Food Delivery Intelligence System with Predictive Analytics

Applied Data Science and Business Intelligence Project

Project Repository: `https://github.com/soufianboukir/food-delivery-analytics`

Dashboard URL: `https://food-delivery-intelligence-system.streamlit.app/`

Soufian Boukir

Abstract

In this project, we developed an end-to-end data science pipeline for a food delivery dataset containing over 15,000 records. The project covers the complete workflow, including data cleaning, exploratory data analysis (EDA), and predictive modeling. Regression models were implemented to predict delivery time, while classification models were used to determine whether an order is likely to be canceled. In addition, an interactive Streamlit dashboard was developed to visualize insights and allow users to interact with the predictive models in real time.

Keywords: Food-delivery Analytics, Machine learning, Business Intelligence

1 Introduction

The food delivery industry has experienced significant growth in recent years, driven by the increasing adoption of online platforms and mobile applications. As customer expectations continue to rise, companies must optimize delivery operations, reduce delays, and improve customer satisfaction in order to remain competitive. Data science and machine learning provide effective solutions for analyzing operational performance and generating predictive insights that support decision-making.

This project presents an end-to-end food delivery analytics and machine learning platform developed using a dataset containing more than 15,000 records. The project begins with data preprocessing and exploratory data analysis to understand customer behavior, delivery patterns, and operational trends.

Two main predictive tasks were implemented. The first task focuses on delivery time prediction using regression models, while the second focuses on order cancellation prediction using classification models. Multiple machine learning algorithms were evaluated and compared based on appropriate performance metrics in order to identify the most effective models.

In addition to predictive modeling, an interactive dashboard was developed using Streamlit to provide data visualization, operational insights, and real-time interaction with the machine learning models through a user-friendly interface. The project also integrates modern software engineering practices, including model serialization, testing, Docker containerization, CI/CD pipelines, and API deployment via AWS EC2, resulting in a complete and production-oriented data science workflow.

2 Data Understanding and Preparation

2.1 Dataset description

Food Delivery Operations and Customer Analytics, This dataset presents a realistic synthetic representation of modern food delivery operations across urban environments. It captures key operational, customer, and delivery-related metrics that influence overall platform performance and customer experience.

The dataset includes information related to delivery timings, traffic conditions, weather impact, customer behavior, order values, discounts, ratings, cancellations, refunds, and delivery efficiency. The data has been structured to reflect realistic business patterns and operational relationships commonly observed in food delivery platforms.

Designed with a strong focus on realism and analytical depth, the dataset offers balanced numerical features suitable for exploratory analysis, visualization, operational monitoring, and predictive modeling.

Dataset Highlights

- 15,000 realistic synthetic records
- 30 business-oriented numerical features
- Operational and customer-focused metrics
- Realistic delivery and order behavior patterns
- Suitable for modern analytics workflows
- Structured for visualization and dashboarding

Key Areas Covered

- Delivery Performance
- Customer Behavior
- Order Trends
- Traffic & Weather Impact
- Ratings & Satisfaction
- Operational Efficiency
- Discounts & Revenue Metrics
- Cancellation & Refund Analysis

2.2 Data Cleaning

Data cleaning is a fundamental step in the data analysis process that involves identifying and correcting inaccurate, inconsistent, or incomplete data in order to improve data quality and ensure reliable analysis and modeling results.

For this dataset, several data cleaning operations were performed, including checking for missing values, removing duplicate records, validating data types, and verifying that numerical variables fall within logical and expected ranges. These steps helped ensure the consistency, integrity, and usability of the dataset before proceeding to exploratory data analysis and machine learning tasks.

2.3 Exploratory Data Analysis

Exploratory Data Analysis is an important step in data analysis where we explore, summarize, and visualize data to understand its structure, detect patterns, identify anomalies, test assumptions, and check relationships between variables before applying any machine learning or statistical models.

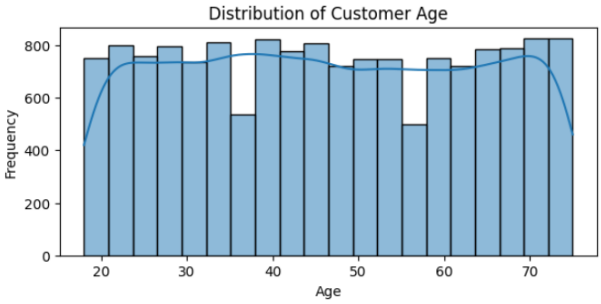

Figure 1: Customer age distribution.

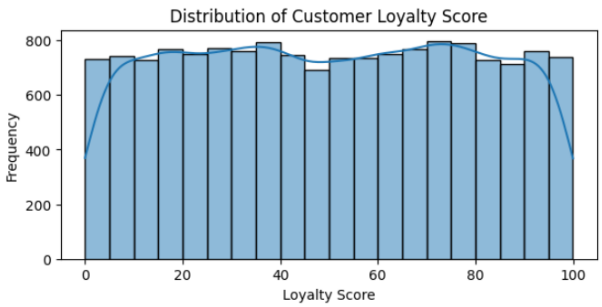

Figure 2: Customer loyalty score.

Figure 1: Customers' ages range between 18 and 75 and are approximately uniformly distributed. The intervals 35-38 and 55-58 represent relatively lower frequencies compared to the rest of the population.

Figure 2: Customer loyalty scores range between 0 and 100 and are approximately uniformly distributed across the entire population.

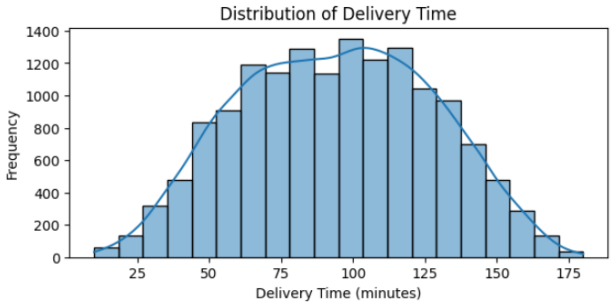

Figure 3: Delivery time distribution.

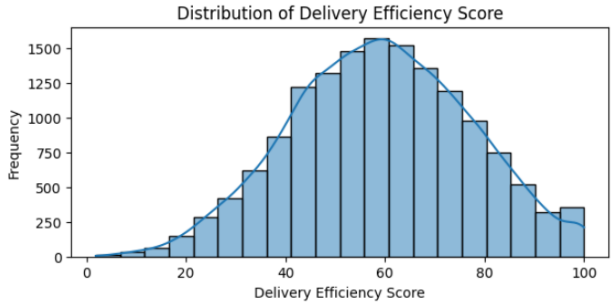

Figure 4: Customer loyalty score.

