

Eating good in the neighborhood

by Amit Ohana &
Leor Marshall
Holon institute of
technology





A wise man once said:
"knowledge is power"



PROLOGUE

Owning a restaurant is about a passion to food.

These days, with the variety of restaurants available,
and the easy access of the internet.

Getting a good review is very important to the business
owner.

Every platform matters ,Yelp, Google, Trip advisor ,Zap,
and etc.



our research question is:

"can we predict if a new restaurant in the US will get a good grade in YELP reviewing website given we know her amenities ?"



In our project, we took each restaurant data, explored it while using various data models and graph, and tried to figure out from the data useful information for the future business owner.



Data acquisition

<https://www.yelp.com/>



while using Selenium

The Selenium logo consists of a green square containing the letters "Se" in white, with a white checkmark positioned above and to the right of the "e". To the right of the logo, the word "Selenium" is written in a large, green, sans-serif font.

for each restaurant we
scrape data to help us figure
out our question.

restaurant page example

Lily
4.5 stars 312 reviews
Claimed • Vietnamese Edit
Closed 5:00 PM - 9:00 PM Hours updated 3 weeks ago See hours
Write a review Add photo Share Save

Menu

Popular dishes

View full menu >

Garlic Noodles 44 Photos • 63 Reviews
 Shaking Beef Salad 40 Photos • 35 Reviews
 Banh Mi 38 Photos • 37 Reviews
 Jackfruit Pizza 31 Photos • 26 Reviews

[Website menu](#)

Location & Hours

Map data ©2023 Google

225 Clement St, San Francisco, CA 94118 4th Ave & 3rd Ave, Inner Richmond

Get directions

Day	Open	Close
Mon	11:00 AM - 2:00 PM 5:00 PM - 9:00 PM	
Tue	Closed	
Wed	5:00 PM - 9:00 PM	
Thu	5:00 PM - 9:00 PM	Closed now
Fri	11:00 AM - 2:00 PM 5:00 PM - 9:00 PM	
Sat	11:00 AM - 2:00 PM 5:00 PM - 9:00 PM	
Sun	11:00 AM - 2:00 PM 5:00 PM - 9:00 PM	

[Edit business info](#)

You Might Also Consider Sponsored

Sowl Bowls 4.5 stars 72
Gabriel C. said "This place is sooo good. Wish I could help them out more (they open in the day during the weekdays) and order more from here, they seem like this cute couple (don't know if they are) running a humble little place in the middle of a..." [read more](#)

Dumpling Kitchen 4.5 stars 2001
Lena R. said "I'll be honest, I was a little hesitant about whether or not to write a review for this place, since I didn't want to hike up the number of Yelp reviews for this awesome place and then have it become one of those "over hyped"..." [read more](#)

Amenities and More

Health Score PASS Powered by Hazel Analytics

Takes Reservations

Offers Delivery

Offers Takeout

[29 More Attributes](#)

we scraped things like:

Name,
Number Of Reviews
Rating
Health Score
Takes Reservations
Offers Takeout
Accepts Credit Cards
Outdoor Seating
Offers Delivery
Vegan Options
and many more.

Amenities and More

 Offers Delivery

 No Reservations

 Offers Takeout

 Pescatarian Options

12 More Attributes

problem in extracting data

solving it using selenium



Initial analysis and reclamation

working on the DataFrame :



