

International Journal Research Publication Analysis

Page: 01-16

A DATA-DRIVEN APPROACH TO ANALYZING AND PREDICTING LAST-MILE DELIVERY FAILURES

***Maddala Thanmayee, Dr. Ramchander (Guide)**

Department of MCA (Master of Computer Applications) Chaitanya Bharathi Institute of
Technology (CBIT), Gandipet, Hyderabad, Telangana, India.

Article Received: 05 September 2026

Article Revised: 25 September 2026

Published on: 15 October 2026

***Corresponding Author: Maddala Thanmayee**

Department of MCA (Master of Computer Applications) Chaitanya Bharathi
Institute of Technology (CBIT), Gandipet, Hyderabad, Telangana, India.

DOI: <https://doi-org/101555/ijrpa.8349>

ABSTRACT

Last-mile delivery is a critical stage of logistics in which delays, unsuccessful delivery attempts and customer-related issues can affect service quality and operational cost. This paper presents a data-driven workflow for analyzing delivery failures using structured dataset and interactive analytics. Python and pandas are used for data preparation and analysis, SQL supports structured querying, and a Streamlit dashboard presents key performance indicators and failure patterns. The project dataset contains 5,000 delivery records, with attributes such as distance, vehicle type, weather, traffic, customer availability, address accuracy, delivery attempts and failure reason. The recorded overall delivery failure rate is 19.24%. Chi-square tests show that customer availability and address accuracy have the strongest association with failure, followed by traffic level and weather. The workflow provides descriptive operational insights and can be extended with real-time data and validated machine-learning models for failure prediction.

INDEX TERMS: Last-mile delivery, delivery failure analysis, data analytics, Python, pandas, SQL, Streamlit, data visualization, chi-square analysis, predictive analytics.

1. INTRODUCTION

Last-mile delivery is the final movement of a shipment from a distribution point to the customer. Problems at this stage may lead to unsuccessful attempts, additional transportation costs, customer dissatisfaction and repeated delivery operations. Delivery outcomes can be affected by distance, traffic, weather, vehicle type, customer availability and address accuracy. Manual spreadsheet-based analysis can make it difficult to identify patterns across

large sets of delivery records. This project applies a repeatable workflow using Python, SQL and a Streamlit dashboard to convert delivery records into interpretable information.

Because every failed delivery typically requires at least one additional attempt, understanding where and why failures occur is a practical starting point for improving service quality. The dataset analyzed in this work covers 5,000 delivery orders placed during 2026 across six delivery locations in and around Hyderabad, and records the operating conditions and the final outcome of each order. The goal is not only to report how often deliveries fail, but also to show which recorded conditions are associated with higher failure rates and whether those associations are statistically distinguishable from chance.

The main contributions of this paper are:

- (i) a reproducible pipeline from a raw delivery CSV file to SQL summaries and an interactive dashboard;
- (ii) a quantitative description of failure reasons and of failure rates across operational attributes;
- (iii) chi-square tests and Cramér's V effect sizes that rank the attributes by strength of association with failure; and
- (iv) a discussion of how the findings can guide operational action and a future, separately validated prediction model.

The remainder of the paper is organized as follows. Section 2 reviews related work, Sections 3 and 4 state the research gap and objectives, Section 5 describes the dataset, Sections 6 to 8 present the methodology, architecture and implementation, Sections 9 and 10 report the results and dashboard, and Sections 11 to 14 provide the discussion, limitations, future work and conclusion.

2. RELATED WORK

Last-mile logistics research and industry practice commonly examine route planning, delivery time, vehicle utilization, customer availability and failed-delivery reduction. Historical delivery records can reveal recurring patterns, while visualization systems help users monitor performance indicators. An end-to-end workflow combining data cleaning, structured queries and interactive visualization can provide a consistent view of delivery outcomes.

Work on route planning and vehicle utilization generally treats the delivery plan as an optimization problem, where the objective is to reduce distance, time or cost. Failed

deliveries are a separate concern: an order that reaches the address but cannot be handed over adds cost without completing the service. Commonly discussed causes include the recipient not being present, incomplete or inaccurate address information, congestion and adverse weather, which correspond to the attributes examined in this study.

