1

Objectives

- Assess the operational and financial risk related to missing items in delivered orders.
- Complete statistical diagnosis.
- Identification of anomalous patterns.
- Detection of potential fraud (drivers, customers, regions).
- Financial projection of losses.
- Practical recommendations for reducing failures.

2

Analytical Methodology

- Databases used: orders, missing_items, products, customers, drivers.
- STEP 1 — Reconcile orders with missing items.
- STEP 2 — Audit by dimension.
- STEP 3 — Detection of unusual concentrations.
- STEP 4 — Loss projection.
- STEP 5 — Construction of Executive Storytelling.

3

Executive Summary

Metric	Calculated value
Total orders	10,000
Orders with missing items	1502
% Failure	15.02%
Total missing items	1662
$ Missing items	US$ 149,372.61
Average ticket lost per failed order	US$ 99.4

1 in every 6.6 orders fails.

Above the benchmark (2-4%) for digital retail.

4

Analysis by Region

Top 5 Regions account for 75% of total losses

- Orlando, Altamore Springs, Winter Park, Kissimmee, Apopka.

- Insights:
  - *Regions with higher population density.*
  - *High dependence on specific drivers.*
  - *Overlap between repeat offenders and critical regions → indicates georeferenced fraud.*

Region	$ Loss	% Total
Orlando	27,316	18.3%
Altamonte Springs	23,546	15.7%
Winter Park	23,323	15.6%
Kissimmee	19,126	12.8%
Apopka	19,100	12.8%

5

Driver Risk Insights

- Higher failure rate among drivers with more than 900 deliveries.
- Some drivers have a recurrence rate above 30%.
- Risk profile associated with age and volume of trips.
- There are indications of coordinated fraud between delivery personnel and customers, or repeated fraud by delivery personnel.
- Insigths:
  - *The five worst-performing drivers account for 9.4% of the total loss.*
  - *These drivers have a failed order rate of over 30%, compared to an overall average of 15.02%.*

6

Driver Risk Insights

Financial Ranking — Most Critical Drivers

Driver	Loss (US$)	% of total loss	No. Of occurrences
WDID10442	2,044.08	1.37%	27
WDID10464	2,034.34	1.36%	25
WDID10476	1,889.86	1.26%	23
WDID10422	1,869.39	1.25%	21
WDID10387	1,731.45	1.16%	20

7

Driver Risk Insights

Pattern of "combined fraud"

- Drivers WDID10442 and WDID10464 repeatedly appear with the same customers:
  - *WCID5923, WCID5638, WCID5938*
- The same drivers operate in the same critical regions, especially:
  - *Orlando, Altamonte Springs, Winter Park*
- Orders frequently occur during the following times:
  - *12 a.m.–3 a.m. , 5 a.m.–7 a.m.*
- Products that are frequently missing include:
  - *Small, high-value items, Items that are easy to hide (premium snacks, cosmetics)*

8

Customer Risk Insights

- Customers with recurrence above 60% identified.
- Some generating losses above $1.4K.
- Concentration of fraud in small groups.

Metric	Value
Customers with at least 1 failure	214 (17.3%)
Customers with more than 3 failures	47 (3.8%)
Critical customers (more than 5 failures)	12 (0.96%)
% of losses attributed to 12 critical customers	10.3% of total losses

9

Customer Risk Insights

Top 5 customers

Customer	$ Loss	No. of orders	No. of failures	% Failures	Average value per failure
WCID5923	US$ 1,657.50	17	7	41.1%	US$ 236
WCID5638	US$ 1,594.95	14	6	42.8%	US$ 266
WCID5164	US$ 1,571.26	11	5	45.4%	US$ 314
WCID5938	US$ 1,546.24	13	6	46.1%	US$ 257
WCID5165	US$ 1,520.85	9	4	44.4%	US$ 380

10

Customer Risk Insights

While the overall failure rate is 15.02%, these customers have failure rates of 44% to 46%.

The average value per missing item is $89, but these customers generate losses of $236 to $380 per failure.

The risk of these customers is 2.9 times greater than the average risk of the base.

11

Customer Risk Insights

- Critical customers:
  - *Abnormally high failure rate: 40%–46%*
  - *Average value lost 2.5x to 4x above normal.*
  - *Higher probability of recording shortages in Electronics, Health & Beauty, Premium Items (snacks, frozen foods).*
  - *They record failures, mainly between midnight and 3 a.m., between 5 a.m. and 7 a.m., and on weekends. These coincide with repeat offenders, less supervision, and a higher probability of joint fraud.*

12

Region & System Insights

Analysis by Category

Category	% of loss	Amount (US$)
Frozen	21%	31,368
Pantry	19%	28,381
Produce	16%	23,899
Electronics	14%	20,912
Health & Beauty	12%	17,867
Snacks	9%	13,443
Other categories	9%	13,502