we filled missing numbers



delete duplicate restaurants

deleted parameters that do not exist

left out only the relevant data

Name	Reviews	Rating	Health Sco	Takes Rese	Offers Takeout	Proof of va	All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears	Heated ou	Covered ou	Vegan Opti	Many Vege	Classy	Casual	Moderate	Good for G
0 Moon Rab	125 review	4.5 star rating	Takes Rese Offers Takeout					Offers Delivery							Vegan Opti	Many Vegetarian Options				
1 dâ€™Alba	186 review	4.5 star rating	Takes Rese Offers Takeout				Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks					Vegan Opti	Many Vegetarian Options			Moderate	Good for G
2 Squable	417 review	4.5 star rating	Takes Rese Offers Takeout				Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks					Vegan Opti	Many Vegetarian Options			Moderate	Good for G
3 La Lucha	691 review	4.5 star rating	Takes Rese Offers Takeout				Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks					Vegan Opti	Many Vegetarian Options			Moderate	Good for G
4 The Gypsy	38 reviews	4.5 star rating	Takes Rese Offers Takeout				Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks					Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
5 Gatlin's Fin	88 reviews	4 star rating	Takes Rese Offers Takeout				Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks					Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
6 Dona Leti's	45 reviews	4 star rating	Takes Rese Offers Takeout				Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks					Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
7 Hi Trendy	151 review	4.5 star rating	Takes Rese Offers Takeout				All staff full	Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks				Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
8 Millies Kitc	135 review	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks				Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
9 Coltivare	1109 review	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks				Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
10 Crisp	1023 review	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks				Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
11 Sao Lao Th	143 review	4.5 star rating	Takes Rese Offers Takeout				All staff full	Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks				Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
12 LES BA'GET	316 review	4.5 star rating	Takes Rese Offers Takeout				All staff full	Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks				Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
13 Bluestone	188 reviews	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Credit Cards	Outdoor Se	Offers Deli	Staff wears masks				Vegan Opti	Many Vegetarian Opti Casual			Moderate	Good for G
14 Alice Blue	317 review	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
15 Heights & C	50 reviews	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
16 Piperâ€™s	73 reviews	4.5 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
17 The Union	469 review	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
18 Elliot's Tab	227 review	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
19 Gagootz	15 reviews	4.5 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
20 Ploughmar	180 review	4.5 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
21 Killer Nood	128 review	4 star rating	Takes Rese Offers Takeout				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
22 Savoir	373 review	4 star rating	Takes Rese Offers Takeo				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
23 The Post Br	58 reviews	4.5 star rating	Takes Rese Offers Ta				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
24 Field & Tid	581 review	4 star rating	Takes Rese Offers T				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
25 Traveler's	11268 review	4.5 star rating	Takes Rese Offer				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
26 Ribeye & R	279 review	3.5 star rating	Takes Rese Offe				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G
27 Hughie's	982 review	4.5 star rating	Takes Rese C				All staff full	Accepts Cr	Street Park	Outdoor Se	Offers Deli	Staff wears masks			Vegan Opti	Many Vege Classy			Moderate	Good for G

