CAPSTONE PROJECT

Battle of Neighbourhood – Report

1. Introduction to business problem

Problem Background: -

Tourism in France has directly contributed 78.9 billion euros to total **Gross Domestic Product (GDP)** in the past years**.** 30% of which comes from international visitors and 70% comes from domestic tourism spending. France was visited by 89 million foreign tourists in 2018, making it most popular tourist destination in the world. In this project, we will be exploring two famous cities of France, *Paris* and *Strasbourg*. Even though France was most populous tourist destination, considering the number of nights spent in the country, it is in sixth place after *United States*, *United Kingdom, China*, *Spain* and *Italy.*

Paris, the Capital City of France is the third most visited city in the world. It has some of the world's largest museums including *Louvre* which is most visited art museum in the world. It hosts some of the world recognizable landmark such as ***Eiffel Tower, The Arc de Triomphe*** and many more.

Strasbourg is one of the four main capital of European Union alongside Brussels, *Luxembourg* and *Frankfurt.* It is among the few cities in the world not being a state capital and hosting international organization of the first order. Economically, it is an important centre of manufacturing and engineering. It is the second largest river port in France after Paris. The city is chiefly known for its *sandstone Gothic Cathedral* with its famous astronomical clock.

Evidently, both of these cities are rich in cultural heritages and thus attract millions of international tourists every year. As France stands at sixth position in terms of nights spent by these tourists, even though it is most popular tourist destination in the world, it will be helpful for tourists to have a rough idea about luxurious apartments, hotels and restaurants, pub, café etc. to make their stay more comfortable. This might change the current scenario by improving it's rank from sixth to 2nd or 3rd. If possible, France might stand at first in terms of total number of nights spent by tourists.

Problem description: -

It is obvious that people who visit these places are somewhere in need of a physical/virtual guide. Through this project, I have explored these two main tourist places to dig some of the useful information about all those luxurious amenities, tourist would be looking for. These basic luxurious resources could be: -

- Hotels
- Restaurants
- Multiplexes
- Opera House
- Mountains
- Museums
- Night Club
- Super Market etc.

  In addition to these, it can be quite helpful for those people who all are international migrants and are looking for perfect place to rent apartments. So, our project could be proven helpful for these immigrants as virtual guide. Our main aim is to provide an outlook of all these available venues within these cities so that people would be less reliable on local guides who often charge these immigrants huge amount in exchange of service.

2. Data

***Data source* - 1**

In this project, we will be exploring **Paris** and **Strasbourg.**

The dataset has been collected from Kaggle. It can be downloaded from this link. The dataset prepared by INSEE. It is the official French institute gathering data of many types around the France.

There were four files in the dataset, but as per the requirements, I have only used ***name_geographic_information.csv*** dataset. Given Dataset contains following features: -

- EU_circo: name of the European Union Circonscription
- Code_region: code of the region attached to the town
- nom_région: name of the region attached to the town
- chef.lieu_région: name the administrative center around the town
- numéro_département : code of the department attached to the town
- nom_département : name of the department attached to the town
- préfecture : name of the local administrative division around the town
- numéro_circonscription : number of the circumpscription
- nom_commune : name of the town
- codes_postaux : post-codes relative to the town
- code_insee : unique code for the town
- latitude : GPS latitude
- longitude : GPS longitude
- éloignement : I couldn't manage to figure out what was the meaning of this number

out of above features, only few were helpful. So, I performed Data wrangling to extract useful features, so that appropriate Machine-Learning algorithm could be used to extract useful information with more accuracy. Those features which were used as primary features for our models are listed below: -

- prefecture – renamed as Borough
- nom_commune – renamed as Neighborhood
- codes_postaux – renamed as Postal-codes
- latitude
- longitude

Given dataset contains many prefectures out of which only ***Paris*** and ***Strasbourg*** have been taken into consideration as only we are interested in exploring only these two cities. Since the dataset contains missing values, and we are only interested in exploring the venues, dropping the missing rows would be appropriate choice. After data cleaning, dataset looks like –

	Borough	Neighborhood	Postal-codes	latitude	longitude
0	Strasbourg	Strasbourg	67000	48.583333	7.75
1	Strasbourg	Strasbourg	67000	48.583333	7.75
2	Strasbourg	Bischheim	67800	48.616667	7.75
3	Strasbourg	Hoenheim	67800	48.616667	7.75
4	Strasbourg	Schiltigheim	67300	48.600000	7.75

Data source – 2

we will be using Foursquare API to leverage neighbourhood venues by providing geographical coordinates along with user credentials. Once, the neighbourhood venues are explored, data frame looks like –

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Strasbourg	48.583333	7.75	Amorino	48.581489	7.749795	Ice Cream Shop
1	Strasbourg	48.583333	7.75	Place de la Cathédrale	48.581544	7.750195	Plaza
2	Strasbourg	48.583333	7.75	Le Saint-Sépulcre	48.582451	7.749090	Alsatian Restaurant
3	Strasbourg	48.583333	7.75	Au Crocodile	48.583712	7.747542	French Restaurant
4	Strasbourg	48.583333	7.75	Maison Lorho	48.582866	7.748701	Cheese Shop