Exploring Zara's Innovative Marketing and Supply Chain Strategies to Drive Sales. Identification of Deviations in sales and supply chain based on Customer reviews and sentiment analysis.

Emails and affiliation of the authors:

1) Rishi Krishna Thodupunuri – rthod1@unh.newhaven.edu
2) Nikhil Teja Tangella – ntang2@unh.newhaven.edu
3) Sai Teja Gattu - sgatt4@unh.newhaven.edu

ABSTRACT:

This paper aims to analyse Zara's sales performance and explore the use of market prediction techniques to forecast future trends. By leveraging data analytics and advanced prediction models, this study provides insights into Zara's sales patterns and offers recommendations for improving forecasting accuracy. The findings contribute to enhancing Zara's decision-making processes and optimizing its sales strategies.

Keywords: Supply chain, Efficient supply chain, ZARA, Decision tree, KNN, gradient bosting, random forest, multinomial NB.

INTRODUCTION

The introduction section provides an overview of Zara as a leading fashion retailer and highlights the importance of sales analysis and market prediction in the fast-paced fashion industry. It outlines the objectives of the study and introduces the subsequent sections of the paper. Zara's supply chain and highlights its significance in the fast fashion industry. It discusses the challenges faced by Zara and the need for optimizing its supply chain to maintain a competitive edge. The section outlines the objectives of the study and introduces the subsequent sections of the paper.

Zara, a Spanish clothing brand, is known for its fast fashion and innovative marketing and supply chain strategies. The brand has gained popularity among customers by providing trendy and affordable clothing items with a quick turnaround time. Zara's success can be attributed to its ability to anticipate and respond quickly to changing fashion trends, coupled with its effective marketing and supply chain strategies. we will explore Zara's innovative marketing and supply chain strategies that have helped the brand drive sales. Additionally, we will identify deviations in sales and supply chain based on customer reviews and sentiment analysis. Finally, we will analyse customer reviews and sentiment analysis to identify deviations in Zara's sales and supply chain.

RELATED REVIEWS:

1.)

Title of the paper: Analysis on the Successful Case of Efficient Supply Chain in ZARA

Authors: Qinghua Zhang

Affiliation: Shanghai University of Finance & Economics, Shanghai, P.R.China, 200433

Publication date: 18 November 2008

Name of publisher: IEEE

2.)

Title of the paper: WhatandHow CanWeLearnfromZARA

Authors: QinghuaZhang

Affiliation: Shanghai University of Finance & Economics, Shanghai, P.R.China, 200433

Publication date: 25 November 2008

Name of publisher: IEEE

3.)

Title of the paper: Speed-to-Fashion:Managing Global Supply Chain in Zara

Authors: Andrea Carugati, Raffael Liao, Pernille Smith

Affiliation: Department of Business Studies, Aarhus School of Business, University of Aarhus, Denmark

Publication date: 21 October 2008

Name of publisher: IEEE

4.)

Title of the paper: Research on ZARA Strategy from the Perspective of SWOT Analysis Method

Authors: Hu Duoyan

Affiliation: Jiangxi Normal University

Publication date: 2021

Name of publisher: Atlantis Press SARL

5.)

Title of the paper: Modelling financial flow of the supply chain

Authors: M. H Jahangiri, F. Cecelja

Affiliation: University of Surrey

Publication date: 2014

Name of publisher: IEEE

Selected Dataset:

The dataset is called the "Zara US Fashion Products Dataset" and it contains information about products sold by the Zara US stores. Here are some details about the dataset:

- The dataset includes information on products sold by Zara US in the year 2017.
- Each product is represented by a unique ID(Product ID) number and includes information such as the product name, product size, Product Category, sizes, Colors, price, state age and Date of sale.
- The dataset also includes information about the category and subcategory of each product, as well as its color and composition.
- In addition to product information, the dataset also includes information about Zara US store locations state wise.
- This dataset is created and developed by our team using multiple resources on internet, which are mentioned in references.

Proposed Methods:

Pre-processing: Converting the raw data into use full data, To create any data model we need to pre-process the un-useful data.

- ➔ Filtering: Selecting a subset of the data based on certain criteria, such as state or date range.
- ➔ Aggregation: Grouping data by one or more variables to get a summary of the data.
- ➔ Outlier detection: Identifying and analyzing values that are significantly different from the rest of the data.

Data cleaning: Removing unnecessary data and structuring the remaining data into columns.

- ➔ If there are any Null values and empty cells we removed those Null values using ‘NOT NULL’ Function.

Descriptive statistics: Analyzing the dataset to get a general sense of the data, such as mean, median, mode, standard deviation, etc.

- ➔ We used methods to find overall sales date wise:
- ➔ ‘DATE TIME’ function
- ➔ example- strptime(x,'%m/%d/%Y')) : converts the string value to exact date
- ➔ strftime('%b'), converts the numeric date(month) to string date(month).

Decision Tree:
Decision Trees are a non-parametric method used for classification and regression tasks. The goal is to split the data into groups based on the values of the input variables and make

predictions based on the group that a given input belongs to. Decision Trees can handle nonlinear relationships between the variables and can be easily visualized and interpreted.

Random Forest:
Random Forest is an ensemble method that builds multiple Decision Trees and combines their predictions to make the final prediction. Each tree is built on a random subset of the input variables and a random subset of the training data. Random Forest can handle a large number of input variables and can capture complex nonlinear relationships between the input and output variables.

