

Statistical-Modelling- Project

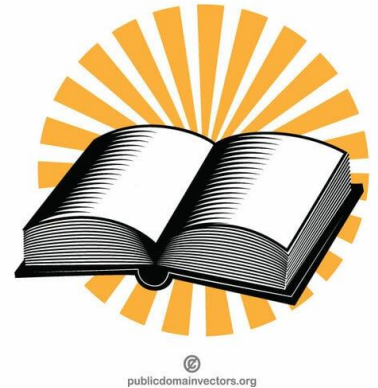
By: Ibadet Azemi

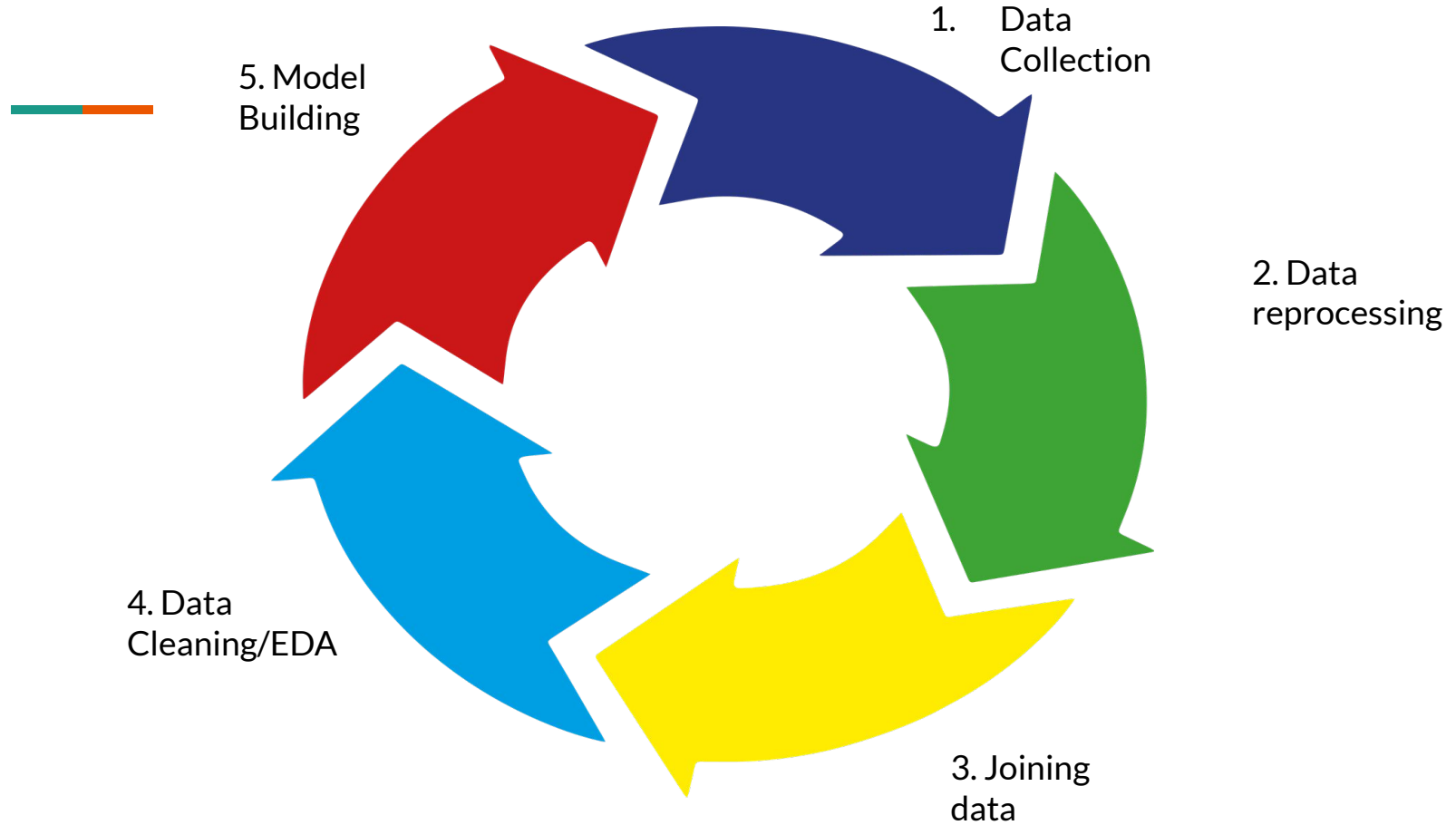


FOURSQUARE

PROJECT/GOALS:

- ❑ Determine the strength of the relationship between number of bikes + POI's (Characteristics: reviews, ratings, distance, location)

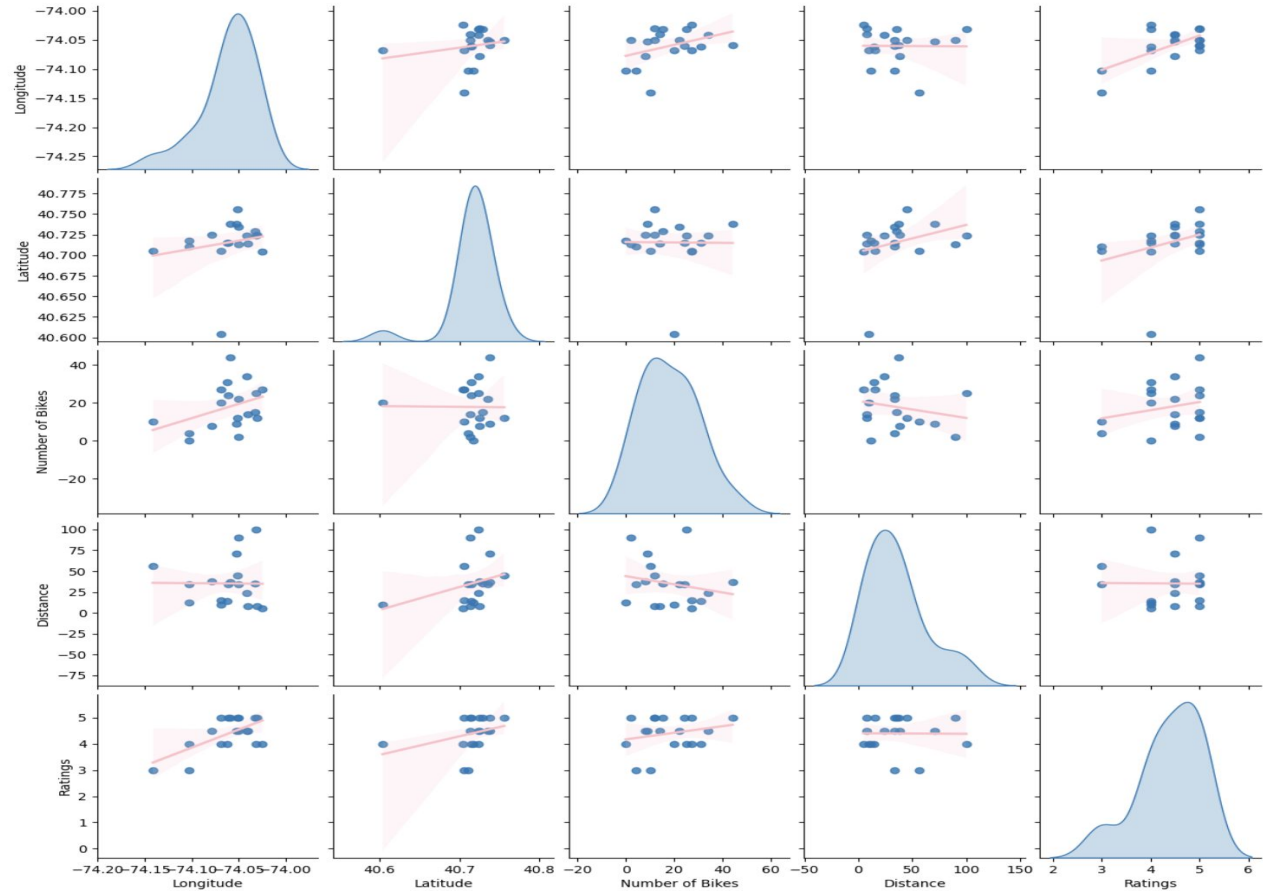




