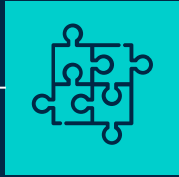


# Data Technology Final Project

By: Nora Trapp, Jing Xin,  
Junchen Xia, Xinran Luo, and  
Ying Zheng

# Agenda



01

BUSINESS QUESTION  
& HYPOTHESES



02

DATA  
ANALYSIS



03

FINAL  
RECOMMENDATION

# Two Datasets Used for Analysis

## 1. Yelp Dataset

- Same one we have analyzed in class

## 2. 2017 Census Dataset From Kaggle

- Contains data on many demographics including total population, gender, ethnicity, income, job type etc. by state and county

# Business Question

How do stars differ with income and population by state level?

01

# Hypotheses

1. States with a higher population are going to have a higher average star rating because they will have more individuals leaving good reviews to balance out the bad ones
2. Average stars are going to be lower for states with higher average income because those individuals have higher expectations for businesses
3. Stars are going to be of wider range with states that have a higher population due to population diversity

01

# Data Analysis

02

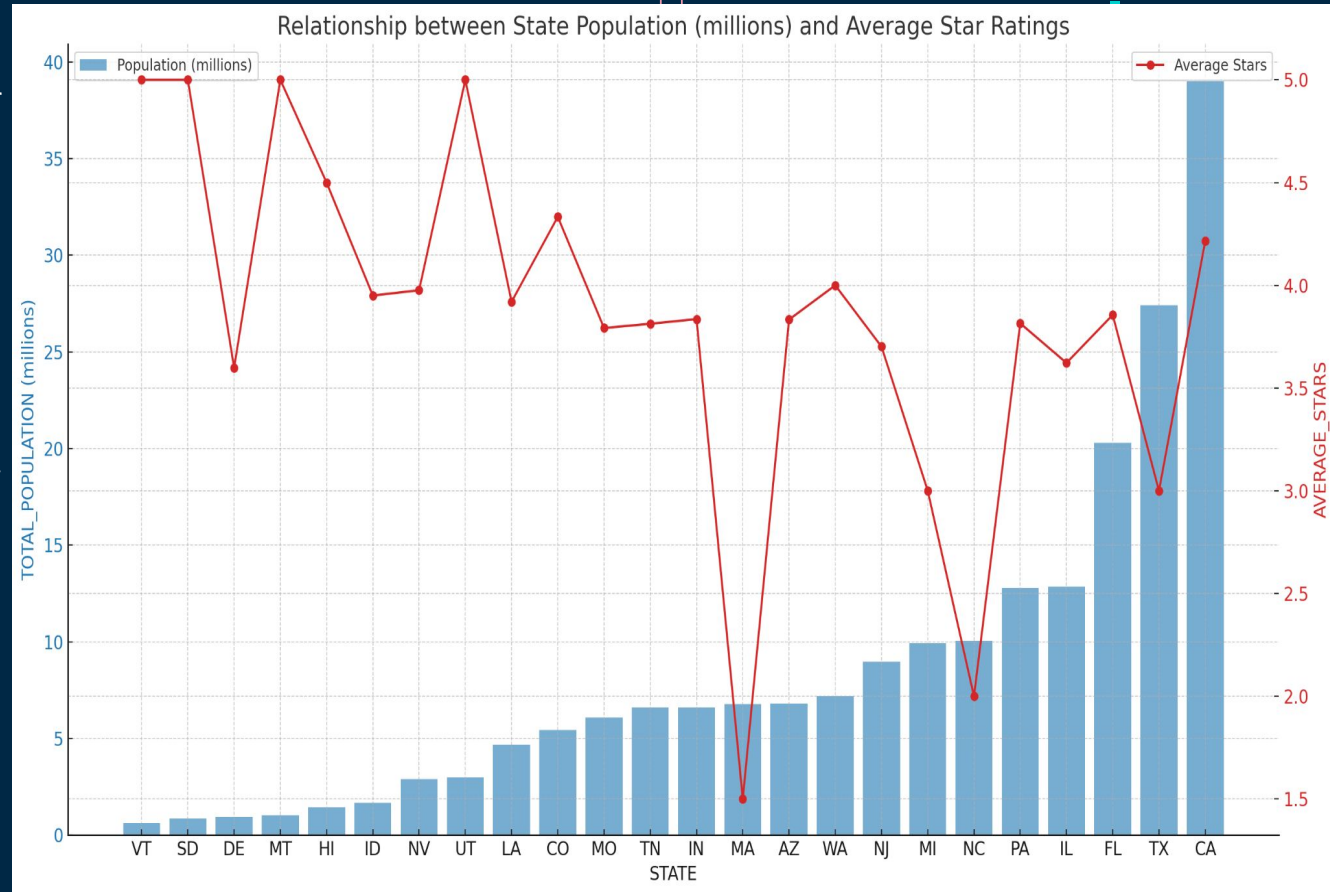
# Hypothesis #1 Analysis

States with a higher population are going to have a higher star average rating because they will have more individuals leaving good reviews to balance out the bad ones

- California's high population correlates with a high average star rating (4.21), which may support the hypothesis that larger populations lead to higher ratings due to a balance of reviews.

