Coursera Capstone

IBM Applied Data Science Capstone

Finding an apartment in Paris, France

By : Avni Kothari

Introduction

Finding accommodation in Paris is kinda like filling in job applications: a lengthy, sometimes stressful and somewhat tedious process, but the result is SO worth the effort you put in!

But for someone who is in hurry finding an apartment can be a serious challenge especially for business workers.

Scenario and Background

I am a Data Science Student going for further studies in paris. My classes are going to start next week and due to my laziness ,i was unable to find an accommodation on time. Now i am not having so much time that a use websites to compare different accommodation places in paris. So i need to use my Data Science Skills that i gain during my course ***IBM Data Science Professional Certificate.*** Using these skills i have to find an apartment in paris and visualize different parameters that effects the expensive of the apartment in **Paris**.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Paris to have an apartment as my personal accommodation during my visit to Paris. Using data science methodology

and Geographical plotting , this project aims to provide the solutions to answer the business questions like:

- What is impact of apartment size on price of apartment
- What features are responsible for addition charge for same size apartments.
- What is appropriate location in paris where i can get an affordable apartment for my visit to paris.

DATA

To solve the problem, we will need the following data

- List of different apartments available in different parts of the city. This will contain the details about all the available apartments.
- Price of apartment or rent of apartment is most essential feature of our dataset.
- Our features of Data will be like:
    - Size of apartment
    - Floors
    - Balcony
    - Postal Code
    - Position (part of city where apartment is)
- INSEE_code of apartments will allow us to plot apartments geographically on the basis of price.

Source of data

The data source will be a french website ***www.seloger.com***.

This website is use to find different living places that are available in different part of France. We will use this website to scrape the data about apartments in Paris.

Methods to Extract data

We will use ***BeautifulSoup*** to extract the data about apartments from the source code of requested page.
When a result page load for any query then there is script tag in source code of the response page. This ***script*** tag contain the dictionary which contain details of different apartments that are available.

Here is image that how data frame will look like when we will finish our scraping:

```
In [15]: df.head()
```

Out[15]:

	INSEE_code	codepostal	floor	idagence	idannonce	idtiers	nb_photos	position	price	balcony	surface
0	750116	75116	0	156068	147793291	178982	8	0	1500	0	30.05
1	750116	75016	0	227641	148083277	263422	6	1	1360	1	41.00
2	750113	75013	0	109207	147585167	164701	13	2	1470	0	41.31
3	750117	75017	0	188694	147988845	244572	9	3	1580	1	47.00
4	750115	75015	0	102541	145643953	153256	8	4	1695	0	60.00