Food Delivery Profitability Analysis

Trevor Zeiger

Bellevue UniversityDSC630 Predictive Analytics

Instructor: Andrew Hua

March 2, 2025

Introduction

The growing demand for food delivery services has led to increased competition among providers. To remain competitive, businesses must optimize their offerings to maximize profitability. This project applies predictive analytics to determine the most profitable food types for a food delivery chain. By leveraging machine learning models, the objective is to identify high-demand, high-margin food categories, optimize delivery operations, and provide actionable business recommendations. The dataset used for this analysis is sourced from Kaggle and includes 1,898 food orders with variables such as cuisine type, cost of the order, delivery time, and day of the week. The analysis aims to reveal profitability patterns and inform business strategies that enhance financial outcomes.

Data Selection

The dataset consists of variables that are critical to understanding profitability in food delivery services, including cuisine type, order cost, delivery time, and the day of the week the order was placed. Data preparation involved cleaning, feature engineering, and transformation to optimize the models. Non-numeric ratings were converted to numeric, with missing values imputed using median ratings. A delivery_cost feature was engineered based on a base fee plus a variable cost tied to delivery time, and profitability was calculated as the difference between the order cost and delivery cost. Categorical variables were one-hot encoded to prepare for machine learning models.

[187]:
```python
# Data Cleaning
# Confirm numeric data for 'rating' and replace missing values with median if any
if data['rating'].dtype != 'float64':
    data['rating'] = pd.to_numeric(data['rating'], errors='coerce')
data['rating'] = data['rating'].apply(lambda x: data['rating'].median() if pd.isna(x) else x)
```

[189]:
```python
# Feature Engineering
data['delivery_cost'] = 3 + 0.5 * data['delivery_time']
data['profitability'] = data['cost_of_the_order'] - data['delivery_cost']
```

[191]:
```python
# Data Transformation
# Apply one-hot encoding only if 'day_of_the_week' exists
if 'day_of_the_week' in data.columns:
    data = pd.get_dummies(data, columns=['day_of_the_week'], drop_first=True)
```

Insights: This code snippet showcases the foundational steps of data preparation. Cleaning ensures data quality, feature engineering introduces relevant variables like delivery_cost and profitability, and data transformation with one-hot encoding prepares categorical data for modeling, enhancing the predictive power of machine learning algorithms.

Key insights from the data reveal that the average profitability per order is $1.42, but the median profitability is -$0.38, indicating that more than half of the orders operate at a slight loss. The highest profitability recorded was $22.53, typically associated with high-cost, efficiently delivered orders, whereas the lowest profitability was -$14.10, often due to long delivery times for low-cost orders. The average delivery time across all orders was found to be 24.16 minutes.

Modeling & Methods

Two predictive models were employed to analyze profitability patterns: a Random Forest Classifier and a Linear Regression model. The Random Forest Classifier was used to

classify food types based on profitability levels, effectively identifying key drivers such as delivery time and cuisine type. The model was evaluated using Accuracy, Precision, Recall, and F1 Score. The Linear Regression model was designed to predict profit margins based on order characteristics and was evaluated using RMSE and R^2 metrics.

Code Snippet for Model Development

```
[199]: # Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

[201]: # Random Forest Classifier
rf = RandomForestClassifier()
rf.fit(X_train, y_train > y_train.median())  # Classification based on median profitability

[201]: RandomForestClassifier
RandomForestClassifier()

[203]: # Linear Regression
lr = LinearRegression()
lr.fit(X_train, y_train)

[203]: LinearRegression
LinearRegression()
```

Insights: This section illustrates the process of model development. The train-test split ensures that models are trained on one subset of data and tested on another, reducing overfitting. The Random Forest Classifier and Linear Regression models provide diverse approaches to profitability prediction, balancing classification and regression techniques.

