



Impact of Food Delivery Services and Population Density on Yelp Restaurant Ratings

JUNJIE XU
(JACKY)○○○○



DATASET

- **YELP DATASET CHALLENGE (2017)**

The screenshot shows the Kaggle interface. On the left, there's a sidebar with icons for Home, Competitions, Datasets, Models, Code, and Discussions. The main area features a search bar at the top. Below it, a circular icon with a red asterisk contains the text "YELP, INC. AND 1 COLLABORATOR · UPDATED 6 YEARS AGO". The title "Yelp Dataset" is prominently displayed, followed by a subtitle "A trove of reviews, businesses, users, tips, and check-in data". At the bottom, there are links for "Data Card", "Code (238)", and "Discussion (25)".

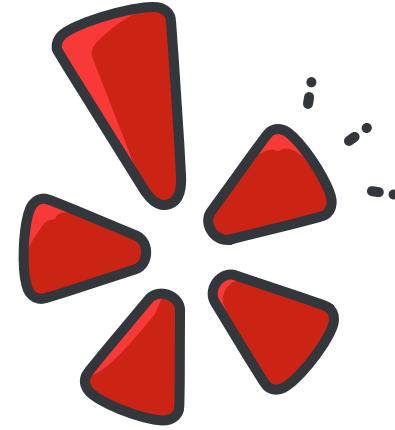
*Provide Business information
11 areas across 4 countries*

- **UNITED STATES CENSUS BUREAU**

The screenshot shows the United States Census Bureau website. At the top, it says "An official website of the United States government [Here's how you know](#)". The logo consists of a stylized American flag icon followed by the text "United States Census Bureau". The main heading is "Explore Census Data" with the subtitle "Learn about America's People, Places, and Economy". Below this is a search bar with placeholder text "Find Tables, Maps, and more ...". At the bottom, there are links for "Help", "Feedback", and "Advanced Search".

*Detailed Census data of US in Zip code level
Population, Shapefile, Income, Age, Sex...*