before and after cleaning

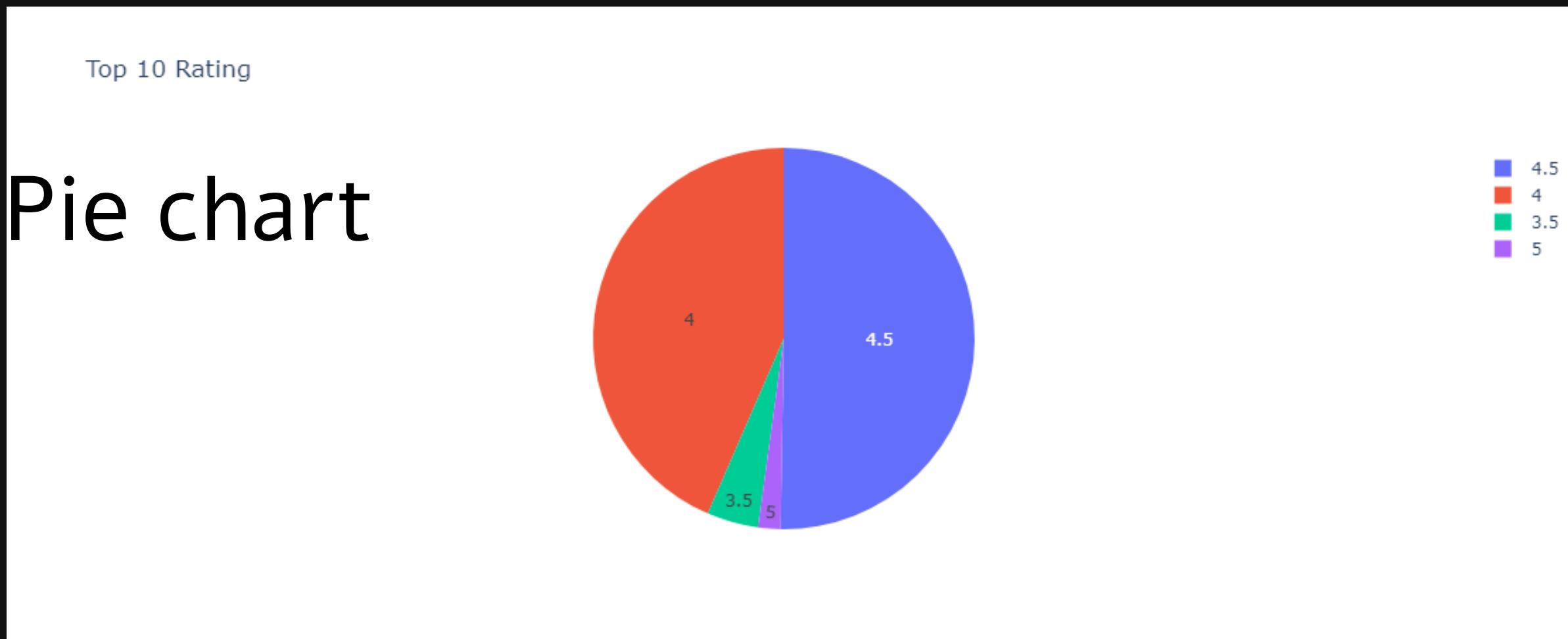
Visualization & EDA

creating a heat map,histogram,scatter plot,bar plot &Pie plot
that shows relations between restaurant amenities

