



**Muzaffer Estelik**

**Prediction of Ratings  
Based on Women's E-Commerce Reviews**

**Data Science Career Track Capstone-2 Project**

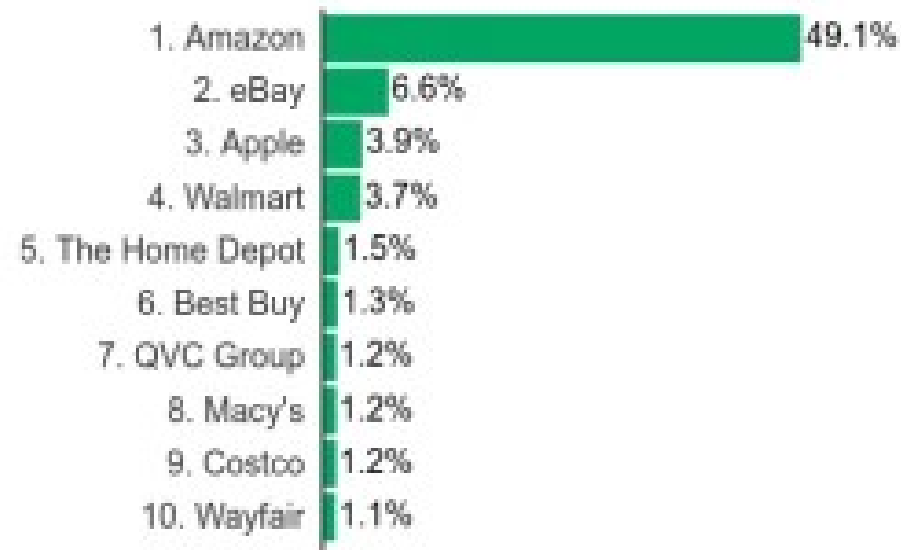
**Oct 29<sup>th</sup>, 2018**

# Table of Contents

- \* **Problem**
- \* **Data Set**
- \* **Exploratory Data Analysis (EDA)**
- \* **Machine Learning**
- \* **Conclusion**
- \* **Appendix**

# Problem

## Top 10 U.S. companies based on % of e-commerce sales



Source: eMarketer, July 2018



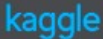

E-commerce is becoming more important day by day, together with all the aspects of it.


# Problem



- ✓ Promotion and discounts
- ✓ Content management
- ✓ Shopping cart and checkout process
- ✓ Search engine optimized code and layout
- ✓ Reporting tools
- ✓ Email marketing


# Data Set

 Search kaggle  Competitions Datasets Kernels Discussion Learn ... [Sign In](#)

 Dataset

## Women's E-Commerce Clothing Reviews

23,000 Customer Reviews and Ratings


 Nick Brooks • updated 9 months ago (Version 1)

188 voters


share

[Data](#) [Overview](#) [Kernels](#) [Discussion](#) [Activity](#) [Download \(3 MB\)](#) [New Kernel](#)

Data (3 MB)

[API](#) `kaggle datasets download -d nicapotato/womens-ec...` [Download All](#) 