MY GOAL



Yelp Ratings



Population Density

Food Delivery



Background Knowledge

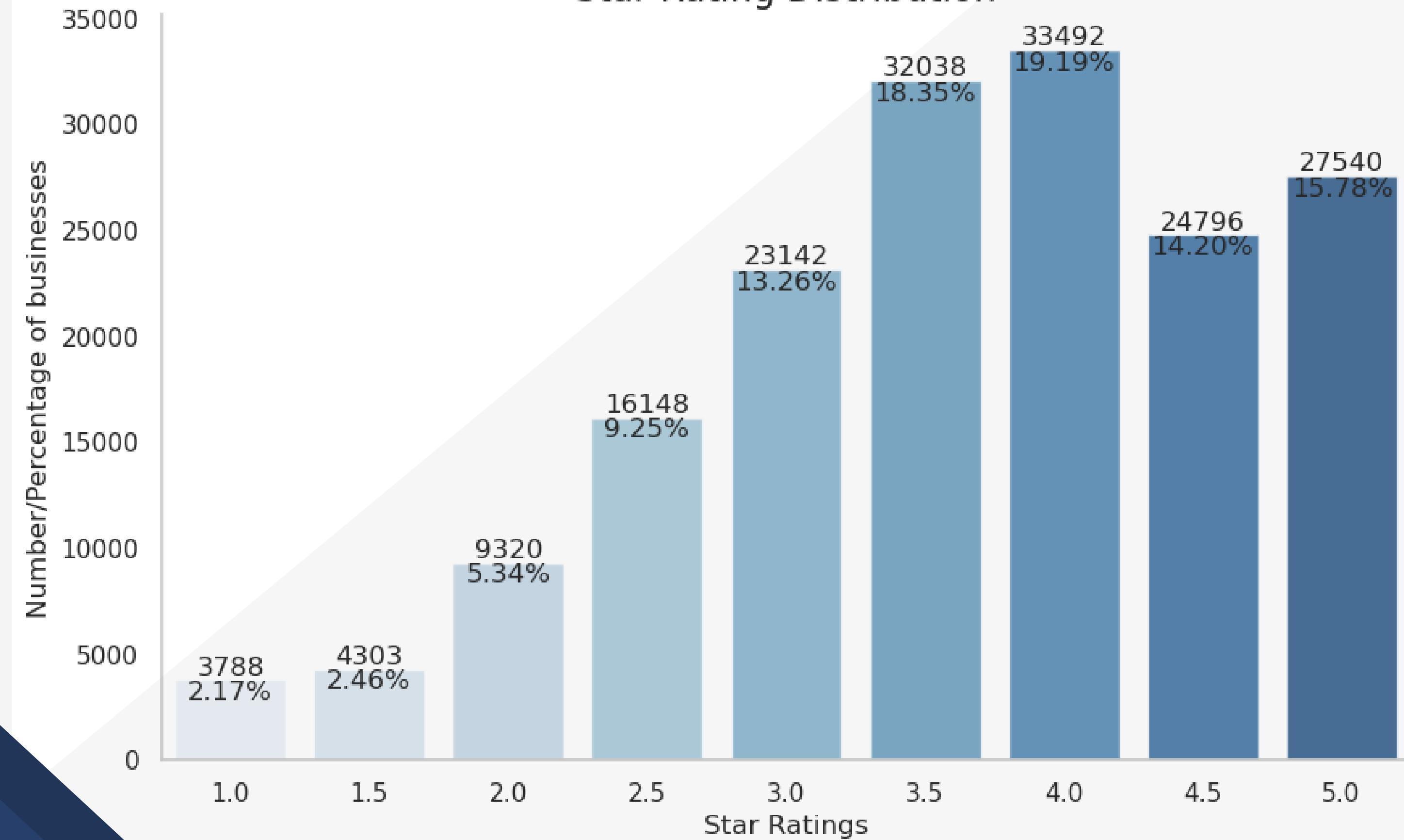
- Luca's 2011 findings emphasize the significant impact of Yelp star ratings on business outcomes and revenue generation.
- Higher population density can intensify competition, prompting better quality control and a narrower restaurant performance range, potentially influencing ratings (Mossay et al., 2020).
- Urbanization (High population density) in China drives food delivery and 20%+ population engages in O2O food delivery (Maimaiti et al., 2018). => USA.