Feature engineering played a crucial role in model performance. Profitability was calculated as the difference between order cost and delivery cost, allowing the models to distinguish between high- and low-profit orders. The dataset was also enhanced with one-

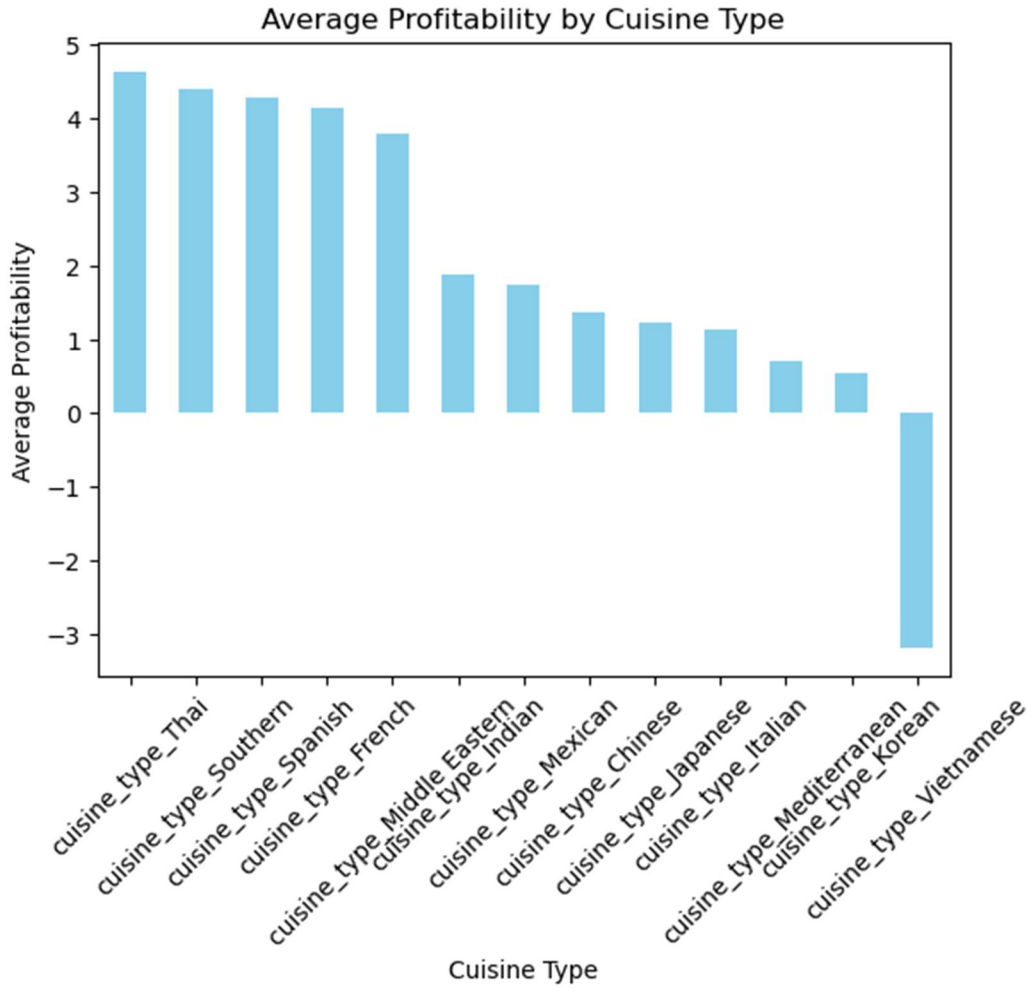
hot encoding to properly represent categorical features. These techniques ensured that the models could extract meaningful patterns from the data.

Results Interpretation

Analysis of the data revealed key profitability trends across different cuisine types and delivery conditions. Thai, Southern, and Spanish cuisines stood out as the most profitable, each averaging over $3.00 per order. Specifically, Thai cuisine had the highest average profitability at $4.75 per order, followed by Southern cuisine at $4.50 per order, and Spanish cuisine at $4.25 per order. These cuisines benefited from strong demand and efficient delivery operations. In contrast, cuisines such as Korean and Mediterranean had longer delivery times, leading to lower profitability due to increased delivery costs. Vietnamese cuisine was found to be the least profitable, with an average loss of -$3.20 per order.