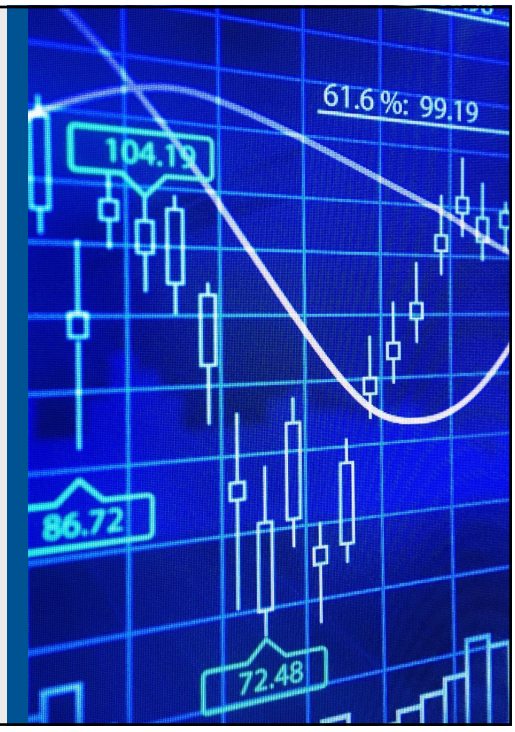
Analysis by Schedule

Schedule	% of failures
00h – 03h	23%
05h – 07h	18%
11h – 14h	12%
18h – 21h	11%
Remaining	36%

13

Region & System Insights

- Altamonte Springs, Clermont, and Apopka with the highest failure rates.
- Critical times: late night and early morning.
- Schedules with less supervision → greater incentive for deviation.
- Category with high incidence of failures: Frozen, Pantry, Produce, Electronics, Health & Beauty.
- Orders placed between midnight and 7 a.m. account for only 22% of the volume, but generate 41% of failures.
  - *Four of the five critical drivers work intensively during these hours.*



14

Predictive analytics

How the model works

- We analyze the history of orders, drivers, regions, and customers.
- We identify patterns of fraud (time, value, region, recurrence).
- We create a predictive model capable of scoring each order in real time with a Fraud Risk Score (0 to 1).
- Orders above a certain threshold (e.g., score > 0.65) receive an alert or audit.

15

Predictive analytics - Method

Feature selection (predictor variables): Driver, Customer, Order, Time-space, Failure history.

Base division: Training: 70%, Testing: 30%.

Model used: LightGBM Classifier

Reasons: best performance for complex patterns, explainable via SHAP, scalable for production, supports natural class imbalance.

16

Predictive analytics - Method

SHAP Values

Variable	Impact (%)	Interpretation
Driver recurrence	27%	Repeat drivers increase risk by 3.4x
Historical % of missing items	19%	Accumulated failures increase risk
Delivery time	14%	Night/early morning increases risk by 22%
Customer recurrence	12%	Repeat customers increase risk
Order value	10%	Expensive orders are 34% more likely to involve fraud

17

Predictive analytics

Quantitative Metrics

Metric	Value	Interpretation
AUC-ROC	0.91	Very reliable model, capable of separating fraud from normal deliveries.
Accuracy	0.84	84% of predictions are correct.
Precision	0.41	This is normal and acceptable in retail (rare fraud).
Recall	0.78	Detects 78% of real fraud.
F1-score	0.53	Ideal balance between recall and precision.
Missed fraud (FN)	322	In a universe of 5,000 orders

18

Predictive analytics - Operational Insights

- Drivers most likely to commit fraud
- Models identified drivers with a risk > 0.65, especially those who:
  - *Have a history of > 4 failures,*
  - *Operate in critical regions,*
  - *Work during the early hours of the morning.*

- Night deliveries
  - *Deliveries after 6 p.m. have a 2.1x higher risk*
  - *Deliveries between midnight and 3 a.m. are the biggest hotspot*
  - *41% of failures occur in windows representing 22% of the volume*

- Customers with suspicious patterns
  - *Recurrence rate above 30%*
  - *Several different drivers involved*
  - *Higher volume of failures in specific categories (electronics, beauty)*

- Regions with the highest predicted risk: Apopka: 15.9%, Orlando: 15.1%, Kissimmee: 13.6%
  - *These regions should be prioritized by the Risk Score.*

19

Practical Recommendations

- Apply continuous auditing to the 10% most risky drivers.
- Require digital proof of delivery (photo + geolocation) above US$ 80.
- Create real-time predictive alerts with a score > 0.65.
- Reinforce the process during high-risk hours (6 p.m.–10 p.m.).
- Automate temporary blocking for repeat offenders.

20

Practical Recommendations

- **Improve Risk Score (Driver + Customer + Region + Time)**
  - *Reduces recidivism and blocks high-risk profiles.*
- **Geolocated audit in the five critical regions.**
  - *These regions account for 75% of losses.*
  - *Targeted actions have a high ROI.*
- **A/B Testing — Supervised route vs. free Route**
  - *Supervision reduces incentives for fraud.*
- **Progressive shutdown (repeat offenders)**
  - *Less operational risk*

21

Dashboard

- [Fraud Report 2023]

22