



# Employee Reviews

Roohullah Mansoor

Data Scientist

May 17, 2019

# Introduction

## Employee Reviews 67K

Amazon 26,430

Microsoft 17,930

Apple 12,950

Google 7,819

Facebook 1,590

Netflix 810

What we will talk about?

Why is this important?

Explore data to find meaningful information

Predictive modeling

# HELLO!

## I am Roohullah Mansoor

Data Scientist

You can find me at [roymansoor.com](http://roymansoor.com)

# Information about Dataset

Employee Reviews 67K Rows

Company Name

Summary

Pros

Cons

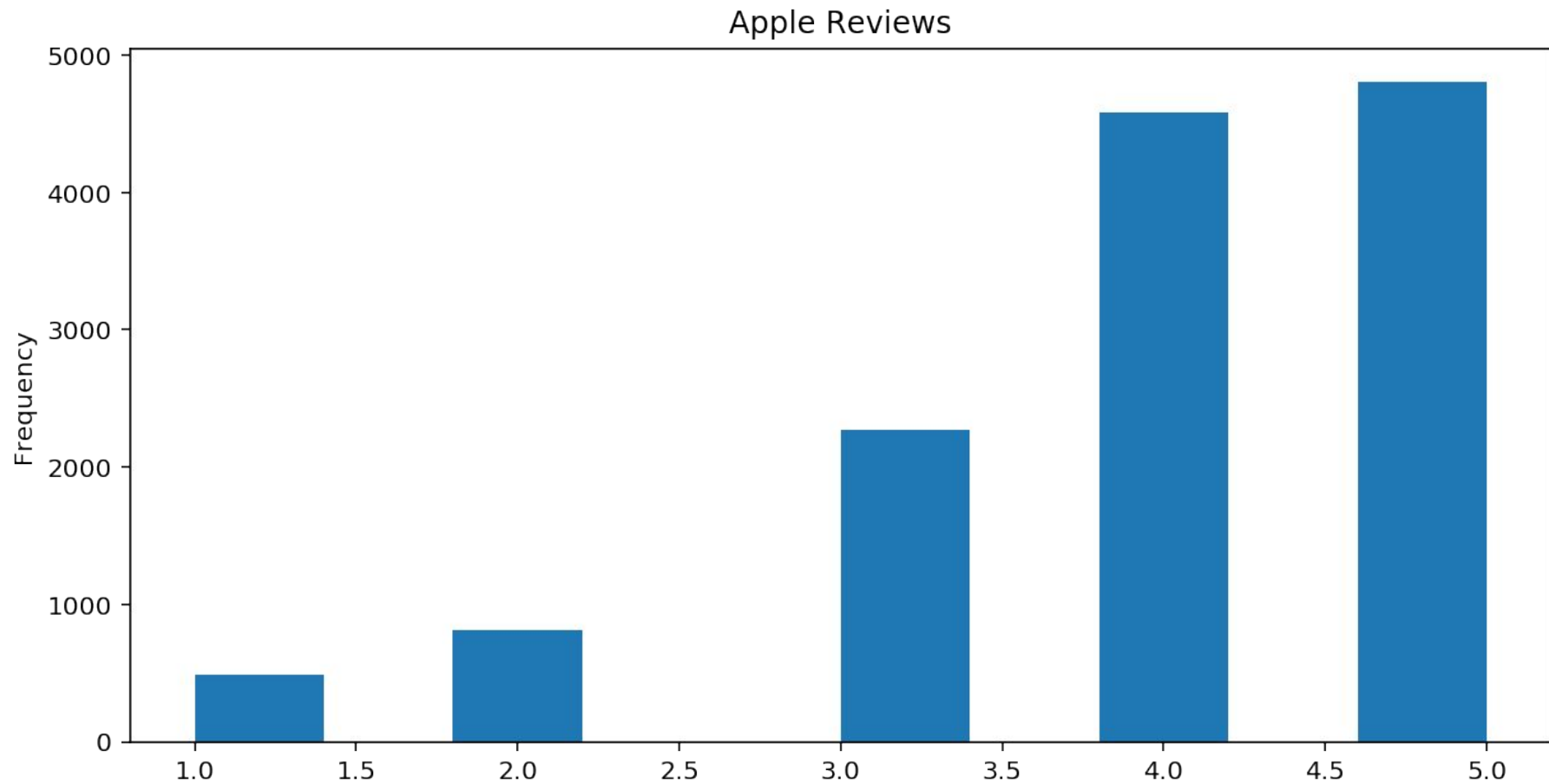
Overall ratings

Job title

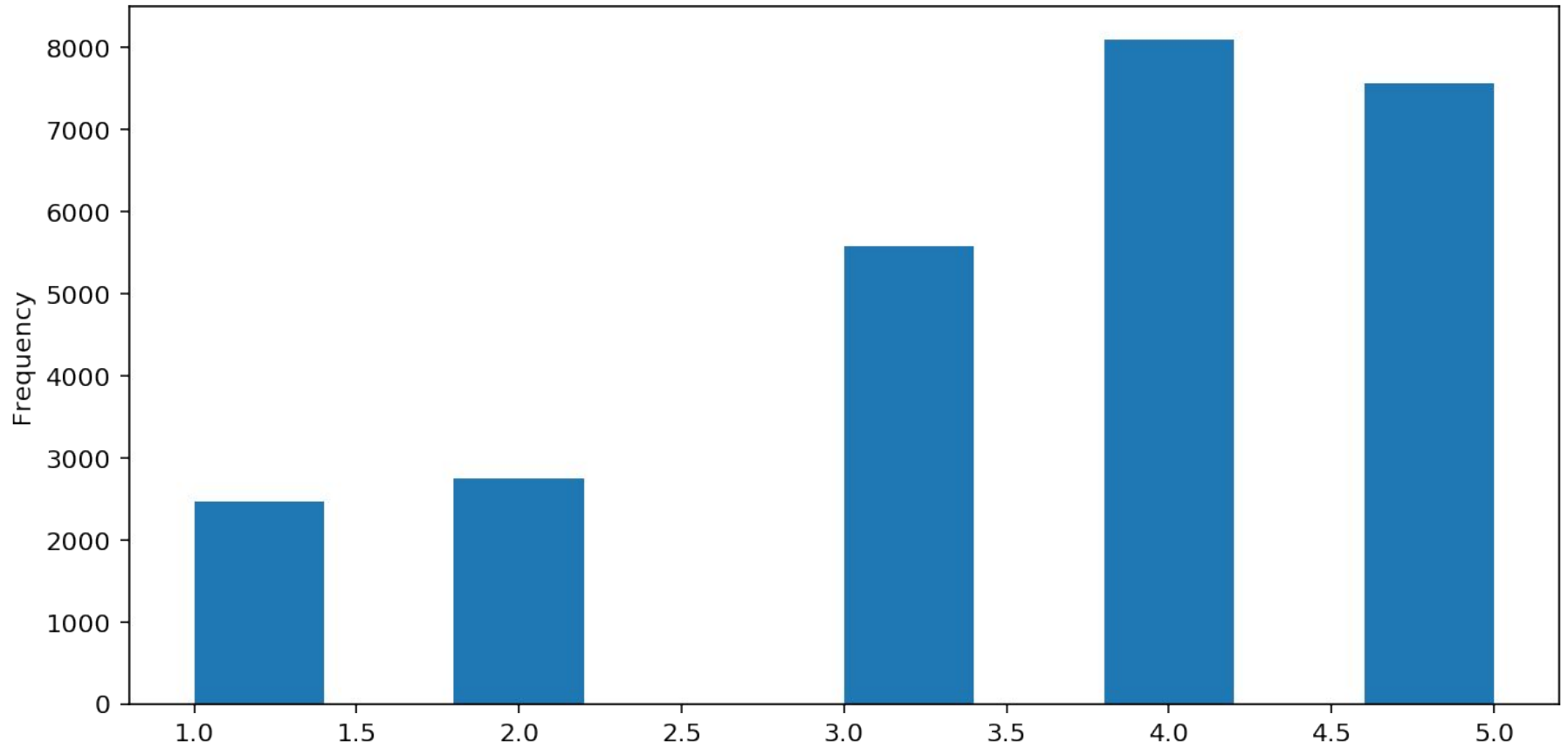
Date

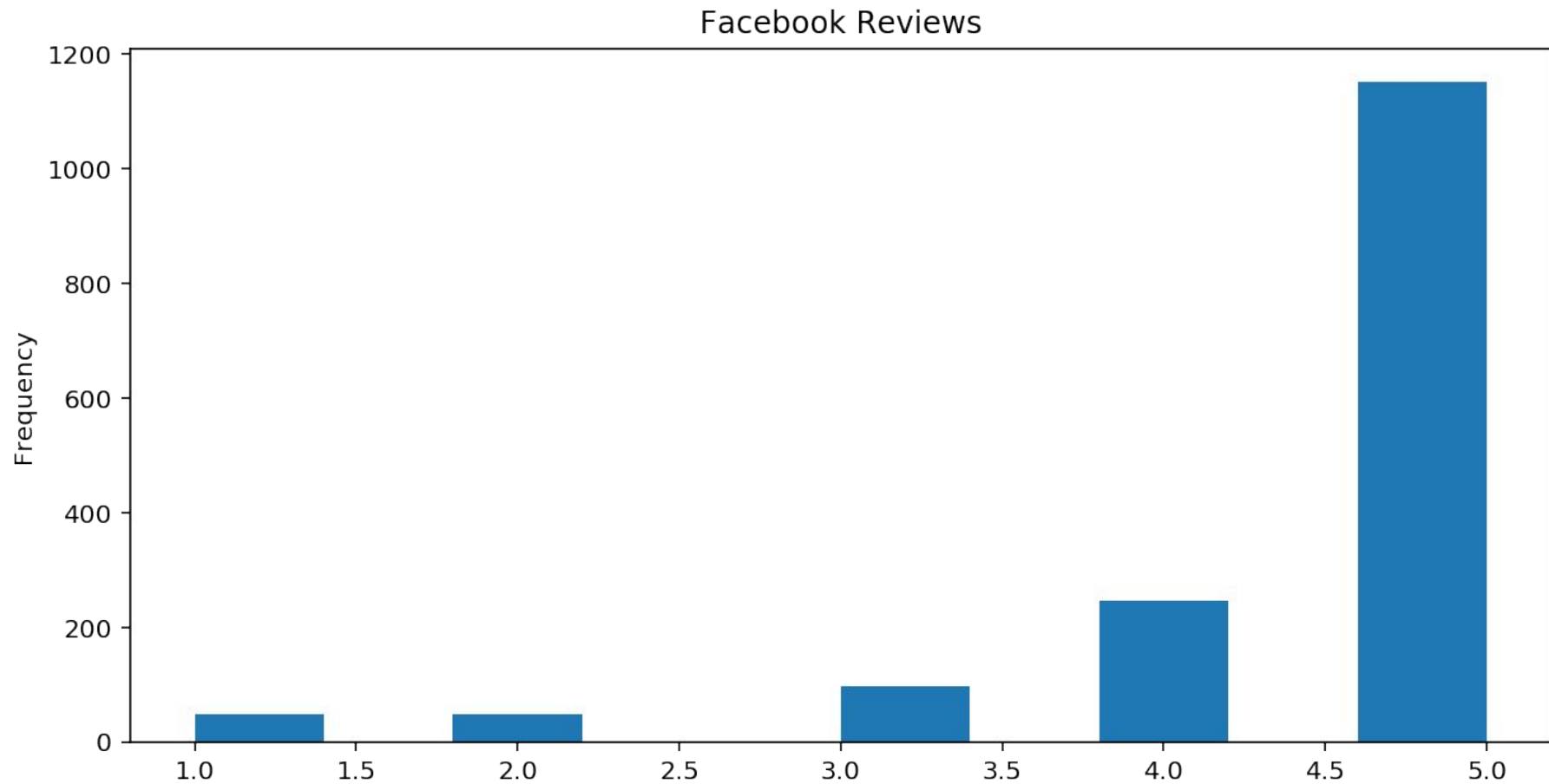
Location

...