Descriptive analytics and dashboards are widely used to monitor such indicators because they are transparent and easy for operations teams to interpret. Predictive models can complement them, but their usefulness depends on careful validation and on using only information that is available at the time a decision must be made. This paper therefore concentrates on the descriptive and statistical foundation on which such a model could later be built.

3. RESEARCH GAP AND PROBLEM STATEMENT

Delivery reporting may focus on isolated metrics without documenting a reproducible path from raw data to interactive analysis. This project addresses the need for a compact workflow that prepares delivery records, analyzes failure-related attributes, summarizes results using SQL and presents findings through a dashboard. The problem is to identify delivery failure patterns and calculate relevant indicators from structured delivery data, while establishing a foundation for future validated prediction.

Three questions guide the analysis. RQ1: what share of deliveries fail, and for which recorded reasons? RQ2: which operational attributes are associated with higher failure rates? RQ3: which of these associations are statistically significant, and how strong are they?

4. OBJECTIVES

The objectives are to clean and validate delivery records; calculate delivery failure indicators; explore patterns associated with distance, vehicle type, weather, traffic, customer availability, address accuracy and delivery attempts; support structured SQL analysis; develop an interactive Streamlit dashboard; and provide a foundation for future predictive modeling.

5. DATASET AND PREPROCESSING

A. Data Description

The project dataset contains 5,000 records. The original dataset contains 21 attributes and the processed analytical data contains 17 attributes. Representative fields include delivery identifier, location, distance, vehicle type, weather, traffic level, customer availability, address accuracy, delivery attempts and failure reason. The orders are dated from 1 January to 31 December 2026 and are served by 20 delivery partners in six locations (Hyderabad,

Gachibowli, Madhapur, Kukatpally, LB Nagar and Secunderabad). Table I summarizes the project dataset and Table II lists the key attributes used in the analysis.

Table I. Project Dataset Summary.

Parameter	Value
Total records	5,000
Original attributes	21
Processed attributes	17
Observed failure rate	19.24%
Tools	Python, pandas, SQL, Streamlit

Table II. Key Attributes Used in the Analysis.

Attribute	Type	Values / description
Delivery_ID	ID	Unique order identifier
Order_Date	Date	1 Jan – 31 Dec 2026
City	Category	Six delivery locations
Vehicle_Type	Category	Bike, Scooter, Car, Van
Distance_KM	Numeric	Delivery distance, 1.0–25.0 km
Weather	Category	Clear, Cloudy, Rainy, Stormy
Traffic_Level	Category	Low, Medium, High
Order_Value	Numeric	Value of the order
Expected_Time_Min	Numeric	Planned delivery time (min)
Actual_Time_Min	Numeric	Recorded delivery time (min)
Customer_Available	Binary	Yes / No
Address_Accuracy	Binary	Correct / Incorrect
Delivery_Attempts	Numeric	1, 2 or 3 attempts
Delivery_Partner	Category	20 delivery partners
Payment_Method	Category	UPI, COD, Card
Failure_Reason	Category	Six reasons (failed orders only)
Failure_Flag	Binary	1 = failed, 0 = delivered

B. Data Quality and Preprocessing

Preprocessing checks data types, handles missing or invalid values as needed, removes duplicates, standardizes categories and prepares fields for analysis. In the supplied CSV file, no field other than the failure reason has a missing value, and every delivery identifier is unique. The failure reason is empty only for the 4,038 successfully delivered orders, and the binary failure flag agrees with the delivery status for all 5,000 records. The observed failure

rate is 19.24%, which describes this project dataset only and should not be generalized to the entire logistics industry.

C. Descriptive Statistics

Table III gives descriptive statistics of the numeric attributes. The average delivery distance is about 13 km, the average recorded delivery time exceeds the planned time by roughly 22 minutes, and most orders (3,696 of 5,000) were completed in a single attempt.

Table III. Descriptive Statistics of Numeric Attributes.

