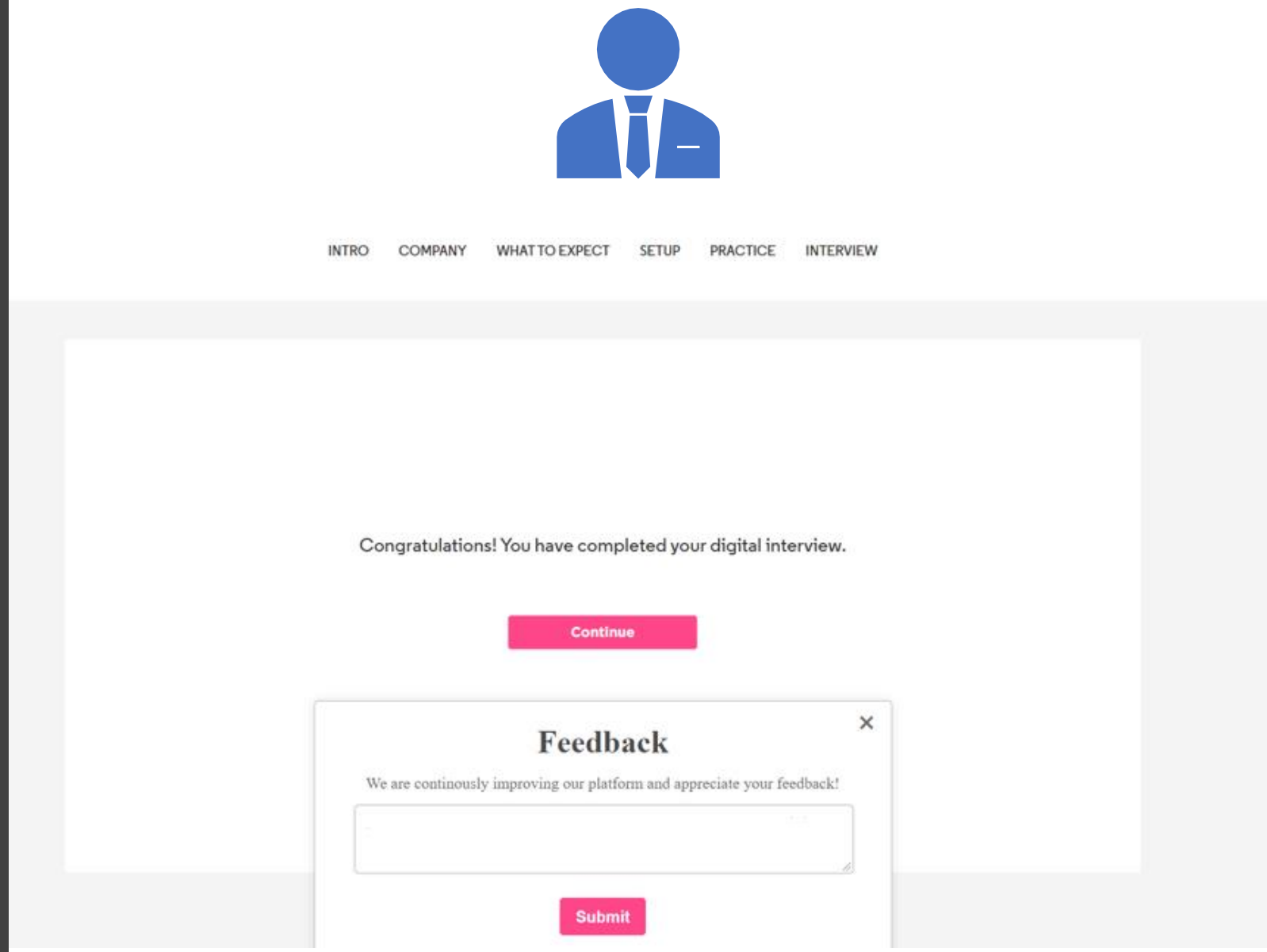
Sentiment analysis for Interview reviews

John Lim
DSI 17

Problem Statement

- A company has approached my team to review their interview process.
- The client wants to give their candidates the best interview experience regardless of whether they were accepted for the job.
- By doing so, they hope to achieve more positive interview reviews, strengthen the company branding and ultimately attract more talent.



Content

METHODOLOGY

DATA GATHERING

EDA AND PRE-PROCESSING

MODELLING

FINDINGS AND LIMITATIONS

Methodology

- Analysing companies with thousands of interview reviews on glassdoor
- Build a model that is optimized and tuned based on accuracy score.
- The model will be able to tell if an interview feedback is a positive or negative sentiment
- Companies can then use the insights to train their HR and hiring managers to adopt or avoid certain attributes during an interview process

Data Gathering...

Web Scraping glassdoor interview reviews with python, selenium and headless chrome

Software Engineer - New Grad Interview

Anonymous Interview Candidate in Seattle, WA (US)

No Offer | Negative Experience | Easy Interview

Application

I applied online. I interviewed at Microsoft (Seattle, WA (US)) in October 2020.

Interview

I was emailed by a recruiter around 2 weeks after applying for a first interview. That went pretty well, only one LC easy and background questions. I was called back for a second round which was supposed to be a day with 4 interviews. They canceled the last one and told me only after I emailed them during the supposed time. The three other interviews were pretty bad, interviewers didn't have their cameras on and didn't seem interested in what I have to say at all.
I also haven't heard back from them yet.

