PRESENTATION

Datacamp Certified Data Scientist Certification

Recipe Site Traffic

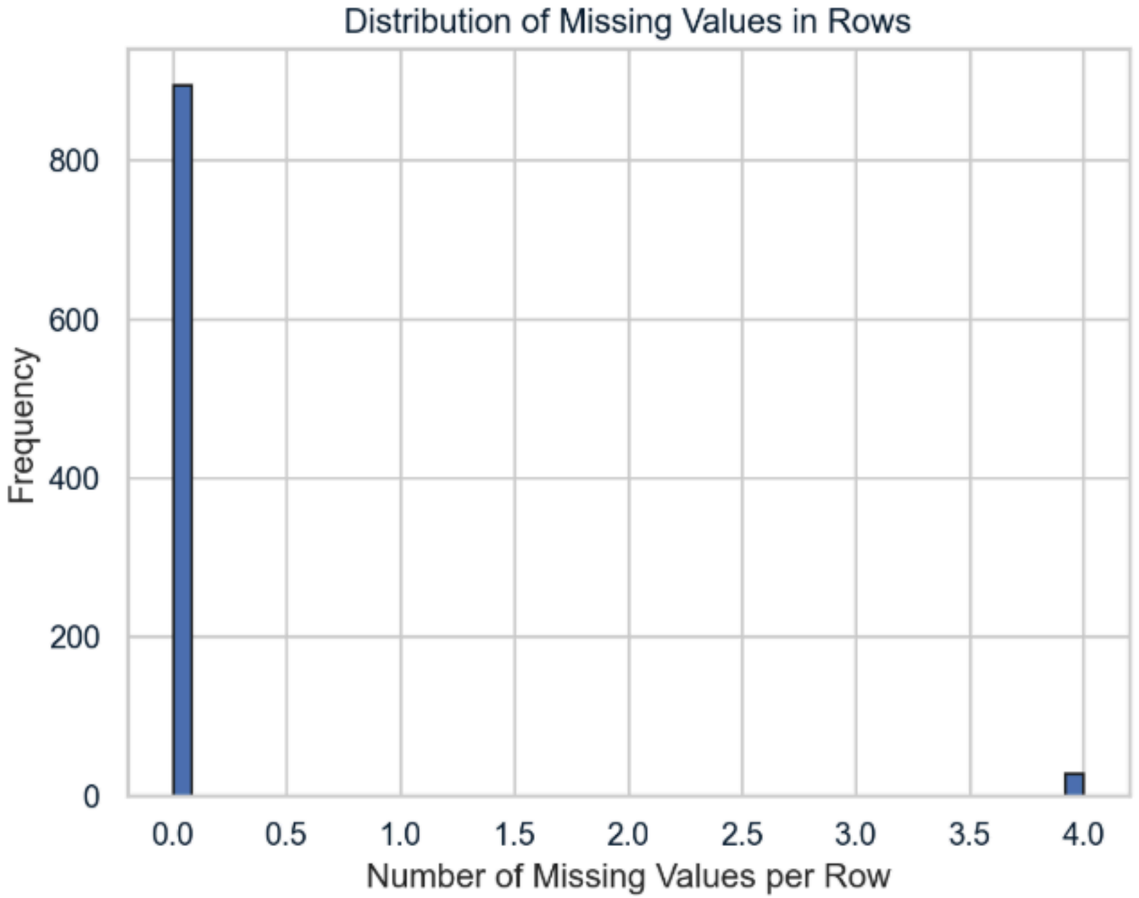
Presented By : Radib Bin Kabir

Recipe Site Traffic Analysis

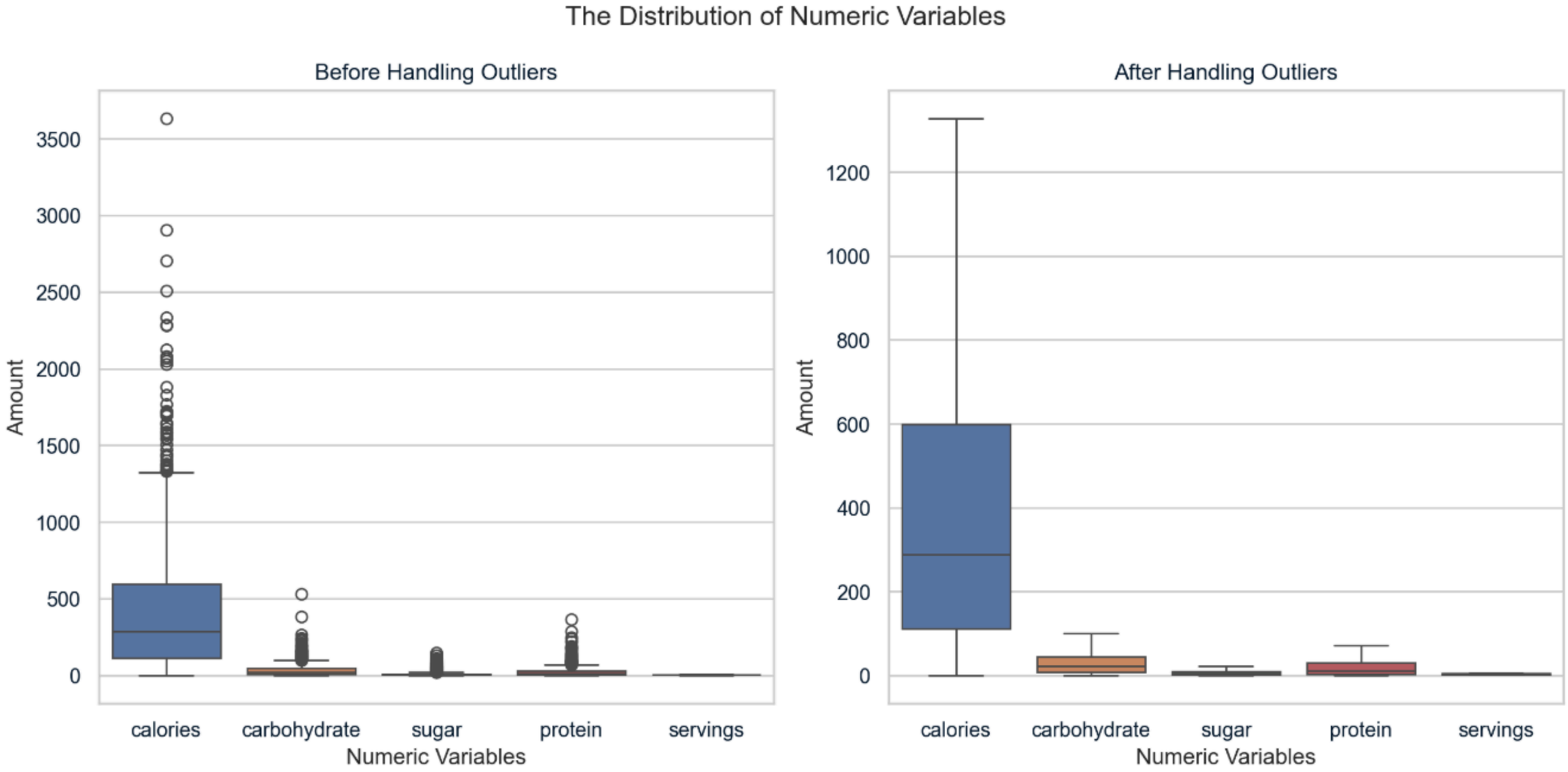
- There were **two objectives** given to us to fulfill-
  - To forecast recipes that will generate **high traffic**
  - Correctly predict high traffic recipes **80%** of the time
- A dataset was provided to us containing **947 rows** and **8 columns**
- **Used tools :** Python, data visualization libraries, and machine learning models

Data Validation

- Converted **object** columns to the appropriate **data type**.
- Adding a **new category 'Low'** for **‘traffic_high’** column.
- **Standardized inconsistent values** in the **'servings'** column.
- **Removed duplicate** entries to ensure data integrity.
- **Dropped rows** with four or more **missing values** to maintain quality
- Performed **Outlier Analysis**

