**Title: Predicting Food Delivery Time: A Formal Analytical Report**

Contents

[Introduction 1](#_Toc211846360)

[Purpose of the Data Analysis 1](#_Toc211846361)

[Choice and Justification of Programming Environment 2](#_Toc211846362)

[Machine Learning Algorithms: Types and Purposes 2](#_Toc211846363)

[Variables: Identification and Justification 3](#_Toc211846364)

[Sampling Strategy: Decision and Justification 3](#_Toc211846365)

[Rationale for Descriptive Analysis 4](#_Toc211846366)

[Detailed Descriptive Statistics for the Dependent Variable 4](#_Toc211846367)

[Hypothesis Testing: Correlations and Association Tests 8](#_Toc211846368)

[Representativeness Assessment: One-Sample T-test 9](#_Toc211846369)

[Regression Modelling: Simple and Multiple Linear Regression 9](#_Toc211846370)

[Classification Modelling and Comparative Evaluation 10](#_Toc211846371)

[Using the Best-Fit Classifier for Prediction 11](#_Toc211846372)

[Cluster Analysis and Interpretive Strategy 11](#_Toc211846373)

[Strategy for System Improvement Based on Cluster Insights 12](#_Toc211846374)

[Conclusion 13](#_Toc211846375)

## Introduction

Food delivery firms, logistics managers, and couriers struggle to anticipate delivery timeframes. Customer satisfaction depends on systems' ability to predict delivery windows, optimize routes, manage resources, and establish prices. This study examined delivery time determinants such distance, preparation time, traffic, weather, courier experience, vehicle type, and time of day. Characterize the distribution and central tendencies of delivery time, identify and quantify relationships between delivery time and its candidate predictors, develop, evaluate, and compare predictive models for continuous (regression) and categorical (classification) delivery time, and cluster orders to reveal operational groupings that could inform targeted interventions. According to academic standards, the methodology section justifies the programming environment and sampling strategies, the results section integrates descriptive statistics, visualizations, hypothesis tests, regression and classification models, and clustering, and the discussion contextualizes these findings and draws managerial conclusions.

## Purpose of the Data Analysis

This analysis aims to develop an evidence-based, operationally practical understanding of how order-level factors affect delivery time and predictive tools for anticipating delivery length at order placement. Customers get more accurate ETAs, couriers can plan routes and accept orders better, and platform operators can improve matching, incentives, surge pricing, and staffing. Beyond prediction, the research seeks to discover actionable levers like reducing preparation time, changing fleet deployment times, or changing routing techniques for longer distances to lower mean delivery time or variability. The analysis also clusters instructions to reveal latent groups (e.g., short-distance fast-prep vs. long-distance slow-prep) to support differential operational approaches.

## Choice and Justification of Programming Environment

Python was used for this analysis. Python is frequently used in data science and machine learning because to its developed ecosystem of libraries (Numerical computation with NumPy, data manipulation with pandas, statistical tests and models with SciPy and statsmodels, and machine learning methods with scikit-learn). Data cleansing, exploratory analysis, statistical testing, model training, and visualization may be done in one place. Python's scriptable workflows and notebooks enable rapid prototyping and replication and integrate well with production pipelines. Python has the right combination of numerical accuracy, model evaluation APIs, and visualization capabilities (matplotlib, seaborn) for the moderate-sized dataset and statistical inference and machine learning pipeline analysis.

## Machine Learning Algorithms: Types and Purposes

This project implements two distinct families of predictive tasks, each with an appropriate algorithmic approach. For the continuous target variable, delivery time in minutes, regression algorithms are appropriate. Simple linear regression is used to evaluate the marginal explanatory power of a single predictor (in this case distance), while multiple linear regression is used to quantify the combined linear effects of several continuous and categorical predictors. The purpose of regression modelling here is twofold: to explain variance in delivery time (in-sample explanatory power) and to produce point predictions for future orders.