DATA SUMMARY

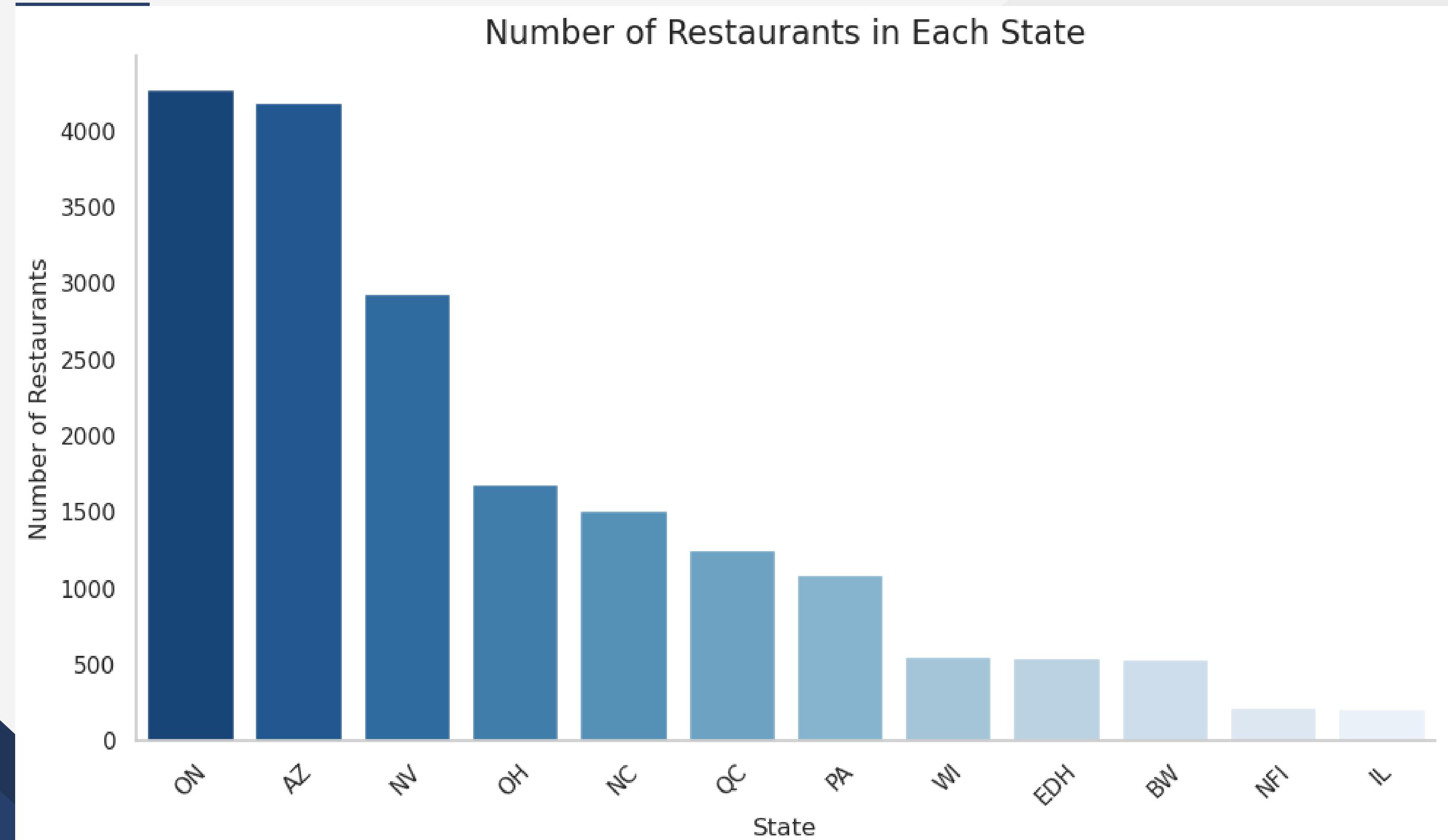
Star Rating Distribution





DATA SUMMARY