finding positive correlations

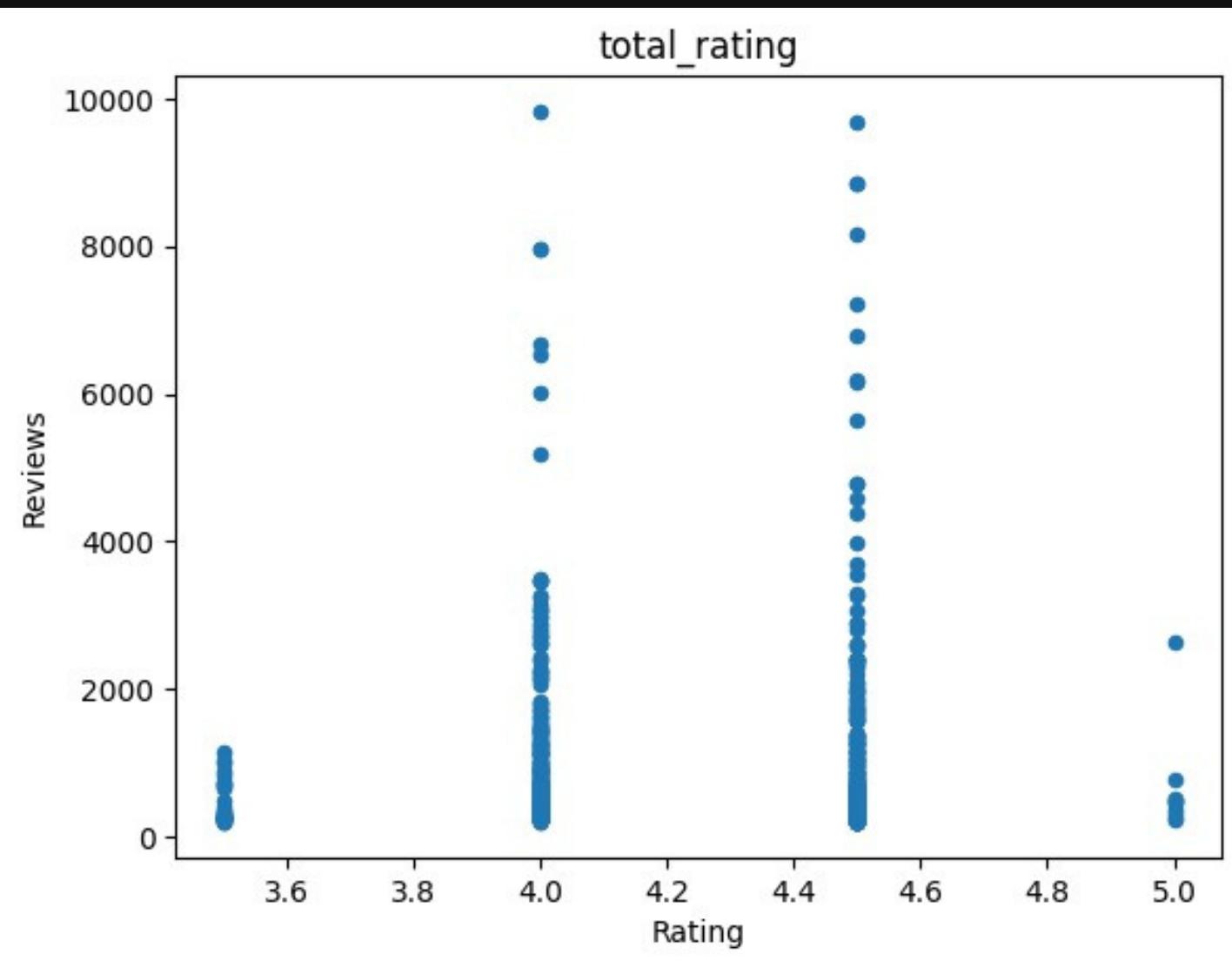
finding negative correlations

Visualization & EDA

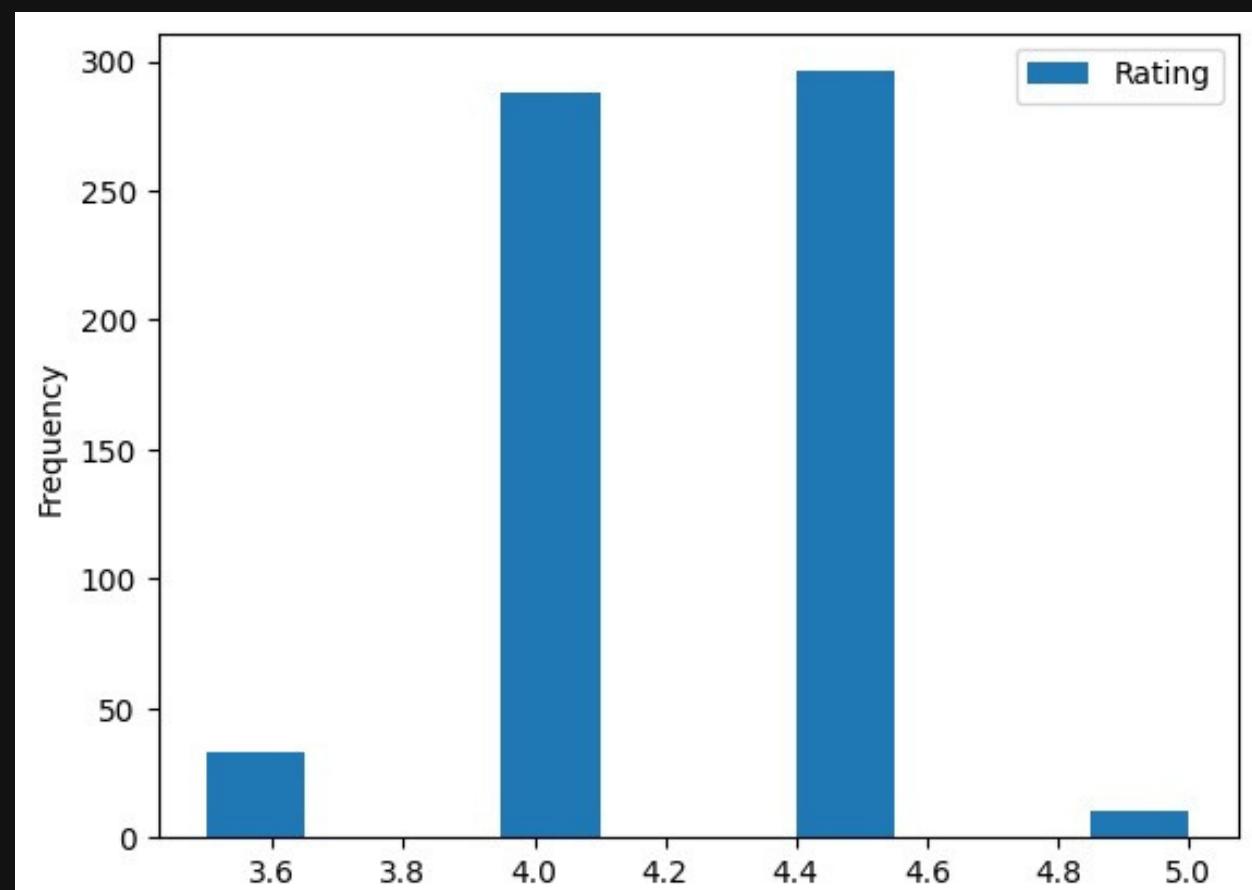


Most of the ratings are between 4 and 4.5, we will check if there are characteristics that affect this