Attribute	Mean	Std. dev.	Min	Max
Distance (km)	12.97	6.91	1.00	24.99
Order value	1535.71	830.29	100.32	2999.12
Expected time (min)	45.9	13.8	22.0	70.0
Actual time (min)	67.9	21.2	18.0	118.0
Delivery attempts	1.33	0.61	1.00	3.00

6. PROPOSED METHODOLOGY

The workflow includes data input, cleaning, transformation, database storage, SQL analysis, Python-based aggregation and interactive visualization. The CSV dataset is processed with Python and pandas. Cleaned records are made available for structured SQL queries. Delivery outcomes are grouped by location and relevant operational variables. Failure rate is calculated as failed deliveries divided by total deliveries, multiplied by 100.

A. Failure Metrics

For a group of deliveries with F failed orders out of N orders, the failure rate is

$$FR = (F / N) \times 100 \quad (1)$$

To compare a category g (for example, high traffic) with the dataset as a whole, the relative failure ratio is

$$RFR_g = FR_g / FR_{all} \quad (2)$$

where a value above 1 indicates a failure rate higher than the overall rate of 19.24%. The recorded failure reasons are ranked by their share of all failures, that is, the number of failed orders with a given reason divided by the total number of failed orders.

B. SQL Analysis

Assuming the cleaned records are loaded into a table named `deliveries`, the headline indicator and the grouped summaries are obtained with queries such as the following. The results agree with the pandas aggregation (for example, 27.20% for high traffic).

```
-- overall failure rate
SELECT COUNT(*) AS total,
       SUM(Failure_Flag) AS failed,
       ROUND(100.0*AVG(Failure_Flag),2) AS fr
FROM deliveries;

-- failure rate by traffic level
SELECT Traffic_Level, COUNT(*) AS n,
       ROUND(100.0*AVG(Failure_Flag),2) AS fr
FROM deliveries
GROUP BY Traffic_Level ORDER BY fr DESC;

-- combined effect of two attributes
SELECT Customer_Availability, Address_Accuracy,
       COUNT(*) AS n,
       ROUND(100.0*AVG(Failure_Flag),2) AS fr
FROM deliveries
GROUP BY 1, 2 ORDER BY fr DESC;
```

C. Statistical Association Testing

To judge whether an observed difference in failure rate could be due to chance, a chi-square test of independence is applied between each categorical attribute and the failure flag (numeric distance is grouped into five 5-km bands). The strength of association is expressed with Cramér's V ,

$$V = \sqrt{\chi^2 / (N(k-1))} \quad (3)$$

where k is the smaller of the number of rows and columns of the contingency table. Following common practice, V values around 0.1, 0.3 and 0.5 are read as small, medium and large effects. A significance level of 0.05 is used, and because twelve attributes are tested, the more conservative Bonferroni threshold of $0.05/12 \approx 0.004$ is also reported.

7. SYSTEM ARCHITECTURE

The architecture follows this pipeline: Delivery Dataset → Python/Pandas Preprocessing → Cleaned Dataset or Database → SQL Analysis → Analytical Metrics → Streamlit Dashboard → Interactive Visualizations and Insights. The stages separate data preparation, analysis and presentation.

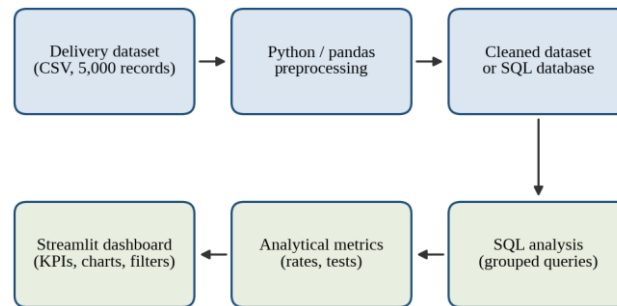


Fig. 1. System architecture of the delivery failure analytics workflow.

As illustrated in Fig. 1, the data layer holds the raw CSV file and the cleaned dataset or database. The processing layer applies the cleaning and transformation steps. The analytics layer computes failure rates, grouped summaries and statistical tests using pandas and SQL. The presentation layer exposes the resulting indicators and charts through the Streamlit dashboard. Keeping these layers separate means that updated delivery records can be re-processed without changing the dashboard code.

