Capstone Project

Menu Sales Forecast Final Report

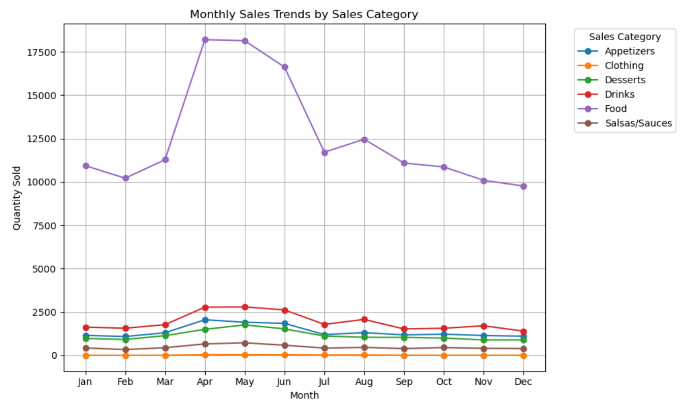
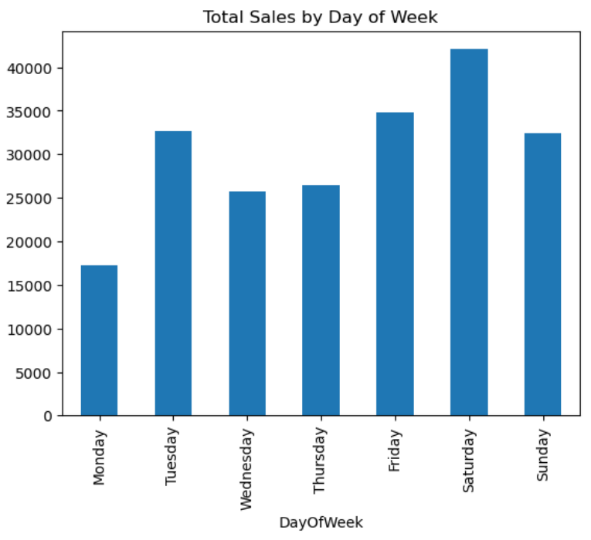
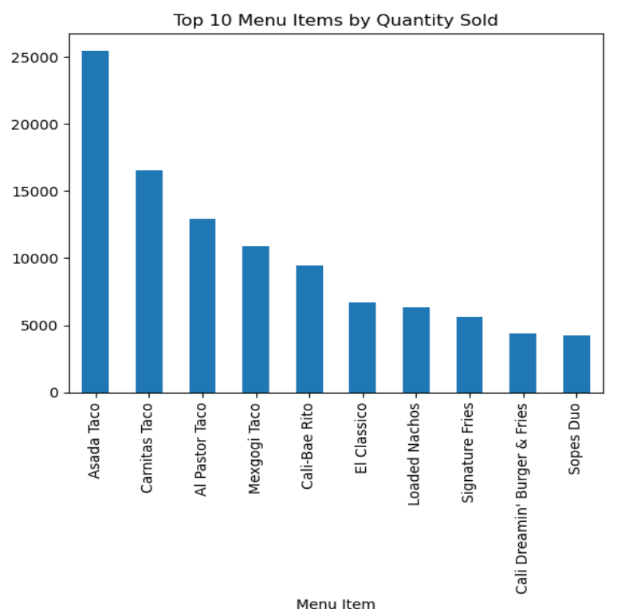
1. **Define the Problem Statement:** The objective of this project is to support data-driven decision-making for a restaurant by analyzing historical sales data. By identifying which menu items generate the highest revenue and understanding seasonal or quantity-based trends, we aim to forecast future sales performance. These insights will enable the restaurant to better manage inventory, reduce waste, optimize purchasing decisions, and ensure popular items remain in stock — ultimately improving operational efficiency and profitability.
2. **Model Outcomes or Predictions:** In this project, we used regression-based machine learning to forecast revenue generated by individual menu items. The goal was to predict the net price of items sold using historical data, taking into account key factors such as:

* Quantity sold (top 10)
* Seasonal trends (daily)
* Monthly sales performance

By training the model on past transaction data, we were able to generate reliable revenue estimates for future sales scenarios. These predictions support more informed decision-making around inventory planning, staffing, and promotional strategies, ultimately helping the business improve operational efficiency and reduce waste.

1. **Data Acquisition:** We utilized a comprehensive sales dataset covering transactions from 2023 to 2025. The data included key details such as order date, item description, quantity sold, and net price. This dataset provided a strong foundation for identifying high-performing menu items, analyzing seasonal trends, and uncovering patterns in customer purchasing behavior.

Visualizations below were created to highlight which items contributed most to revenue and during which time periods and insights that are critical for forecasting and strategic planning.



1. **Data Preprocessing/Preparation:** Before building predictive models, the dataset was carefully cleaned and prepared to ensure accuracy and reliability. Key steps included:

* Data Cleaning: We removed voided (canceled) transactions and standardized the price column by converting currency values to numeric format. We also checked for and handled missing or inconsistent entries to avoid errors during modeling.
* From the order dates, we extracted new features like Month and Year to help the model recognize seasonal trends.
* Data Splitting: The dataset was divided into two parts — a training set to teach the model using past data, and a test set to evaluate how well it performs on unseen data. This approach helps ensure the model’s predictions are generalizable and not just memorizing old patterns.

These preprocessing steps were essential in preparing the data for machine learning and enabling accurate revenue forecasting.

1. **Modeling:** To forecast future menu item revenue, we applied two machine learning models: **Linear Regression and Random Forest**

Linear Regression helped identify clear, linear trends in the data, making it useful for understanding baseline relationships between time, quantity sold, and price.

Random Forest model captured more complex, nonlinear patterns by combining multiple decision trees. This approach allowed for more nuanced predictions, especially in cases where customer behavior or sales patterns varied over time.

Using both models enabled us to compare performance and choose the most effective strategy for revenue prediction — with the goal of improving planning accuracy and decision-making.

1. **Model Evaluation:** Because our goal was to predict revenue values, we used regression models to evaluate performance. To assess how well each model performed, we used two key metrics:

* R² Score: Measures how well the model explains the variation in sales data. Higher values indicate stronger predictive power.
* Mean Squared Error (MSE): Indicates the average difference between the model’s predictions and actual sales. Lower values reflect higher accuracy.

After evaluating both models, the Random Forest consistently outperformed Linear Regression across both metrics. Based on its results, we recommend Random Forest for future revenue forecasting tasks.