**Predictive Analysis Report on Customer Lifetime Value, Spend Probability, and Behavior Analysis**

**Introduction**

This report provides a comprehensive analysis of customer behavior and spending patterns using predictive modeling techniques. The project leverages the **CDNow Dataset** to address critical business questions about **customer retention, spending probability, and lifetime value**. By utilizing advanced **feature engineering, machine learning models, and detailed visualizations,** the report provides actionable insights to guide **strategic decision-making**.

**Dataset Overview**

The dataset used is the CDNow Dataset, offering real-world transactional data that includes:

- **Customer ID:** Unique identifier for each customer.

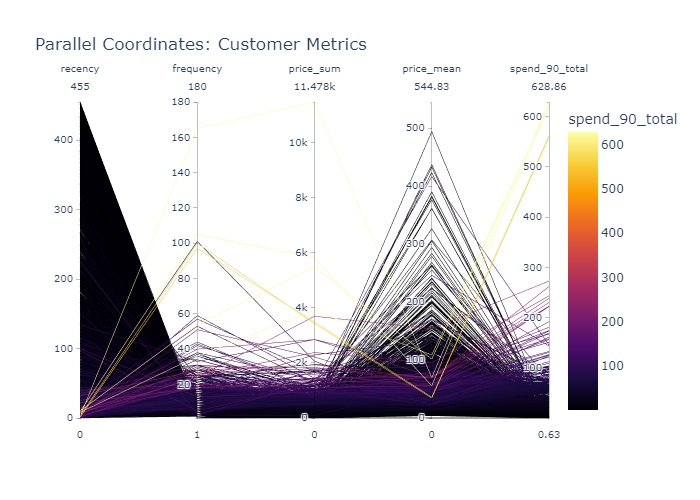
- **Purchase Date:** Date of each transaction.

- **Purchase Quantity**: Items purchased per transaction.

- **Order Value:** Total value of each transaction.

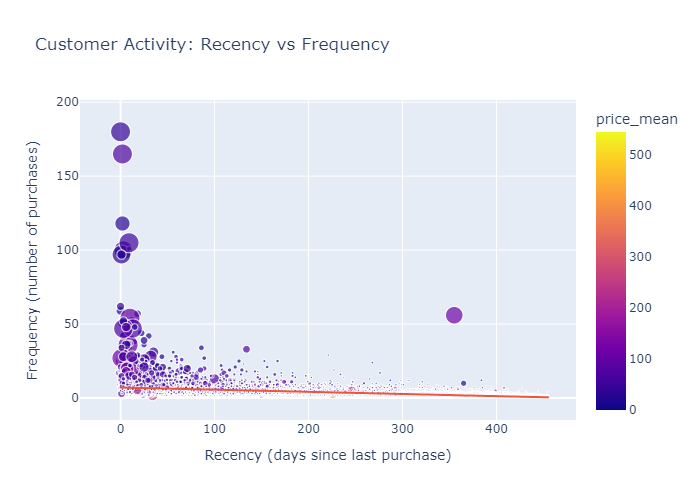
**Customer Metrics**

1. **Customer Metrics Including Predicted CLV - Parallel Coordinates**



**Analysis Summary:** The parallel coordinates plot visualizes multiple customer metrics and their relationship with each other, helping to identify key factors that drive customer value.

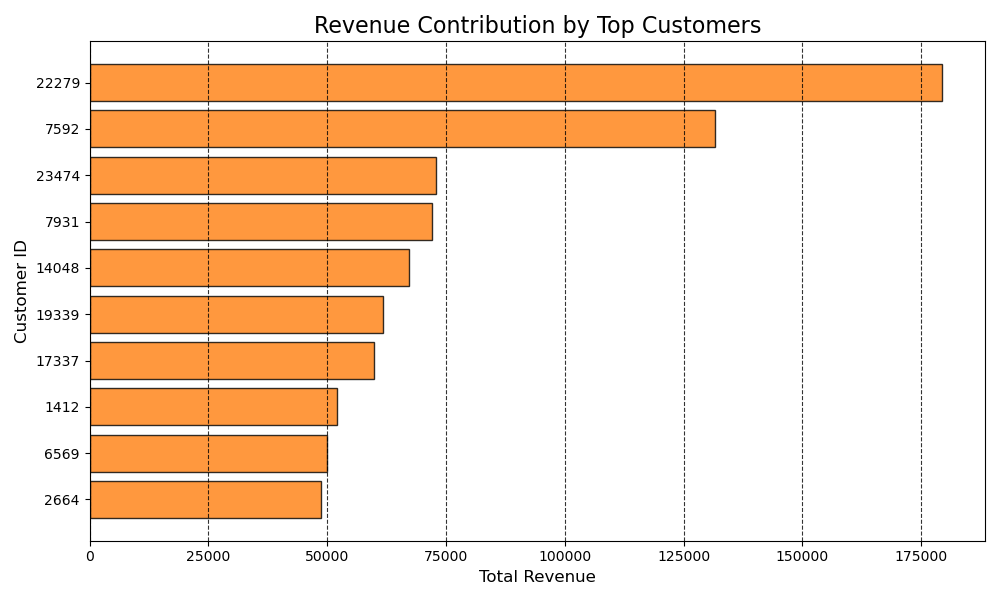
1. **Customer Activity Trendline**



**Analysis Summary:** The trendline of customer activity over time provides insights into the overall engagement levels and helps in identifying periods of high or low activity. Here the trendline shows a slight increase in the frequency of purchases.