Interview Questions

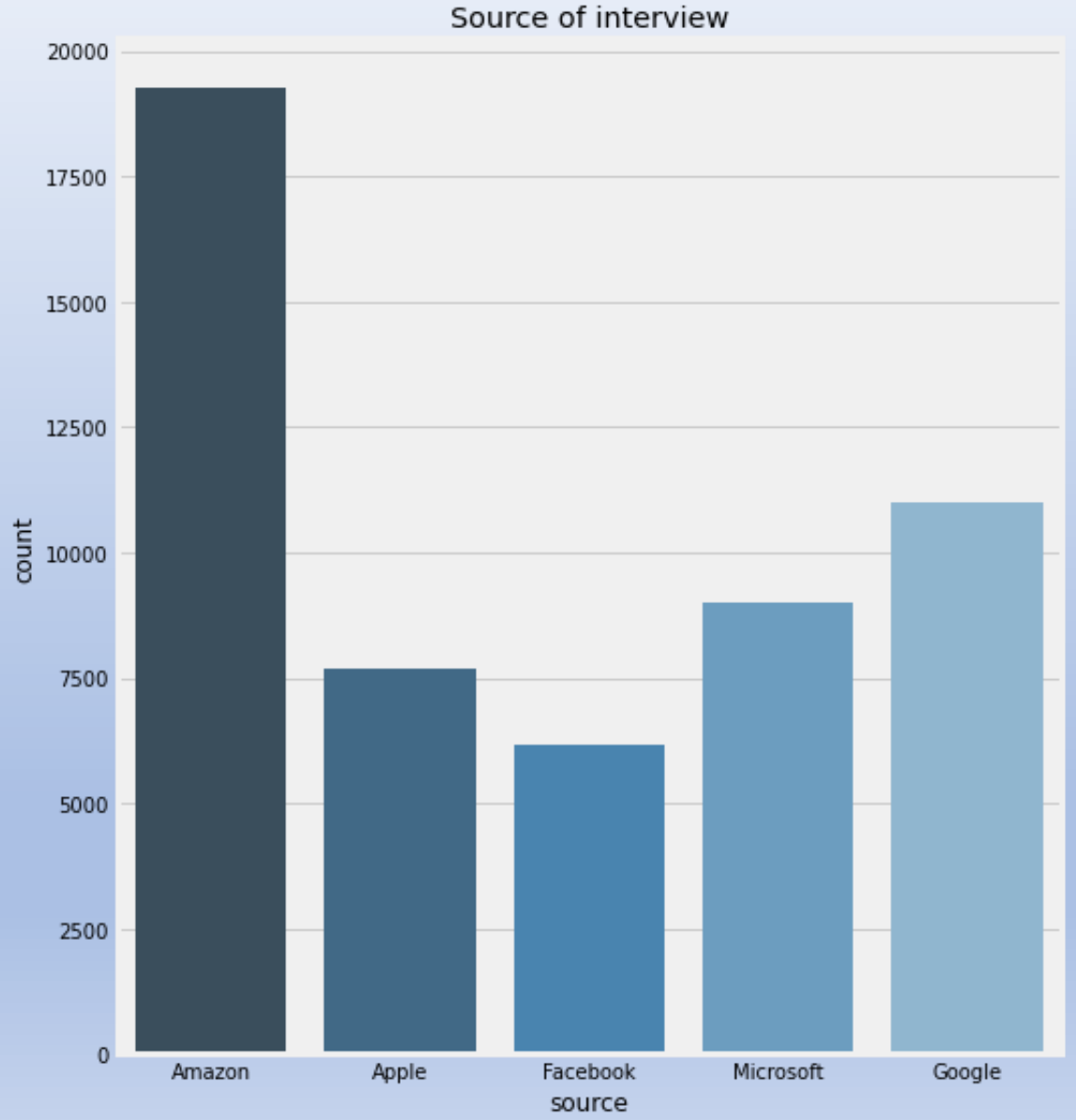
Top 5 numbers in an array without sorting, Remove duplicates from array in place, Binary search algo, Valid parans
What are distributed systems, microservices, APIs
What do you think about Agile development