Number of Restaurants in Each State



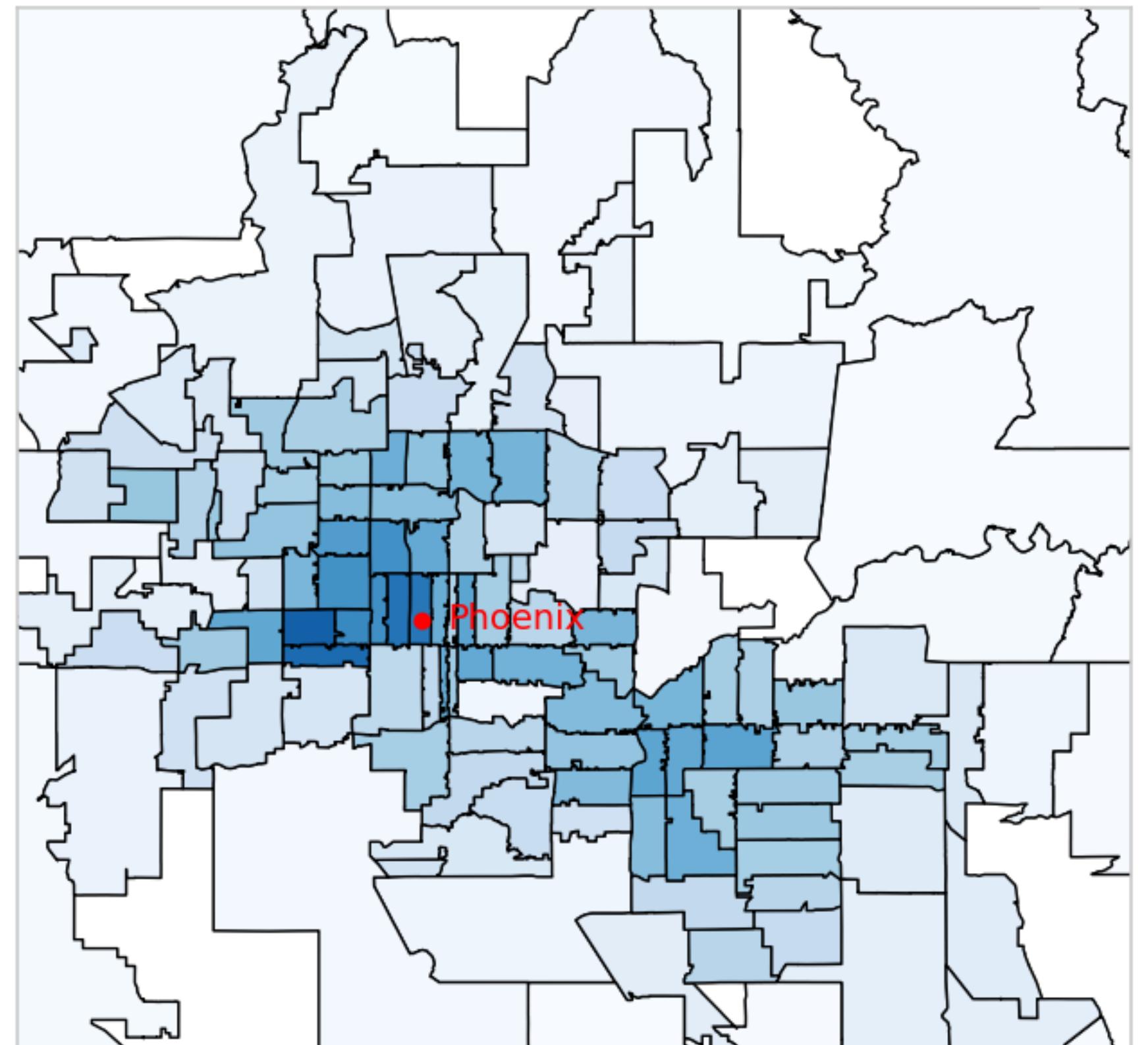


ARIZONA

