PIZZAS SALES ANALYSIS AND PREDICTION

PRESENTED BY : C. KOWSALYA

TABLE OF CONTENTS

PROBLEM STATEMENT

- The Pizza Sales Analysis problem aims to extract valuable insights and predictive patterns from a comprehensive dataset capturing the sales dynamics of a pizzeria. The dataset contains detailed information on individual pizza orders, encompassing aspects such as pizza name, quantity, order date and time, unit price, total price, pizza size, category, ingredients, and more.

- With this rich dataset at hand, the objective is to develop a data-driven machine learning model that can predict future pizza sales based on a combination of factors. These factors include the pizza's attributes such as size, category, and ingredients, as well as temporal variables like order date and time. By considering these variables collectively, the model should provide an accurate estimation of the sales figures for various pizzas over time.

- By harnessing the power of machine learning and predictive analytics, this problem aspires to empower pizzeria owners and managers with actionable insights. The model's predictions can aid in optimizing inventory management, production planning, and marketing strategies. Furthermore, understanding the impact of different pizza attributes on sales can guide menu enhancements and special promotions, thereby driving customer satisfaction and boosting business revenue. Short the paragraph without content loss .

DATA SET AND FEATURES

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48620 entries, 0 to 48619
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   pizza_id           48620 non-null  int64
 1   order_id           48620 non-null  int64
 2   pizza_name_id      48620 non-null  object
 3   quantity           48507 non-null  float64
 4   order_date         48620 non-null  datetime64[ns]
 5   order_time         48620 non-null  object
 6   unit_price         48620 non-null  float64
 7   total_price        48620 non-null  float64
 8   pizza_size         48526 non-null  object
 9   pizza_category     48553 non-null  object
 10  pizza_ingredients  48620 non-null  object
 11  pizza_name         48620 non-null  object
dtypes: datetime64[ns](1), float64(3), int64(2), object(6)
memory usage: 4.5+ MB
```

1)Pizza ID: This attribute serves as a unique identifier for each distinct pizza.

2)Order ID: This attribute functions as an identifier for individual pizza orders within the system.

3)Pizza Name ID: This attribute corresponds to the identifier associated with the specific name of a pizza.

4)Quantity: This attribute quantifies the number of pizzas included in an order.

5)Order Date: This attribute the date on which a customer placed a pizza order.

6)Order Time: This attribute specifies the exact time when a customer's pizza order was placed.

DATA SET AND FEATURES

7)Total Price: Representing the cumulative cost of the pizzas, this attribute is denominated in dollars. The dataset is derived from U.S. pizza sales data.

8)Unit Price: This attribute signifies the price of a single unit of pizza.

9)Pizza Size: Indicating the dimensional classification of the pizza, this attribute describes its size.

10)Pizza Category: Offering insight into the classification of the pizza, this attribute conveys its category.

11)Pizza Ingredients: This attribute lists the specific constituents utilized in the preparation of the pizza.

12)Pizza Name: This attribute articulates the unique name attributed to a specific pizza variant.

EXPLORATORY DATA ANALYSIS

Missing Values Treatment

In this dataset, we have 3 variables that are treated with some filling method for null values. Since these 3 variables are object values, we can fill them with the mode option.

Here are some extra points about filling null values with the mode option:

- The mode is the most frequent value in a column.
- Filling null values with the mode is a good option when the data is categorical.