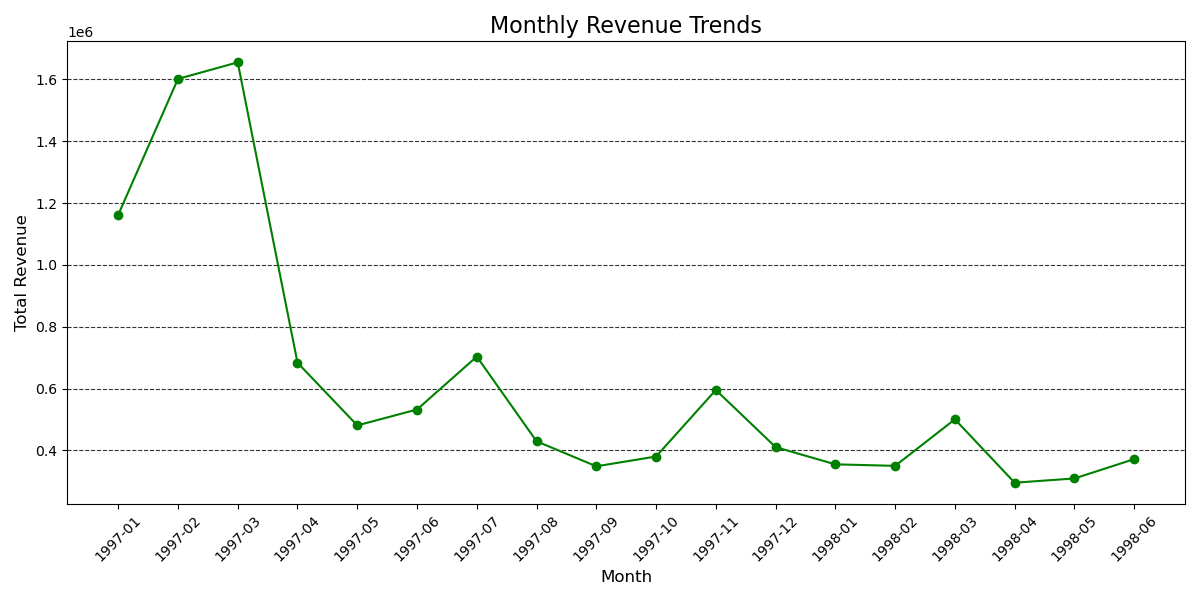
**Insights from Dataset Analysis**

1. **Revenue Contribution by Top Customers**



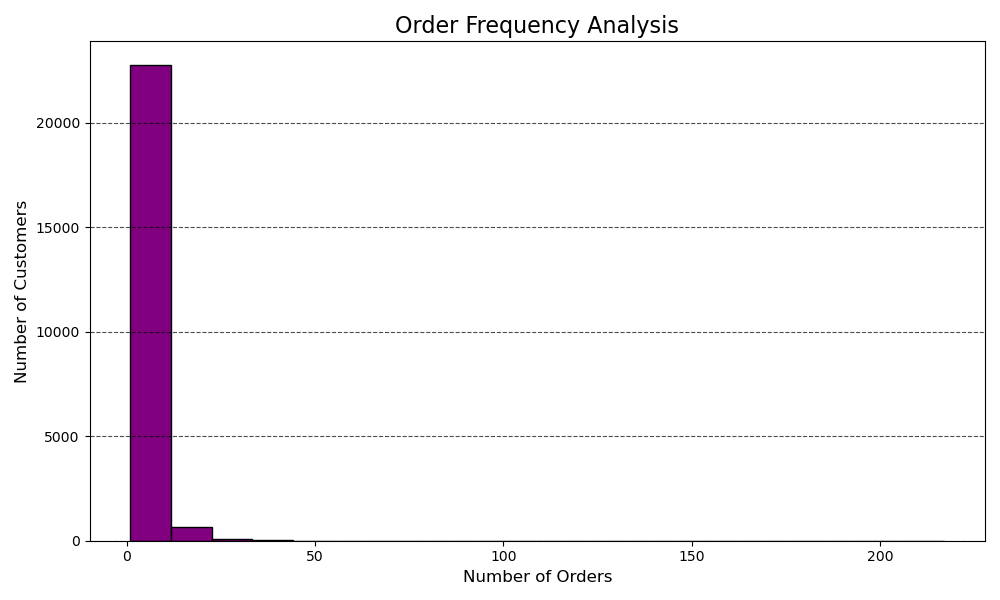
**Analysis Summary**: The top 10 customers contribute significantly to the overall revenue, highlighting the importance of focusing retention efforts on high-value customers.

**2. Monthly Revenue Trends**



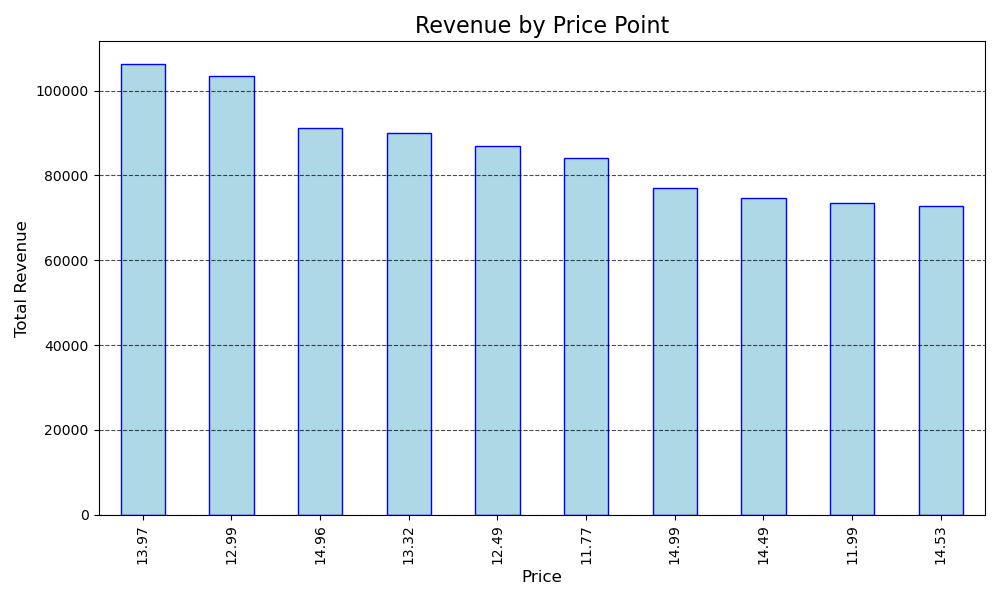
**Analysis Summary**: Monthly revenue trends show seasonal variations and peaks during specific periods, which can inform marketing strategies. Here the graph shows decrease in the overall revenue, which makes sense as the dataset we are working with is from a company that went bust.

**3. Order Frequency Analysis**



**Analysis Summary:** Understanding order frequency helps in identifying customer purchasing patterns and predicting future behavior. The company has more customers which have low number of orders which indicates that we should work getting existing customers to repurchase.

**4. Revenue by Product Price**



**Analysis Summary**: Higher-priced products contribute more to revenue, indicating the potential for upselling and cross-selling strategies.

**Feature Engineering**

**Objectives**

- Create features encapsulating customer behavior, such as recency, frequency, and monetary value (RFM).

- Generate advanced predictors like average spending, purchase trends, and segmentation metrics.

**Key Features Created**

- **Recency:** Days since the last purchase.

- **Frequency:** Total number of purchases.

- **Monetary Metrics:**

- **price\_sum:** Total spend across all transactions.

- **price\_mean:** Average spend per transaction.

- **spend\_90\_total:** Predicted customer lifetime value (CLV) for the next 90 days.

**Methods**

- Aggregation and grouping based on customer\_id.

- Handling missing data and outliers.

- Normalization and scaling of numerical features to enhance model performance.

**Temporal Splitting**

- Split the dataset into two segments:

- 90 days: Used for short-term CLV predictions.

- Remaining period: Used for training and evaluation.

