Case of Study:

**Location, Location and Location: A Machine Learning approach to choose  
location for a new business in NYC.**

Author: **P.A. Spring**

Applied Data Science Capstone  
by IBM

**I. Introduction.**

Without doubt choosing the right location for a new business is one of the most crucial steps in the life of a project. Determine which factor must be taken in consideration is core to solve this problem. Our case of study is centered in one of the business with the highest rate of failures within a year of operation, the small and mid-restaurants, this enterprise have percentages of failure of around the 26% within a year. Therefore, for restaurants owners and entrepreneurs is decisive to identify how different variables are related with the success or failure of his business.

The goal of this work is to develop a machine learning model with the objective of predict and describe how the right location influence the success or failure of a restaurant. Specifically, we are going to analyze the restaurants distribution around the neighborhoods of New York City using foursquare, google and opencity data. The target audience for this study will be restaurants owners, entrepreneurs and researchers

**II. Data.**

With the objective of predict how the location influence in the success or failure of a small/mid restaurant in the neighborhoods of New York City we are going to collect data from different sources. In the following lines is a description of all the necessary data for our project.

Restaurant descriptions: it is necessary to create a database with restaurants around the city of NY, with information about geo-location, name, address, type of restaurant, average rating, customers review and level of price. For this we will need to get the data from the foursquare api and complementary information from the google maps api.

Geo-location data from the neighborhoods of the city: Every neighborhood of the city has relevant and useful information for our case of study. For this we need the planning data of the city, in a geojson file with the geometry of every neighborhoods, with the specific name, codes and boroughs. This data will be extracted from the NYC data site.

Information about point of interest around the selected restaurants: In the NYC data site we can find different kind of locations of interest around the city, how college, schools, touristic points and metro stations. This kind of places and them closeness with the restaurants could be in some way related with the success or failure of them, because of that we will try to include them in the analysis.

Description of the groups of people living in the neighborhoods of the city: Demographic data is another important source of information for this study, some researches point at some specific demography characteristics around the success of a small restaurant, because of this we will use the demographic information of the neighborhoods of the city as: average age, income, education level. All this information is available in the census2020 website.

**III. Methodology.**

section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

3.1 Foursquare data acquisition.

3.2 Type of restaurants selection.

3.3 Restaurants distribution around the city.

The next step in the characterization of our dataset was to group and count the number of different restaurants by neighborhood. We can see in the next table which neighborhoods have the greater number of restaurants.

3.4 Datasets merging.

3.4 Statistical analysis and initial visualization.

3.5 Machine learning model selection.

**III. Results section.**

where you discuss the results.

**IV. Discussion section.**

where you discuss any observations, you noted and any recommendations you can make based on the results.

**V. Conclusion.**

section where you conclude the report.