Figure 3: Delivery time follows an approximately normal distribution, ranging between 10 and 175 minutes, with about 68% of orders falling within the interval of 50 to 130 minutes.

Figure 4: Delivery efficiency scores are also approximately normally distributed between 0 and 100. Higher values are observed near the upper bound, and about 68% of the observations lie between 35 and 80.

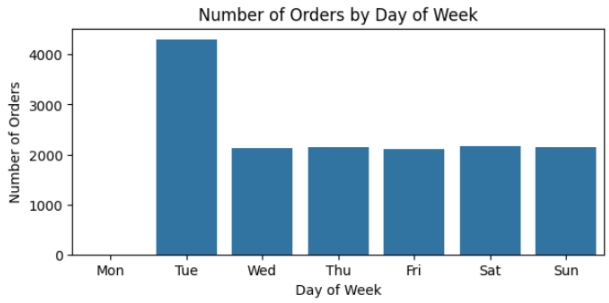

Figure 5: Number of orders per day of week.

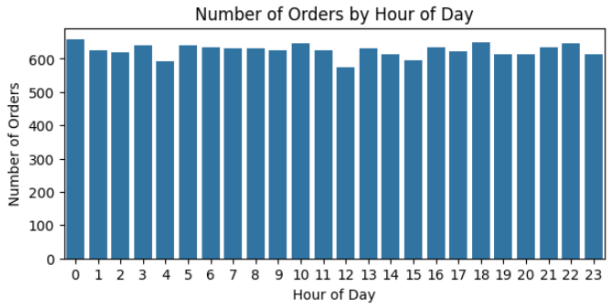

Figure 6: Number of orders per hour of day.

Figure 5: Tuesday has the highest number of orders, with more than 4,000 orders. Monday shows no recorded orders, while the remaining days are approximately uniformly distributed, with around 2,000 orders per day.

Figure 6: All hours of day are approximately uniformly distributed, with roughly 600 orders per hour.

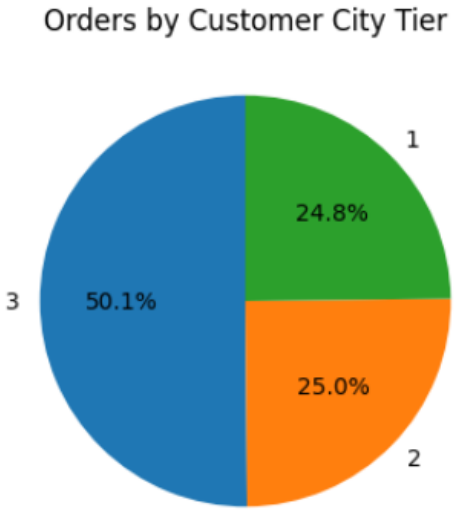

Figure 7: Orders by customer city tier.

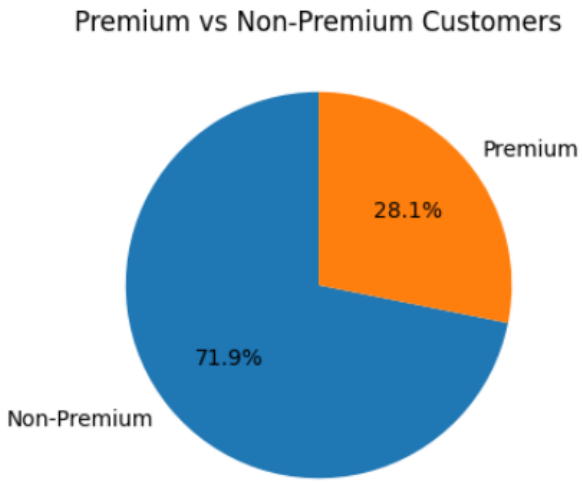

Figure 8: Premium vs non premium customers.

Figure 7: City Tier 3 has the highest number of customer orders, accounting for more than 50% of all orders, while other City Tiers contribute approximately 25%.

Figure 8: Non-premium customers account for more than 71% of all orders, while premium customers represent approximately 28%.

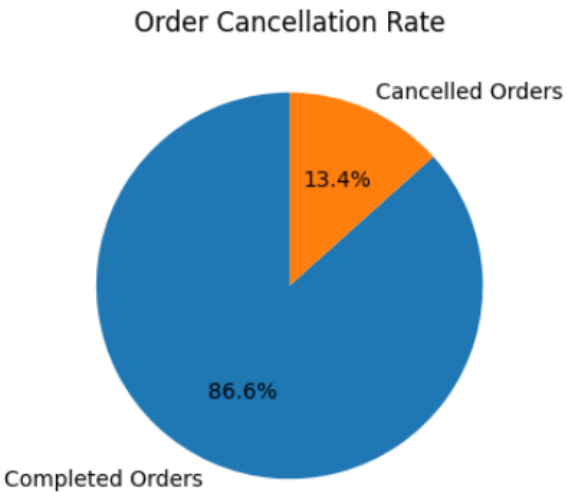

Figure 9: Orders cancellation rate.

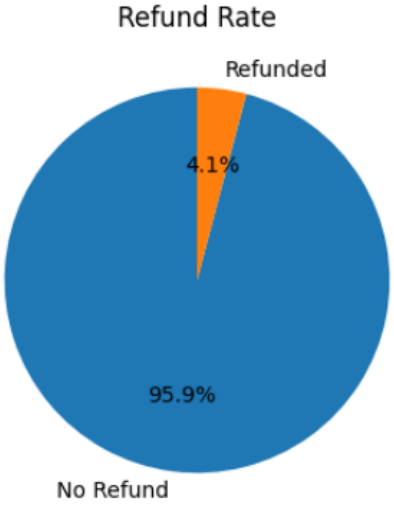

Figure 10: Refund rate.

Figure 9: More than 85% of orders are successfully completed, while the remaining orders are cancelled.

Figure 10: Refunded orders account for approximately 4% of the total orders.