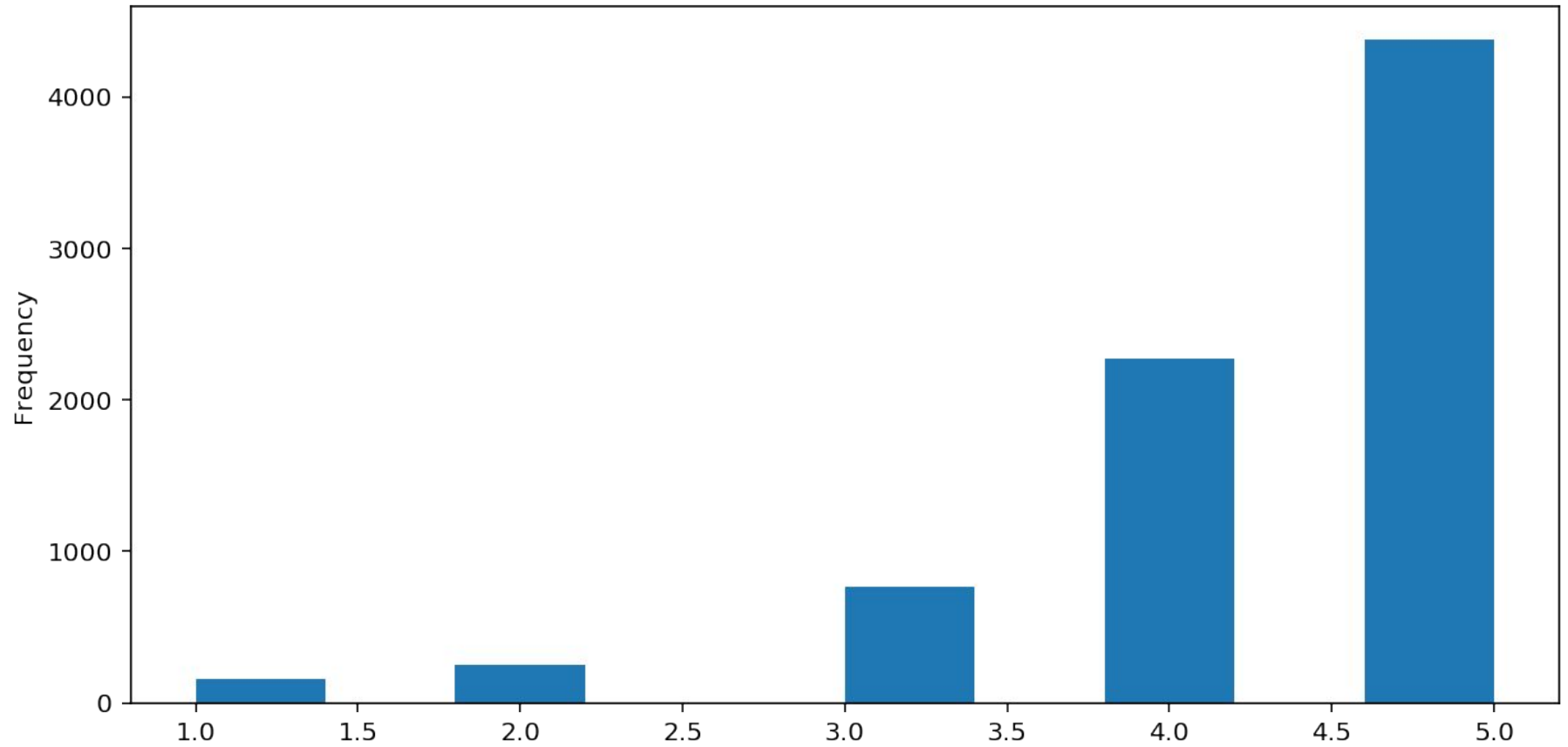


Amazon Reviews



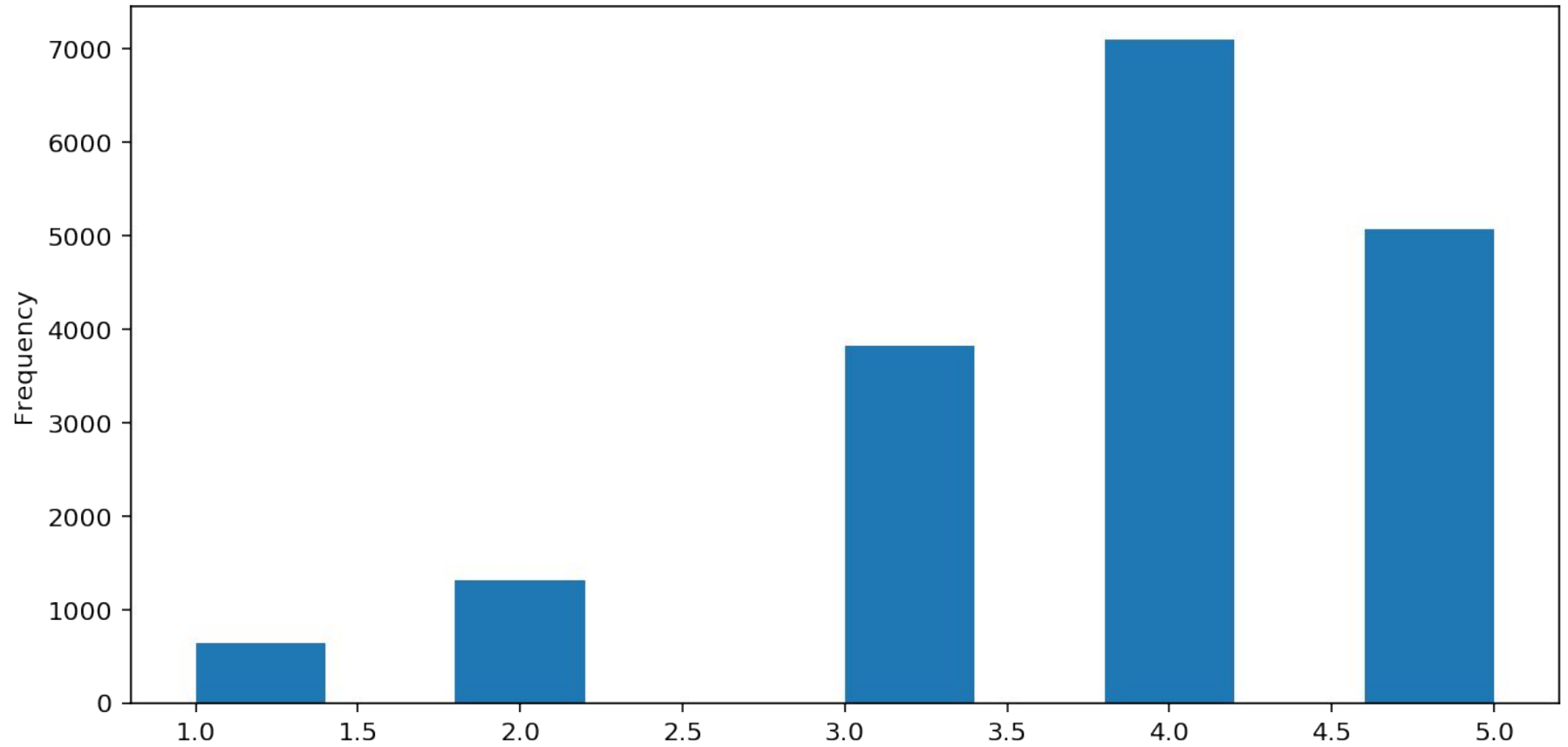


## Google Reviews

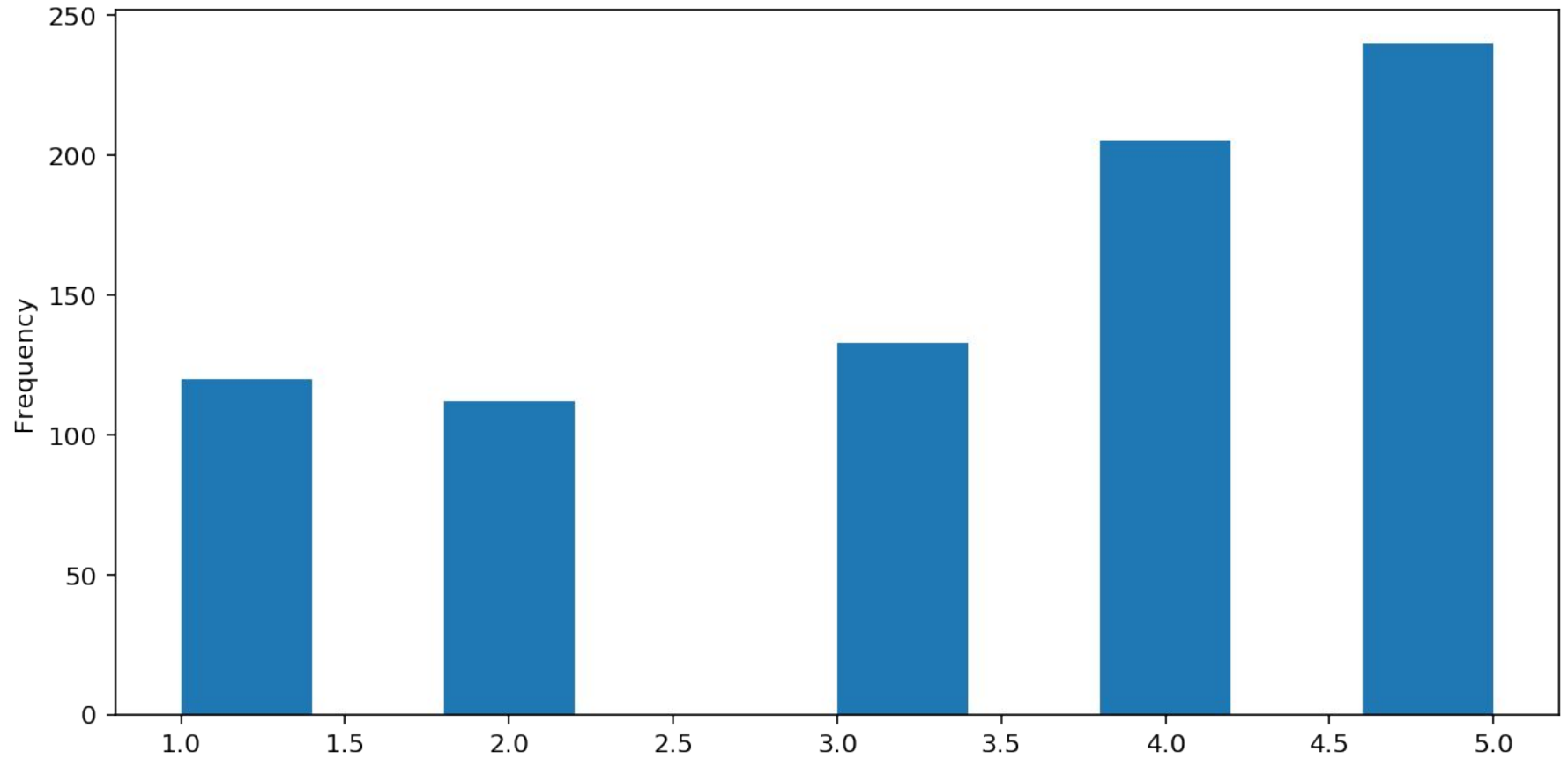