**Predictive Models**

1. **Spend Amount Prediction Model**

**Objective**: Predict customer spend over the next 90 days.

**Model:** Gradient Boosting Regressor (XGBoost)

**Key Steps:**

**1. Feature Selection:** Utilized RFM metrics, purchase trends, and segmentation data.

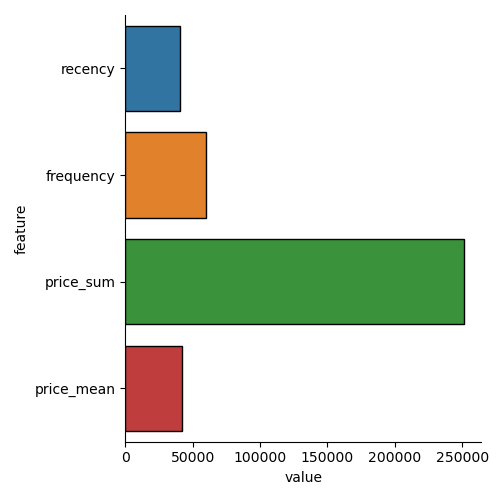
**2. Evaluation Metrics:**

- Mean Absolute Error (MAE): ~9.65 units. (9-10 dollars)

- Root Mean Square Error (RMSE).

**3. Feature Importance Analysis:**

- Key drivers of customer spending included Recency, Frequency, and price\_sum.



**PTO**

1. **Spend Probability Prediction Model**

**Objective:** Classify customers likely to make a purchase in the next 90 days.

**Model:** Gradient Boosting Classifier (XGBoost)

**Key Steps:**

**1. Label Encoding:** Binary classification for spend\_90\_flag.

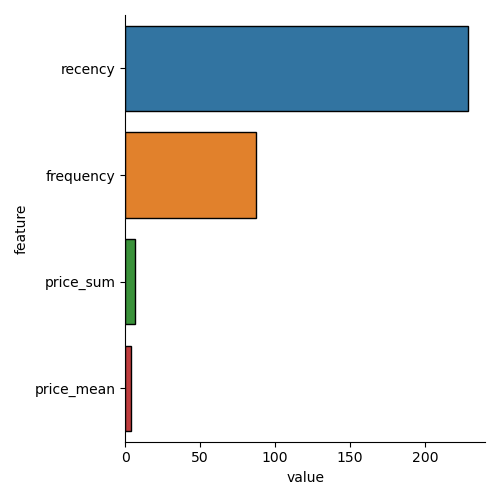
**2. Metrics:**

- Accuracy: ~84%

- ROC-AUC: 0.84.

**3. Feature Importance:**

- Behavioral metrics like Recency and Frequency were top contributors.

