# Prediction of Best Business Type to Open on a Specific Location

**Maria Inês Mota**

**November, 15, 2019**

## Introduction/Business Problem

For many entrepreneurs who are looking to find a location to open their business, or business owners who are looking to expand, there is a common struggle - to find the perfect place.
The perfect place is the one which maximizes their chances of success, or, in other words, the one which maximizes their popularity. In this scenario popularity will lead to profitability, which will eventually lead to success.

To develop this project we are starting with the following premisse: the popularity of a certain type of business is correlated with their physical location within a city. For example, let's say we are a trading company looking for the best spot in New York. It's obvious we should look for a place near Wall Street (as it is known as the Financial District).

## Business Description

The main goal of this project is to find the best location for a given business within a city. The idea is to use the Foursquare API to collect data on a chosen city - location, venue category, etc. The purpose of this is to predict the amount of popular hours of our business location based on existing similar venues in the area.

## Target Audience

The development of this project will be especially useful for business owners and entrepreneurs looking to open their first physical venue. As they probably won't have enough data of their own to give an accurate prediction of location, it will be useful using data from similar businesses to get the best possible prediction

However, this could also be useful for business owners looking to open new venues in other cities they don't know much about. In essence, they would have a better idea of the response their business would get before investing in places with less probability of success.

## Data Section

As explained before, we are going to predict the popular hours of new venue spots according to category and the similar venues existing (eg.: "in location A the most popular venues are of category «restaurant»" - this would be the best location to open a restaurant - with an indication of the most popular hours).

For this we are going to use the Foursquare API to gather the data related to our location, mainly related to venue categories, venue coordinates, popular hours, etc.

## Methodology

For the purpose of this report we are going to use “Regent Street, London, UK” as an example of how the script works.

To achieve our goal, firstly, we are going to need to get all categories of venues from Foursquare, and create a dataframe. After sorting the categories we need to use Nominatim to convert an address into latitude and longitude values - thus getting the location of the city we are going to analyse.

As we now have the values of latitude and longitude we can request a limited number of venues within a chosen radius through the Foursquare API - creating a second dataframe with the existing venues. We can combine the two and get the list of venues with the respective categories in the limited region.

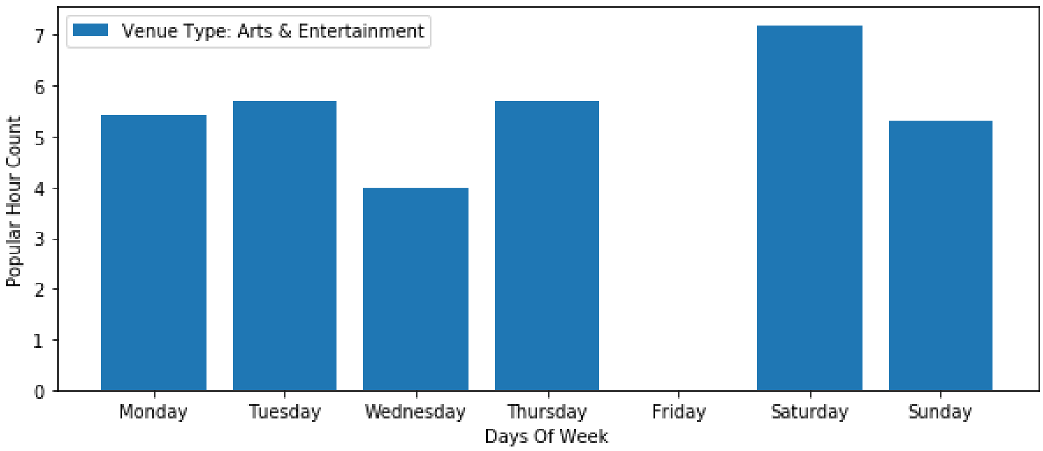
Next we need to collect the data related to the working hours and the time frames in which each business is more popular. From this step we create another dataframe with the days of the week and the time intervals in which they are more popular, for each venue.