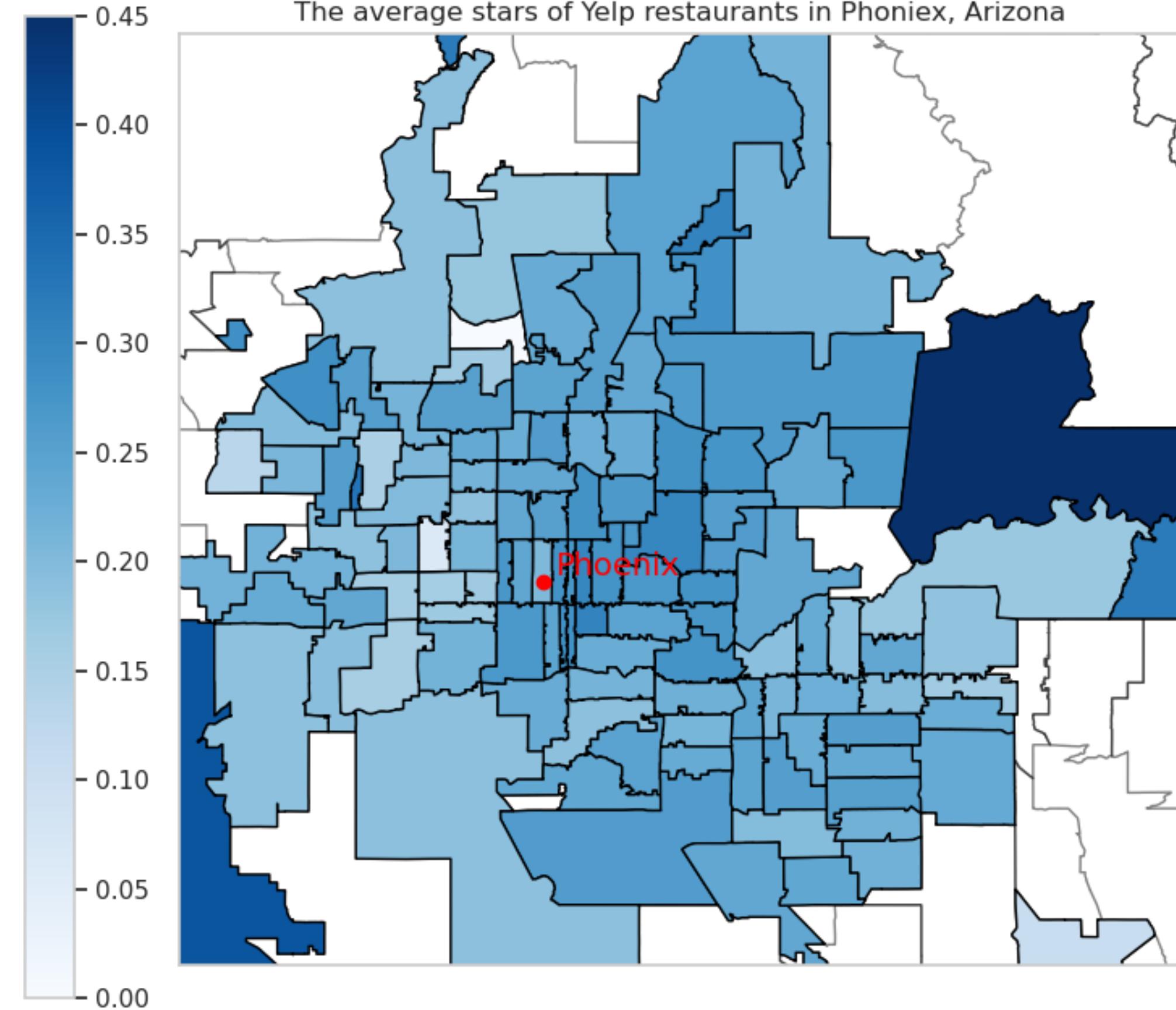
MAPS



The Population Density in Phoenix, Arizona



