Project 1: Coffee Sales Analysis

Project Objective

The goal of the project was to analyse coffee sales data to uncover trends, customer preferences, and insights. The project also aimed to create a predictive model for sales, leveraging machine learning techniques, and provide actionable recommendations to optimize inventory, sales strategies, and customer satisfaction.

Approach

The project followed a structured data analysis and machine learning workflow:

1. **Data Exploration and Cleaning**
    - Loaded and explored the dataset.
    - Handled missing values and ensured data types were correct.
    - Conducted feature engineering to extract meaningful attributes.
2. **Exploratory Data Analysis (EDA)**
    - Conducted in-depth analysis using visualizations and statistical methods to identify trends.
3. **Machine Learning Modelling**
    - Implemented and evaluated three machine learning models: Linear Regression, Decision Tree Regressor, and Random Forest Regressor.
    - Tuned model parameters to enhance predictive performance.
4. **Insights and Recommendations**
    - Synthesized insights from analysis and modelling.
    - Provided actionable recommendations for business optimization.

Steps & Results

Data Handling

- **Dataset Summary**:
    - **1133 rows** with information on sales transactions including date, payment method, product, and sales amount.
    - Missing values in the card column were replaced with 'Unknown'.
    - Extracted additional features like month, day_of_week, and hour for enhanced analysis.
- **Data Validation**:
    - Checked for duplicates (none found).
    - Data types were updated to datetime where applicable.

Exploratory Data Analysis (EDA)

Revenue Analysis by Product:

- Top Performers: Latte and Americano with Milk generated the highest revenue.
- Underperformers: Cocoa and Espresso had the least sales.

Monthly Sales Trends:

- Revenue showed steady growth from March to July 2024, with variations in product performance across months.

Distribution of Sales:

- Most transactions were concentrated in the range of **28–38 currency units**.

Payment Method Preferences:

- 92% of transactions were via **card payments**, reflecting a cashless customer base.

Sales by Day of the Week:

- **Tuesdays** had the highest sales, indicating increased post-weekend activity.

Hourly Trends:

- Peak hours: **Morning (10 AM)** and **Evening (7 PM)**, reflecting customer habits.

Correlation Analysis:

- Strong correlation between coffee type, time, and sales revenue.

Machine Learning Models

Model Implementation

- Models evaluated:
    - Linear Regression
    - Decision Tree Regressor
    - Random Forest Regressor

Model Performance

- Linear Regression:
    - $R^2 = 0.83$, MSE = 3.23
- Decision Tree Regressor:
    - $R^2 = 0.79$, MSE = 3.97
- Random Forest Regressor:
    - $R^2 = 0.80$, MSE = 3.75

Hyperparameter Tuning

- Random Forest Regressor (Best Parameters):
    - max_depth: 10, min_samples_split: 10, n_estimators: 50
    - Final $R^2 = 0.83$, MSE = 3.19

Feature Importance

- Most impactful features: Coffee type, hour of purchase, and weekend indicator.

Key Insights

The analysis of coffee sales data revealed several significant trends and behaviours that are crucial for understanding customer preferences and optimizing business operations. These insights can guide strategic decisions to enhance performance and customer satisfaction:

-Sales Trends and Seasonality:

- Sales patterns demonstrate **two daily peak periods**, one in the **morning around 10 AM** and another in the **evening around 7 PM**. These periods correspond to times when customers are likely to seek refreshments, such as before starting work or winding down for the day.
- Weekly sales trends show that **Tuesdays** generate the highest sales volume, suggesting increased customer activity early in the workweek. In contrast, weekends see comparatively lower sales, offering an opportunity for targeted promotions.

-Product Performance:

- **Latte** and **Americano with Milk** are the most popular products, contributing significantly to revenue. This indicates strong customer preference for creamy and milk-based coffee options.
- **Cocoa** and **Espresso** are the least purchased items, highlighting a need to either adjust their pricing, improve their visibility, or market them to a niche audience.
- Product popularity varies by time of day; for example, **Latte** is more preferred in the morning, while **Hot Chocolate** and **Cappuccino** are favored during evening hours.