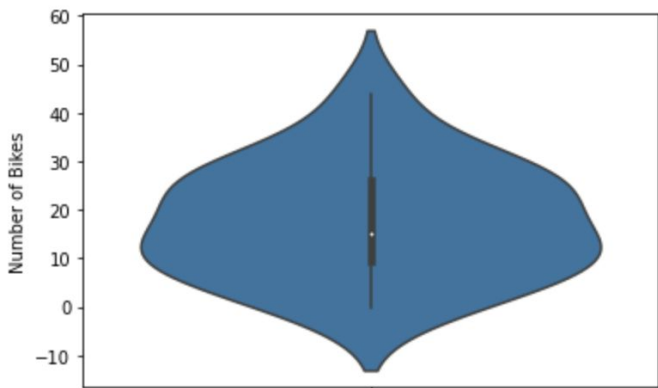
VISUALIZATIONS: EDA

PAIRPLOT:

- ❑ No correlation was found
- ❑ R-Squared: is weak

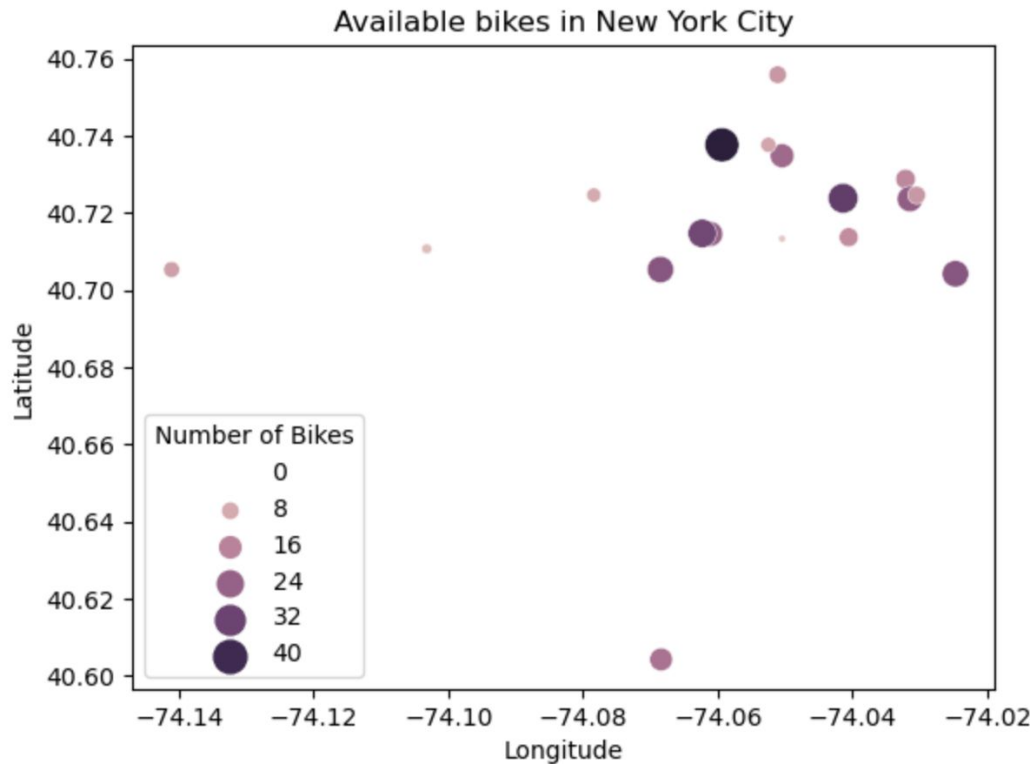


VISUALIZATIONS:EDA



Visual EDA Analysis:

- ❑ There does not seem to be a relationship
- ❑ There are outliers



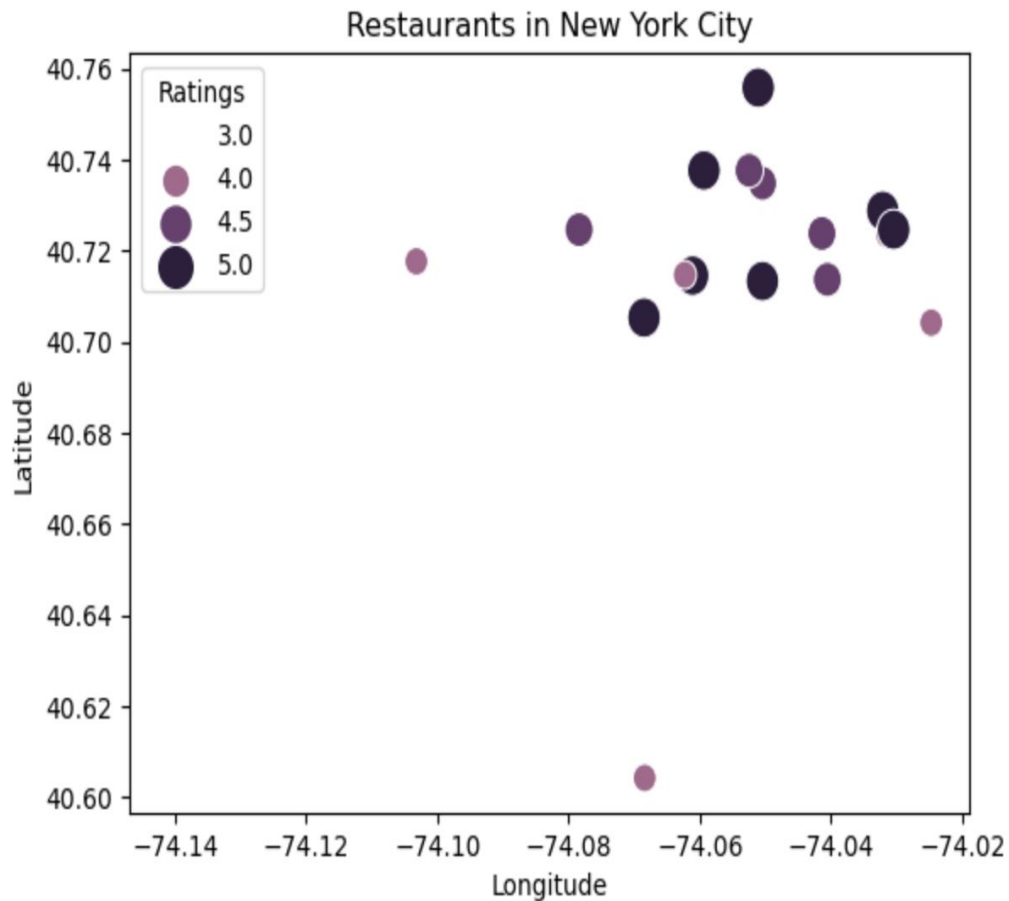
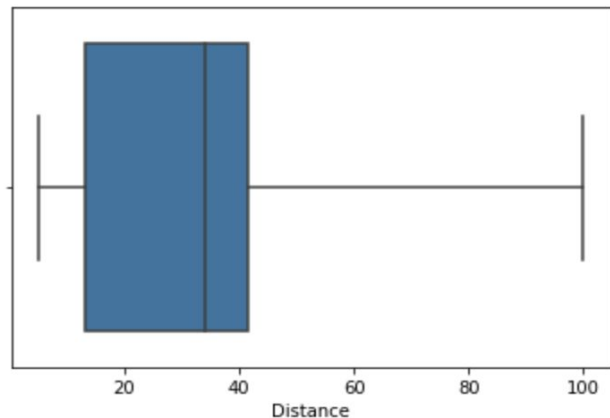
VISUALIZATIONS:EDA