### Data Sources

	Womens Clothing E-... 23.5k x 11
---	----------------------------------

### About this file

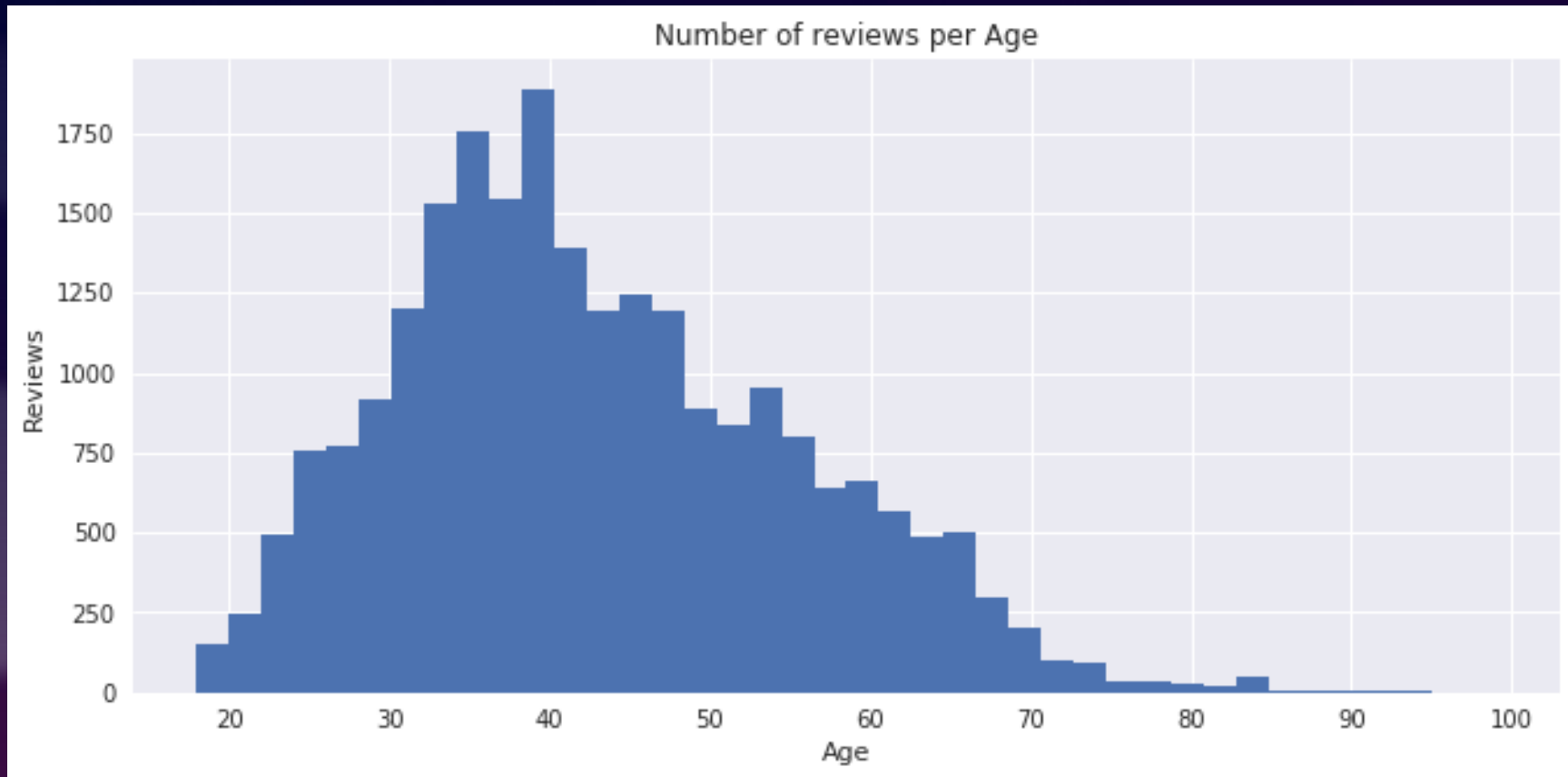
This is a Women's Clothing E-Commerce dataset revolving around the reviews written by customers

### Columns

-  **Clothing ID** Unique ID of the product
- # Age** Age of the reviewer
- Δ Title** Title of the review
- Δ Review Text** review
- # Rating** Product rating by reviewer
- ✓ Recommended IND** Whether the product is recommended or not by the reviewer