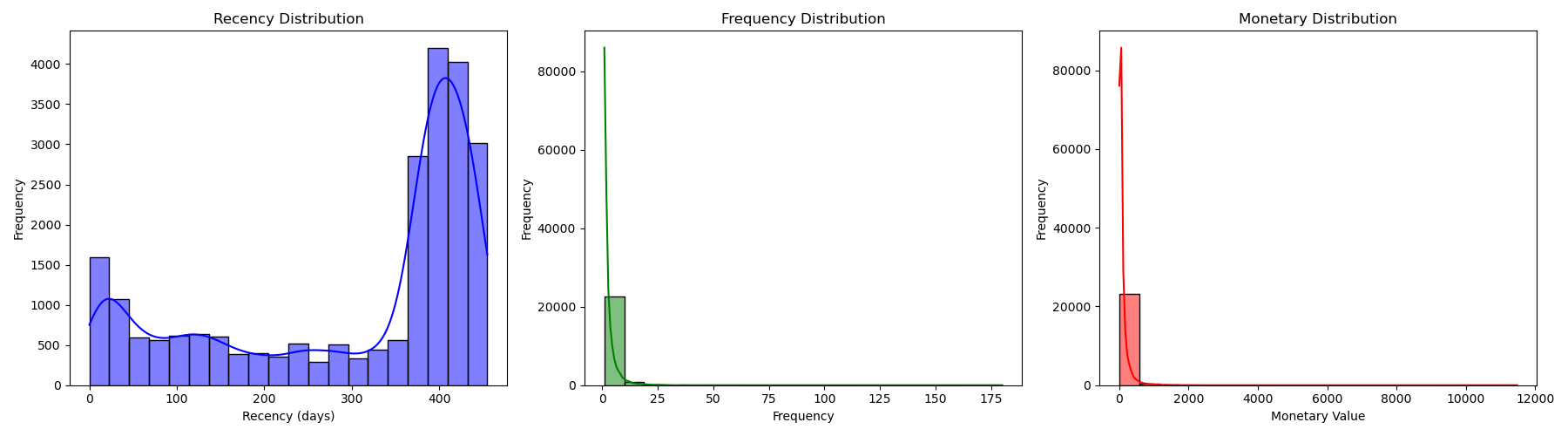


**PTO**

**Analysis**

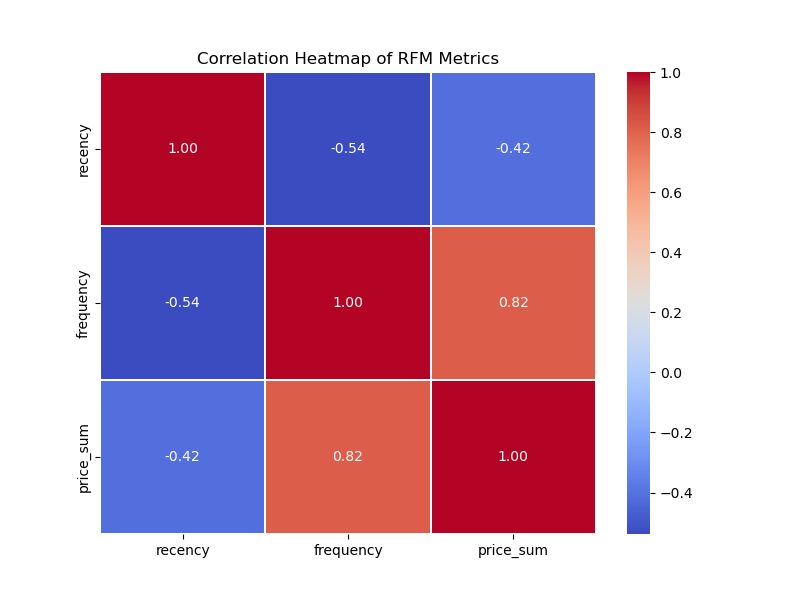
**RFM Analysis**

**1. Distribution of RFM Metrics**



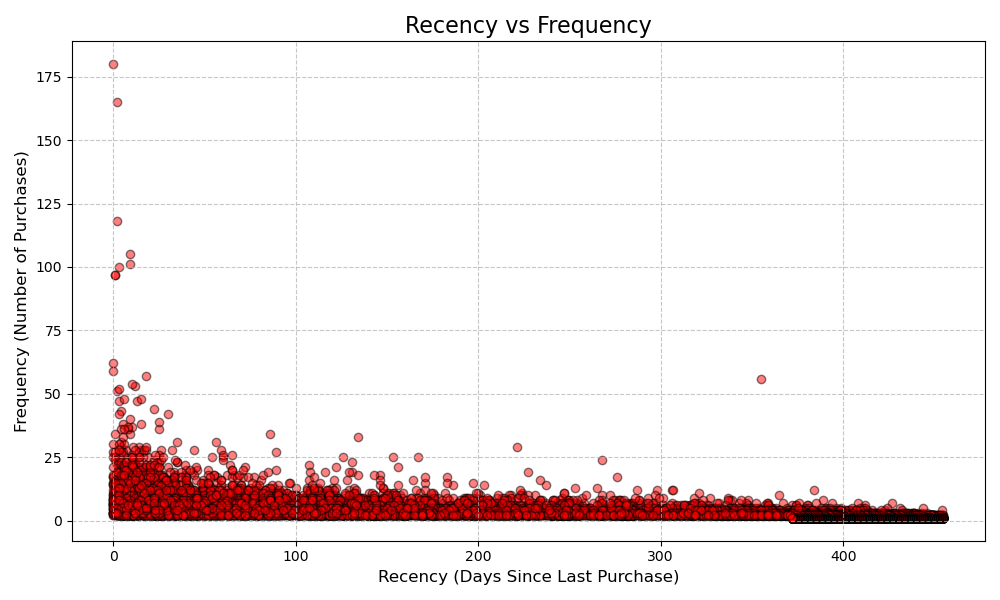
**Analysis Summary:** The distribution of Recency, Frequency, and Monetary metrics provides insights into customer segmentation and behavior.

**2. Relationship Between RFM Metrics**



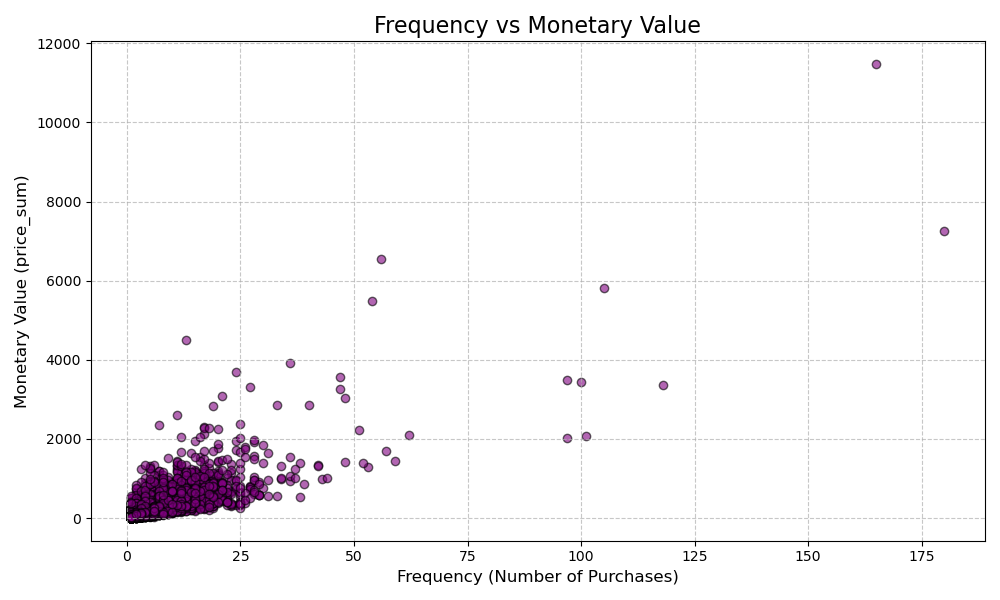
**Analysis Summary:** The heatmap shows the correlation between RFM metrics, helping to identify key drivers of customer value.

**3. Customer Spread by Recency and Frequency**



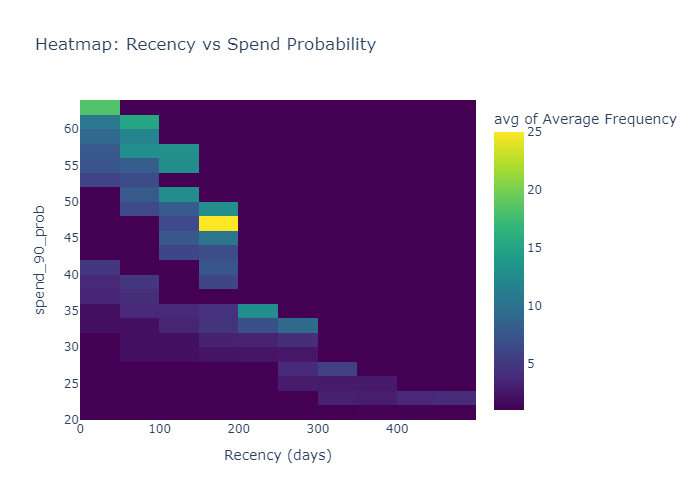
**Analysis Summary:** Visualizing customer spread by recency and frequency helps in identifying loyal customers and those at risk of churn.

**4. Customer Spread by Monetary Value and Frequency**



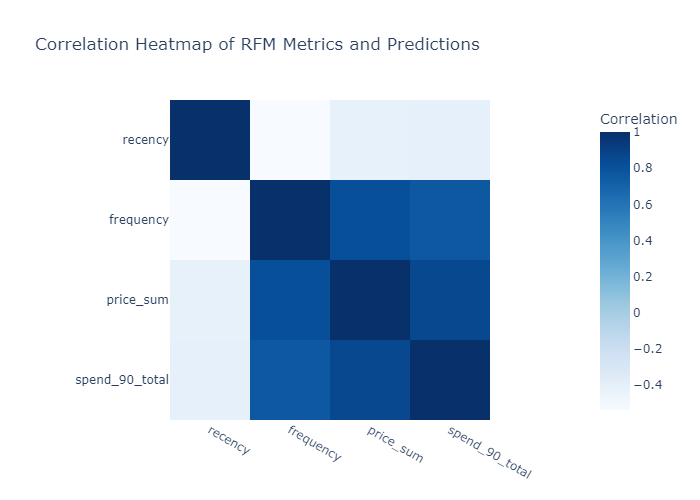
**Analysis Summary:** This visualization highlights the relationship between purchase frequency and monetary value, aiding in customer segmentation.