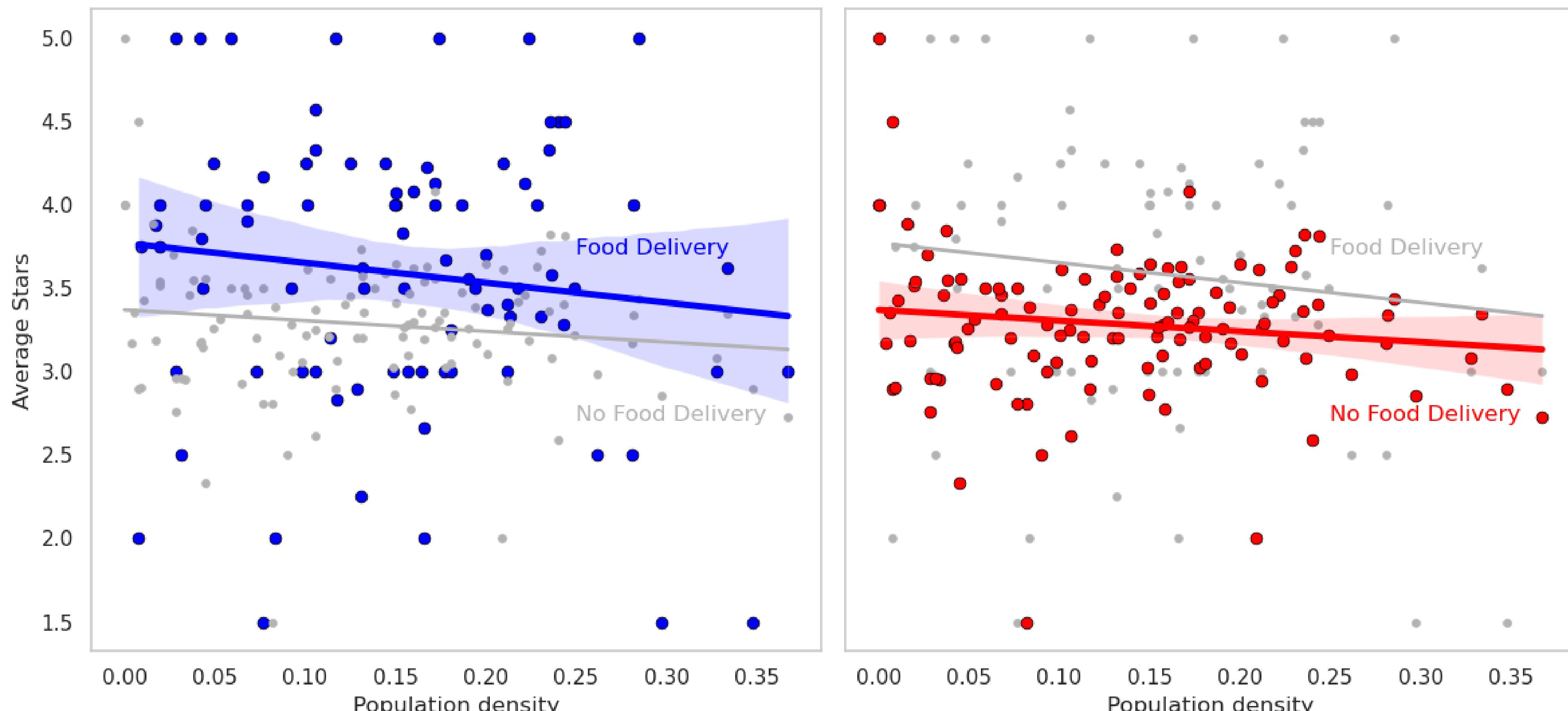
The average stars of Yelp restaurants in Phoenix, Arizona





