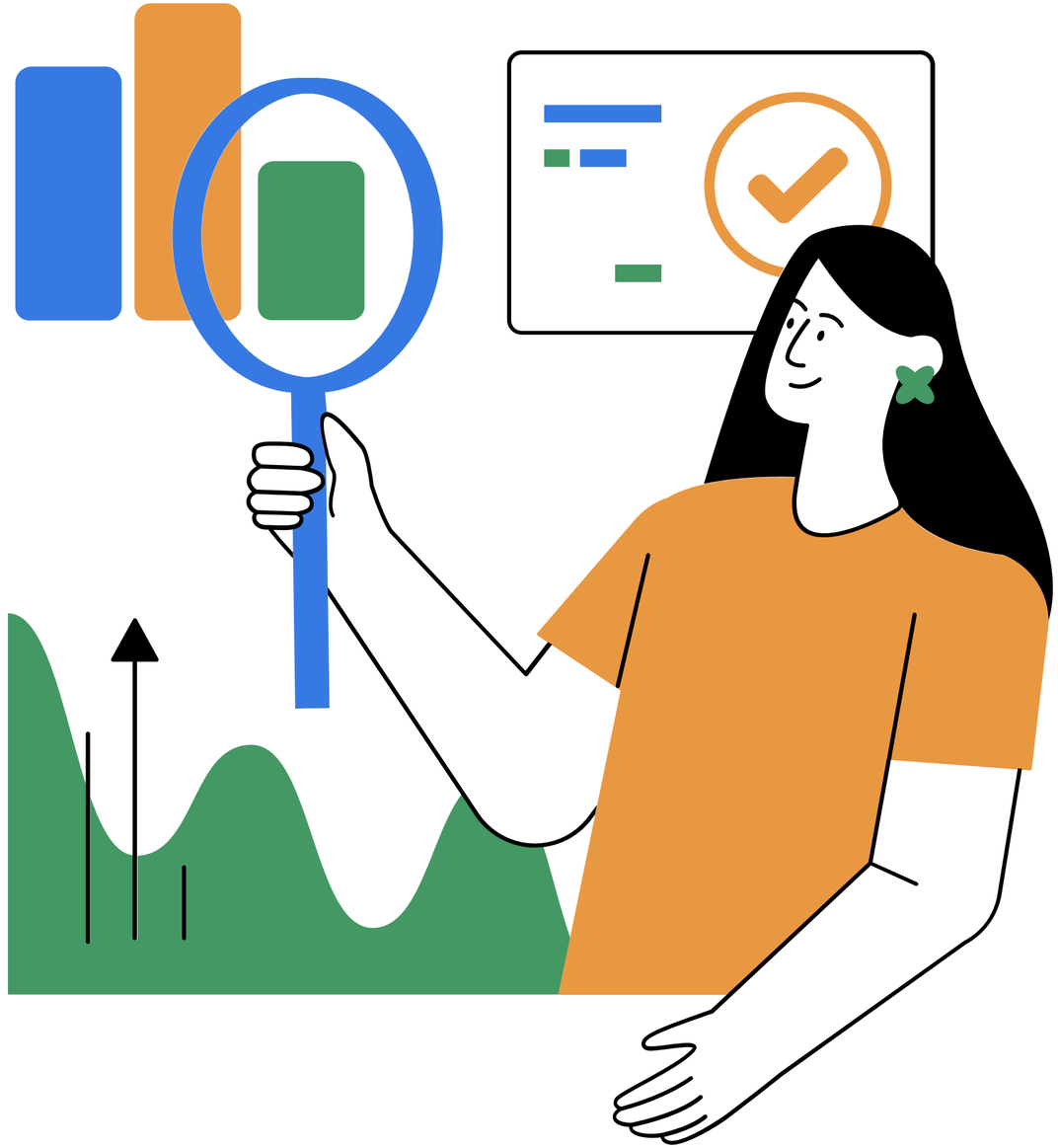
Predicting Recipe Traffic to Boost Engagement

Leveraging Machine Learning to Identify Popular Recipes



Overview of the Project and Business Goals

- **Objective:** Identify popular recipes to increase website traffic and subscriptions.
- **Current Issue:** Personal preferences lead to inconsistent results.
- **Solution:** Use data analysis and machine learning models to predict high-traffic recipes.

