

# **Analyzing women's E-commerce clothing reviews**

By: Ruqaiah Abdullah  
Rawan Alshehri  
Alanoud Alzahrani





# Outline

- -Introduction
- -Dataset
- -Mythology
- -Results and Analysis
- -Conclusion





# Introduction

in this project, we'll be using NLP with Topic Modeling and Clustering to find out some good information about customers reviews to take advantage of e-commerce

answering some questions:

- What is the best selling clothing?
- What is the most type of clothes that has more positive reviews?
- At which age womens buy clothes most?

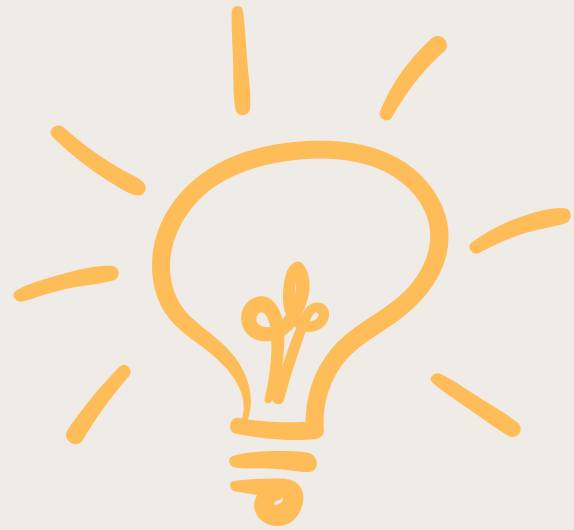


# Dataset

- It is a Women's Clothing E-Commerce dataset revolving around the reviews written by customer
- It has 23,486 columns and 11 rows
- **Features:** Clothing ID, Age, Review Text, Title, Rating, Recommended IND, Positive Feedback Count, Division Name, Department Name, Class Nam



