**Final Report for the Capstone Project in Data Science**

1. Introduction:

Starting a new business venture by opening an establishment requires a lot of effort and careful consideration. It might be useful to use precedence in this matter to ascertain the success of your new business.

The problem I’m trying to describe and solve in this project is: if you had to start a venture in the city of Luxembourg what location would be most appropriate for you to do so, based on the success of previous businesses of the same category. I believe this maybe of interest to any party who wishes to start a business in the country.

1. Data Acquisition and Cleaning:

The Dataset used in this project has been sourced via Kaggle as a csv file named *luxembourg\_places* (a copy exists in the repository). It has been read into a pandas dataframe (*df\_places)* and the columns are as follows: *index, id, name, location.lat* , *location.lng, location.postalCode, location.cc, location.city, location.state, location.country, location.address, location.formattedAddress, location.crossStreet, Category, Nearest City, Category type, location.neighbourhood and VenuePage.*

The data was cleaned to more precisely fit exclusively to Luxembourg and not surrounding countries. Further, all venues where state or category wasn’t available were dropped from the dataframe. There was also a redundancy in features including country code, cross street, country, etc. which were also dropped considering the area in question was entirely within Luxembourg.

*First five rows of the dataset before and after cleaning:*



