Index - AS2021934

Name – K.P.R.D.Pitawala

Project Proposal: Data Science for Coffee Shop Sales Analysis

1. Executive Summary

This project aims to analyze coffee shop sales data using data science techniques to uncover sales trends, customer purchasing behaviors, and potential business insights. The findings will help optimize product recommendations, forecast sales, and improve customer experience.

2. Problem Statement

Understanding customer preferences and sales patterns is crucial for coffee shops to optimize inventory, maximize revenue, and enhance customer satisfaction. This project will address challenges such as fluctuating demand, product bundling, and sales forecasting.

3. Objectives

- Identify sales trends and peak sales periods.
- Analyze customer purchasing behavior.
- Perform product affinity analysis to recommend frequently purchased product combinations.
- Predict future sales trends using machine learning models.
- Provide actionable insights to improve business strategies.

4. Business Impact

- Optimized inventory management to reduce waste.
- Improved sales forecasting for strategic planning.
- Enhanced customer experience through personalized recommendations.
- Increased revenue through better pricing and promotional strategies.

5. Literature Review / Related Work

Several studies have shown that data-driven decision-making in the food and beverage industry improves profitability and customer retention. Techniques like association rule mining and time series forecasting have been effectively used in retail analytics to boost sales.

6. Methodology

- **Data Preprocessing**: Cleaning and transforming the dataset, handling missing values, and ensuring data quality.
- **Exploratory Data Analysis (EDA)**: Visualizing data distribution, identifying patterns, and detecting anomalies.
- **Association Rule Mining**: Implementing techniques such as Apriori or FP-Growth to identify frequently purchased product combinations.
- **Time Series Analysis**: Using models like ARIMA or LSTM to forecast future sales trends.
- **Clustering & Segmentation**: Applying K-Means or DBSCAN to segment customer purchasing behavior.
- **Predictive Modeling**: Implementing regression or classification models to predict sales trends and customer preferences.
- **Visualization & Reporting**: Creating dashboards and reports to present insights.

7. Risk Analysis & Mitigation

- **Data Quality Issues**: Address by thorough preprocessing and validation.
- **Model Accuracy Concerns**: Optimize hyperparameters and test multiple algorithms.
- **Scalability**: Ensure models can handle future growth in data volume.
- **Ethical Considerations**: Ensure customer data privacy and compliance with regulations.

8. Ethical Considerations & Data Privacy

- Compliance with GDPR and data protection laws.
- Anonymization and encryption of sensitive customer information.
- Ethical use of data without bias or discrimination.

9. Project Timeline (1.5 Months)

Week	Task Description
1	Data collection, cleaning, and preprocessing
2	Exploratory data analysis and visualization
3	Association rule mining and product recommendation analysis
4	Time series forecasting and customer segmentation
5	Predictive modeling and performance evaluation
6	Visualization, report preparation, and finalization

10. Deliverables & Milestones

- Cleaned and preprocessed dataset.
- EDA report with key insights.
- Association rule mining and clustering results.
- Time series forecasting models.
- Predictive models with performance metrics.
- Interactive dashboards and final report.

11. Tools & Technologies

- Python (Pandas, NumPy, Scikit-learn, TensorFlow, Matplotlib, Seaborn)
- SQL for data querying
- Tableau or Power BI for visualization

12. Budget & Resource Requirements

- Cloud computing services (AWS, GCP, or Azure) for model training and storage.
- Software tools for visualization and machine learning.
- Access to additional customer behavioral datasets (if required).

13. Conclusion

This project will provide data-driven insights to improve coffee shop operations and enhance customer experiences through data science techniques. The findings will be valuable for decision-making, revenue optimization, and customer retention.