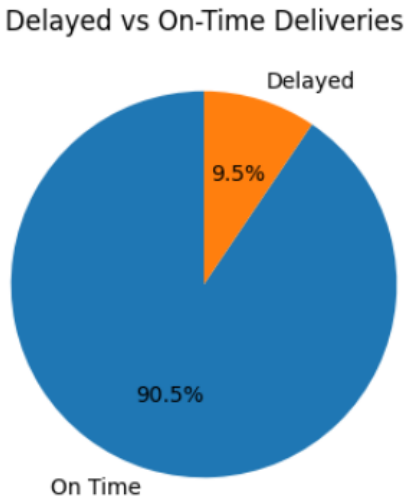

Figure 11: Distribution of delayed vs on-time orders.

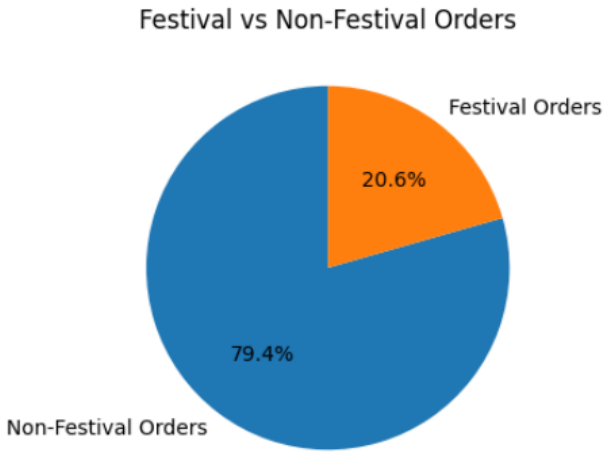

Figure 12: Festivale vs non festivale distributions.

Figure 11: More than 90% of orders are delivered on time, indicating a high level of service reliability and time efficiency.

Figure 12: Non-festival orders account for more than 79% of the total orders.

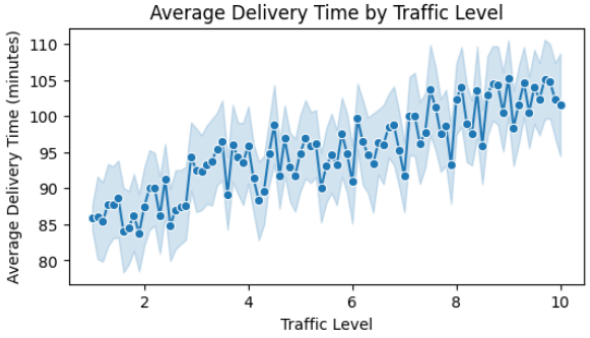

Figure 13: Average delivery time by traffic level.

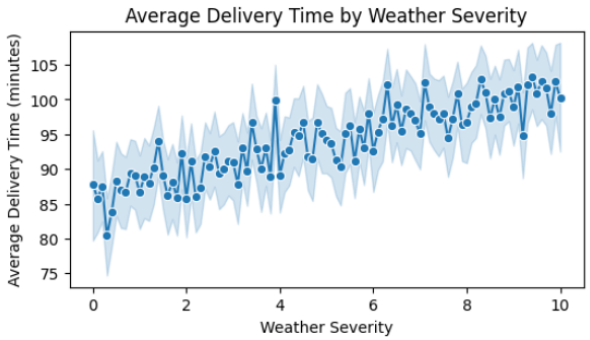

Figure 14: Average delivery time by weather severity.

Figure 13: Average delivery time increases steadily with traffic level, rising from around 85 minutes at level 1 to approximately 100-105 minutes at level 10.

Figure 14: Delivery time follows a similar upward trend with weather severity, starting near 85-90 minutes at low severity and reaching around 100-105 minutes at severity 10.

Figure 15: Distribution of customer ratings.

Figure 16: Distribution of restaurant ratings.

Figure 15: Customer ratings are concentrated between 3.5 and 4.5, with a clear peak at around 4.0, suggesting most customers are fairly satisfied.

Figure 16: Restaurant ratings show a bimodal-like pattern with notable peaks around 3.8 and 4.25, indicating a split in restaurant quality perception among customers.

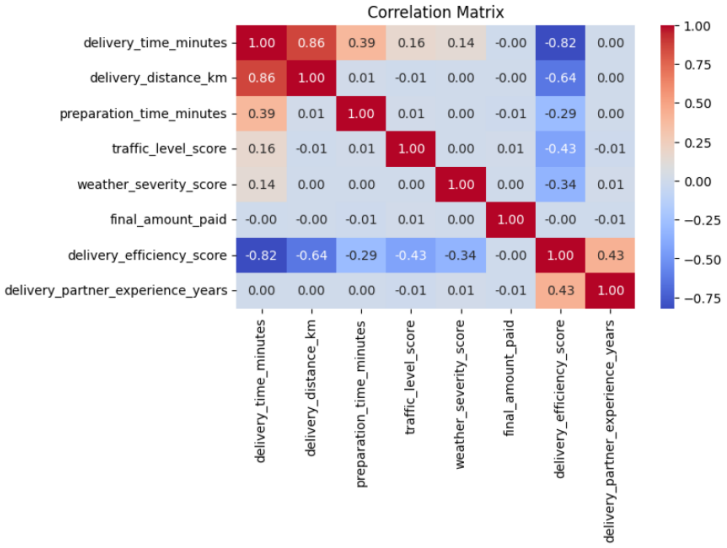

Figure 17: Distribution of customer ratings.

Figure 17: Correlation matrix of num colunns

Delivery time is most strongly influenced by delivery distance (0.86) and negatively by delivery efficiency score (-0.82), meaning longer distances increase delivery time while higher efficiency reduces it. Preparation time, traffic level, and weather severity show moderate positive correlations with delivery time, while final amount paid and partner experience years have virtually no impact on it.

3 Modeling

To predict whether a customer will cancel an order and to estimate the delivery time for a given order, multiple machine learning approaches were applied for each task in order to compare their performance and select the most suitable model.

3.1 Regression Task

For the regression task, which consists of predicting the delivery time of an order, three machine learning models were implemented and evaluated: Linear Regression, Random Forest Regressor, and XGBoost Regressor.

Linear Regression was used as a baseline model due to its simplicity and interpretability. Random Forest was applied to capture non-linear relationships between features, while XGBoost was selected for its strong predictive performance and ability to handle complex patterns efficiently.

To predict the delivery time in minutes, six features were selected based on their strong relationship with the target variable and their direct impact on delivery operations:

- delivery_distance_km
- preparation_time_minutes
- traffic_level_score
- weather_severity_score
- delivery_efficiency_score
- delivery_partner_experience_years

These features represent operational, environmental, and driver-related factors that significantly influence the estimated delivery duration.

