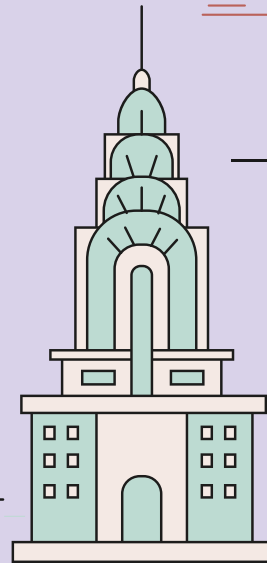
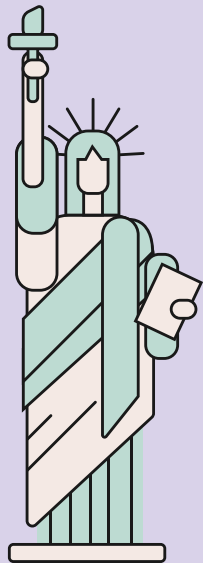


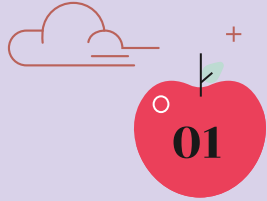
F.R.I.E.N.D.S

**The One with Natural Language
Processing**

By Gabrielle Clavell



Contents of Presentation



01

Introduction

Background on Friends and why I picked this project



02

Problem Statement

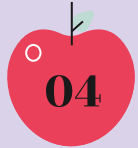
The basis of this project



03

EDA

All the exploratory analysis that was done



04

Modeling & Predictions

All the models built and their performances



05

Streamlit App

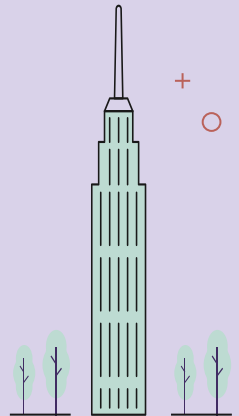
A look at the Web Application!



06

Conclusion

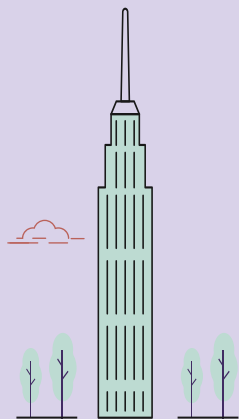
Results and Expectations





Introduction

- 90s sitcom based in Manhattan NYC (woohoo)
- Six friends: Phoebe, Chandler, Bing, Rachel, Ross, Monica, Joey
- Go through life experiences together; experience everything together; close-knit group
- Reason I chose this project: I love Friends and Natural Language Processing!

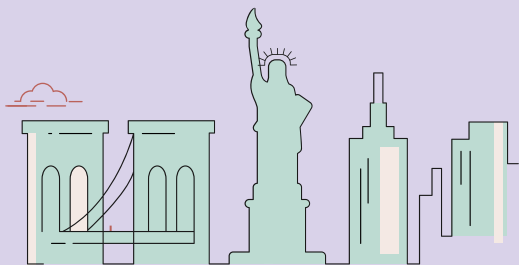




Problem Statement

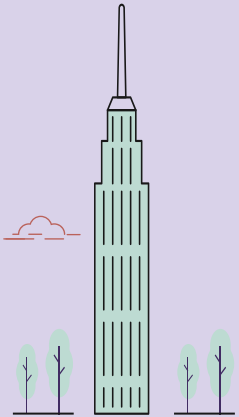


Through Natural Language Processing, people can give computers to understand text and spoken words. This project is aimed to read in the Friends dataset from Enmory NLP's repository (Character Mining) with the season, episode, character, and transcript columns, and build different models to see if each one can correctly determine a character's dialogue.



Exploring the Data

To the web app for some Graphs!