Summary of Work Undertaken

- **Data Collection:** Gathered recipe data including calories, carbohydrates, sugar, protein, category, servings, and traffic levels.
- **Data Cleaning:** Removed duplicates, handled missing values, and corrected errors in the dataset.
- **Exploratory Analysis:** Studied ingredient distributions, recipe categories, and their relationship to traffic.
- **Model Building**: Developed Logistic Regression and Random Forest models to predict high traffic.

Data Preprocessing

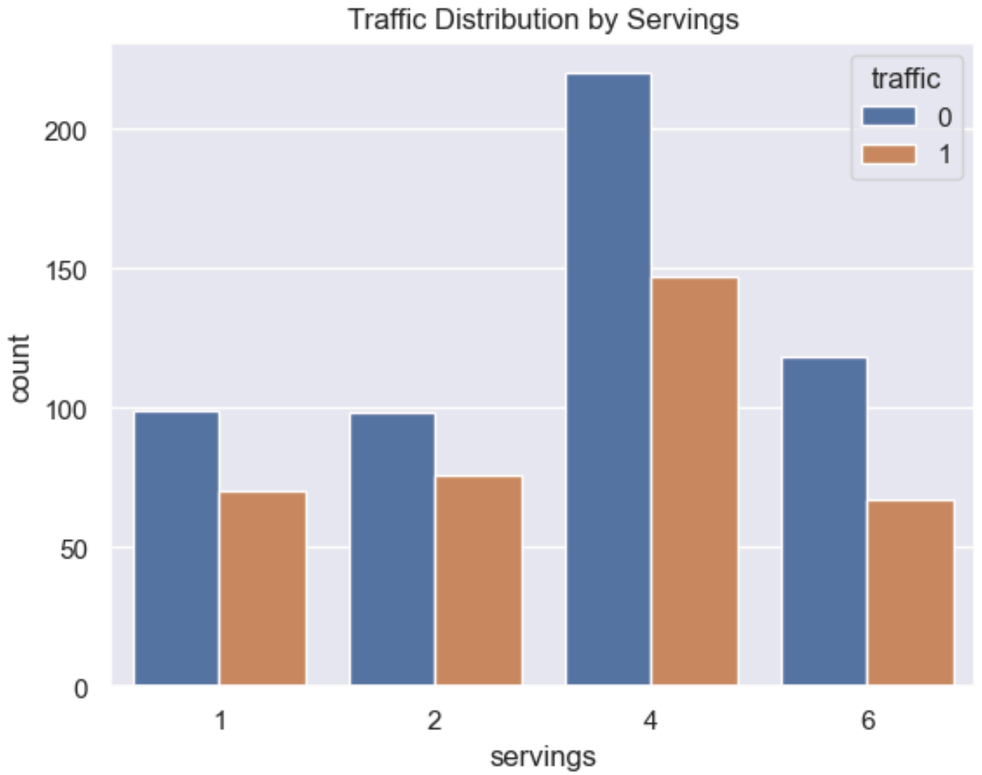
Missing Values: Initially, some rows had missing values. These were dropped after cleaning.

Outliers: Handled using the Interquartile Range (IQR) method.

Categorical Variables: Transformed into dummy variables for better model performance.

