**Machine Learning helps me in my Venue hopping Tour Venture**

**Gurpratap Singh (**[**gpsingh07331@yahoo.com**](mailto:gpsingh07331@yahoo.com)**)**

**Date Published: 15th June 2019**

**Table of Contents:**

1. Introductory section

Problem Description

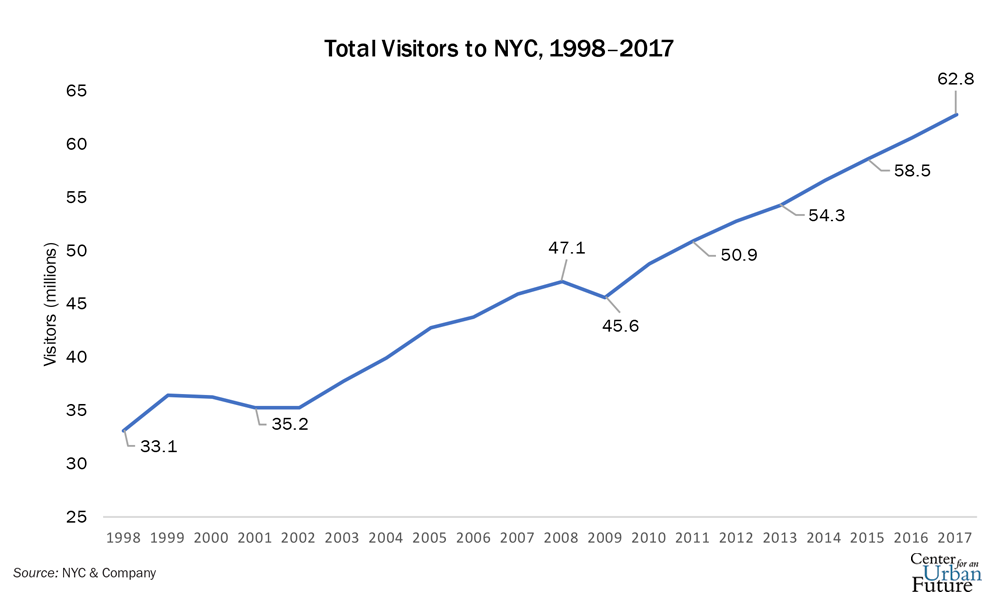
1. Methodology section

Data and Research and Approach

1. Exploratory Data Analysis
2. Results section
3. Discussion section
4. Conclusion section
5. References
6. Acknowledgment
7. Appendix
8. **Introduction: Problem Description.**

New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. It is one of world’s top tourist spots due to its diversity in Arts, Architecture, Shopping and Dining.

New York City received an eighth consecutive annual record of approximately 62.8 million tourists in 2017, counting not just overnighters but anyone visiting for the day from over 50 miles away, including commuters. Overall the city welcomed 38 million visitors who stayed overnight of which 13.3 million were international in 2018

