

Google Maps: Unsupervised Learning on Restaurants in PA

Zimeng Liu

Data Analytics Department at Dickinson College

Problem Statements

Data Source





Restaurant Data:

- Information on restaurants in Carlisle, Pittsburgh, Harrisburg, and Philadelphia.
- (restaurant name, rating, number of ratings, category, and address)
- 118 restaurants in Carlisle, 191 restaurants in Pittsburgh, 160 restaurants in Harrisburg, 181 restaurants in Philadelphia
- By 03/30/2023

Review Data:

- Information on restaurant reviews in Carlisle.
- (restaurant name, reviewer, time, rating, comment)
- 32 / 118 restaurants in Carlisle
- 16897 reviews
- By 04/12/2023

* minimum number of rating = 500 and category ≠ Fast Food

	 Carlisle	 Pittsburgh	 Harrisburg	 Philadelphia
# of Restaurants	118	191	160	181
Mean of Rating	4.2	4.4	4.3	4.4
Mean of Rating Number	529	777	453	791
Restaurant Max #	Middlesex Diner (3000)	Primanti Bros. Restaurant (5900)	Fiesta Mexico (3100)	Fogo de Chão Brazilian Steakhouse (5000)
Dominant Category	Pizza (14)	Italian (14)	Mexican (13)	Italian (21)
# of Restaurants*	37	88	48	85
Highest Rating*	4.7	4.8	4.8	4.9
	Hamilton Restaurant Redd's Smokehouse BBQ Seve-N-Dots	Arepittas	Alvaro Bread & Pastry Shoppe	Cilantro Mediterranean Cuisine

Data Processing

Restaurant Data:

- Remove duplicated data
- Modify the `category` variable by keeping 7 main categories: Italian, Pizza, American, Mexican, Chinese, Sandwich, and Fast Food.
- Convert the data type of the `number of rating` variable from string to integer.

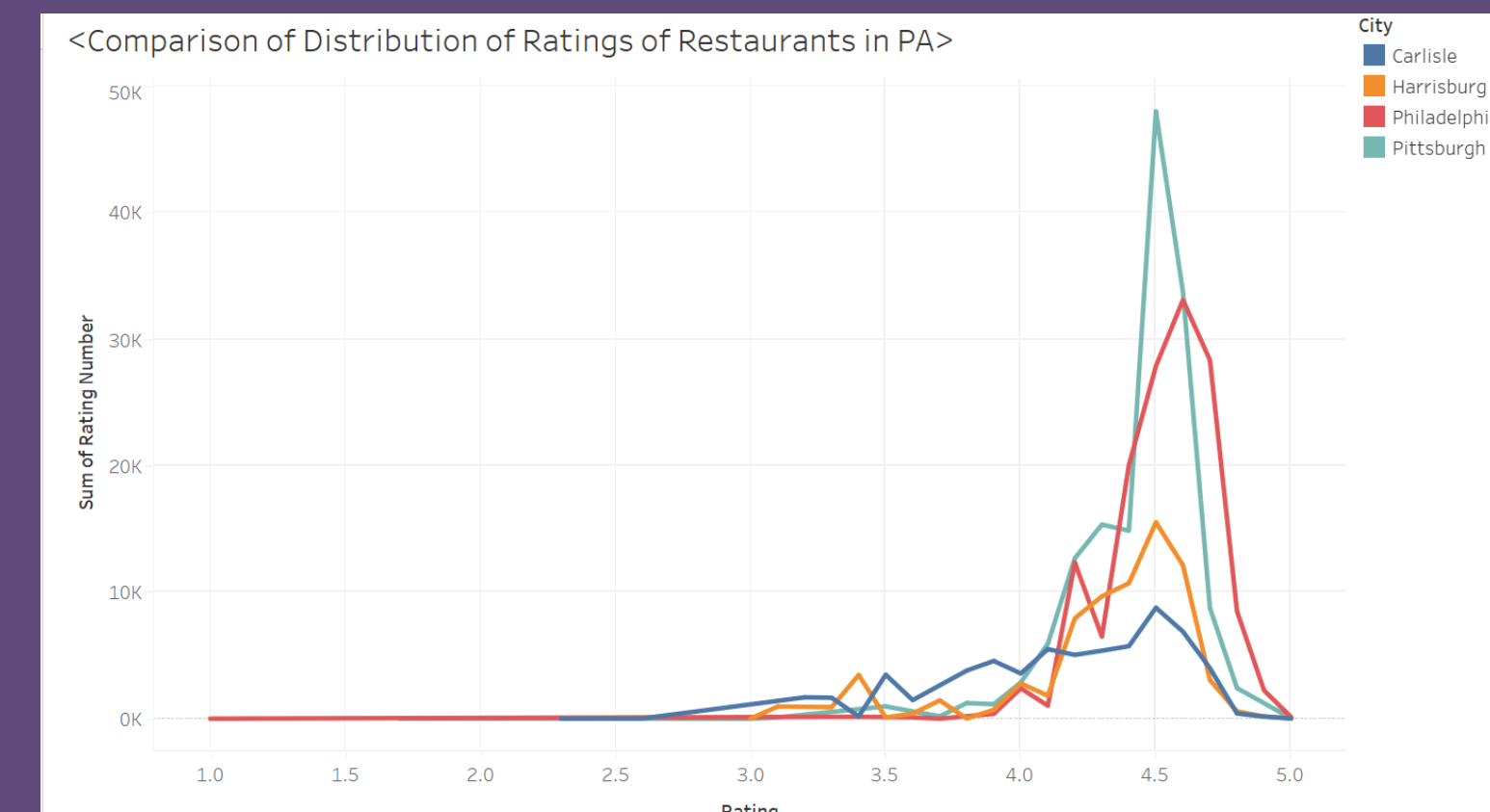
Review Data:

- Create a new variable `period` based on the `time` variable.
- The `period` variable has 3 values: `before` (before COVID-19), `covid` (during COVID-19), and `post` (after 75% of the capacity of restaurants is allowed in PA).
- Convert the data type of the `rating` variable from string to integer.

Analysis

Results

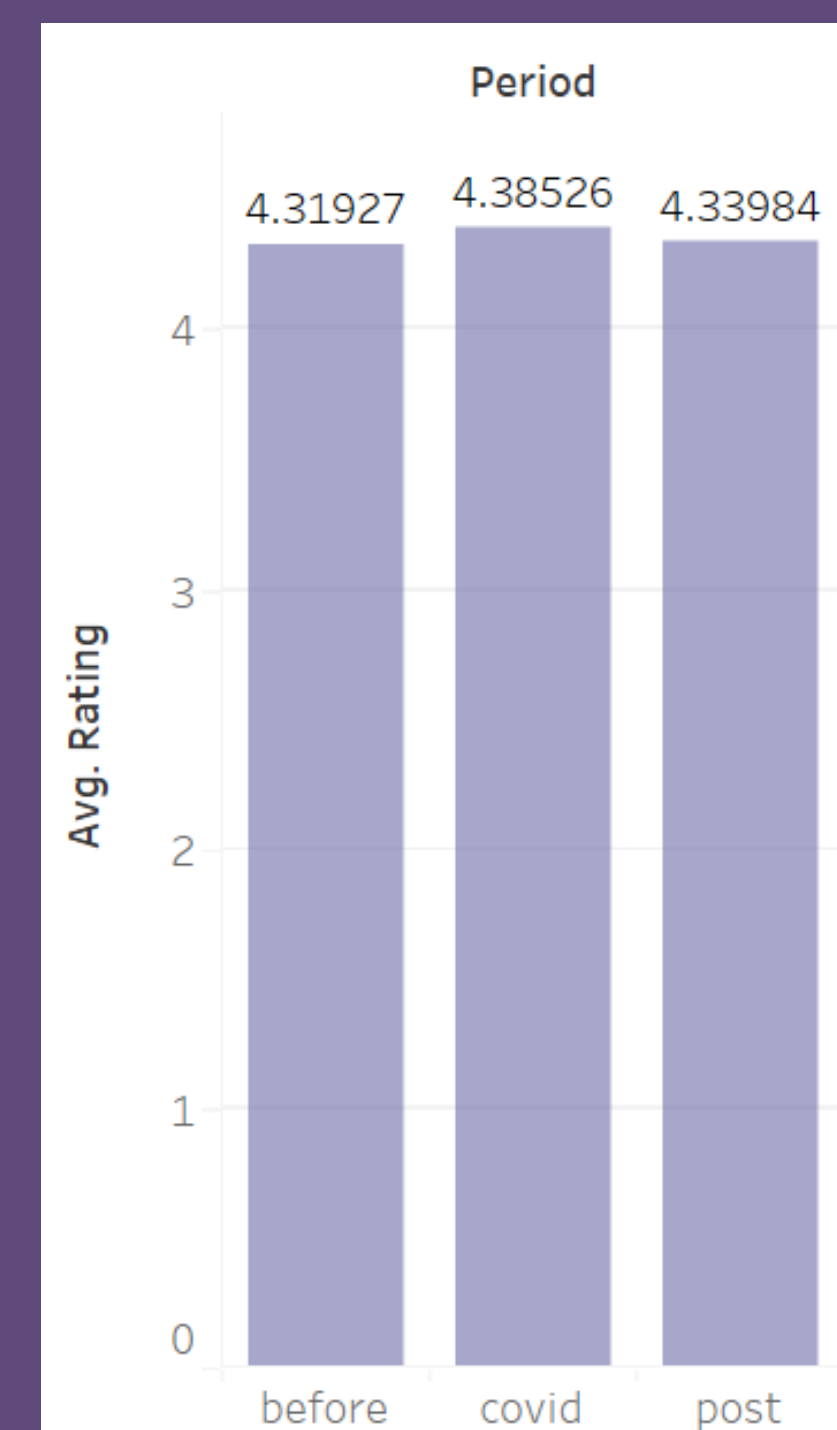
- ## 1. Comparison of Distribution of Ratings of Restaurants in 4 Cities in PA