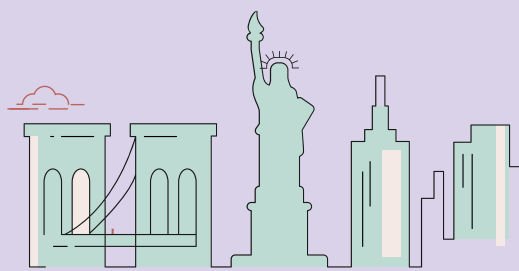
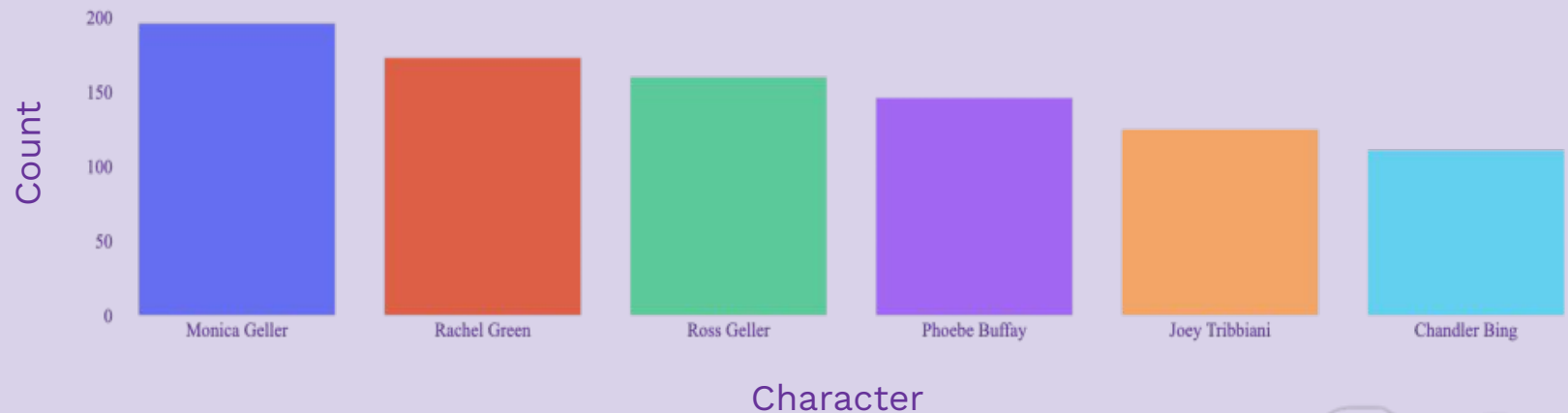
F•R•I•E•N•D•S