Figure 18: Linear regression Actual vs Predicted.

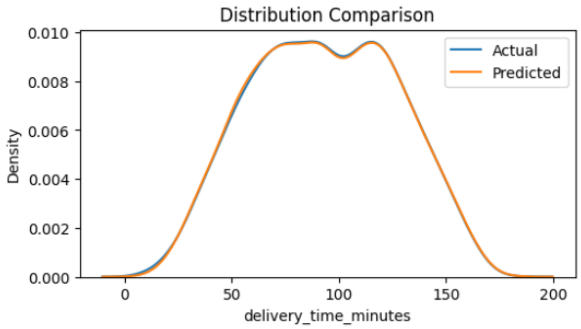

Figure 19: Linear regression: distribution comparison.

Figure 18: Predicted values align very closely with actual delivery times, with most points falling tightly along the red diagonal line, indicating a highly accurate model. A small cluster of outliers appears in the lower range, suggesting slight over-prediction for very short deliveries, but overall model performance is excellent.

Figure 19: The density curves for actual and predicted delivery times overlap almost perfectly across the full range (0-200 minutes), confirming that the model successfully captures the overall distribution shape.

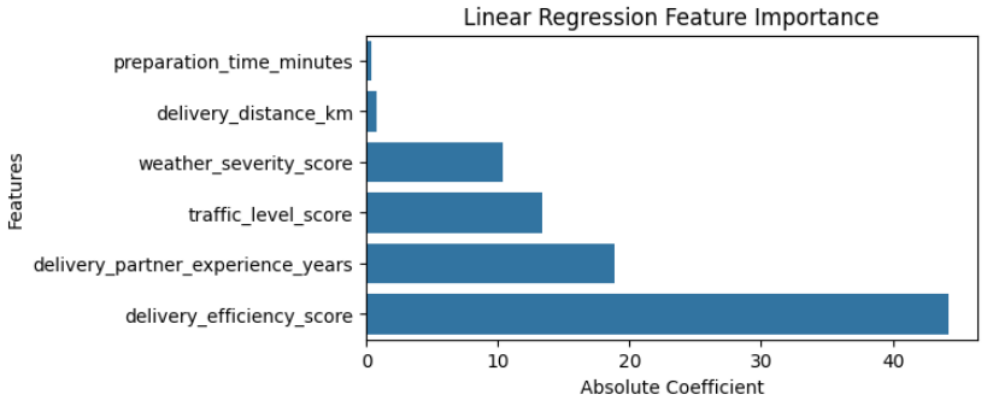

Figure 20: Linear regression: feature importance.

Figure 20: The Linear Regression model identified the delivery efficiency score as the most influential feature, assigning it the highest coefficient value. This was followed by delivery partner experience and preparation time. In contrast, delivery distance had the lowest impact on the predicted delivery time among the selected features.

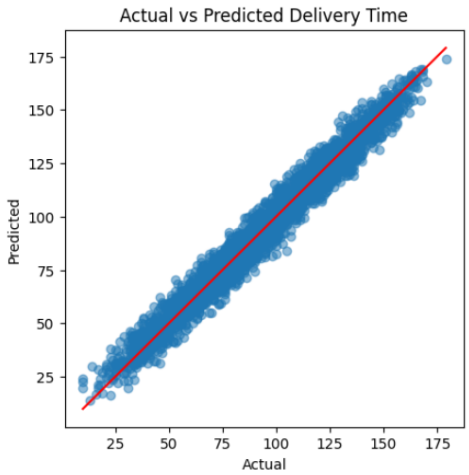

Figure 21: Random forest regressor Actual vs Predicted.

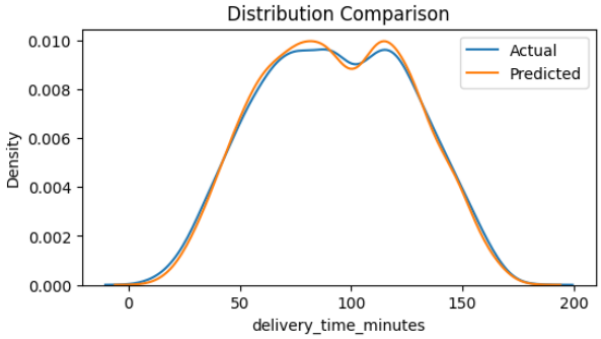

Figure 22: Random forest regressor: distribution comparison.

Figure 21: Predicted values align closely with actual delivery times, with approximately 50% points falling tightly along the red diagonal line, indicating a not highly accurate model.

Figure 22: The density curves for actual and predicted delivery times overlap almost perfectly across the full range (0-200 minutes), the plot show that between delivery time 75-125, the model prdecit them not correct.

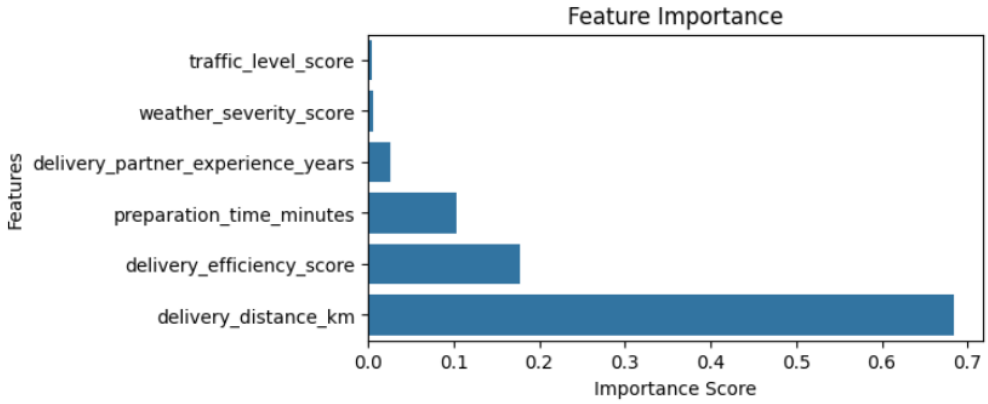

Figure 23: Random forest regressor: feature importance.

Figure 23: The Random forest regressor model identified the delivery distance as the most influential feature. This was followed by delivery efficienty score. In contrast, traffic level score had the lowest impact on the predicted delivery time among the selected features.

