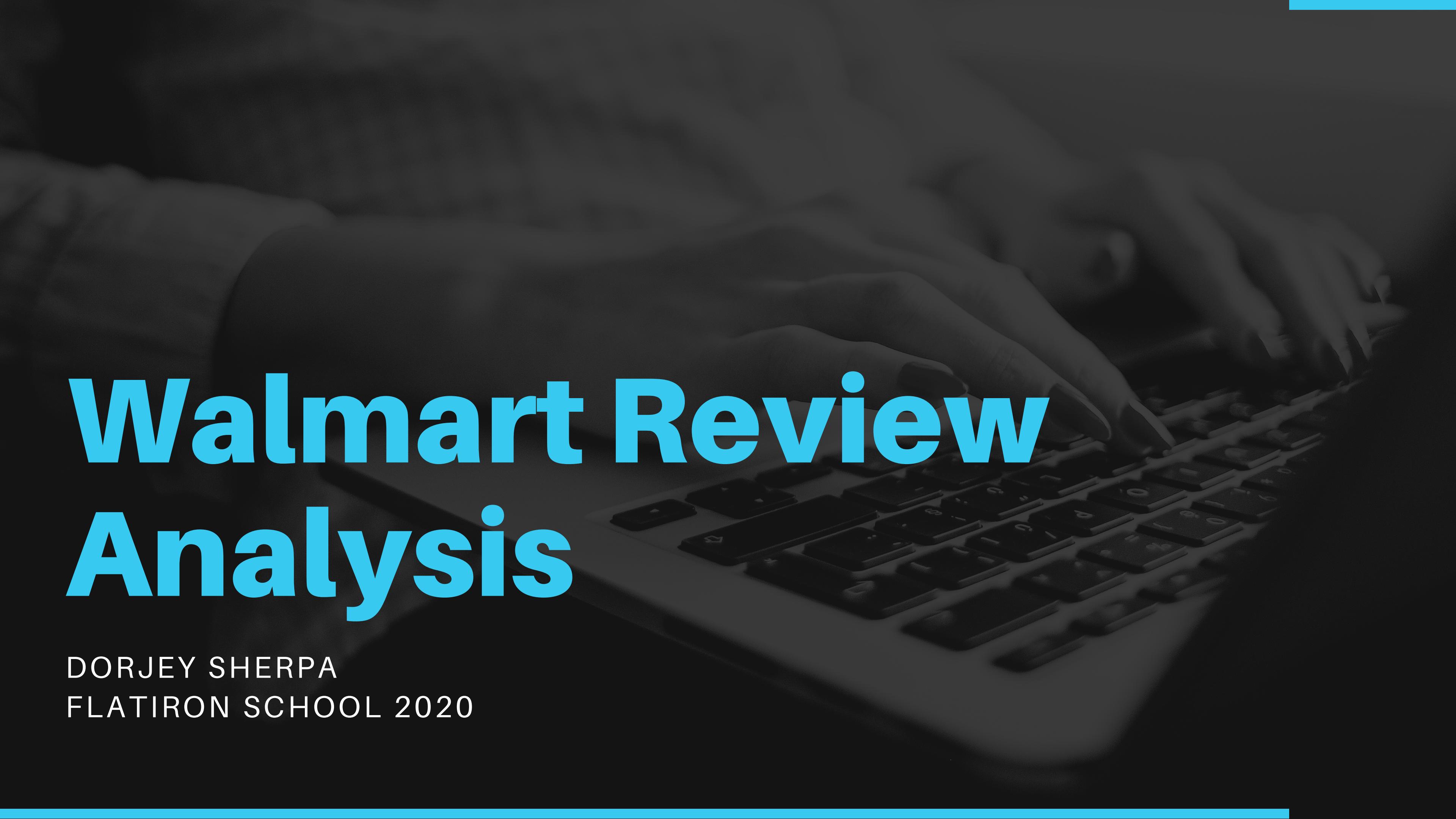