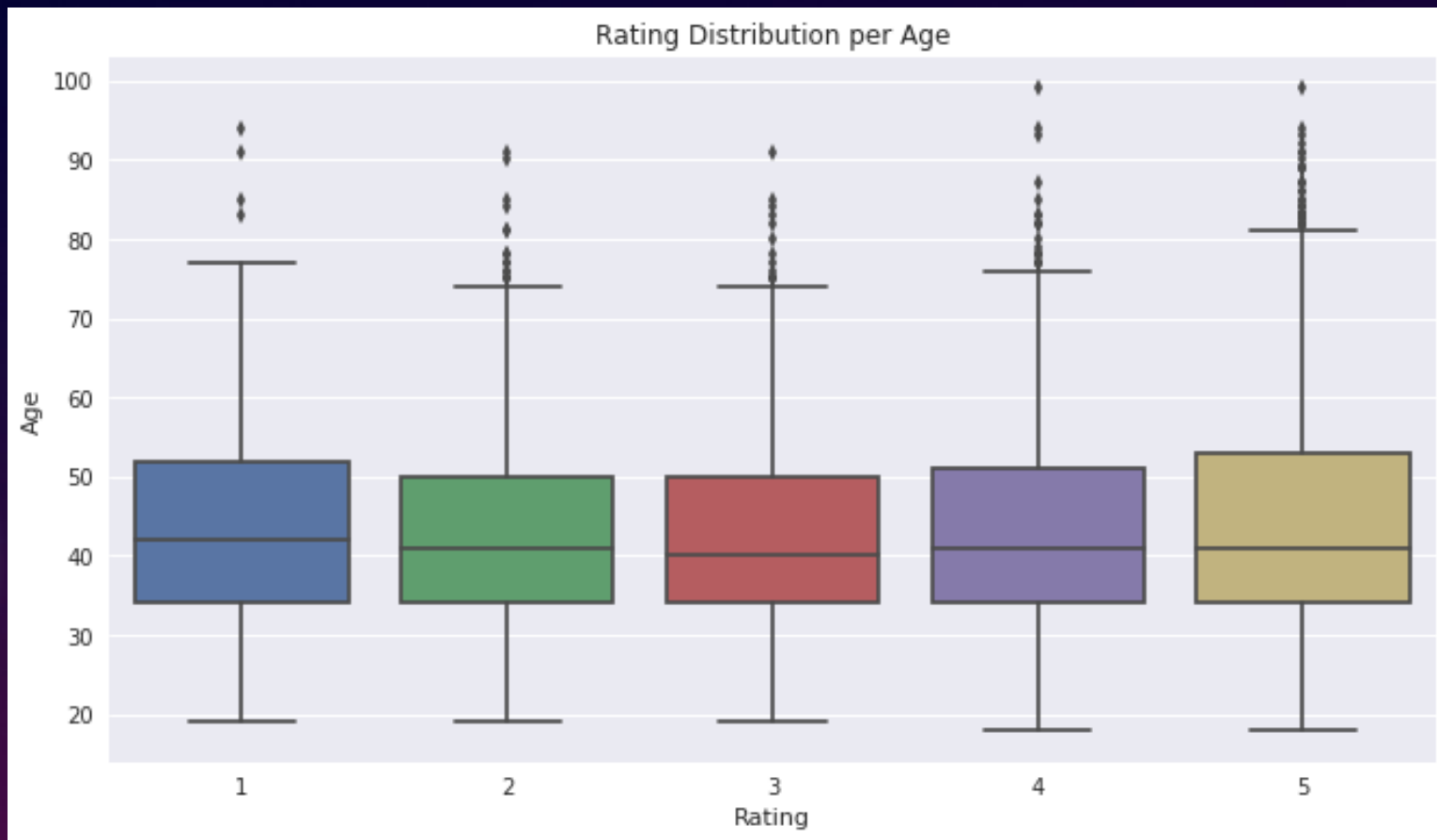
<https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>