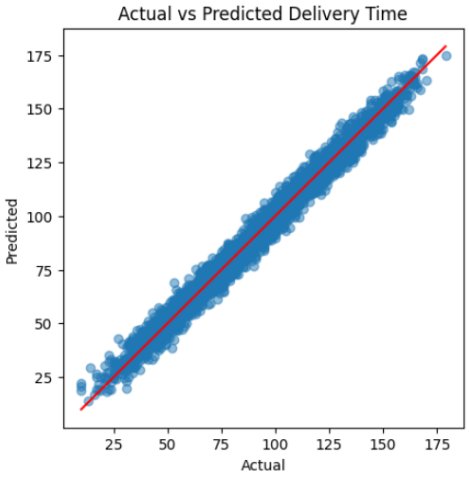

Figure 24: XGboost Actual vs Predicted.

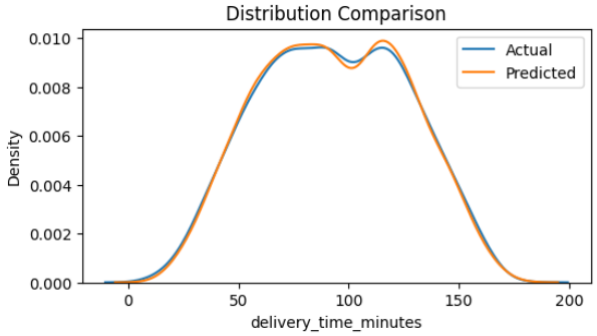

Figure 25: XGboost: distribution comparison.

Figure 24: Predicted values align closely with actual delivery times (better than random forest regressor), with approximately 70% points falling tightly along the red diagonal line, indicating an accurate model.

Figure 25: The density curves for actual and predicted delivery times overlap almost perfectly across the full range (0-200 minutes), the plot show that between delivery time 75-125, the model prdecit them not correct.

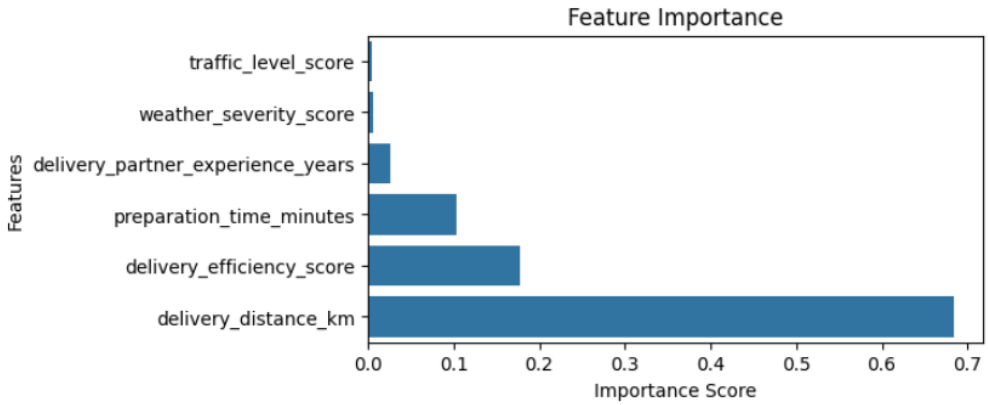

Figure 26: XGboost: feature importance.

Figure 26: Same as random forest regressor since they are models based on trees archeticture.

Model comparison and selection

Table 1: Regression model evaluation metrics.

Model	MAE	RMSE	R^2
Linear Regression	**0.42**	**1.41**	**0.9983**
Random Regressor	4.47	5.63	0.9721
XGboost	3.41	4.28	0.9839

The evaluation results presented above indicate that Linear Regression achieves the best overall performance across all metrics, with the lowest MAE and RMSE as well as the highest R^2 score. This suggests that the relationship between the selected features and the target variable is largely linear, which aligns well with the assumptions of the model.

In contrast, both Random Forest and XGBoost exhibit lower performance in this specific setting. This may be due to the underlying data structure being relatively simple, where complex non-linear models are not necessary and may introduce unnecessary variance.

Based on these results, Linear Regression was selected as the final model for the delivery time prediction task and deployed within the API for inference.

3.2 Classification task

For the classification task, which consists of predicting the order cancellation, three machine learning models were implemented and evaluated: Logistic Regression, Random Forest, and XGBoost.

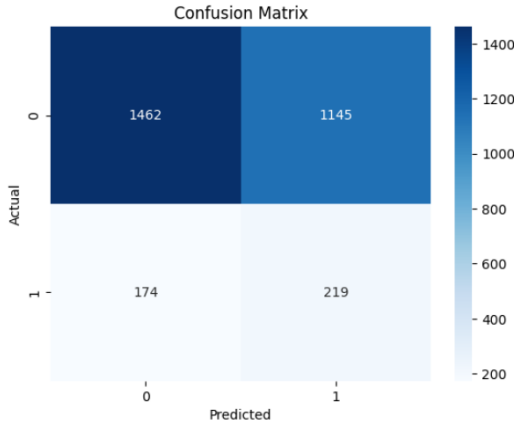

Figure 27: Logoistic regression: Confusion matrix.

Figure 27: The model strongly predicts class 0 but struggles with class 1. Out of 393 actual class-1 cases, it only correctly identified 219, missing 174. It also misclassified 1,145 actual class-0 cases as class 1. Overall, it has a significant false positive rate and moderate recall for the minority class.

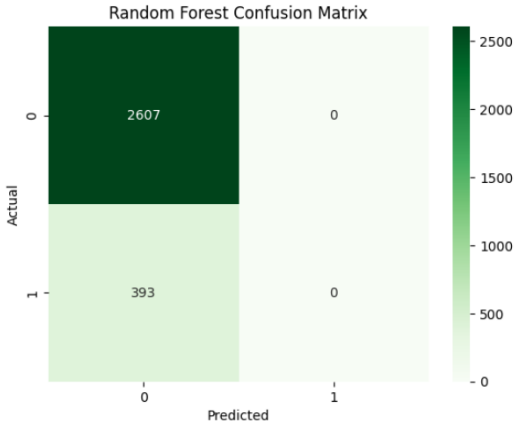

Figure 28: Random forest classifier: Confusion matrix.

Figure 28: The model predicted everything as class 0 — it classified all 393 positive cases as negative (zero true positives). While it perfectly identified all 2,607 negatives,

it completely failed to detect any class-1 instance. This suggests the model is heavily biased toward the majority class, likely due to class imbalance.

