**Group 6: Viraj Vijaywargiya, Samara Fernandez, Kathleen Overbaugh, Anirudh Rajeev, Vincent Perez**  
**BANA 288 Winter 2025**





## Table of Contents

| **Section** | **Title** | **Page Number** |
| --- | --- | --- |
| 1 | Data Source and References | 3 |
| 2 | Data Transformations and Feature Engineering | 3 |
| 3 | Assumptions Justification | 4 |
| 4 | Presentation Exclusions | 4 |
| 5 | Data Split Details | 4 |
| 6 | Statistical Methods and Analysis | 4 |
| 7 | Key Findings and Model Results | 5 |
| 8 | Strategic Conclusions | 6 |
| 9 | Project Insights and Limitations | 6-9 |
|  |  |  |

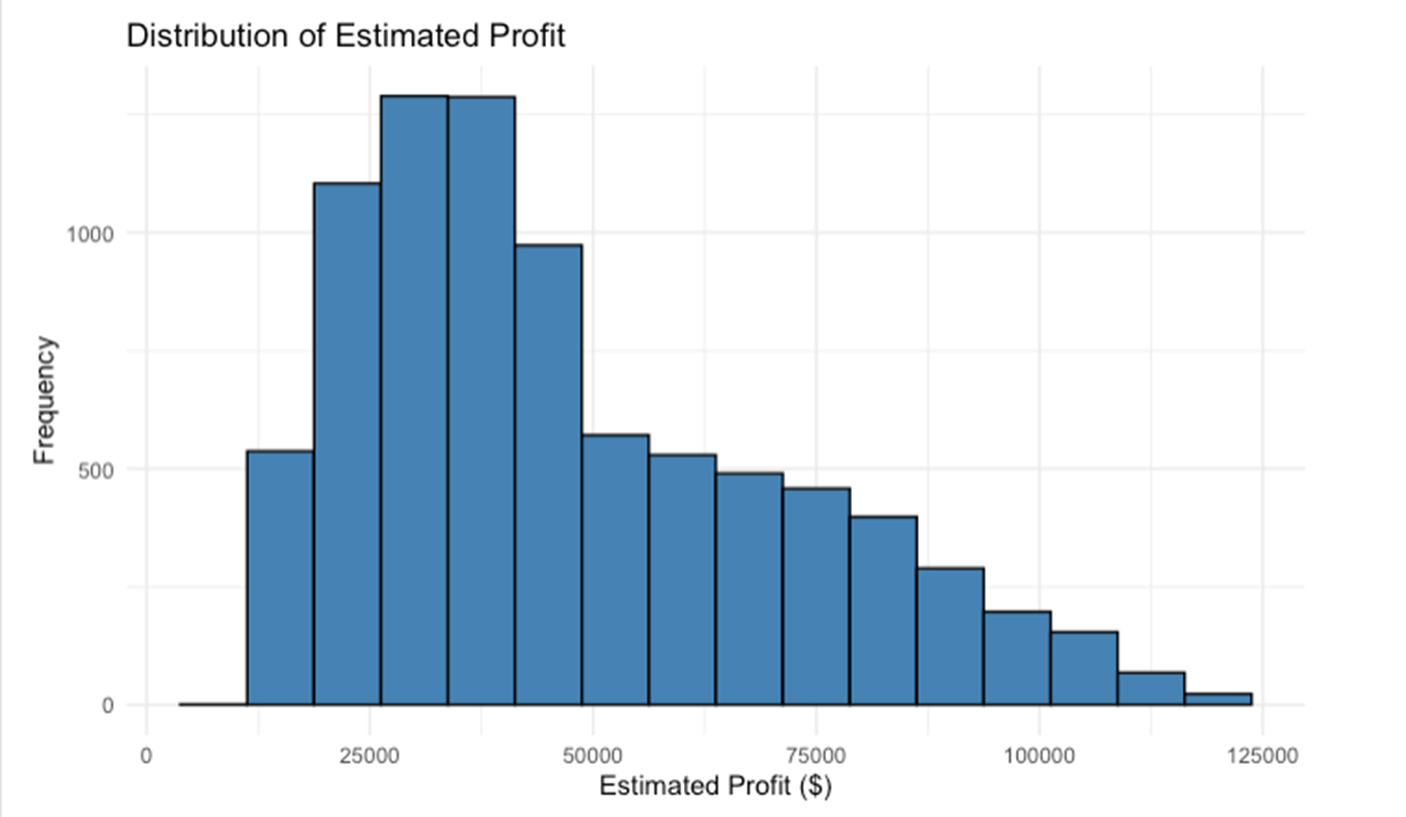
### 

### Data Source and References

* **Dataset**: [Restaurant Revenue Prediction Dataset](https://www.kaggle.com/datasets/anthonytherrien/restaurant-revenue-prediction-dataset)
* **Scope**: 8,368 records with 17 features.
* **Predictors**:
  + *Primary Drivers*: Cuisine type, seating capacity, average meal price.
  + *Secondary Drivers*: Chef experience, customer reviews, reservations.
* **Target Variable**: Profitability (Revenue - Operating Costs).

### Data Transformations and Feature Engineering

1. **Profitability Derivation**:
   * **Fine Dining** (meal price > median): Profit = **8% of revenue**.
   * **Casual Dining** (meal price ≤ median): Profit = **6% of revenue**.
   * *Rationale*: Reflects pricing strategy and margin assumptions for restaurant types.
2. **Feature Engineering**:
   * **Restaurant Segmentation**: Classified into *fine* or *casual dining* based on median meal price.
   * **Distribution Insights**: Identified a **right-skewed profit distribution**—most restaurants yield moderate profits, with limited high-profit outliers.



3. **Variable Selection**:

* + **Retained Features**:
    - *Cuisine type* (French, Japanese, Italian, Indian, American – top predictors).
    - *Seating capacity* and *average meal price* (statistically significant).
    - *Chef experience* (small but consistent impact).
  + **Excluded Features**:
    - *Location*, *marketing budget*, *social media followers* (minimal contribution via Lasso/Ridge regularization).

### Assumptions Justification

* **Profit Margins**:
  + Fine dining (8%) and casual dining (6%) profit margins align with industry benchmarks.
  + *Source*: [Upserve Restaurant Profit Margins Report](https://www.upserve.com/restaurant-insider/restaurant-profit-margins/) notes fine dining typically operates at 5–10% margins, while casual dining ranges at 3–7%.
* **Cuisine Impact**: Assumed French, Japanese, and Italian cuisines command premium pricing, validated by model outcomes.

### Presentation Exclusions

* **Excluded Models**:
  + **All-in Linear Regression**: Overfit due to redundant predictors (e.g., location).
* **Excluded Variables**:
  + *Location*, *marketing budget*, and *social media followers*: Coefficients reduced to near-zero in regularized models, indicating negligible impact on profit.