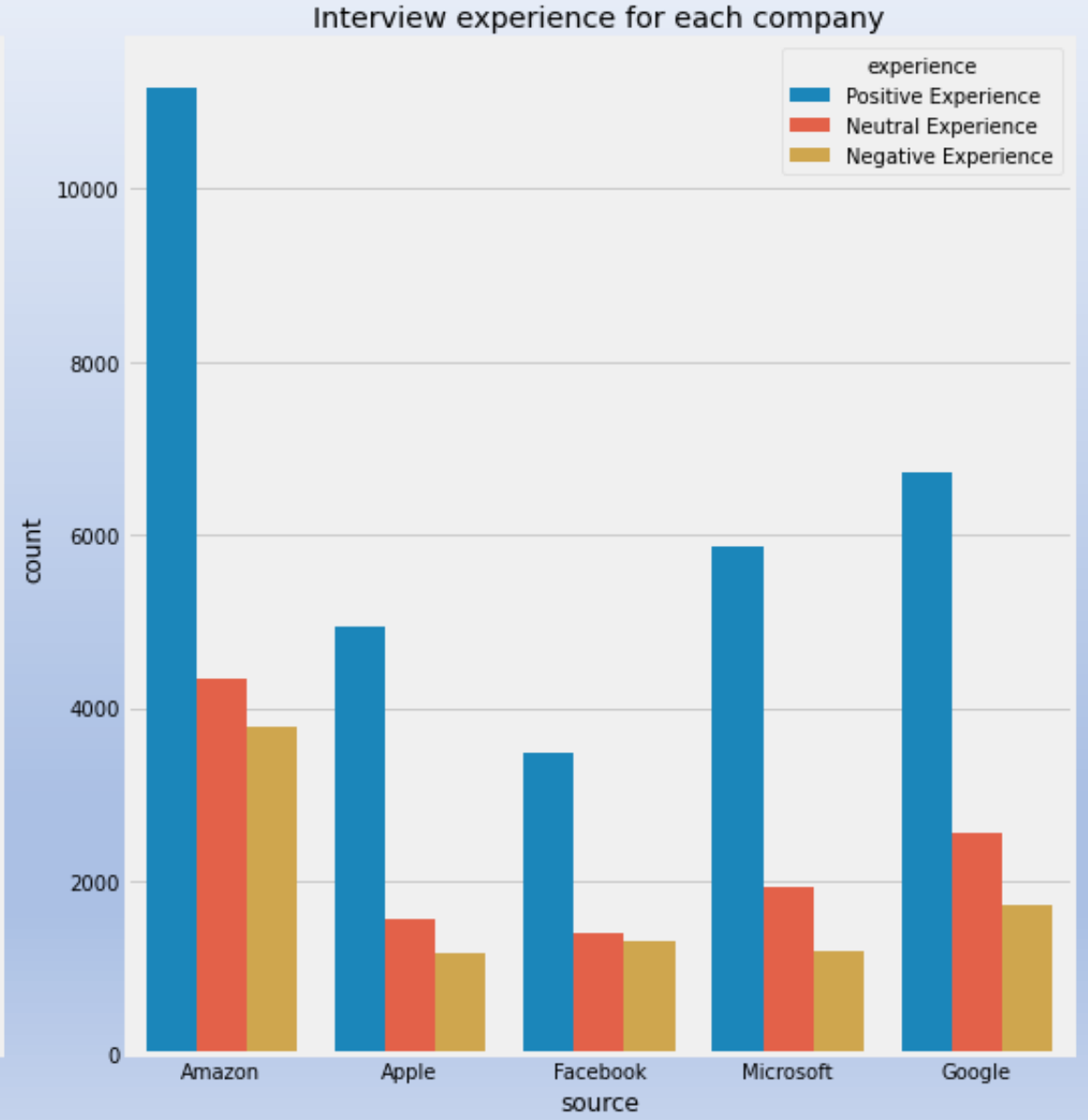
↳ Answer Question

Data Gathering

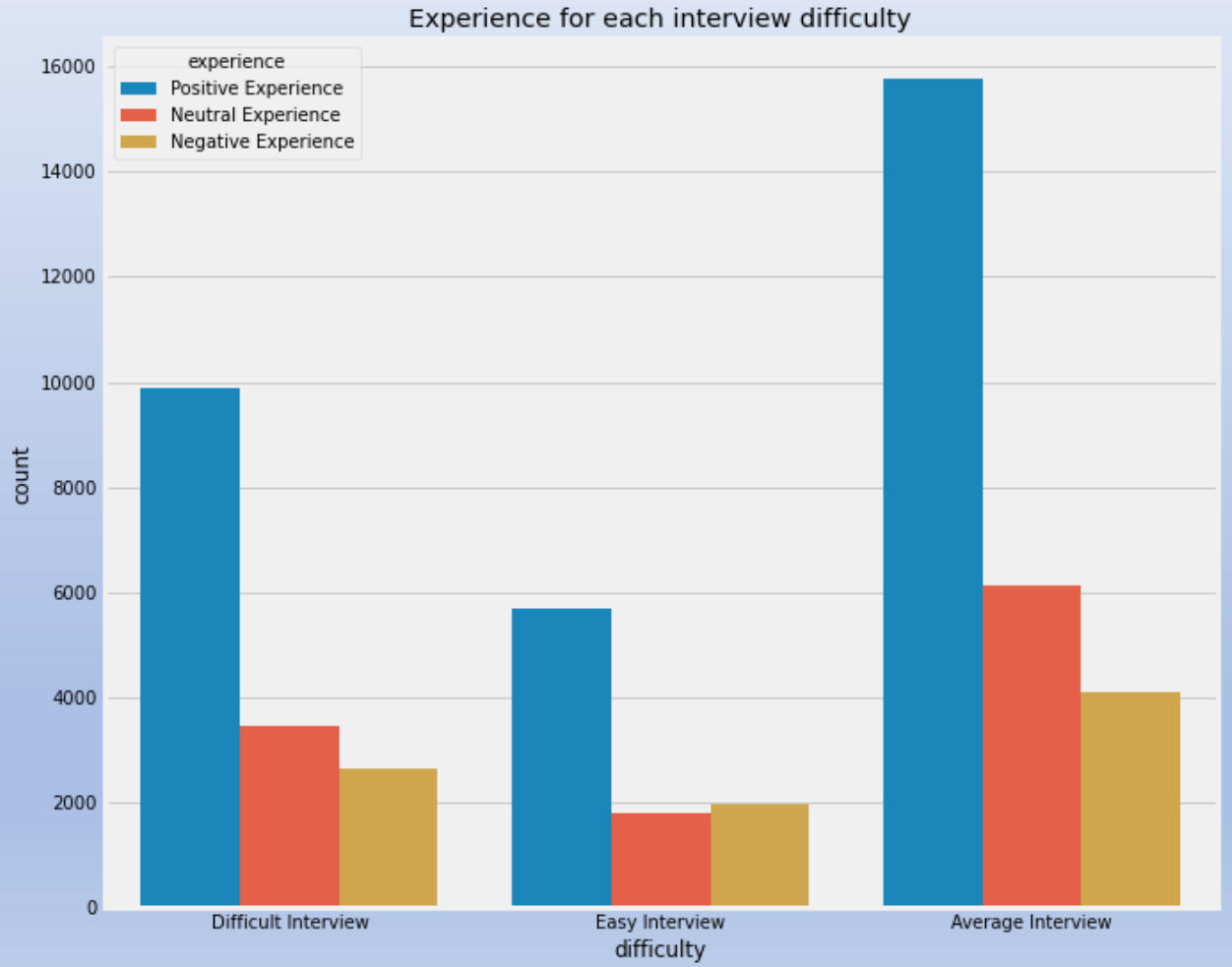
Web Scraping Apple, Microsoft, Facebook, Google and Amazon glassdoor profiles for their interview reviews

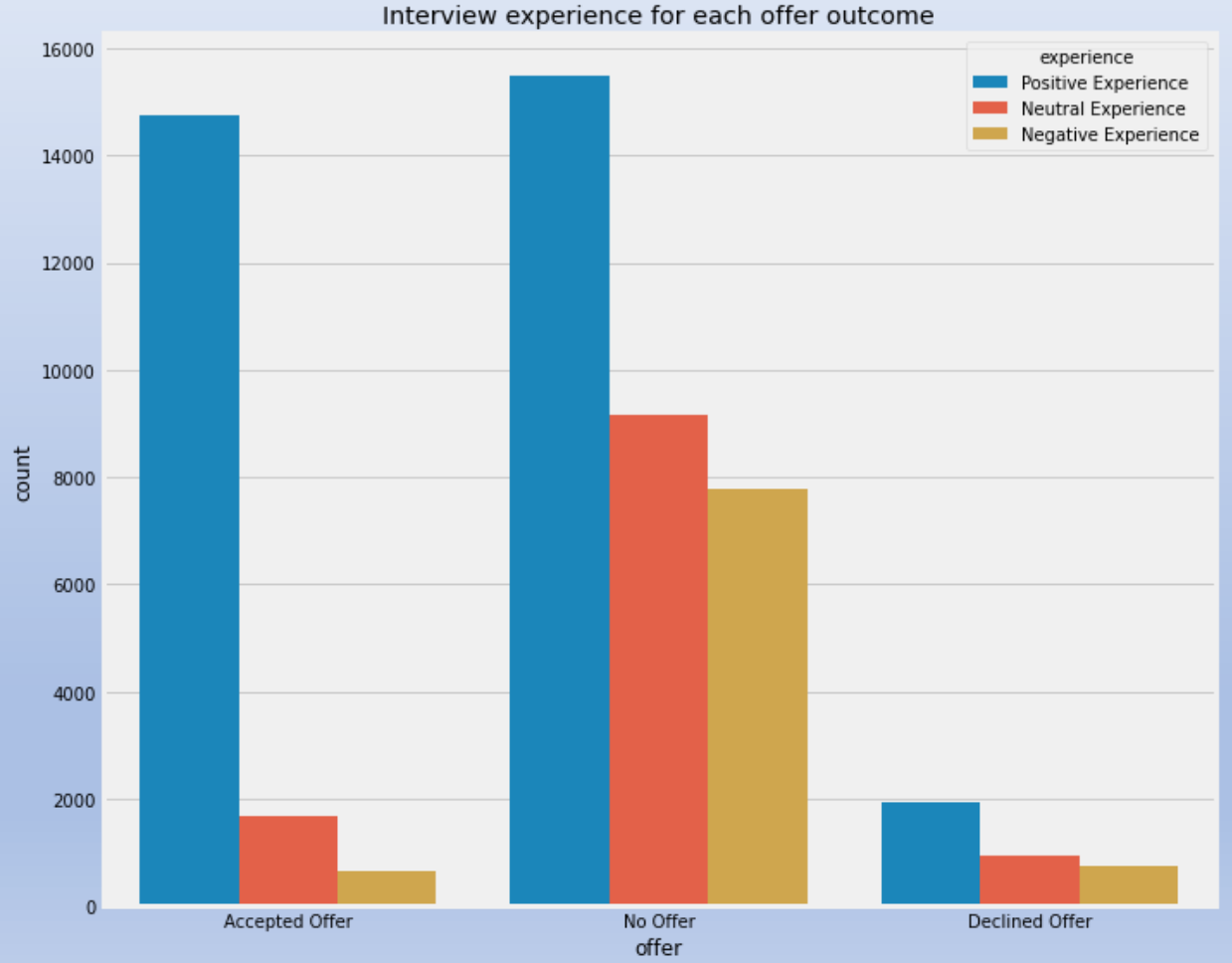
Interview experience from respective companies