Exploring the Data



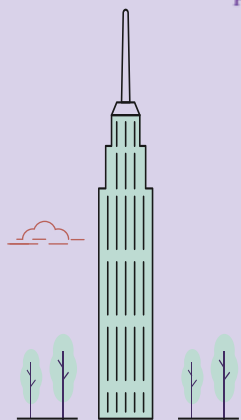
Maximum Word Count per Character





Exploring the Data

Average Word Count per Character



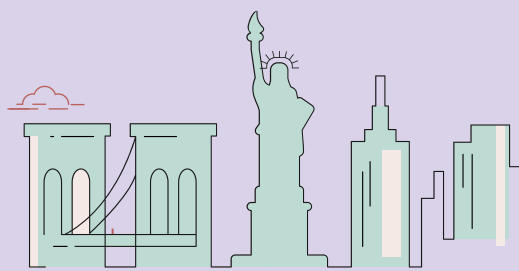
F•R•I•E•N•D•S



Exploring the Data

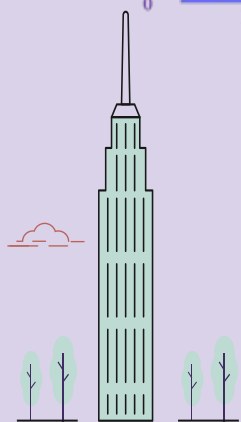


Amount of Unique Words per Character



Exploring the Data

Average Neutral Sentiment Score per Character



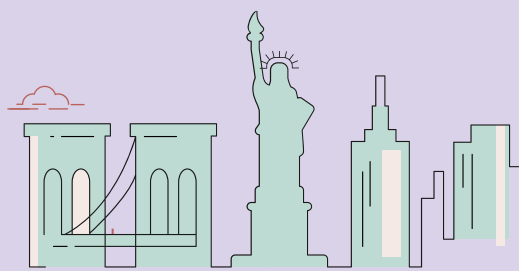
F.R.I.E.N.D.S



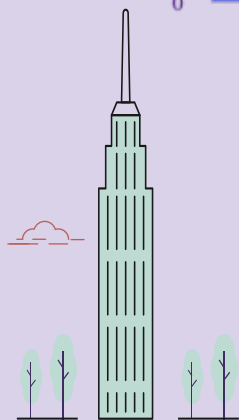
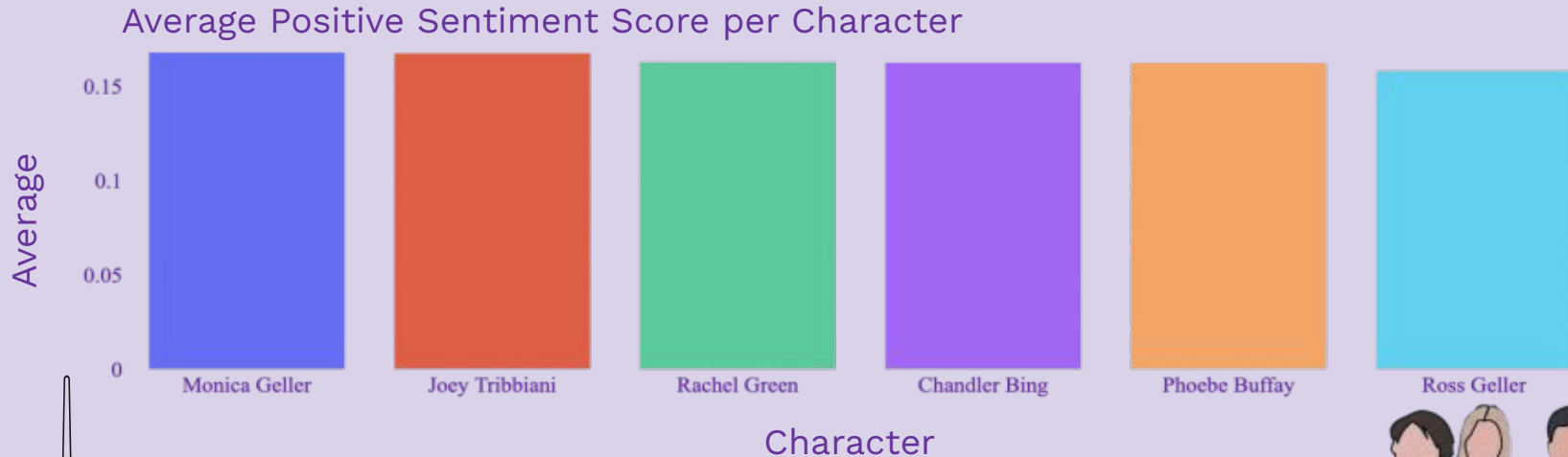
Exploring the Data