For categorical prediction tasks where delivery times are binned into discrete classes such as fast, normal, and slow classic supervised classification algorithms are used. Logistic regression serves as a baseline probabilistic classifier with interpretability; k-nearest neighbours (KNN) provides a non-parametric local decision rule that can capture nonlinear boundaries; Gaussian Naïve Bayes offers a simple generative approach assuming conditional independence and Gaussian feature distributions; and decision trees provide an intuitive, rule-based classifier that can capture interactions and nonlinearities. The purpose of comparing these classifiers is to identify the algorithm that yields the best predictive accuracy and reliability when delivery time is discretized, enabling classification use cases such as rapid triage of orders into expected service-level bins.

Clustering methods are also applied for segmentation. K-means clustering is used to identify spherical, variance-based clusters in continuous feature space, and hierarchical (agglomerative) clustering referred to here as horizontal clustering in the project brief is employed to reveal nested groupings and dendrogram structure. The purpose of clustering is not predictive per se but descriptive and prescriptive: to find natural groups of orders that can be targeted with different operational strategies.

## Variables: Identification and Justification

The dataset contains several variables. The dependent variable, or outcome of interest, is Delivery\_Time\_min, recorded in minutes. This variable is directly operationally relevant and serves as the target for regression and classification. The independent variables are chosen because they are theory-driven or intuitively linked to delivery duration. Distance\_km captures the physical separation between restaurant and customer, and it is a primary driver of travel time. Preparation\_Time\_min captures time that must elapse before the courier can depart, directly inflating total delivery time. Weather and Traffic\_Level are contextual risk factors that affect travel speed; time-of-day can proxy for diurnal patterns in demand and congestion; Vehicle\_Type captures differences in average speeds, maneuverability, and access limitations; Courier\_Experience\_yrs may influence route choice and driving efficiency, reflecting learning effects.

## Sampling Strategy: Decision and Justification

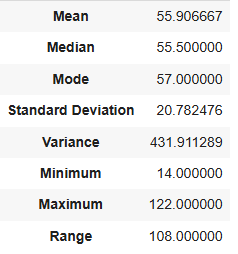
Sampling decisions must weigh the goals of representativeness and computational tractability. This study uses two sampling methods as part of the analysis: random sampling and systematic sampling. Random sampling provides an unbiased subset of observations and is appropriate when the dataset does not show time-ordered heterogeneity or strong temporal drift that would bias random draws. Systematic sampling selecting every k-th observation after a random starting point was also used with the same target sample size. Systematic sampling is particularly useful when the dataset is ordered but lacks periodic patterns aligned with the sampling interval; it is easy to implement and provides coverage across the entire dataset span. The dual use of random and systematic sampling is defensible because it offers a robustness check: similar descriptive statistics across the two sampling methods indicate that the dataset does not suffer from a severe ordering bias or sampling artifact and that samples of moderate size (n = 150) capture the central tendency and dispersion of the full dataset satisfactorily.

## Rationale for Descriptive Analysis

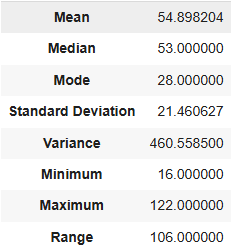
In this case, descriptive statistics for delivery time revealed a mean around the mid-to-high fifties of minutes, moderate dispersion (standard deviation in the low twenties), and a range spanning approximately 14 to 122 minutes across samples. Visualizations (scatter plot, boxplot, histogram, and heatmap of correlations) further elucidate shape (slight positive skew), spread (majority between 20–80 minutes with tails beyond), and relationships (positive correlations of distance and preparation time with delivery time). These descriptive insights justify modelling choices: for example, the positive skew suggests that regression residuals should be examined and potentially transformed if non-normality of errors impairs inference; the strong correlation between distance and delivery time motivates inclusion of distance as a key predictor (and possibly interaction terms with traffic or vehicle type). Descriptive analysis also informs feature engineering for classification: defining cut-points to create classes for delivery times, understanding which ranges constitute a meaningful “slow” or “on-time” delivery, and assessing class balance for classifier training. Overall, descriptive work supports both statistical inference and predictive modelling, and it anchors interpretation of model coefficients in the data’s empirical realities.