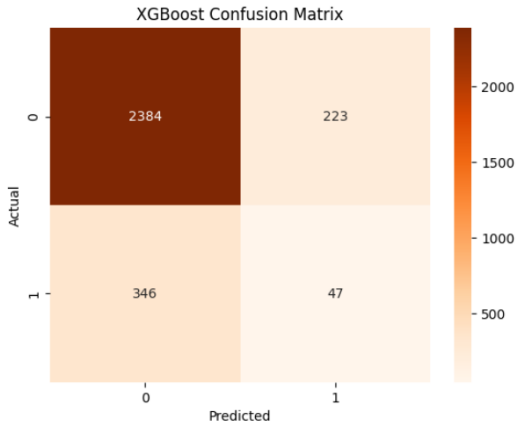

Figure 29: XGboost: Confusion matrix.

Figure 28: XGboost Still struggling with class 1. It correctly identified 2,384 negatives and caught 47 true positives, but missed 346 actual class-1 cases and had 223 false positives. It at least detects some positives, unlike Random Forest, but recall for class 1 remains low (12%).

Model comparison and selection

Table 2: Classification model evaluation metrics.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.5603	0.1606	**0.5573**	**0.2493**
Random Regressor Classifier	**0.8693**	0	0	0
XGboost	0.8103	**0.1741**	0.1196	0.1418

It is important to note that the dataset is inherently imbalanced, comprising approximately 12,000 completed orders versus only 3,000 cancelled orders, which directly influenced the models' behavior and performance.

The Random Forest Classifier achieved the highest accuracy (86.93%), however this result is misleading — the model predicted every instance as the majority class (completed), resulting in a precision, recall, and F1-score of 0, meaning it entirely failed to detect any cancelled order. This is a classic symptom of class imbalance, where a model maximizes accuracy by ignoring the minority class.

Logistic Regression showed the lowest accuracy (56.03%) but obtained the highest recall (55.73%) and F1-score (24.93%) among the three models, indicating that it was at least partially capable of identifying cancelled orders, at the cost of many false positives (precision of 16.06%).

XGBoost achieved a moderate accuracy of 81.03%, with a precision of 17.41% and a recall of only 11.96%, yielding an F1-score of 14.18%. While it outperforms Random Forest in detecting cancellations, its recall remains low, suggesting it still heavily favors the majority class.

Based on the evaluation metrics, Logistic Regression was selected as the best model for predicting order cancellations. Although it has the lowest accuracy (56.03%), it achieved the highest recall (55.73%) and F1-score (24.93%), making it the most capable of detecting actual cancellations. The Random Forest Classifier was disqualified despite its high accuracy (86.93%), as it failed to identify any cancelled order due to class imbalance. XGBoost also underperformed with a recall of only 11.96%, confirming Logistic Regression as the most suitable choice for this imbalanced classification task.

4 Dashboard overview

The dashboard provides an interactive interface that consolidates the results of the analysis and predictive modeling into a set of visual insights. It is designed to support decision-making by presenting key business metrics, delivery performance, customer insights, revenue discounts insights, and predictive results in a clear and accessible format. Through dynamic visualizations and filters, the dashboard enables users to explore the data from multiple perspectives and understand both historical trends and future revenue projections.

Figure 30: Page-1 Operations Overview.

Based on the results presented on this page, the dataset contains approximately 15,000 orders, with an average delivery time of 94.1 minutes across all deliveries. This relatively high average may be influenced by preparation time and customer delivery distance. The cancellation rate is approximately 13%, indicating that the majority of orders are successfully completed. In addition, the platform generated a total revenue of nearly $1.79M, with an average order value of $113.95.

Customers from City Tier 3 represent nearly 50% of all orders and contribute approximately $900K in revenue. In comparison, City Tiers 1 and 2 each account for roughly 25% of total orders, generating nearly $445K each.

Around 13.5K orders were delivered on time, reflecting strong operational performance and contributing positively to customer satisfaction and restaurant ratings. On the other hand, approximately 1.5K orders experienced delays, which may be associated with factors such as severe weather conditions or high traffic levels.

Tuesday recorded the highest number of orders, while Monday showed no recorded activity. The remaining days were distributed relatively uniformly. In addition, order activity across hours of the day appeared to be fairly balanced without major peaks.

The customer satisfaction correlation analysis shows a moderate positive relationship between delayed deliveries and delivery time minutes (approximately 0.16), which is logically consistent, as longer delivery durations are more likely to result in delayed orders.

