# Import necessary libraries

import numpy as np

import pandas as pd

import nltk

import random

import os

from textblob import TextBlob

import matplotlib.pyplot as plt

# Read file

e\_data = pd.read\_csv("Womens Clothing E-Commerce Reviews.csv")

# Printing e\_data

print(e\_data.shape)

e\_data.head()

# null value elimination

e\_data.info()

e\_data.nunique()

column\_data\_types = e\_data.dtypes

column\_data\_types

null\_counts\_per\_column = e\_data.isnull().sum()

# count of null values in data

null\_counts\_per\_column

e\_data.drop(e\_data.columns[0], inplace=True, axis=1)

# eliminating unnneccessary first column in data

null\_counts\_per\_column = e\_data.isnull().sum()

null\_counts\_per\_column

# Eliminating Values

for x in ["Title", "Division Name", "Department Name", "Class Name", "Review Text"]:

e\_data = e\_data[e\_data[x].notnull()]

e\_data.isnull().sum()

duplicates = e\_data.duplicated()

print(duplicates)

data\_no\_duplicates = e\_data.drop\_duplicates()

print(data\_no\_duplicates)

summary\_statistics = e\_data.describe().T.drop('count', axis=1)

summary\_statistics

e\_data['Clothing ID'].unique()

e\_data['Clothing ID'].value\_counts()

popular\_products = pd.DataFrame(e\_data.groupby('Clothing ID')['Rating'].count())

most\_popular = popular\_products.sort\_values('Rating', ascending=False)

top\_10\_popular\_products = most\_popular.head(10)

# sample evaluation for data changes

top\_10\_popular\_products

most\_popular.head(30).plot(kind = "bar")

least\_popular = popular\_products.sort\_values('Rating', ascending=True)

least\_popular.head(30).plot(kind = "bar")

# Assumption for Review Text Column

# Polarity and Subjectivity of Data

def get\_sentiment(text):

analysis = TextBlob(text)

polarity = analysis.sentiment.polarity

subjectivity = analysis.sentiment.subjectivity

return pd.Series([polarity, subjectivity])

# Setiments in Review Text

e\_data[['Polarity', 'Subjectivity']] = e\_data['Review Text'].apply(get\_sentiment)

# Printing of Polarity and Subjectivity

print(e\_data[['Review Text', 'Polarity', 'Subjectivity']])

def get\_sentiment\_label(score):

if score > 0:

return 'Positive'

elif score < 0:

return 'Negative'

else:

return 'Neutral'

e\_data['Sentiment Label'] = e\_data['Polarity'].apply(get\_sentiment\_label)

sentiment\_counts = e\_data['Sentiment Label'].value\_counts()

print(sentiment\_counts)

# Display of Sentimental Model

plt.figure(figsize=(8, 6))

sentiment\_counts.plot(kind='bar', color='orange')

plt.title('Sentiment Analysis Of User Reviews')

plt.xlabel('Sentiment Label')

plt.ylabel('Count')

plt.grid(axis='y')

plt.xticks(rotation=45)

plt.show()

e\_data['Sentiment Label'] = e\_data['Subjectivity'].apply(get\_sentiment\_label)

sentiment\_counts = e\_data['Sentiment Label'].value\_counts()

print(sentiment\_counts)

file\_path = 'cleaned\_data.csv'

# Saving the Data Cleaned

e\_data.to\_csv(file\_path, index=False)

sp\_data = pd.read\_csv("cleaned\_data.csv")

sp\_data.head()