## Detailed Descriptive Statistics for the Dependent Variable

Delivery\_Time\_min has various descriptive traits. The random sample of 150 observations has a mean of 55.9067 minutes, a median of 55.5 minutes, and a mode of 57 minutes. Moderate dispersion is indicated by the 20.7825-minute standard deviation and 431.9113 variance. The sample range is 108 minutes, with lowest and maximum of 14 and 122 minutes.



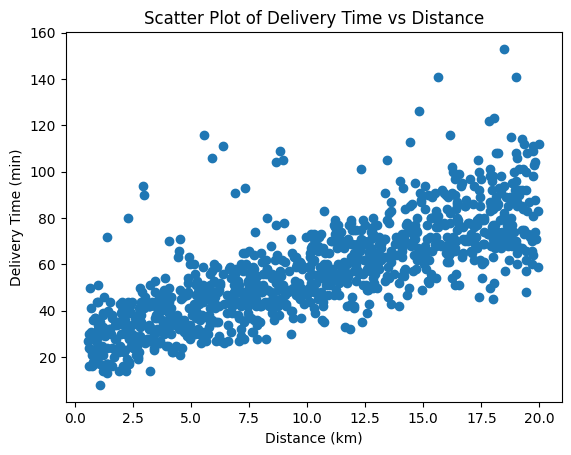
A mean of 54.8982 minutes, a median of 53 minutes, and a mode of 28 minutes are found in the systematic sample, with a somewhat greater standard deviation of 21.4606 minutes and variance of 460.5585. Minimum and maximum values of 16 and 122 minutes produce a range of 106 Means and medians are similar across sampling procedures, ensuring central tendency estimates' stability. The systematic sample's mode and slightly higher variance may be attributable to sample-specific inclusion or removal of severe outliers.



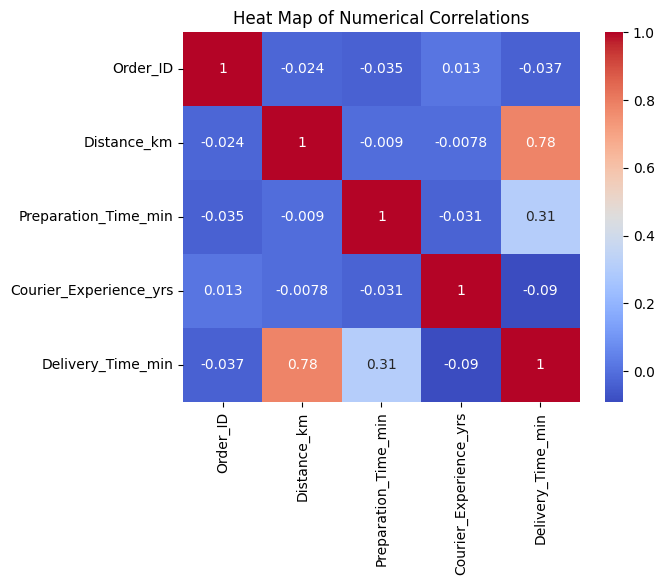
Summary statistics are enhanced by visual inspection. The histogram has a moderate right skew, with most observations between 20 and 80 minutes and fewer at higher delivery times. The boxplot confirms outliers, mostly at the top.

|  |  |
| --- | --- |
| C:\Users\Muhammad Husnain\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\516DC99F.tmp | C:\Users\Muhammad Husnain\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\99D0A405.tmp |