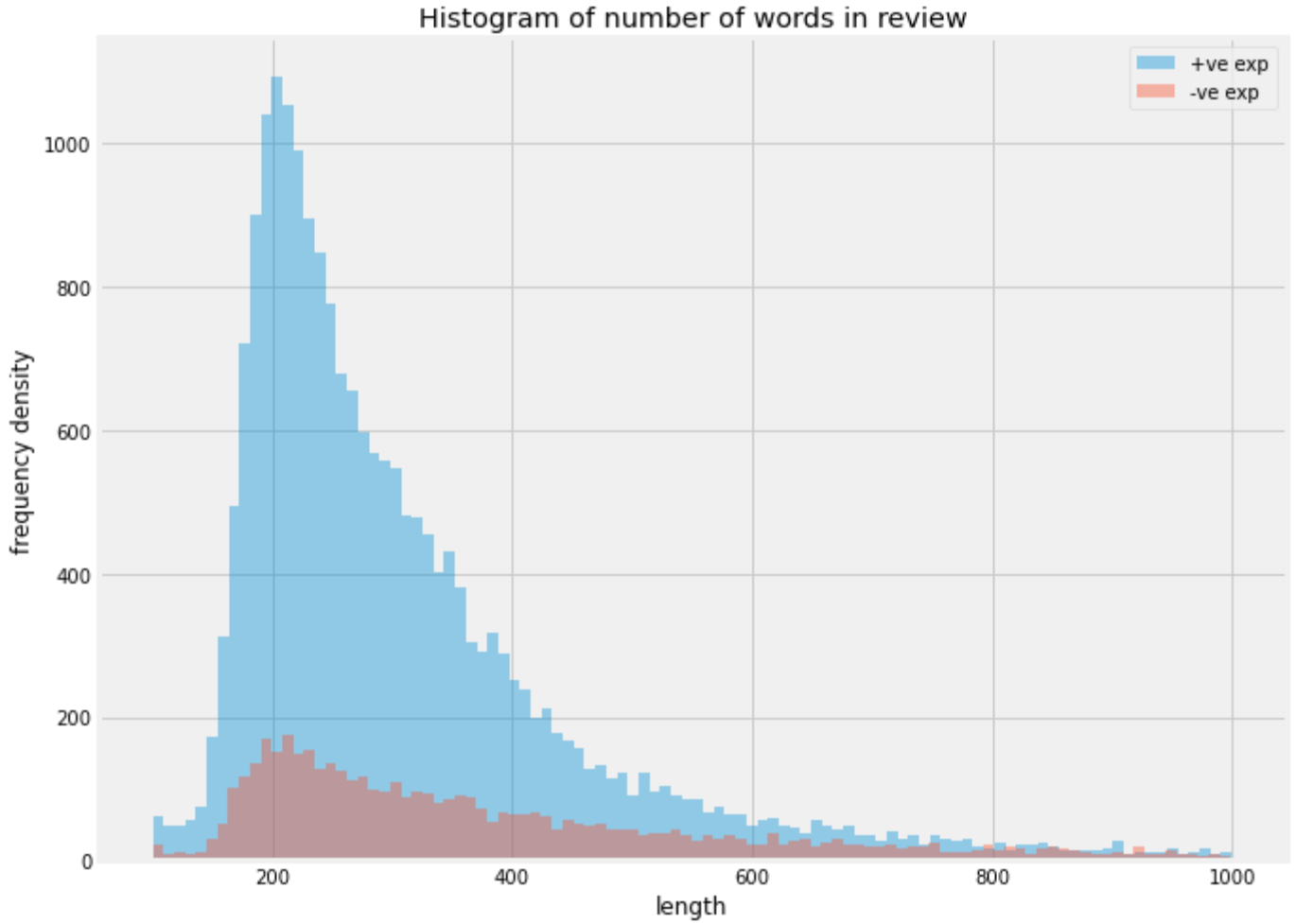
Interview experience and their difficulty/acceptance





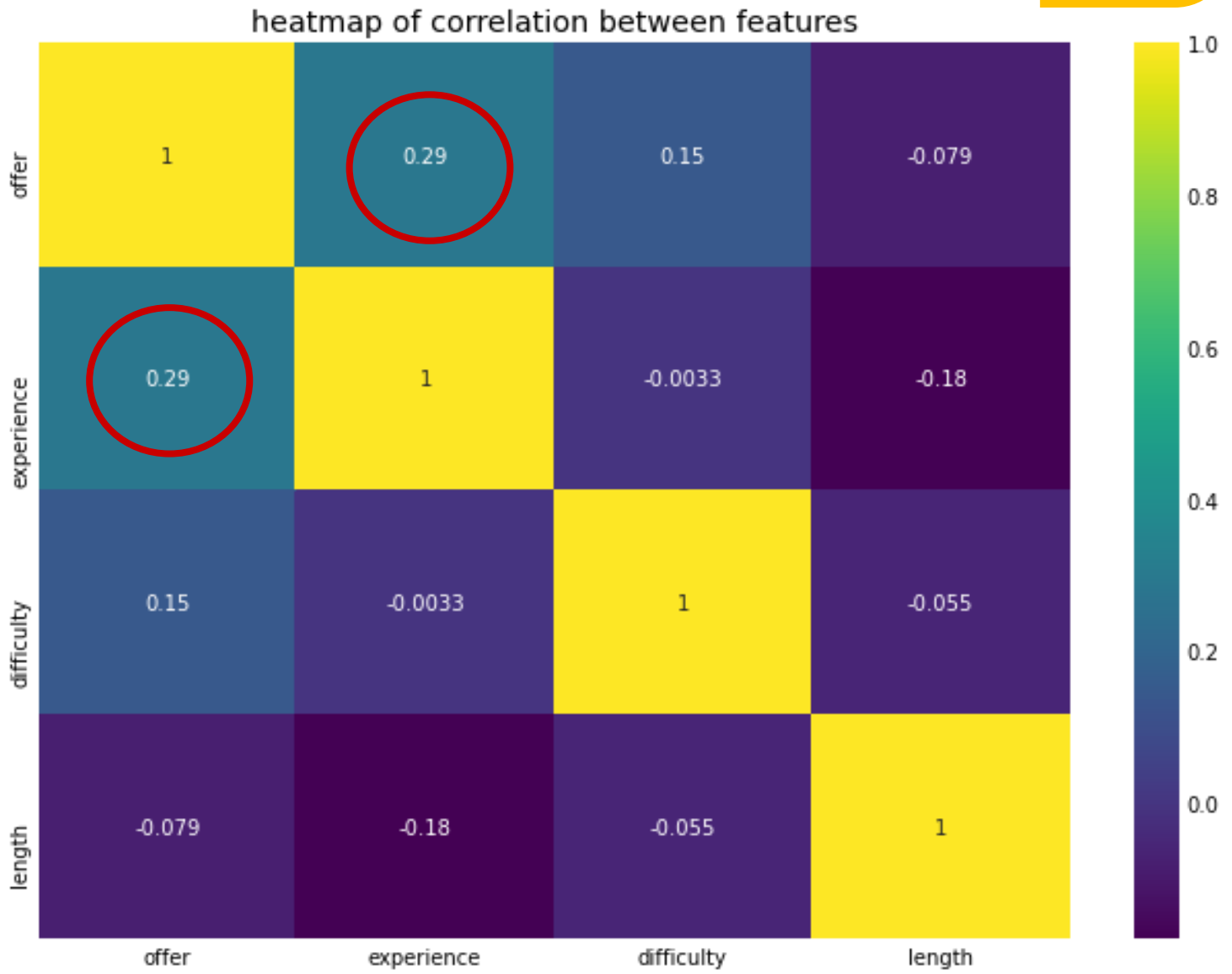
Count of words for positive and negative reviews