5. **Recency vs Spend Probability Heatmap**



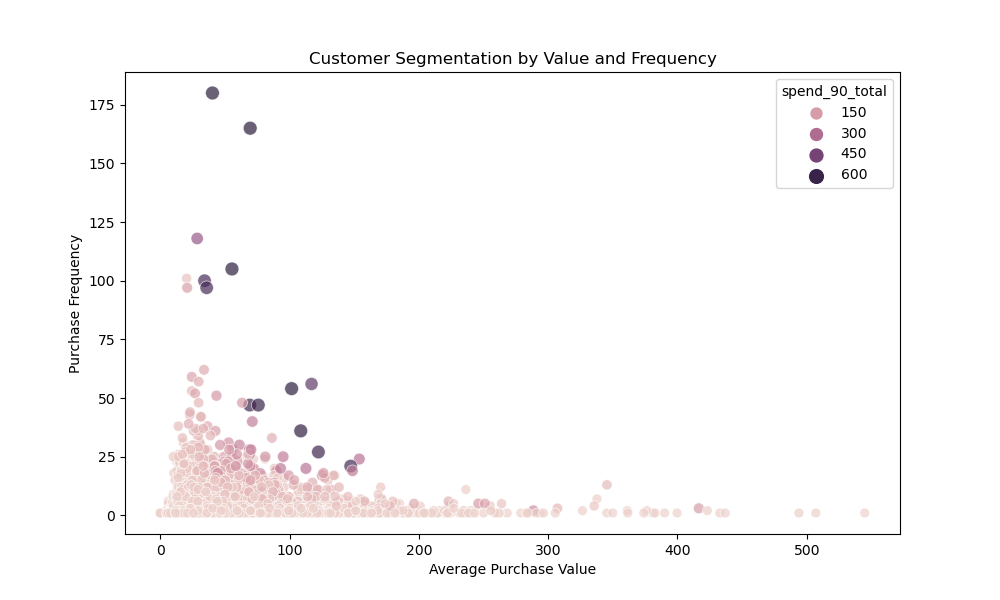
**Analysis Summary:** The heatmap illustrates the correlation between recency and spend probability, helping to identify how recent purchases influence the likelihood of future spending.

6. **RFM Predicted Spend Correlation Heatmap**



**Analysis Summary:** This heatmap shows the correlation between RFM metrics and predicted spend, providing insights into which metrics are most strongly associated with future spending.

7. **Segmentation by Value and Frequency**

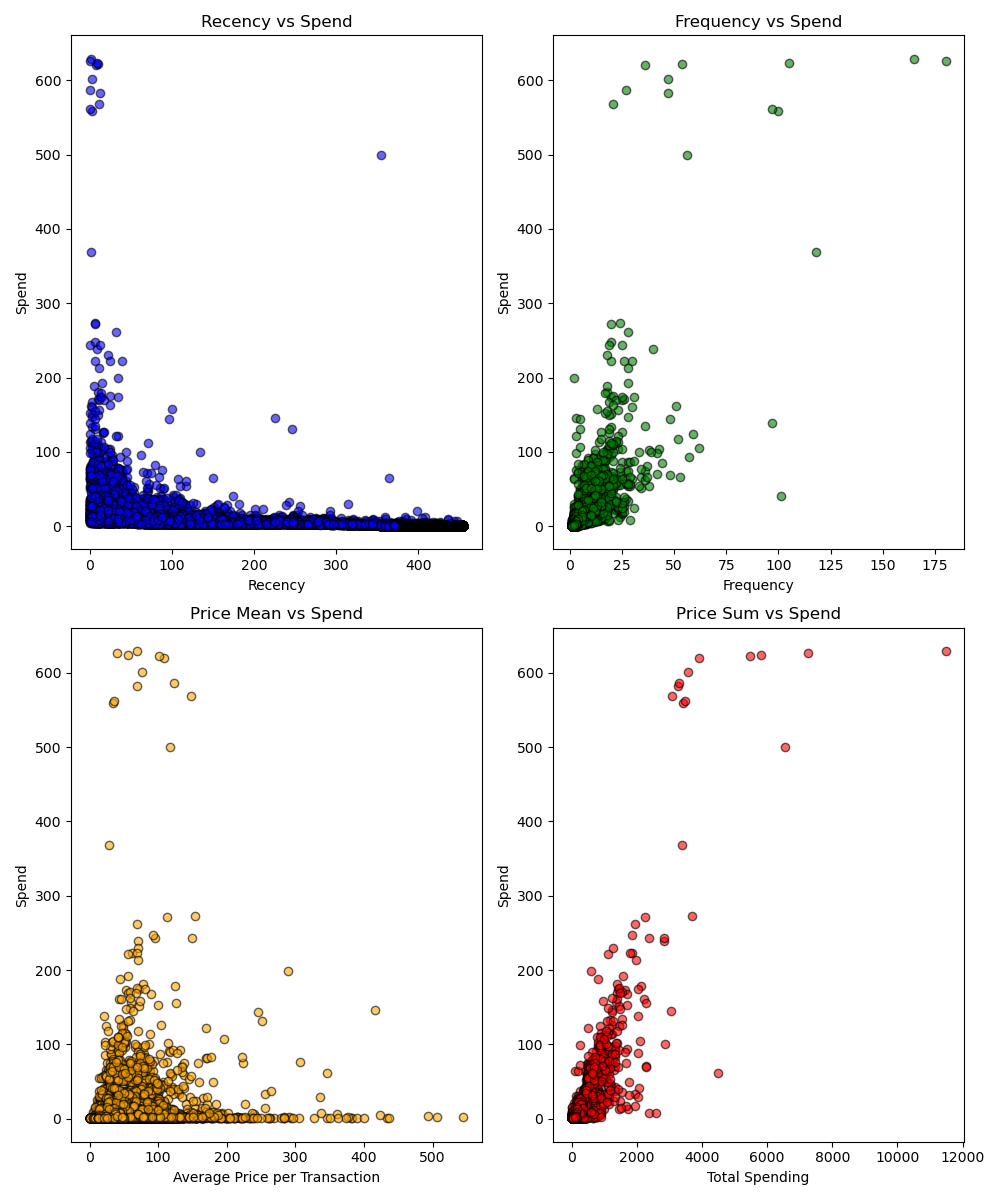


**Analysis Summary:** This visualization highlights the segmentation of customers based on their monetary value and purchase frequency, aiding in identifying high-value and high-frequency customers.