From here we have every information we need to apply a machine learning algorithm to predict the amount of popular hours on our business location based on existing similar venues in the area.

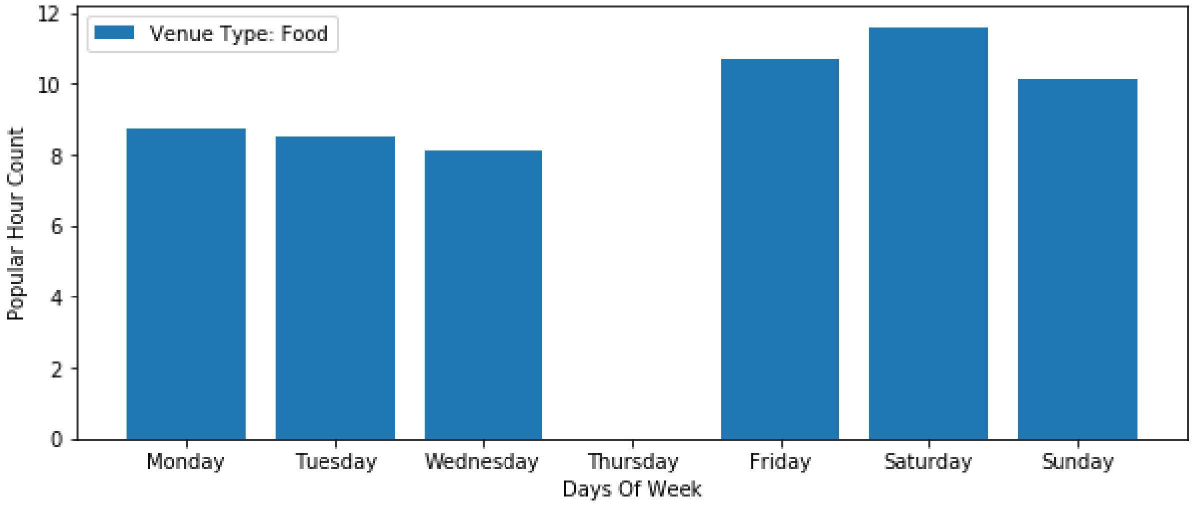
## Machine Learning Algorithm - KNN

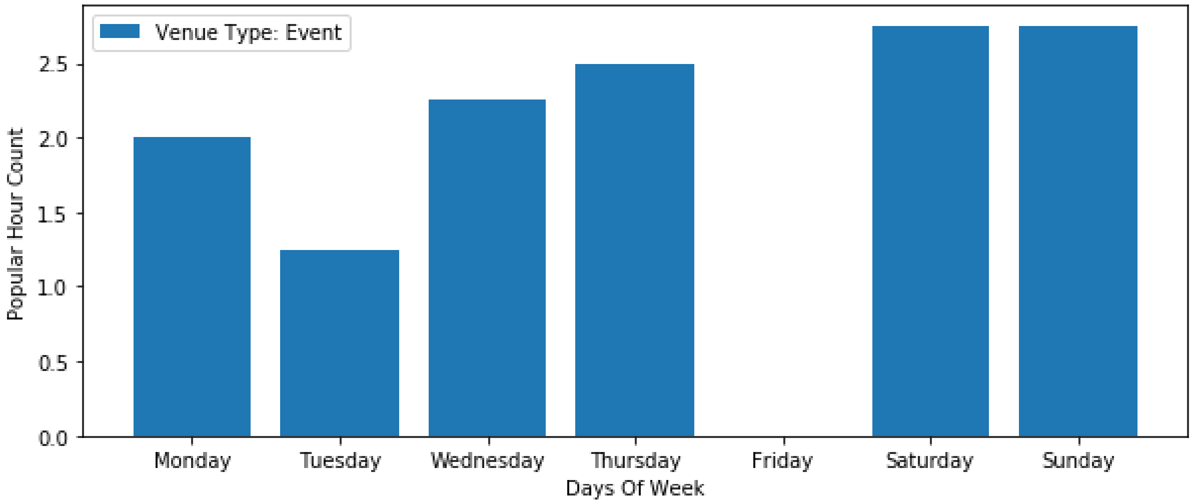
As the name suggests the k-nearest neighbors algorithm predicts a given metric, in our case the amount of popular hours of a venue, for any given day based on the nearest venues form our dataset. As such it was the best method to the solution of our problem.