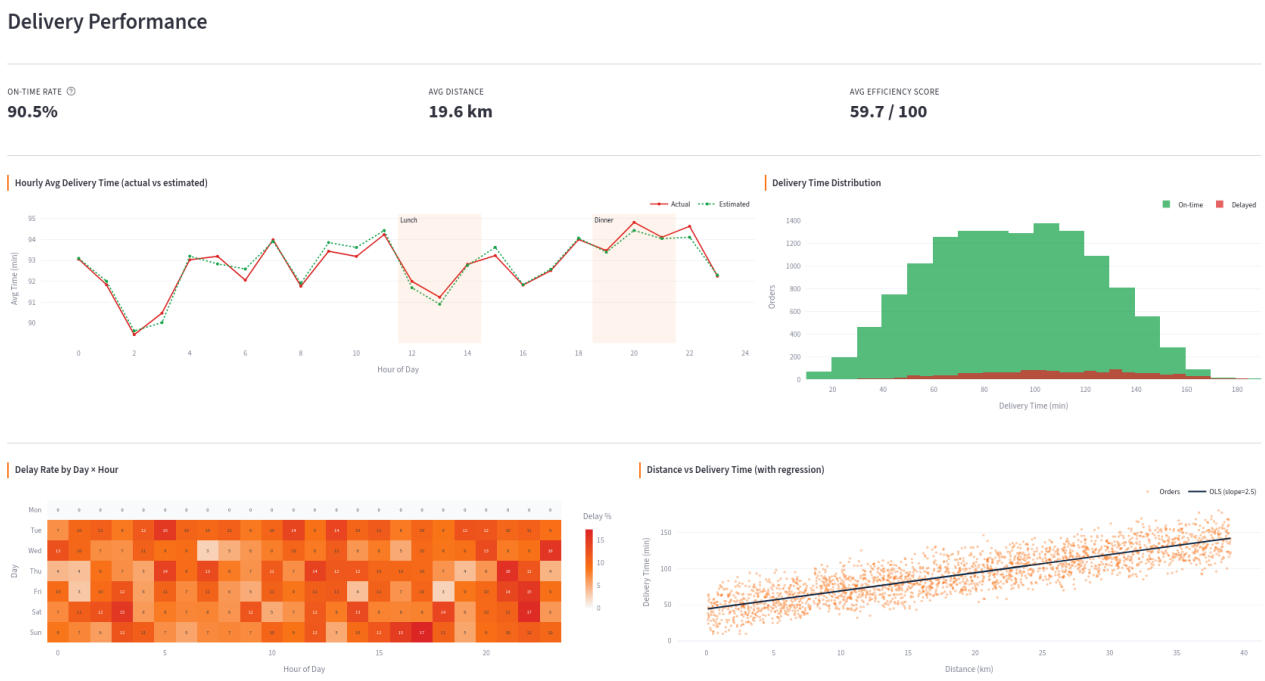

Figure 31: Page-2 Delivery Performance.

Based on the results presented on this page, the on-time delivery rate is approximately 90.5%, which reflects strong operational performance. The average delivery distance is around 19.6 km, indicating that many customers are located relatively far from restaurants or delivery centers. In addition, the average delivery efficiency score is approximately 59%, suggesting a moderate level of delivery performance across orders.

The “Hourly Average Delivery Time (Actual vs Estimated)” plot demonstrates that the estimated delivery times closely follow the actual delivery durations, with only minor deviations. This indicates that the estimation system performs reliably and produces reasonably accurate predictions.

The delivery time distribution follows an approximately normal distribution, showing that nearly 68% of deliveries are completed within the range of 60 to 130 minutes.

The relationship between delivery distance and delivery time appears to be positively linear, which is logically expected, as longer distances generally require more delivery time.

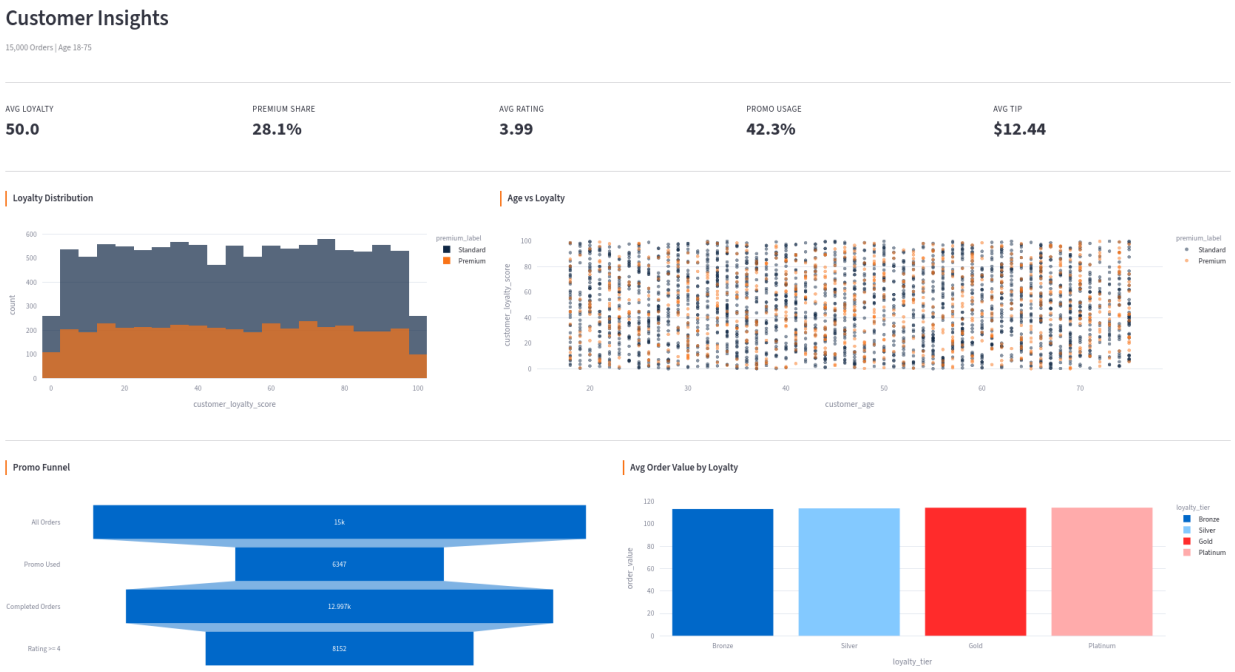

Figure 32: Page-3 Customer Insights.

Based on the results presented on this page, the average loyalty score is 50/100, and About 1 in 4 customers is a Premium member, also Customers are generally satisfied (4/5) as shown on average rating, and Nearly half of orders used a promo code and Moderate tipping behavior across orders.

Loyalty Distribution Shows how loyalty scores (0-100) are spread across customers. Both Standard and Premium customers are fairly evenly distributed, with no strong concentration at high or low scores.

Age vs Loyalty Plots customer age against loyalty score shows that there is no clear pattern — loyalty scores are scattered across all age groups, suggesting age alone does not determine loyalty.

Promo Funnel plot Shows the order journey: 15K total orders: 6,347 used a promo: 12,997 completed: 8,152 rated 4 stars. Shows healthy completion and satisfaction rates.

Avg Order Value by Loyalty Tier: Compares average order value across Bronze, Silver, Gold, and Platinum tiers. All tiers show similar order values ($100), meaning higher loyalty tier doesn't necessarily mean higher spending.

Figure 33: Page-4 Revenue and discount analysis.

The dashboard reveals that the average order value stands at $113.95, with an average discount of $14.93 applied per order and a final average amount paid of $119.08, slightly higher than the gross order value due to added delivery fees and tips.

The Revenue Flow Breakdown waterfall chart illustrates this clearly: starting from a gross revenue of $1.5M, discounts reduce it, while delivery fees and tips ($1M combined) partially recover it, resulting in a comparable final revenue.

The Discount vs Order Value scatter plot shows no strong relationship between discount size and order value, meaning discounts are applied fairly uniformly regardless of how much a customer spends.

Finally, the Discount Distribution histogram indicates that discount amounts are spread evenly between $0 and $30, with no particular discount value being significantly more common, suggesting a randomized or varied promotional strategy across the 42.3% of orders that used a promo code.

