**Capstone Project - The Battle of the Neighborhoods**

**WEEK 1:**

For this week, you will required to submit the following:

-A description of the problem and a discussion of the background. (15 marks)

[Introduction: Business Problem](https://render.githubusercontent.com/view/ipynb?commit=86ac2973bbdb50491e34450f2ecc07590d2bbe2f&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f6d69737332306b6d2f49424d2d43617073746f6e652d50726f6a6563742f383661633239373362626462353034393165333434353066326563633037353930643262626532662f43617073746f6e6525323046696e616c25323050726f6a6563742e6970796e62&nwo=miss20km%2FIBM-Capstone-Project&path=Capstone+Final+Project.ipynb&repository_id=255153975&repository_type=Repository#introduction)

-A description of the data and how it will be used to solve the problem. (15 marks)

[Data](https://render.githubusercontent.com/view/ipynb?commit=86ac2973bbdb50491e34450f2ecc07590d2bbe2f&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f6d69737332306b6d2f49424d2d43617073746f6e652d50726f6a6563742f383661633239373362626462353034393165333434353066326563633037353930643262626532662f43617073746f6e6525323046696e616c25323050726f6a6563742e6970796e62&nwo=miss20km%2FIBM-Capstone-Project&path=Capstone+Final+Project.ipynb&repository_id=255153975&repository_type=Repository#data) , [Methodology](https://render.githubusercontent.com/view/ipynb?commit=86ac2973bbdb50491e34450f2ecc07590d2bbe2f&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f6d69737332306b6d2f49424d2d43617073746f6e652d50726f6a6563742f383661633239373362626462353034393165333434353066326563633037353930643262626532662f43617073746f6e6525323046696e616c25323050726f6a6563742e6970796e62&nwo=miss20km%2FIBM-Capstone-Project&path=Capstone+Final+Project.ipynb&repository_id=255153975&repository_type=Repository#methodology)

## Introduction: Business Problem

**Background:**

Beringer Vineyards Inc. is one of the top wine companies in the U.S. The stakeholders are planning to open a flagship store in Europe for branding and market shares expansion. Paris is famous for its cuisine and wine industry in the world. The stakeholders chose Paris as the city of the flagships store. The competition is aggressive and we need an objective reasoning while choosing the locations, therefore, data science is adopted to help with the decision making.

This project aims to find optimal neighbourhoods for **a fine wine flagship store in Paris, France**. The result of this project is to find a couple of options, rather than a specific store, because this project will be followed by an on-site narrow-down research and finally help with stakeholders' decision-making.

**Criteria:**

The idea locations should meet these requirements:

* **To attract potential clients and for the best brand advertisement practice, locations should be close to restaurants and bars to guarantee heavy people traffic.**
* **We want to avoid the most expensive neighbourhoods like Ave Champs-Elysees because previous market research found the operation costs in these expensive neighbourhoods won't cut the books even. Medium high-end areas would be ideal.**
* **Locations are as close to city centre as possible, while meeting the other two criteria.**

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final neighbourhood can be chosen by stakeholders.

## Data

Paris is divided into 20 Arrondissements Municipaux (or administrative districts), shortened to just arrondissements. They and normally referenced by the arrondissement number rather than a name.

Following data sources will be needed to extract/generate the required information:

Arrondissements Municipaux for Paris CSV (administrative districts)

Data for the arrondissements is necessary to select the most suitable of these areas for new stores.

Initially looking to get this data by scraping the relevent Wikipedia page (<https://en.wikipedia.org/wiki/Arrondissements_of_Paris>), fortunately, after much research, this data is available on the web and can be manipulated and cleansed to provide a meaningful dataset to use.

Data from Open|DATA France: <https://opendata.paris.fr/explore/dataset/arrondissements/table/?dataChart>

Also available from Opendatasoft: <https://data.opendatasoft.com/explore/dataset/arrondissements%40parisdata/export/>

**Methodology**

The methodology and work flow includes the following:

* Outline the initial data that is required:

District data for Paris including names, location data if available, and any other details required.

* Obtain the Data:
* Research and find suitable sources for the district data for Paris.

Access and explore the data to determine if it can be manipulated for our purposes.

* Data Cleaning:

Clean the data, drop redundant columns; extract and convert the useful ones to pandas dataframe.

* Data Analysis and Location Data:
* Foursquare location data will be leveraged to explore districts in Paris.
* Data manipulation and analysis to derive subsets of the initial data.

Identify high traffic areas.

* Visualization:

Analysis and plotting the maps to demonstrate the analysis result.

* Conclusions:
* Summary the findings of the project.
* Recommendation making.

Final report to present to the stakeholders.

## Results and Discussion

From this visualization, it is clear that on a practical level, with no data to base decisions on, the circle of the 20 districts is very large, and researching and then visiting them all would be a daunting and time-consuming task. We have narrowed the search area down significantly from 20 potential districts to 3 that should suit the client's business. We have made inferences from the data in making the location recommendations, but that is exactly the point. There is no right or wrong answer or conclusion for the task at hand. The job of data analysis here is to steer a course for the location selection of new stores (i) to meet the criteria of being in neighborhoods that are lively with abundant leisure venues, and (ii) to narrow the search down to just a few of the main areas that are best suited to match the criteria.

## Conclusion and Report Link

There are many ways this analysis could have been performed based on different methodology and perhaps different data sources. I chose the method I selected as it was a straight forward way to narrow down the options, not complicating what is actually simple in many ways – meeting the the criteria for the surrounding venues, and in my case, domain knowledge I have on the subject. I originally intended to use the clustering algorithms to cluster the data, but as it progressed it became obvious that this only complicated the task at hand. The analysis and results are not an end point, but rather a starting point that will guide the next part of the process to find specific store locations. The next part will involve domain knowledge of the industry, and perhaps, of the city itself. But the data analysis and resulting recommendations have greatly narrowed down the best district options based on data and what we can infer from it. Without leveraging data to make focus decisions, the process could have been drawn out and resulted in new stores opening in sub-standard areas for this retailer. Data has helped to provide a better strategy and way forward, these data-driven decisions will lead to a better solution in the end.