# Exploratory Data Analysis (EDA)



**25-50 is the most reviewing age group**

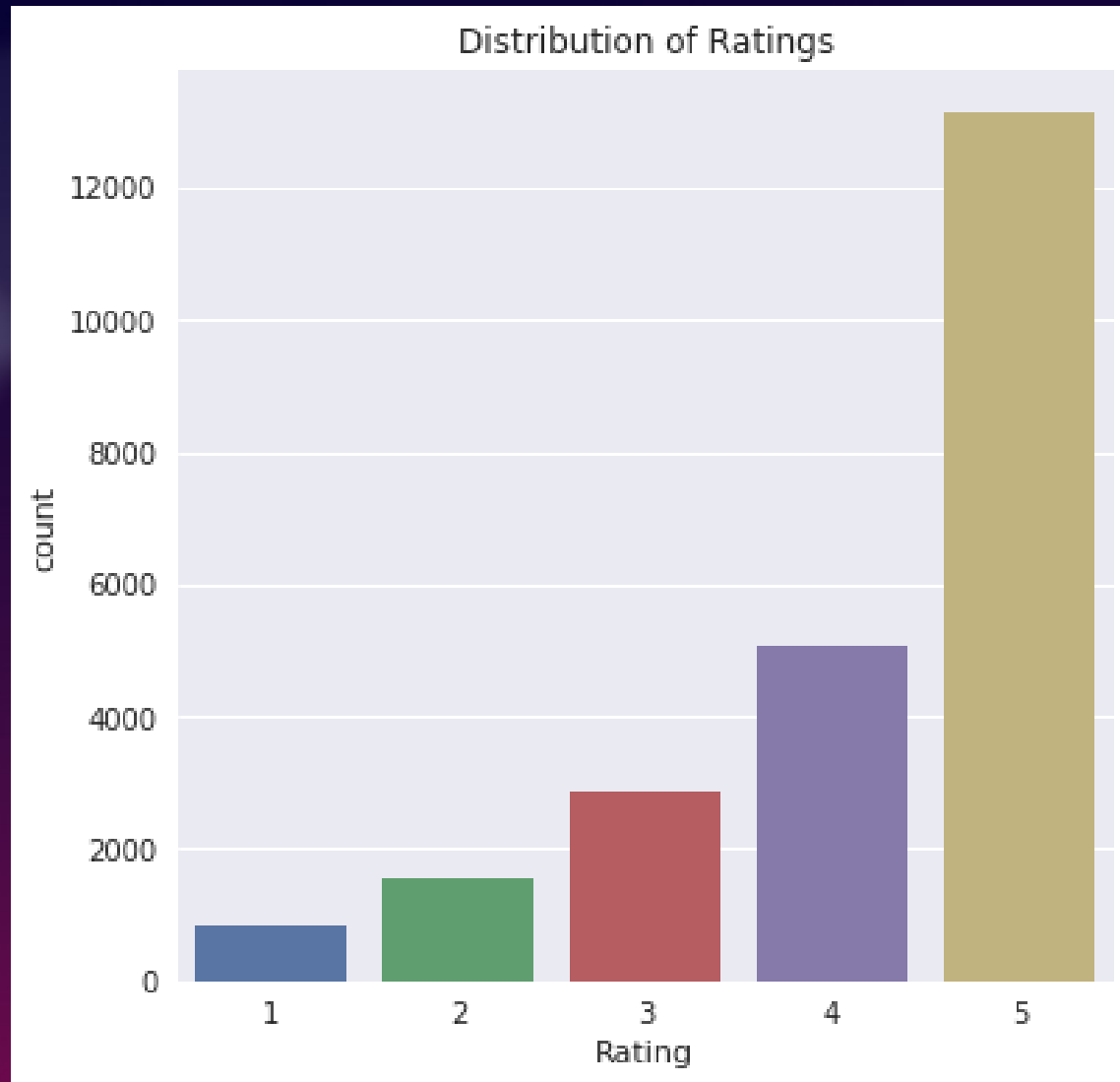
# Exploratory Data Analysis (EDA)