-Customer Payment Behaviour:

- A notable 92% of transactions are made through **card payments**, showcasing a strong preference for cashless transactions. This emphasizes the importance of having a reliable card payment system and the potential to introduce **mobile payment options** for added convenience.

-Revenue Distribution:

- Most transactions are concentrated in the sales range of **28–38 currency units**, reflecting consistent spending patterns among customers.

-Time-Based Insights:

- The hour of purchase significantly influences sales trends. Morning and evening peaks suggest opportunities for targeted promotions during these hours to capitalize on customer footfall.
- Day-of-week analysis indicates that **weekdays generally outperform weekends**, likely due to routine office-goers purchasing coffee.

-**Feature Importance in Predictive Modeling**:

- In the machine learning analysis, features like **coffee type**, **hour of the day**, and whether the purchase was made on a **weekend** emerged as the most impactful in predicting sales performance.

-**Correlation and Relationships**:

- Correlation analysis revealed that numeric features like hour and money have meaningful relationships, providing insights into how time-based attributes influence spending patterns.

These insights form a strong foundation for actionable recommendations, enabling the business to address customer needs effectively, optimize resources, and enhance sales performance. By leveraging these findings, the business can make informed, data-driven decisions to achieve sustainable growth.

Recommendations

-**Inventory Optimization**:

- Stock popular products like Latte during peak hours.
- Increase inventory of evening favorites like Cappuccino and Hot Chocolate.

-**Promotional Strategies**:

- Launch discounts for low-performing items like Cocoa and Espresso.
- Offer loyalty rewards to boost off-peak sales.

-**Payment System Enhancements**:

- Maintain a seamless card payment experience.
- Introduce mobile payment options for added convenience.

-**Data-Driven Adjustments**:

- Regularly analyse sales data to adapt to customer preferences.

Conclusion

The comprehensive analysis of coffee sales data provided valuable insights into customer behaviors, product performance, and sales trends. By leveraging advanced data exploration techniques and predictive modeling, the project revealed key factors driving revenue and customer preferences.

The data showcased that **sales were highly influenced by time and day**, with peak periods in the **morning (10 AM)** and **evening (7 PM)** and the highest sales volumes occurring on **Tuesdays**. These patterns emphasize the need for inventory and operational adjustments to cater to customer demands during these high-traffic periods.

Product performance analysis indicated that **Latte and Americano with Milk** were the top-selling items, contributing significantly to overall revenue. Conversely, products like **Cocoa and Espresso** underperformed, suggesting an opportunity for targeted promotional strategies to boost their sales. Additionally, customer payment preferences overwhelmingly leaned toward **card transactions**, with 92% of payments made via cards. This highlights the importance of maintaining reliable and efficient cashless payment systems, potentially integrating mobile payment options to enhance convenience further.

From a machine learning perspective, the **Random Forest Regressor** emerged as the most robust model for predicting sales, achieving an **R^2 score of 0.83** after hyperparameter tuning. Feature importance analysis highlighted that **coffee type, time of purchase, and whether the transaction occurred on a weekend** were critical factors influencing sales. These insights offer a deeper understanding of customer purchasing behavior, enabling more precise sales forecasting and strategy formulation.

The findings from this project pave the way for actionable recommendations, such as optimizing inventory management to align with demand patterns, introducing promotional campaigns for underperforming products, and enhancing the overall customer experience. By implementing these strategies, the business can capitalize on customer preferences, reduce wastage, and maximize revenue potential.

In conclusion, the project underscored the power of data-driven decision-making in the coffee sales domain. By continuously monitoring and analyzing sales trends, the business can remain agile in responding to market demands and maintaining a competitive edge. The combination of analytical insights and predictive modeling positions the business for sustainable growth and improved customer satisfaction.