DATA SUMMARY

Average Star Ratings and Population Density

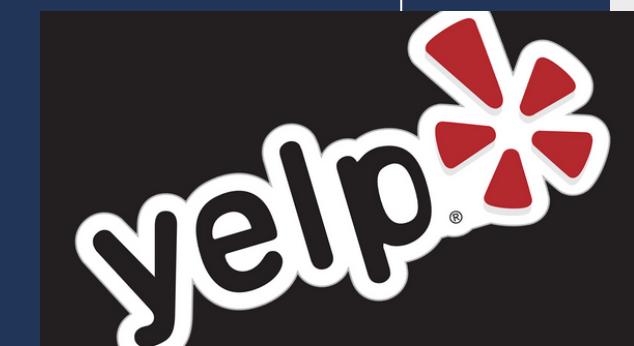


**Ordinary Least
Squares regression
(OLS)**

ARIZONA

	Model 1	Model 2	Model 3	Model 4	Model 5
Mean_Income		0.280 *** (0.047)	0.352 *** (0.052)	0.245 *** (0.054)	0.255 *** (0.055)
Pop_15_24			-0.002 (0.003)	-0.013 *** (0.003)	-0.013 *** (0.003)
Sex_ratio			0.016 *** (0.002)	0.015 *** (0.002)	0.015 *** (0.002)
const	3.346 *** (0.033)	-0.054 (0.553)	-2.236 *** (0.632)	-0.884 (0.663)	-1.016 (0.674)
count		0.003 *** (0.000)		0.003 *** (0.001)	0.003 *** (0.001)
has_food_delivery	0.341 *** (0.058)	0.342 *** (0.058)	0.347 *** (0.058)	0.342 *** (0.057)	0.331 *** (0.058)
percentage_delivery					0.004 (0.003)
population_density	-0.015 (0.192)	0.404 * (0.230)	0.370 (0.231)	0.145 (0.232)	0.166 (0.233)
Observations	4149	4149	4149	4149	4149
R ²	0.000	0.297	0.307	0.425	0.437
Adjusted R ²	-0.000	0.296	0.307	0.424	0.436
Residual Std. Error	0.273 (df=4146)	0.229 (df=4144)	0.227 (df=4143)	0.207 (df=4142)	0.205 (df=4141)
F Statistic	0.793 (df=2; 4146)	436.762 *** (df=4; 4144)	367.871 *** (df=5; 4143)	509.367 *** (df=6; 4142)	458.702 *** (df=7; 4141)

Linear Regression





Linear Regression

2. Stars ~ Population Density + Food Delivery + Income + Count:

$$\text{Stars} = \beta_0 + \beta_1 \text{population_density} + \beta_2 \text{has_food_delivery} + \beta_3 \text{Mean_Income} + \beta_4 \text{count} + \epsilon$$

$$\text{Stars} = -0.054 + 0.404 * \text{population_density} + 0.342 * \text{has_food_delivery} + 0.280 * \text{Mean_Income} + 0.003 * \text{count} + \epsilon$$

This suggests that restaurants situated in areas with high population densities, higher incomes, and more local restaurants, particularly if providing food delivery services, tend to achieve higher Yelp ratings.