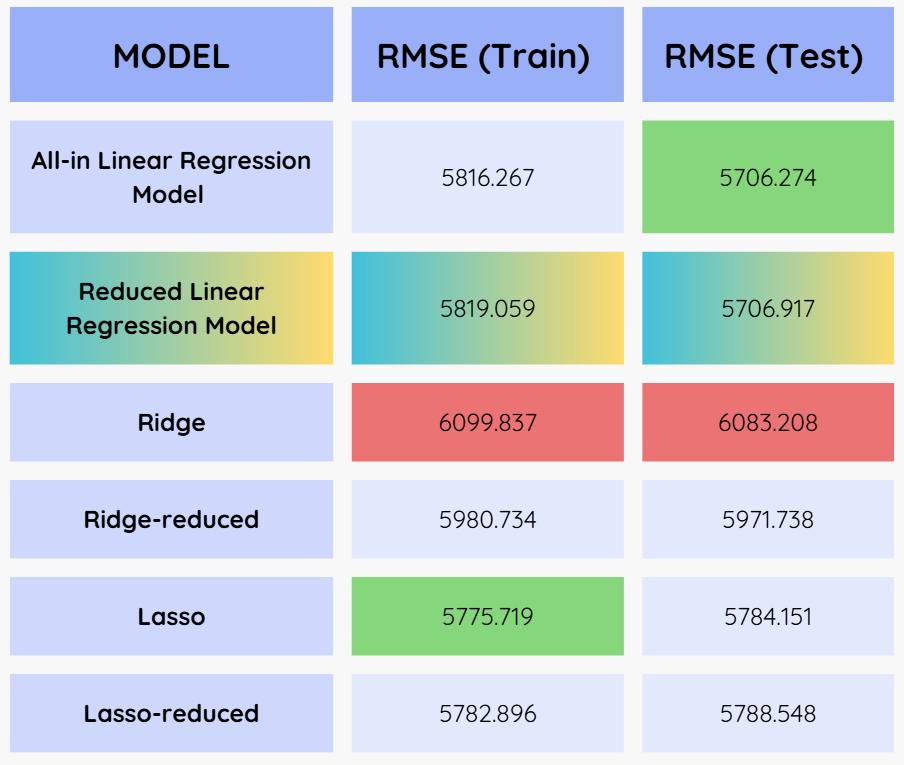
### Data Split Details

* **Split Strategy**: Standard 80:20 ratio (training: testing) for regression tasks.
* **Validation Focus**: Test RMSE prioritized to ensure model generalizability.
* **Rationale**: Avoided complex cross-validation due to dataset size and simplicity of linear models.

Statistical Methods and Analysis

1. **Regression Models**:
   * **All-in Linear Regression**: Included all 17 predictors (R² = 0.9417).
   * **Reduced Linear Regression**: Retained key variables (Adjusted R² = 0.9415).
2. **Regularization Techniques**:
   * **Ridge Regression (L2)**: Penalized coefficients to mitigate overfitting.
   * **Lasso Regression (L1)**: Eliminated irrelevant predictors by shrinking coefficients to zero.
3. **Evaluation Metrics**:
   * **RMSE**: Primary metric for comparing train/test performance.

Key Findings and Model Results



1. **Profitability Drivers**:
   * **Cuisine Type**: French, Japanese, and Italian cuisines drive **~30% higher profits** than others.
   * **Seating Capacity**: Directly proportional to profit (e.g., +100 seats ≈ +12% profit).
   * **Meal Pricing**: Higher prices correlate with profitability, especially in fine dining.
2. **Model Performance**:
   * **Optimal Model**: *Reduced Linear Regression* achieved the lowest test RMSE (**5,706.917**).
   * **Regularization Outcomes**:
     + Ridge and Lasso models reduced overfitting while retaining predictive power (test RMSE: **5,971–6,083**).
3. **Sector-Specific Trends**:
   * **Fine Dining**: Higher absolute profits but narrower margins (8% vs. 6% for casual dining).
   * **Casual Dining**: Dominates market share but is constrained by lower pricing.