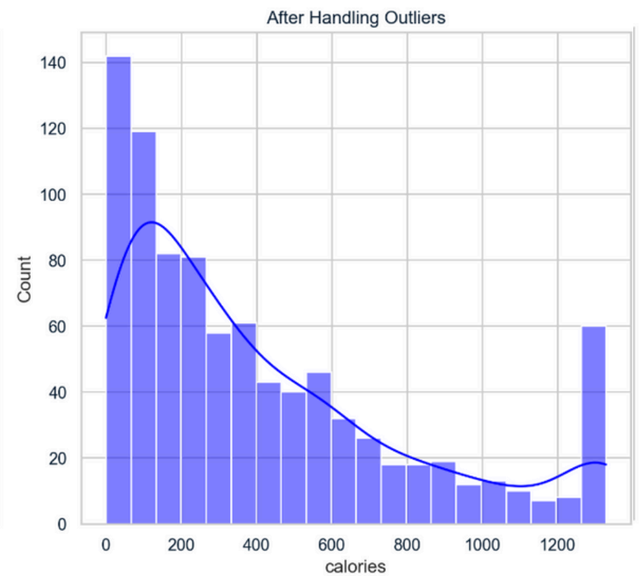


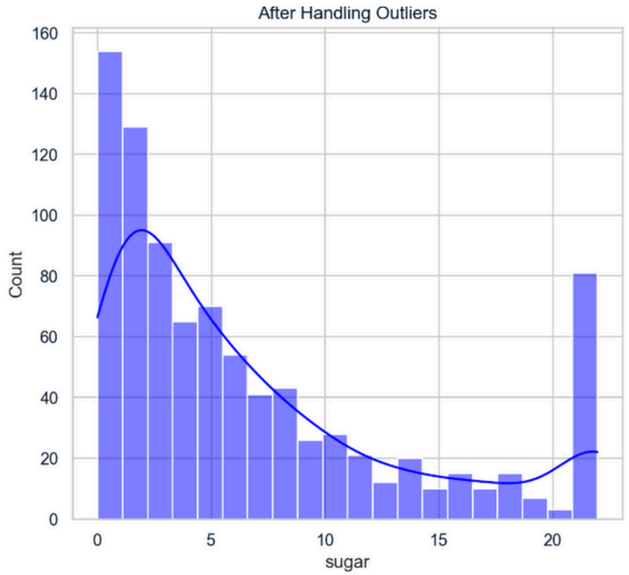
Outlier Analysis

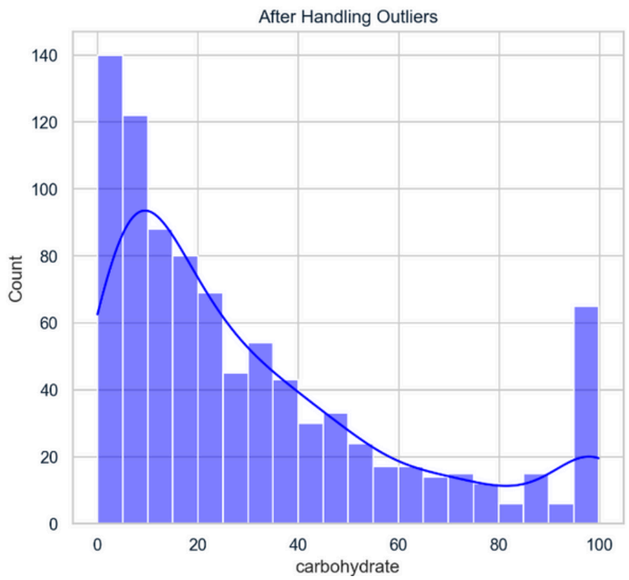


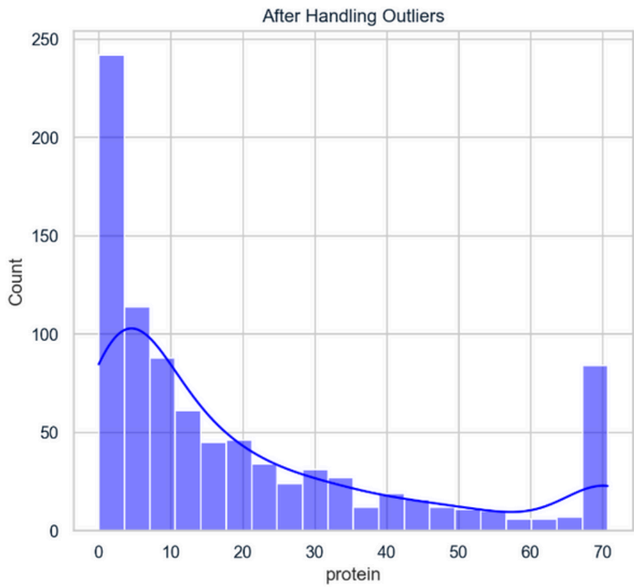
Exploratory Data Analysis

- The numeric column values are **right-skewed**, meaning most values are clustered on the lower end, with a long tail of higher values.
- Possible need for transformation, such as **log transformation**, to improve model performance and normalization.