- Contrary to the hypothesis, less populous states like Montana and South Dakota achieve the highest average star ratings (5.0), possibly due to more personalized services in tighter-knit communities.

- Mid-sized populations in states like Arizona and New Jersey show variable star ratings, indicating that population size does not consistently predict the average star rating.





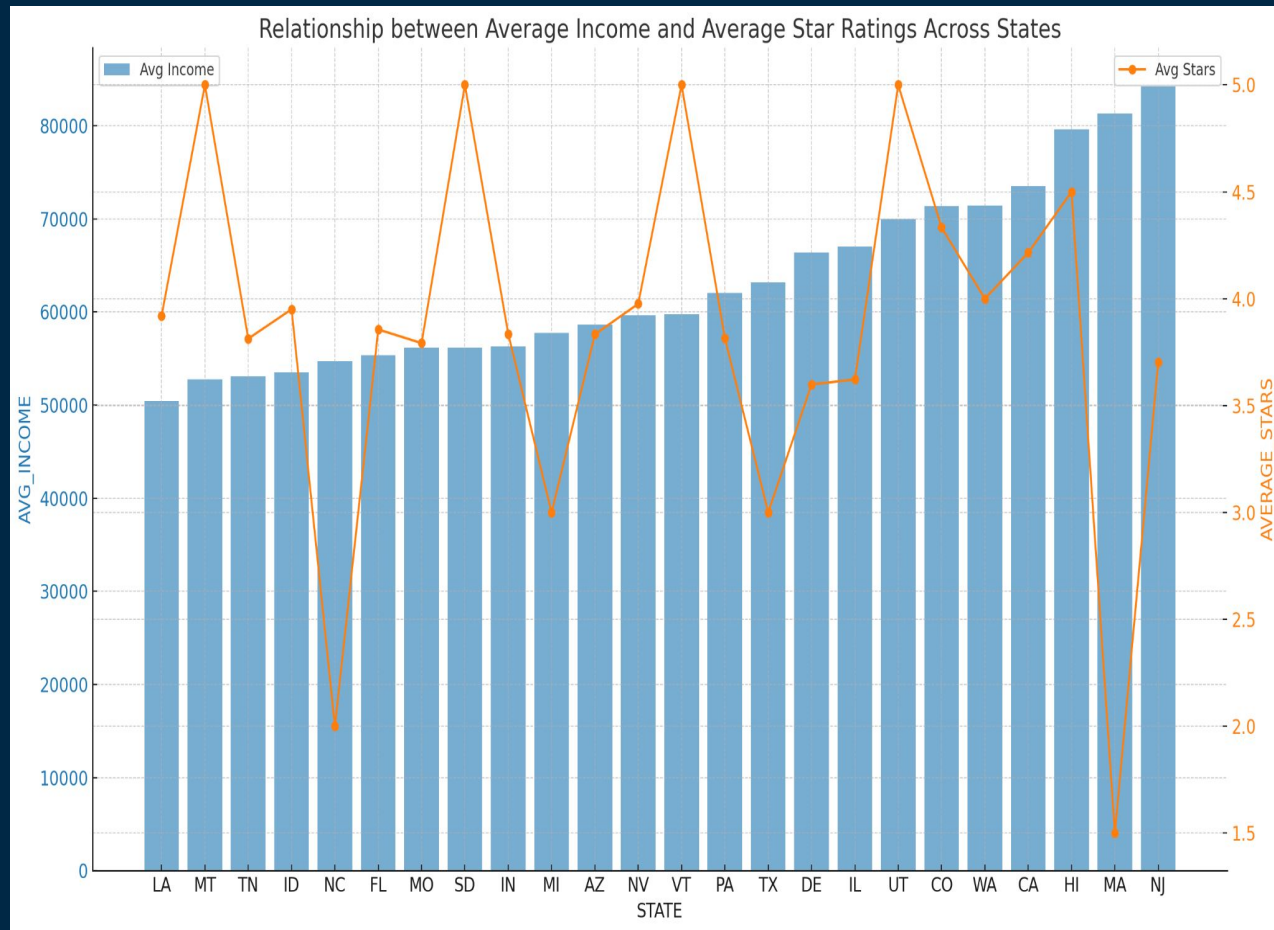
# Hypothesis #2 Analysis

Average stars are going to be lower for states with higher average income because those individuals have higher expectations for businesses

- Massachusetts has the second highest average income but the lowest average star rating (1.5). This aligns with the hypothesis that higher income leads to lower ratings due to increased expectations.