- ## 2. Regression Table of rating ~ category + number_of_rating for All Restaurants

=====						
Dep. Variable:	rating	R-squared:	0.102			
Model:	OLS	Adj. R-squared:	0.091			
Method:	Least Squares	F-statistic:	9.074			
Date:	Thu, 13 Apr 2023	Prob (F-statistic):	7.46e-12			
Time:	15:00:29	Log-Likelihood:	-180.64			
No. Observations:	650	AIC:	779.3			
DF Residuals:	641	BIC:	819.6			
DF Model:	8					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.3643	0.026	168.891	0.000	4.314	4.415
category[T.American]	-0.0930	0.072	-1.300	0.194	-0.233	0.047
category[T.Chinese]	-0.3480	0.094	-3.788	0.000	-0.532	-0.164
category[T.Fast Food]	-0.5609	0.103	-5.462	0.000	-0.763	-0.359
category[T.Italian]	0.0173	0.069	0.252	0.801	-0.117	0.152
category[T.Mexican]	-0.0722	0.088	-0.817	0.414	-0.246	0.101
category[T.Pizza]	-0.0895	0.072	-1.248	0.213	-0.230	0.051
category[T.Sandwich]	-0.4522	0.094	-4.794	0.000	-0.637	-0.267
rating_number	5.543e-05	2.2e-05	2.591	0.012	1.23e-05	9.86e-05

- ### 3. Means of Average Rating for All Restaurants in 3 Different Periods



- #### 4. Word Clouds for Comments with Rating 1.0 and Rating 5.0



Conclusion

1. We can know that Philadelphia has the largest number of high-quality restaurants with rating higher than 4.5 among the 4 cities in PA. Most of the restaurants in Pittsburgh have a 4.5 rating, which is a good score. Unfortunately, according to the rating, the restaurants in Carlisle may have the lowest quality among the 4 cities in general.
2. The rating, number_of_ratings, and category are not correlated with each other for restaurants in PA in general since the r-squared is too small. The results do not change for each city.
3. COVID-19 has no significant impact on the rating of restaurants in PA since the average ratings of restaurants in PA before, during, and after COVID-19 are almost the same.
4. It seems like comments with a 1.0 rating intend to complain about the length of the waiting time because words like "order," "never," "back," and "minute" are obvious in the word cloud. The appearance of "good" is out of my expectation. As was expected, comments with a 5.0 rating have more positive words included, such as "great," "good," "friendly," "best," and so on. It is interesting that "carlisle" appears in the word cloud of highly rated comments.