DA 204o: **Data Science in Practice**

Course Project Deck

Price Optimization and Customer Satisfaction Prediction

Barani Ranjan S, *baranis@iisc.ac.in*
Ishika Saxena, *ishikasaxena@iisc.ac.in*
Prasanna Kumar B V, *prasannabv@iisc.ac.in*
Siva S, *sivas12@iisc.ac.in*

Introduction:

Project Overview

Objectives

Significance

Project Overview :

The restaurant industry is highly competitive, with customer satisfaction being a key driver of success. This project leverages data science techniques to address two critical challenges:

- Predicting customer satisfaction based on restaurant attributes.
- Optimizing pricing strategies to maximize customer satisfaction and profitability.

Objectives:

- Build a classification model to predict high/low customer satisfaction.
- Develop a regression model to recommend optimal price ranges for different operational conditions.

Significance:

These models empower restaurants to enhance customer experiences, improve ratings, and optimize revenue strategies.

Dataset Overview

Data source

Description of the data

Key Features: Input Variables

Target Variable

Data source : The data is sourced from Kaggle's unique "27000 Indian Restaurant Dataset", which contains detailed information on various Indian restaurants.

Description of the data

Features:

- **Restaurant Name**: The name of the restaurant.
- **Cuisine Type**: The type of cuisine served.
  (South Indian, North Indian, Fast Food, Street Food, Bakery)
- **Avg. Price** : The average price of the food in the Restaurant.
- **Avg. Delivery time** : The average time taken for delivery by the restaurant.
- **Location** : City where the restaurant is located.
- **Avg. Rating :** The average of Ratings provided by the customer for a particular restaurant.

Key Features: Input Variables:

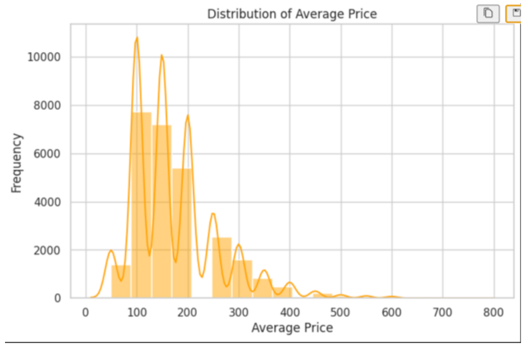
- Average price, delivery time, and cuisine type (e.g., South Indian, North Indian, Fast Food).

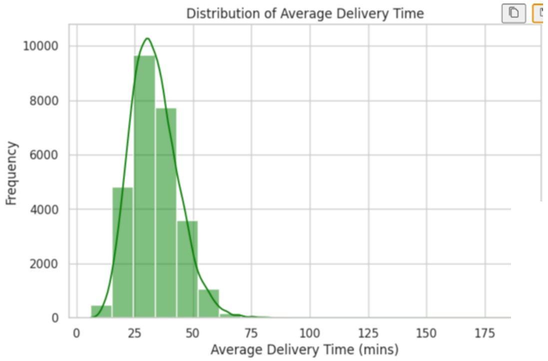
Target Variable:

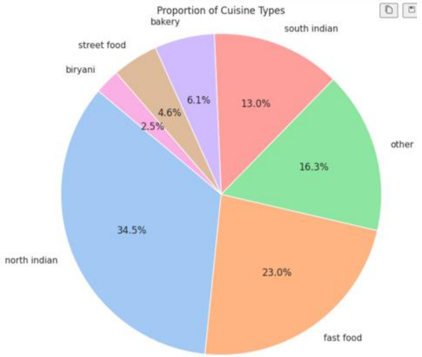
- **For Classification**: It is binary (0 or 1), where a value of 1 indicates a rating of 4 or higher, and 0 indicates a rating below 4. Transform the original rating column into a binary outcome.
- **For Regression**: Optimal price (continuous).