- New Jersey, the state with the highest average income, contradicts the hypothesis with a moderate star rating (3.70). This suggests that higher income does not uniformly lead to lower star ratings across all states.

- States like Montana, South Dakota, and Utah, despite having lower average incomes, exhibit the highest star ratings (5.0), indicating that factors other than income might significantly influence star ratings.



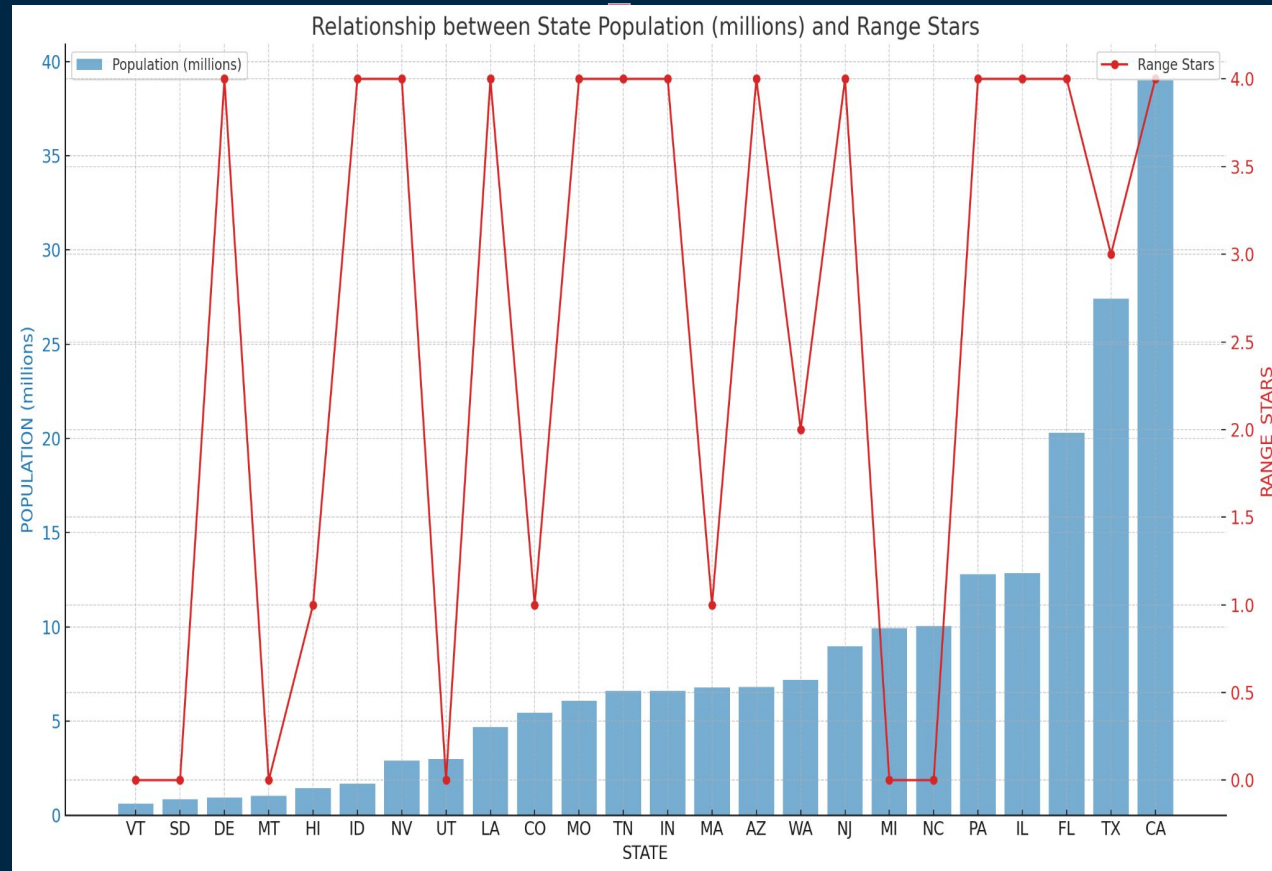
# Hypothesis #3 Analysis

Stars are going to be of wider  
range with states that have a  
higher population due to  
population diversity

- States like California and Texas with higher populations show a wider range of star ratings (4 and 3 respectively), supporting the hypothesis that population diversity leads to varied ratings.

- Less populous states like Montana and South Dakota have a narrow range of star ratings (0 for both), aligning with the idea that lower diversity in smaller populations leads to more consistent ratings.

- Some states with moderate populations, like Washington, exhibit a range of ratings (2), indicating that factors other than just population size and diversity might influence the range of star ratings.