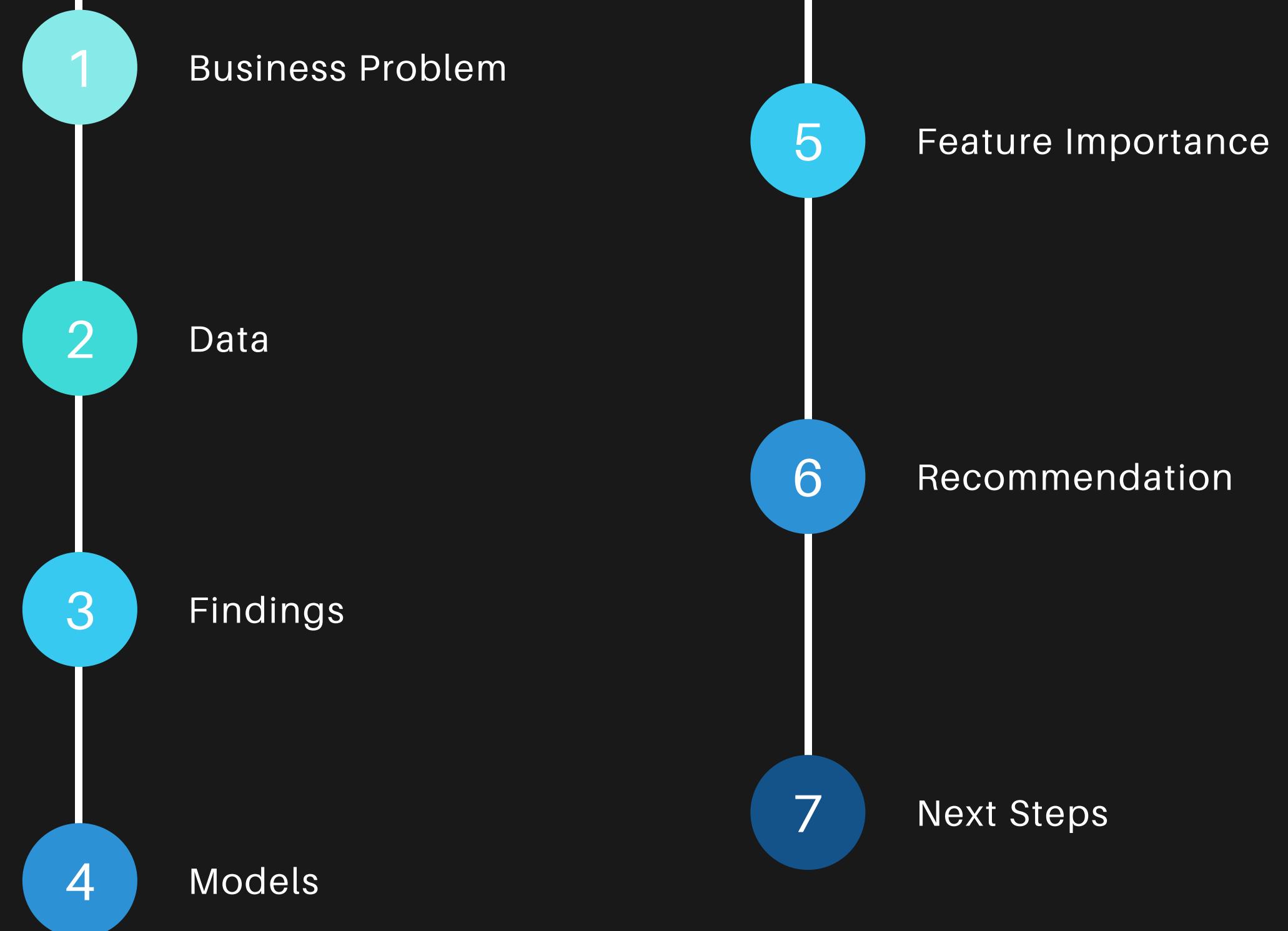