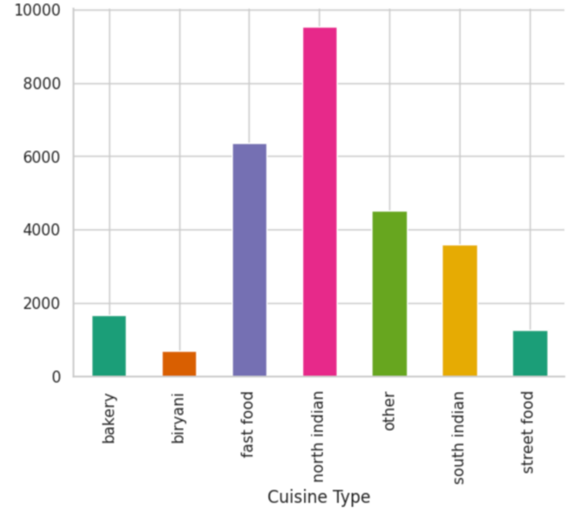
Exploratory Data Analysis (EDA)

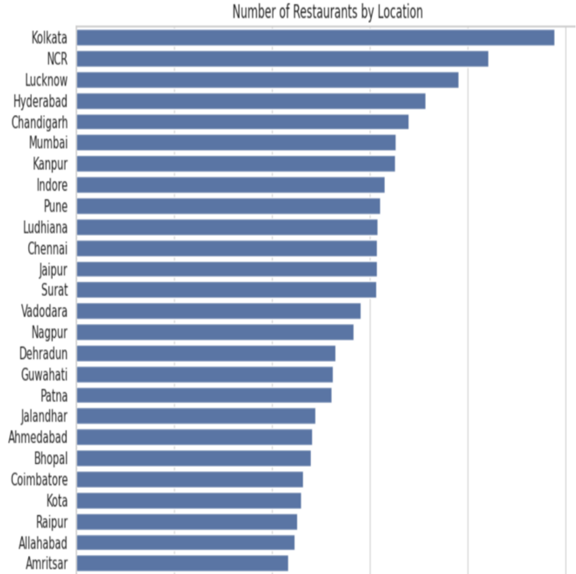
Univariate Analysis & Bivariate Analysis

