IBM Data Science Capstone Project

The Battle of Neighborhoods

——The sharing economy of power banks in New York

I. Introduction

It's a Friday night, after a long and tedious week, you decide to have some fun with friends. After flicking through a few Facebook pages, you and your friends decide to go to a newly opened restaurant-bar, that seem to be quite popular among your Facebook acquaintances. You guys were having a great time taking photos of the food, of each other and instagraming these all night long. Suddenly a friend says, do you guys have chargers, I didn't bring mine because they are so heavy and inconvenient. Me neither, you replied, and looking at your phone's battery going red, you start worrying about how to get an Uber later when your phone goes out!

Although seeming a little dramatic, but the above scenario is quite common nowadays with mobile devices becoming a much more essential part of many people's (especially the younger generation's) life. People rarely carry mobile chargers or power banks, not because they are pricey but because they are inconvenient and heavy. Wouldn't it be nice to have easily accessible power banks when you need them the most?

Indeed, a number of companies have started market testing power banks for rent. They typically cost the user less than 50 cents per hour of charging.

Figure 1.1 An user returning the power bank to the rental station after usage. Fees are automatically charged through Paypal or Alipay.

II. The Business Problem

After observing some business success of power bank rental business in Asia, our company (hypothetically) decided that it could be a business opportunity in the US. However, as a start-up company, we are tight on marketing budget and can only test market in a handful of neighborhoods in New York at a time. After considering our constraints and the potential users, we have summarized the following information:

1. Our potential customers are 20~35 year olds.
2. Potential customers are more likely to use our service, the more time they spend at a venue (thus the more likely that their phone's battery running out).
3. The power bank rental stations need some but not extensive maintenance (assuming a monthly checkup).
4. The power bank rental stations cost $3,000 each to produce, and we only have a budget to produce a limited amount for pilot testing.
5. Our marketing team would like to be able to go frequently around the rental stations to do promotions and observe how people actually adopt or interact with our service.

Therefore, our business problem can be described as a two stage process:

1. Identify the neighborhoods and venues (such as restaurants and bars) most likely to have a lot of potential customers spending longer hours after work to maximize potential customer adoption of our service.
2. From these select the top few most densely concentrated neighborhoods and venues for our pilot testing, to minimize our marketing and production budget.

II. Data Description

In order to address our business problem, we will need data on popular places among our target customers (20~35 year olds). A good source of this kind of data is the Foursquare location data.

I believe the Foursquare data is appropriate for a number of reasons:

1. Foursquare users overlap with our target customers in terms of age-group.
2. Foursquare data are user generated and updated, which is more relevant to our business, because we don't want outdated places that are recommended by critics but irrelevant for our actual target customers.

Specifically, we will need to explore the following aspects of the Foursquare data:

1. Popular venues such as bars, restaurants, live music houses, and nightclubs
2. We will explore the venue's stats with Foursquare API, this will gives us a lot of useful information such as:
    i)   Sharing: number of checkins on Facebook and Twitter
    ii)   AgeBreakdown: to see the age of users
    iii) HourBreakdown: to see number of checkins each our of the day
3. We will explore the venue's time series data with Foursquare API:

i) TotalCheckins: to see if a venue is popular over time.

ii) ageXX-XX: to see the count of checkins by age group.

4. After identifying particular venues, we will explore venues that are similar.