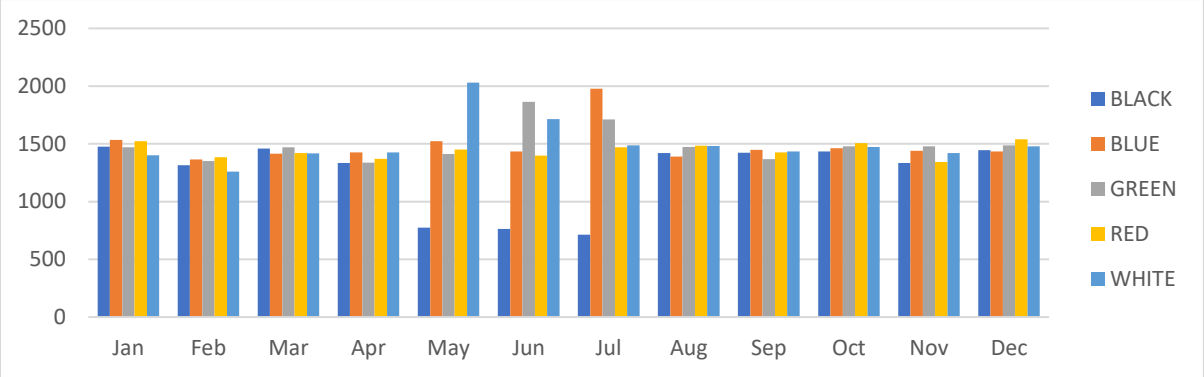
Gradient Boosting:
Gradient Boosting is another ensemble method that builds an additive model by sequentially adding weak learners to the ensemble. Each new learner tries to correct the errors made by the previous learners. Gradient Boosting can also capture complex nonlinear relationships and has been shown to perform well on a wide range of datasets.

K-Nearest Neighbors (KNN):
K-Nearest Neighbors (KNN) is a non-parametric method that makes predictions based on the k nearest training examples in the feature space. The value of k is a hyperparameter that needs to be set. KNN can work well on small datasets and can capture local patterns in the data.

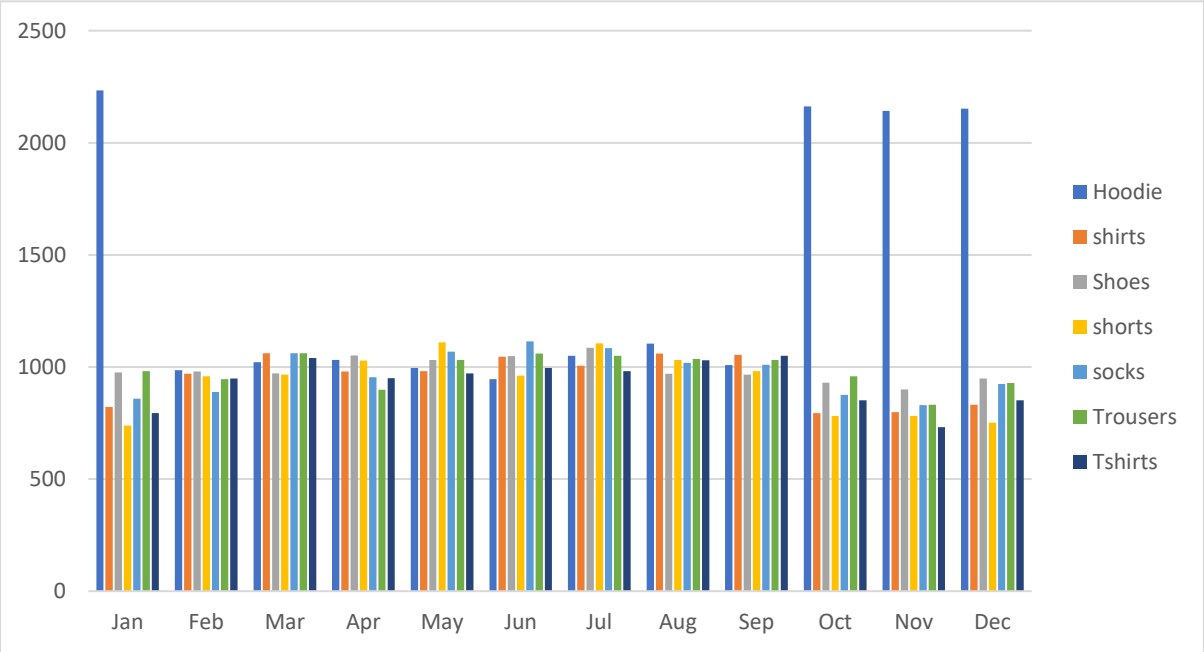
Multinomial Naive Bayes (Multinomial NB):
Multinomial Naive Bayes (Multinomial NB) is a probabilistic classification algorithm that is commonly used for text classification tasks, such as sentiment analysis, spam filtering, and topic classification. It is a variant of the Naive Bayes algorithm, which is based on Bayes' theorem of probability.