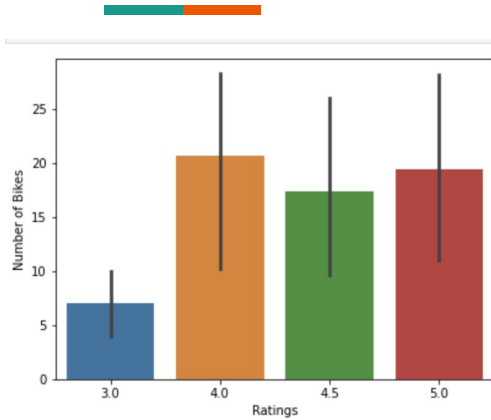
Visual EDA Analysis:

- ☐ There does not seem to be a relationship
- ☐ There are outliers

<Axes: xlabel='Distance'>

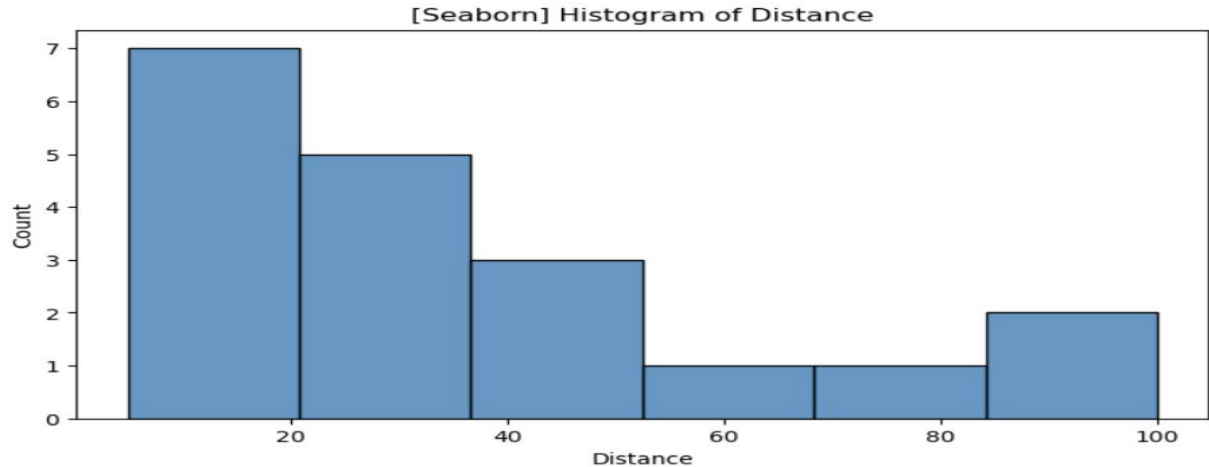
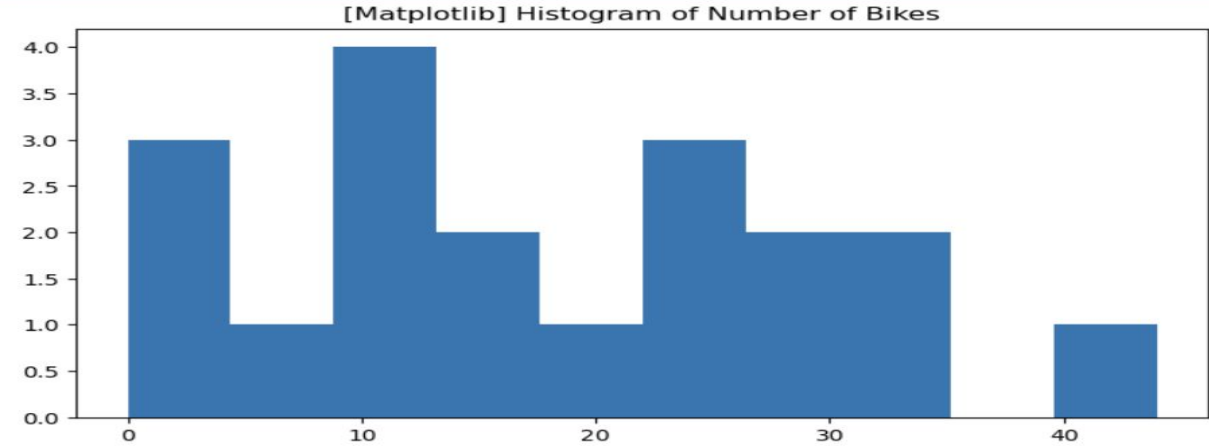


