visit. This can be done by having good arrangement about the places to visit and making good balance between traditional fixed-schedule tourist group and individual tourists.

In short, the main idea for the proposal is to figure out a list of places to recommend for visiting for Travel Corporation/Travel Agency to achieve customer Extreme Experience based on the feedback and rating from previous visitors by fully leveraging Foursquare volume geo data available.

**2.Solution proposal**

For tourism, history data about customer experience and feedback rating for various places will be very helpful and can be the foundation for the future new trip planning and arrangement. Therefore, we can select top places with excellent customer feedback and experience for each category (hotel, museum, restaurants etc.) based on historic customer feedback and rating data to achieve extreme experience for future trip arrangement.

Foursquare is a technology company that built a massive dataset of location data. Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services like Apple Maps, Uber, Snapchat, Twitter and many others, and is currently being used by over 100,000 developers, and this number is only growing.

There are volume customer satisfaction and rating data as well as tips for each venue in Foursquare and we can investigate through Foursquare location data to dig out and recommend top places for sightseeing with excellent customer feedback and extreme customer experience to visit for each category like hotels, museums, restaurants and shopping centers etc.

For simplicity of the project, we just use customer rating data as the basis for venue selection. Later, we can add customer tips analysis through machine learnings for deeper analysis to help for venue selection, but it’s not included in the scope of this Capstone project.

**3.Data to be used**

Foursquare offers real-time access to Foursquare’s global database of rich venue data and user content to power your location-based experiences in your app or website through Places RESTful API. The location data is mainly divided into 3 categories: venues, users and tips and there are lots of detailed data available under each category. There are different functions you can use for investigation like venue search or explore, user investigation and tips collection etc. Following I just list some data that will be used for this Capstone project.

Venue search data: This is regular access and it covers data about venue Name, ID, Location and Category. This data can be used to get value list as the beginning of the project based on selected category, that is, we can collect a list of venues for each category which will be further analyzed to get Top Venue list with excellent customer feedback rating and experience. Venue ID will be used for further venue data collection and analysis.

Venue Explore data: This needs premium license for data access and is limited for free developers. It covers detailed data for the selected venue including venue Name, ID, Tips, URL, Location, Rating, Statistics, Contact Info. For this Capstone project, we are mainly interested in Venue Rating data which is used for venue ranking and selection.

Customer average rating data for the venue: This is based on Premium/Enterprise version and data is not provided for free developer license. For demo/capstone project purpose we can use the rating data for Top Venue Recommendation to get the venue list of recommendation for each category. After we consolidated top venue list for all the selected categories, we can apply clustering algorithms based on the venue location to divide the venue to several segments which is linked to each day as daily arrangement.

GPS data for the venue: This is basic venue data and will be used for city segmentation and clustering algorithms after Top Venues for each category are selected.

Tips: This needs Premium license and the data can be used for detailed Top Venue analysis and selection with certain Machine Learning algorithms to further support and fine tune venue rating data ranking. As this is premium call requirement and data access is limited for free developers and the ML algorithms are very complicated, the tips data will not be used for Capstone project this time but can be used later for venue selection accuracy enhancement.

Users: This is regular call data, the data can be further analyzed to help to enhance venue selection accuracy, but the data will not be used for Capstone project.

**4.Methodology and process**

We will select certain number of places for each category as recommend list and then get the total recommend list by consolidating all selected categories. In that sense, we will get a recommended top venue list with location information covering each category.

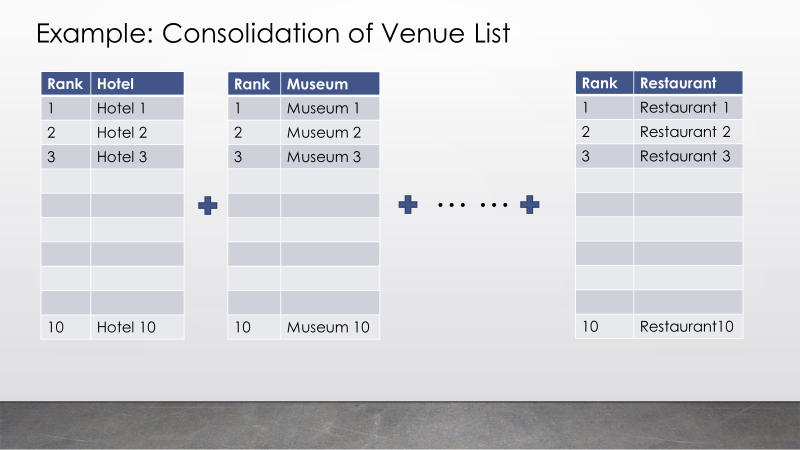
Based on how many days we are planning for the tourist trip, we can divide the recommended top venue list to different segments using Machine Learning clustering algorithms which is mapped to each day based on location data. Then we can tailor make trip arrangement for everyone based on their preference and interest of the places, that is, they can just simply select the places to visit for each day from the recommended list.