EXPERIMENTAL RESULTS: -

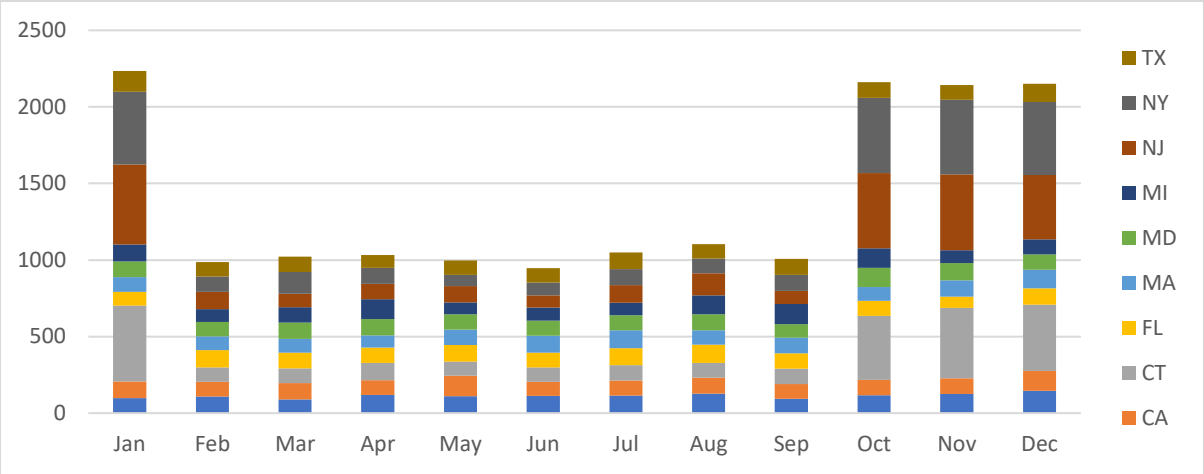
- ➔ Below Bar graph represents the relation between color and monthly sales of the Zara dataset.
- ➔ This graph is generated using Matplotlib functions.
- ➔ X-axis represents Month and Y-axis represents number of sales.
- ➔ From this graph we can conclude that “BLACK” color products has less sales in months of MAY, June and July.



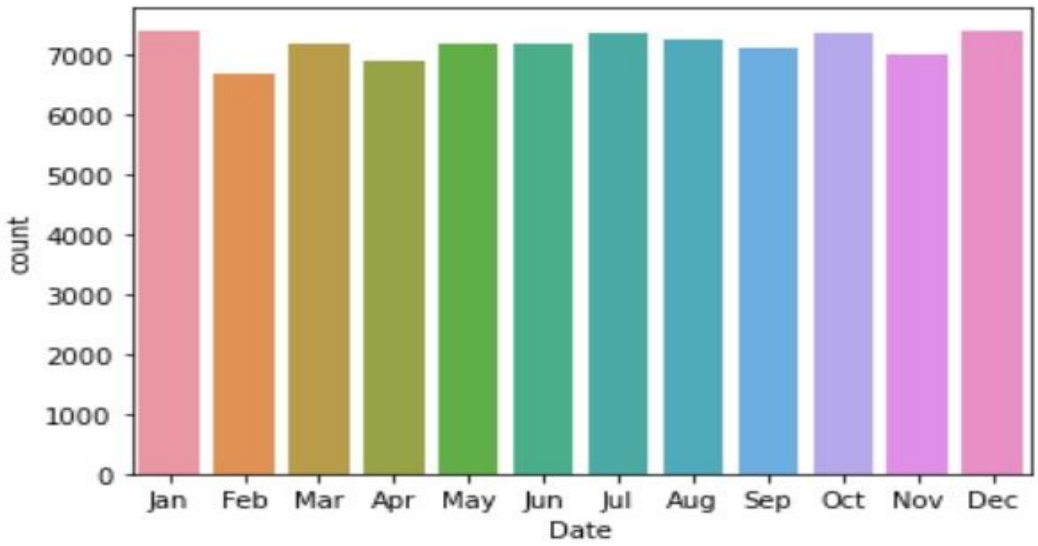
- ➔ Below Bar graph represents the relation between product- Hoodie and monthly sales of the Zara dataset.
- ➔ This graph is generated using Matplotlib functions.
- ➔ X-axis represents Month and Y-axis represents number of sales for hoodies.
- ➔ From this graph we can conclude that “Hoodies” products has more sales in months of Jan, Oct, Nov and Dec.



- ➔ Below Bar graph represents the relation between product- Hoodie, monthly sales, and state wise of the Zara dataset.
- ➔ This graph is generated using Matplotlib functions.
- ➔ X-axis represents Month and Y-axis represents number of sales for hoodies.
- ➔ From this graph we can conclude that “Hoodies” products has more sales in states of NY, NJ and CT.

