Formulating Data from Yelp and Foursquare API

Contents

- Data Results from City Bikes, Yelp, Foursquare
- Compare which businesses had the most bike stations near them
- Compare catalogs that had the most bike stations
- Regression Model
- Challenges

Data Results from City Bikes, Yelp and FSQ

City Bikes

Total Locations: 223

Total Bikes: 1616

Yelp

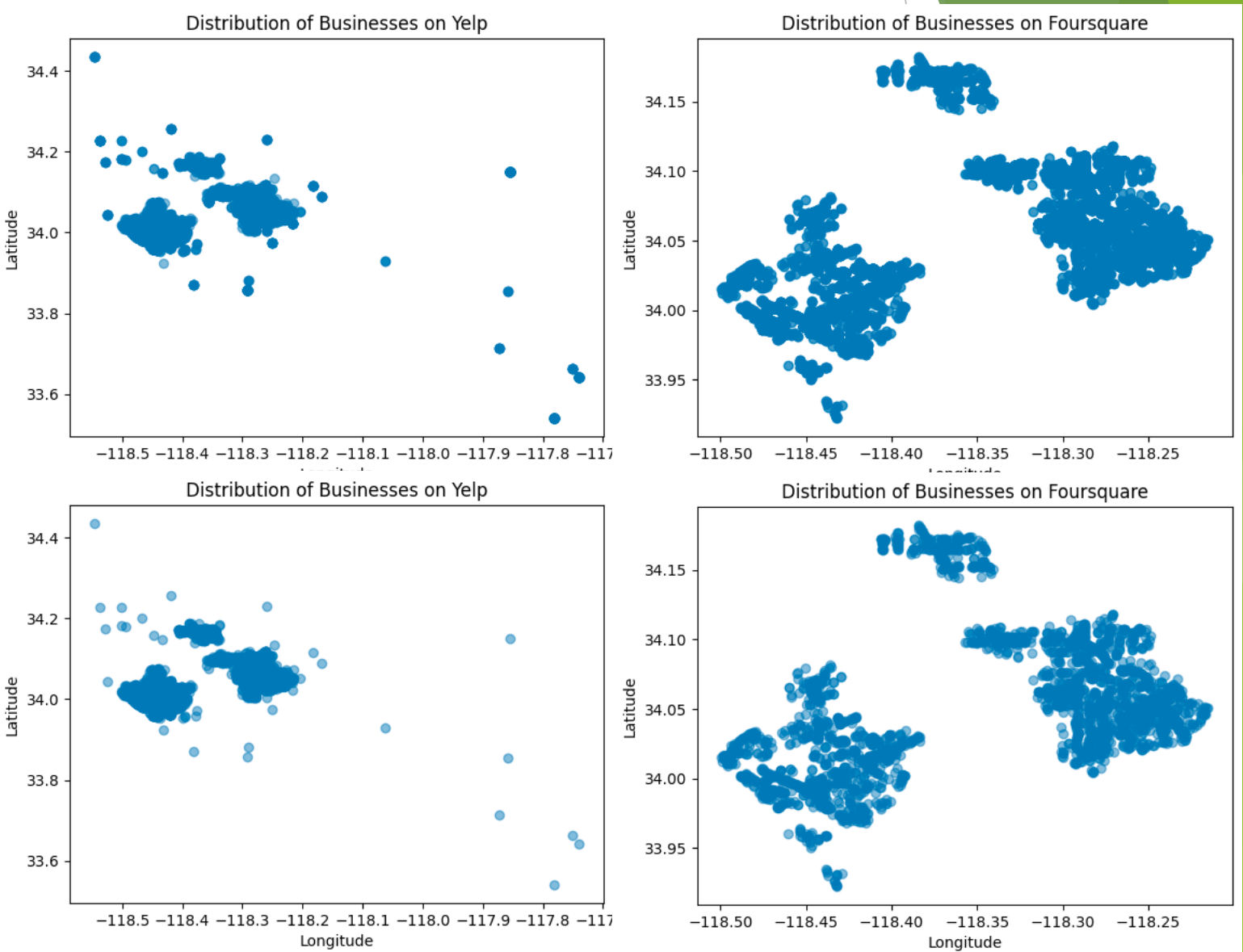
Results from API: 13696

Cleaned Dataframe: 3422

