



GLASSDOOR COMPANY REVIEW ANALYSIS

ZEHRA KEZER

INTRODUCTION

Glassdoor is an online platform where old or new employees can comment on companies and is also used for job search. Our project talks about analyzing the reviews and feedbacks from the customer.

Why Reviews Matter

Analyzing company reviews to improve reputation

For job seeker

- Looking up reviews of a company can provide valuable insight in deciding whether to continue with a job application.

For company

- It helps employers gain real insights on their employee engagement and create a structure for future similar tasks

How

- Using NLP techniques to uncover patterns
- Enhancing workplace culture through analysis
- Scraping Glassdoor reviews for data analysis

Process

- Gaining insights from customer/employee feedback
- Understanding sentiment and topics in reviews

Tools

The programming is done in Python. For web scraping, we used Beautiful Soup. The Python libraries used are NLTK, TextBlob and Gensim. For data analysis and visualization, we used Numpy, Pandas, matplotlib and seaborn

DATA EXPLORATION

5.0 ★★★★

Former Employee, more than 1 year

Easy driving job

Mar 15, 2023 - Driver in Newark, NJ

Recommend CEO Approval Business Outlook

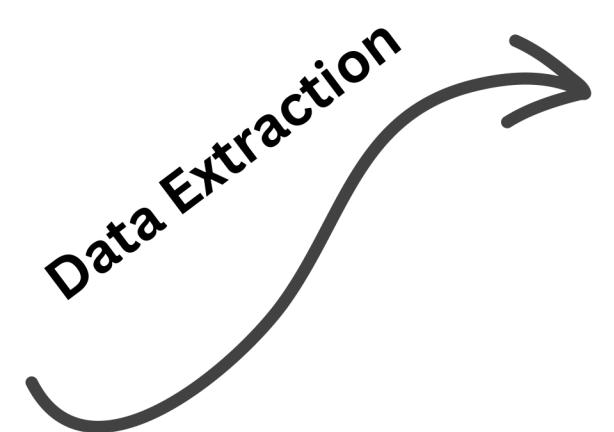
Pros

You can easily start making 1,000-2,000 per week if you work full time. Usually 2000 would take about 70 hours, so gotta grind to make that much. Even 40 hours should yield you at least 1,000. There are also "Quests" that give you up to 400 a week in bonuses.

Overall good pay and chill job if you like to drive.

Cons

Wear and tear on vehicle. You can put a lot of miles on your vehicle driving for Uber full time. Very occasionally you will have a rude customer, but I only had that happen 2 times after 3500 rides. It is something that will occasionally happen but certainly a outlier.



```
Summary  
Date  
City  
State  
OverallRating  
Pros      nice company great people and benefits  
Cons      non cons great culture and  
Name: 1, dtype: object
```

