CAPSTONE

PROJECT

DATA SCIENCE

Food Delivery Dataset

GET STARTED →

The dataset focuses on food delivery, a service where food is delivered to customers by restaurants, stores, or independent companies. Orders are typically placed via websites or mobile apps. The delivery can include various food items and is usually made using cars, bikes, or scooters, depending on the locations.

How to Use the Data:

This dataset is intended for enhancing skills in feature engineering and model stacking. The primary evaluation metric is the R^2 score. It is ideal for Data Science capstone projects.

What you can do with this Dataset ?

- UNDERSTAND THE DATASET: REVIEW THE STRUCTURE, VARIABLES, AND TYPES OF DATA PROVIDED.
- DATA CLEANING: HANDLE MISSING VALUES, DUPLICATES, AND OUTLIERS.
- DESCRIPTIVE STATISTICS: CALCULATE MEAN, MEDIAN, MODE, AND STANDARD DEVIATION FOR KEY VARIABLES.
- EXPLORATORY DATA ANALYSIS (EDA): CREATE VISUALIZATIONS LIKE HISTOGRAMS, BOX PLOTS, AND SCATTER PLOTS.
- FEATURE ENGINEERING: DEVELOP NEW FEATURES THAT COULD IMPROVE MODEL ACCURACY.
- CORRELATION ANALYSIS: INVESTIGATE RELATIONSHIPS BETWEEN DIFFERENT VARIABLES.
- PREDICTIVE MODELING: USE MACHINE LEARNING MODELS TO PREDICT DELIVERY TIMES.
- MODEL EVALUATION: ASSESS MODEL PERFORMANCE USING THE R^2 SCORE AND OTHER METRICS.
- HYPERPARAMETER TUNING: OPTIMIZE MODEL PARAMETERS TO IMPROVE PREDICTIONS.
- CROSS-VALIDATION: IMPLEMENT CROSS-VALIDATION TO CHECK MODEL RELIABILITY.
- DATA SPLITTING: DIVIDE THE DATASET INTO TRAINING AND TESTING SETS FOR BETTER MODEL EVALUATION.

- Dimensionality Reduction: Apply techniques like PCA to reduce the feature space.
- Outlier Detection: Identify and manage outliers in the dataset.
- Customer Segmentation: Group customers based on their ordering behavior.
- Distance Analysis: Examine how the distance between restaurant and customer affects delivery time.
- Impact of Discounts: Study the influence of discounts on order frequency and volume.
- Restaurant Performance: Compare delivery efficiency across different restaurants.
- Visualization: Use tools like Tableau or Power BI to create interactive dashboards.
- Customer Retention: Analyze factors contributing to customer loyalty.
- Order Size Analysis: Explore patterns in the size of food orders.
- Model Stacking: Combine different models to improve prediction accuracy.
- Residual Analysis: Check residuals to ensure model assumptions hold.
- Ensemble Learning: Use methods like Random Forest or Gradient Boosting for predictions.
- Final Presentation: Prepare a detailed report and presentation summarizing the findings.

CONCLUSION:

THIS DATASET OFFERS RICH OPPORTUNITIES FOR STUDENTS TO APPLY AND ENHANCE THEIR DATA SCIENCE SKILLS, FROM BASIC EXPLORATORY ANALYSIS TO ADVANCED MACHINE LEARNING TECHNIQUES, MAKING IT A VALUABLE RESOURCE FOR CAPSTONE PROJECTS.