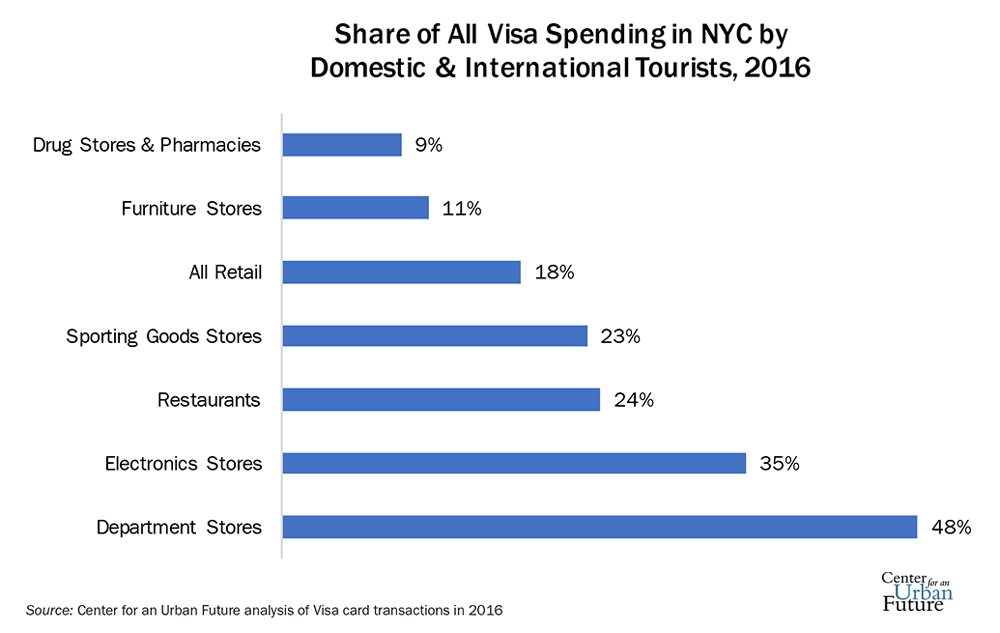


**Fig 1.**

[**https://www.google.com/search?client=firefox-b-1-d&q=New+York+City#**](https://www.google.com/search?client=firefox-b-1-d&q=New+York+City)

[**https://nycfuture.org/research/destination-new-york**](https://nycfuture.org/research/destination-new-york)

And the expenditure by these tourists have been mostly in Departmental stores and also restaurants. See Fig 2.



**Fig 2.**

Now imagine you are one of those tourists and you want to spend your time in NYC wisely. With size of NYC and so many choices available and gems of them hidden around and you want to cover the best of everything, how would you do it. How can you optimize your day so that you can experience the best of everything?

Here I offer you an optimized tour so you can spend your time wisely and get the best experience of whole of NYC in least amount of time.

Well at least let me start small and give you best of Eating, departmental shop and electronics experience across NYC in smallest possible time and allow you to select the section of NYC as a choice.

**Stakeholders:**

* **I, as the owner of Travel related online business.**
* **Tourists who have come to get relief from daily stress.**

Questions I need answer for so as to provide the above experience.

* How can divide the NYC into sections so as to get the location data for each of these sections?
* Where can I find the trusted data source for the best rated and liked Eating, Departmental and electronics shop location in NYC.
* How can I optimize the time taken to visit each of the top venue categories in each section of NYC?
* Which section should I visit first or should be the only section to visit?

1. **Methodology**

**Data**

For this project we need the following data:

* **New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.**

Data source: <https://geo.nyu.edu/catalog/nyu_2451_34572>

Description: This data set contains the required information. And we will use this data set to explore various neighborhoods of New York city.

* **Venues in each neighborhood of New York city with category, ratings and likes.**

Data source: Fousquare API

Description: By using this api we will get all the venues in each neighborhood. To keep things simple, if there is a tie, I will pick up one randomly.

* **GeoSpace data**

Data source: <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

#### Or Use geopy library to get the latitude and longitude values of New York City

Description: By using this geo space data we will get the New York Borough boundaries that will help us visualize choropleth map.

* **Route Time Travel Data**

Data source: <https://docs.microsoft.com/en-us/bingmaps/rest-services/routes/calculate-a-route>

*Or* [*https://github.com/Project-OSRM/osrm-backend/blob/master/docs/http.md#responses*](https://github.com/Project-OSRM/osrm-backend/blob/master/docs/http.md#responses)

Description: A waypoint is a specified geographical location defined by longitude and latitude that is used for navigational purposes. The route includes information such as route instructions, travel duration, travel distance or transit information.

### Approach

* Collect the new York city data from <https://geo.nyu.edu/catalog/nyu_2451_34572>
* Using FourSquare API, we will find all venues for each neighborhood.
* Filter out all venues that are of type Eatery, Shopping and Electronics.
* Using rating=5 or 4 and above and likes > 100 for each above venue, we will sort that data.
* Get their Lat and Long coordinates.
* Use K-Means Clustering to create our own sections with top of each type in each section.
* Using above Lat and Long and using Time Travel API, get the optimized route in each cluster.