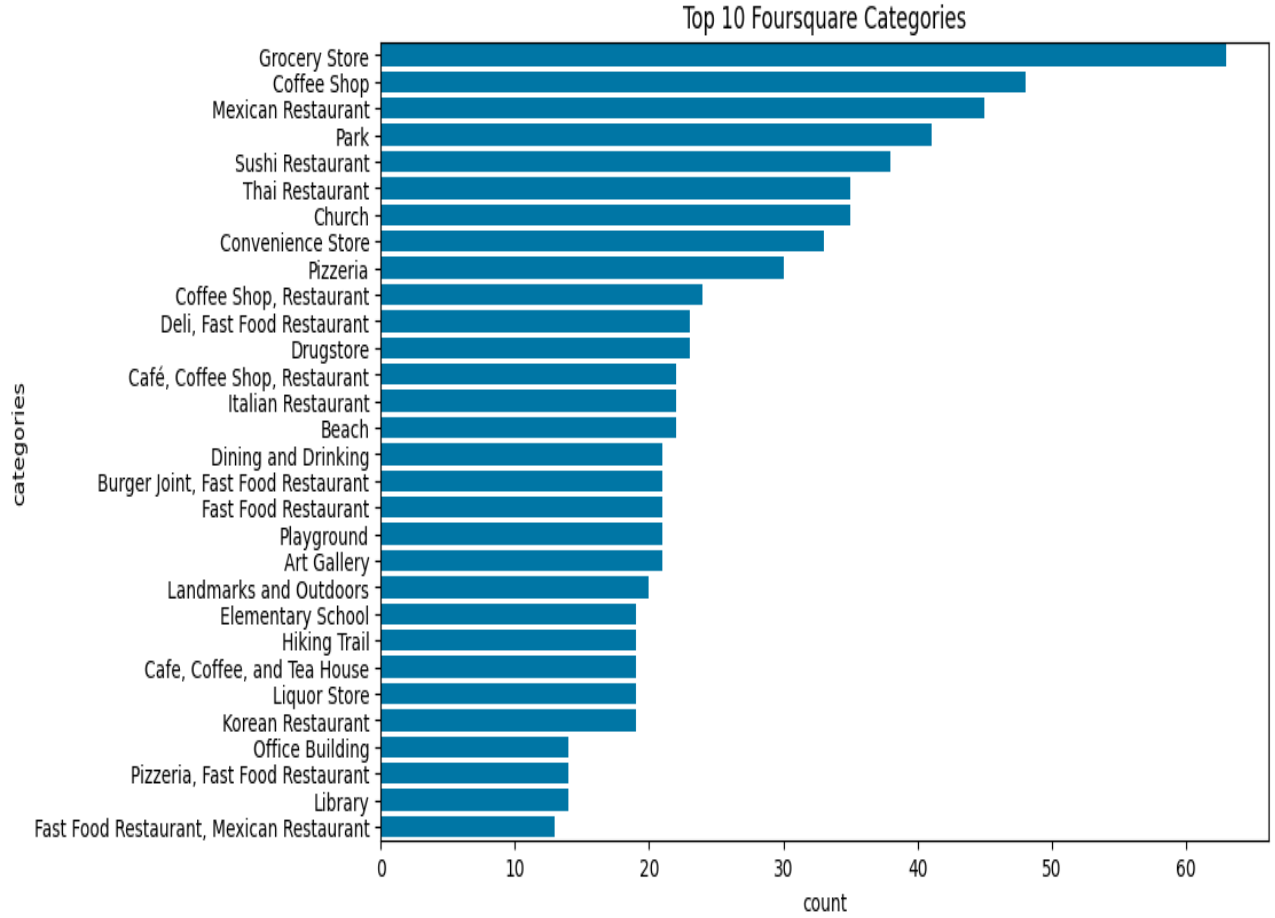
Foursquare

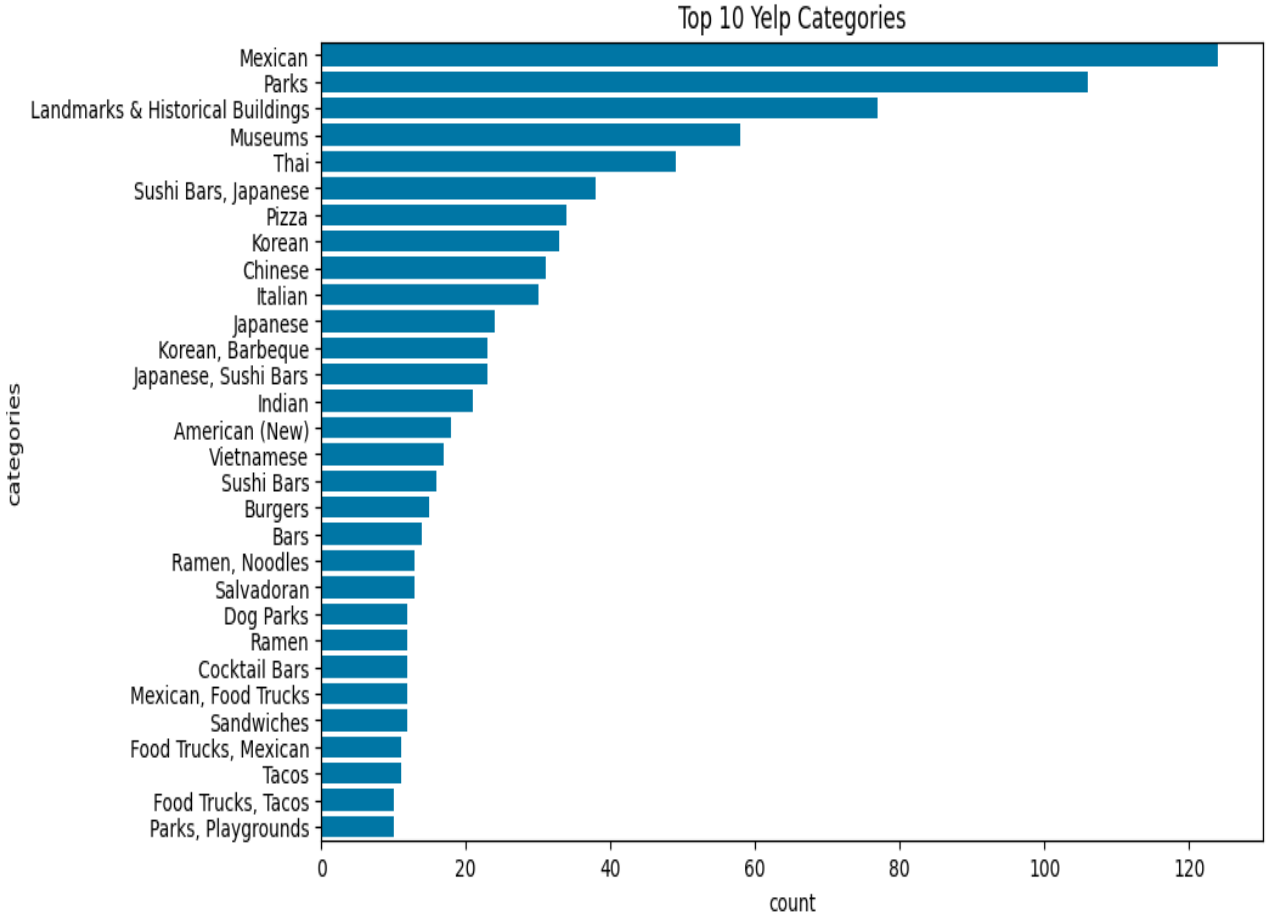
Results from API: 18142

Cleaned Dataframe: 2465