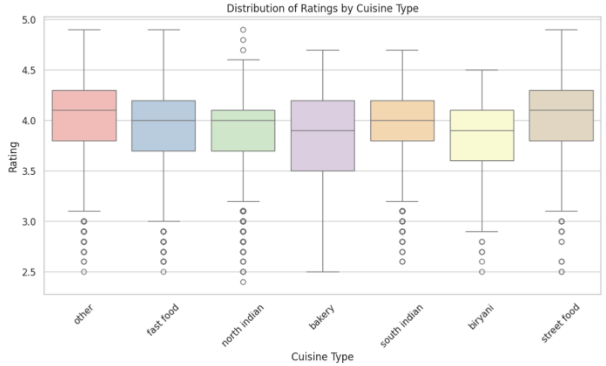


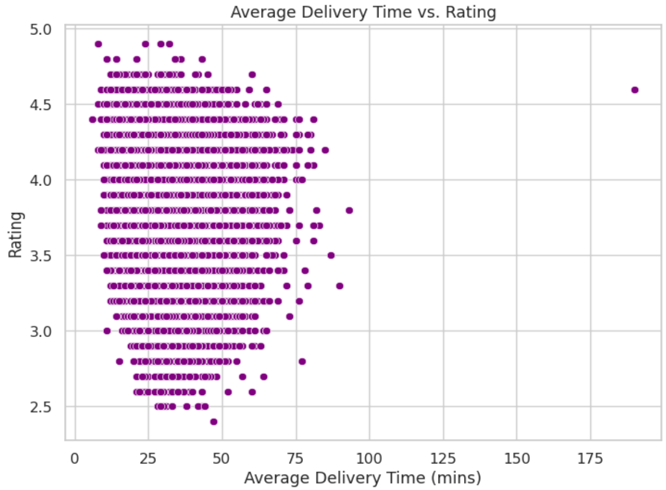


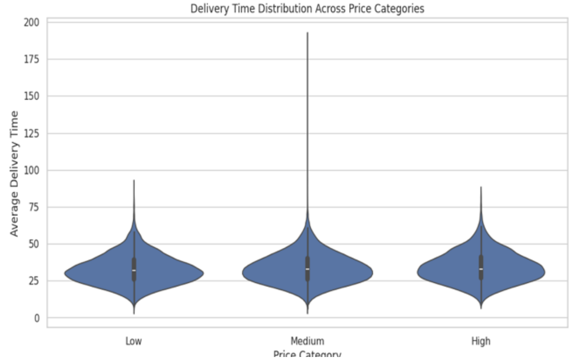












Multivariate Analysis
Correlation Analysis and Heatmap

Multivariate Analysis

Ratings Consistency Across Price Categories: The analysis shows that ratings remain consistent across different price categories.

Clear Price Distinction: The average price effectively differentiates between various price levels.

Delivery Time Stability: Delivery time remains relatively stable across all price categories.

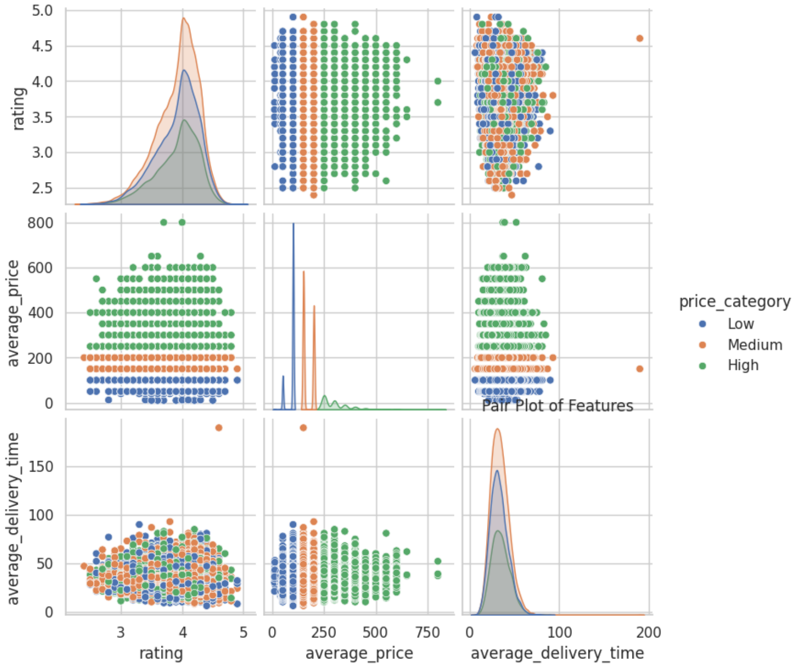
Low Feature Correlation: There is little correlation between the features, suggesting that the factors (ratings, price, and delivery time) operate independently of each other.

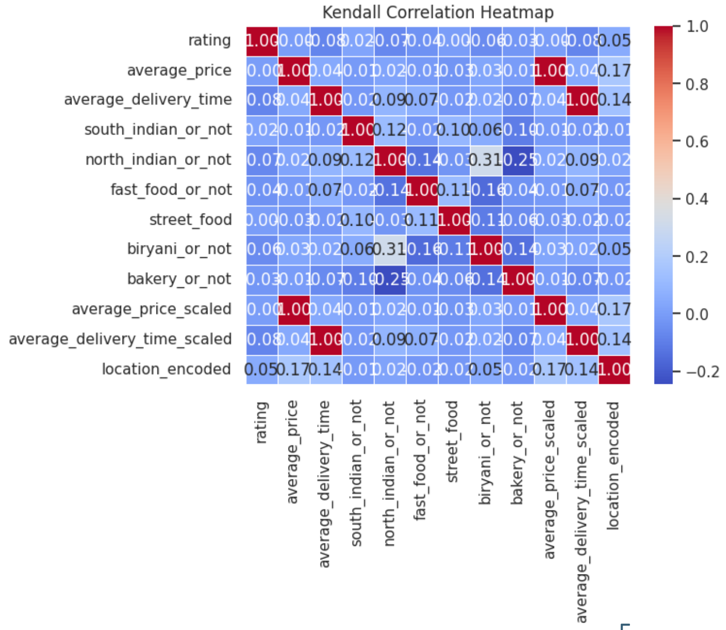
The correlations indicate that:

1. Price and rating are weakly related, suggesting ratings are not heavily influenced by price.

2. Cuisine types show minor correlations with price and ratings but are largely independent.

3. Location has a moderate impact on price but a minimal effect on delivery time or ratings, while delivery time is weakly related to other factors.



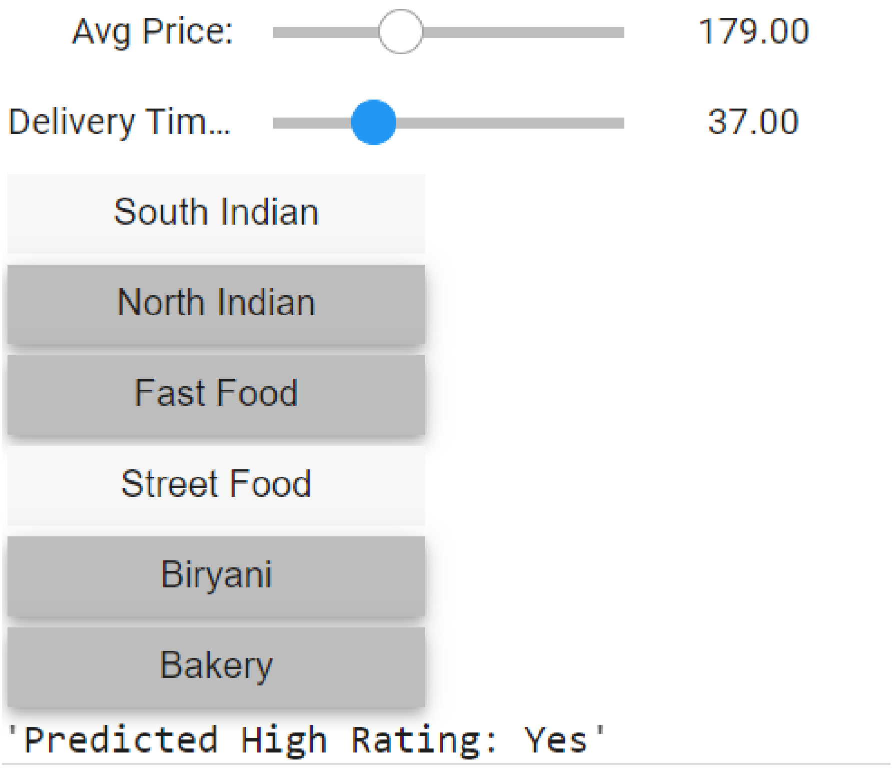


Classification Model Evaluation

- Weak correlation of Rating
- Scaled numeric features
- Class Imbalance : Slight Imbalance towards Positive class ,but not extreme (1.2 : 1)
- Models
  - ✓ Logistic Regression
  - ✓ Decision Trees
  - ✓ Random forest
  - ✓ XG Boost

