

Instructions :

- Ask the following questions and look for the answers using code and plots:
 1. Can you count something interesting?
 2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?
 3. Can you make a bar plot or a histogram?
 4. Can you compare two related quantities?
 5. Can you make a scatterplot?
- Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you'd like to investigate further? What other questions do the insights lead you to ask?
- Now that you've asked questions, hopefully you've found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?

Actions in Jupyter Notebook:

Step 1: import modules and files

Step 2: create functions to find top brands in a given df

Step 3: plot cbgs on map

3.1 plot most popular cbgs (according to the % or x wanted) (with column raw **visitor** count column)

3.2 plot most popular cbgs (according to the % or x wanted) (with column raw **visit** count column)

Step 4: create high level histograms of the popularity by day and hour

4.1 Plot popularity of days of the week (histogram) by summing across "common3"

4.2 Plot popularity of hours in a day (histogram) by summing across "common3"

Step 5: plot top x cbgs based on column "visit count"

Top 10, 20, 100, 500, 1000

→ visualize location on top 5 + which cbgs are top 5?

Step 6: Day classification:

6.1 Split days into categories: week days (Monday-Thursday) and weekends (Friday to Sunday)

6.2 Find most popular brand across all cbgs for week days

6.3 Find most popular brand across all cbgs for weekends

Step 7: Hour classification:

7.1 Split hours into categories:

1. Early morning: 0 to 7am
2. Morning: 7am to 12pm

3. Afternoon: 12pm to 7pm
4. Night: 7pm to 0am

7.2 Find most popular brand across all cbgs for the 4 categories

Step 8: Data clustering

8.1 Implement OPTICS algorithm

8.2 Save the clustered dataset to a csv file

Step 9: Second implementation of clustering with HDBSCAN

Step 10: EDA with the clustered data

10.1 Plotting clusters on map