# Mythology



**Gathering  
Data**

**Text  
Preprocessing**



**EDA**

**Topic  
Modeling**



# Text Preprocessing

## Clean Text

**Remove:** punctuation,  
lowercasing, non-  
alphabetic, stop words

## Tokenize

Count Vector  
TFIDF

## Split Data

80% Train  
20% Test



# EDA

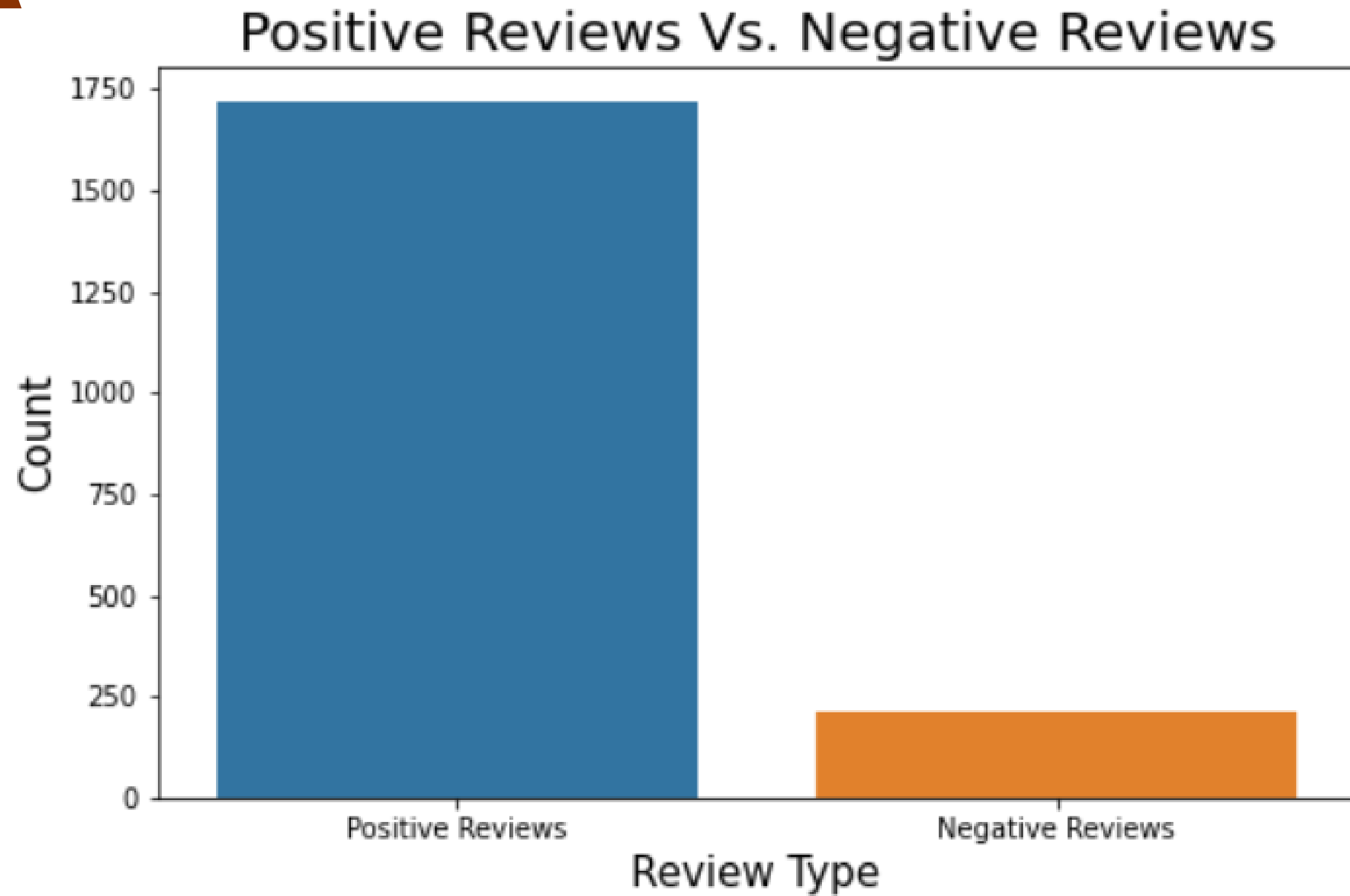