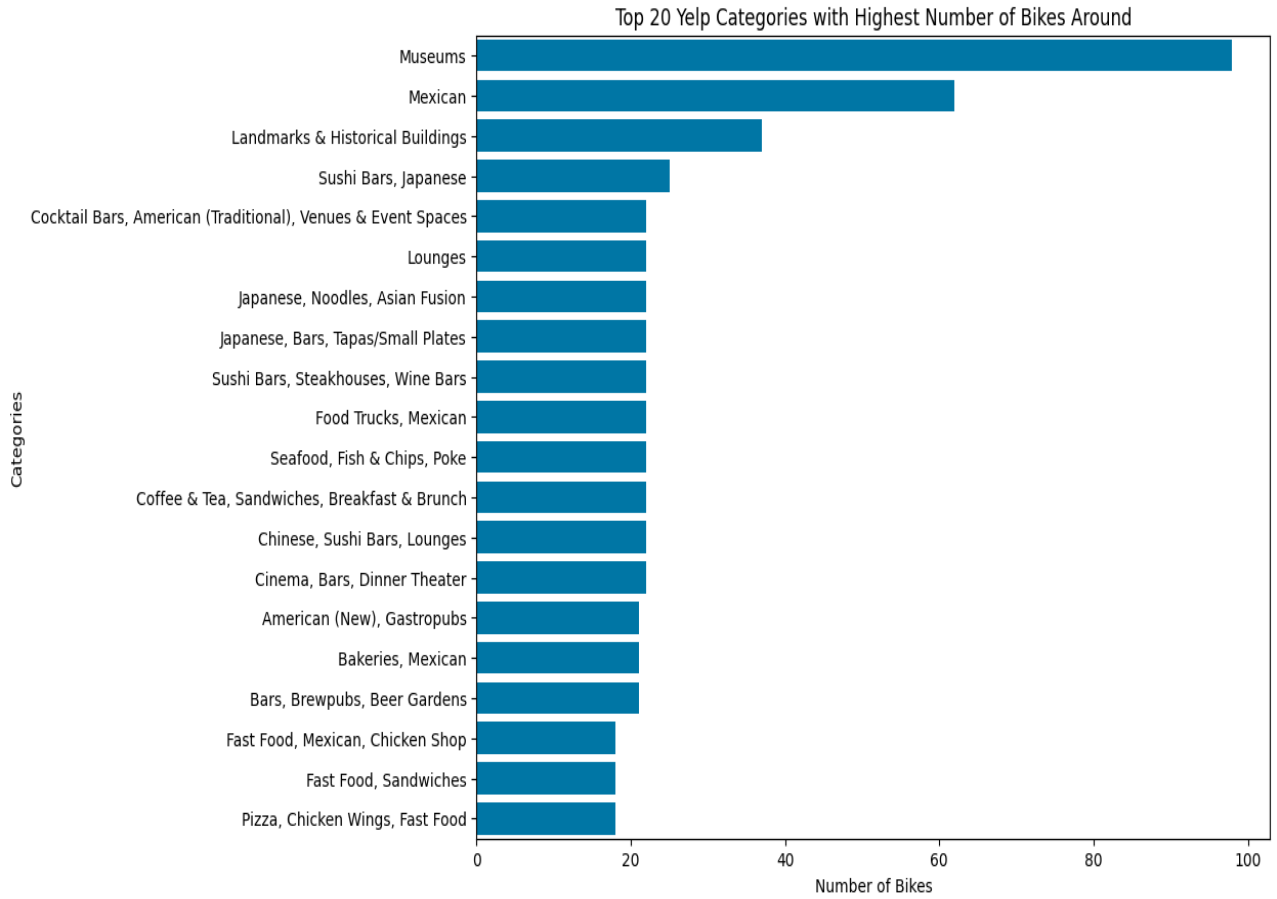


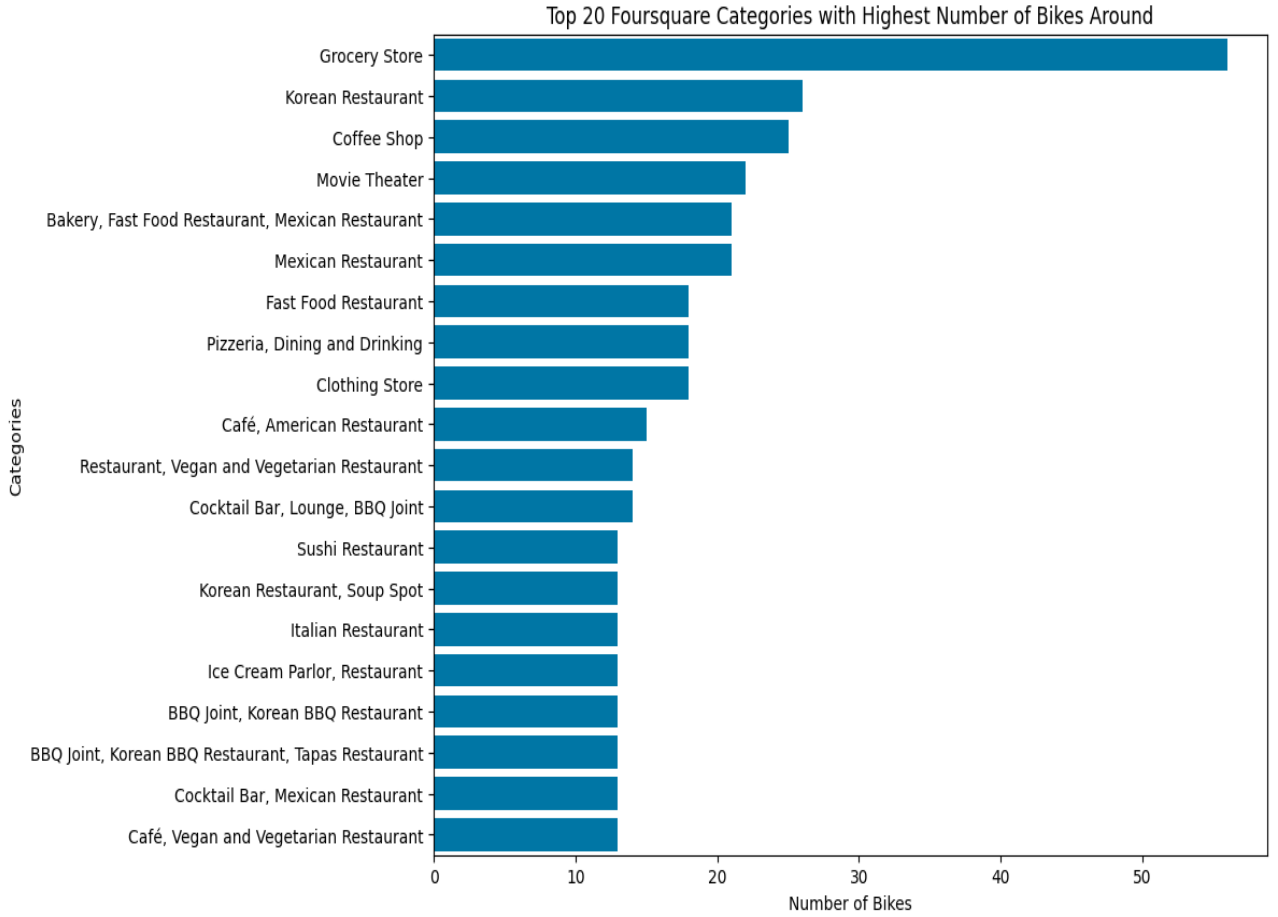
Top Categories



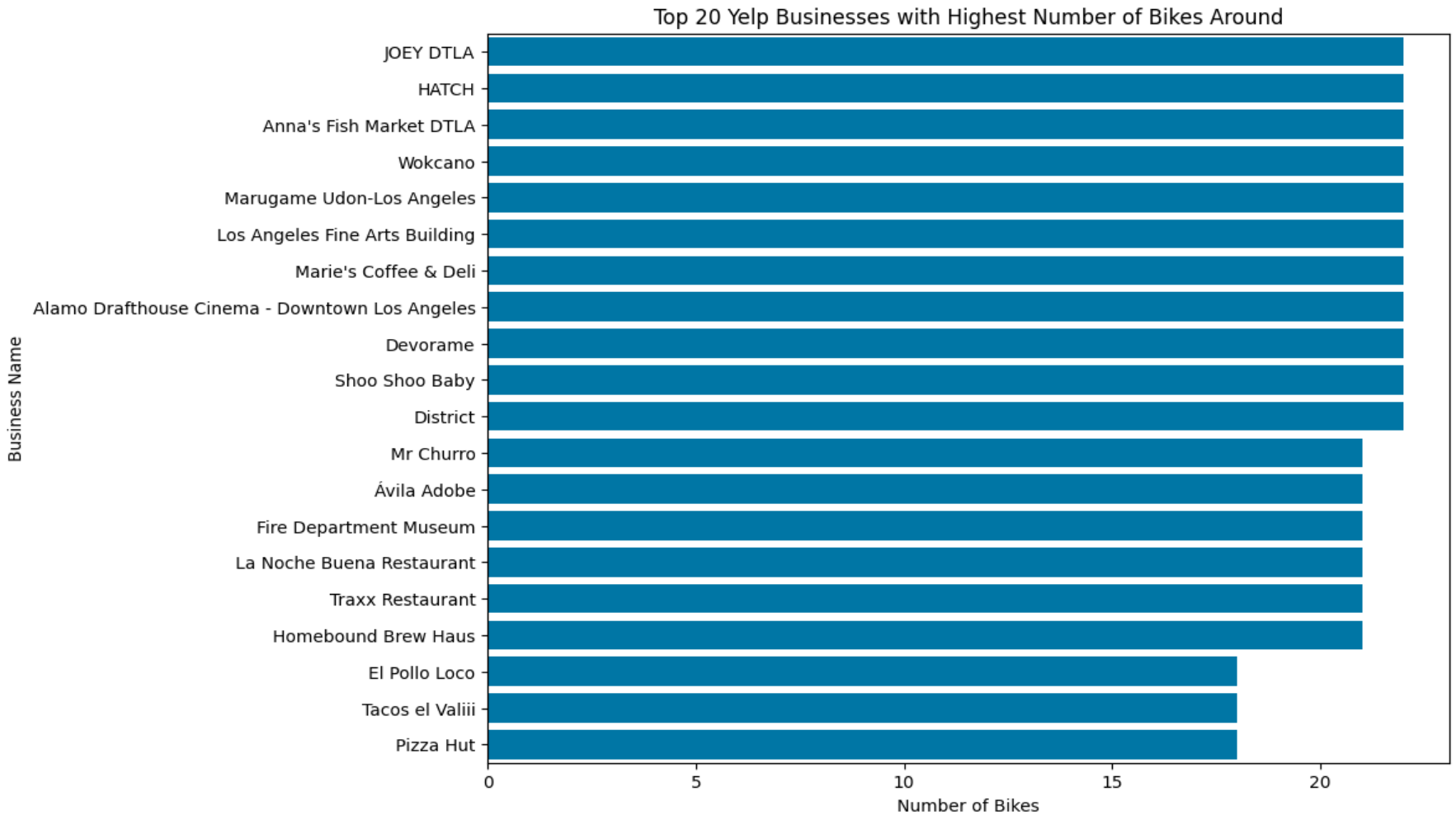


Top Categories For Bike Locations