Data Cleaning

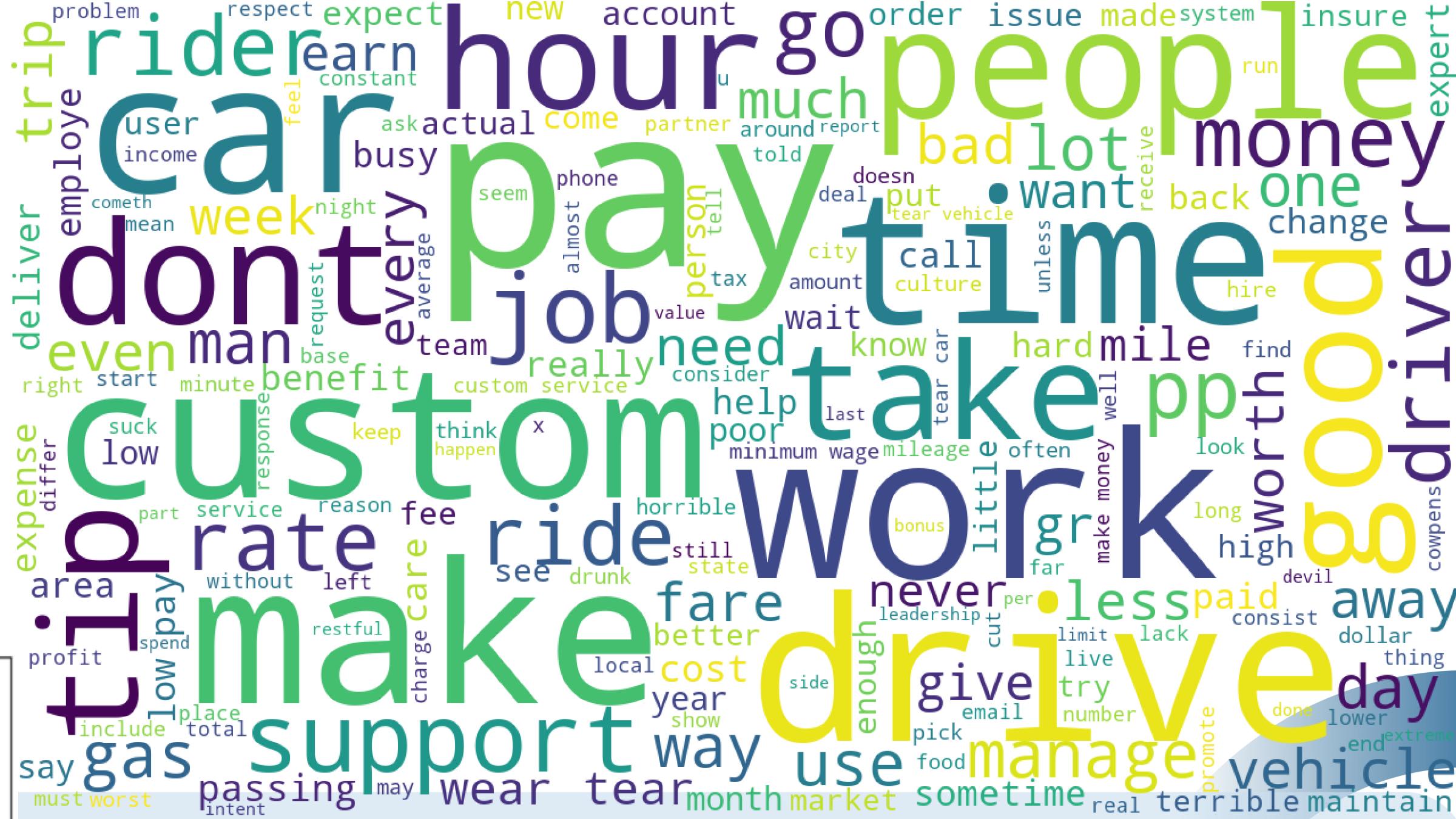
