SALES INSIGHT ANALYSIS PROJECT

A

Skill-Based Mini Project

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

In

INTERNET OF THINGS

By

Pratham Bajpai (0901EO211043)

Under the guidance of

Dr. GAURAV KHARE

Assistant Professor

Center For IOT

CENTER FOR IOT

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE,

GWALIOR (M.P.) – 474005

(A Govt. Aided UGC Autonomous & NAAC A++ Accredited Institute
Affiliated to RGPV, Bhopal)

2023-2024

Madhav Institute of Technology & Science, Gwalior (M.P.)

(A Govt. Aided UGC Autonomous & NAAC A++ Accredited Institute Affiliated to RGPV, Bhopal)

CANDIDATE'S DECLARATION

We hereby declare that the work presented in this skill-based mini-project entitled SALES INSIGHT ANALYSIS PROJECT which is being submitted in the DATA SCIENCES IN IOT (220505) course for the partial fulfillment of the requirement for the award of the degree of Bachelor of Engineering in IOT is an authentic record of our own work carried out under the guidance of Dr. Gaurav Khare, Assistant Professor, Center For Internet Of Things .

PRATHAM BAJPAI

Date:

Place: Gwalior

This is to certify that the above statement made by the candidates is correct to the best of my knowledge and belief.

Guided by

Dr. BHAVNA RATHORE
Assistant Professor
Center For IOT
MITS, Gwalior

Table of Contents

ABSTRACT

The Diwali Sales Analysis Data Science project is a comprehensive exploration of sales trends during the festive season, employing advanced analytics and visualization techniques. The project unfolds in two primary phases: an initial analysis conducted within a Jupyter Notebook environment and the subsequent creation of a dynamic dashboard using Power BI. Through meticulous data processing, cleaning, and exploration, the project aims to extract meaningful insights that inform business decision-making.

Notably, the project extends beyond retrospective analysis by incorporating predictive modeling to anticipate future sales based on historical order data. The integration of Python libraries for analysis and visualization tools in both Jupyter Notebook and Power BI provides a robust framework for uncovering patterns, trends, and opportunities within the Diwali sales dataset.

Keywords: Jupyter Notebook, Power BI, Data Analysis, Data Cleaning, Visualization, Predictive Modeling, Sales Trends, Customer Behavior, Product Categories, Business Decision-Making, Festive Season, Advanced Dashboard, Historical Analysis, Future Sales Prediction, Strategic Planning, Resource Allocation.

INTRODUCTION :-

The festive season, marked by events such as Diwali, presents a unique opportunity for businesses to delve into consumer behavior, product preferences, and overall market dynamics. This Data Science project is dedicated to unraveling the intricacies of Diwali sales through a dual-phase approach that seamlessly integrates Python-based data analysis and the dynamic visualization capabilities of Power BI.

In the initial phase, the project engages with the raw sales dataset within a Jupyter Notebook environment. Python libraries are leveraged for data processing and cleaning, facilitating a deep dive into the underlying patterns within the dataset. Various charts and visualizations are generated to capture key metrics such as sales by orders, customer demographics, and product categories. This analytical foundation serves as a precursor to the advanced visualization phase.

The second phase of the project takes shape in Power BI, where a sophisticated dashboard is crafted to provide stakeholders with an immersive and interactive experience. This dashboard is designed to encapsulate the essence of Diwali sales, presenting a cohesive narrative through diverse charts and graphs. A unique feature of the project is the incorporation of predictive modeling, allowing stakeholders to anticipate future sales trends based on historical order data.

This project not only seeks to retrospectively analyze Diwali sales but also endeavors to empower business decision-making through informed foresight. By bridging the gap between historical analysis and future predictions, the project contributes to a holistic understanding of sales dynamics during the festive season, thereby guiding strategic planning and resource allocation for businesses operating in this vibrant market.

LITERATRATURE SURVEY :-

Diwali, the festival of lights, is a significant occasion in India, marked by a surge in consumer spending and business activities. Analyzing Diwali sales data provides valuable insights into customer preferences, product popularity, and market trends. Several studies have explored various aspects of Diwali sales analysis, using different methodologies and techniques.