**PTO**

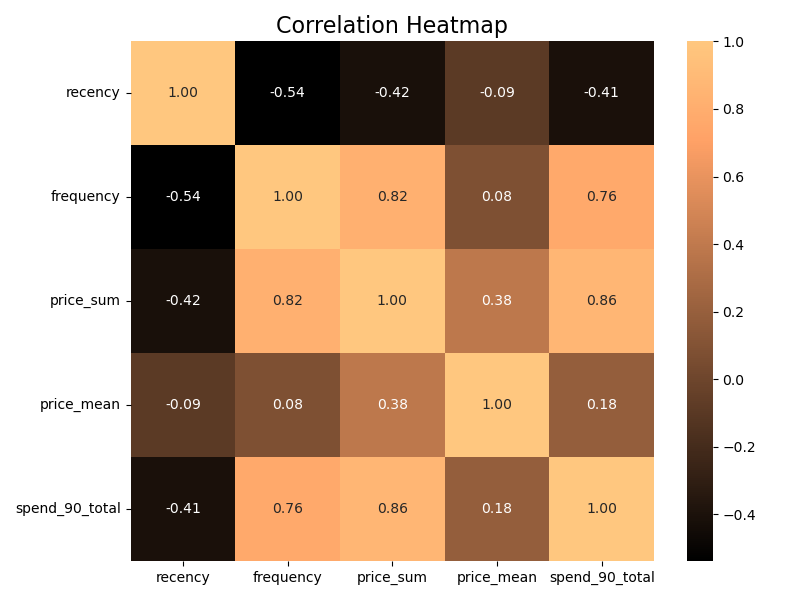
**CLV Analysis**

**1. Predicted Spend vs Features**



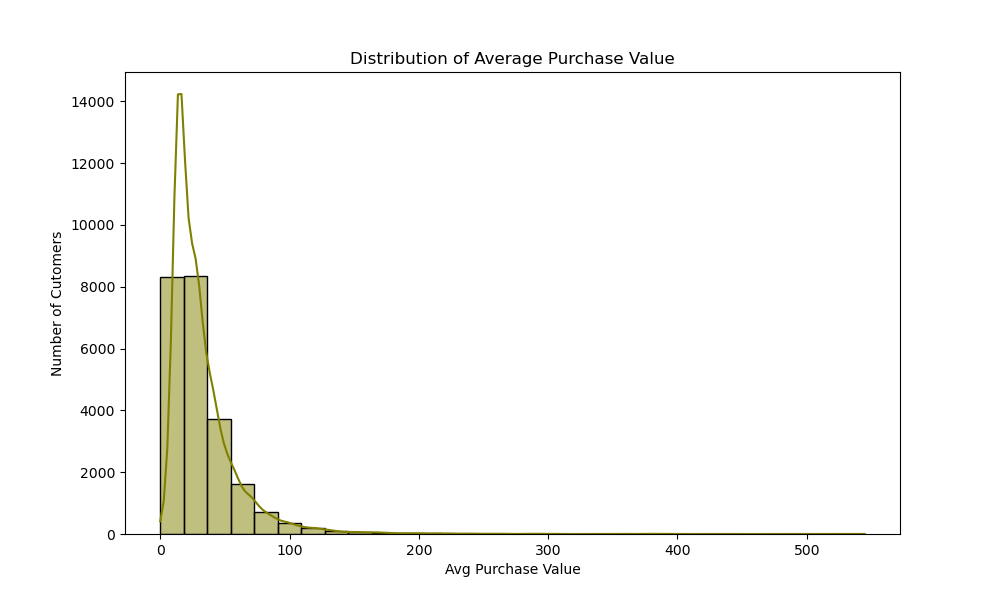
**Analysis Summary:** The scatter plots show the relationship between predicted spend and key features, providing insights into customer behavior.

**2. Correlation Heatmap**



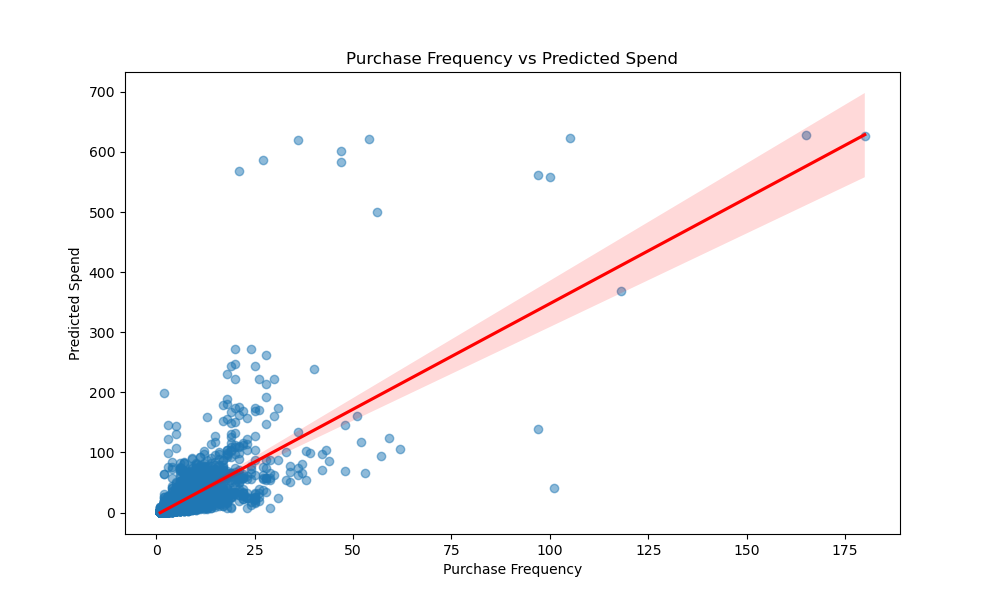
**Analysis Summary**: The heatmap highlights the correlation between different features and predicted spend, helping to identify important predictors.

**3. Customer Spending Distribution**



**Analysis Summary**: The distribution of customer spending provides insights into the overall spending behavior and helps in identifying high-value customers.