With this approach, we can ensure that all the selected places are top places with excellent customer feedback and experience and every tourist can have their tailor-made trip plan and they can have Extreme Experience trip in New York. Of course, there can be different kinds of practical arrangement for the trip based on Travel Corporation/Agency strategy, but this methodology can be the basis for trip planning.

Overall process is as following:

1. Limit the data to Manhattan and surrounding area as show case.
2. Define major categories for the travel arrangement based on typical tourists’ interest e.g. Hotels, Museums, Restaurants, Shopping Centers etc.
3. First select some central location in Manhattan as starting point. Then select max 50 venues for each category using Venue Search API with defined radius (e.g. 5km).
4. Collect rating for each venue using Venue Explore API and then select Top 10 venues with highest rating score for each category.
5. Combine all the venues for the selected categories and get the venue list with location data which will be used for clustering algorithms.
6. Use K-means algorithm or other clustering algorithms to divide the venues into K segments/clusters based on location data so that all the venues in one cluster are very close and can be used as recommendation list for one day trip arrangement. The travel corporation/travel group or tourists can choose the places they want to visit for each day from K segments. K segments corresponding to K days trip arrangement.

Following is the methodology for top venue consolidation.



Following is one expected result.



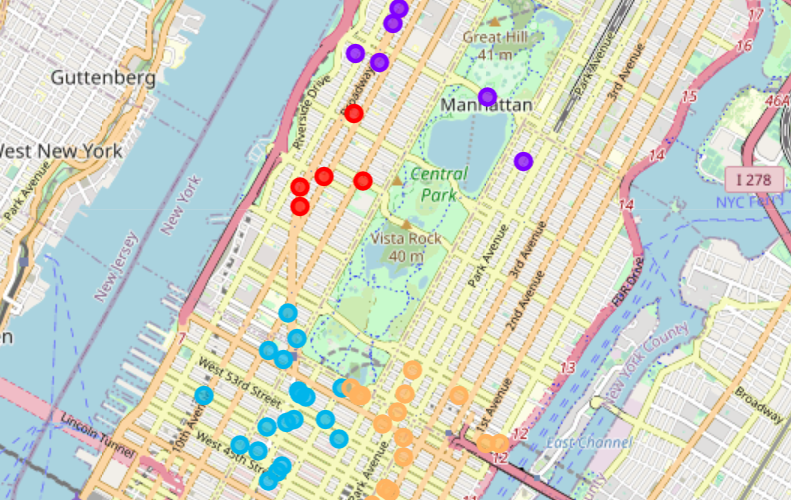
**5.Results**

I chose Manhattan as starting point and got the location data and started hotel search as first interested category using Foursquare API and got 50 hotels. Then I collected the rating data for the first hotel. In theory, I can collect the rating data for all the hotels and then can get Top 10 hotel list based on rating ranking as recommendation. Then I can search for other categories and get Top 10 list for each interested category and then consolidate all the venues as total recommendation list. Then I can segment the venues using K-means clustering with location data.

As premium license is needed for venue rating data collection and it’s limited for certain calls for free developer license, I didn’t have further rating data collection for other hotels and other interested categories. The process is the same and we just need to loop through all the venues using venue ID for rating data collection.

And I didn’t collect other categories data for simplicity purpose and just used the collected 50 hotels as example for venue clustering to show the methodology and machine learning algorithm. It will be the same process for the consolidated venue clustering just with more venues.

Following is the simplified sample results based on hotels only, that is clustering based on location data.



And the following is venue list showing venue names and cluster labels as well as other critical information:



**6.Discussion**

This is simplified process due to license limitation and time limitation etc., but it fully proved the concept and methodology as well as the process. For sure, it can be further optimized with more data and analysis.

Tip and user data are not used for this Capstone project due to premium license requirement for the data access and complicacy of Machine Learning algorithms. There is lots of information can be achieved from tip and user data which can add much extra value for venue selection accuracy on top of rating only approach to ensure that the selected venues are really top venues with excellent customer feedback and experience.

After top venue selection, there are several alternatives for Travel Corporation or Travel Agency to plan and arrange their travel groups. They can plan trip themselves by selecting some venues recommended and they can have flexible travel group so that tourists can choose their places to visit from the list themselves.

**7.Conclusion**

Based on the exercise of the project, we can conclude that the proposal, methodology and the process is feasible. So as Proof of Concept (PoC) project, we can think that it is successful. There are lots of data available in Foursquare website, but due to license and time limitation, we just used some very basic data for the project. With more data used, we can have more reliable and accurate selection for the top venues for the interested categories and it will be more practical for solution implementation.

**8.Appreciation**

This is last module of IBM Data Science certificate program besides other 8 courses. I really learned a lot from these courses as non-IT professional and this will be a great help for my future work and I’m sure lots of knowledge will be used in my daily work. I’d like to take this opportunity to appreciate for all the help and support I got from the trainers and technical support people as well as from the discussion group. It’s your help make the learning journey much more interesting and a little bit easier.