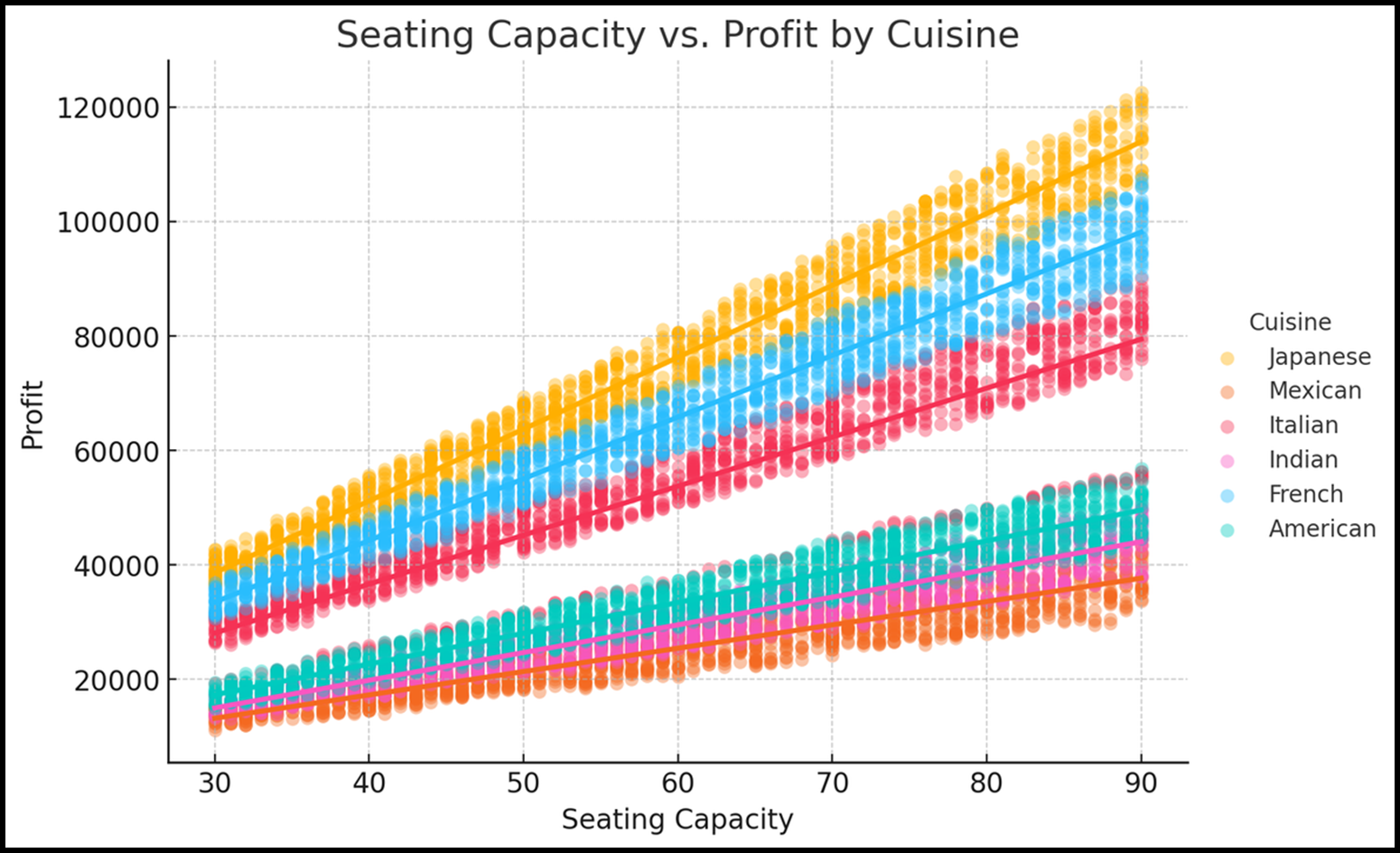
### Strategic Conclusions

1. **Prioritize High-Impact Cuisines**: French, Japanese, and Italian cuisines yield premium returns.
2. **Optimize Capacity and Pricing**:
   * Expand seating for high-demand cuisines.
   * Fine dining establishments should focus on premium pricing strategies.
3. **Scalable Framework**: The model supports dynamic adjustments for menu pricing, capacity planning, and operational efficiency.

Project Insights and Limitations

1. **Insights**:

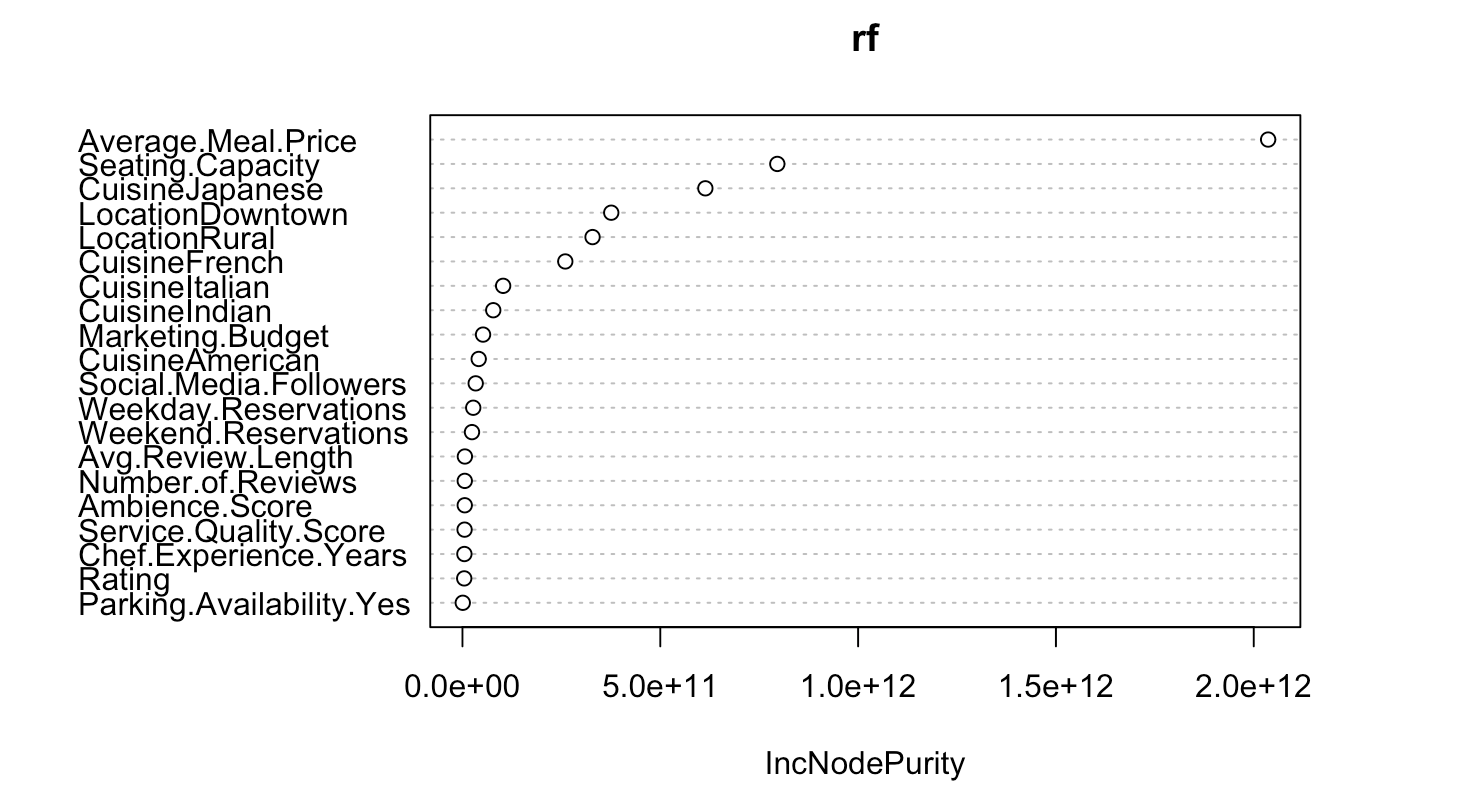
* It shows a positive correlation when meal prices increase, profit also increases.
* A wider spread at higher prices might indicate that restaurants with higher-priced meals have more variation in profit.
* A sudden shift in profit after the median average. meal price
* A clear positive correlation between seating capacity and profit, confirming it as a major profitability driver.
* Japanese and French cuisines have a steeper trend, meaning they benefit more from increased seating than other cuisines.
* American and Indian restaurants show a lower increase in profit per additional seat.