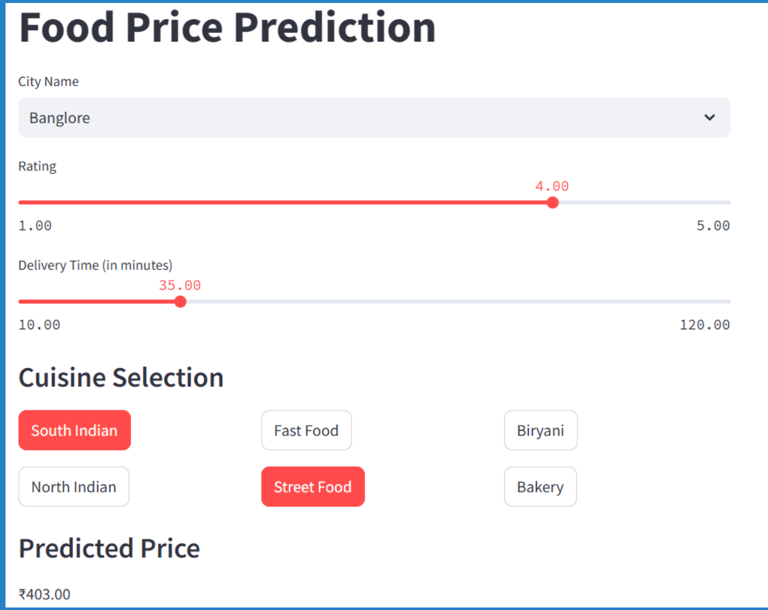
Model	Accuracy	Precision (Class 0)	Recall (Class 0)	F1-Score (Class 0)	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)	ROC AUC	Confusion Matrix
Random Forest	54.86%	0.49	0.47	0.48	0.59	0.61	0.6	0.56	[[1134, 1295], [1202, 1901]]
Decision Tree	53.36%	0.47	0.54	0.5	0.6	0.53	0.56	0.54	[[1316, 1113], [1467, 1636]]
Logistic Regression	57.43%	0.52	0.37	0.44	0.6	0.73	0.66	0.6	[[908, 1521], [834, 2269]]
XGBoost	60%	0.54	0.43	0.51	0.6	0.73	0.67	0.61	[[1050, 1379], [872, 2231]]

Classification Model Results



- XGBoost performed the best in terms of accuracy (~60%)
- Best Hyperparameters: {'classifier__colsample_bytree': 1.0, 'classifier__learning_rate': 0.01, 'classifier__max_depth': 5, 'classifier__n_estimators': 200, 'classifier__scale_pos_weight': 1, 'classifier__subsample': 0.8}
- Performance for class 1 is better than for class 0
- Random Forest and Decision Tree are with lower accuracy (~54%) and struggled with Class 0 classification

Pricing Optimization
(Regression Model)



Objective:
Recommend optimal price points based on operational factors to balance customer satisfaction and profitability.

Features Used:
- **Rating:** Customer rating of the restaurant.
- **Average Delivery Time:** Average time taken to deliver food.
- **Cuisine Flags:** Types of cuisine offered.
- **Location:** Geographical location of the restaurant.

Target Variable:
- **Average Price:** Predicted based on the features above.

Data Transformations:
- **Encoding:**
  - ✓ Location (categorical to numerical encoding)
- **Scaling:**
  - ✓ Rating
  - ✓ Average Delivery Time

Pricing Optimization

(Model Evaluation and Key insights)

Models Trained:

- Random Forest
- Gradient Boosting
- XGBoost
- Support Vector Regressor (SVR)

Model comparison and Metrics

Model	MAE	MSE	R^2	Best Hyperparameters
Random Forest	65.62	7335.76	0.3	'max_depth': 10, 'min_samples_split': 10, 'n_estimators': 200
Gradient Boosting	64.53	7199.27	0.4	'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 200
XGBoost	64.58	7199.75	0.4	'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 200
Support Vector Regressor	61.08	7745.82	-0.3	'C': 10, 'gamma': 'scale', 'kernel': 'linear'

Key Insights:

- Significant features: **Rating, Average Delivery Time**
- Top-performing models: **Gradient Boosting** and **XGBoost**
- Improvement: Robust hyperparameter tuning led to enhanced performance for Gradient Boosting and XGBoost.

Recommendations, Conclusion & Future Work

Recommendations:

- Predict Ratings with Classification Model: Use the classification model to forecast whether a customer will rate the restaurant highly and adjust operational strategies (e.g., delivery time, pricing) accordingly.
- Optimize Pricing Dynamically: Use the regression model to recommend dynamic price adjustments based on factors like delivery time and cuisine type.
- City specific insights

Conclusion:

- Data-driven models enhance restaurant profitability and customer satisfaction through optimized pricing and satisfaction prediction.

