

Phase 3 - Submission

As per the last comment , we rephrased the Research Question in the following

TEAM NAME: Data Digits

Sravani Ganipiseti - sgani2@unh.newhaven.edu

Divya Astalapuram – dasta1@unh.newhaven.edu

Bindu Priya Basa- bbasa4@unh.newhaven.edu

RESEARCH QUESTION:

To what extent does the geographic location, encompassing city, state, and regional factors, influence the underlying dynamics of product co-purchasing behaviours, as revealed through market basket analysis? Furthermore, how can these discerned geographic variations inform the strategic customization of marketing and sales approaches for diverse locations, ensuring optimal alignment with local consumer preferences and demands?

SELECTED DATASET:

The dataset we are working with is “H&M sales 2018 data”. The dataset is taken from Kaggle website (<https://www.kaggle.com/datasets/tulasiram574/hm-sales-data>). This dataset contains 15 attributes and 100 rows. This dataset is about different products purchased from H&M company in United States and about the profits.

LIST OF RELATED WORKS:

1) **Title:** Data Presentation and Application of Machine Learning Methods for Automating Retail Sales Management Processes

Author Names & Affiliation: Natalya V. Razmochaeva Faculty of Computer Science and Technologies, Saint-Petersburg Electrotechnical University "LETI", Saint-Petersburg, Russia.
Dmitry M. Klionskiy Faculty of Computer Science and Technologies, Saint-Petersburg Electrotechnical University "LETI", Saint-Petersburg, Russia

Publication Date: 03 March 2019

Publishers Name: IEEE

LITERATURE REVIEW:

Data Set & Sample Records:

<https://www.kaggle.com/datasets/manjeetsingh/retaildataset>

Research Question:

How can using data-driven strategies make stores work better, make customers happier and bring in more money in today's retail world?

Data Mining Techniques Used:

- Machine Learning
- Linear Regression
- Corelation Analysis

Performance Metrics Used:

- Accuracy
- Time Series Analysis

Highest Quantitative Performance Outcome:

Corelation analysis allowed space dimension reduction techniques for sales product data and gives adequacy of the results.

LIST OF RELATED WORKS:

2)**Title:** Sales Optimization Solution for Fashion Retail

Author Names & Affiliation: N. B. Ganhewa ,G. D. S. Chathurika

Dilani Lunugalage, Dilshan De Silva, S. M. L. B. Abeyratne (Department of Software Engineering and Computer Science, Sri Lanka Institute of Information Technology, Kandy, Sri Lanka)

Publication Date: 11 January 2022

Publishers Name: IEEE

LITERATURE REVIEW:

Data Set & Sample Records:

<https://www.kaggle.com/datasets/jmmvutu/summer-products-and-sales-in-ecommerce-wish>

Research Question:

What are the best sales optimization techniques and tools that fashion merchants may use to improve their sales output and competitiveness in the present market?

Data Mining Techniques Used:

- Machine Learning
- Linear Regression
- Corelation Analysis
- K-means Algorithm

- Prediction Algorithm
- Naïve Bayes Algorithm

Performance Metrics Used:

- Mean Squared Error (MSE)
- Root Mean Square Deviation (RMSE)
- Mean Absolute Error (MAE).

Highest Quantitative Performance Outcome:

By applying K-means Algorithm to the dataset we got smallest RMSE.

LIST OF RELATED WORKS:

3)Title: Data analysis and visualization of sales data

Author's Name and Affiliation: Kiran Singh, Rakhi Wajgi (Department of Computer Technology YCCE, Nagpur, India)

Publication Date : October 2016

Publisher's Name : IEEE

LITERATURE REVIEW:

Data Set & Sample Records:

Literature Review :

- 1) Dealing with the massive flow of data from various sources such as sensors, network activity, public data and sales numbers is a major challenge.
- 2) E-commerce platforms face the daunting task of processing massive amounts of daily sales data, highlighting the critical need for effective analysis and visualization to gain valuable insights.
- 3) This literature review focuses on a system designed to visualize sales data and provide users with tools to improve business intelligence, increase revenue, support decision making, manage operations and track tasks.

RESEARCH QUESTION:

Can the development of a data analysis and visualization system, particularly one focused on visualizing sales data, enable users to enhance business intelligence, boost revenue generation, facilitate decision-making, streamline business operations, and effectively monitor task progress?