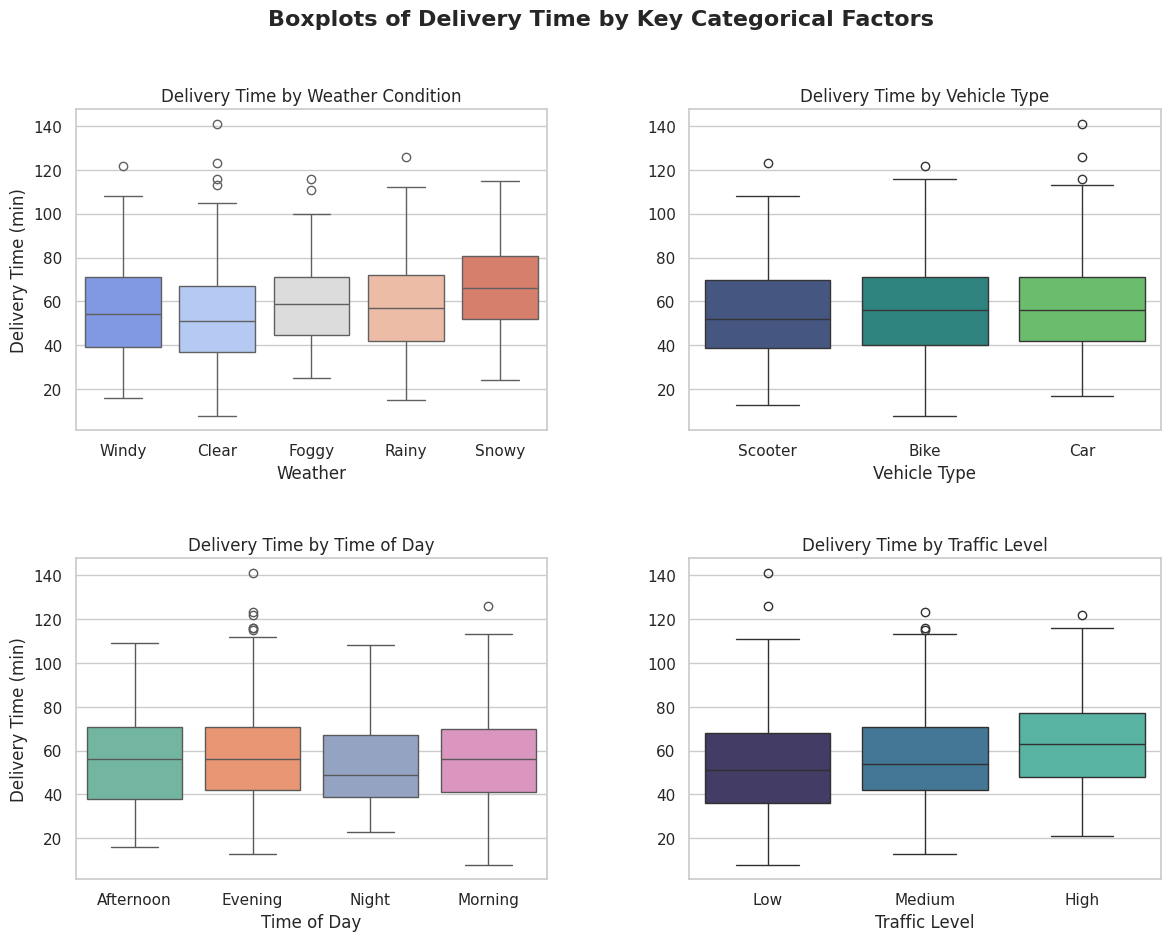
The scatter plot of delivery time against distance exhibits a clear positive trend, reflecting the intuitive relationship that greater distances typically increase delivery durations.



The heatmap of correlations demonstrates that distance and preparation time positively correlate with delivery time, whereas courier experience has no linear correlation.

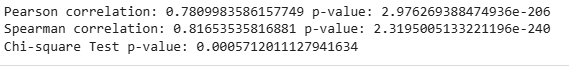


Further anlaysis shows that the delivery time is highest in snowy weather and lowest in clear weather. There is no significant difference in average delivery time by vehicle type but scooter records slightly lower delivery time compared to bike and car. The delivery time is lowest in night and no significant difference observed between delivery time of morning, evening and afternoon. Finally, higher traffic leads to higher delivery time and lower traffic leads to lower delivery time.