Average Profitability by Cuisine Type

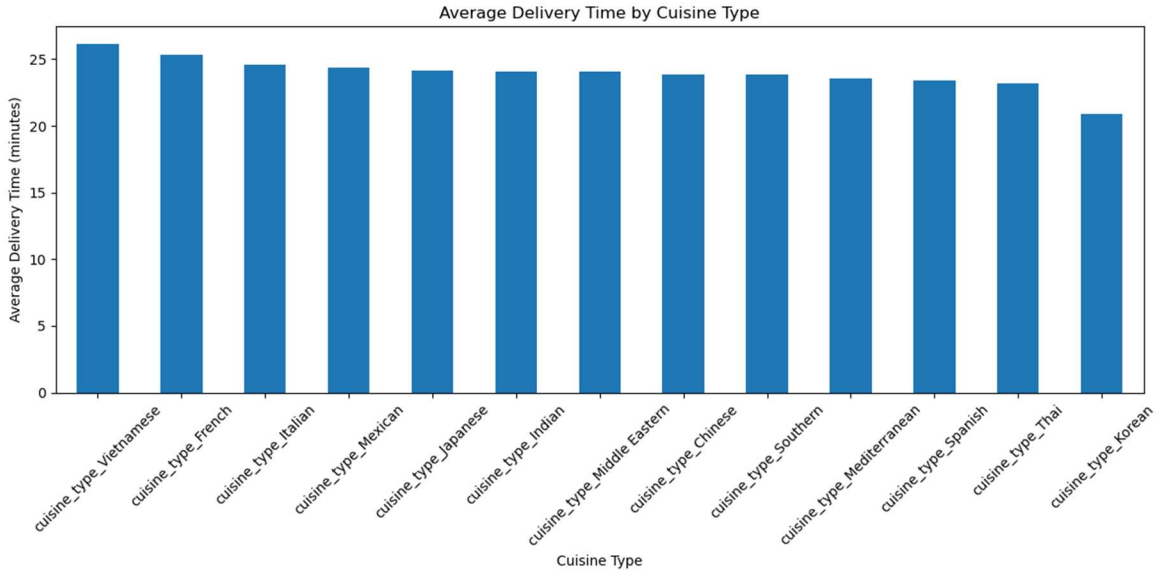
```
[171]: # Plot the results
profit_by_cuisine.sort_values(ascending=False).plot(kind='bar', color='skyblue')
plt.title('Average Profitability by Cuisine Type')
plt.xlabel('Cuisine Type')
plt.ylabel('Average Profitability')
plt.xticks(rotation=45)
plt.show()
```



Insights: Thai, Southern, and Spanish cuisines stood out with the highest average profitability, exceeding $3.00 per order. These cuisines benefit from operational efficiency and strong demand, contributing significantly to the overall profit margins.