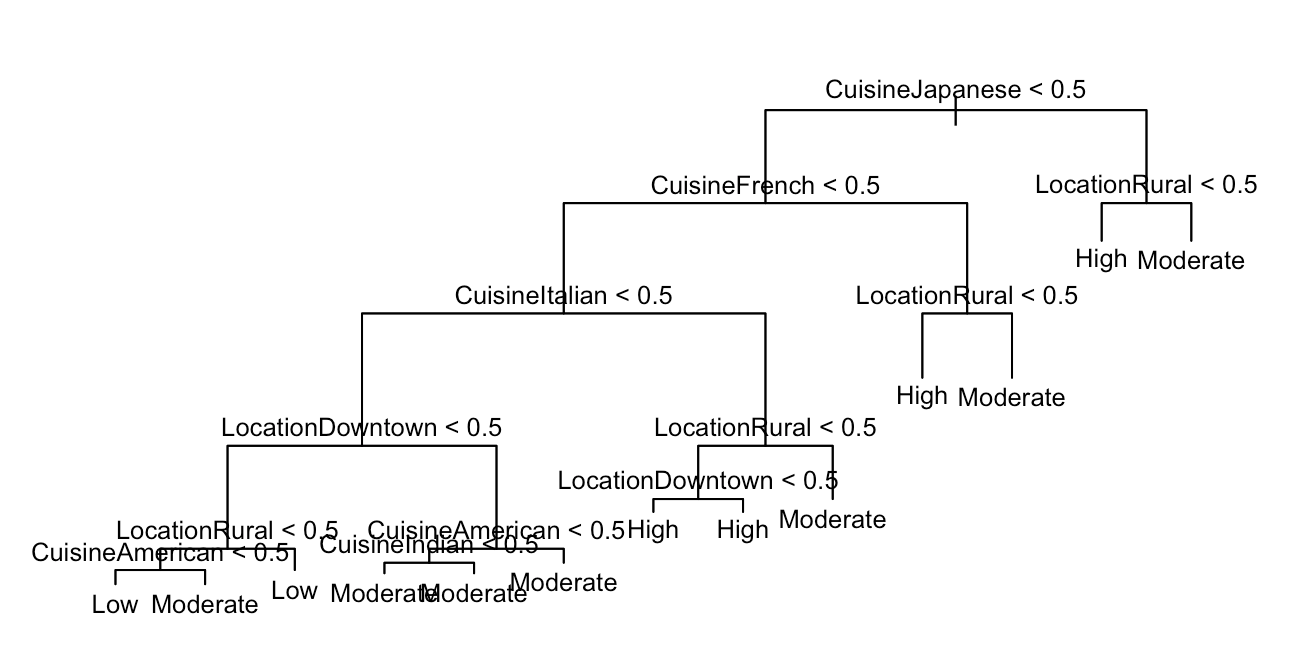
1. **Limitations**:

* Profit calculations exclude operational costs (labor, rent).
* Cuisine trends (e.g., saturation of French/Japanese restaurants) may reduce future returns.
* Irrelevant variables (e.g., location) dilute model accuracy; regularization improves focus.
* Overfitting risks exist in complex models (e.g., All-in Regression).
* Assumed profit margins (6–8%) may not reflect actual cost structures.
* Excluded variables like *competitor density* or *seasonality* could enhance predictions.

1. **Classification Models Not Used for Final Modeling:**
   1. **Random Forest:**
      1. Although we tested a Random Forest model using Estimated Profit as the dependent variable, we ultimately did not use it for our final analysis. Here’s why:
         1. Random Forest is primarily a classifier and does not provide clear insights into the magnitude or direction of variable influence.
      2. Our project goal was to understand which specific variables significantly impact profitability and to interpret their contributions. We therefore chose linear regression models.



* + 1. While Random Forest confinement that Average Meal Price, Seating Capacity, and certain Cuisine Types are strong predictors of profit, it was less useful for interpretation, which was critical for a strategic business recommendation. This information would’ve been more useful if we were to classify our restaurants into profitability tiers (“High” vs. “Low”) then Random Forest or other classifiers would have been great.
  1. **Decision Tree:**
     1. Although we initially explored classification trees as a way to analyze restaurant profitability, we ultimately decided not to use them in our final analysis. In this model, profit was categorized into three levels—Low, Moderate, and High—and a decision tree was constructed based on these classes. While the visual output of the tree was interpretable and highlighted factors like cuisine type and location, it came with notable limitations that made it less suitable for our objectives.



* + - 1. The most significant issue was the loss of information, by converting profit values into categorical. This reduced the granularity of the data and weakened our ability to measure nuanced differences in financial outcomes.
      2. The classification tree is primarily split based on categorical variables, such as whether a restaurant served Japanese cuisine or was located in a rural area. However, it failed to fully leverage important numerical predictors like average meal price and seating capacity. Two variables that our regression models consistently identified as strong drivers of profitability. This omission further limited the tree’s usefulness for generating detailed, actionable insights.

1. **Future Enhancements**:
   * Integrate competitor analysis and hyperparameter tuning.
   * Expand the dataset to include regional economic factors.