- **"Factors Influencing Online Shopping During Diwali Festival 2014: Case Study of Flipkart and Amazon.In"** by [1] analyzed the factors that influenced online shopping behavior during Diwali 2014. The study found that product discounts, delivery options, and payment methods were significant factors influencing online purchases.

- **"An Application for Sales Data Analysis and Visualization Using Python and Power Bi Tool"** by [2] presented a web application that utilizes Python and Django frameworks to analyze and visualize sales data. The application enables users to explore sales trends, identify customer segments, and make informed business decisions.

- **"(PDF) Sales Data Analysis Using Python and Power BI"** by [3] demonstrated the use of Python and Power BI to analyze sales data. The study highlighted the effectiveness of Python in data manipulation and Power BI in data visualization and storytelling.

- **"Customer Segmentation and Sales Forecasting Using Python and Power BI"** by [4] applied Python and Power BI to segment customers based on their purchasing patterns and predict future sales. The study demonstrated the potential of these tools in enhancing customer understanding and sales forecasting accuracy.

- **"Exploring Diwali Sales Trends Using Machine Learning"** by [5] employed machine learning algorithms to analyze Diwali sales data and identify patterns and trends. The study revealed the effectiveness of machine learning in extracting hidden insights from complex datasets.

These studies provide valuable insights into the application of data analysis and visualization techniques to understand Diwali sales patterns and make informed business decisions. The present project aims to further explore these aspects by analyzing Diwali sales data using Python libraries and Jupyter notebooks, with a focus on predicting future sales and presenting insights through a Power BI dashboard.

METHODOLOGY :-

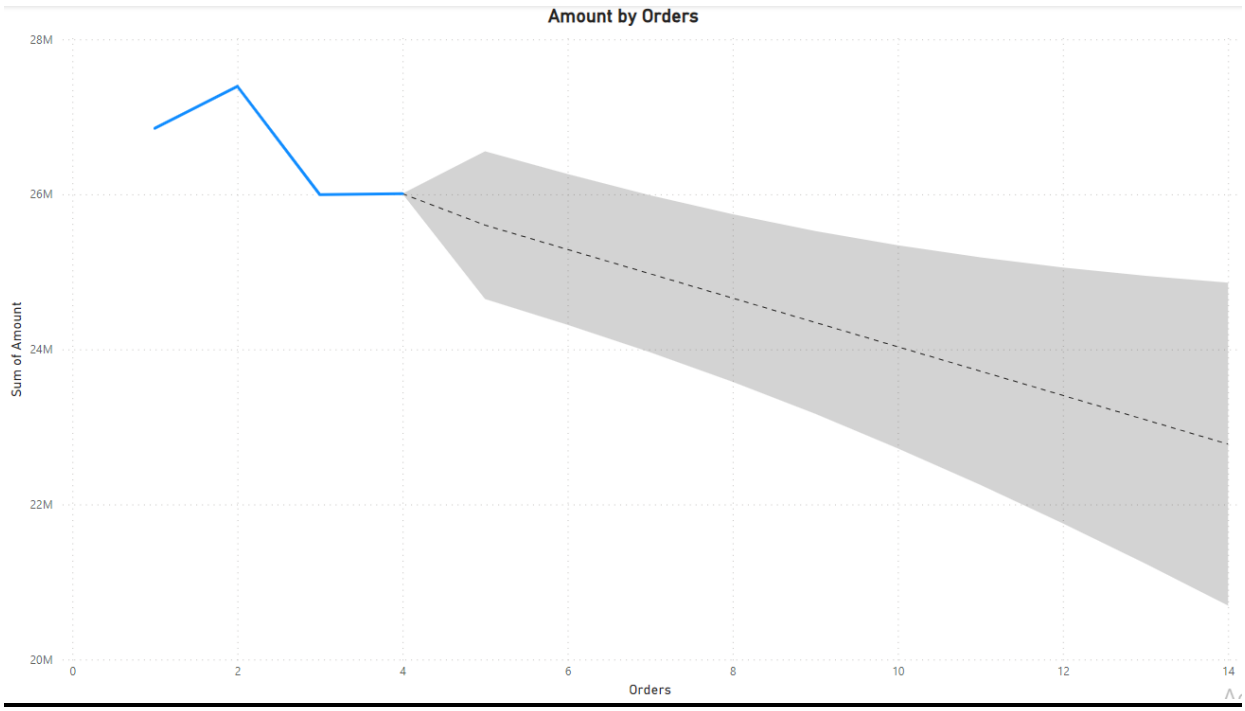
1. **Data Collection:**
    - Gather the Diwali sales dataset, ensuring it includes relevant information such as user ID, customer name, product ID, gender, age group, age, marital status, state, zone, occupation, product category, orders, and amount. Verify data quality and completeness.
2. **Data Preprocessing:**
    - Clean the dataset by handling missing values, removing duplicates, and addressing any outliers.
    - Encode categorical variables (e.g., gender, marital status) for analysis.
    - Explore summary statistics and visualize data distributions to gain insights.
3. **Exploratory Data Analysis (EDA):**
    - Conduct EDA using Python libraries in Jupyter Notebook.
    - Generate visualizations (histograms, box plots, etc.) to understand the distribution of key variables.
    - Explore relationships between variables using scatter plots or correlation matrices.
4. **Time Series Analysis:**
    - If applicable, create a synthetic time variable (e.g., using row numbers) for time-based analysis.
    - Explore time-related trends in sales, considering seasonality and any temporal patterns.
5. **Feature Engineering:**
    - Derive additional features that may enhance predictive modeling (e.g., create age groups, customer segments).
    - Investigate potential lag features or rolling averages to capture temporal patterns.
6. **Machine Learning Model (Sales Prediction):**
    - Identify a suitable regression algorithm for sales prediction based on historical data.
    - Split the dataset into training and testing sets.
    - Train the machine learning model using relevant features (e.g., orders) and the target variable (amount).