Distribution of word count for positive and negative reviews are generally the same



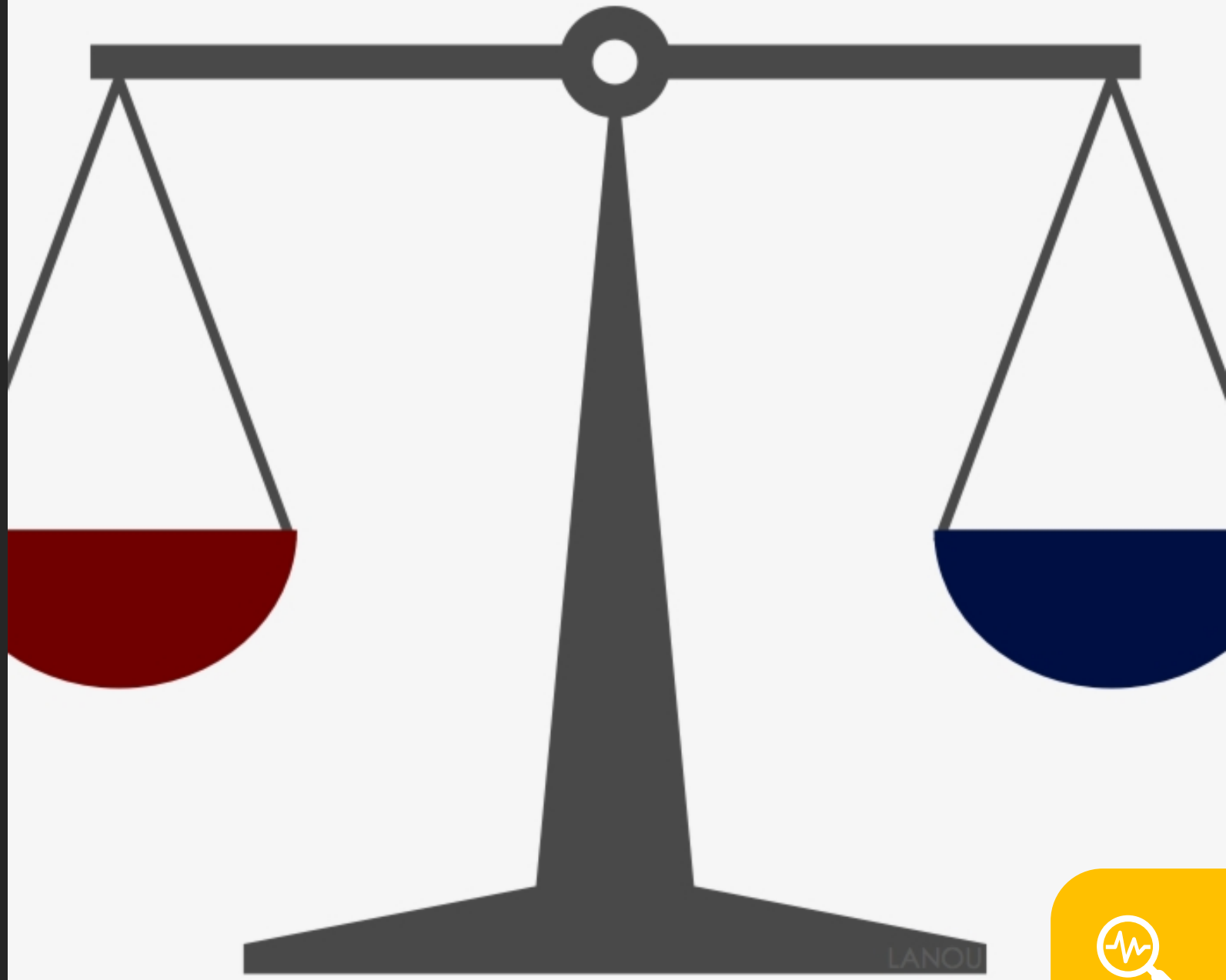
Correlation between different features and interview experience

The highest correlation among the features and interview experience is offer



Dealing with imbalanced class

- Scraped reviews from other companies on glassdoor and added the negative reviews into original data.
- Increase our negative reviews from 4762 to 11800
- There are 21079 positive reviews for our majority class