## Positive Words



## Negative Words



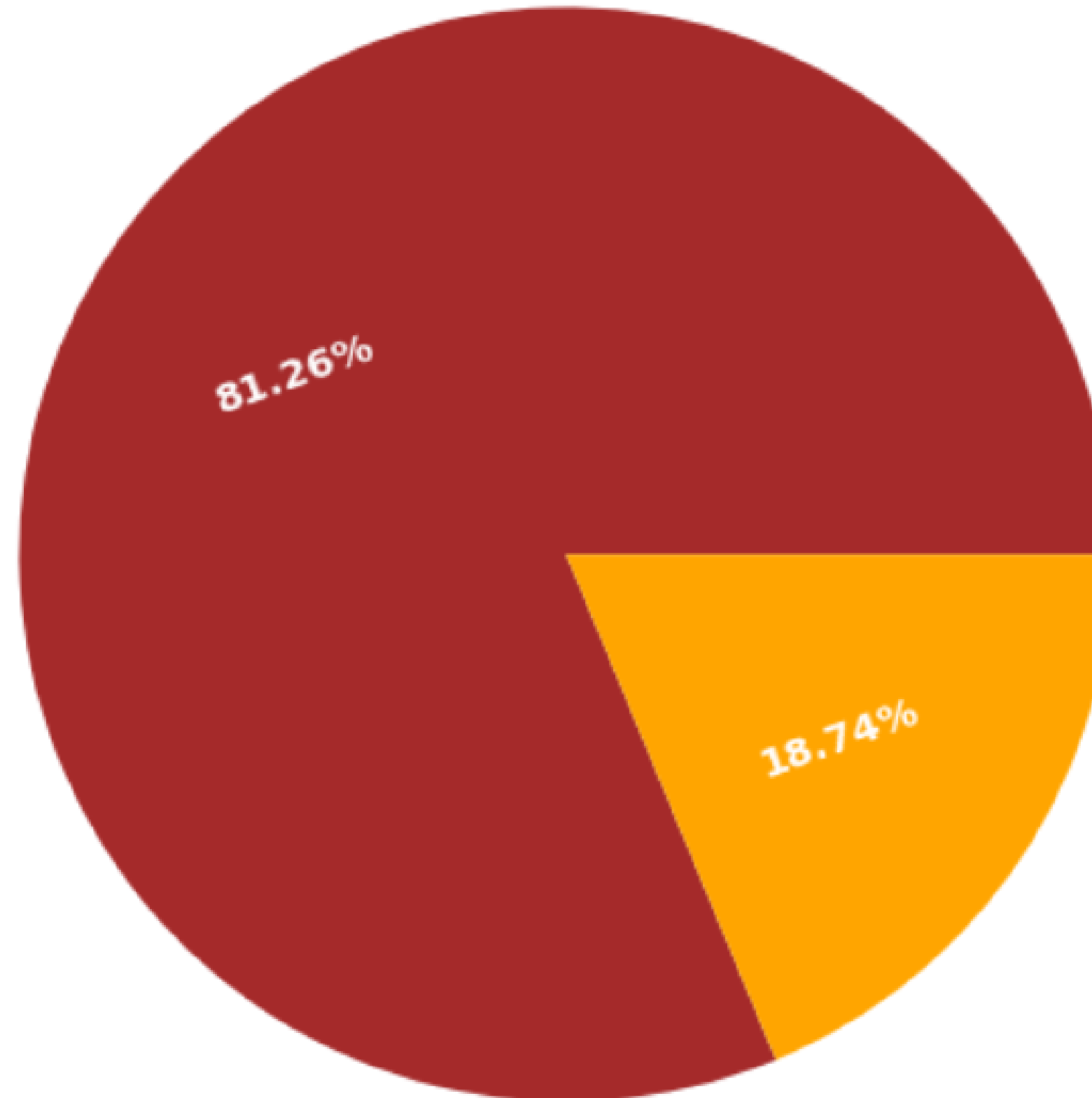
# EDA





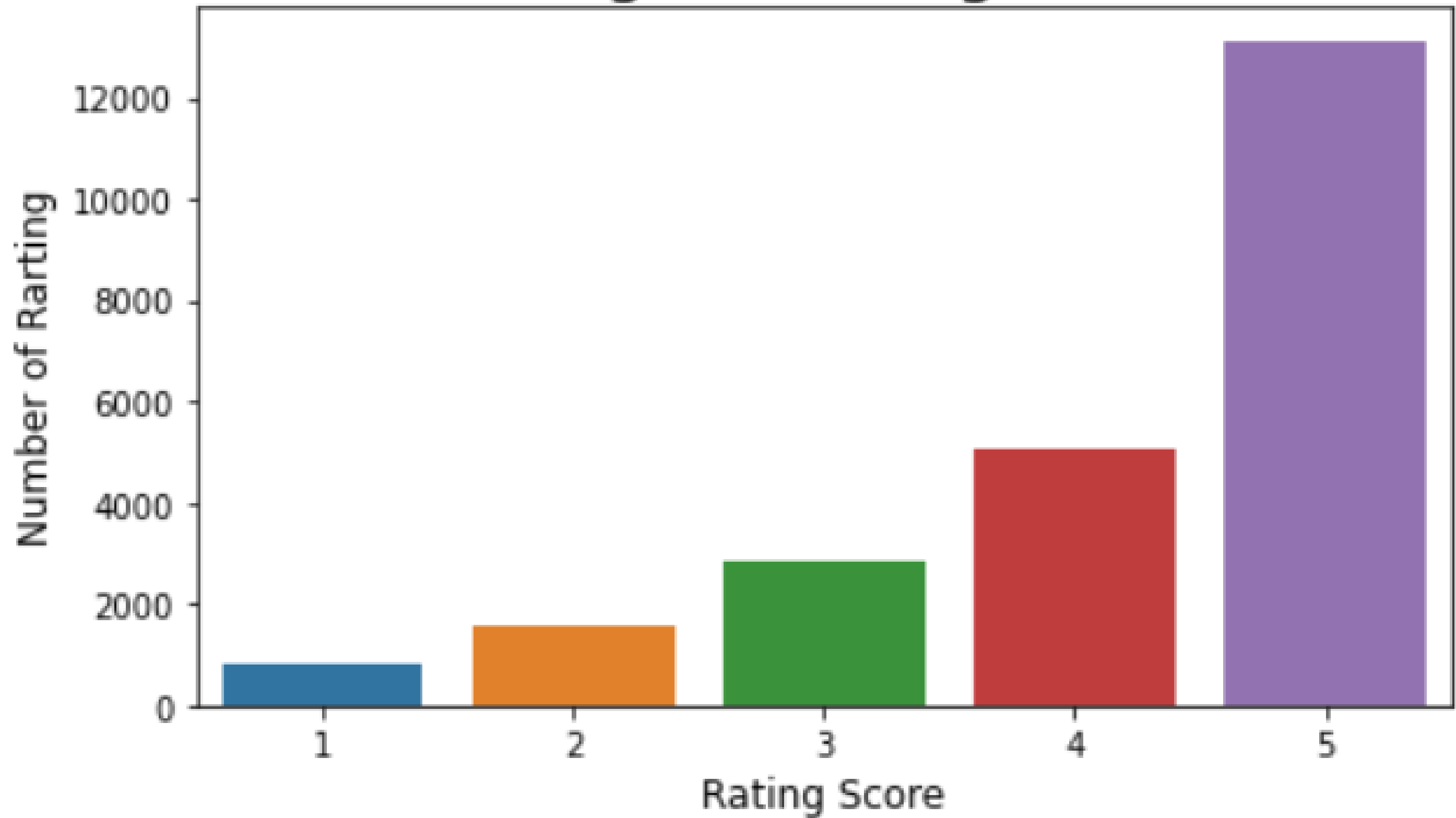
# EDA

## Recommended Products Vs. Not Recommended Products



# EDA

The highest rating scores





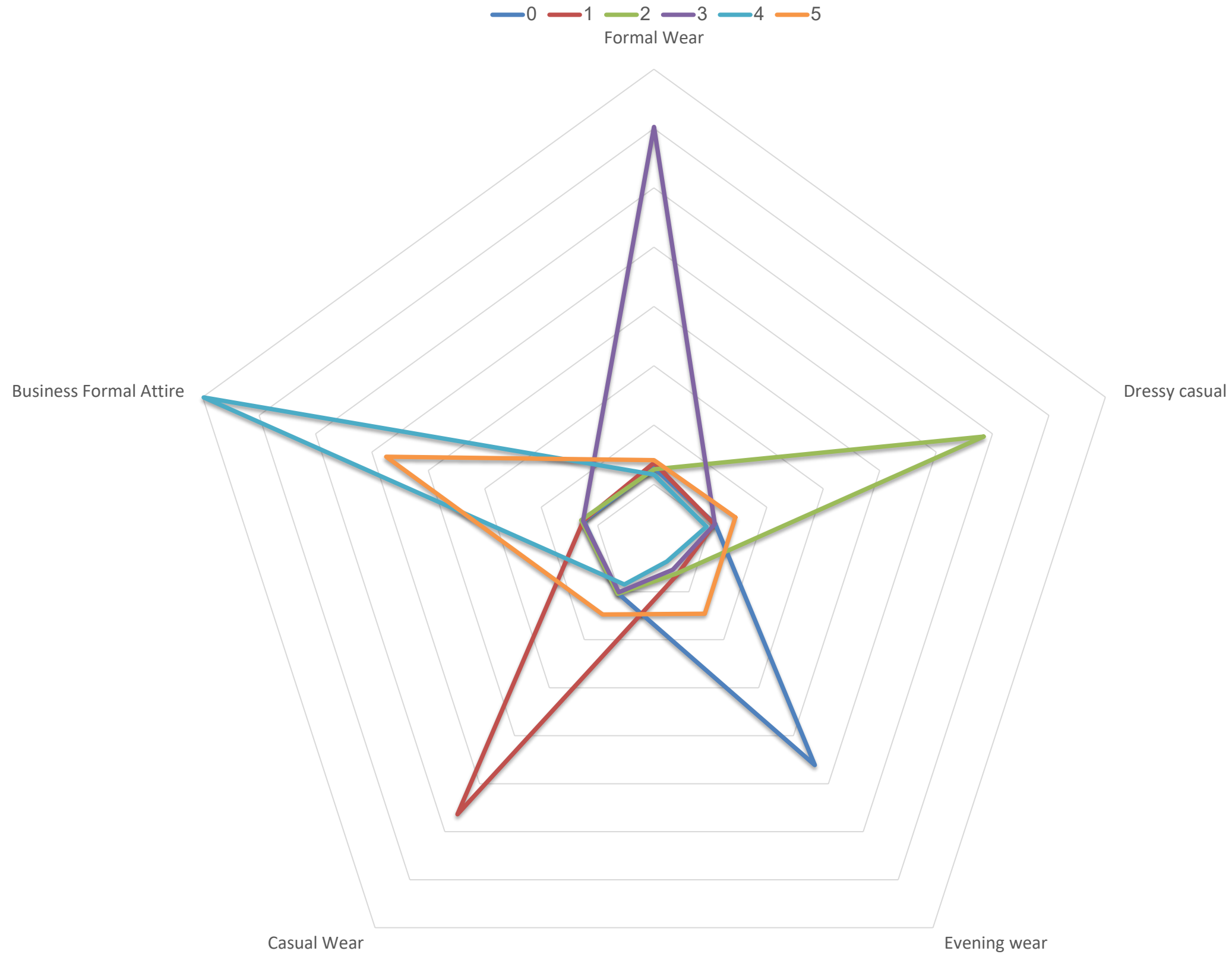
# Topic Modeling



- **LSA Model**
- **NMF Model**
- **LDA Model**

# Results and Analysis

## clustering





# Conclusion

- Best model LDA
- Based on the results, women usually prefer Business Formal Attire, Casual Wear, Followed by Formal Wear On this basis, we prefer to choose these types of clothes to start our project.
- future work ...