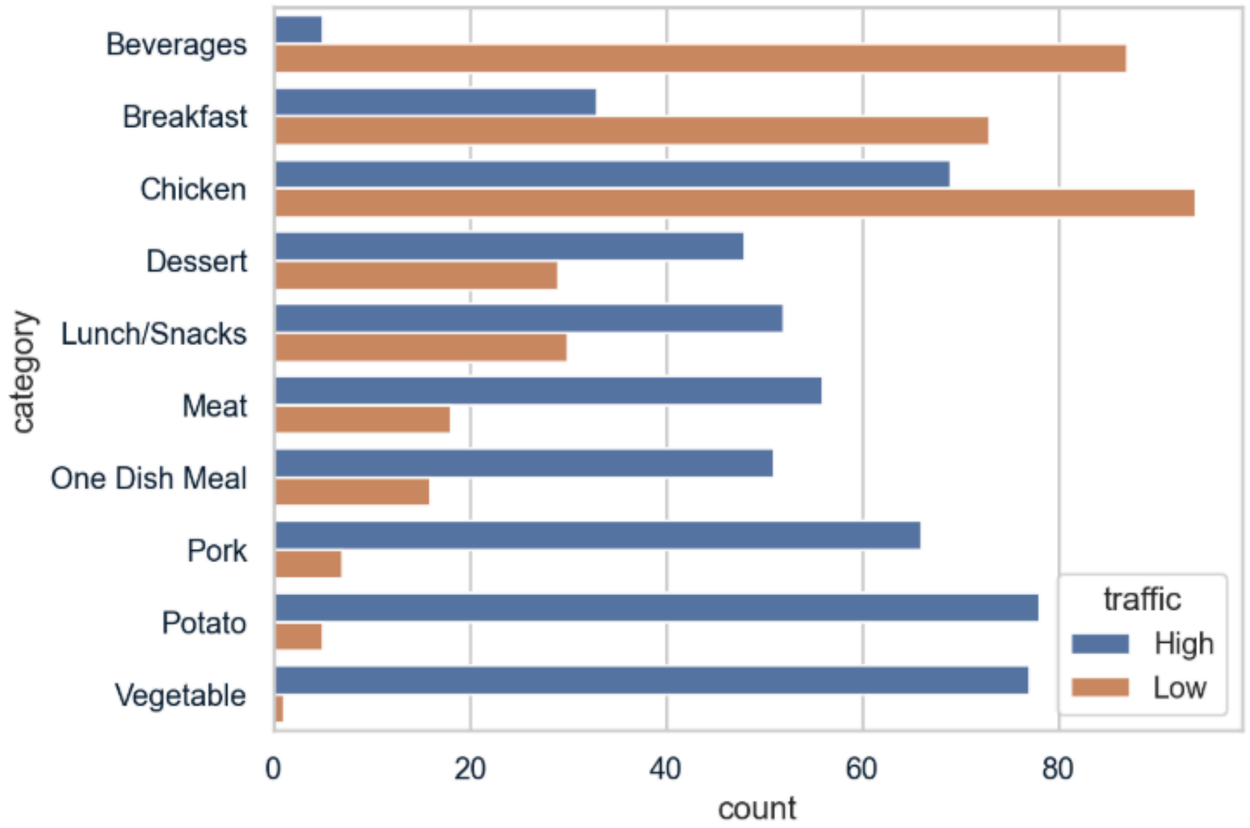






Exploratory Data Analysis

- **Potato** recipe category generates the most **high** traffic
- **Vegetable and Pork** recipe category also genrates **high** traffic
- **Beverage and Breakfast** recipe category generates the lowest traffic

