DATA VISUALIZATION

DESIGN AND IMPLEMENTATION OF A DATA-DRIVEN BI FRAMEWORK: FROM DATA TO DECISIONS IN DELIVERY PERFORMANCE

Le Huynh Thuy Vy, Student

AGENDA

01 Introduction

03 BI Solutions

05 Conclusions and Future Work

02 Building Data Warehouse

óc

04 Data Visualization Using Power BI

Chapter 01: Introduction

1.1. Pain Point & Objectives

OBJECTIVE REASON

- Fragmented and unstructured operational data
- Difficult to monitor performance across drivers, orders, and regions
- Lack of integrated system for real-time KPI tracking

SUBJECTIVE REASONS

Build an end-to-end data visualization system using Microsoft Fabric
Centralize and standardize operational data from multiple sources
Enable real-time KPI monitoring (AR, FR, CR, Productivity)

1.2. Research Question & KPIs

Research Question

How can data visualization help identify and explain the decline in operational performance and high cancellation rates at Ahamove?

AR

Acceptance Rate

FR

Fulfillment Rate

CR

Cancellation Rate

Productivity

Completed order per driver

1.2. Data Preparation

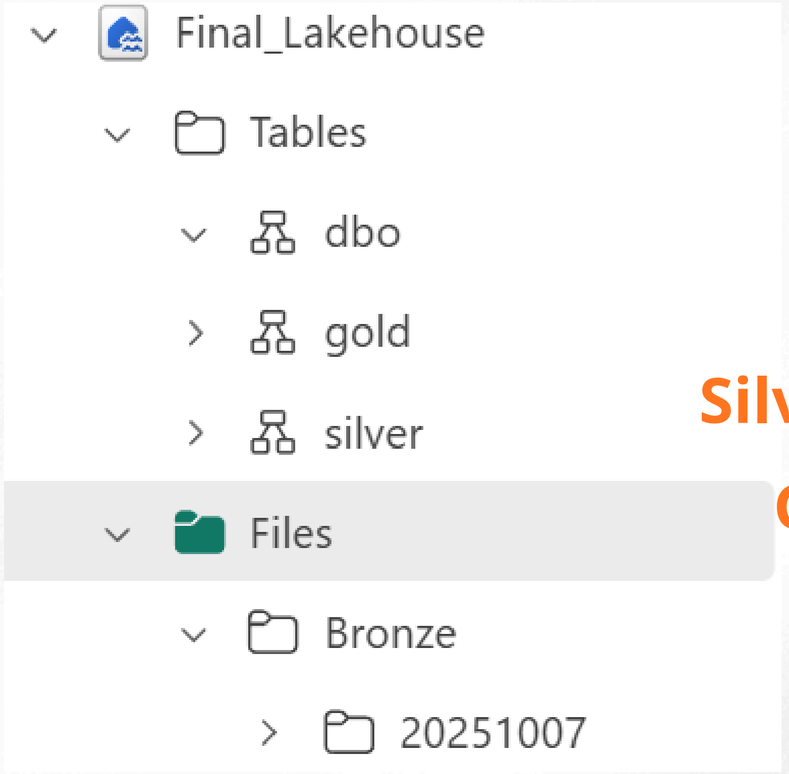
General Objectives:

DATA SCOPE	DATA SOURCE TOOLS	KEY FINDINGS

Dataset provided by one business department of Ahamove
Covers orders in Hanoi and one specific service type
Focused operational scope for deeper analysis
	Internal Ahamove system – preprocessed and standardized dataset
	Tools: Microsoft Fabric, PySpark, Python (pandas, numpy, seaborn)
		• No duplicated records
		• Null values appear logically based on order status:
		  ○ Cancelled orders: missing driver/time fields
		  ○ Completed orders: missing cancellation fields
		• Data is clean and ready for modeling and visualization

Chapter 02: Pipeline in Fabric

2.1 Fabric & Lakehouse Architecture



Medallion Architecture:

Bronze – Raw data ingestion (raw file format)

Silver – Cleaned & standardized Delta tables (Delta Parquet)

Gold – Analytical Delta tables (fact & dimension) for BI

2.2. Tools & Services

Microsoft fabric

OneLake (Lakehouse): centralized storage for all layers

Delta Parquet for Silver & Gold: ACID, schema evolution, time-travel

- PySpark Notebooks: scalable transformations and SCD handling
- Fabric Data Pipeline: scheduling & orchestration (daily)
- Power BI Direct Lake: near real-time analytics without data duplicatio

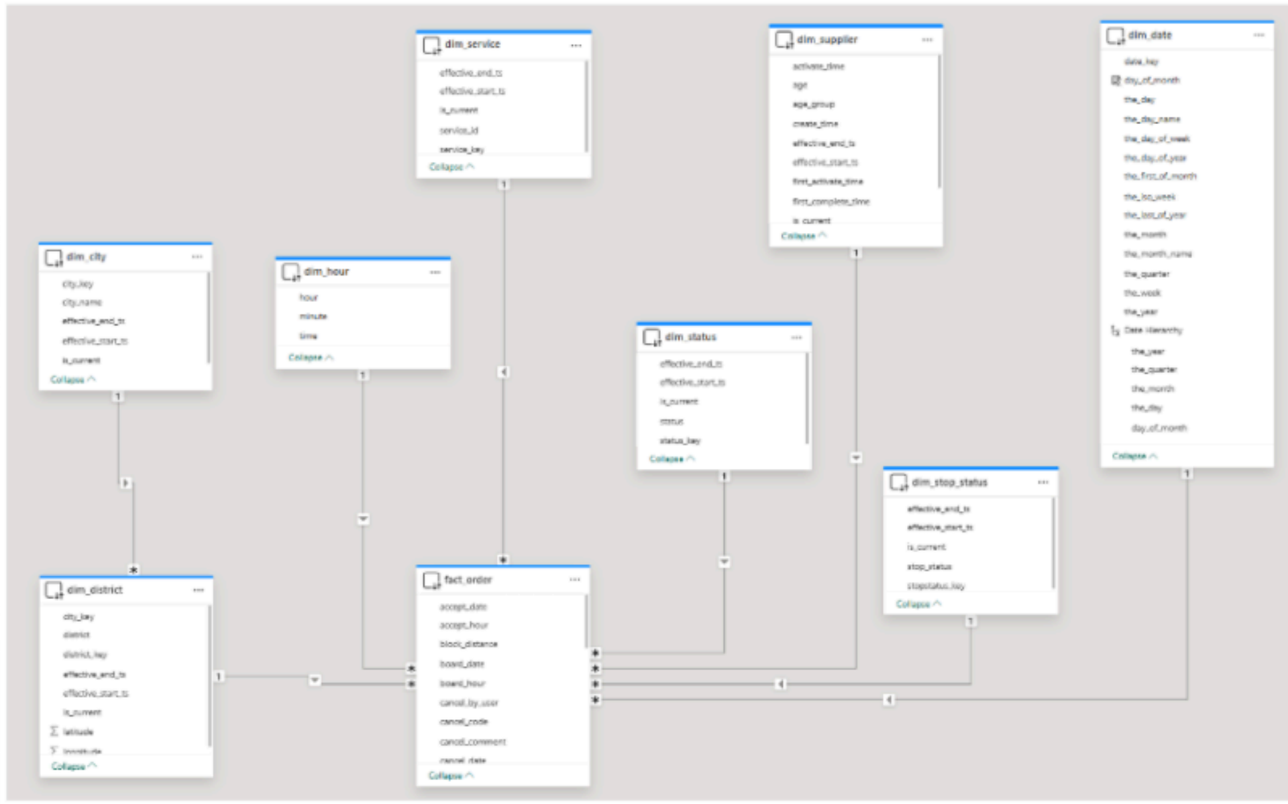
Python
PySpark
Delta Parquet

2.3. Snowflake Schema & Data Modeling

Data Modeling – Snowflake Schema
Fact Table:
Fact_Order – core transactional data