age is not a factor on the ratings of the reviews

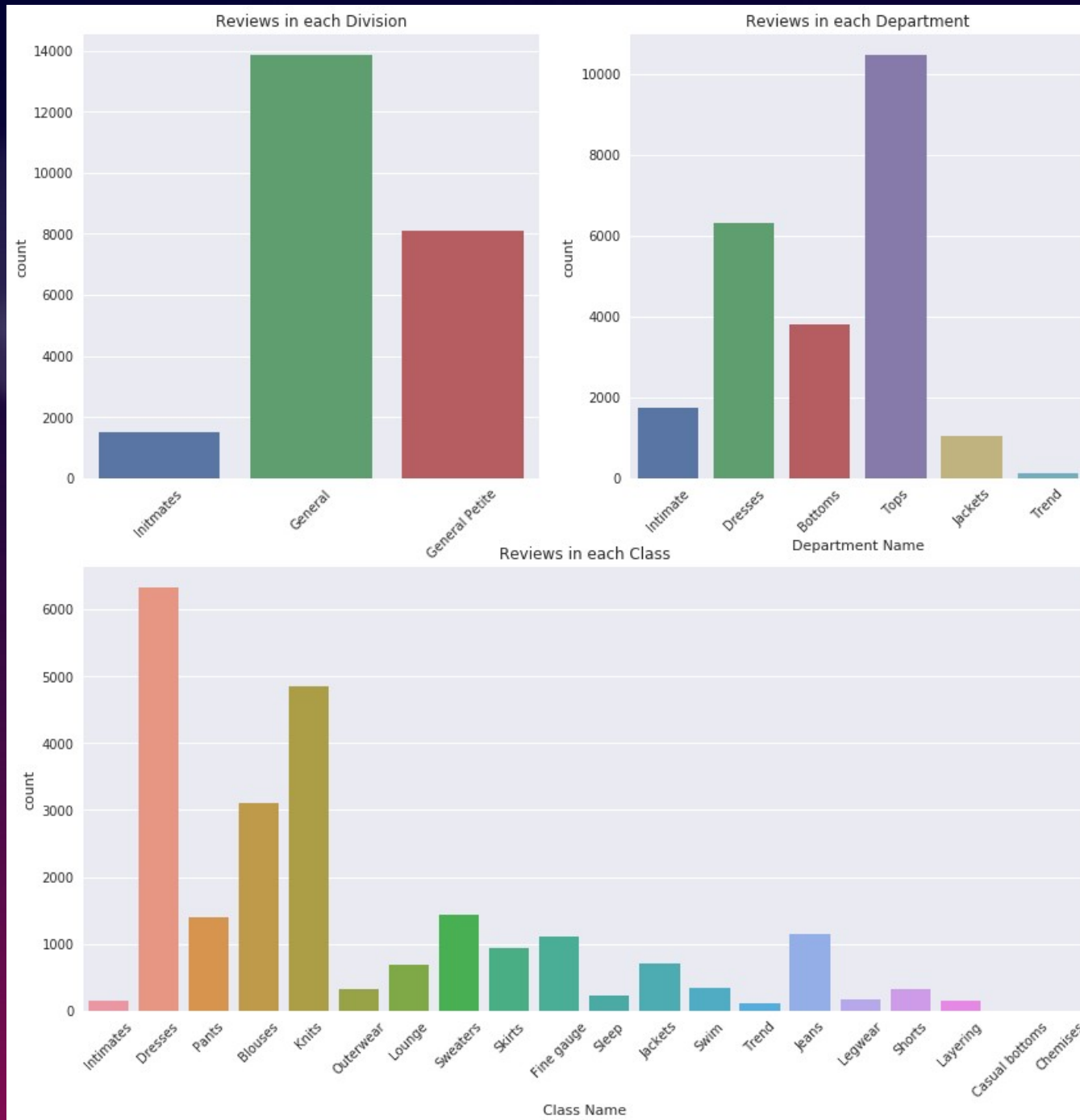