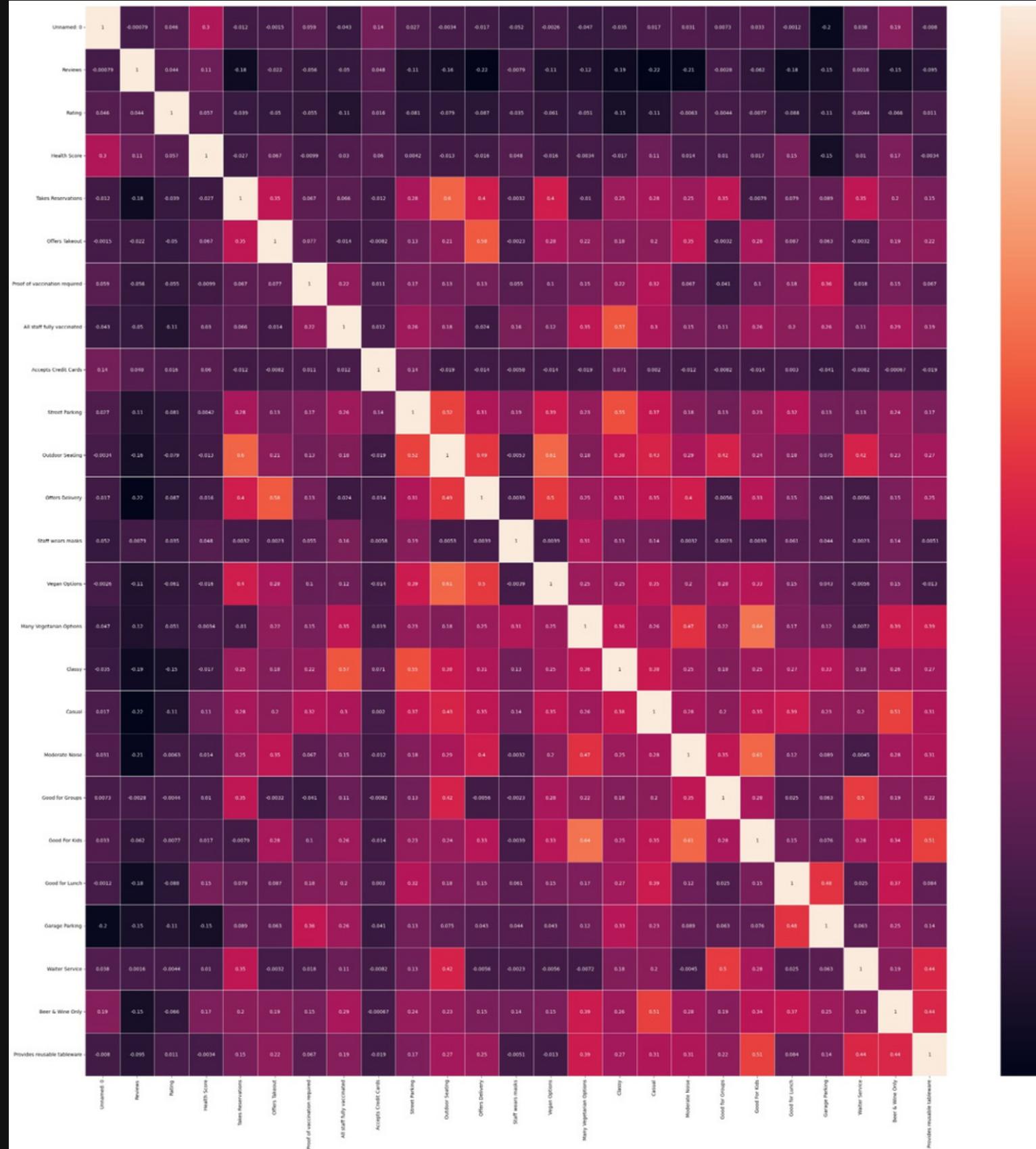
Visualization & EDA



scatter plot & bar plot
**majority of the reviews
are between 4-5- and
there is no connection
between number of
reviews to the grade
itselfe**