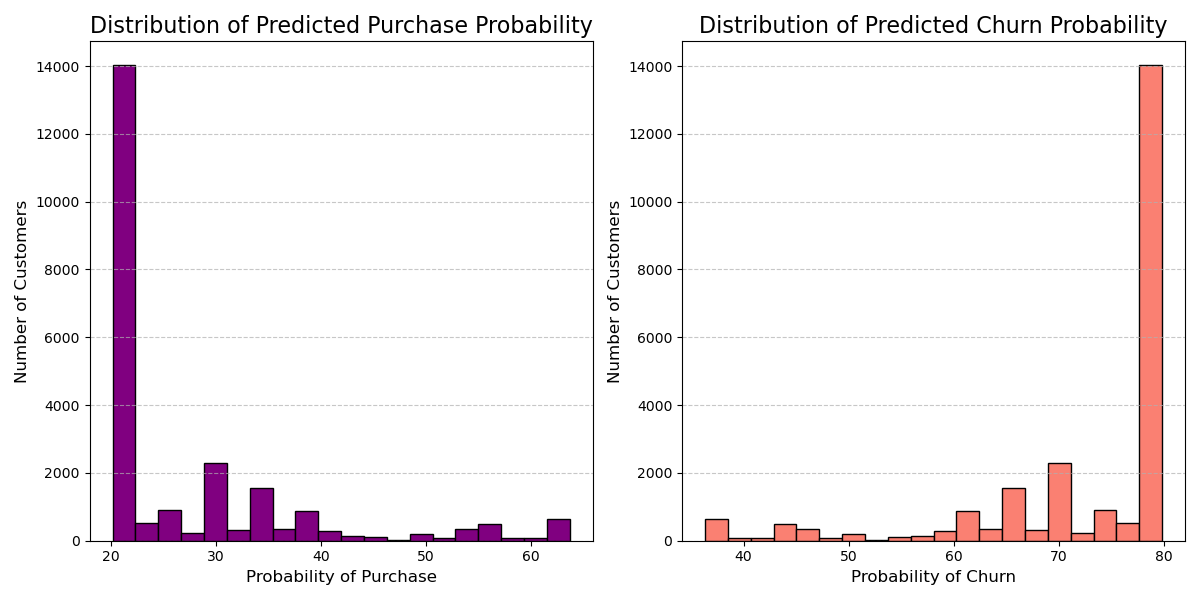
**4. Relationship Between Predicted Spend and Frequency**

****

**Analysis Summary:** This scatter plot shows the relationship between purchase frequency and predicted spend, highlighting how frequently purchasing customers tend to have higher predicted spend values.

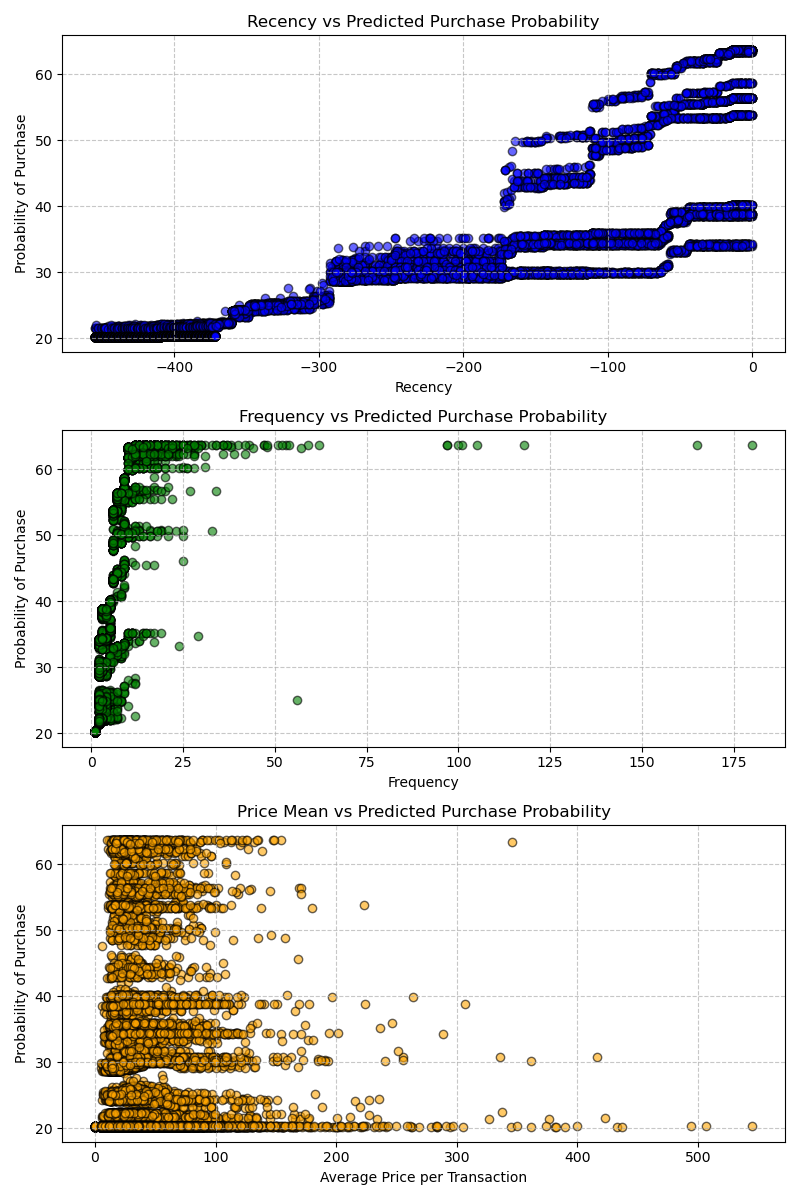
**Churn Probability Analysis**

**1. Purchase Probability Distribution**

[Placeholder for plot]

**Analysis Summary**: The distribution of purchase probability helps in understanding the likelihood of future purchases and identifying customers at risk of churn.

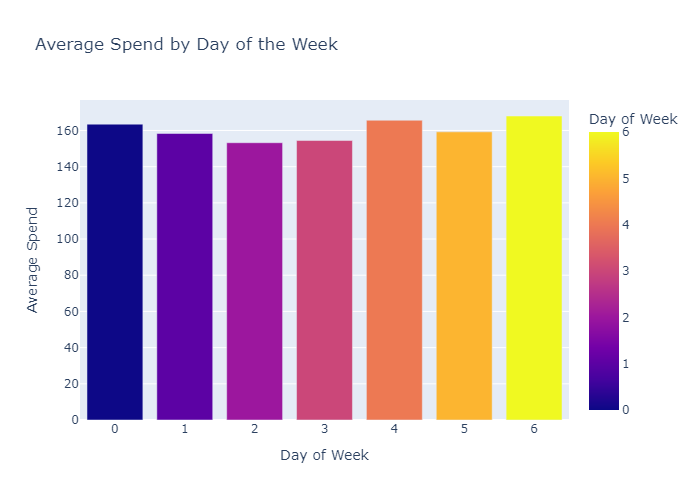
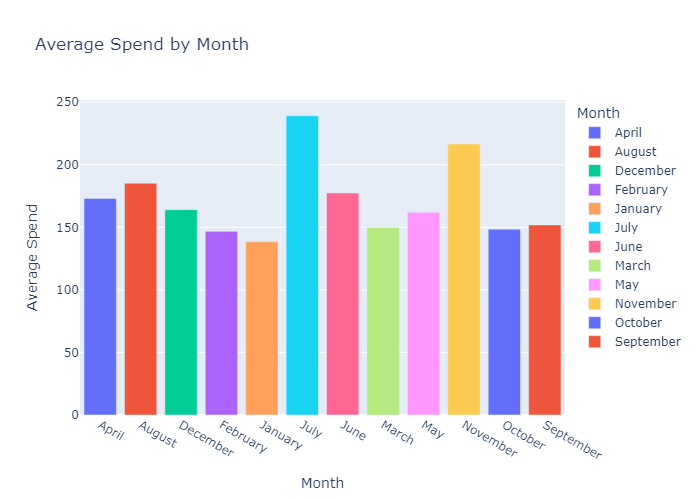
**2. Purchase Probability over RFM**



**Analysis Summary**: This visualization shows the relationship between purchase probability and RFM metrics, aiding in customer segmentation and targeting.