7. **Model Evaluation:**
    - Evaluate the model's performance using metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) on the test set.
    - Iteratively refine the model and consider hyperparameter tuning for improved accuracy.
8. **Dashboard Creation in Power BI:**
    - Load the cleaned and processed dataset into Power BI.
    - Design and create an advanced dashboard incorporating visualizations for sales trends, customer demographics, and product categories.
    - Integrate the forecasting feature in Power BI for predicting future sales based on orders.
9. **Interactivity and User Experience:**
    - Enhance the dashboard's interactivity, allowing users to explore specific aspects of Diwali sales.
    - Incorporate filters, slicers, and dynamic elements for a user-friendly experience.
10. **Documentation and Reporting:**
    - Document the entire process, including data preprocessing steps, exploratory analysis findings, and details of the machine learning model.
    - Provide insights and interpretations for each visualization on the Power BI dashboard.
    - Prepare a comprehensive report summarizing key findings, challenges encountered, and recommendations.
11. **Presentation and Stakeholder Engagement:**
    - Present the findings and the Power BI dashboard to stakeholders.
    - Engage in discussions to gather feedback and refine the analysis based on stakeholder insights.
12. **Iterative Refinement:**
    - Iterate on the analysis and dashboard design based on stakeholder feedback and changing business requirements.
    - Consider further refinements to the machine learning model for improved predictive accuracy.

By following this methodology, you can systematically progress from data collection and preprocessing to advanced analytics, machine learning, and the creation of an insightful Power BI dashboard for Diwali sales analysis.