Dimension Tables:
- Dim_Supplier, Dim_Date, Dim_Hour, Dim_Service, Dim_Status, Dim_City, Dim_District

Why Snowflake Schema?
- Reduces data redundancy
- Supports City → District hierarchy
- Optimized for Power BI joins and queries

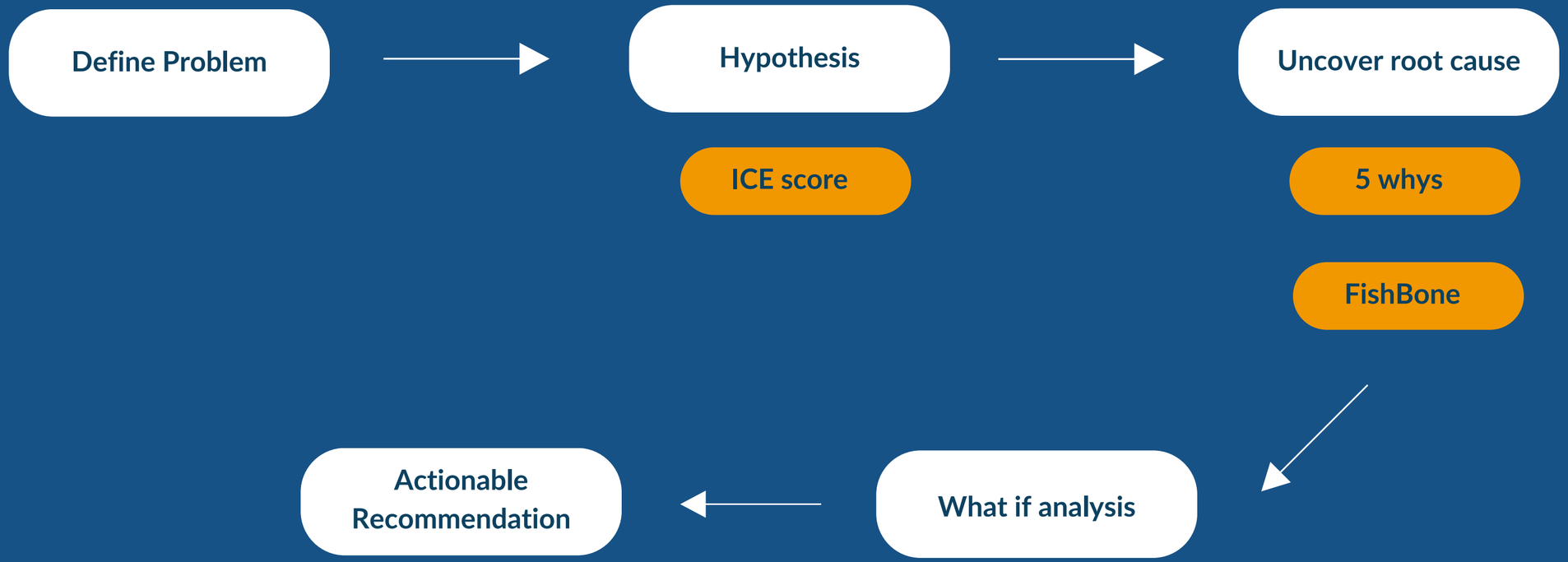


Chapter 03: Data Visualization Using Power BI

What Happened

Why Happened

Finding Insight



DashBoard Flow

3.2. Problem Statement

PROBLEM 1

Ahamove's operational performance showed a significant decline from mid-year onwards, with the Acceptance Rate (AR) dropping sharply in September and failing to consistently meet the 75% target, while the Fulfillment Rate (FR) remained persistently below the 70% warning threshold.

