# **Introduction / Business Problem**

Paris, France and Millan, Italy are the two well know and renowned destinations of Europe known as shopping and Fashion paradise. Both the city houses most world-famous apparel and accessories brand in the heart of the city. There are different studies available in literature regarding the effectiveness of both the cities as shopping destination but some of them may be lacking in data driven exercise during the decision-making process.

As part of this current study my intention is to help a small boutique start-up firm Smart Fashion inc. to take a holistic decision of opening a shop in either Paris or Italy main city or neighbourhood area. While making the data driven decision, both the city’s neighbourhood and the cities will be compared to see how similar or dissimilar they are and based on that location for boutique shop opening decision will be taken.

The objective of this capstone project is described as follows:

* While carrying out the comparison process, shopping centre data of both Paris and Millan’s neighbourhood will be collected and they will be plotted in a plot for visualization purpose.
* Based on the statistical average of shopping centre data in both the city’s neighbourhood area we will compare the similarity of the Paris and Millan.
* It is evident that city have more population within the city limit and the countries having more GDP per capita may have more spending power towards fashion or luxury items. In order to take the shop opening decision, both city’s population data and corresponding countries GDP (Gross Domestic Product) per capita data will be extracted from external data sources.
* Combining all the data mentioned above a concise decision will be suggested to Smart Fashion Inc.

# **Data Source & Collection**

The data used for this study are secondary data sources related to Paris and Millan’s location and neighbourhood data, data related to the two countries GDP per capita and city’s population data. During the process three different data sources are being considered and they are explained below:

* **Foursquare**: This data source will be access to get the shopping venue for the neighbourhood location of Paris, France and Millan, Italy. Foursquare provides REST (Restful Webservice) based API (Application Programming Interfaces) to get the data in JSON format. My own user account with Foursquare will be used for providing access credentials to get the data.
* **Wikipedia Page for City’s Population Data**: The Wikipedia page (<https://en.wikipedia.org/wiki/List_of_cities_in_the_European_Union_by_population_within_city_limits> ) with European city’s population data will be used to get an idea of population living within the city limits of Paris and Millan. Pandas library will be used for web scrapping of the page to het the tabular data during analysis.
* **Wikipedia Page for Country’s GDP per Capita:** The Wikipedia page (<https://en.wikipedia.org/wiki/Economy_of_the_European_Union> ) will be web scrapped to fetch the GDP per capita data of France and Italy. Pandas library will be used for web scraping the page and retrieving tabular data.

Both the Wikipedia pages are publicly available and accessible and there are no licencing agreements in place for accessing those data.