VISUALIZATIONS:EDA



Visual EDA Analysis:

- ❑ There does appear to be a relationship
- ❑ Outliers



Step 4.) Model Building

OLS REGRESSION RESULTS:

Interpretation


Based on the interpretation of the results, it seems that the regression model that was built to predict number of bikes, ratings and distance is performing very well.

The low R-squared value appears to reflect that there is a strong correlation

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified

OLS Regression Results						
Dep. Variable:	Number of Bikes		R-squared:	0.097		
Model:	OLS		Adj. R-squared:	-0.016		
Method:	Least Squares		F-statistic:	0.8584		
Date:	Mon, 21 Aug 2023		Prob (F-statistic):	0.442		
Time:	05:11:46		Log-Likelihood:	-72.294		
No. Observations:	19		AIC:	150.6		
Df Residuals:	16		BIC:	153.4		
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	2.3139	19.832	0.117	0.909	-39.729	44.357
Ratings	4.2563	4.387	0.970	0.346	-5.045	13.557
Distance	-0.0886	0.102	-0.871	0.396	-0.304	0.127
Omnibus:	0.875		Durbin-Watson:	2.087		
Prob(Omnibus):	0.646		Jarque-Bera (JB):	0.850		
Skew:	0.399		Prob(JB):	0.654		
Kurtosis:	2.339		Cond. No.	331.		

R-SQUARED/ADJ R. SQUARED & F-STATISTICS



R-squared:	0.097
Adj. R-squared:	-0.016
F-statistic:	0.8584
Prob (F-statistic):	0.442
Log-Likelihood:	-72.294
AIC:	150.6
BIC:	153.4

R-SQUARED: Is a weak correlation

ADJ. R-SQUARED: Shows model is good at predicting as is equal to R-squared

F-STATISTIC: Is greater than Prob(F-Statistic)

Prob (F-Statistic): Is less than F-Statistic

COEFF & P-VALUE

	coef	std err	t	P> t	[0.025	0.975]
const	2.3139	19.832	0.117	0.909	-39.729	44.357
Ratings	4.2563	4.387	0.970	0.346	-5.045	13.557
Distance	-0.0886	0.102	-0.871	0.396	-0.304	0.127

<> Coefficient: Is positive and has a positive affect on Y

<> T-Statistic: Strong coefficient

<> P-Value: Significance of each and is a strong coefficient



CHALLENGES:

- ❑ Time consuming for a large city
- ❑ 1119 bike stations in NYC
- ❑ Lack of bikes
- ❑ Limited calls for Yelp
- ❑ Limited API calls
- ❑ Learning and understanding API's
- ❑ Converting json response to csv
- ❑ Combining multiple data frames/csv



RESULTS:

- ❑ <> Coefficient: Is positive and has a positive affect on Y
- ❑ <> T-Statistic: Strong coefficient
- ❑ <> P-Value: Significance of each and is a strong coefficient

The regression model that was built to predict number of bikes, ratings and distance is performing very well.

The low R-squared value appears to reflect that there is a strong correlation



FUTURE GOALS:

- ❑ More data cleaning
- ❑ Going more in depth with the process
- ❑ More POI's
- ❑ Better/more visualizations(EDA)



THANK YOU!



FOURSQUARE