## Hypothesis Testing: Correlations and Association Tests

Parametric and non-parametric correlation tests assessed predictor-delivery time connections. Distance and delivery time had a strong positive linear relationship, rejecting the null hypothesis of no linear association at conventional significance levels, according to the Pearson correlation test. Strong monotonic connection resilient to outliers and nonlinearity is confirmed by the Spearman rank correlation coefficient of 0.817, likewise highly significant. A chi-square test was performed to assess association between categorical variables and delivery time categories. The p-value is 0.0005712, which is below conventional significance thresholds and indicates a statistically significant association.



## Representativeness Assessment: One-Sample T-test

A one-sample t-test was performed to assess whether the sample mean of delivery time differs significantly from a hypothesized population mean. The computed t-statistic was -0.4848 with an associated p-value of 0.6286, leading to a failure to reject the null hypothesis that the sample mean equals the hypothesized population mean. Practically, this suggests that the sampled observations are consistent with the prespecified benchmark mean and that there is no statistical evidence of systematic deviation in average delivery time for the sample relative to that population parameter. This result supports the representativeness of the sample with respect to its central tendency.



## Regression Modelling: Simple and Multiple Linear Regression

Regression analysis quantified relationships and predicted delivery time continuously. A simple linear regression model with distance as the single predictor yielded a R² score of 0.6293 and an MSE of 151.5151. Distance alone predicts 62.9% of delivery time variance, demonstrating its dominance. The residual mean squared error shows a non-trivial absolute prediction error in squared terms, but with a dependent variable of mean 55 minutes and standard deviation 21 minutes, these mistakes are operationally important and expected in a single predictor model.



Adding relevant factors to the multiple linear regression resulted in a higher R² of 0.8325 and a lower MSE of 68.4478. This shows that the multivariate model explains 83.3% of delivery time variance, an improvement above the single-variable model. MSE reduction indicates better prediction. The empirical results and heatmap suggest that distance and preparation time are the main contributors to this improvement. Categorical variables like traffic level and time of day likely added explanatory power by accounting for systematic variations in travel speed and waiting time. Courier experience, on the other hand, displayed negligible correlation and therefore likely yielded a small or insignificant coefficient in the multiple regression.



The practical implication of the regression results is that when multiple, complementary predictors are used, delivery time can be modelled with reasonable accuracy. For operational deployment, the multiple linear regression model provides a transparent, interpretable mapping from features to predicted minutes and could be embedded in real-time ETA calculators with retraining as more data accumulate.

## Classification Modelling and Comparative Evaluation

When delivery time is discretized into classes (for example short, medium, long), classification algorithms were trained and evaluated. Four classifiers were compared: logistic regression, k-nearest neighbours (KNN), Naïve Bayes, and decision tree. The confusion matrices and accuracies reported show that KNN achieved the highest accuracy at approximately 78.53% (accuracy = 0.7853), closely followed by logistic regression with accuracy about 77.97% (0.7797). Naïve Bayes achieved approximately 74.01% accuracy, and the decision tree yielded approximately 66.67% accuracy.

Beyond accuracy, confusion matrices show more nuanced performance disparities. Logistic regression and KNN performed well on the first class (probably “short” deliveries) with higher true-positive numbers. KNN surpassed logistic regression in some cells, improving its accuracy. Naïve Bayes underperformed in comparison possibly due to violated conditional independence assumptions or non-Gaussian feature distributions, while the decision tree may have overfit if grown fully or underfit if pruned excessively; its lower accuracy suggests instability relative to the other methods.