Negative Experience
Positive Experience

Word pre-processing

Remove swear words, flags, emojis and html texts

Convert all words to lower case

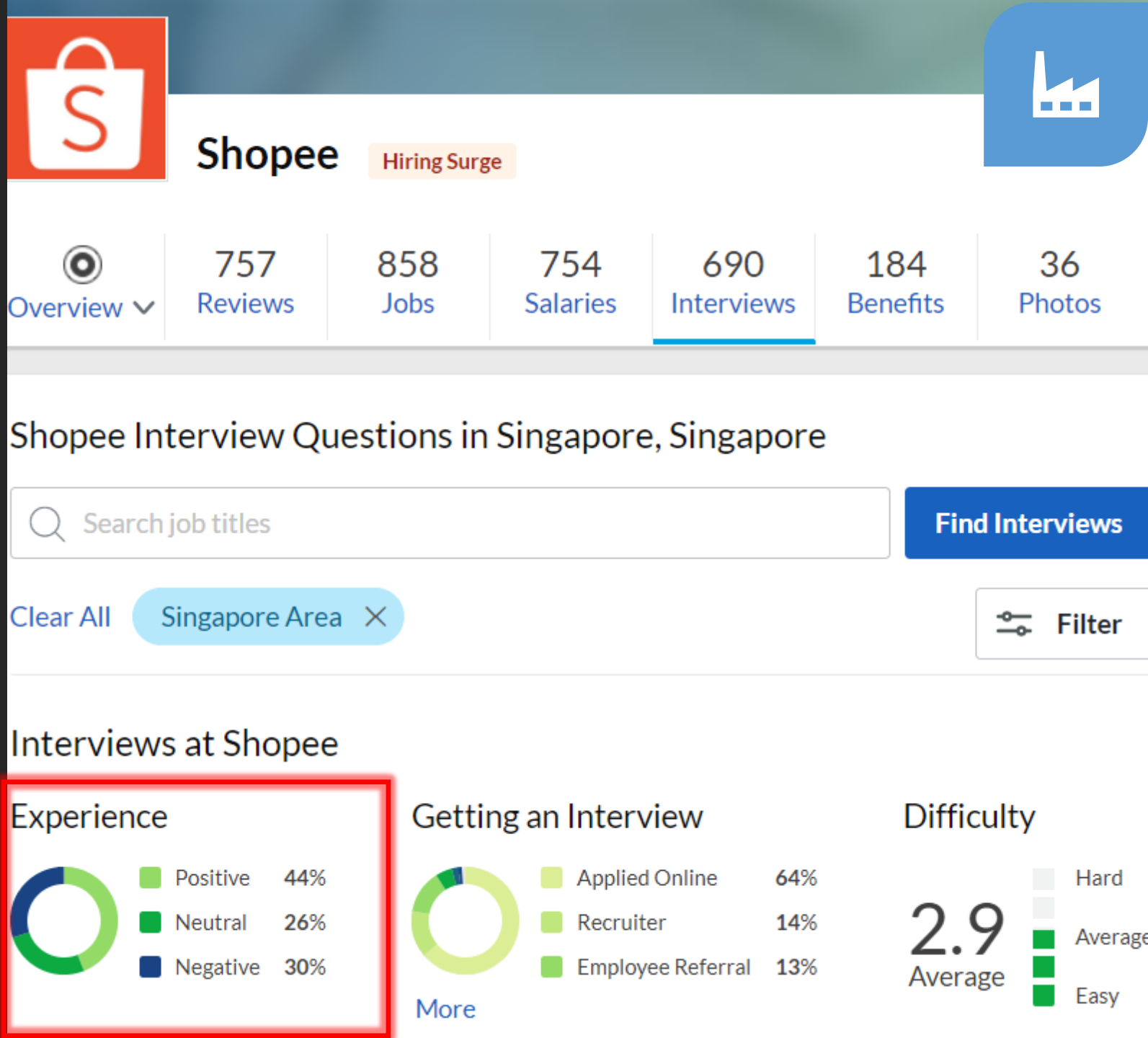
Lemmatizing each word

Removing words with less than 2 characters

Model Selection and evaluation

Model	Train Acc	Val Acc	Sensitivity	Specificity	ROC-AUC
K Nearest Neighbours	0.7534	0.67609	0.9412	0.2025	0.6402
Multinomial Naïve Bayes	0.8243	0.8239	0.8833	0.7178	0.8791
Logistic Regression	0.8337	0.8349	0.9018	0.7152	0.8978
Random Forest	0.7803	0.785	0.8254	0.7127	0.8487
XGBoost	0.7298	0.7293	0.9957	0.2533	0.8637