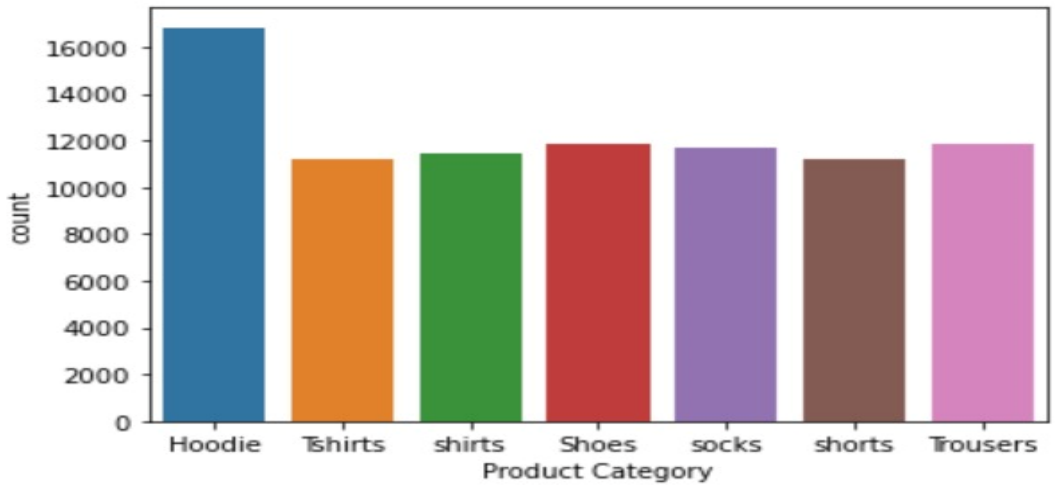


- ➔ Below Bar graph represents the relation between total sales and monthly wise of the Zara dataset.
- ➔ This graph is generated using seaborn functions.
- ➔ X-axis represents Month and Y-axis represents number of sales of all products.
- ➔ From this graph we can get total sales for every month in a year

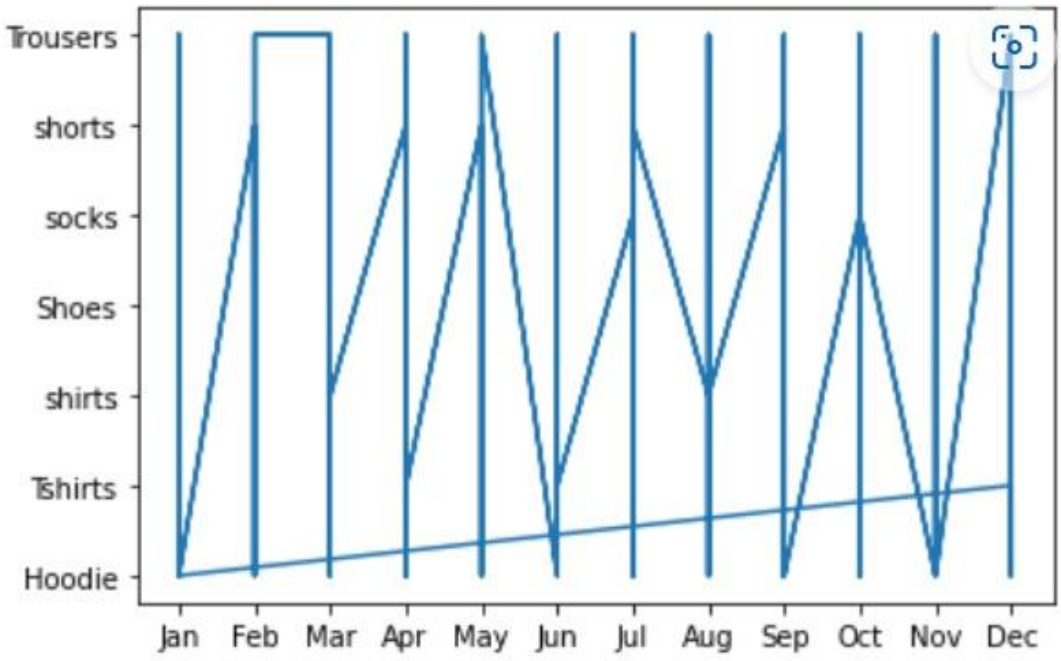


- ➔ Below Bar graph represents the relation between total sales and product category wise of the Zara dataset.
- ➔ This graph is generated using seaborn functions.

➔ X-axis represents Product category and Y-axis represents number of sales(count) of all products.
➔ From this graph we can get total sales of all products based on category.