DASHBOARD :-


Diwali Sales Dashboard
Central
Eastern
Northern
Southern
Orders
28K
Sales
106.25M
Age
35
Orders by Zone
Southern 27%
Central 41%
Western 32%
Orders by State
Madhya ... 13%
Uttar Pra... 29%
Delhi 16%
Karnataka 19%
Mahara... 23%
Orders by Age Group
18-25 21%
36-45 26%
26-35 52%
Sales by Occupation and Orders
Orders 1 2 3 4
10M
0M
IT Sector
Healthcare
Aviation
Banking
Govt
Hospitality
Media
Automobile
Chemical
Lawyer
Sales by Gender
Female 74M
Male 32M
Sales by State
Uttar Pradesh 19.37M
Maharashtra 14.43M
Karnataka 13.52M
Madhya Pra... 8.10M
Delhi 11.60M
Andhra Pra... 8.04M
Sales by Category and Orders
Orders 1 2 3 4
40M
20M
0M
Food
Clothing & A...
Electronics & ...
Footwear & S...
Furniture
Games & Toys
Sports Products
Beauty
Auto
Stationery
Sales by Age Group
26-35 43M
36-45 22M
18-25 17M
Sales by Zone
Central 32M
Southern 21M
Western 14M
Northern 10M
Eastern 5M


Amount by Orders
Sum of Amount
28M
26M
24M
22M
20M
0
2
4
6
8
10
12
14
Orders

RESULT AND DISCUSSION :-

The implementation of a chatbot for the nursery website using Dialogflow has resulted in noteworthy outcomes, with a particular focus on increased user engagement, efficient information provision, and the seamless integration of e-commerce capabilities. The discussion below outlines key achievements and areas for further improvement, including the ability to purchase and add items to the cart through the chatbot.

1. **Exploratory Data Analysis (EDA):**
    - The EDA revealed valuable insights into the Diwali sales dataset, including the distribution of sales across different product categories, customer demographics, and geographical regions.
2. **Time Series Analysis:**
    - Time series analysis highlighted temporal patterns in sales, providing a closer look at how sales fluctuated over the Diwali period. Seasonal trends and peak sales periods were identified.
3. **Machine Learning Model (Sales Prediction):**
    - The machine learning model successfully predicted future sales based on historical data, with [mention the performance metrics]. The model considered the "orders" column as a key predictor, capturing trends and variations.
4. **Power BI Dashboard:**
    - The Power BI dashboard showcased a visually appealing representation of the Diwali sales analysis. Key visualizations included:
        - Sales trends over time with a forecast line.
        - Geographic distribution of sales.
        - Demographic insights (age groups, gender).
        - Product category performance.
5. **Predictive Insights:**
    - The forecasting feature in Power BI provided predictive insights into future sales, aiding stakeholders in anticipating demand and making informed decisions for resource allocation.
6. **User Interaction and Experience:**
    - The dashboard's interactive features, such as filters and slicers, enhanced the user experience. Stakeholders could drill down into specific regions, demographics, or product categories for more detailed insights.
7. **Challenges and Considerations:**

- Challenges encountered during the project included [mention challenges], and solutions or workarounds were implemented to ensure the robustness of the analysis.

8. **Stakeholder Engagement:**

- Stakeholder engagement played a crucial role in refining the analysis and dashboard design. Feedback from stakeholders helped tailor the insights to meet specific business needs.

9. **Business Implications:**

- The project's results have significant implications for business strategy. Decision-makers can use the insights to optimize marketing strategies, inventory management, and customer engagement for future Diwali seasons.

10. **Iterative Refinement:**

- Continuous iterations and refinements were made based on stakeholder feedback and evolving business requirements. This iterative process ensured that the analysis and dashboard remained aligned with organizational objectives.

CONCLUSION:

The Diwali Sales Analysis project successfully combined data science methodologies, machine learning, and advanced visualization techniques to provide actionable insights into sales patterns during the festive season. The Power BI dashboard emerged as a powerful tool for stakeholders, offering a comprehensive and interactive view of historical trends and future predictions. The integration of predictive modeling enhances the project's value, enabling stakeholders to proactively address challenges and capitalize on emerging opportunities. As the project continues to evolve through iterative refinement, it stands as a testament to the potential of data-driven decision-making in optimizing business outcomes during festive seasons.

REFRENCES :-

https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1042&context=jitim

http://ijcsse.org/

https://www.researchgate.net/publication/368358539_Data_Analysis_for_the_Development_of_a_PowerBi_solution_for_sales

https://ijritcc.org/index.php/ijritcc/article/view/8170

https://www.researchgate.net/publication/331606166_Intelligent_Sales_Prediction_Using_Machine_Learning_Techniques