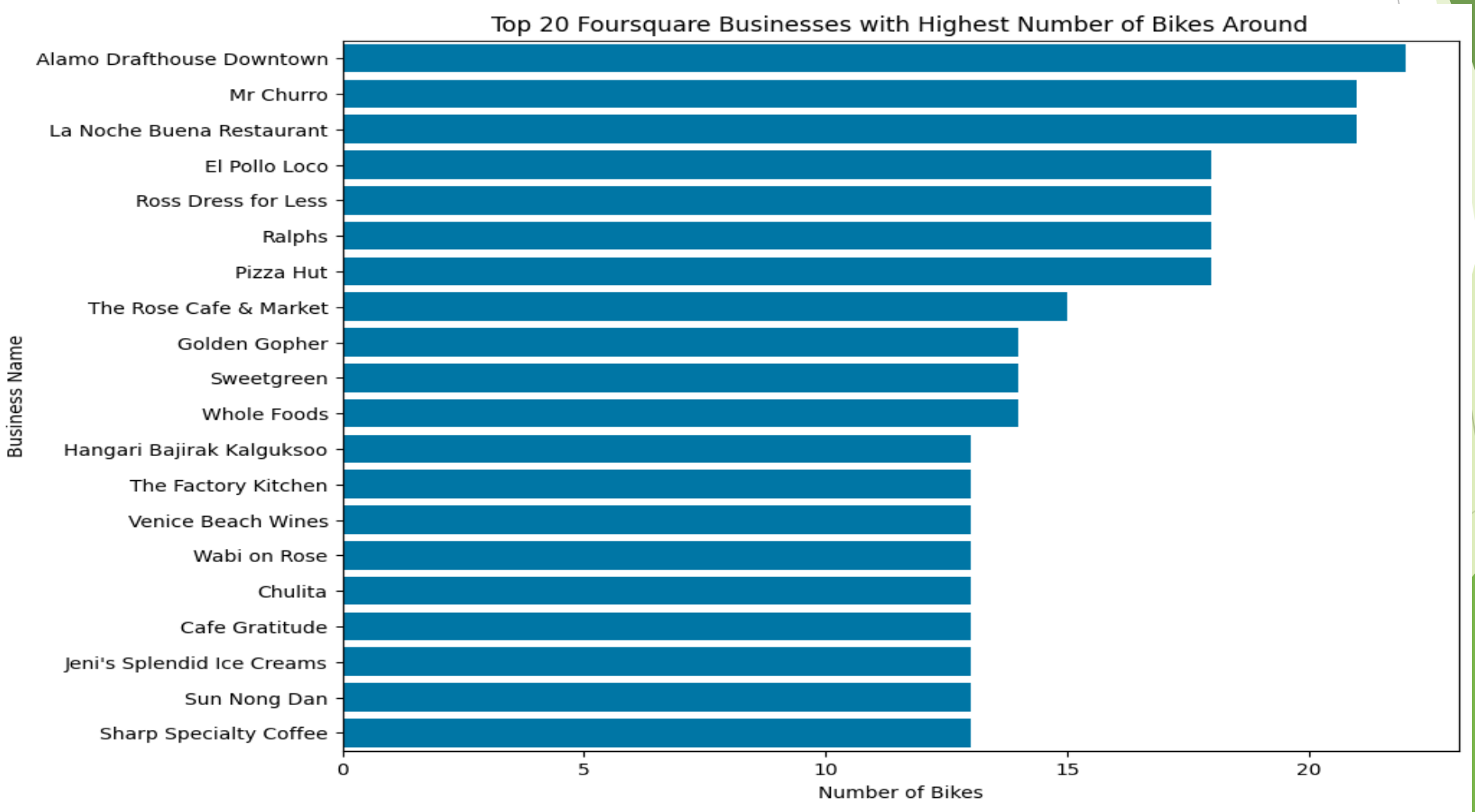




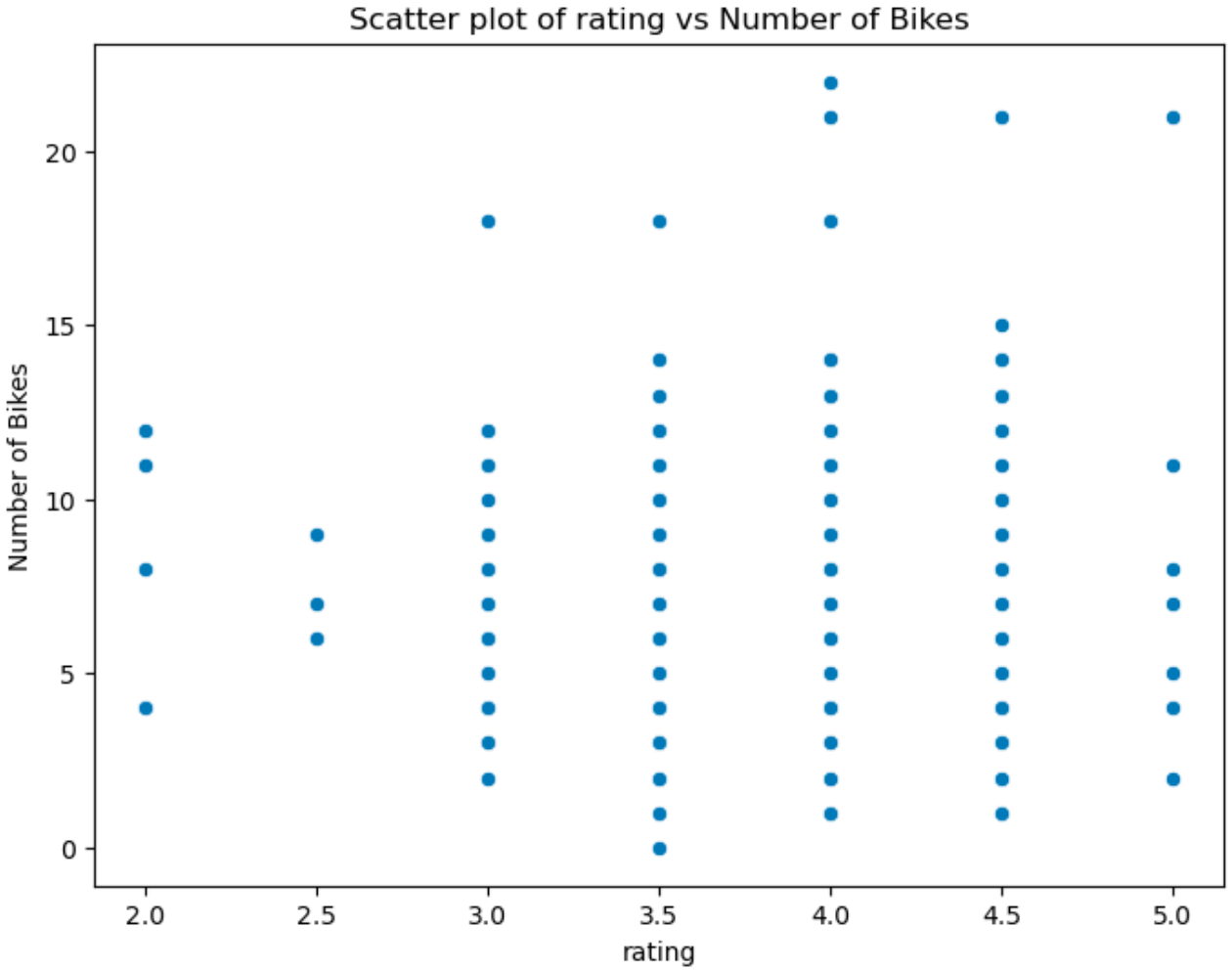
Yelp Top Businesses with the Highest Bike Count

