EXPLORATORY DATA ANALYSIS (EDA) PROJECT REPORT

1. Introduction

***Project Title**: Exploratory Data Analysis of Pizza Sales Dataset*

***Objective**:-*

The goal of this project is to perform exploratory data analysis on a pizza sales dataset to uncover insights into sales trends, customer preferences, and key factors influencing sales.

***Dataset**: The dataset used for this analysis includes the following columns:*

- *order_details_id*
- *order_id*
- *pizza_id*
- *quantity*
- *order_date*
- *order_time*
- *unit_price*
- *total_price*
- *pizza_size*
- *pizza_category*
- *pizza_ingredients*
- *pizza_name*
-

2. Methodology

2.1 Data Cleaning

- ***Missing Values****: Identified and handled missing values by either filling them with appropriate values or removing rows/columns with excessive missing data.*
- ***Data Types****: Converted columns to appropriate data types (e.g., date and time columns).*

2.2 Feature Engineering

- ***Date and Time Extraction****: Extracted day, month, and year from order_date and hour from order_time to create new features.*
- ***Aggregated Features****: Calculated aggregated metrics such as total sales and average unit price per pizza category.*

2.3 Data Exploration

- ***Descriptive Statistics****: Calculated summary statistics to understand the central tendency, dispersion, and distribution of the data.*
- ***Correlation Analysis****: Explored relationships between numerical variables to identify correlations.*

2.4 Visualization

- ***Histograms****: Used histograms to visualize the distribution of continuous variables like total_price and quantity.*
- ***Scatter Plots****: Created scatter plots to examine relationships between variables such as total_price vs. quantity.*
- ***Heatmaps****: Attempted to visualize correlations between numerical variables (if applicable).*

3. Findings

3.1 Descriptive Statistics

- ***Average Total Price****: The average total price of an order is $X.*
- ***Average Quantity****: The average quantity of pizzas ordered is Y.*

3.2 Key Insights from Visualizations

- ***Distribution of Total Price****: The histogram shows that the majority of orders have a total price in the range of $X to $Y.*
- ***Relationship Between Total Price and Quantity****: Scatter plots indicate a positive/negative correlation between total price and quantity.*

- ***Pizza Category Preferences****: The distribution of pizza categories shows that the most popular categories are A, B, and C.*

3.3 Correlation Analysis

- ***Correlation Between Variables****: Heatmap analysis reveals that total_price has a strong/weak correlation with quantity.*
-

4. Visualizations

4.1 Scatter Plot of Actual vs. Predicted Values

- ***Description****: This scatter plot compares actual vs. predicted values for the quantity of pizzas ordered. The green line represents the regression line.*

4.2 Histogram of Total Price

- ***Description****: This histogram displays the distribution of total prices in the dataset.*

4.3 Correlation Heatmap

- ***Description****: This heatmap shows the correlation between numerical variables.*

5. Conclusion

- ***Summary****: The EDA provided insights into pizza sales trends, customer preferences, and factors influencing total sales. Key findings include the distribution of total prices, relationships between price and quantity, and popular pizza categories.*
- ***Recommendations****: Based on the analysis, it is recommended to focus on improving offerings in popular pizza categories and consider pricing strategies to increase sales.*

6. Future Work

- ***Further Analysis****: Investigate other features such as customer demographics or time-based trends.*
- ***Advanced Modeling****: Apply predictive modeling techniques to forecast future sales and customer behavior.*

7. References

- ***Dataset Source****: [Link to dataset or source]*
- ***Tools Used****: Python, Pandas, NumPy, Matplotlib, Seaborn*