PROBLEM 2

Internal Ahamove system – preprocessed and standardized dataset
Tools: Microsoft Fabric, PySpark, Python (pandas, numpy, seaborn)

3.3. PROBLEM 2 & HYPOTHESIS

Hypothesis Evaluation for Problem 2 (ICE Framework):

ID	Hypothesis	Impact	Confidence	Ease	ICE Score
H1	Cancellation BEFORE Driver Acceptance	5	5	5	15
H2	Primary Cause: USER-Initiated Actions	4	5	5	14
H3	Excessive Waiting Time	5	4	4	13
H5	Key Account (KA) Group Performance Skew	5	4	4	13
H4	Longer Delivery Distance Correlation	5	2	2	9

3.2. Problem Statement

PROBLEM 1

Ahamove's operational performance showed a significant decline from mid-year onwards, with the Acceptance Rate (AR) dropping sharply in September and failing to consistently meet the 75% target, while the Fulfillment Rate (FR) remained persistently below the 70% warning threshold.

PROBLEM 2

Internal Ahamove system – preprocessed and standardized dataset
Tools: Microsoft Fabric, PySpark, Python (pandas, numpy, seaborn)

3.4. Root Cause Analysis: Problem 1

Method: 5 Whys to uncover the true root cause beyond surface symptoms

Step	Original Question / Analysis Focus	Finding & Data	Conclusion
1.	Why did AR/FR drop sharply from June onwards?	Requested Stops ↑ but Complete Stops ↔. AR plummeted in Sep. FR consistently below 70% threshold.	Decline is a Systemic Issue, not just seasonal.