# Exploratory Data Analysis (EDA)

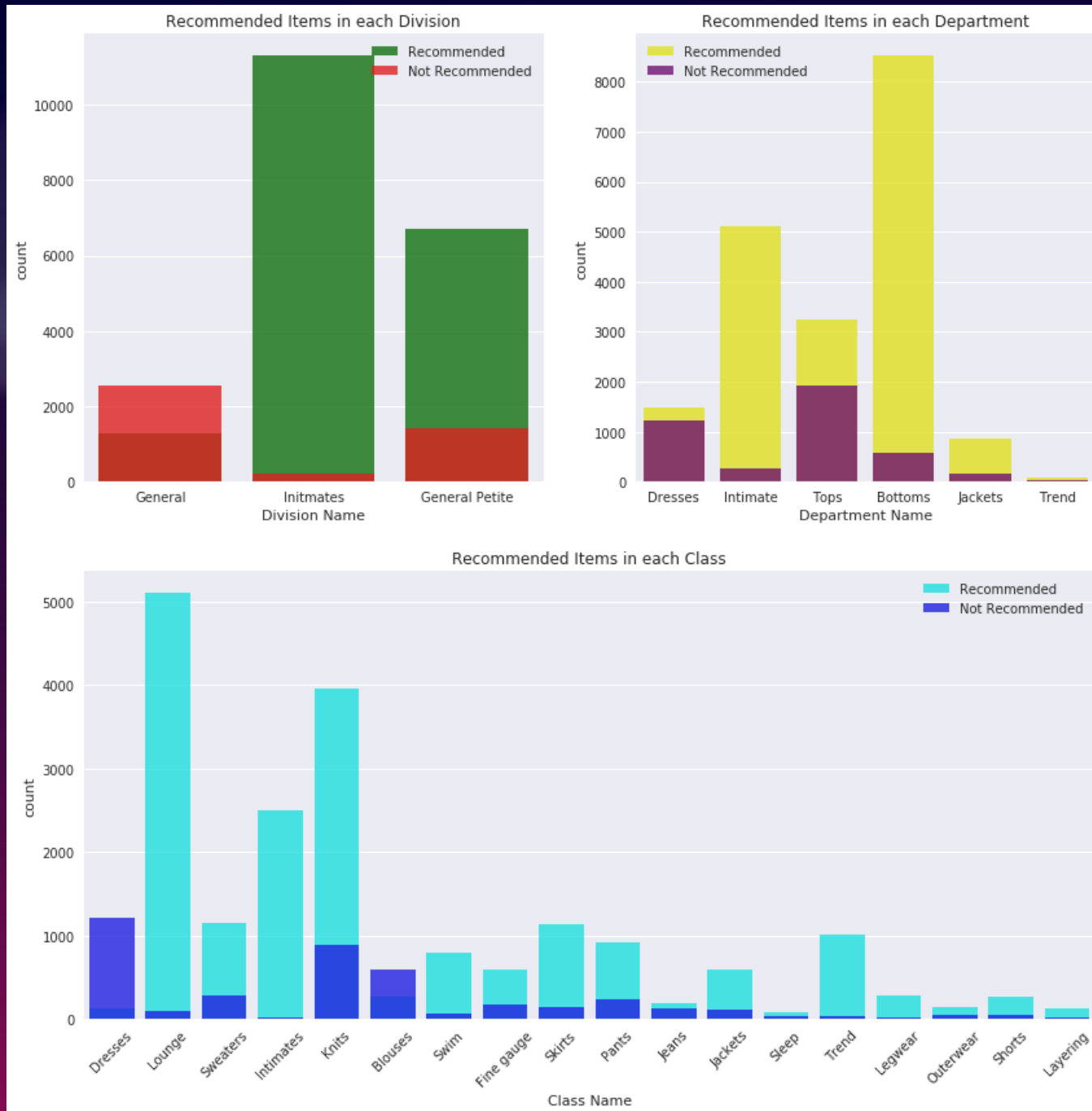