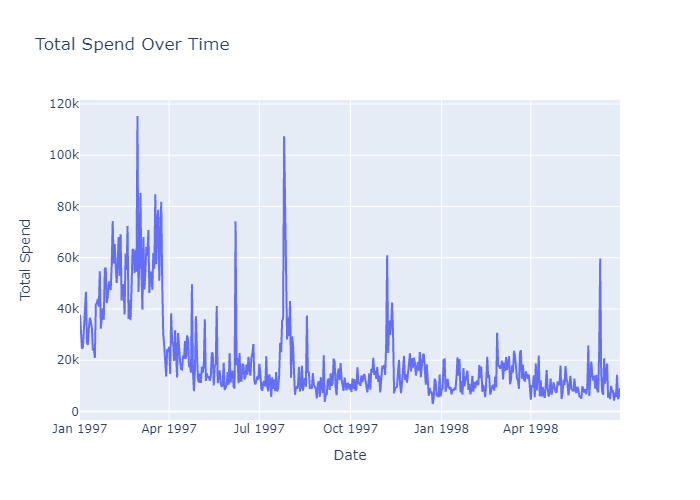
**Timeseries and Forecasting**

**1. Spend Trends Over Time**



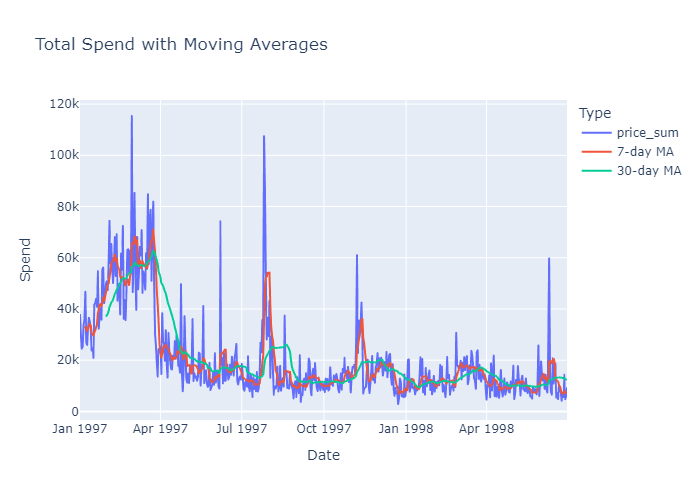
**Analysis Summary**: The timeseries analysis of spend trends helps in understanding seasonal variations and predicting future revenue.

**2. Total Spend Over Time**

****

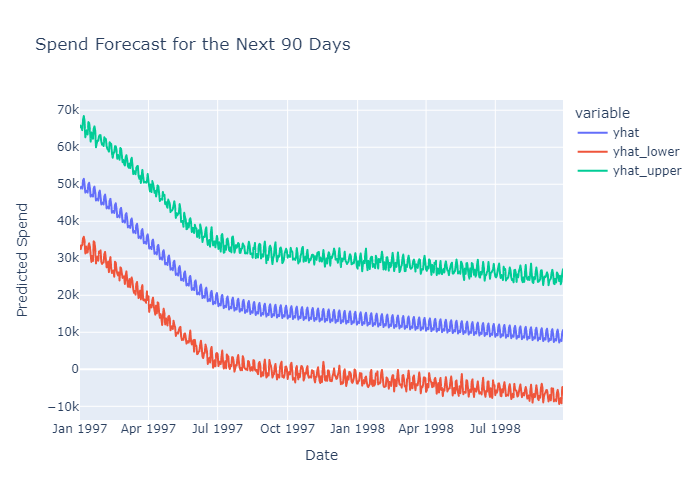
**Analysis Summary**: The timeseries analysis of total spend over time helps in understanding the overall spending trends and identifying any significant peaks or troughs in customer spending behavior.

**3. Total Spend with Moving Average**

****

**Analysis Summary:** The plot of total spend with a moving average provides a smoothed view of the spending trends, helping to identify long-term patterns and seasonal variations.

**4. Revenue Forecast Over Time**



**Analysis Summary**: The forecast for the next 90 days provides insights into expected revenue and helps in planning marketing strategies.

**Solutions to Business Problems**

**1. Identify high-spending customers in the next 90 days.**

- By using the Spend Amount Prediction Model, we can accurately predict which customers are likely to spend the most in the next 90 days. This allows for targeted marketing efforts to maximize revenue from these high-value customers.

**2. Detect customers who have recently purchased but are unlikely to buy again.**

- The Spend Probability Prediction Model helps identify customers who have made recent purchases but have a low probability of making another purchase. This insight can be used to design retention strategies to re-engage these customers.

**3. Highlight customers predicted to purchase but who did not (missed opportunities).**

- By comparing predicted spend probabilities with actual purchase behavior, we can identify customers who were expected to make a purchase but did not. This information can be used to understand potential reasons for missed opportunities and to improve future predictions.

**PTO**

**High-Level Insights and Conclusions**

- High-value customers contribute significantly to overall revenue, highlighting the importance of targeted retention strategies.

- Seasonal variations in revenue trends can inform marketing campaigns and promotional activities.

- Understanding customer purchasing patterns and behavior through RFM analysis aids in effective customer segmentation and targeting.

- Accurate predictions of spend probability and CLV enable data-driven decision-making and strategic planning.

**Suggestions for Further Actions**

**1. Targeted Marketing Campaigns:** Use the insights from the analysis to design targeted marketing campaigns for high-value customers.

**2. Customer Segmentation:** Implement customer segmentation based on RFM metrics to tailor marketing strategies and improve customer engagement.

**3. Retention Strategies:** Focus on retention strategies for customers with high spend probability and those at risk of churn.

**4. Upselling and Cross-Selling:** Leverage the insights on revenue by product price to design upselling and cross-selling strategies.

**5. Continuous Monitoring**: Continuously monitor customer behavior and update the models to ensure accurate predictions and effective decision-making.

**Conclusion**

This report provides a detailed analysis of the dataset, including all visualizations, model descriptions, solutions to our business problems, high-level insights, conclusions, and suggestions for further actions. The analysis highlights the importance of leveraging data-driven insights for effective customer engagement and strategic decision-making.

THANK YOU.