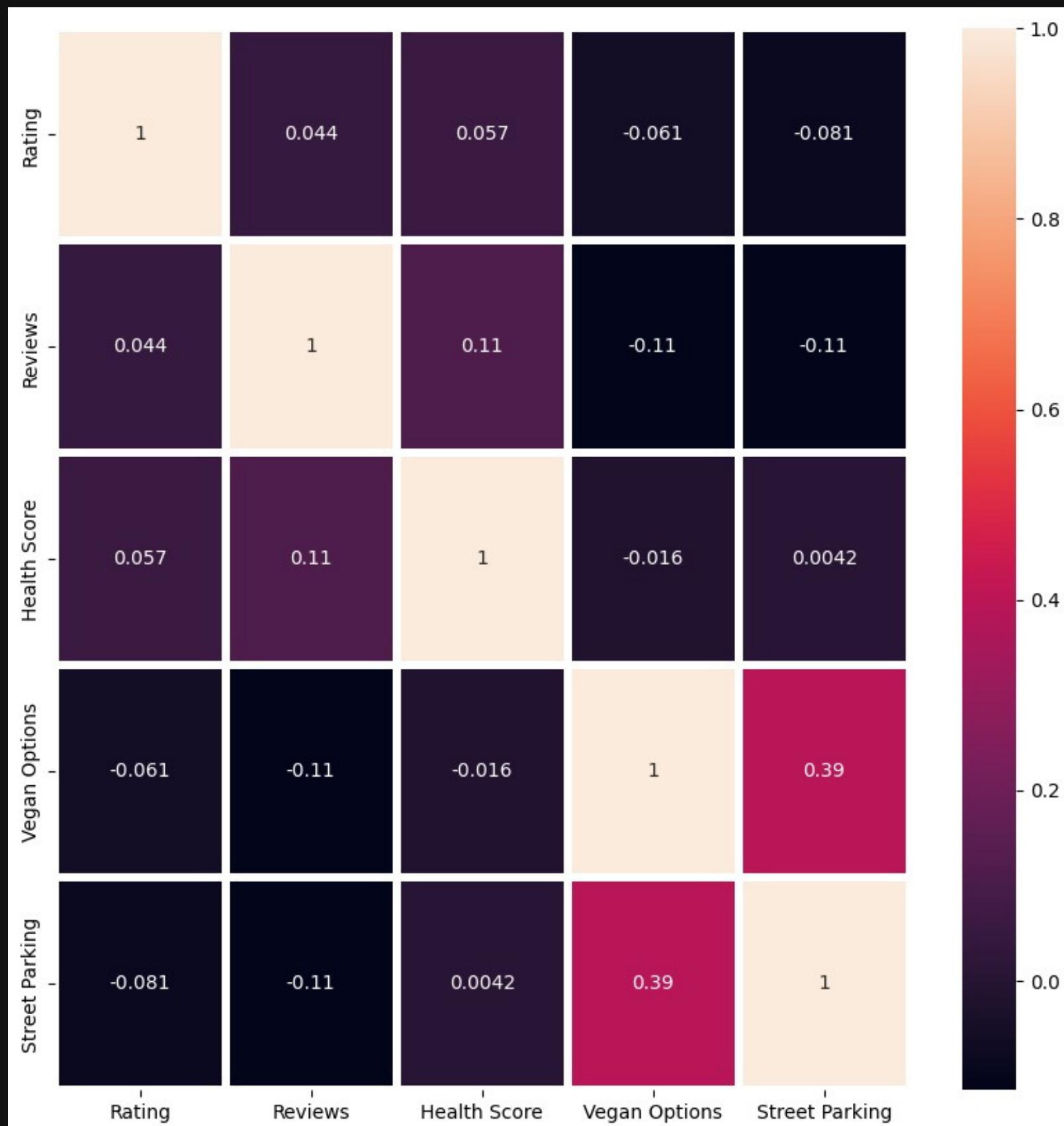


Visualization & EDA



HEAT MAP
**Apparently there is no parameter
 that is
 important to influence
 the rating of the restaurant.**

Visualization & EDA



Heat Map