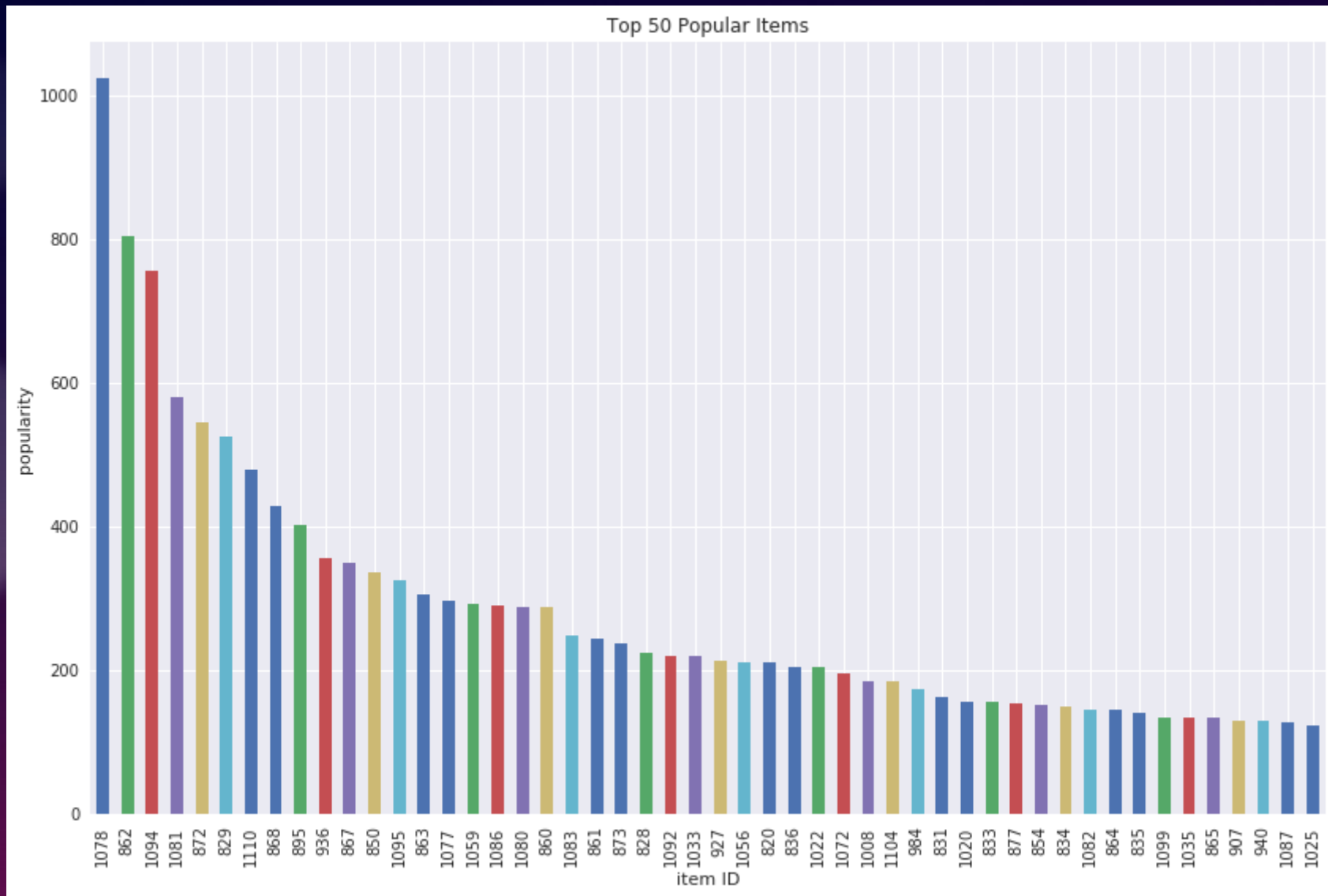
# Exploratory Data Analysis (EDA)