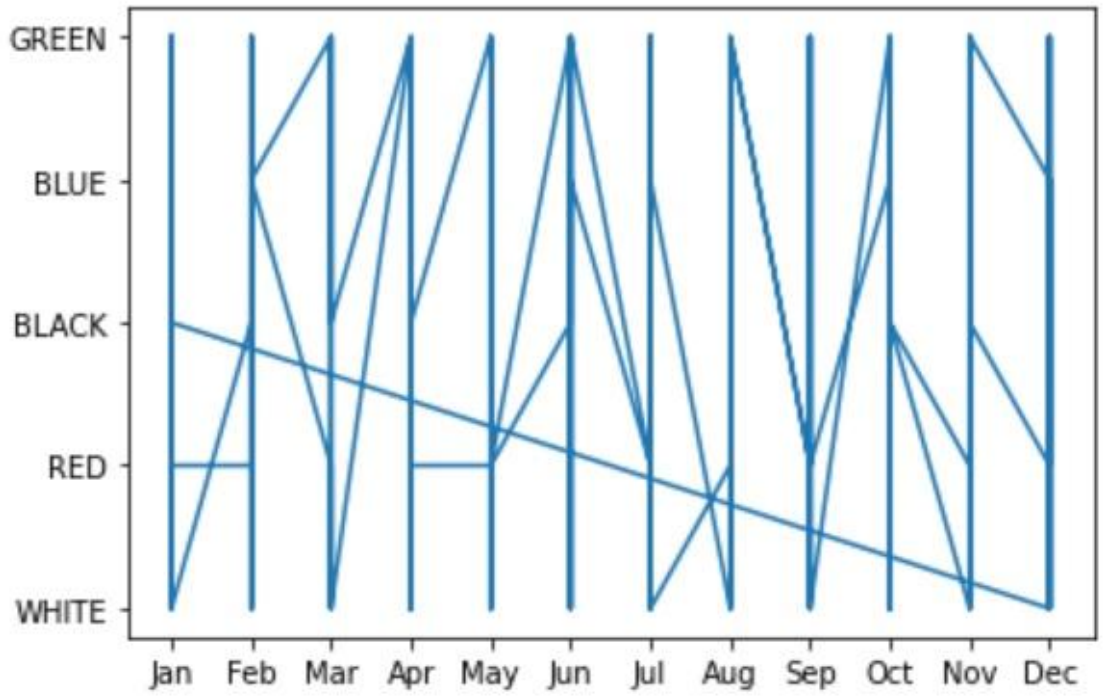


➔ Below Line graph represents the relation between product category and month wise of the Zara dataset.
➔ This graph is generated using matplotlib functions.
➔ X-axis represents product category and Y-axis represents Product category of all products.
➔ From this graph can get sales variation across the year for particular category, all presented in one graph.

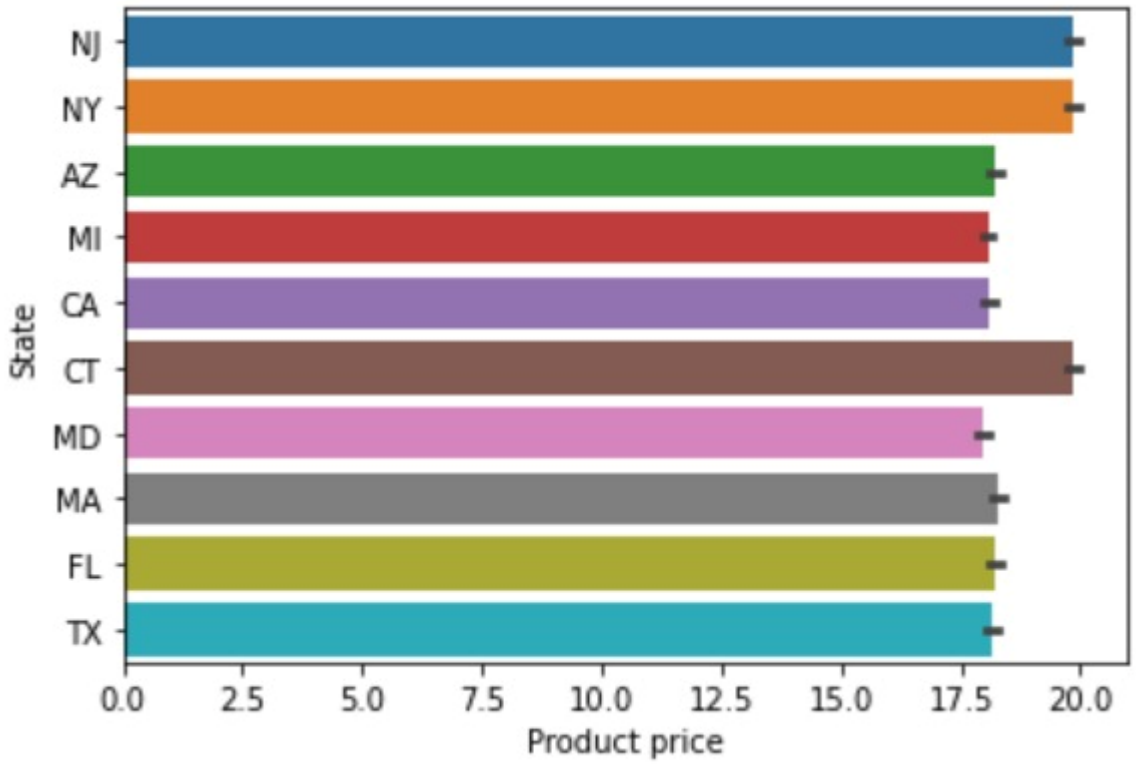


➔

➔ Below Line graph represents the relation between product color and month wise of the Zara dataset.
➔ This graph is generated using matplotlib functions.
➔ X-axis represents product category and Y-axis represents Product color of all products.
➔ From this graph can get sales variation across the year for particular product color, all presented in one graph.
➔



Below bar graph represents the relation between product price and state wise of the Zara dataset.
➔ This graph is generated using seaborn functions.
➔ X-axis represents product prices and Y-axis represents states.
➔ From this graph can get sales variation across the states, all presented in one graph.