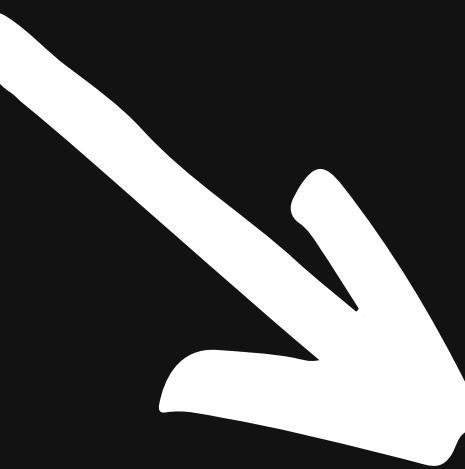
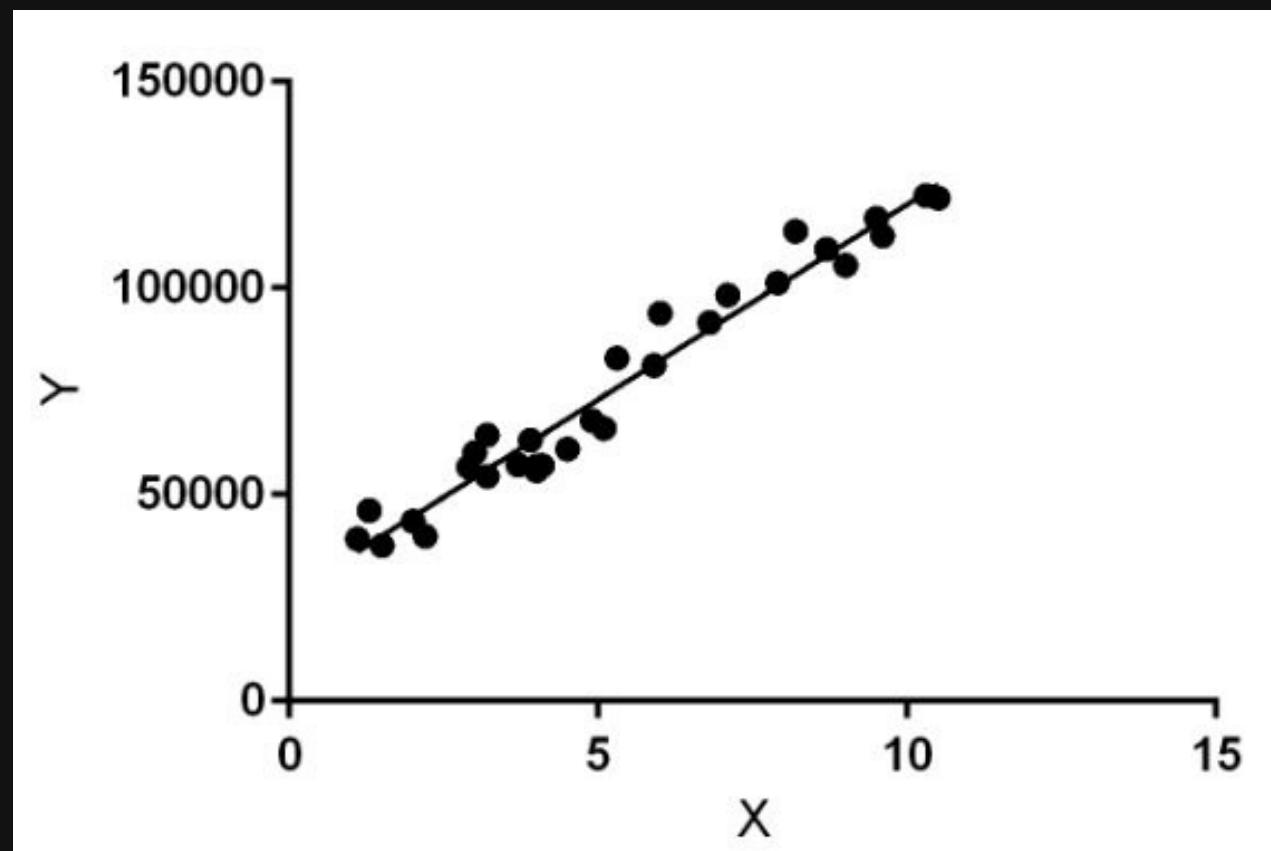
The values that we think will have the greatest impact on the restaurant's rating, and here too you can see that there is no influence on each other