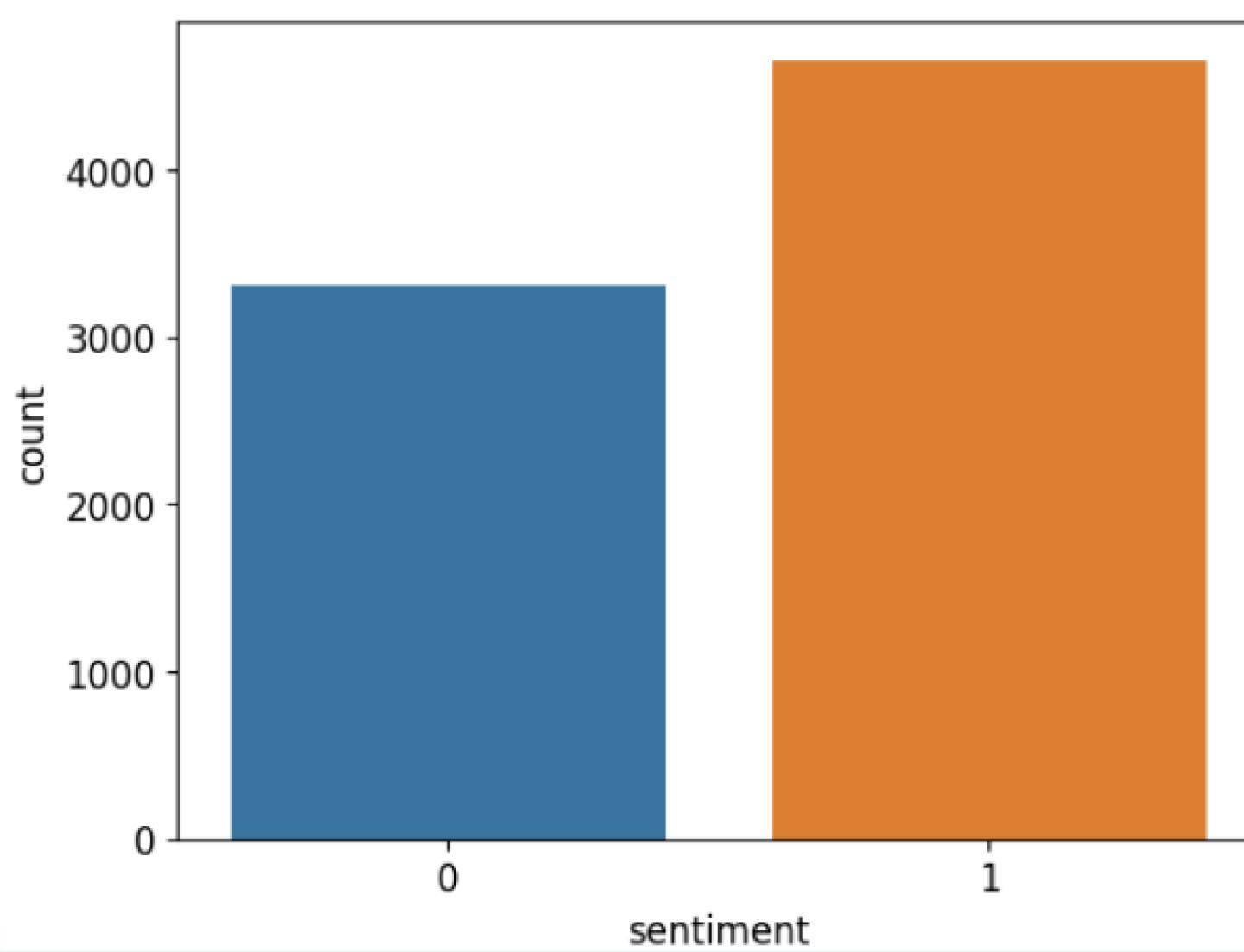
Feature Engineering