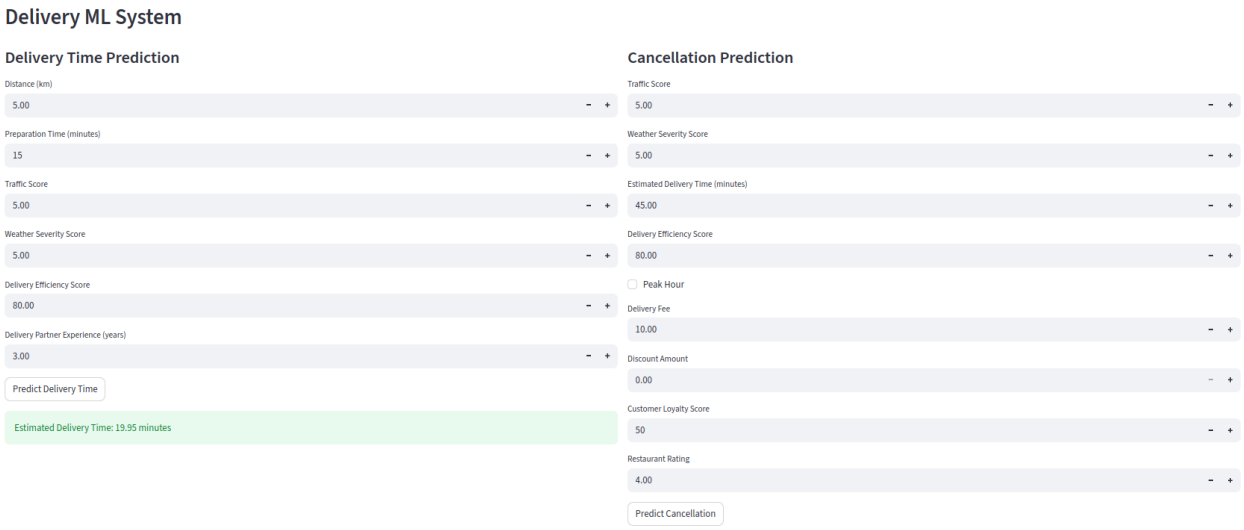

Figure 34: Page-5 Delivery ML system.

This interface presents a machine learning-powered tool with two prediction modules.

The **Delivery Time Prediction** takes inputs such as distance, preparation time, traffic score, weather severity, delivery efficiency, and driver experience to estimate how long a delivery will take — in this example, predicting 19.95 minutes.

The **Cancellation Prediction** uses a different set of features including traffic, weather, estimated delivery time, delivery fee, discount amount, customer loyalty score, and restaurant rating to predict whether an order is likely to be cancelled.

Both models run independently, allowing users to adjust input values using the +/- controls and instantly generate predictions by clicking the respective buttons.

5 Business Recommendations

1. Invest in City Tier 3: City Tier 3 represents an underserved market with significant growth potential. Expanding operations by partnering with local restaurants and recruiting delivery partners could increase order volume, while targeted promotions could accelerate adoption in these regions.

2. Reduce Delivery Time: The ML model estimated an average delivery time of approximately 20 minutes, which can be further optimized by prioritizing experienced delivery partners, improving route planning algorithms, and better scheduling restaurant preparation times to reduce cancellation risk.

3. Offer Delivery Zones Due to Average Distance of 19 km: Given the average delivery distance of approximately 19 km, introducing zone-based delivery hubs and clustering orders by geographic proximity would reduce fuel costs, minimize travel time, and improve overall delivery efficiency.

4. Address Class Imbalance to Improve Cancellation Detection: The imbalance between completed (12,000) and cancelled (3,000) orders negatively impacted model performance. Applying techniques such as SMOTE or class-weight adjustment during model training would improve the system's ability to proactively detect cancellations.

5. Leverage Promo Codes Strategically: With 42.3% of orders using a promo code, discounts should be targeted toward high-risk cancellation orders or low-loyalty customers rather than applied uniformly, in order to maximize impact while protecting revenue margins.

6. Retain and Reward High-Loyalty Customers: The average loyalty score of 50/100 indicates room for improvement. A tiered rewards program offering benefits such as free delivery or personalized discounts to Gold and Platinum customers would incentivize repeat orders and improve long-term retention.

7. Optimize Peak Hour Operations: Orders during peak hours face higher delays and cancellation risks. Deploying additional delivery partners during these periods and offering incentives for off-peak orders would help balance demand and maintain consistent service quality.

6 Conclusion

This project presented an end-to-end data science and machine learning pipeline applied to a food delivery dataset containing over 15,000 records. The work covered the complete analytical workflow, including data cleaning, exploratory data analysis, predictive modeling, and dashboard development.

For the regression task, Linear Regression outperformed both Random Forest and XGBoost, achieving an R^2 of 0.9983, an MAE of 0.42, and an RMSE of 1.41, demonstrating that the relationship between the selected features and delivery time is largely linear. For the classification task, the dataset's inherent class imbalance — approximately 12,000 completed orders versus 3,000 cancelled orders — significantly challenged all three models. Logistic Regression was selected as the most suitable model, achieving the highest recall (55.73%) and F1-score (24.93%), making it the most capable of detecting actual cancellations despite its lower accuracy.

An interactive Streamlit dashboard was developed to consolidate the analytical results into accessible visual insights across five pages, covering operational overview, delivery performance, customer behavior, revenue analysis, and real-time ML predictions. The system also integrates production-ready practices including model serialization, Docker containerization, CI/CD pipelines, and API deployment via AWS EC2.

Key findings reveal that City Tier 3 customers represent over 50% of total orders, Tuesday is the busiest day, and traffic and weather conditions moderately influence delivery time. Future work should focus on addressing the class imbalance through techniques such as SMOTE, retraining XGBoost with class-weight adjustments, and introducing zone-based delivery hubs to improve operational efficiency given the average delivery distance of approximately 19 km.

References

Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32.

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research, 16, 321–357.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.

Streamlit Inc. (2023). *Streamlit: The Fastest Way to Build and Share Data Apps*. Retrieved from `https://streamlit.io`

Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 9(3), 90–95.

Kaggle. (2023). *Food Delivery Operations and Customer Analytics Dataset*. Retrieved from `https://www.kaggle.com/datasets/deepeshkansotia/food-delivery-oper` `data`

Amazon Web Services. (2023). *Amazon EC2 Documentation*. Retrieved from `https://docs.aws.amazon.com/ec2`