Statistical modeling Project

G.Manvelyan 31July2023

Description / Main goal:

Generate, preprocess, and explore a DataFrame by merging data from City Bikes, Foursquare, and Yelp sources. Develop a regression model and provide business-oriented recommendations based on the analysis.

My goal:

At least complete the project by utilizing fundamental knowledge, without perfection, but showing a basic understanding of the subject

PROCESS

STEP #1

Connecting to CityBikes API

Chosen Network: bixi-toronto

Result: DataFrame df_bike / 674 rows (bike stations)

Columns: Station_id, Name, Latitude, Longitude, Free_Bikes

Challenge: Difficulties to connect , getting errors 400

STEP #2

Connecting to Foursquare and Yelp

Result: creating fqr_restaurant_df and yelp_restaurant_df providing information about restaurants in bike stations area (r=1km)

Number of rows: 977 (fqr) + 994(yelp)

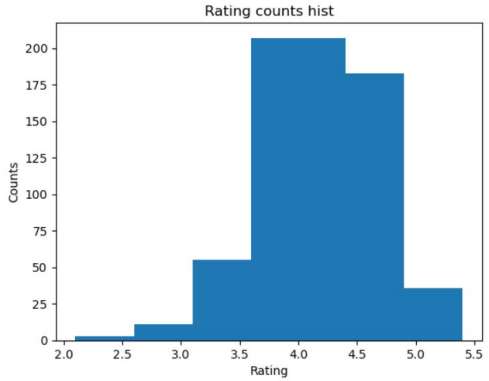
Columns: Station_id, Restaurant Name, Latitude, Longitude, Rating, Distance

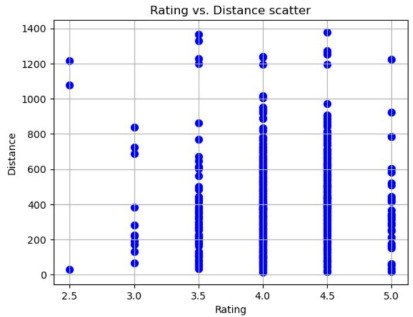
Challenge: Difficulties to connect. Lack of knowledge. Long waiting time.

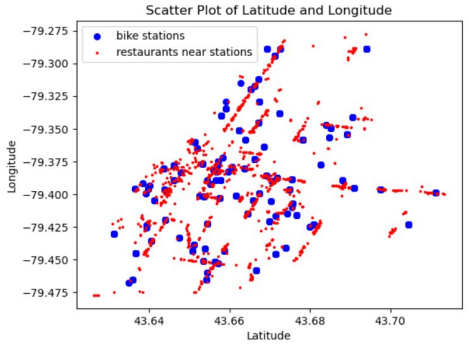
STEP #3

Joining Data

Result: combined (stacked) yelp and foursquare dataframes with further merging with bike_df. Data visualisations. Creation of database combined_data.db.







Challenge: Lack of knowledge. Difficulties to ask the right questions.

STEP #4

Building a Model

Result: Build a regression model that demonstrates a relationship between the number of bikes in a particular location and the characteristics of the restaurant location (distance from bike station).

Interpretations:

R=0.006 in linear regression model means that relationship between 'Distances' and 'Free Bikes' is weak.

Challenge : Lack of time