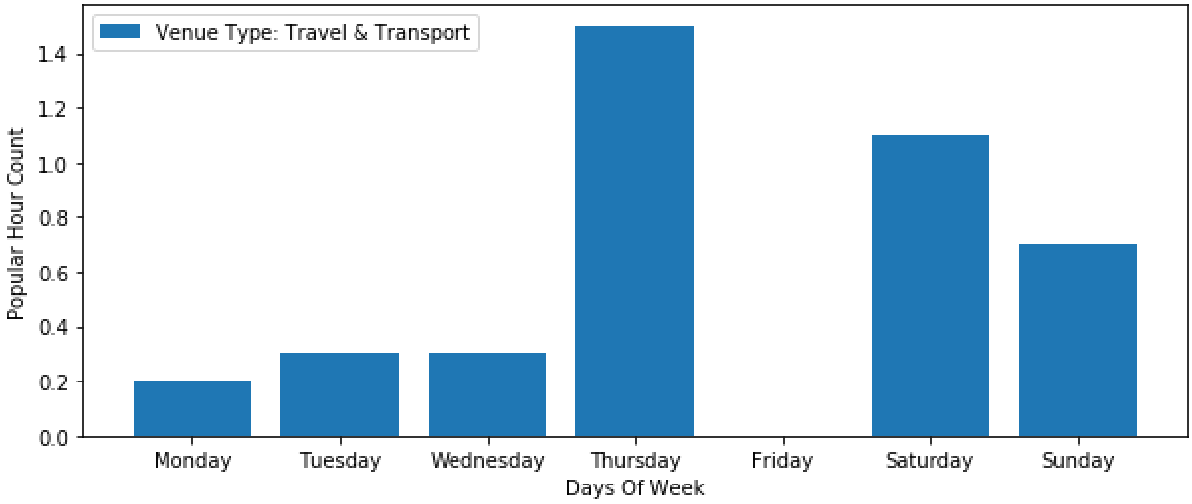
The way the algorithm was developed all parameters are set in the beginning of the script, so it could easily be used for any other address, and it would work just as well. To continue with our example of Regent Street, here are some examples of the charts created for different venue categories:



In this chart, for venues classified as “Arts & Entertainment” we can see the there is a good distribution of popular hours as it is fairly the same, except on fridays when the venues are probably closed, or it can also be an indicator that friday isn’t a good day to be open.
As we analysed this chart, the same could be done with the following ones:







Some interesting pointers from these graphs:

- The popular hours in food places slowly decrease until thursday (when it goes to zero) and from friday to saturday they reach their maximum and starts decreasing again on sunday.
- In the events category we can see there is a big concentration of popular hours in the weekends as the general public enjoys their free time in events venues, in the work week these venues are most likely being visited by tourists.
- From the travel and transport graph we can see a peak on popular hours on thursdays as it is habit for many people who work away from their home to leave on thursday and work from home, thus explaining the zero popularity on fridays.

**Discussion**

An interesting follow up to this project would be to cross reference the datasets on location with profitability of different business types, combining the two to get your most profitable option for this street.

**Conclusions**

In this project we analysed geographical location data to predict the popular hours of different venues in Regent Street, London, UK to come to a conclusion on which would be the most profitable venue. For this it was used a KNN algorithm.

After analysing the graphs and as a conclusion to our example in Regent Street, we came to the conclusion that the type of venue with most popular hours throughout the week were the food venues (restaurants).

Of course there is more to a business than it's location as such this project looks only to create a tool to help with a small aspect of a very complex matter. As said before