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 7933 entries, 0 to 7963  
Data columns (total 3 columns):  
 #   Column       Non-Null Count Dtype  
---  
 0   sentiment    7933 non-null   int64  
 1   reviews      7933 non-null   object  
 2   state        7933 non-null   object  
dtypes: int64(1), object(2)  
memory usage: 247.9+ KB
```

SENTIMENT ANALYSIS

How positive or negative the comments are?

1	0.584129
0	0.415871



If the rating is 1 or 2 implies that the sentiment is 0.

If the rating is 4 or 5 implies that the sentiment is 1.

PROS

- work 1857
- gr 1325
- good 945
- company 642
- flexible 640
- people 622
- hour 501
- make 489
- time 438
- pay 435
- schedule 422
- job 403
- want 374
- get 344
- money 335
- benefit 319
- opportune 299
- lot 289
- drive 279
- place 279
- place work 154
- work want 149
- gr company 131
- gr place 122
- flexible hour 119
- flexible schedule 113
- good pay 95
- flexible work 79
- company work 77
- work environs 75
- fast pace 71
- work hour 70
- part time 68
- make hour 67
- make money 66
- work time 65
- gr benefit 64
- make schedule 63
- gr job 63
- good money 58

CONS

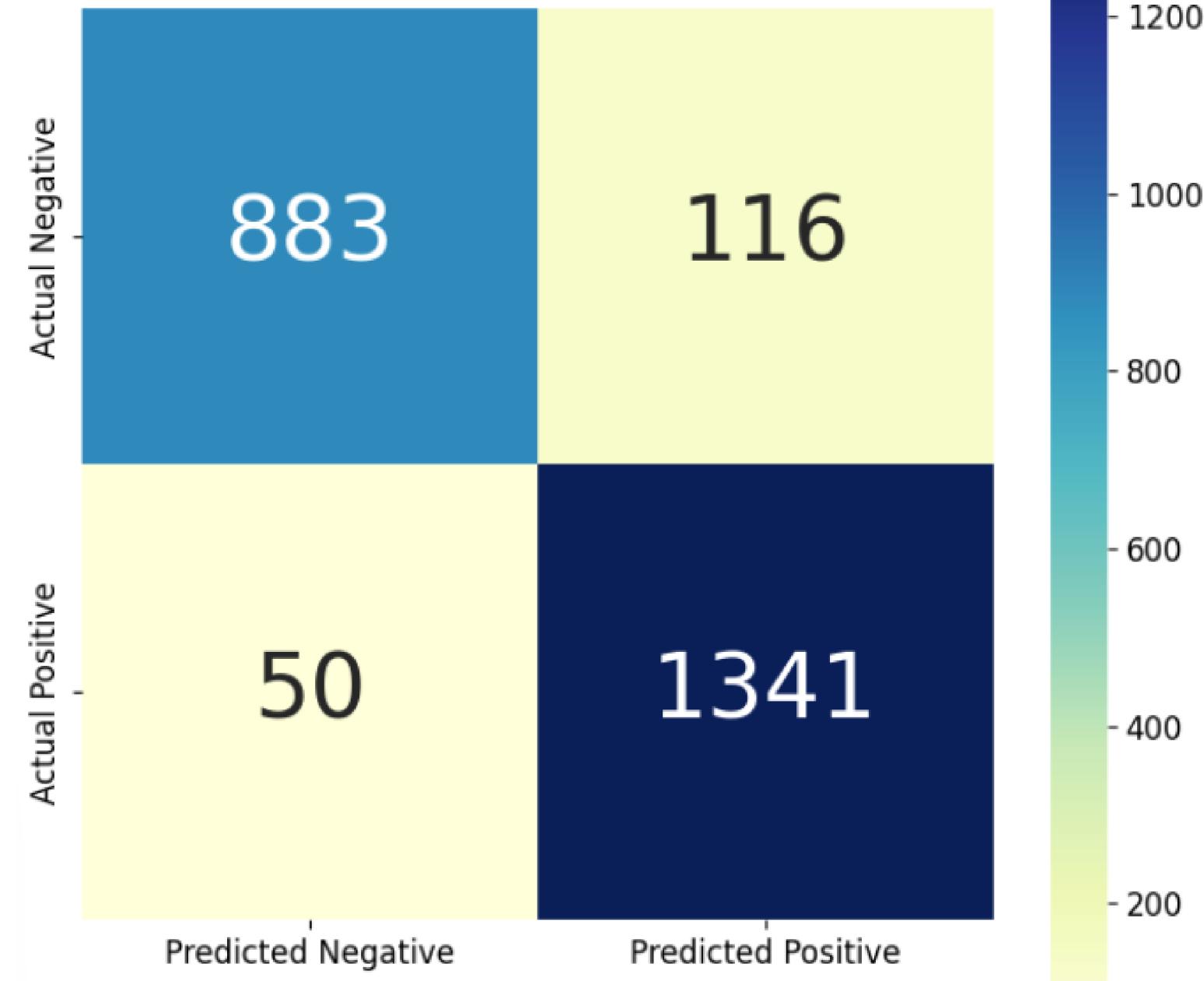
- pay 866
- get 706
- work 640
- time 636
- car 615
- make 613
- drive 560
- company 422
- custom 414
- money 409
- hour 407
- dont 396
- people 349
- good 339
- take 323
- gas 311
- manage 289
- tip 286
- rider 280
- rate 276
- wear tear 154
- low pay 111
- minimum wage 87
- custom service 82
- make money 78
- get paid 64
- tear car 63
- tear vehicle 59
- per mile 45
- long hour 43
- full time 40
- work hour 37
- per hour 37
- wast time 35
- dont get 34
- less minimum 34
- part time 34
- pay gas 32
- feel like 31
- car maintain 28

BEST MODELLING

	Accuracy	Precision	Recall	F1 Score
LR(Word)-CV	0.926778	0.930063	0.891892	0.910577
LR(N-Grams)-CV	0.889540	0.930832	0.794795	0.857451
MNB(Word)-CV	0.906276	0.925357	0.843844	0.882723
MNB(N-Grams)-CV	0.894979	0.951691	0.788789	0.862616