EDA conclusion

As we have seen there is no substantial relationship between the different variables

Machine learning methods

Regression problem



LinearRegression

Linear Regression

```
# fit the model on the training data
trained_model = linear_model.LinearRegression().fit(X_train, y_train)

# predict on the training and test data
y_pred_train = trained_model.predict(X_train)
y_pred_test = trained_model.predict(X_test)

# evaluate the performance of the model
mse_train = mean_squared_error(y_train, y_pred_train)
mse_test = mean_squared_error(y_test, y_pred_test)

r2_train = r2_score(y_train, y_pred_train)
r2_test = r2_score(y_test, y_pred_test)
### as we can see the correlation is vary low in all models
print("Mean squared error on train data: {:.2f}".format(mse_train))
print("Mean squared error on test data: {:.2f}".format(mse_test))
print("R^2 score on train data: {:.2f}".format(r2_train))
print("R^2 score on test data: {:.2f}".format(r2_test))
```

```
Mean squared error on train data: 0.13
Mean squared error on test data: 0.12
R^2 score on train data: 0.03
R^2 score on test data: -0.06
```

Linear Regression

```
1 model = LinearRegression()
2 X_train, X_test, y_train, y_test = train_test_split(df_X, df_y, test_size=0.2, random_state=42) # reducing test_size to 0.2
3 model.fit(df_X,df_y)
4
5 scores = cross_val_score(model, df_X, df_y, cv=5)
6 score = model.score(X_test, y_test)
7 y_pred = model.predict(X_test)
8
9 print("Test score:", score)
10 print("Cross-validation scores:", scores)
11 print("Mean of cross-validation scores:" ,scores.mean())
12 print("Standard deviation of cross-validation scores::" ,scores.std())
13 ###cross-validation also Low, Mean squared error was the stronger but does not say anything.
```

```
Test score: 0.056757186022171724
Cross-validation scores: [-0.15060774 -0.14537576 -0.07834077 -0.24131182 -0.11970912]
Mean of cross-validation scores: -0.1470690418349117
Standard deviation of cross-validation scores:: 0.0536082688427717
```

summary & conclusion

Our research question was: "Is it possible to predict the rating of a new restaurant in the US on the Yelp website given the restaurant's amenities ?"

After the machine learning in which we attacked the problem from several different angles, we see that it is not possible to predict, given the data from the website, whether a new restaurant that opens in the US will receive a high rating on the Yelp website.

THE END

The end