Given these results, KNN is identified as the best-fit classifier by the metric of accuracy on the available evaluation set. Logistic regression offers interpretability and calibrated probabilities, while KNN is simple but can be computationally heavy at prediction time for large datasets. Therefore, although KNN has the highest raw accuracy here, logistic regression may be preferred when interpretability and probabilistic outputs are prioritized.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | TP | TN | FP | FN | Accuracy |
| Logistic Regression | 46 | 113 | 22 | 21 | **0.7797** |
| K-Nearest Neighbors (KNN) | 46 | 115 | 20 | 21 | **0.7853** |
| Naïve Bayes | 44 | 109 | 26 | 23 | **0.7401** |
| Decision Tree | 39 | 100 | 35 | 28 | **0.6667** |

## Using the Best-Fit Classifier for Prediction

The KNN classifier having the highest observed accuracy was used to predict delivery time classes for new orders. Operationally, KNN’s simplicity permits it to be used as a quick benchmarking algorithm, but for scalability, approximate nearest neighbours or indexing structures should be considered if the full dataset grows large. Calibration of K and standardization of features prior to distance computations are essential for stable KNN performance, and weighting schemes (distance-weighted votes) can improve predictions when class boundaries are uneven.

## Cluster Analysis and Interpretive Strategy

Both K-Means and Hierarchical clustering algorithms are used to find natural categories in the data. The K-Means clustering scatter plot, which used distance and delivery time, showed three separate clusters. The first cluster was for short-distance, fast delivery; the second was for medium-distance deliveries with intermediate timeframes; and the third was for long-distance deliveries that took a long time. The hierarchical clustering dendrogram validated these three main groups by showing that the linkages were clearly separate in distance. The results show that delivery data may be split up into useful operational groups that can help with future optimization efforts.

|  |  |
| --- | --- |
| C:\Users\Muhammad Husnain\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\DD253721.tmp | C:\Users\Muhammad Husnain\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\63B2A6D7.tmp |

## Strategy for System Improvement Based on Cluster Insights

A cluster-driven operational strategy starts with targeted interventions for clusters associated with the highest mean delivery times and highest variance. For clusters characterized by long distances, strategies include optimizing acceptance radius, deploying more cars during times when long-distance orders concentrate, or pooling nearby orders for batch deliveries when feasible. For clusters where preparation time is a primary contributor, the platform could institute restaurant-focused performance metrics, offer training or incentives to restaurants for on-time preparation, or integrate preparation time predictions into courier dispatch so couriers pick up other nearby orders while waiting.

Traffic-sensitive clusters suggest the need for real-time routing adjustments and dynamic estimated times that incorporate live traffic feeds. Weather-driven clusters, especially in severe weather like rain or snow, should encourage safety-oriented measures like longer guaranteed times, higher courier pay, or temporary vehicle type bans. Clusters with negligible courier experience effects show that hiring or retention tactics based only on years of experience may not reduce average delivery time. Training should stress route familiarity and performance-measurable best practices.

These tactics must be implemented iteratively using pilots to test interventions in a subset of clusters, assessment of delivery time mean and variance, and cost-benefit analysis to balance customer satisfaction with courier remuneration. To keep forecasts accurate as market conditions and behavioral patterns change, predictive models must be monitored for data drift and retrained.

## Conclusion

This paper summarizes food delivery data analytical pipeline findings. According to descriptive analysis, delivery times average 50–56 minutes with a moderate dispersion and a right tail of lengthier deliveries. Distance and preparation time predicted delivery time the most, with strong Pearson and Spearman correlations. The multivariate approach improves explanatory and predictive power, as a multiple linear regression model achieved a R² of 0.833, outperforming a simple regression on distance alone. Decision trees underperformed in classification studies, while KNN and logistic regression had the greatest accuracy at 78.5%. Clustering analyses identified actionable segments for differential operational tactics like routing, vehicle allocation, and restaurant interventions.

The findings provide strong, actionable evidence that delivery times can be predicted with reasonable accuracy using available features and that operational improvements targeting preparation times and long-distance deliveries could significantly reduce average delivery duration and improve service reliability. For production deployment, employ multiple linear regression for continuous ETA estimations and KNN or logistic regression for rapid service-level classification. Maintaining model performance and realizing the practical benefits of the analysis requires continuous monitoring and iterative refinement, especially with real-time traffic and restaurant-specific factors.