BNB(Word)-CV	0.820502	0.930514	0.616617	0.741722
BNB(N-Grams)-CV	0.674477	1.000000	0.221221	0.362295
LR(Word)-TFIDF	0.928452	0.924180	0.902903	0.913418
LR(N-Grams)-TFIDF	0.930544	0.946409	0.883884	0.914079
MNB(Word)-TFIDF	0.914226	0.957373	0.831832	0.890198
MNB(N-Grams)-TFIDF	0.914226	0.957373	0.831832	0.890198
BNB(Word)-TFIDF	0.737238	0.967254	0.384384	0.550143
BNB(N-Grams)-TFIDF	0.737238	0.967254	0.384384	0.550143
LR(Word)-TFIDF	0.928452	0.924180	0.902903	0.913418
LR(N-Grams)-TFIDF	0.930544	0.946409	0.883884	0.914079
MNB(Word)-TFIDF	0.914226	0.957373	0.831832	0.890198
MNB(N-Grams)-TFIDF	0.914226	0.957373	0.831832	0.890198
BNB(Word)-TFIDF	0.737238	0.967254	0.384384	0.550143
BNB(N-Grams)-TFIDF	0.737238	0.967254	0.384384	0.550143

- **LR(n-grams)-TFIDF**

Splitting the data into Training data(%70) and Testing data(%30).
Apply the Logistic Regression model and TFIDF(n-grams) fit it with the training data.



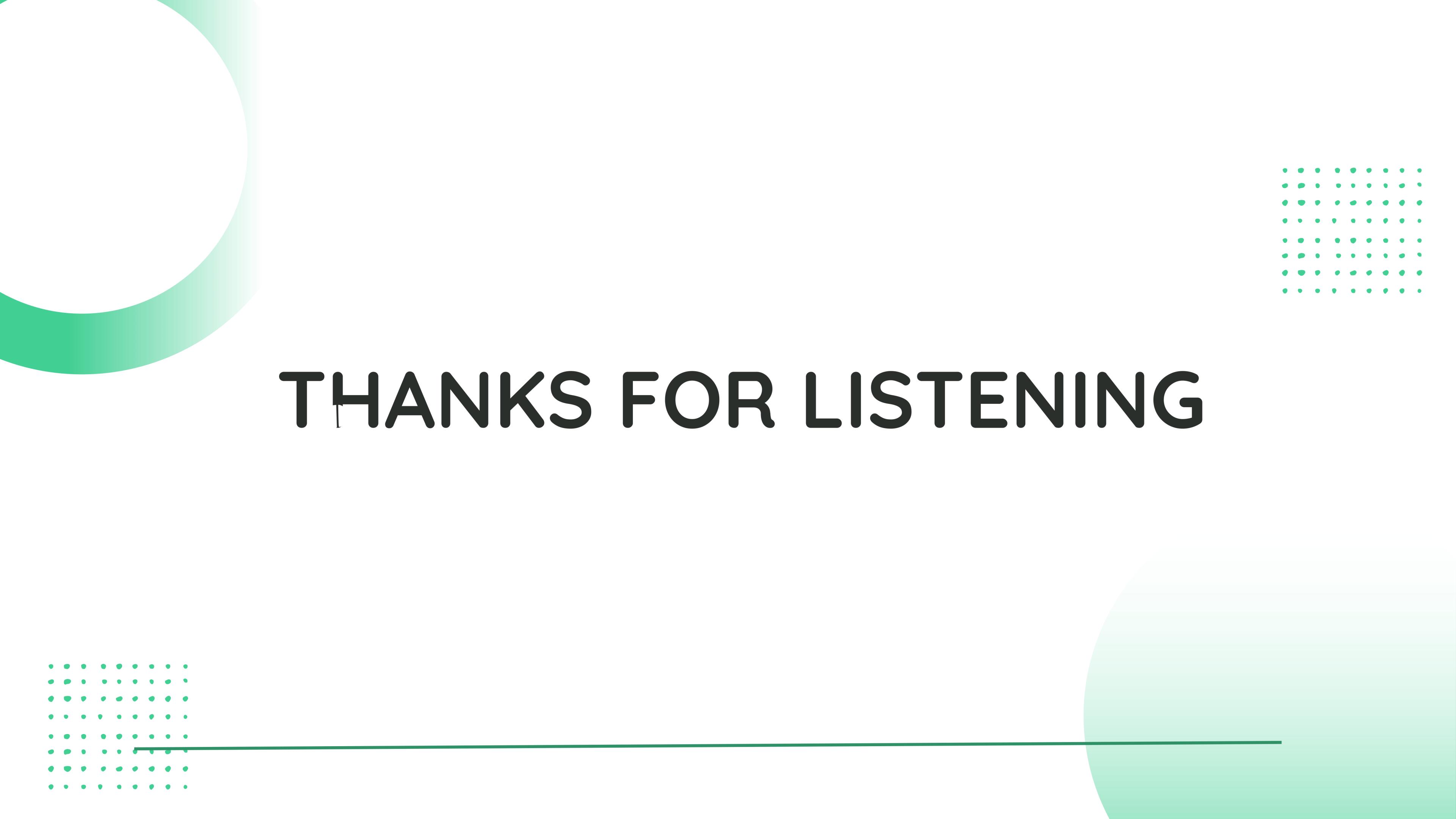
CONCLUSION

The goal of this analysis is to analyze anonymous employee reviews submitted on Glassdoor, with the hope to help employers gain real insights on their employee engagement.

Fitting with Logistic Regression model and TFIDF(n-grams)

The majority of employees are very satisfied with the company. Employees love the flexible hour, work place and work-life balance. Employees dislike a lot of wear and tear, low pay and tear of vehicle.

Given any company's Glassdoor review data, the same structure and code can be applied with little tweaks, especially the data cleaning code blocks can be used without change since all Glassdoor review data follow the same format.



A large white circle overlaps a smaller green circle on the left side of the slide. On the right side, there is a vertical column of green dots and a horizontal row of green dots at the bottom, intersecting near the center. A thin green horizontal line runs across the bottom of the slide.

THANKS FOR LISTENING