Walmart Review Analysis



DORJEY SHERPA
FLATIRON SCHOOL 2020

Overview

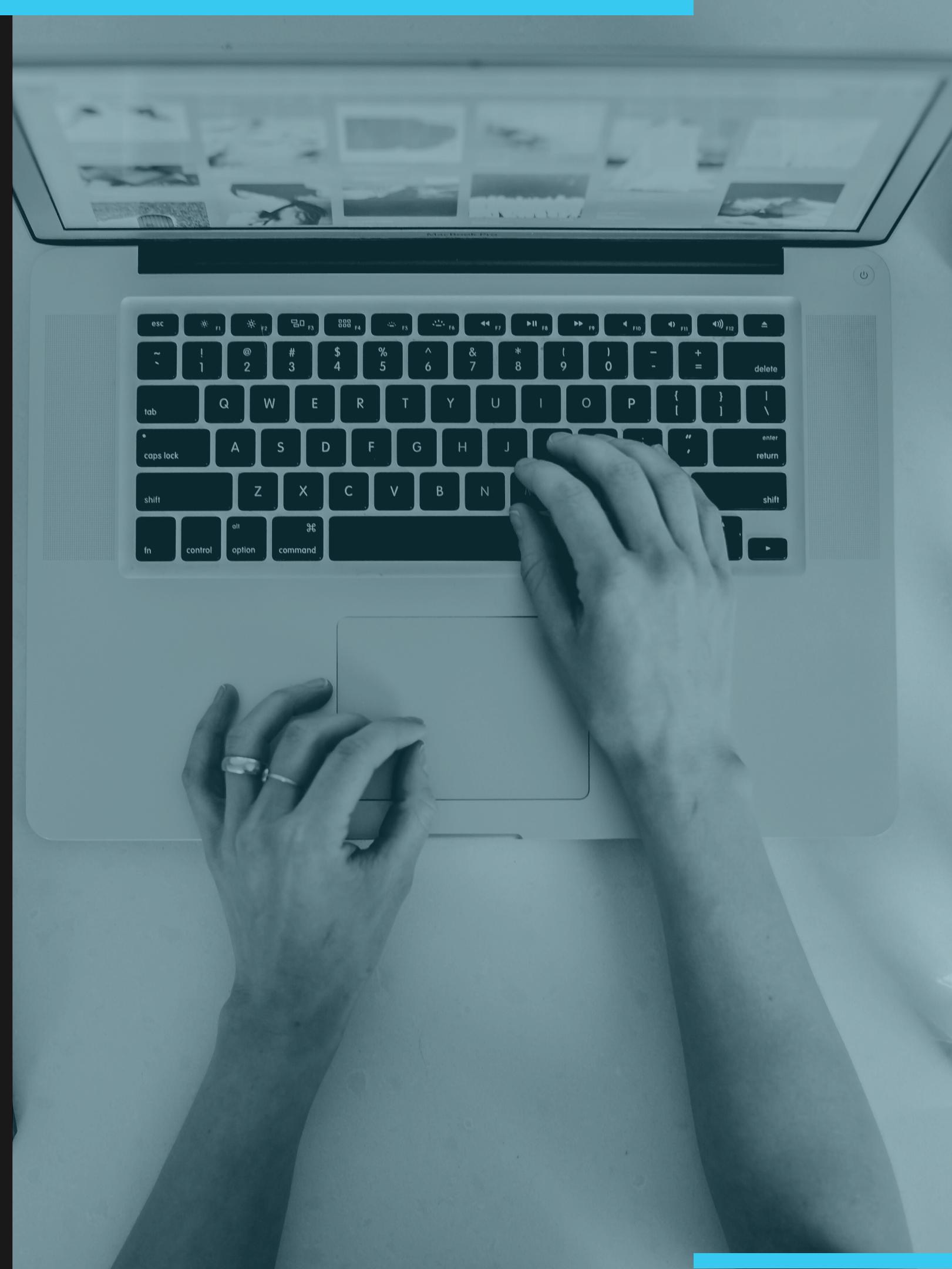


Business Problem

- Currently: price range and brand.
- Best seller items do not have ratings
- Sense of the online market

Goal:

- Analyze Walmart Clothing reviews
- Predict whether item is recommend or not by the customers
- Find the features that explain item recommendation



Data



DATA GATHERING

scraped directly from Walmart
webpage



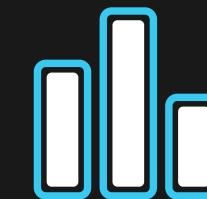
4 SECTIONS

Men's clothing
Women's clothing
Boys' clothing
Girls' clothing



TIMELINE

scraped between November 27 -
November 30



OBSERVATIONS

1225 observations with 10
features

Data Cleaning, Preprocessing and Feature Engineering

- Remove noise
- Regex
- Tokenizer
- Stopwords
- Lemmatizer
- LDA Topic Modeling

End Result:

- 1149 observations with 29 features

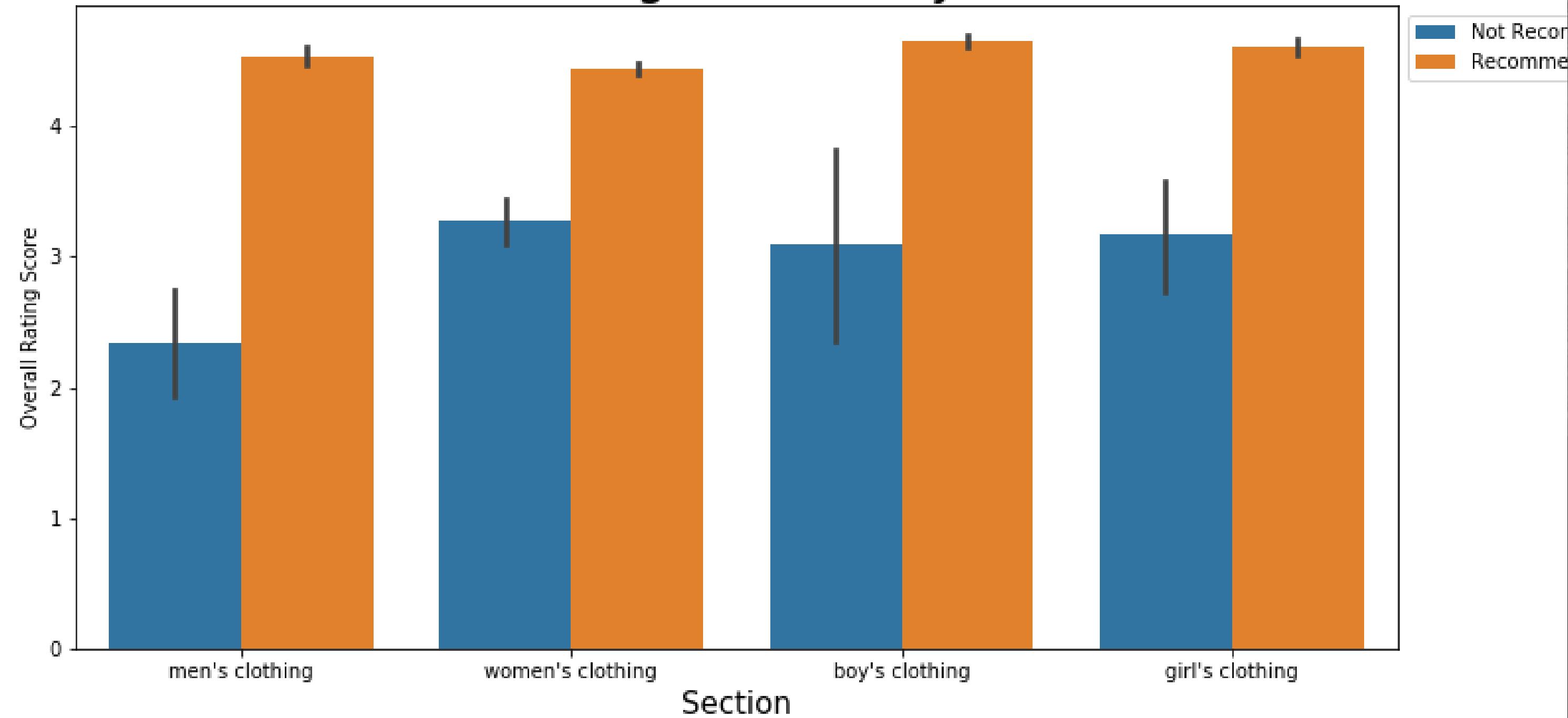
Findings

Categorize the Recommended Score:
Above 85% (inclusive) and
below 85%



Findings

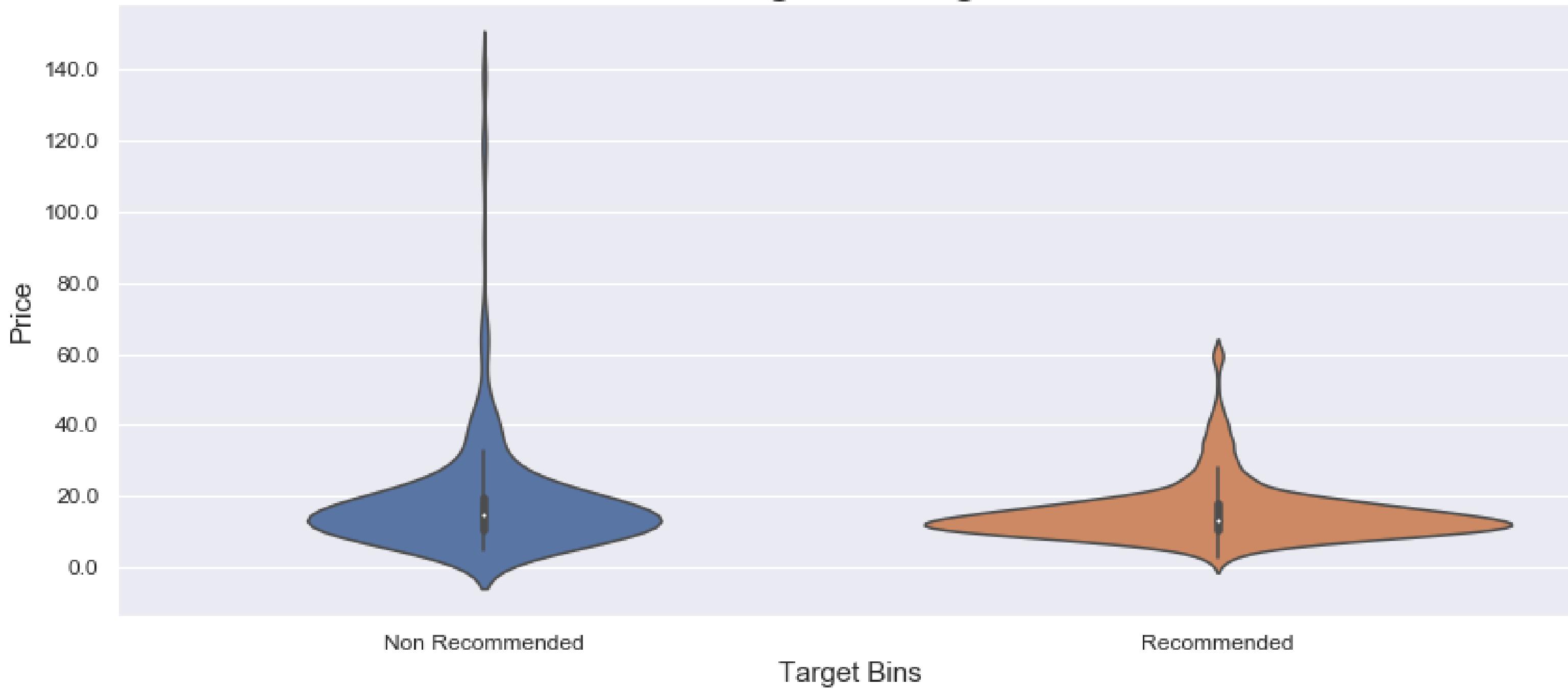
Overall Rating of items by Section



Higher the rating, more likely the item is to be recommended