Future Enhancements:

- **Feature Expansion:** Add more variables like customer demographics and weather to improve model accuracy.
- **Real-Time Pricing:** Implement dynamic, real-time pricing recommendations based on customer behaviour and external factors.

Data Science Canvas

Project: Price Optimization and Customer Satisfaction Prediction

Team: Barani Ranjan S, Ishika Saxena, Prasanna Kumar B V, Siva S

Problem Statement

Business Case & Value Added

Optimizing restaurant pricing based on cuisine type and delivery times to maximize profitability while enhancing customer satisfaction is the business case analysed. It reduces churn, and improve overall operational efficiency, ultimately leading to increased revenue and customer loyalty.

Data Landscape

The dataset already provides key data like **restaurant prices**, **cuisine types**, and **delivery times**. Additional data on customer preferences or operational costs might be needed for deeper insights.

Model Selection

For this analysis, **regression models** can predict optimal prices, and **classification models** can categorize customer satisfaction levels based on price, cuisine, and delivery time.

Model Requirements

Ensure the dataset includes all required variables: city, cuisine, price, delivery time, and any additional features (e.g., customer satisfaction scores).
Clean and preprocess data, handle missing values, and use feature encoding for categorical variables.
Ensure balanced classes for classification problems (e.g., satisfaction prediction).

Software & Libraries

Python
Scikit-learn
Matplotlib/Seaborn
Jupyter Notebook/
Google Colab

Skills

Data Analysis
Machine Learning
Scikit-learn.
Statistical Knowledge
Programming
Data Visualization

Execution & Evaluation

Model Evaluation

Regression (Price Optimization): Evaluate using **MAE**, **MSE**, or **RMSE** to understand prediction accuracy.
Classification (Satisfaction Prediction): Evaluate using metrics like **accuracy**, **precision**, **recall**, **F1-score**, and **confusion matrix**.

Data Storytelling

The target group needs **clear**, **visual** insights, focusing on **actionable recommendations** like optimal pricing. Use **charts**, **graphs**, and **simple explanations** to effectively communicate key findings and highlight the business impact.

Data Collection & Preparation

Data Selection & Cleansing

Relevant data includes **restaurant prices**, **cuisine types**, **delivery times** and **ratings**. Yes, the data needs cleaning to handle **missing values**, remove **duplicates**, and ensure consistency in **categorical values** like cuisine and location.

Data Integration

Combining restaurant sales data with customer reviews or feedback).

Data Collection

Additional data like **customer preferences** can be gathered through **surveys** or **restaurant reports** or **API integrations** from food delivery platforms.

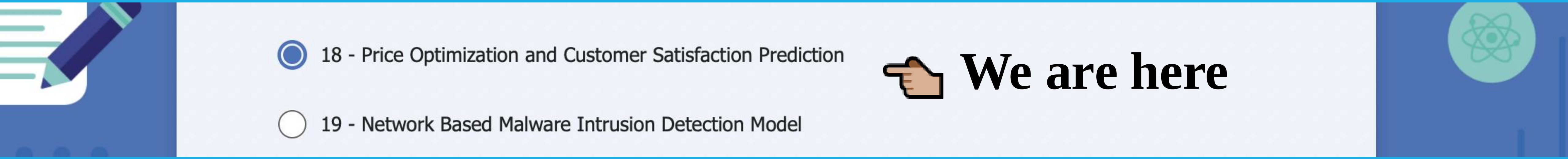
Explorative Data Analysis

Generate descriptive statistics & inferential statistics to guide the model-building process.

Adopted from: https://github.com/tomalytics/datasciencecanvas

Any Questions ??

A Gentle reminder to Provide your feedback for our work on the Peer Review Form


18 - Price Optimization and Customer Satisfaction Prediction

19 - Network Based Malware Intrusion Detection Model

We are here

THE END