- This is because the mode is the most likely value for a missing value in a categorical column.

I have performed Exploratory Data Analysis (EDA) on the dataset and created a Power BI dashboard to visualize the insights derived from the data. Additionally, I have conducted an in-depth analysis of pizza sales, uncovering valuable patterns and trends. The Power BI dashboard provides a comprehensive view of the data, offering actionable insights that contribute to a better understanding of both the dataset and the pizza sales dynamics.

DATA VISUALIZATION

We need to analyze key indicators for our pizza sales to gain insights into our business performance. Specially, we want to calculate the following metrics:

1. Total Revenue: The sum of the total price of all pizzas order
2. Average Order Value: The average amount spent per order, calculated by dividing the total revenue by the total number of orders.
3. Total Pizzas Sold: The sum of the quantities of all pizzas sold.
4. Total Orders: The total number of orders placed.
5. Average Pizzas Per Order: The average number of pizzas sold per order, calculated by dividing the total number of pizzas sold by the total number of orders.

DATA VISUALIZATION

Chart Requirements:

We would like to visualize various aspects of our pizzas sales data to gain insights and understand key trends. We have identified the following requirements for creating charts:

1.Daily Trend for Total Orders: Create a bar chart that displays the daily trend of total orders over a specific time period. This chart will help us identify any patterns or fluctuations in order volumes on a daily basis.

2.Monthly Trend for Total Orders: Create a line chart that illustrate the hourly trend of total orders throughout the day. This chart will help us identify any patterns or fluctuations in order volume on a daily basis

3.Percentage of Sales by Pizza Orders:Create a pie chart that shows the distribution of sales across different pizza categories. This chart will provide insights into popularity of various pizza categories and their contributions to overall sales

DATA VISUALIZATION

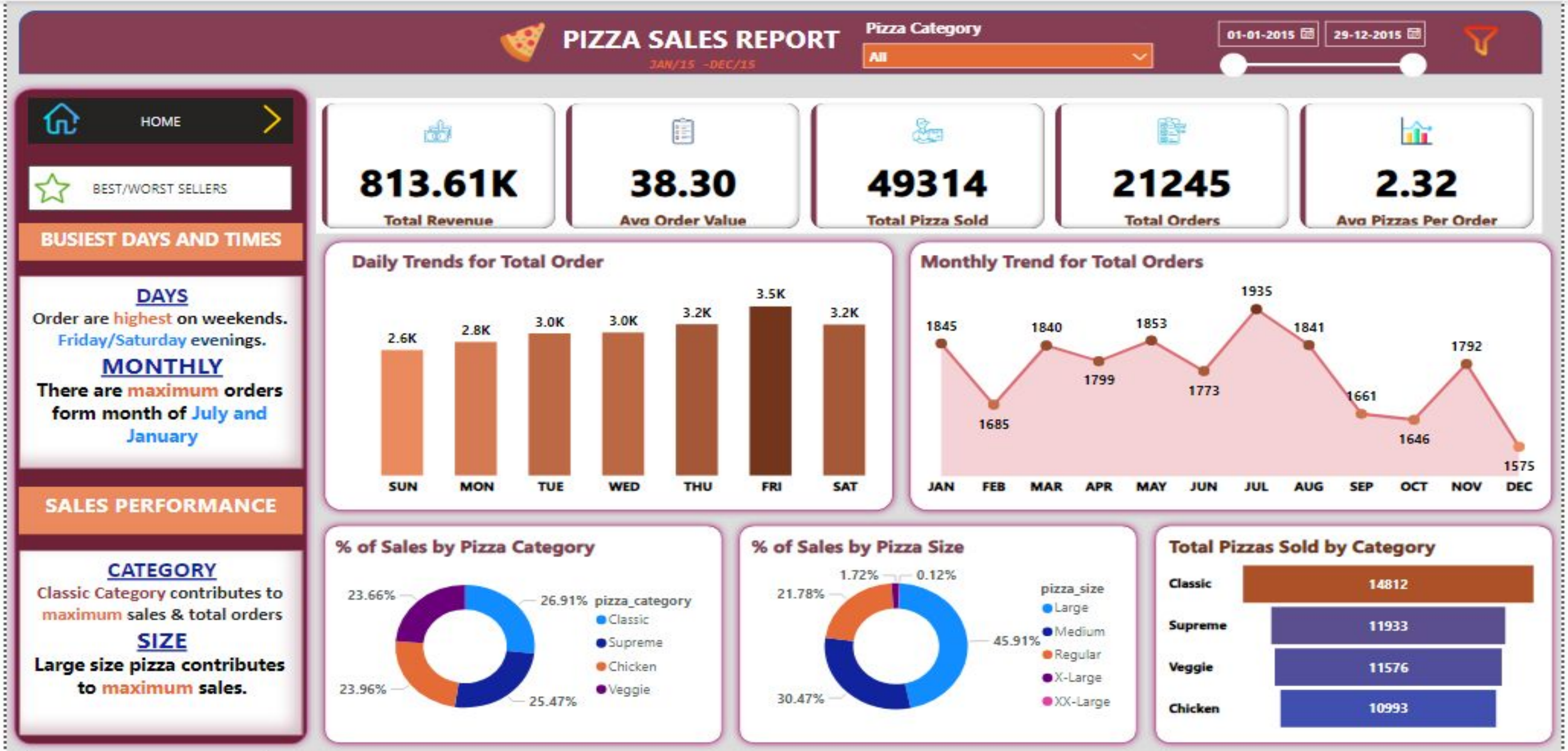
4.Percentage of Sales by Pizza Size: Generate a pie chart that represents the percentage of sales attributed to different pizza sizes. This chart will help us understand customer preferences for pizza sizes and their impact on sales.

5.Total Pizzas Sold by Pizza Category: Create a funnel chart that presents the total number of pizzas sold for each pizza category. This chart will allow us to compare the sales performance of different pizza categories,

6.Top 5 Best Sellers by Revenue, Total Quantity and Total Orders: Create a bar chart highlighting the top 5 best-selling pizzas based on the Revenue, Total Quantity, Total orders. This chart will help us identity the most popular pizza option.

7.Bottom 5 Best Sellers by Revenue, Total Quantity and Total Orders: Create a bar chart showcasing the bottom 5 worst-selling pizzas based on the Revenue, Total quantity, Total orders. This chart will enable us to identify underperforming or less popular pizza options.

DASHBOARD


PIZZA SALES REPORT
JAN/15 -DEC/15
Pizza Category
All
01-01-2015
29-12-2015
HOME
BEST/WORST SELLERS
BUSIEST DAYS AND TIMES
DAYS
Order are highest on weekends. Friday/Saturday evenings.
MONTHLY
There are maximum orders form month of July and January
SALES PERFORMANCE
CATEGORY
Classic Category contributes to maximum sales & total orders
SIZE
Large size pizza contributes to maximum sales.
813.61K
Total Revenue
38.30
Avg Order Value
49314
Total Pizza Sold
21245
Total Orders
2.32
Avg Pizzas Per Order
Daily Trends for Total Order
SUN 2.6K
MON 2.8K
TUE 3.0K
WED 3.0K
THU 3.2K
FRI 3.5K
SAT 3.2K
Monthly Trend for Total Orders
JAN 1845
FEB 1685
MAR 1840
APR 1799
MAY 1853
JUN 1773
JUL 1935
AUG 1841
SEP 1661
OCT 1646
NOV 1792
DEC 1575
% of Sales by Pizza Category
26.91%
25.47%
23.96%
23.66%
pizza_category
Classic
Supreme