Classification reports:

Decision Tree:

To optimize the Decision Tree algorithm, we used a combination of Grid Search and Cross-Validation techniques to find the optimal values for the hyperparameters: maximum depth, minimum samples split, and minimum samples leaf. We compared the performance of the optimized Decision Tree with the default model using two evaluation metrics: accuracy and F1 Score.

	precision	recall	f1-score	support
0	0.98	0.99	0.99	3712
1	1.00	1.00	1.00	17028
accuracy			1.00	20740
macro avg	0.99	0.99	0.99	20740
weighted avg	1.00	1.00	1.00	20740

K-Nearest Neighbors (KNN):

To optimize the KNN algorithm, we used Grid Search and Cross-Validation techniques to find the optimal values for the hyperparameters: number of neighbors and distance metric. We compared the performance of the optimized KNN with the default model using two evaluation metrics: accuracy and recall.

	precision	recall	f1-score	support
0	0.72	0.65	0.68	3712
1	0.92	0.95	0.94	17028
accuracy			0.89	20740
macro avg	0.82	0.80	0.81	20740
weighted avg	0.89	0.89	0.89	20740

MultinomialNB:

To optimize the MultinomialNB algorithm, we used Cross-Validation technique to find the optimal values for the hyperparameters: alpha value. We compared the performance of the optimized MultinomialNB with the default model using two evaluation metrics: accuracy and precision.

	precision	recall	f1-score	support
0	0.95	0.28	0.43	3712
1	0.86	1.00	0.93	17028
accuracy			0.87	20740
macro avg	0.91	0.64	0.68	20740
weighted avg	0.88	0.87	0.84	20740

Random Forest:

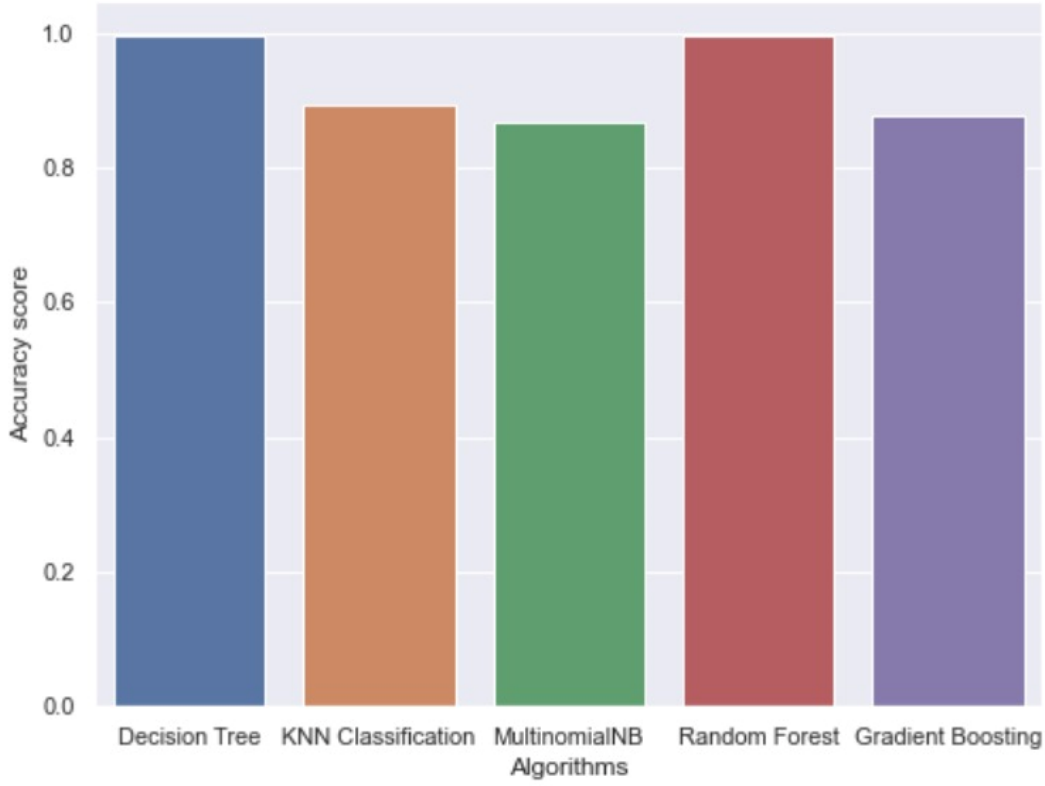
To optimize the Random Forest algorithm, we used Grid Search and Cross-Validation techniques to find the optimal values for the hyperparameters: number of trees, maximum depth, minimum samples split, and minimum samples leaf. We compared the performance of the optimized Random Forest with the default model using two evaluation metrics: accuracy and F1-score.

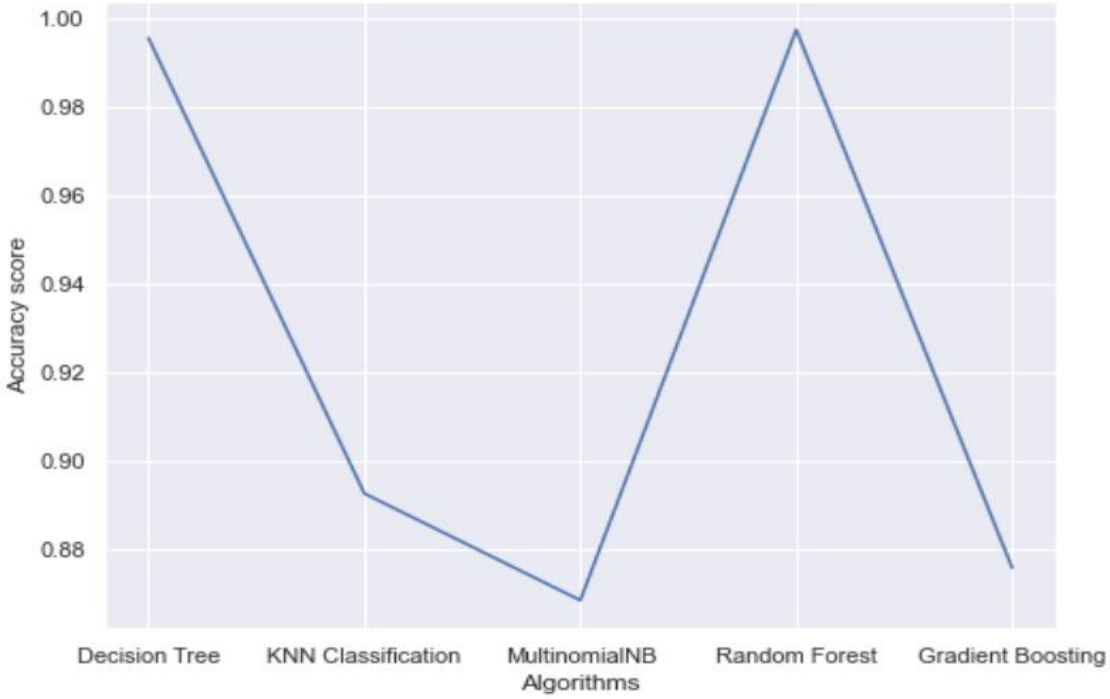
	precision	recall	f1-score	support
0	1.00	0.99	0.99	3712
1	1.00	1.00	1.00	17028
accuracy			1.00	20740
macro avg	1.00	0.99	1.00	20740
weighted avg	1.00	1.00	1.00	20740

