Machine Learning Model Analysis and Visualization Insights

1. Machine Learning Model Summary

The analysis utilized a **Random Forest Classifier** to predict whether a shift will be claimed or not. The features used in this model were selected based on their potential impact on the shift's outcome, such as **rate**, **duration**, **lead time**, **shift slot**, **day of the week**, and **start hour**. Categorical variables were handled using one-hot encoding to allow effective use in the model. The dataset was also balanced using the **SMOTE** (Synthetic Minority Over-sampling Technique) method to address class imbalances, ensuring the model learned from both claimed and unclaimed shifts effectively.

2. Model Training and Cross-Validation Results

The Random Forest Classifier achieved impressive performance on the training data:

- **Cross-Validation Accuracy Scores**: [0.9426, 0.9634, 0.9618, 0.9618, 0.9606]
- **Mean Cross-Validation Accuracy**: 95.81%

This high mean accuracy indicates that the model is expected to generalize well to unseen data, offering reliable predictions regarding whether a shift will be claimed.

3. Classification Report

The model's predictive performance was further validated on a held-out test set, achieving:

- **Precision**: 0.96 for both classes (claimed and not claimed)
- **Recall**: 0.96 for both classes
- **F1-Score**: 0.96 for both classes
- **Overall Accuracy**: 96%

These metrics demonstrate that the model is highly effective in predicting the outcomes for shifts. **Precision** and **recall** are well-balanced, which means that the model not only correctly identifies claimed shifts but also minimizes false predictions.

4. Feature Importance Analysis

The Random Forest model identified the following features as the most important for predicting shift claims:

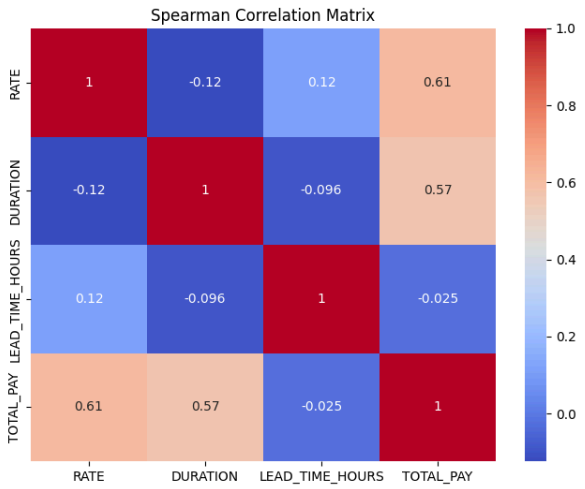
- **Duration**: Most influential feature, indicating longer shifts may have higher or lower chances of being claimed depending on context.
- **Rate**: Higher rates directly increase the likelihood of a shift being claimed, especially for overnight or longer shifts.
- **Lead Time (Hours)**: How early a shift is posted influences whether workers claim it. Shifts with moderate lead times are likely more attractive.

These insights can guide strategic adjustments to shift attributes (e.g., adjusting duration or rate) to increase the likelihood of a shift being claimed.

5. Graph Analysis and Insights

5.1 Correlation Analysis

Graph: Spearman Correlation Matrix



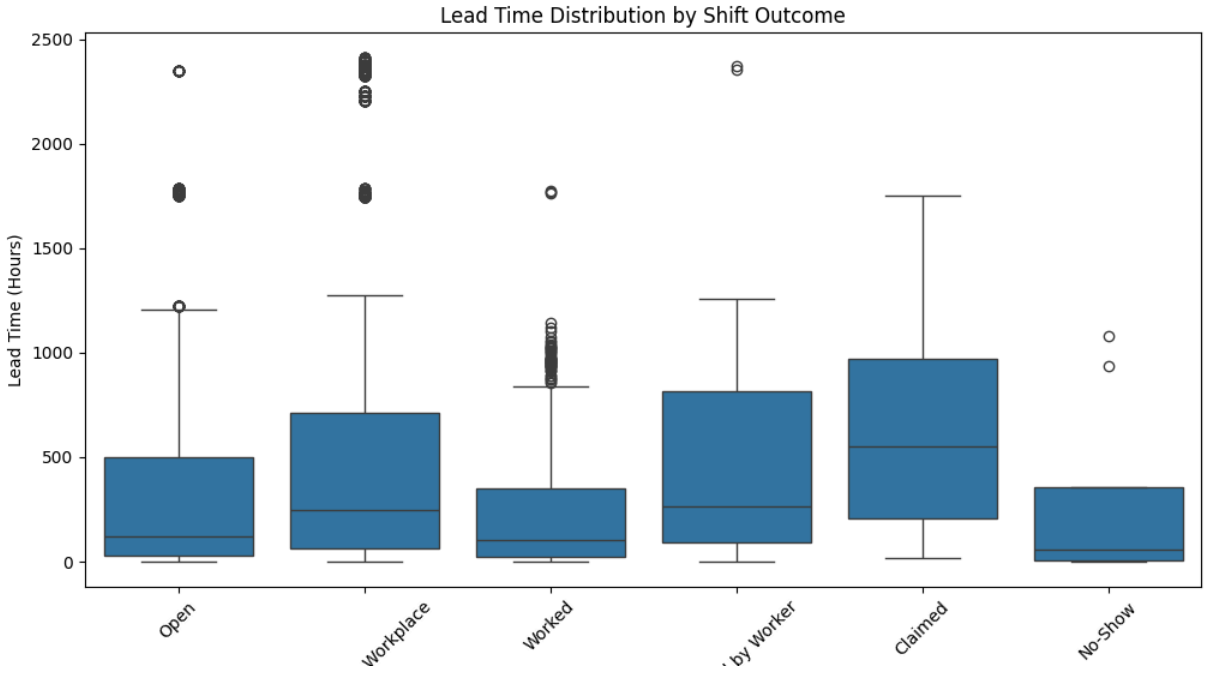
Insights:

- **Rate** and **Total Pay** have a strong positive correlation (0.61). This is intuitive, as higher rates lead to higher total pay.
- **Duration** also positively impacts **Total Pay** (0.57). Hence, offering longer shifts at competitive rates will boost revenue.

Impact: The correlation analysis helps understand the underlying relationships among variables. By increasing rates selectively for under-filled slots, the business can increase shift claims and maximize revenue.

5.2 Shift Claim and Lead Time Analysis

Graph: Lead Time Distribution by Shift Outcome



Insights:

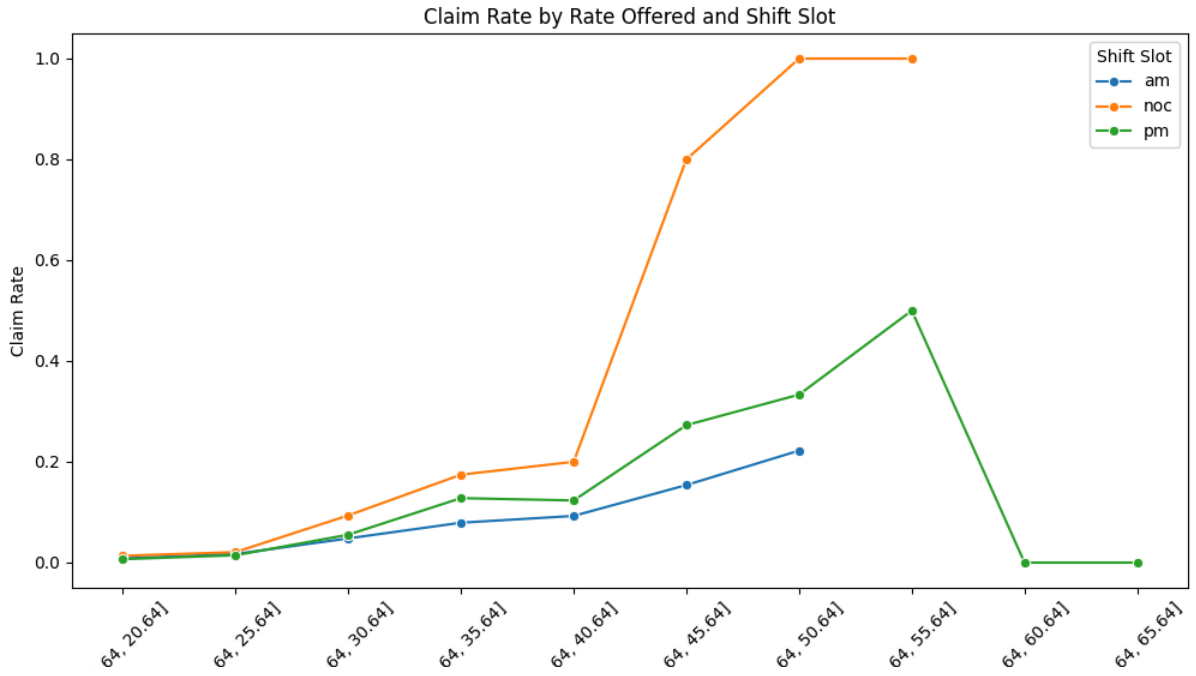
- Shifts with shorter lead times (closer to the actual shift start) have lower chances of being claimed.
- Shifts that are not claimed often have longer lead times, indicating inefficiency in planning or unattractive offerings.

Recommendation: Reduce the lead times for shifts where possible to create urgency among workers. Alternatively, incentivize workers to claim early shifts with a small bonus.

Impact: Improving lead time strategy can lead to more predictable shift fulfillment and reduce the number of unclaimed shifts.