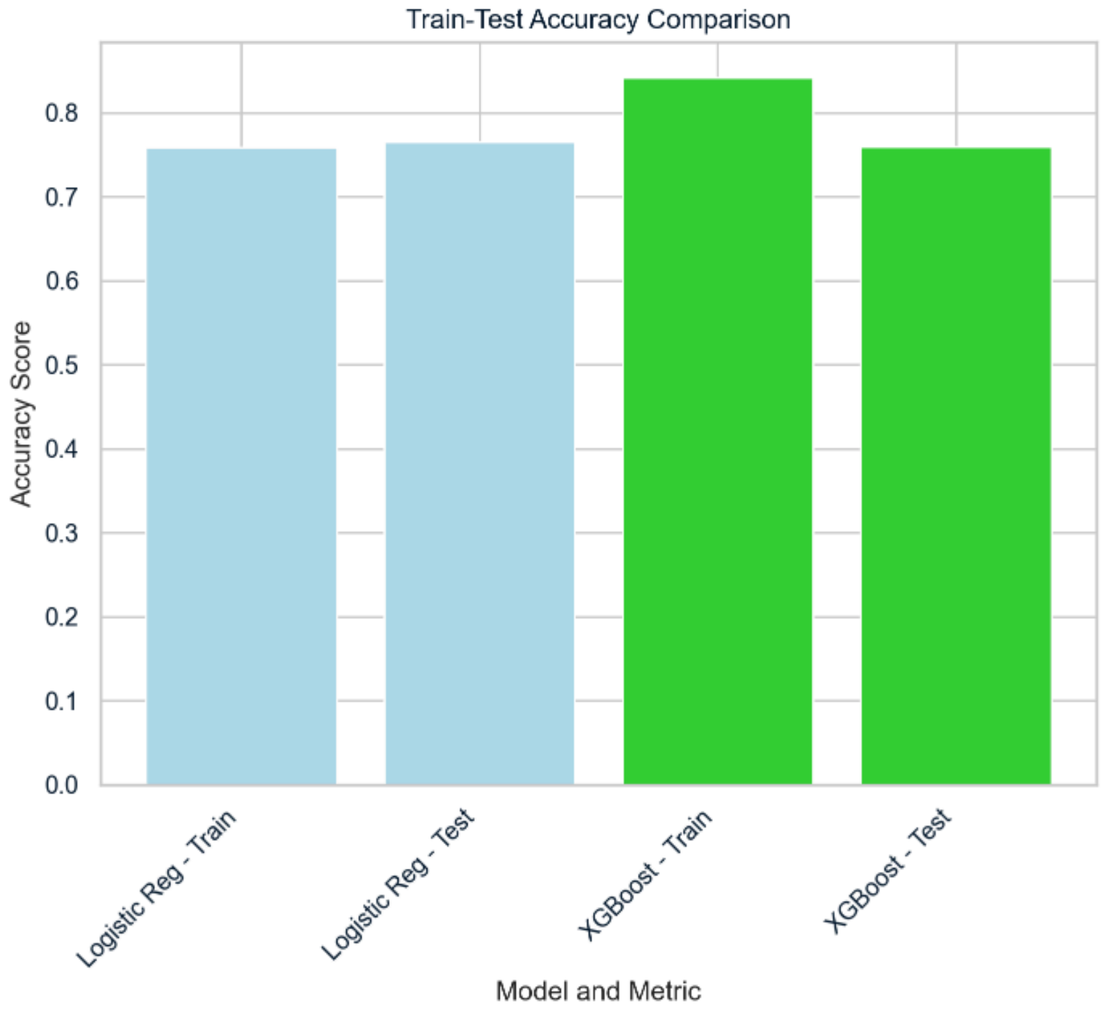


Correlation Analysis

Model Performance Evaluation

Model	Accuracy	F1 Score	Precision	Recall
Logistic Regression	**0.77**	**0.70**	0.72	**0.68**
Random Forest	0.72	0.56	**0.76**	0.44
XGBoost	0.76	0.68	0.74	0.63

Model Performance Evaluation



Recommendations For Business

- Prioritize **high-traffic categories** (Pork, Potato, Vegetable etc.) when choosing homepage recipes.
- Consider **ingredient count and preparation time** as key factors when curating homepage content.
- Deploying the **Logistic Regression** model into production is advised.
- **A/B Testing** – Test different recipe selections using machine learning recommendations vs. manual selections to measure impact.

THANK YOU