Average Negative Sentiment Score per Character

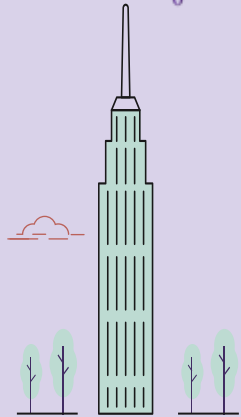
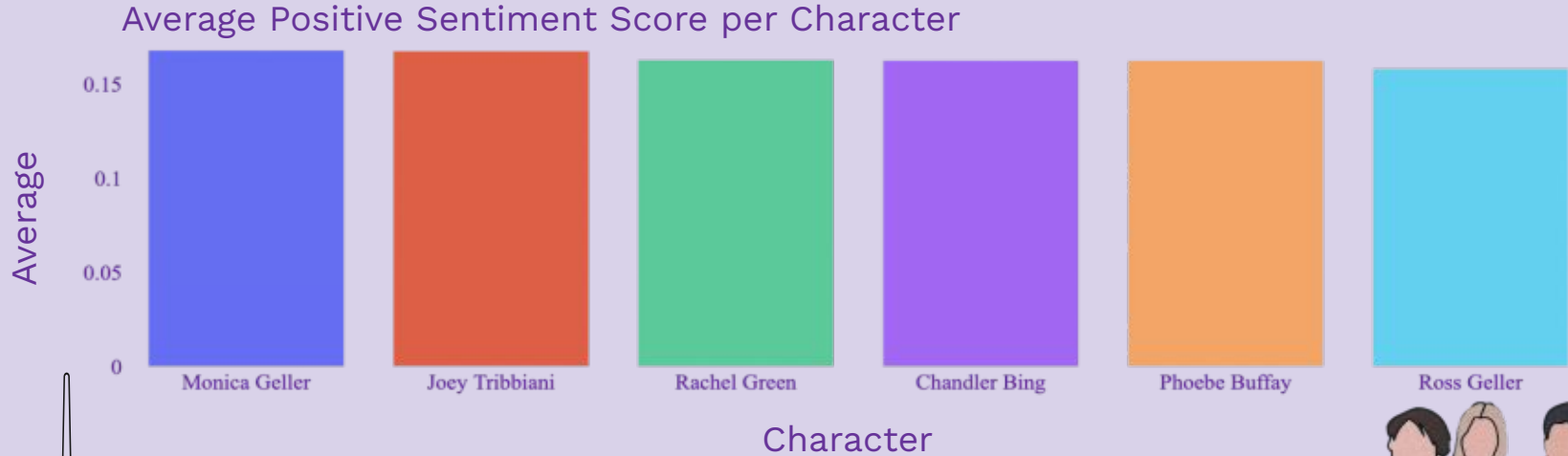


Exploring the Data



F•R•I•E•N•D•S

Exploring the Data



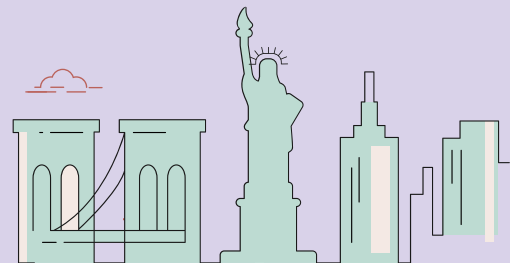
F•R•I•E•N•D•S



Modeling



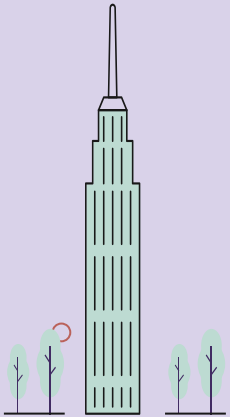
Model Algorithm	Train Score	Test Score
Näive Bayes	0.492	0.31
Logistic Regression	0.546	0.306
Random Forest	0.808	0.295
Ada Boost	0.312	0.287
K-Nearest Neighbors	0.311	0.205



Predictions



Switch back to the web applications
for some predictions!



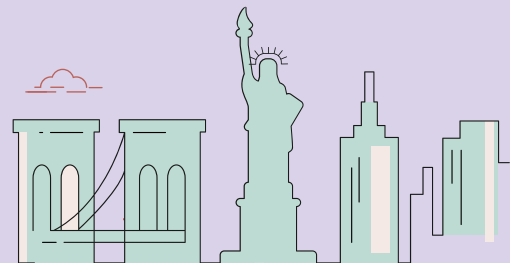
F.R.I.E.N.D.S



Conclusion



This project was a really fun project to do. I love natural language processing and machine learning, so I put both together for this project. I realized with the modeling process that it is definitely difficult to have perfect models, since predicting from a language is not the easiest. I also realized that modeling a multi-classification is difficult to work with and harder to get better results. Being said that, I learned a lot from this project and no matter how much I wanted my models to do better, it did better than the baseline model and I am happy with the final result. I hope you enjoy exploring this project and the web application!



A decorative header illustration featuring a stylized suspension bridge with two green towers and black cables. On either side of the bridge are two small green trees. The background is light purple with a few small decorative elements: a red cloud in the top left, a red star and circle in the top right, and a red cloud in the middle right.

THANKS!

**A special thanks to Adi and Caroline;
you both were amazing instructors!**

Any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

The images were from Google and thank you to Clipping Magic for a free way to edit the images!