Chicken
Veggie
% of Sales by Pizza Size
45.91%
30.47%
21.78%
1.72%
0.12%
pizza_size
Large
Medium
Regular
X-Large
XX-Large
Total Pizzas Sold by Category
Classic 14812
Supreme 11933
Veggie 11576
Chicken 10993

DASHBOARD


PIZZA SALES REPORT
JAN/15 -DEC/15
Pizza Category
All
01-01-2015
31-12-2015
BEST/WORST SELLERS
HOME
BEST SELLERS
REVENUE
The Thai Chicken Pizza Contributes to maximum Revenue
QUANTITY
The Classic Deluxe Pizza Contributes to maximum Total Quantities
TOTAL ORDERS
WORST SELLERS
REVENUE
The Brie Carre Pizza Contributes to minimum Revenue
QUANTITY
The Brie Carre Pizza Contributes to minimum Total Quantities
TOTAL ORDERS
The Brie Carre Pizza Contributes
817.86K
Total Revenue
38.31
Avg Order Value
49574
Total Pizza Sold
21350
Total Orders
2.32
Avg Pizzas Per Order
Top 5 Pizzas by Revenue
The Thai Chi... 43K
The Barbecu... 43K
The Californ... 41K
The Classic ... 38K
The Spicy It... 35K
Top 5 Pizzas by Quantity
The Classic ... 2.5K
The Barbecu... 2.4K
The Thai Chi... 2.4K
The Californ... 2.4K
The Spicy It... 1.9K
Top 5 Pizzas by Total Orders
The Classic ... 2.3K
The Hawaiia... 2.3K
The Pepper... 2.3K
The Barbecu... 2.3K
The Thai Chi... 2.2K
Bottom 5 Pizzas by Revenue
The Spinach... 16K
The Mediter... 15K
The Spinach... 15K
The Green G... 14K
The Brie Car... 12K
Bottom 5 Pizzas by Quantity
The Green G... 997
The Spinach... 970
The Spinach... 950
The Mediter... 934
The Brie Car... 490
Bottom 5 Pizzas by Total Orders
The Chicken... 938
The Calabre... 918
The Spinach... 918
The Mediter... 912
The Brie Car... 480

DATA ENCODING ON OBJECT VARIABLES

Label Encoding

- In this dataset, there are four columns that need label encoding as they contain categorical data representing rankings. Since these four columns are of object data type, I will perform label encoding to convert these categorical variables into numerical values.

One Hot Encoding

- I have applied one-hot encoding to the 'pizza_size' column. This transformation was done because 'pizza_size' contains categorical data, and one-hot encoding converts each categorical value into a separate binary column, making it suitable for machine learning algorithms.

DATA SCALING ON NUMERICAL VARIABLES

- In the given dataset, comprising three numerical attributes, one of which serves as the target variable, the "quantity" feature exhibits outlier values. Consequently, outlier mitigation techniques are not employed within this dataset preprocessing. However, for subsequent stages, we employ the Min-Max scaling algorithm to normalize the numerical features. Notably, despite the presence of outliers and the utilization of Min-Max scaling, the predictive performance of the machine learning model remains unaltered.

MODEL BUILDING

- Split train and test to build machine learning models.
- Then we will run different models on this data.

1) Linear Regression

2) Decision Tree

3) Random Forest

Accuracy Of Different Modules

Model	Train Score	Test Score	RMSE Score	MAE Score
Linear Regression	0.85	0.84	1.67	0.06
Decision Tree	0.83	0.82	1.83	0.06
Random Forest	0.99	0.99	0.14	0.00

Three different models were trained and evaluated on this data: Linear Regression, Decision Tree, and Random Forest. The Random Forest model outperformed the others with a high Train Score and Test Score, indicating strong predictive capabilities. Additionally, the RMSE and MAE scores for the Random Forest model are significantly lower than those of the other models, suggesting that its predictions are more accurate and closer to the actual sales values.

In conclusion, the Random Forest model's strong predictive performance offers valuable insights for optimizing various aspects of the pizzeria's operations. By utilizing the model's predictions and understanding the factors driving sales, the pizzeria can enhance inventory management, marketing strategies, menu offerings, and overall customer satisfaction, ultimately leading to increased business revenue.

FEATURE WORK