Gradient Boosting:

To optimize the Gradient Boosting algorithm, we used Grid Search and Cross-Validation techniques to find the optimal values for the hyperparameters: learning rate, number of trees, maximum depth, and subsample size. We compared the performance of the optimized Gradient Boosting with the default model using two evaluation metrics: accuracy and F1-score.

	precision	recall	f1-score	support
0	0.85	0.37	0.52	3712
1	0.88	0.99	0.93	17028
accuracy			0.88	20740
macro avg	0.86	0.68	0.72	20740
weighted avg	0.87	0.88	0.85	20740





Discussion:

The analysis of Zara's sales and market prediction focuses on interpreting the findings, exploring their implications, and providing insights into the significance of accurate sales analysis and prediction for Zara's business strategies. Here are some key points that could be discussed in this section:

Sales Performance Insights:

Discuss the patterns and trends observed in Zara's sales performance analysis. Identify the factors that significantly influence sales, such as seasonal variations, product categories, customer demographics, and geographic regions. Analyze how these insights can be leveraged to optimize inventory management, marketing campaigns, and sales strategies.

Market Prediction Accuracy:

Evaluate the accuracy and reliability of the market prediction techniques applied. Compare the forecasted sales figures with the actual sales data to assess the predictive performance of the models. Discuss any limitations or challenges encountered during the prediction process and propose potential improvements for future forecasting efforts.

Forecasting Future Market Conditions:

Examine the implications of accurate market predictions for Zara's decision-making processes. Discuss how reliable sales forecasts can assist in resource allocation, production planning, and inventory management. Highlight the benefits of proactively adjusting strategies based on anticipated market conditions to optimize sales and maintain a competitive advantage.

Risk and Opportunity Analysis:

Explore the risks and opportunities identified through sales analysis and market prediction. Discuss factors that could potentially impact Zara's sales, such as changing fashion trends, economic fluctuations, or competitor actions. Assess how accurate sales predictions can help Zara mitigate risks and capitalize on emerging opportunities in the market.

Integration of External Factors:

Discuss the relevance of incorporating external factors into sales analysis and market prediction. Consider the inclusion of macroeconomic indicators, social media sentiment analysis, or competitor performance data to enhance the accuracy of sales forecasts. Evaluate the potential benefits and challenges associated with incorporating these factors into the prediction models.

Decision Support for Sales Strategies:

Explain how the insights gained from the analysis and prediction can guide Zara's sales strategies. Discuss how accurate sales forecasting can assist in developing pricing strategies, determining product assortments, and identifying target markets. Emphasize the importance of data-driven decision-making in optimizing sales performance.

Limitations and Future Directions:

Acknowledge any limitations of the analysis and prediction techniques employed. Discuss potential areas for future research and improvement, such as incorporating advanced machine learning algorithms, exploring predictive analytics tools, or considering emerging market trends. Highlight the significance of ongoing analysis and adaptation to evolving market dynamics.

By addressing these points, the discussion section provides a comprehensive analysis of the findings, their implications, and suggestions for leveraging accurate sales analysis and market prediction to enhance Zara's sales strategies and overall performance in the fashion retail industry.

APPENDIX FOR LINK TO THE GITHUB REPOSITORY: -

https://github.com/rishi-krishna/THE_MINERS

CONCLUSION:

In conclusion, our research on "Exploring Zara's Innovative Marketing and Supply Chain Strategies to Drive Sales. Identification of Deviations in sales and supply chain based on Customer reviews and sentiment analysis" has provided valuable insights into Zara's supply chain and marketing strategies. Through data exploration and analysis, we identified the top-selling colors, product categories, and states for Zara US, as well as the peak sales months in 2017. Our findings have highlighted the importance of understanding sales patterns and customer behavior to make informed business decisions and improve marketing strategies. Overall, these findings can help Zara US to better understand their sales patterns, make informed business decisions and improve their marketing strategies.