# Final Recommendation

03

# Conclusions

1. Lower/ higher population alone don't seem to have a correlation with average star rating
2. Level of income alone doesn't seem to correlate with average star rating
3. Total population alone doesn't seem to correlate with the range of star ratings

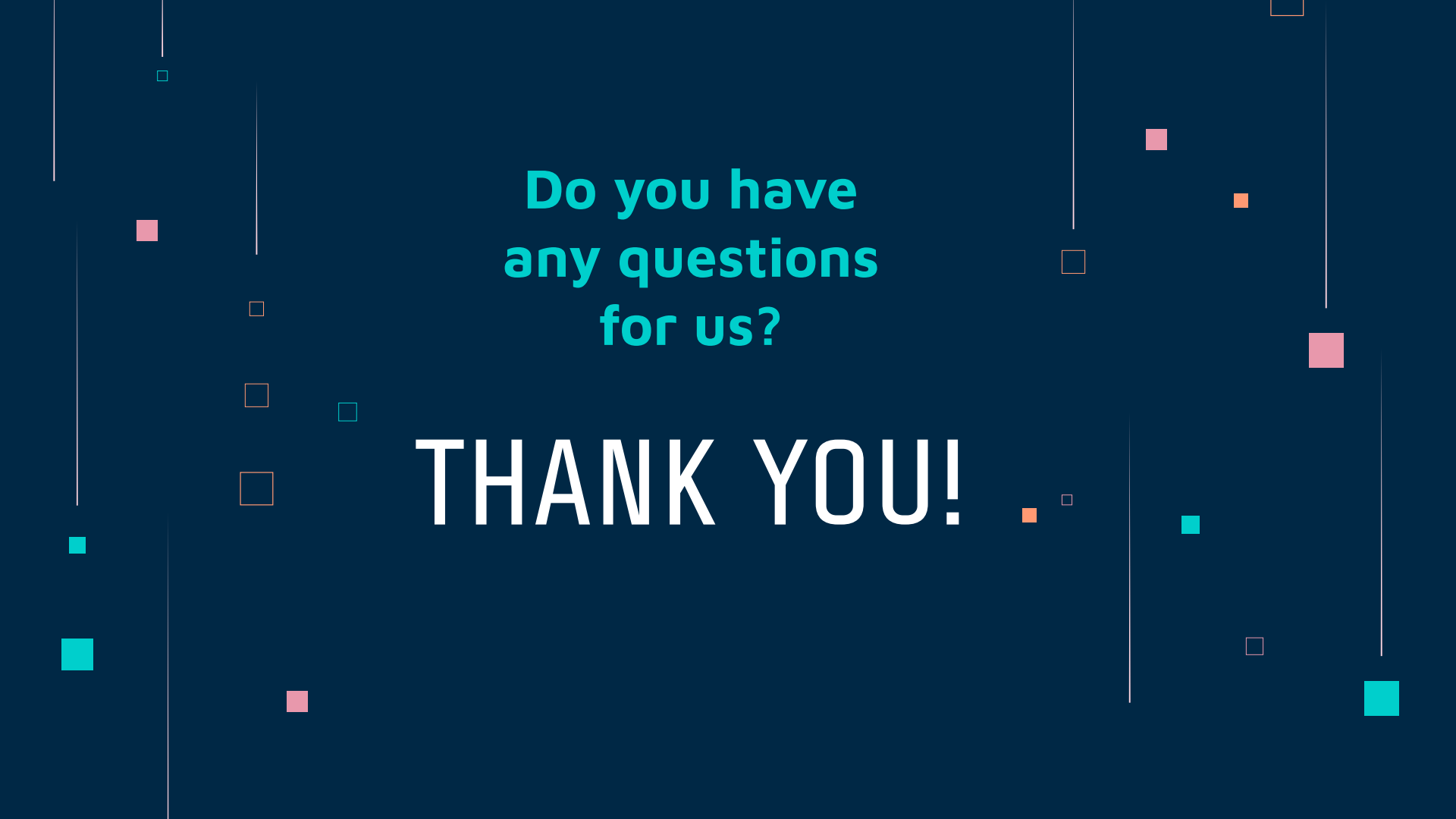
# Key Challenges

1. Finding a secondary dataset
  - Needed a commonality to join on
  - One that we could easily write queries on
2. Downloading secondary dataset into Snowflake

# Final Recommendations Going Forward ...

- Value in pursuing this question further by exploring additional demographic areas to uncover more insights
  - How average star ratings differ by gender, ethnicities, job status, or by different age groups
- Also, value in pursuing the effect of different combinations of demographics on average stars as one demographic is likely not the sole cause of changes in average stars





**Do you have  
any questions  
for us?**

**THANK YOU!**