8. IMPLEMENTATION

Python is used for data preparation and dashboard logic. pandas [1] supports tabular manipulation and aggregation, SQL organizes and queries delivery information, and Streamlit provides the web-based interface. Dashboard filters and visualizations allow users to examine delivery outcomes by available attributes.

The data preparation module reads the CSV file, validates data types, standardizes category labels and derives helper fields such as the failure flag and the delay between actual and expected delivery time. Grouped aggregations are computed with pandas group-by operations and cross-checked with equivalent SQL queries on a SQLite database [5]. The dashboard reads the cleaned data, applies the filters chosen by the user and recalculates the indicators and charts. The statistical tests and static figures reported in this paper were produced with SciPy [7] and Matplotlib [6] from the same CSV file.

9. RESULTS AND ANALYSIS

The project dataset yielded an overall delivery failure rate of 19.24%. Failure patterns can be explored across traffic, weather, customer availability, address accuracy, distance, vehicle type and delivery attempts. Grouped analysis can help identify recurring patterns, but observed associations are descriptive and do not establish causation.

The dataset contains 962 failed deliveries out of 5,000 records. Figs. 2–11 and Tables IV–V summarize the failure patterns, computed directly from the project CSV file; in the rate charts, the dashed line marks the overall failure rate of 19.24% and the darkest bar marks the highest category.

A. Failure Reasons

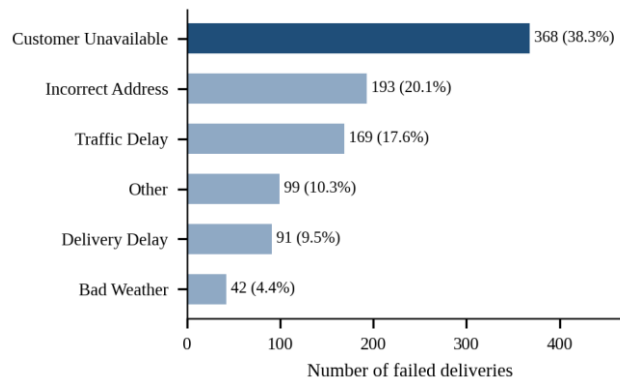


Fig. 2. Distribution of recorded failure reasons among the 962 failed deliveries.

As shown in Fig. 2, customer unavailability is the most frequent recorded reason (368 cases, 38.3% of failures), followed by incorrect address (193, 20.1%) and traffic delay (169, 17.6%). Other reasons, delivery delay and bad weather account for 99, 91 and 42 failures respectively. The reasons are consistent with the recorded conditions: all 169 traffic-delay failures occurred under high traffic, all 42 bad-weather failures occurred in stormy weather, and all 368 customer-unavailable failures were recorded with the customer marked as unavailable.

B. Customer Availability and Address Accuracy

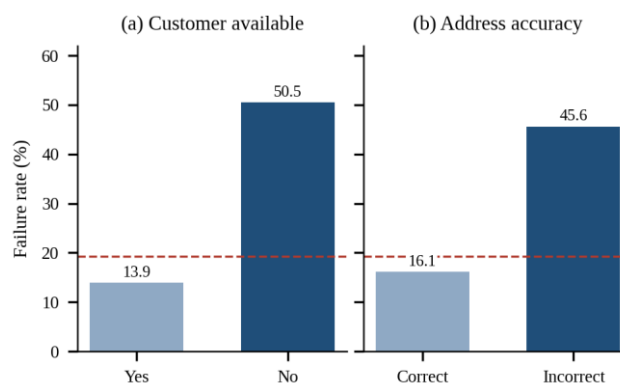


Fig. 3. Failure rate by (a) customer availability and (b) address accuracy. Dashed line: overall failure rate (19.24%).

Customer availability and address accuracy show the largest differences (Fig. 3). Deliveries where the customer was not available failed at 50.5% compared with 13.9% when the customer was available, and deliveries with an incorrect address failed at 45.6% compared with 16.1% for correct addresses.