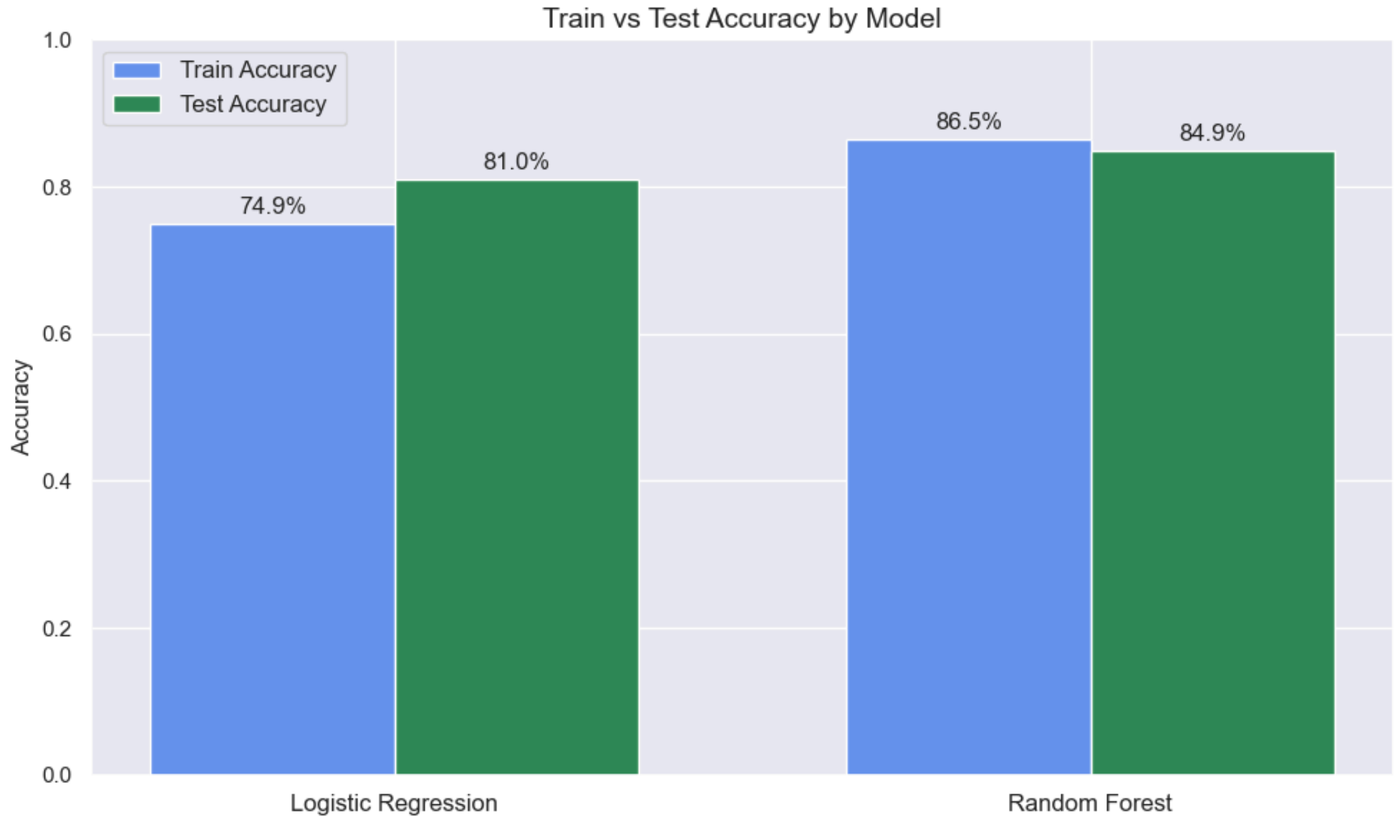
Key Findings - Traffic Distribution

- **Traffic Imbalance:** About 60% of recipes have high traffic, while 40% have low traffic.
- **Servings Impact:** Recipes with 4 servings tend to attract more visitors compared to others.



Model Performance

- **Logistic Regression:** Achieved around 81% accuracy but was less accurate than Random Forest.
- **Random Forest:** Performed better with about 85% accuracy, making it the preferred choice.



Business Recommendations

- **Use Machine Learning for Recipe Selection:** Implement the Random Forest model to predict high-traffic recipes.
- **Regular Updates:** Retrain the model periodically to adapt to changing user behavior.
- **Data-Driven Decision Making:** Set up dashboards to track recipe performance and improve recommendations.
- **Future Enhancements** Collect more details, such as preparation time and cost per serving, to enhance predictions.

Conclusion

This project uses data to help choose the best recipes, making the website more popular and engaging. By using the Random Forest model, updating it regularly, and tracking real-time data, the business can make better choices and keep users happy.