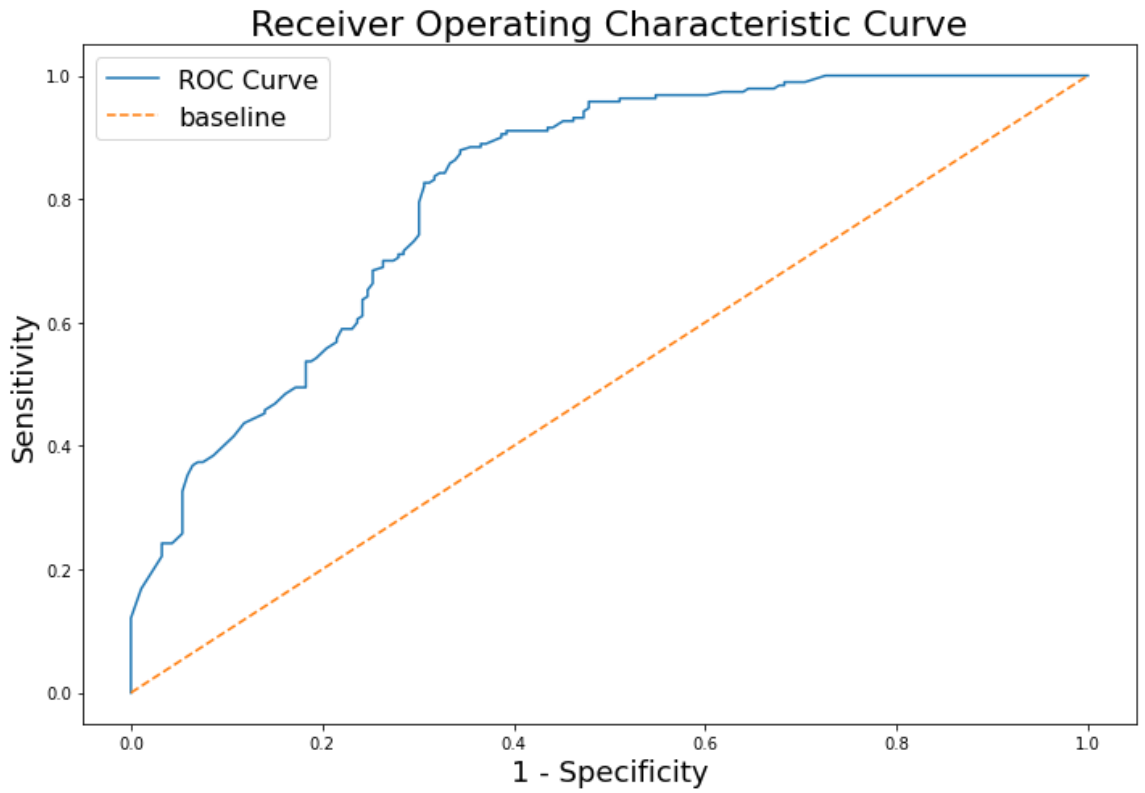
Production Model on real data



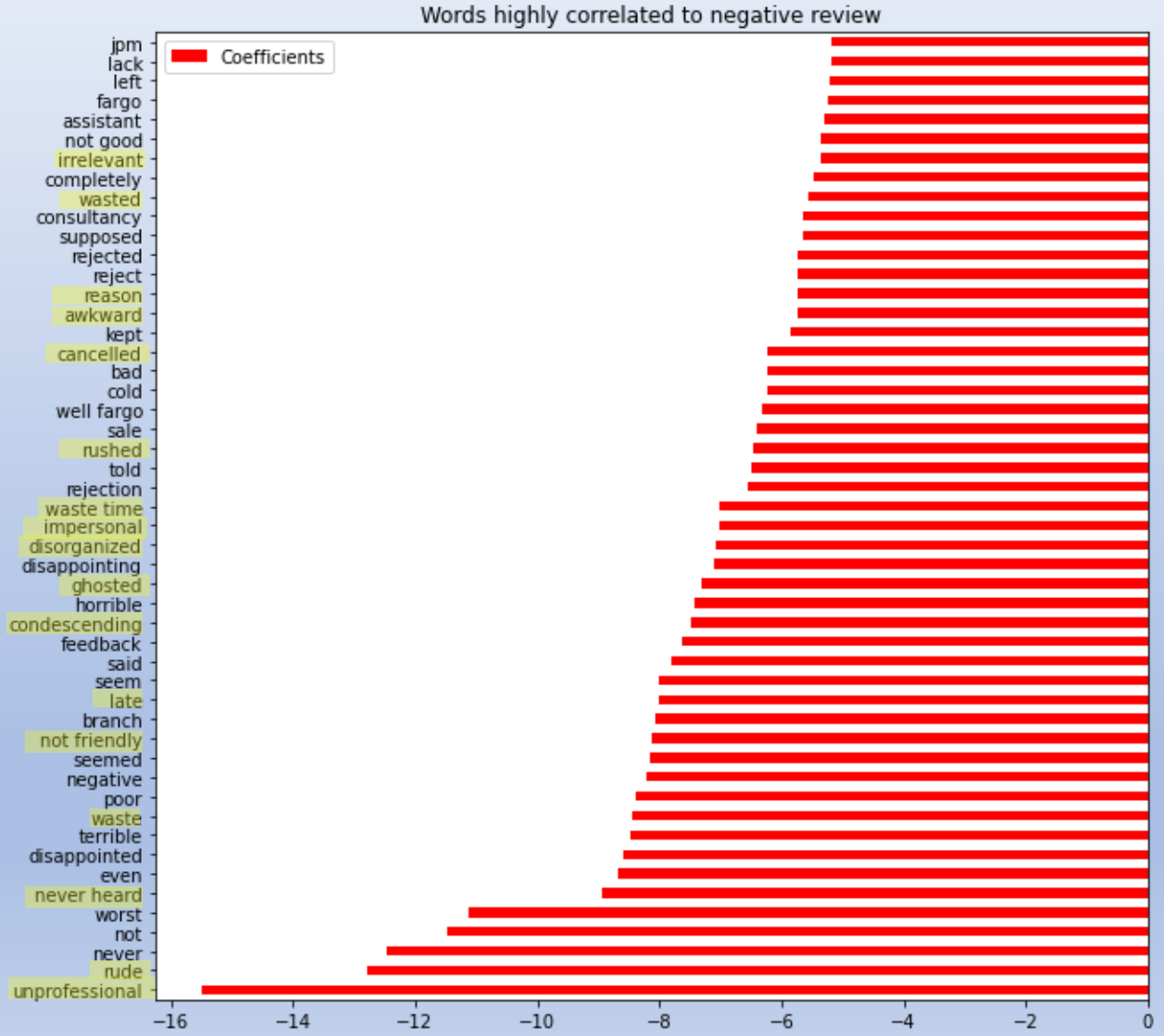
Model Performance on real data

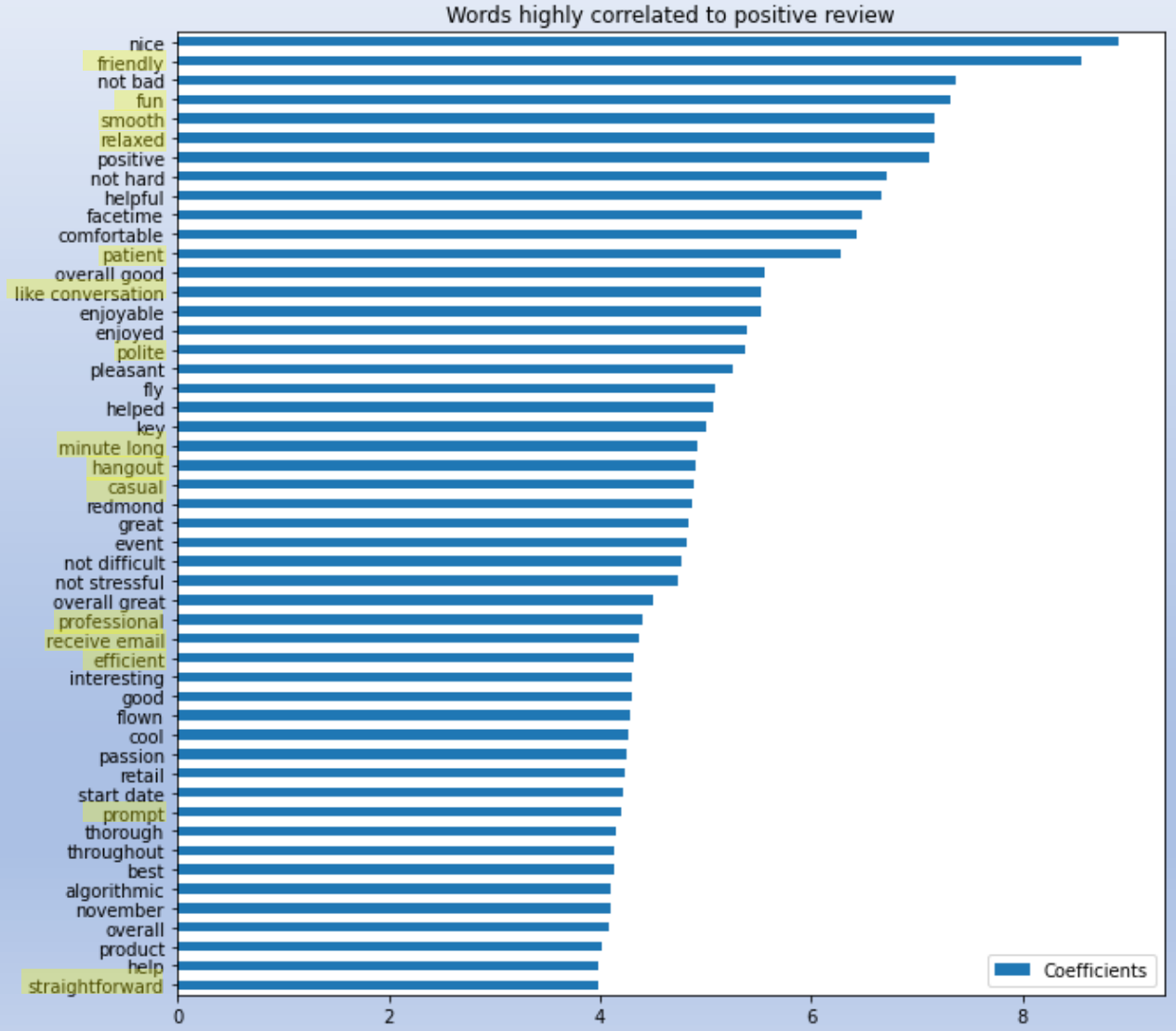
Model	Accuracy	Sensitivity	Specificity	ROC-AUC
Logistic Regression	0.7606	0.8895	0.629	0.8143

Confusion Matrix		Predicted Columns	
		Predicted Negative Review	Predicted Positive Review
Actual Labels	Actual Negative review	117	69
	Actual Positive review	21	169