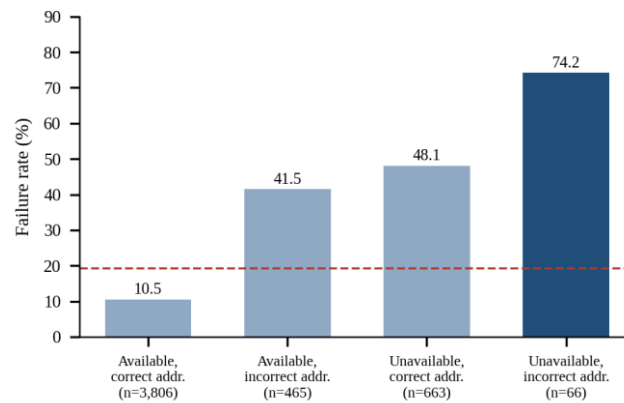


Fig. 4. Failure rate for combinations of customer availability and address accuracy. Dashed line: overall failure rate (19.24%).

The two factors also reinforce each other (Fig. 4). When the customer was available and the address was correct, only 10.5% of 3,806 deliveries failed. The rate rises to 41.5% with an incorrect address alone and 48.1% with an unavailable customer alone, and reaches 74.2% when both problems occur, although this last group is small (66 deliveries).

C. Traffic and Weather

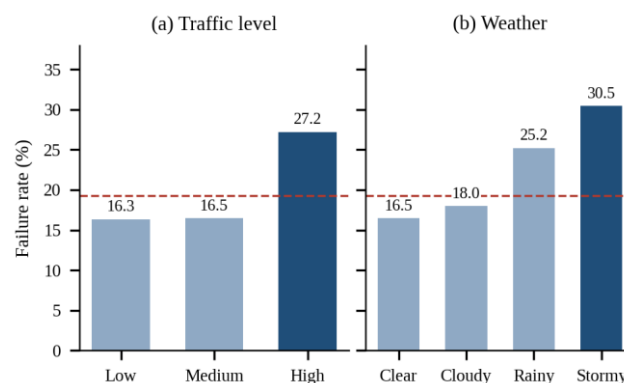


Fig. 5. Failure rate by (a) traffic level and (b) weather condition. Dashed line: overall failure rate (19.24%).

Failure rates rise under adverse external conditions (Fig. 5). High traffic records a failure rate of 27.2%, against 16.3% for low and 16.5% for medium traffic. By weather, the rate

increases from 16.5% for clear conditions to 18.0% for cloudy, 25.2% for rainy and 30.5% for stormy conditions.

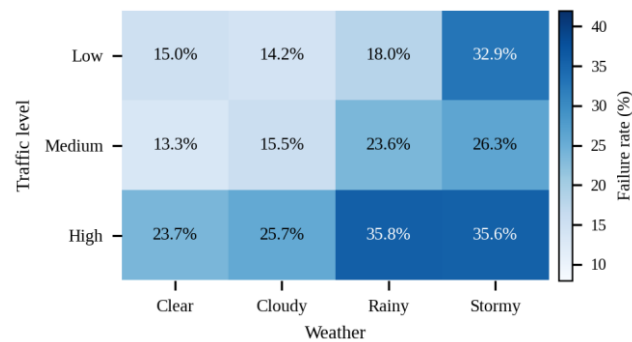


Fig. 6. Failure rate (%) by traffic level and weather condition.

The combined view in Fig. 6 shows that the highest failure rates occur when high traffic coincides with rain (35.8%) or storms (35.6%), whereas medium traffic with clear weather has the lowest rate (13.3%). Stormy weather is also associated with a high rate under low traffic (32.9%), although the storm cells contain only 59 to 114 deliveries each and should be read with care.

D. Vehicle, Attempts, Distance, Location and Time

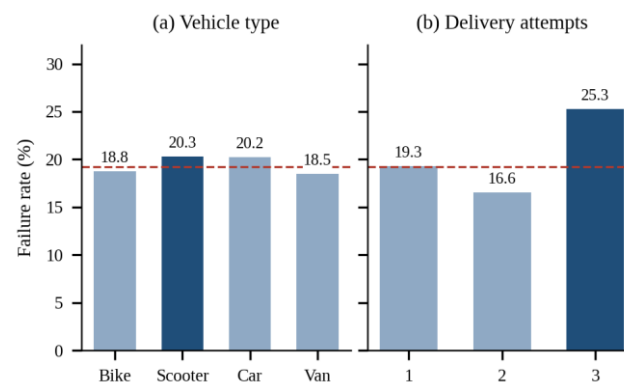


Fig. 7. Failure rate by (a) vehicle type and (b) number of delivery attempts. Dashed line: overall failure rate (19.24%).