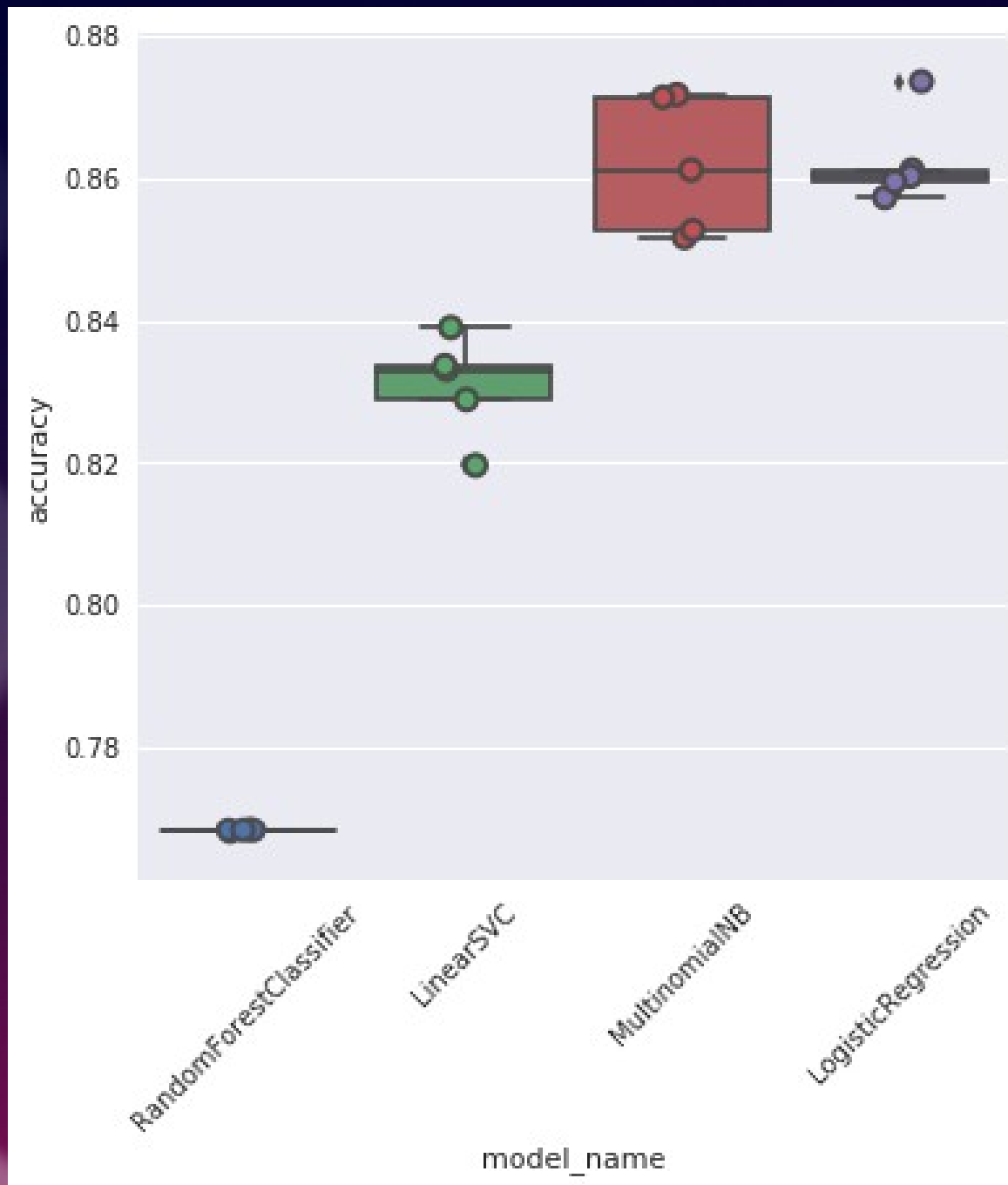
# Exploratory Data Analysis (EDA)



# Exploratory Data Analysis (EDA)



# Machine Learning



TPOT

RNN

# Conclusion

Model Name	Train Accuracy Score	Test Accuracy Score
Linear SVC	0.830866	0.832948
Naive Bayes	0.861688	0.859608
Logistic Regression	0.862371	0.845893
Random Forest	0.768390	0.774387