Findings

Price range in Target Bins



Modeling

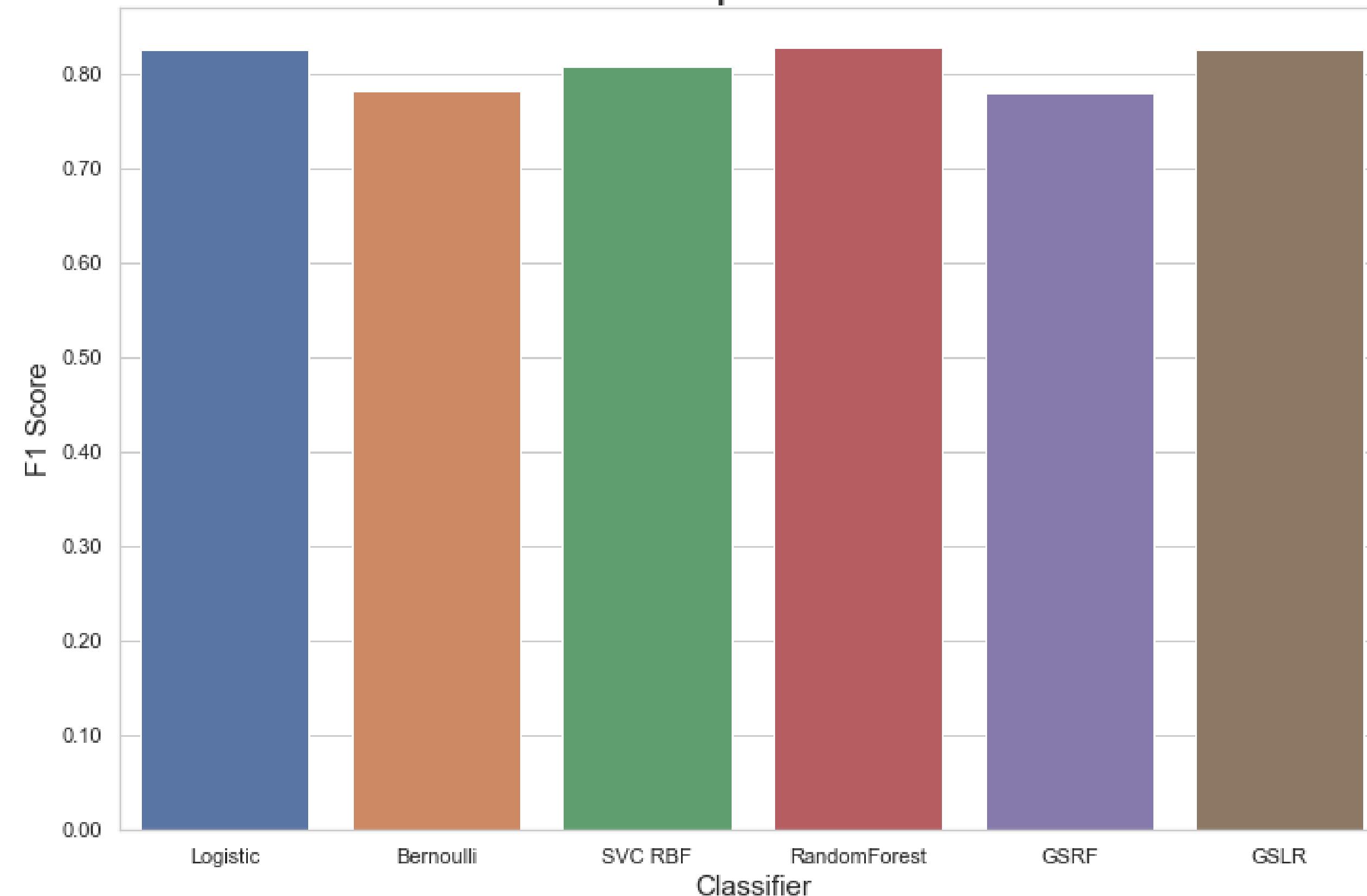
- Logistic Regression
- Bernoulli Naive Bayes
- Support Vector Classifier - Radial Basis Function
- Random Forest Classifier
- GridSearchCV with Random Forest Classifier
- GridSearchCV with Logistic Regression