Vehicle type shows only small differences, with rates between 18.5% (van) and 20.3% (scooter) as seen in Fig. 7(a). For delivery attempts, Fig. 7(b) shows 19.3%, 16.6% and 25.3% for one, two and three attempts respectively; because attempt count is itself recorded as part of the delivery outcome, this pattern should be interpreted cautiously.

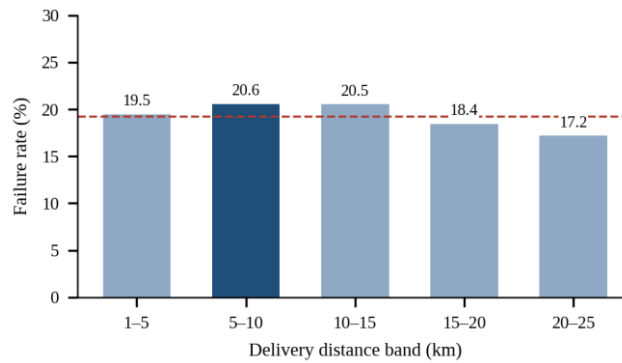


Fig. 8. Failure rate by delivery distance band. Dashed line: overall failure rate (19.24%).

Distance does not show a clear monotonic relationship with failure (Fig. 8). Rates range from 17.2% for the 20–25 km band to 20.6% for the 5–10 km band, all close to the overall rate.

Table IV lists the failure rates by delivery location, payment method and day of week. Location-level rates are relatively uniform, ranging from 17.7% (Hyderabad) to 21.8% (Kukatpally). Cash on delivery has a slightly higher rate (20.6%) than UPI (18.1%), and rates by weekday lie between 18.3% and 20.6%.

Table IV. Failure Rate by Location, Payment Method and Day.

Attribute	Category	Records	Failure rate (%)
Location	Kukatpally	812	21.8
Location	Gachibowli	852	20.1
Location	Secunderabad	803	18.9
Location	Madhapur	842	18.6
Location	LB Nagar	821	18.4
Location	Hyderabad	870	17.7
Payment	COD	1,939	20.6
Payment	Card	1,012	19.0
Payment	UPI	2,049	18.1
Day of week	Monday	726	18.5
Day of week	Tuesday	730	18.6
Day of week	Wednesday	699	19.7
Day of week	Thursday	726	18.3
Day of week	Friday	710	20.6
Day of week	Saturday	686	20.4
Day of week	Sunday	723	18.7

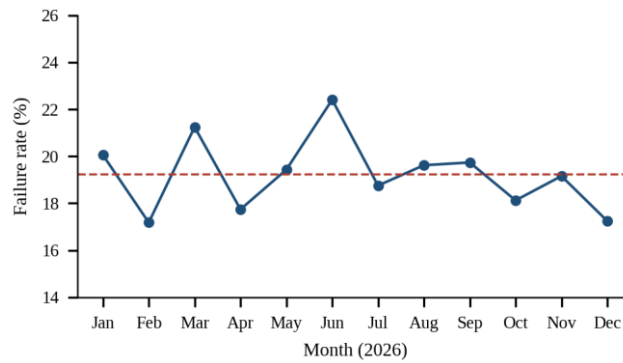


Fig. 9. Monthly delivery failure rate during 2026. Dashed line: overall failure rate (19.24%). The monthly failure rate fluctuates between 17.2% (February and December) and 22.4% (June), without a sustained trend (Fig. 9).