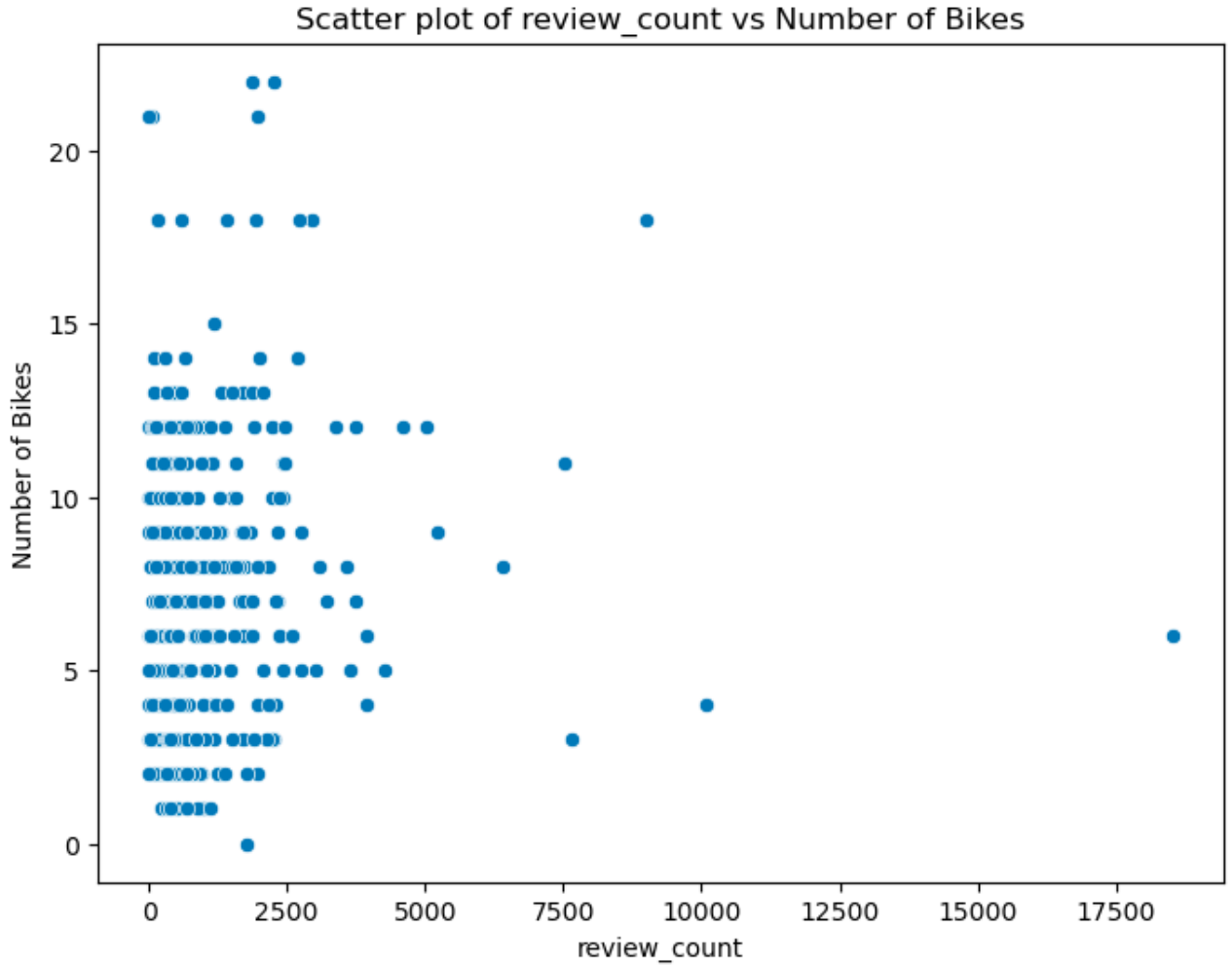


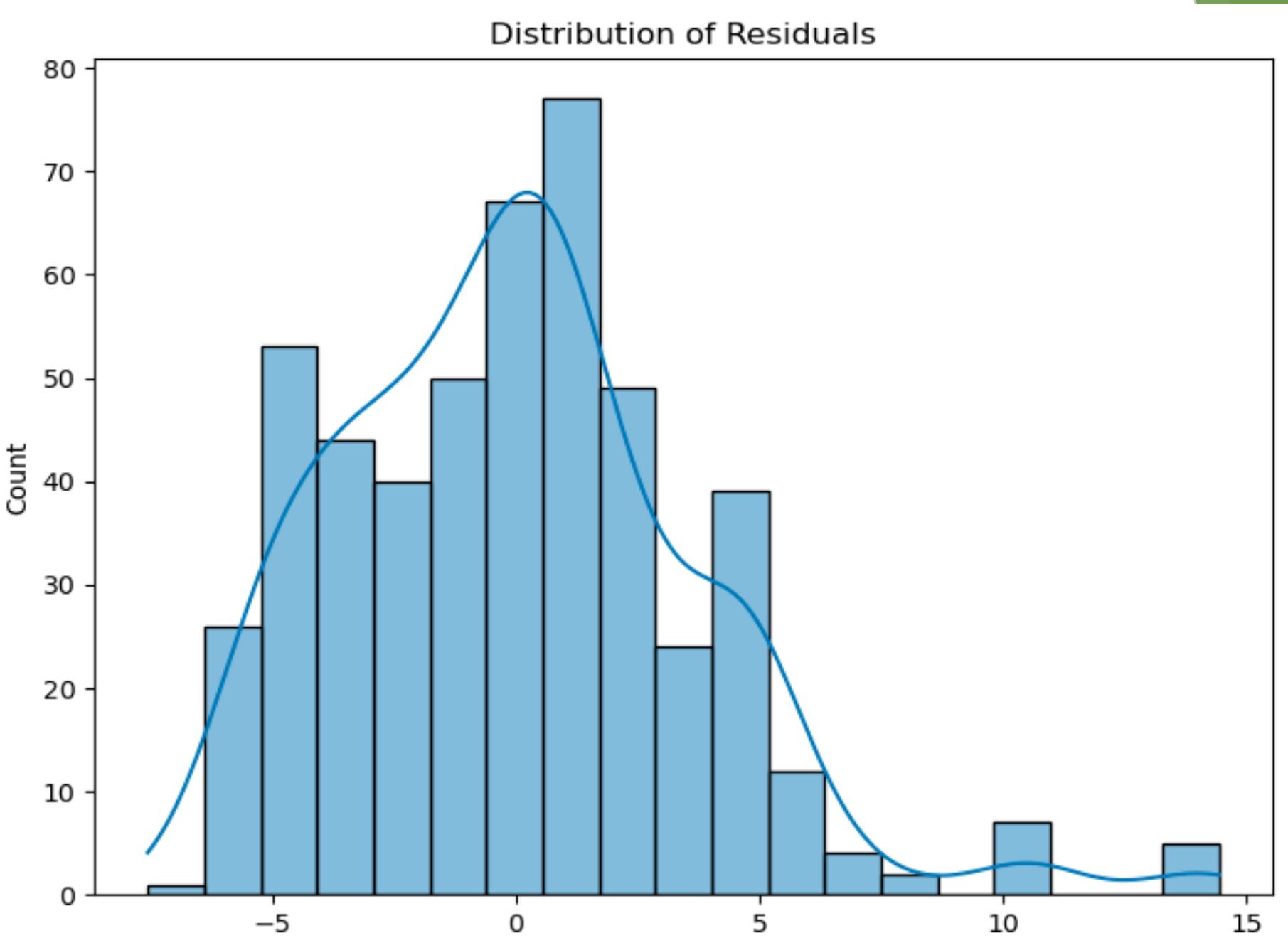
FourSquare Top Businesses with the Highest Bike Count



Regression Model






Distribution of Residuals
Count
80
70
60
50
40
30
20
10
0
−5
0
5
10
15


Residuals vs Fitted Values
15
10
5
0
−5
7
8
9
10
11
12

OLS Regression Results
==========
====== Dep. Variable: Number of Bikes R-squared (uncentered): 0.820 Model: OLS
Adj. R-squared (uncentered): 0.819 Method: Least Squares F-statistic: 565.4 Date:
Fri, 27 Oct 2023 Prob (F-statistic): 3.40e-183 Time: 20:36:08 Log-Likelihood: -
1332.8 No. Observations: 500 AIC: 2674. Df Residuals: 496 BIC: 2690. Df Model: 4
Covariance Type: nonrobust
==========
coef std err t P>|t| [0.025 0.975] ----------
---------- review_count 0.3457 0.157 2.204 0.028 0.037
0.654 rating 0.0006 0.157 0.004 0.997 -0.308 0.309 longitude -5.8266 0.764 -7.629
0.000 -7.327 -4.326 latitude -20.0329 2.654 -7.548 0.000 -25.247 -14.818
==========
Omnibus: 81.917 Durbin-Watson: 2.003 Prob(Omnibus): 0.000 Jarque-Bera (JB):
151.563 Skew: 0.945 Prob(JB): 1.23e-33 Kurtosis: 4.925 Cond. No. 2.18e+03
==========
...
[1] R^2 is computed without centering (uncentered) since the model does not
contain a constant. [2] Standard Errors assume that the covariance matrix of the
errors is correctly specified. [3] The condition number is large, 2.18e+03. This
might indicate that there are strong multicollinearity or other numerical
problems.

Challenges