Findings


Words highly correlated to negative review
Coefficients
jpm
lack
left
fargo
assistant
not good
irrelevant
completely
wasted
consultancy
supposed
rejected
reject
reason
awkward
kept
cancelled
bad
cold
well fargo
sale
rushed
told
rejection
waste time
impersonal
disorganized
disappointing
ghosted
horrible
condescending
feedback
said
seem
late
branch
not friendly
seemed
negative
poor
waste
terrible
disappointed
even
never heard
worst
not
never
rude
unprofessional
−16
−14
−12
−10
−8
−6
−4
−2
0


Words highly correlated to positive review
nice
friendly
not bad
fun
smooth
relaxed
positive
not hard
helpful
facetime
comfortable
patient
overall good
like conversation
enjoyable
enjoyed
polite
pleasant
fly
helped
key
minute long
hangout
casual
redmond
great
event
not difficult
not stressful
overall great
professional
receive email
efficient
interesting
good
flown
cool
passion
retail
start date
prompt
thorough
throughout
best
algorithmic
november
overall
product
help
straightforward
Coefficients
0
2
4
6
8

Examples of positive and negative shopee interview reviews

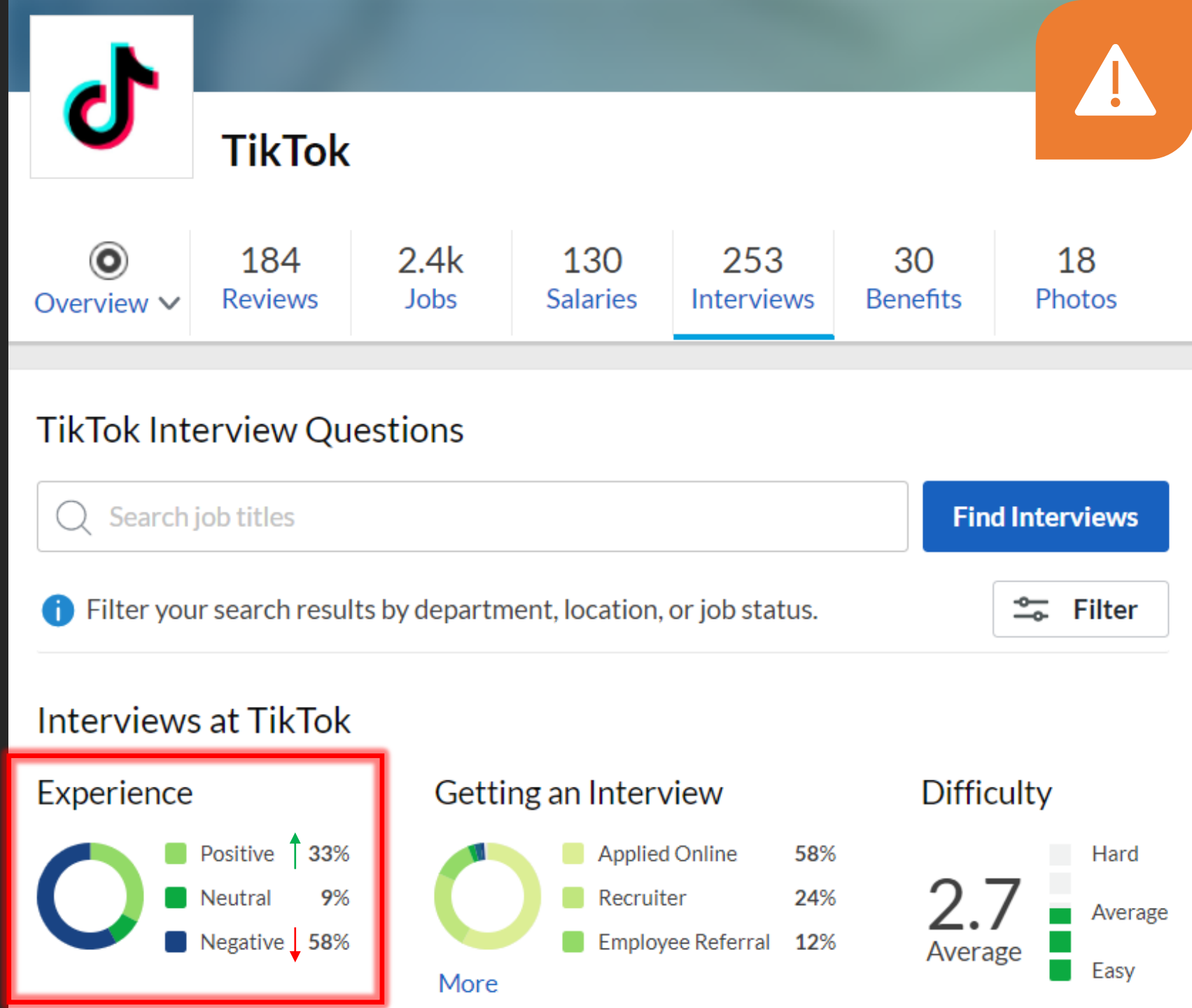
Negative Experience	Positive Experience
• One technical round. The interviewer provided a false link at the beginning. The interviewer does not even know Java and he kept asking how to design a dynamic size array in Java without reallocation. This is utterly disgusting.	• The recruiter is very nice and she will explain to you how the company views the most for this position. Phone interview is not difficult and then she directly gives me the chance for a 2nd round test interview. The test interview is about 1.5 hrs and the Python&SQL questions are very basic.
• The entire interview process was very disrespectful. They don't ask for your free time and simply give you a time. During the interview, they can be up to 30 minutes late! They don't seem to have much good things to say about the company too. A senior manager said work environment is very stressful.	• Phone interview with HR first then direct interview with the cat managers. both interviews were quite fast, within 30 minutes, i think the phone interview was shorter and the hr person was very friendly
• Phone interview, sent me an email to schedule phone interview. Sent email afterwards to ask why I didn't get the role and if there's anything I can work on. Did not get a reply.	• HR Manager was friendly and provided information before the interview took place. There was a better understanding of the happenings and the overall reputation of the company at present. The questions were as expected (i.e. tell me more about yourself, why are you applying for Shopee, etc.

Limitations

- Words that are new to the corpus will not be considered when doing vectorizing transformation and prediction
- If the review is in another language, our model is not able to accurately predict the sentiment
- Sarcasm within a review might result in inaccurate prediction
- Logistic model assumes linear separability between different texts. However in reality, texts or comments are not exactly linearly separable
- Some of the key words with high coefficients tend to be very subjective and scalable. For instance, nice is the word that has the highest coefficient for positive experience, however it is very subjective and there are too many variations of nice. Hence there is still a need for someone to review and validate if certain predictions makes sense

Increase positive interview experience

- Any company who have collected feedback from interviewees and want to understand the sentiments and gather insights can adopt this model
- Analyse feedback gathered from interviewees, and find out areas that they can improvement on
- This will probably help them achieve a higher positive interview experience rating on their public profiles, achieve a better company brand and attract more talent



Questions