E. Delivery Time

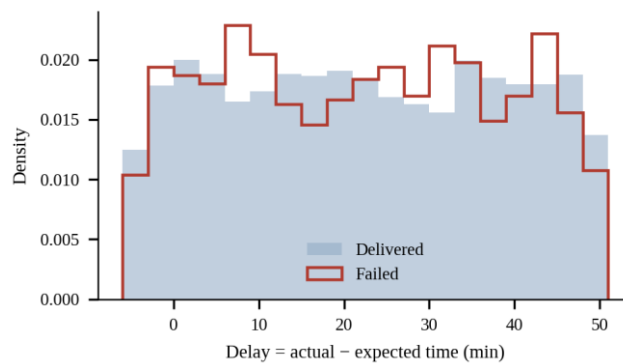


Fig. 10. Distribution of delay (actual minus expected delivery time) for delivered and failed orders.

The recorded delivery time exceeds the planned time for most orders: the mean expected time is 45.9 minutes against a mean actual time of 67.9 minutes, and 88.9% of orders took longer than planned. However, the delay distributions of delivered and failed orders are almost identical (Fig. 10), with mean delays of 22.0 and 21.8 minutes respectively. Delay therefore does not separate failed from successful deliveries in this dataset.

F. Statistical Association

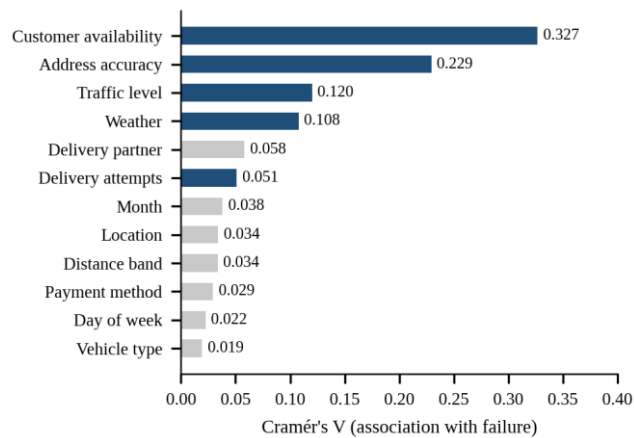


Fig. 11. Strength of association (Cramér's V) between each attribute and delivery failure. Dark bars are significant at the 0.05 level.

Table V and Fig. 11 give the chi-square results. Customer availability ($V = 0.327$) and address accuracy ($V = 0.229$) have the strongest associations with failure, followed by traffic level (0.120) and weather (0.108). Delivery attempts is also significant but weak (0.051). For vehicle type, location, month, day of week, payment method, delivery partner and distance band, the tests do not reject independence at the 0.05 level. The significant results remain significant under the Bonferroni threshold of 0.004.

Table V. Chi-Square Test Results.

Attribute	χ^2	df	p-value	V	Signif.
Customer availability	533.7	1	< 0.001	0.327	Yes
Address accuracy	263.3	1	< 0.001	0.229	Yes
Traffic level	72.3	2	< 0.001	0.120	Yes
Weather	58.1	3	< 0.001	0.108	Yes
Delivery partner	16.8	19	0.603	0.058	No
Delivery attempts	13.0	2	0.002	0.051	Yes
Month	7.3	11	0.777	0.038	No
Location	5.7	5	0.332	0.034	No
Distance band	5.7	4	0.225	0.034	No
Payment method	4.3	2	0.117	0.029	No
Day of week	2.5	6	0.866	0.022	No
Vehicle type	1.9	3	0.596	0.019	No

These observations are descriptive associations within the project dataset and do not establish causation.

10. DASHBOARD AND VISUALIZATION

The Streamlit dashboard can present total deliveries, successful and failed deliveries, and the overall failure rate as key indicators. Charts can display failure reasons, location-level failure rates, delivery outcomes by vehicle type and comparisons across delivery conditions. Filters support both high-level monitoring and focused investigation.

Table VI lists the indicator values for the complete dataset, that is, with no filters applied. When the user selects a subset, for example a single location or weather condition, the same indicators are recalculated for the selected records, which lets an operations team compare a specific segment with the overall figure of 19.24%.