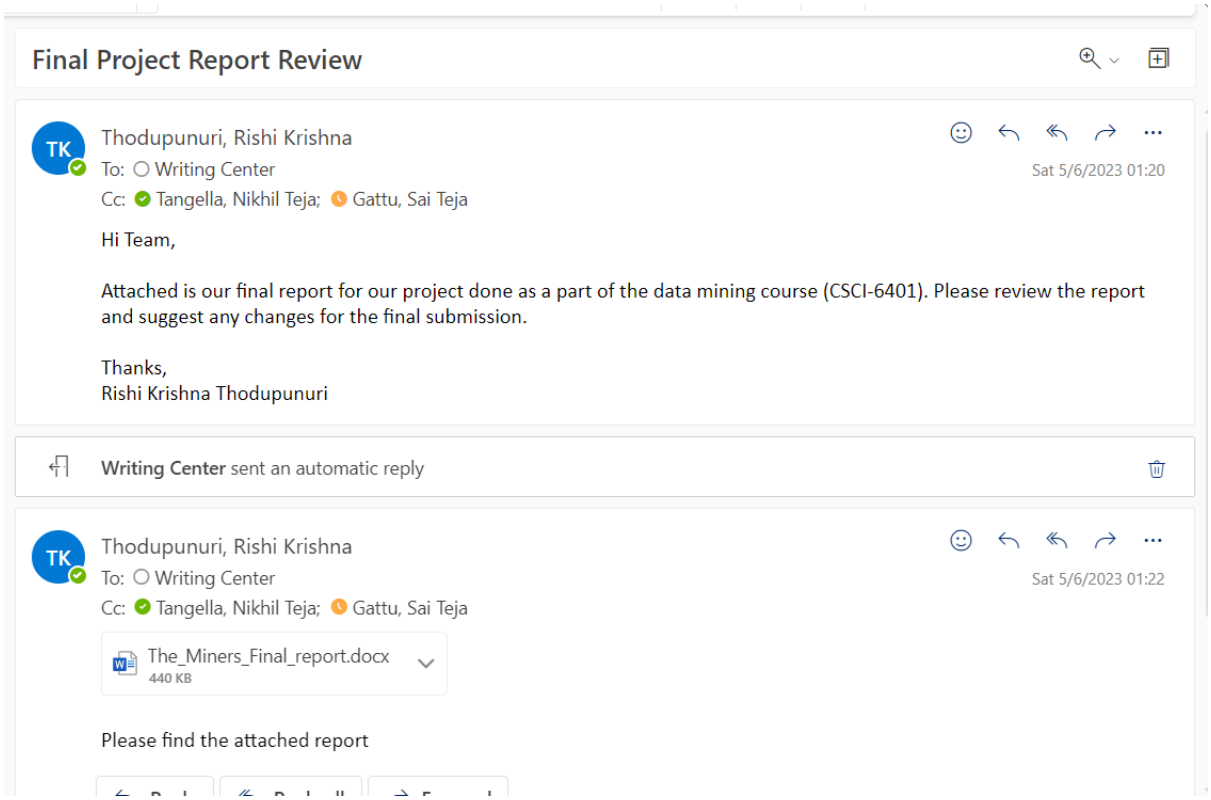
FUTURE WORK:

To improve upon our work, future studies could expand the analysis to other Zara markets and explore the effectiveness of Zara's marketing strategies in different cultural contexts. Additionally, incorporating social media data could provide a more comprehensive understanding of customer sentiment towards Zara. Finally, future studies could examine the impact of Zara's supply chain innovations on sustainability and ethical practices.

Reference:

https://ecommercedb.com/reports/zara-com-brand-report-poland-2021/284
https://www.dataandsons.com/data-market/product-lists/zara-uk-fashion-data-in-csv-format
https://www.kaggle.com/crawlfeeds/zara-us-fashion-products-dataset
https://search.vi-seem.eu/dataset/zara_dataset
https://www.statista.com/forecasts/1218316/zara-revenue-development-ecommercedb
https://www.kaggle.com/datasets/thedevastator/fast-fashion-eco-data
https://www.volza.com/p/zara-clothes/
https://www.globaldata.com/data-insights/consumer/the-sales-of-zara-in-clothing---footwear-industry-in-china-1861343/
https://www.aggdata.com/clothing_store_locations/zara_spain
https://agenty.com/marketplace/stores/zara-retail-store-locations-in-germany
https://ieeexplore-ieee-org.unh-proxy01.newhaven.edu/document/4682950
https://ieeexplore-ieee-org.unh-proxy01.newhaven.edu/document/4679487
https://ieeexplore-ieee-org.unh-proxy01.newhaven.edu/document/4654593
https://github.com/fediazgon/zara-data-challenge-19

Proofreading with an email from Writing Center:



Final Project Report Review

TK Thodupunuri, Rishi Krishna
To: Writing Center
Cc: Tangella, Nikhil Teja; Gattu, Sai Teja
Sat 5/6/2023 01:20

Hi Team,

Attached is our final report for our project done as a part of the data mining course (CSCI-6401). Please review the report and suggest any changes for the final submission.

Thanks,
Rishi Krishna Thodupunuri

Writing Center sent an automatic reply

TK Thodupunuri, Rishi Krishna
To: Writing Center
Cc: Tangella, Nikhil Teja; Gattu, Sai Teja
Sat 5/6/2023 01:22

The_Miners_Final_report.docx
440 KB

Please find the attached report