Metric

F1 Score

Findings

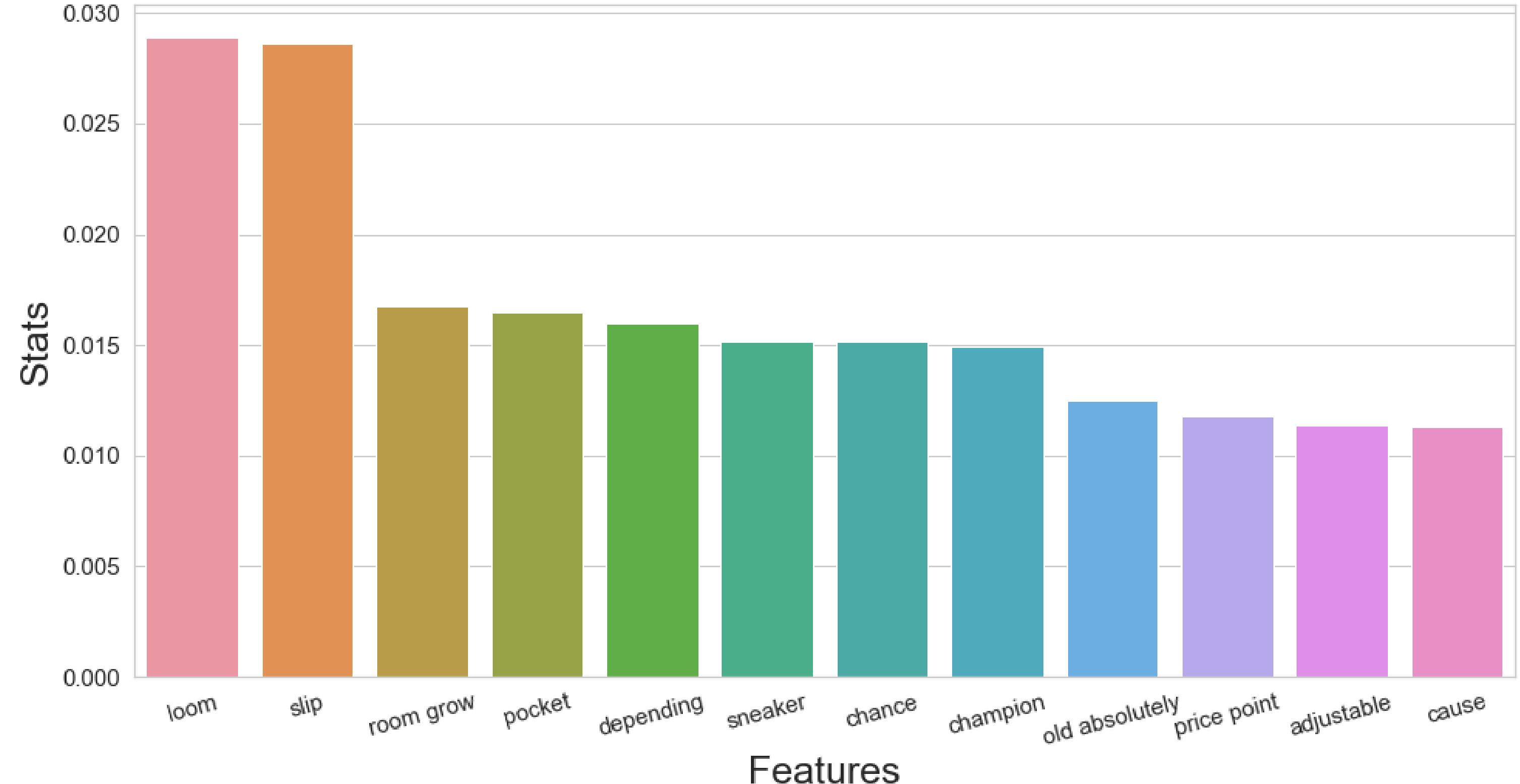
F1 Score per Classifier



Feature Importance

GridSearchCV with Random Forest

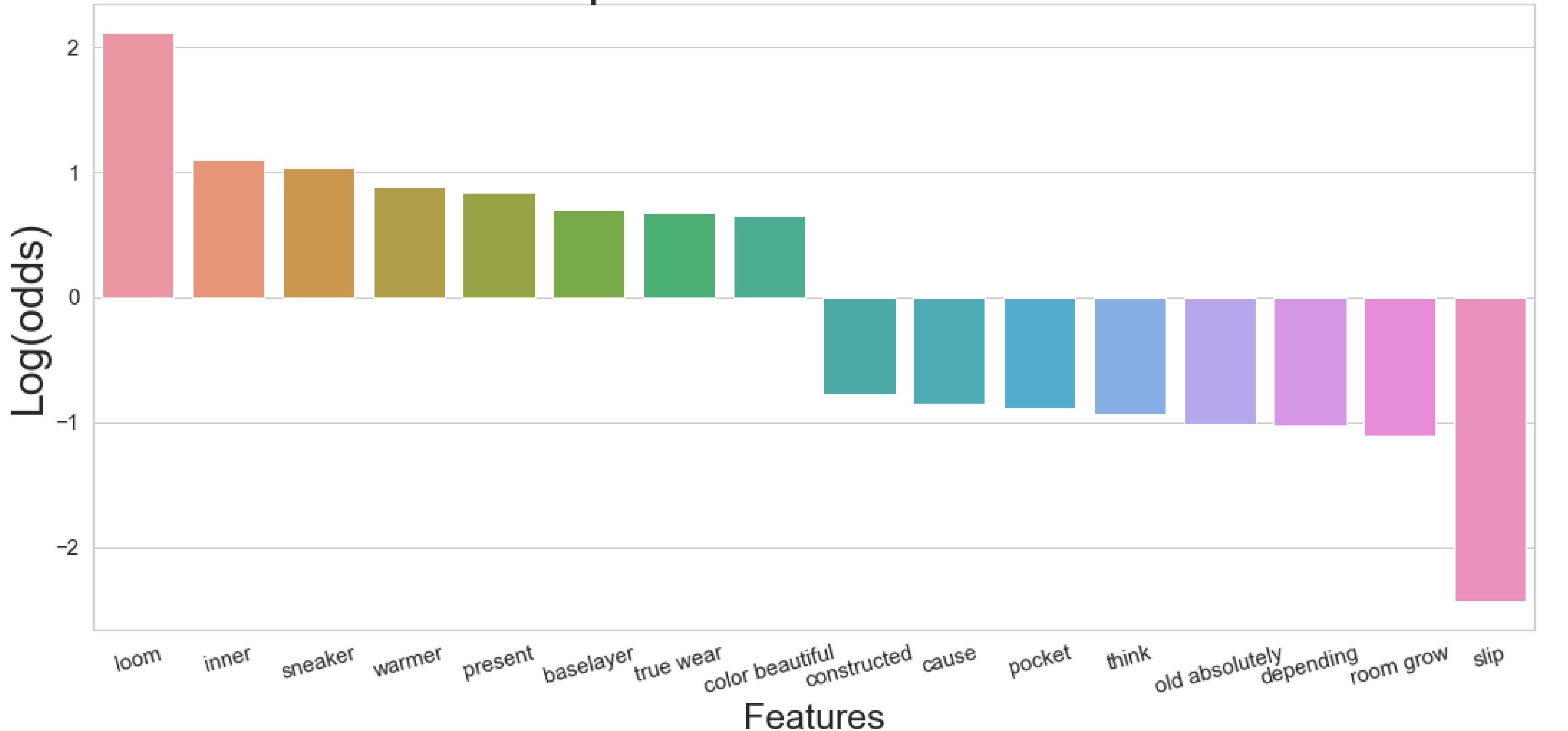
GridSearchCV with RandomForestClassifier



Feature Importance

Logistic Regression

Top and Bottom Features



Feature Importance

Feature Exploration

Loom

"Seams are starting to unravel after first use. Spend the extra money on Hanes brand or Fruit of a Loom..so disappointed..."

Slip

"Hubby also claims they slip down easily if not tied tightly."

Feature Importance

Feature Exploration

*"Did not fit my feet well. **Slipped** off easily"*

Slip

*"Hubby also claims they **slip** down easily if not tied tightly."*

*"These **slippers** run SO Small!"*

*"These are the best **slipper** socks I have ever owned!"*

*"I sadly had to return my ballet **slipper** pink because I ordered the wrong size."*

Recommendation

- **Customers want wider range of color**
- **Fit importance**

"My niece loved the bright pink color and it fit her perfectly!"