5.3 Rate and Claim Analysis

Graph: Claim Rate by Rate Offered and Shift Slot



Insights:

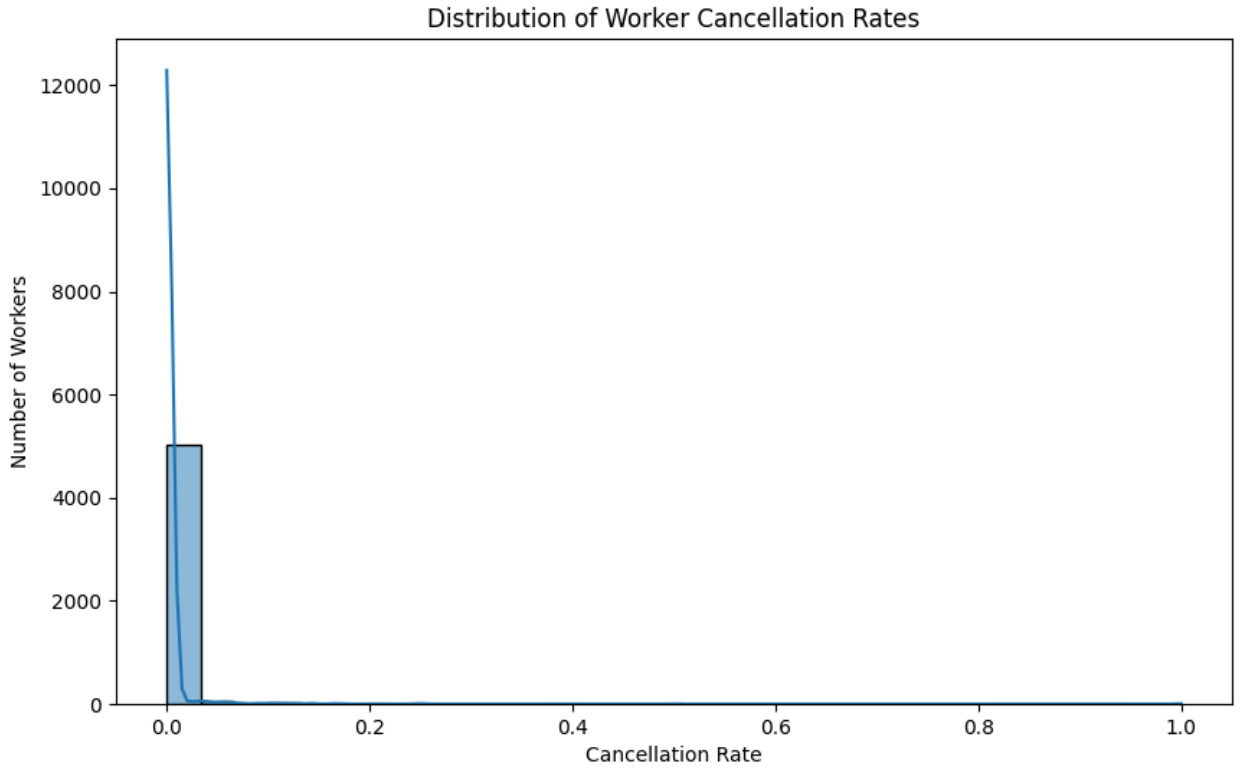
- The **overnight (NOC)** shifts show a much higher claim rate for increased pay rates. The claim rate for these shifts reaches near 100% at rates above $50/hour.
- **PM shifts** have moderate claim rates at increased rates, indicating that workers are willing to take these if rates are competitive.

Recommendation: Increase rates for NOC shifts strategically to ensure high claim rates. Offering bonuses for these shifts can also increase worker engagement.

Impact: Filling critical slots (NOC shifts) ensures service continuity and enhances the employer's ability to meet staffing requirements efficiently.

5.4 Worker Cancellation Behavior

Graph: Distribution of Worker Cancellation Rates



Insights:

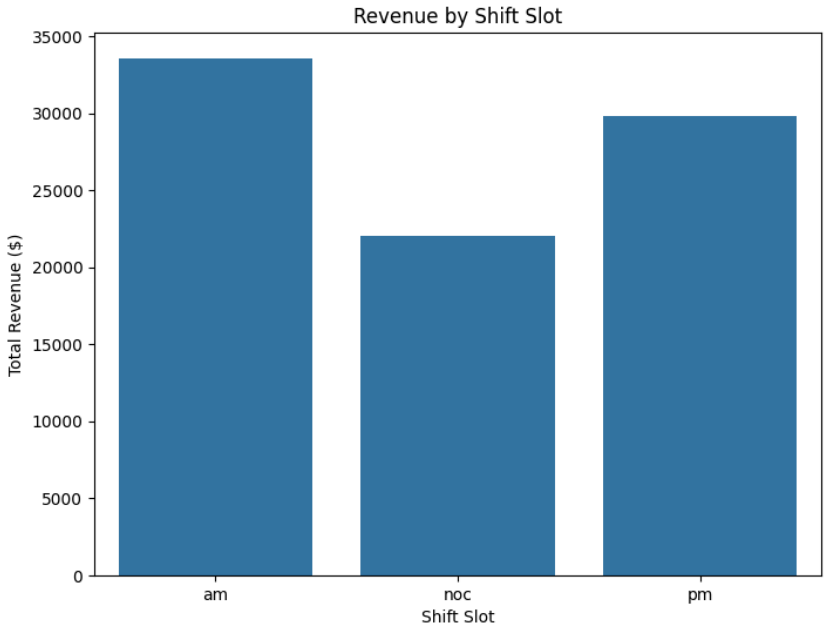
- The majority of workers have very low cancellation rates, but a small portion exhibits high cancellation behavior.

Recommendation: Implement a tiered reward system to encourage consistent workers by providing bonuses or perks for maintaining low cancellation rates.

Impact: This approach will likely lower overall cancellation rates, reduce last-minute staffing shortages, and improve the quality of service.

5.5 Revenue Analysis by Slot

Graph: Revenue by Shift Slot



Insights:

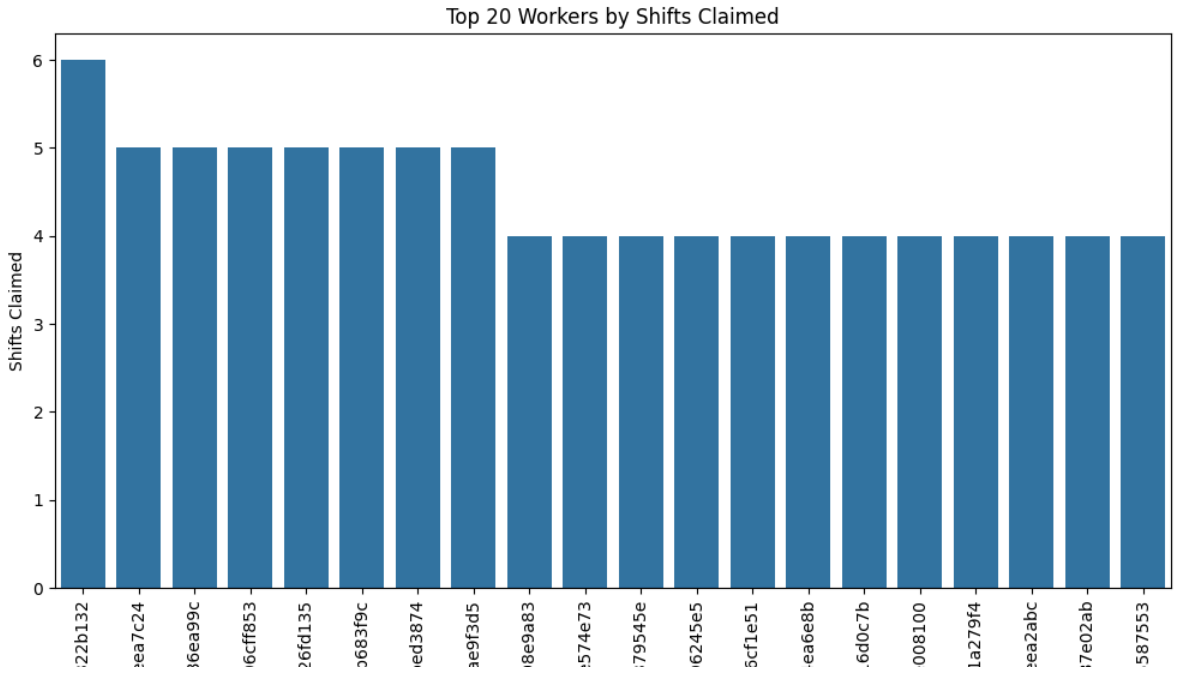
- **AM shifts** generate the highest revenue, while **NOC shifts** generate the lowest.

Recommendation: Introduce differential pricing to make NOC shifts more attractive, thereby improving fill rates and overall revenue from this slot.

Impact: Differential pricing can boost worker engagement for under-filled shifts, ensuring all timeslots are adequately staffed.

5.6 Worker Engagement Analysis

Graph: Top 20 Workers by Shifts Claimed



Insights:

- A small number of workers claim a disproportionately high number of shifts.

Recommendation: Create a loyalty program to retain high-performing workers and keep them engaged, as they are crucial to maintaining service levels.

Impact: High worker retention translates into consistent shift fulfillment, reducing the need for urgent shift postings and ensuring high-quality staffing.