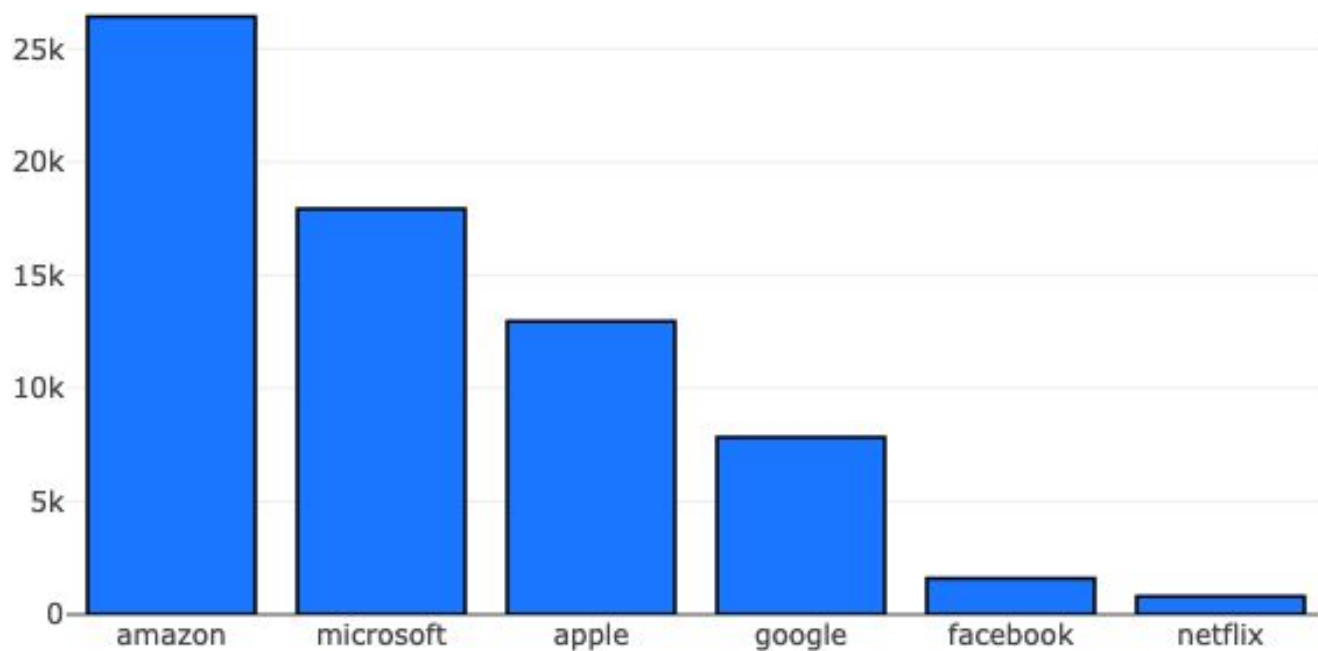
Microsoft Reviews



Netflix Reviews



Number of Reviews of Each Company



## Summary Column, Word Clouds

APPLE



AMAZON



## FACEBOOK





GOOGLE



