**1.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.

**2.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.

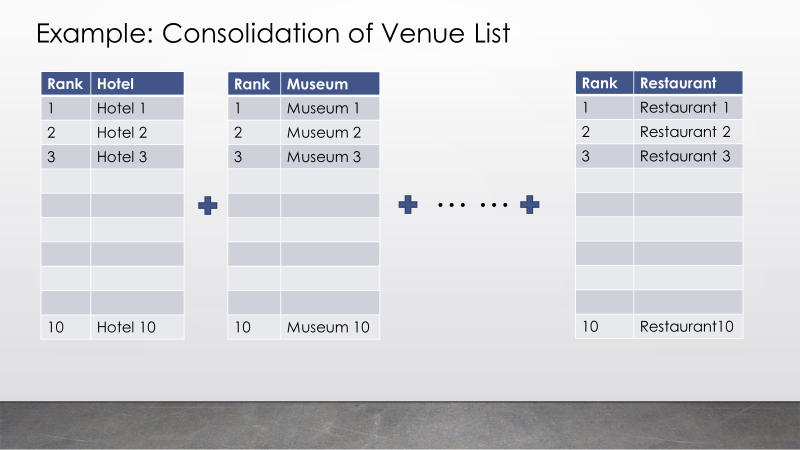
Based on how many days we are planning for the tourist trip, we can divide the recommended top venue list to different segments using Machine Learning clustering algorithms which is mapped to each day based on location data. Then we can tailor make trip arrangement for everyone based on their preference 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.