- In this dataset, we possess a comprehensive record of individual pizza orders, encompassing various attributes like size, category, ingredients, and more. To further enrich this dataset from a business standpoint, we could consider gathering additional information such as customer preferences, delivery times, and feedback. By incorporating such supplementary data, we can enhance the accuracy and effectiveness of our pizza sales prediction model.

- When we amalgamate this supplemental data with the existing pizza sales dataset, the resultant model becomes more robust in its ability to forecast sales trends, customer preferences, and peak demand periods. Consequently, pizzerias can leverage this predictive capability to optimize ingredient procurement, adjust staffing levels, create targeted marketing campaigns, and innovate their menu offerings.

- Furthermore, understanding the impact of different attributes on pizza sales allows pizzerias to create personalized experiences for their customers. For instance, by recognizing that a significant number of customers prefer vegetarian options during certain times, the pizzeria can promote veggie-based specials during those periods, potentially driving higher sales and customer satisfaction.

FEATURE WORK

- In conclusion, integrating additional relevant data into the pizza sales dataset can amplify the accuracy and utility of predictive models. By considering various factors like customer behavior, trends, and preferences, pizzerias can strategically position themselves to anticipate market shifts, optimize operations, and ultimately thrive in the competitive food industry landscape.

Business Perspectives and Insights:

Optimized Inventory and Production Planning: The accurate sales predictions from the Random Forest model can be used for efficient inventory management and production planning. By understanding which pizzas are likely to sell more, the pizzeria can ensure that they have the right ingredients on hand and avoid excess waste.

Tailored Marketing Strategies: With insights into which pizzas are popular and which attributes contribute to higher sales, the pizzeria can tailor their marketing strategies. Promotions, advertisements, and specials can be designed around the most popular pizza categories and attributes.

Menu Optimization: Understanding the impact of different ingredients and pizza attributes on sales allows for menu optimization. The pizzeria can focus on enhancing or promoting the pizzas that customers are more likely to order, potentially boosting revenue.

Business Perspectives and Insights:

Customer Satisfaction and Loyalty: By analyzing the data, the pizzeria can identify customer preferences and patterns. This can lead to the creation of customer-favorite combinations or personalized options that resonate with their audience, enhancing customer satisfaction and loyalty.

Demand Forecasting: The model's predictive power can assist in forecasting future demand, helping the pizzeria make informed decisions about staffing, supply orders, and overall business strategies.

Sales Improvement: Based on the Random Forest model's strong performance, it's evident that attributes such as pizza size, category, and ingredients play a significant role in sales. By focusing on these attributes, the pizzeria can experiment with new combinations and offerings that align with customer preferences.

Special Promotions: The insights from the model can guide the creation of special promotions centered around high-performing pizza categories or attributes. Limited-time offers or discounts on popular pizzas can attract more customers and increase sales.

Customer Segmentation: The pizzeria can segment their customer base based on order patterns and preferences. This segmentation can inform targeted marketing efforts and customization options for different customer groups.

THANK YOU

PERSONAL INFO:

Name : C.KOWSALYA

Batch : PGDA 03