1. The problem I would like to solve is:

let’s assume I would want to start a new business in Jersey City (or any other places), but I am not sure about what is the best business in this location that is most needed. For example, should I open a restaurant, or should I open a pharmacy, etc.?

The goal is to find the best new businesses to open in Jersey City (or any other places).

Clearly this would be an interesting topic to many investors.

1. Data:

Foursquare venue data for all neighborhoods in NY and Jersey City

1. Methodology:

* Aggregate the venue data in each neighborhood, so that we would know the distributions of venues within each neighborhood
* Run K Means algorithm to divide the neighborhoods into different groups
* Pick out the group that Jersey City belongs to, and calculate the intra-cluster similarity
* In order to find the best venue, we will iteratively add a new venue to each of the venue category in Jersey City one by one, while keeping all other venue numbers unchanged, and then update the intra-cluster similarity.
* The most needed business is the venue that will give us the most increment of the intra-cluster similarity.