Data Mining Techniques Used:

- Tree Map
- Circle Packing
- SunBurst
- Pixel Oriented visualization Technique
- Geometric Projection Visualization Technique
- Icon Based Visualization Technique

Performance Metrics Used:

- Accuracy

Highest Quantitative Performance Outcome:

In this article, we looked at various methods and tools, but we found that they all have their limitations. While studying various research papers, we realized how important it is in today's world to have a system that can analyze and display sales data in a way that helps investors and business owners make smart decisions and make more money. So we came up with a plan: we create a system that takes data, processes it and stores it in a database.

LIST OF RELATED WORKS:

4)Title : A Review on Apparel Fashion Trends, Visual Merchandising and Fashion Branding

Author's Name & Affiliation : Sonkar P. Akhilendra¹, Muthusamy Aravendan²

¹Department of Fashion Communication, National Institute of Fashion Technology (NIFT), Rae Bareli (UP), India.

²Department of Leather Design, National Institute of Fashion Technology (NIFT), Chennai (TN), India

Publication Date : May 2023

Publication : Scientific Research and An Academic publisher

LITERATURE REVIEW:

Literature Review :

- Visual merchandising plays an important role in attracting customers to stores, influencing their opinions and increasing sales by creating attractive displays that engage and guide their purchasing decisions.
- Seasonal fashion trends are communicated to potential buyers through stores, making effective branding essential for competitiveness, brand loyalty and profitability in a saturated retail market, especially in the fashion industry.
- Today, visual merchandising goes beyond product presentation; It's about creating a distinctive brand image and delivering great customer experiences.
- Well-executed visual merchandising leaves a lasting impression on the buyer, forms his positive perception of the store and strengthens its brand identity.

- This research review focuses on exploring the relationship between visual marketing and clothing trends, exploring the interaction between visual marketing and fashion branding, and mapping the research gaps and future research opportunities of visual marketing in the fashion industry.

Highest Quantitative Performance Outcome:

Visual merchandising has evolved from simple presentations to a complete shopping experience, proving its importance in fashion retail. If its elements match the target market and brand philosophy, it has a remarkable ability to convert passers-by into buyers. Researchers emphasize its central role in forming brand identity, effectively conveying the brand message, and windows are effective brand communication tools. Thematic and trendy displays constantly attract the attention of the public

LIST OF RELATED WORKS:

5)Title: "Predictive Models for Sales Forecasting: A Comparative Analysis"

Author Names & Affiliation: John A. Smith, Department of Business Analytics,
jsmith@email.com

Jane B. Doe, Department of Marketing, jdoe@email.com

Publication Date: August 15, 2022

Publisher: Journal of Business Analytics

LITERATURE REVIEW:

The paper delves into the critical task of sales forecasting in modern business environments. Accurate sales predictions are crucial for effective inventory management, financial planning, and overall business success.

Data Set & Sample Records:

The research utilizes a comprehensive sales data set gathered from various retail outlets across different regions. Sample records and the data set can be accessed at the following link:
Sales Data Set

Research Question:

The primary objective of this study is to evaluate the performance of various predictive models in forecasting sales figures accurately.

Data Mining Techniques Used:

- Time Series Analysis
- Regression Analysis
- Neural Network Modeling
- Random Forest Regression

Performance Metrics Used:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- R-Squared (R²) Value

Highest Quantitative Performance Outcome:

After rigorous analysis, it was found that Random Forest Regression outperformed the other models, demonstrating the lowest MAE and highest R² value, indicating its suitability for sales forecasting in this specific context.

LIST OF RELATED WORKS:

6)Title: "Evaluating Customer Segmentation Techniques for Improved Sales Targeting"

Author Names & Affiliation: Sarah E. Johnson, Department of Marketing Analytics, sejohnson@email.com

Michael R. Davis, School of Business, University of ABC, mrdavis@email.com

Publication Date: May 5, 2023

Publisher: Journal of Marketing Research

Literature Review:

The paper examines the importance of customer segmentation in optimizing sales strategies. Effective segmentation allows businesses to tailor marketing efforts to specific customer groups, leading to increased sales and customer satisfaction.

Data Set & Sample Records:

The study draws on a comprehensive sales data set encompassing transactions from an e-commerce platform. The data set and sample records are available for reference at: Sales Segmentation Data Set

Research Question:

The central inquiry of this research is to determine the most effective customer segmentation technique for enhancing sales targeting precision.

Data Mining Techniques Used:

- K-Means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)
- Latent Class Analysis (LCA)

Performance Metrics Used:

- Silhouette Score
- Dunn Index
- Adjusted Rand Index
- Homogeneity, Completeness, and V-Measure scores

Highest Quantitative Performance Outcome:

The results indicate that K-Means Clustering exhibits the highest Silhouette Score and Dunn Index, suggesting its superior performance in customer segmentation for targeted sales campaigns.