## MICROSOFT

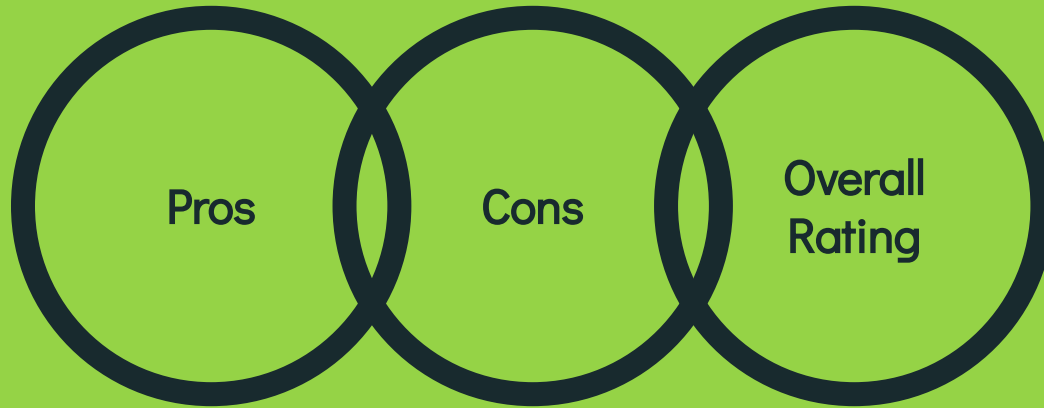




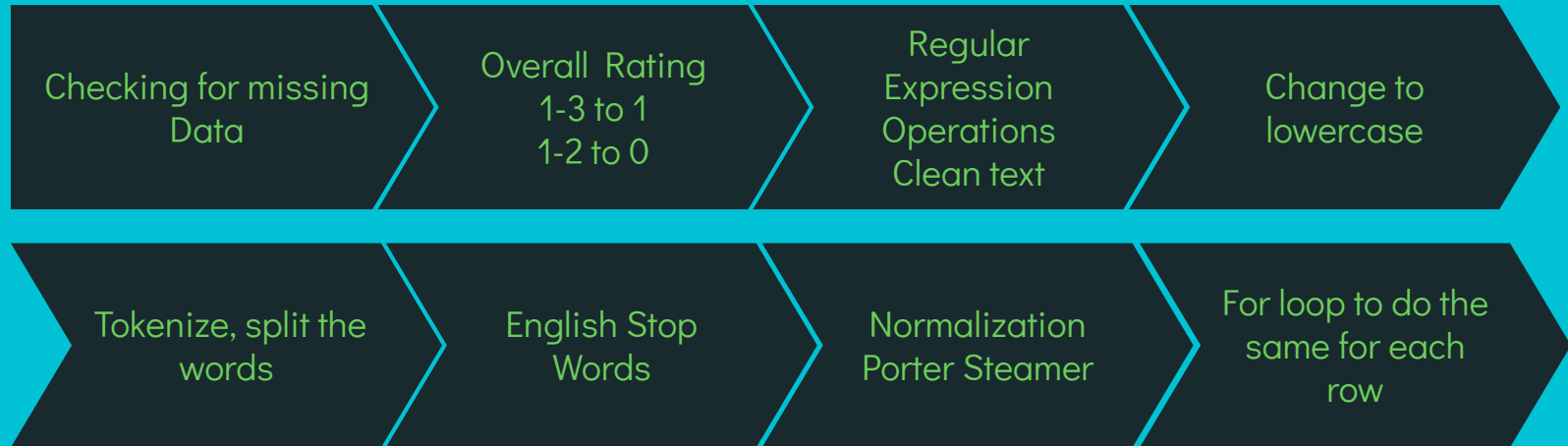
## NETFLIX



# Columns:



# Data Cleaning Process:



# Pros & Cons Data Frames:

## Pros\_df

Pros	Liked
Text	0-1
Text	0-1

## Cons\_df

Cons	Liked
Text	0-1
Text	0-1

# Pre-Processing & Modeling :

- Train Test Split
- Countvectorizer
- Naive Bayes
- Logistic Regression

# How to Improve :

- Try TFIDF
- Ngram Range
- Play with Max Features
- GridSearchCV



# THANKS!

## Any questions?

You can find me at [roymansoor.com](http://roymansoor.com)