"My granddaughter loved the color. I would buy another if they had different colors."

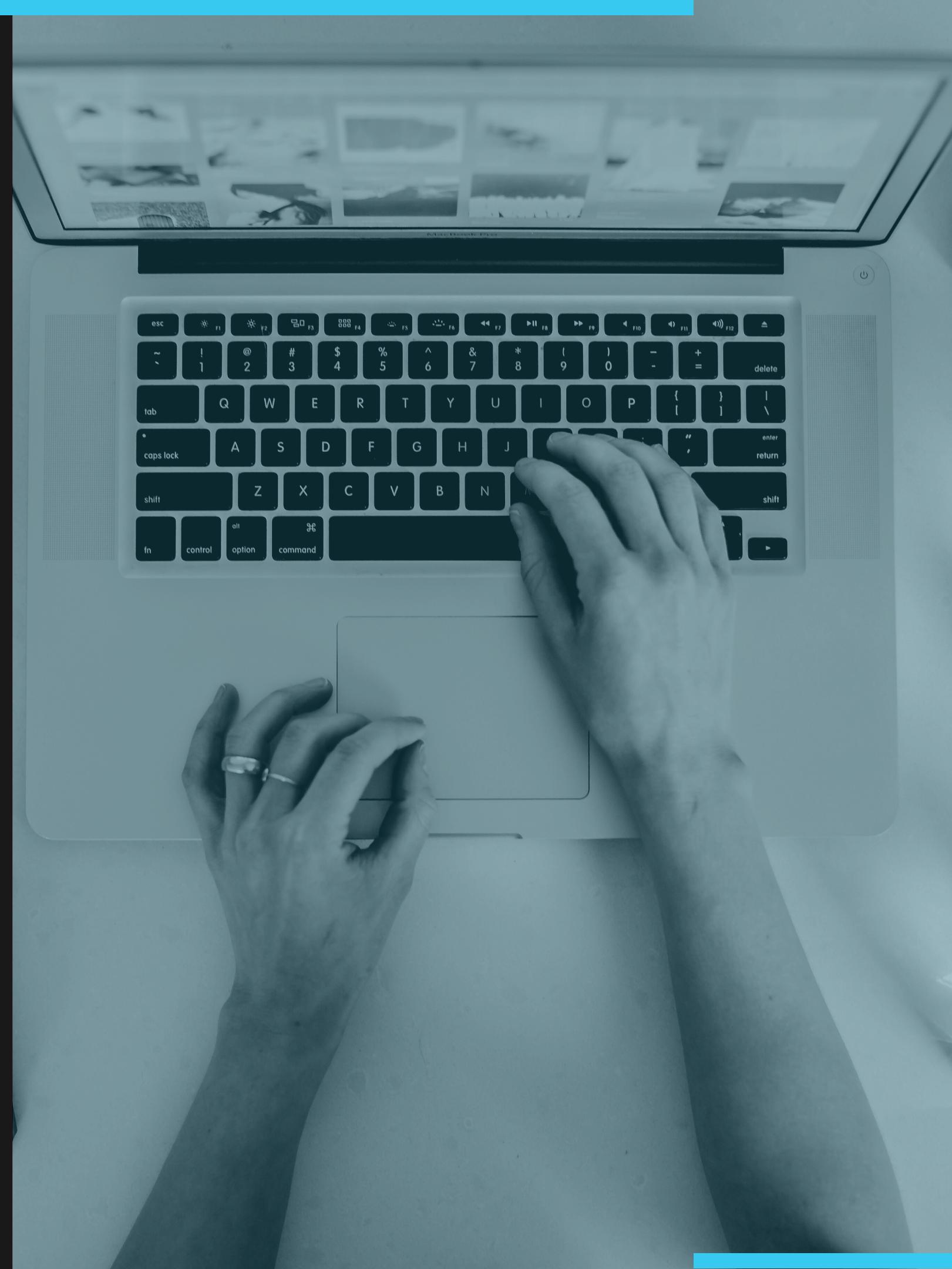
Quality Control

"When we opened the vacuum sealed packaging that this Mario suit came in, we were completely overwhelmed by a smell of mold and mildew."

"First, the sweater came smelling really bad like vinegar. Worst of all, the design came off after just one wash!"

Next Steps

- Lemmatization vs Stemming vs Nothing
- Compare brands
- Scraping reviews across seasons



Thank you

For further clarification or inquiries please
contact below:



dorjeys3@gmail.com



github.com/dorjeys3



twitter.com/dorjeys3