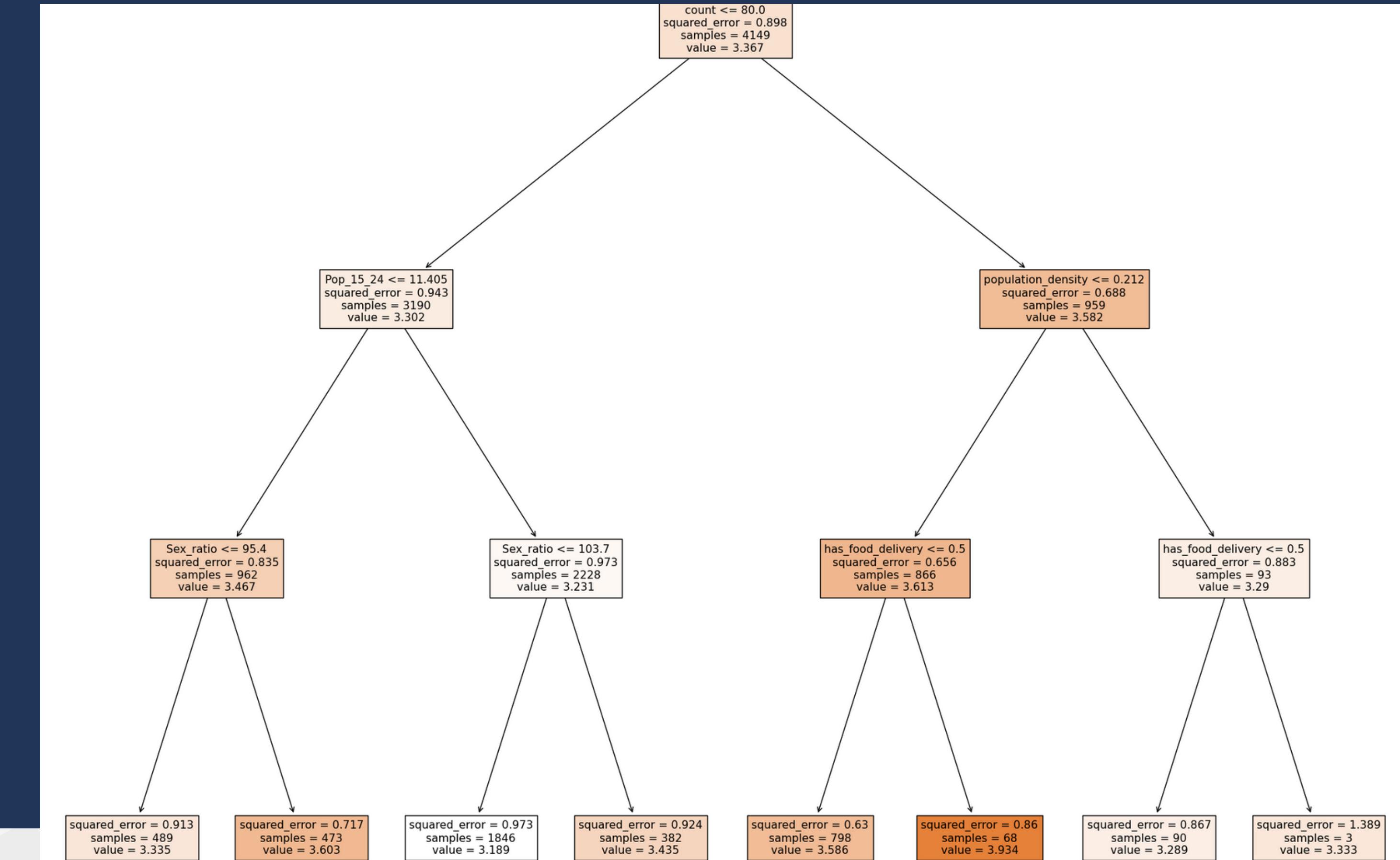
ARIZONA



**Machine Learning:
Regression Tree**

Regression Tree

o o o



RESULT

- This approach highlights population density's importance, linking it to higher ratings in less dense areas.
- Food delivery services boosted ratings, notably in low-population areas.
- The Linear Regression model (Linear relationship) produced an MSE of 0.0524, while the Regression Tree model (capture non-linear) achieved an MSE of 0.0057.





Discussion

In addition to the observed positive correlation, I intend to employ a Difference-in-Differences (DiD) Analysis to establish a causal link between food delivery services and Yelp restaurant ratings. Limited data across Years, Too many NAs.

While study insights guide owners and investors, culinary quality and customer focus remain pivotal for lasting success.





THANK YOU

*We look forward to hearing
feedback from you*

JUNJIE XU



○ ○ ○ ○ **Reference**

Maimaiti, M., Zhao, X., Jia, M., Ru, Y., & Zhu, S. (2018). How we eat determines what we become: opportunities and challenges brought by food delivery industry in a changing world in China. European Journal of Clinical Nutrition, 72(9), 1282–1286. <https://doi.org/10.1038/s41430-018-0191-1>

Matti, J. (2020). Reaching for the Stars: Spatial Competition and Consumer Reviews. Atlantic Economic Journal, 48(3), 339–353. <https://doi.org/10.1007/s11293-020-09679-x>

Mossay, P., Shin, J. K., & Smrkolj, G. (2020). Quality Differentiation and Spatial Clustering among Restaurants. SSRN Electronic Journal, 80(102799). <https://doi.org/10.2139/ssrn.3540202>

United States Census Bureau. (2020). Explore Census Data. Data.census.gov. <https://data.census.gov/table?g=040XX00US04>

Yelp Dataset Challenge. (2017, August 22). Yelp Dataset. Www.kaggle.com. [https://www.kaggle.com/datasets/yelp-dataset VERSIONS/6](https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset VERSIONS/6)