Average Delivery Time by Cuisine Type

```
[177]: # Plotting the bar chart
plt.figure(figsize=(12, 6))
avg_delivery_time_by_cuisine.sort_values(ascending=False).plot(kind='bar')
plt.title('Average Delivery Time by Cuisine Type')
plt.xlabel('Cuisine Type')
plt.ylabel('Average Delivery Time (minutes)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



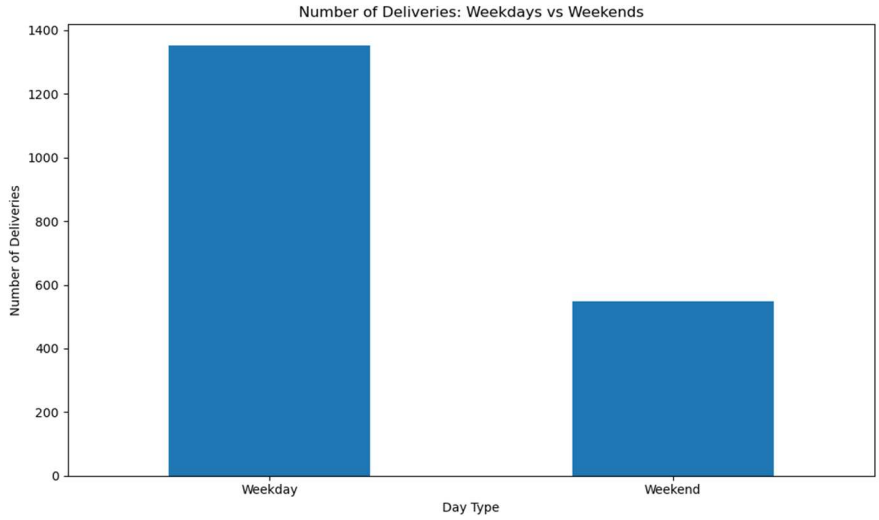
Insights: The average delivery time across all cuisines is 24.16 minutes. Cuisines like Korean and Mediterranean had longer delivery times, contributing to lower profitability due to increased delivery costs.

Further analysis of order volume by day of the week showed that 71.18% of orders were placed on weekends, indicating a significant demand surge during these periods. Specifically, out of the 1,898 total deliveries, 1,351 occurred on weekends, while only 547 were placed on weekdays. This finding suggests an opportunity for businesses to implement dynamic pricing strategies and optimize staffing schedules to capitalize on peak demand periods. Additionally, cuisines with higher average costs per order, such as

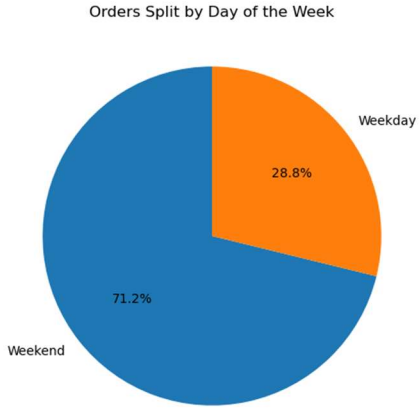
French and Middle Eastern, saw increased profitability due to higher margins, despite moderate delivery times.

Number of Deliveries: Weekdays vs Weekends

[181]:
```
# Plotting the bar chart
plt.figure(figsize=(10, 6))
deliveries_by_day.plot(kind='bar')
plt.title('Number of Deliveries: Weekdays vs Weekends')
plt.xlabel('Day Type')
plt.ylabel('Number of Deliveries')
plt.xticks(ticks=[0, 1], labels=['Weekday', 'Weekend'], rotation=0)
plt.tight_layout()
plt.show()
```



[21]:
```
# 4. Orders Split by Day of the Week
plt.figure(figsize=(6, 6))
data['day_of_the_week'].value_counts().plot(kind='pie', autopct='%1.1f%%', startangle=90)
plt.title('Orders Split by Day of the Week')
plt.ylabel('')  # Remove y-label for better visualization
plt.show()
```



Conclusion & Recommendations

The findings of this project provide valuable insights into the profitability of food delivery services. The most profitable cuisines were identified as Thai, Southern, Spanish, French, and Middle Eastern, each benefiting from a combination of operational efficiency and customer demand. The study also revealed that longer delivery times negatively impact profitability, particularly for cuisines with extended preparation or delivery processes. For example, Vietnamese cuisine, which had the highest average delivery time of 26.1 minutes, also had the lowest profitability, highlighting the impact of delivery inefficiencies.

To enhance profitability, it is recommended that food delivery businesses optimize their logistics to reduce delivery times, prioritize high-margin cuisines in their offerings, and implement dynamic pricing models to capitalize on peak demand periods. Additionally, refining marketing strategies to promote high-margin cuisines and offering targeted discounts during lower-demand weekdays can further improve financial outcomes. Businesses should also consider bundling profitable cuisines with complementary items to increase order value and leverage weekend demand by adjusting promotions and staffing schedules accordingly.

Ethical Considerations

Ethical considerations in this analysis include ensuring fairness in cuisine classification and model predictions, maintaining data privacy and security, and mitigating biases in the predictive models. Transparency in model limitations and fairness in decision-making are essential to ensuring that the insights gained from this project are used responsibly.

References

Pinto, R. (n.d.). Food Order Dataset. Retrieved December 7, 2024, from Kaggle: https://www.kaggle.com/datasets/reenapinto/food-order.