Table VI. Key Performance Indicators. (all records)

Indicator	Value
Total deliveries	5,000
Successful deliveries	4,038
Failed deliveries	962
Overall failure rate	19.24%
Average distance	12.97 km
Average expected time	45.9 min
Average actual time	67.9 min
Average delivery attempts	1.33

The charts in Section 9 correspond to the views that the dashboard can provide interactively: a failure-reason chart, rate charts for each attribute and combined views of two attributes. Presenting these views together with the headline indicators allows users to move from the overall rate to the specific conditions in which failures concentrate.

11. DISCUSSION

The project demonstrates how structured analytics can turn delivery records into operationally useful summaries. The overall failure rate provides a headline metric, while attribute-level analysis adds context. The repeatable workflow can be applied again when updated delivery records become available.

Two groups of factors emerge from the results. The first group concerns the recipient: customer availability and address accuracy are associated with the largest differences in failure rate, and their combined effect is strong. These are conditions that can potentially be influenced before dispatch, for instance through address validation at checkout, delivery-slot confirmation or reminders to the customer. The second group concerns external conditions:

high traffic and rainy or stormy weather are associated with markedly higher failure rates, which suggests that schedules and time buffers could be adjusted for such conditions. These are suggestions for further evaluation rather than findings of this study.

Equally informative is what showed no significant association: vehicle type, location, distance band, payment method, day of week and delivery partner. In this dataset, failure is therefore not concentrated in a particular vehicle class, area or partner, and improvement efforts directed at those dimensions would be unlikely to reduce failures by much.

For future prediction, the results suggest a distinction between information known at dispatch (address accuracy, forecast traffic and weather) and information known only at or after the delivery attempt (customer availability at the door, number of attempts). A model intended for operational use should rely on the former to avoid using information that would not be available when a decision is made.

12. LIMITATIONS

The dataset reflects the available project records and may not represent all real-world delivery conditions. External information about routes, real-time traffic, weather and customer behavior may be limited. The current workflow is descriptive and should not be described as a validated machine-learning prediction system unless a model is trained and evaluated separately.

Further limitations apply. The records cover a single year and six locations. The chi-square tests examine one attribute at a time and do not adjust for the other attributes, so overlapping effects, such as traffic and weather occurring together, are not separated. Delivery attempts and failure reason are recorded as part of the outcome and are not independent explanatory variables. Finally, the failure reason is recorded only for failed orders, so the reasons for unsuccessful first attempts that were later delivered cannot be analyzed.

13. FUTURE WORK

Future work may include larger real-world datasets, live traffic and weather data, GPS and route attributes, automated dashboard refresh, cloud deployment and machine-learning models. Candidate models can be evaluated with suitable classification metrics and validation procedures before operational use.

In particular, a classification model using only dispatch-time attributes could be compared against the simple baseline of always predicting the majority class, using precision, recall, F1-score and ROC-AUC on a held-out test set or with cross-validation. Multivariate methods

such as logistic regression would also allow the joint effect of traffic, weather, customer availability and address accuracy to be estimated. Additional data such as customer contact history or delivery time windows may further improve understanding of recipient-related failures.

14. CONCLUSION

This paper presented a data-driven workflow for analyzing last-mile delivery failures using Python, pandas, SQL and Streamlit. The project processes 5,000 records and reports a dataset-level failure rate of 19.24%. The workflow supports descriptive analysis and interactive exploration of delivery patterns, while providing a foundation for future validated predictive modeling.

The analysis shows that failures are concentrated in identifiable conditions. Customer unavailability and incorrect addresses account for the largest share of recorded failure reasons, and these two attributes, together with traffic level and weather, have the strongest statistical association with failure. Vehicle type, location, distance, payment method and day of the week show no significant association. These findings are descriptive and specific to the project dataset, but they indicate where operational attention and future predictive modeling effort are most likely to be worthwhile.

REFERENCES

1. W. McKinney, "Data Structures for Statistical Computing in Python," in Proc. 9th Python in Science Conf., 2010, pp. 56–61.
2. C. R. Harris et al., "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
3. Pandas Development Team, "pandas Documentation."
4. Streamlit, "Streamlit Documentation."
5. SQLite, "SQLite Documentation."
6. J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.
7. P. Virtanen et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, pp. 261–272, 2020.