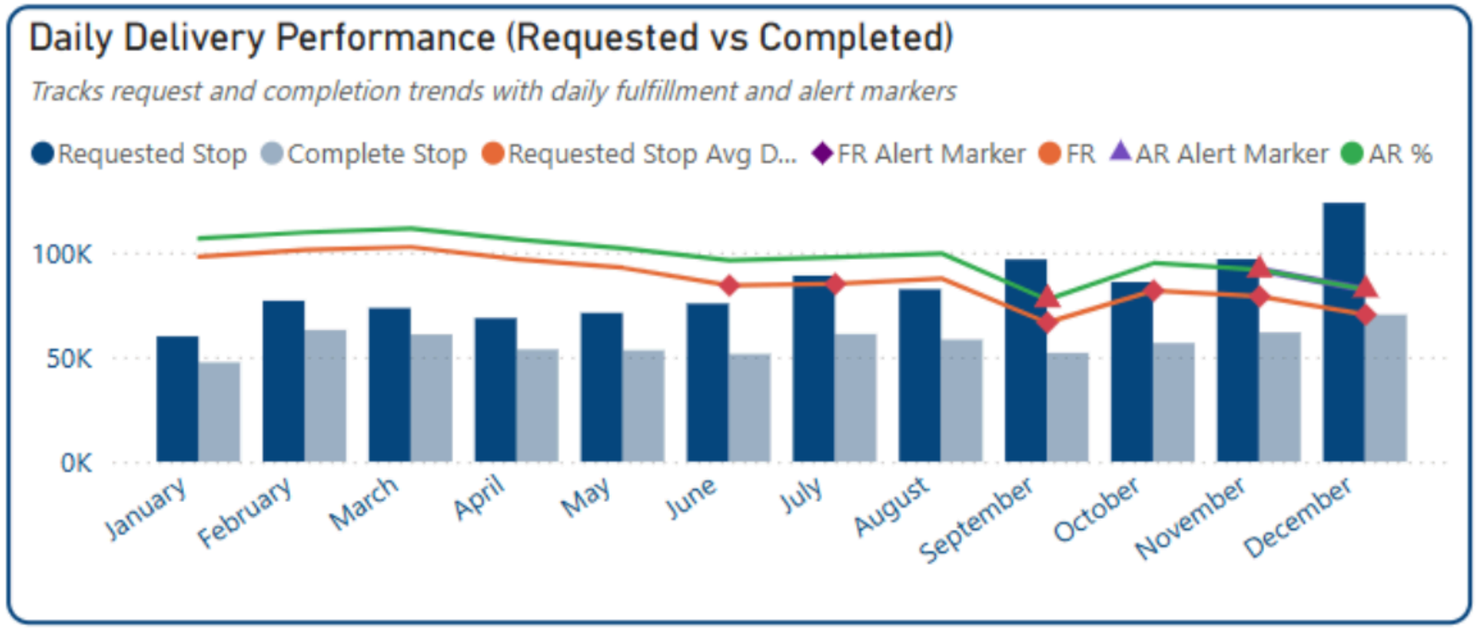


Figure 5.4: Daily Delivery Performance

3.4. Root Cause Analysis: Problem 1

Method: 5 Whys to uncover the true root cause beyond surface symptoms

Step	Original Question / Analysis Focus	Finding & Data	Conclusion
2.	Is driver capacity insufficient to meet demand?	Driver Productivity ↓ severely (5.9 orders/driver in Sep). Demand Growth ≫ Supply Growth.	Drivers are OVERBURDENED, ignoring more orders.

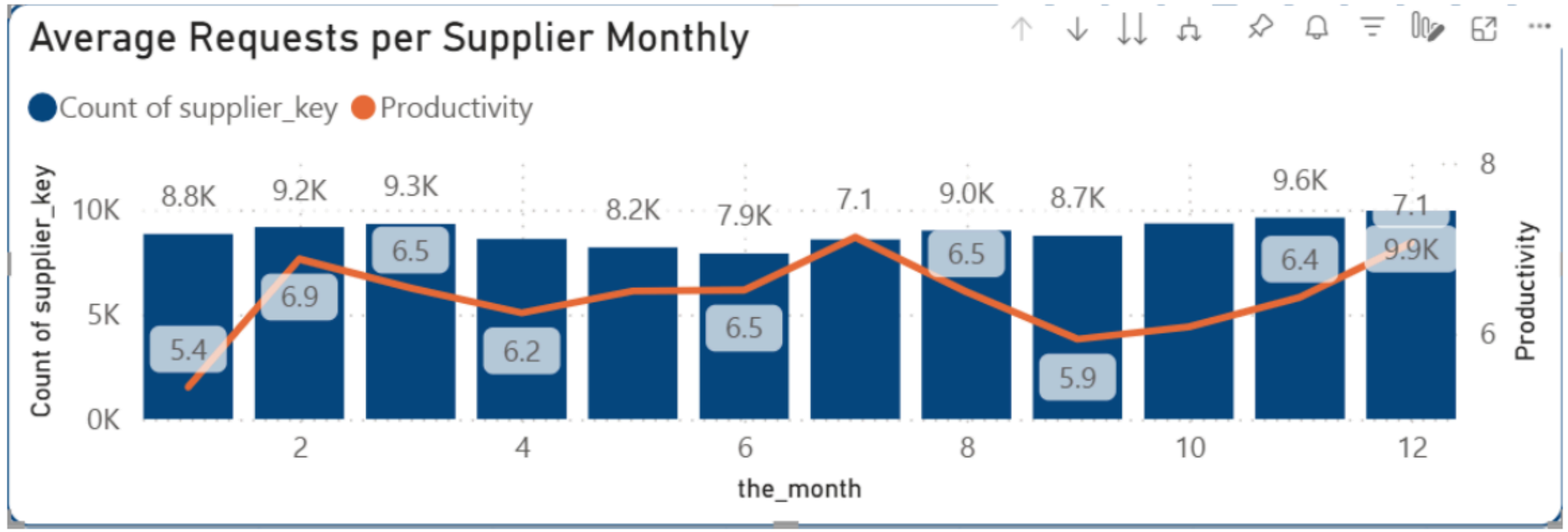
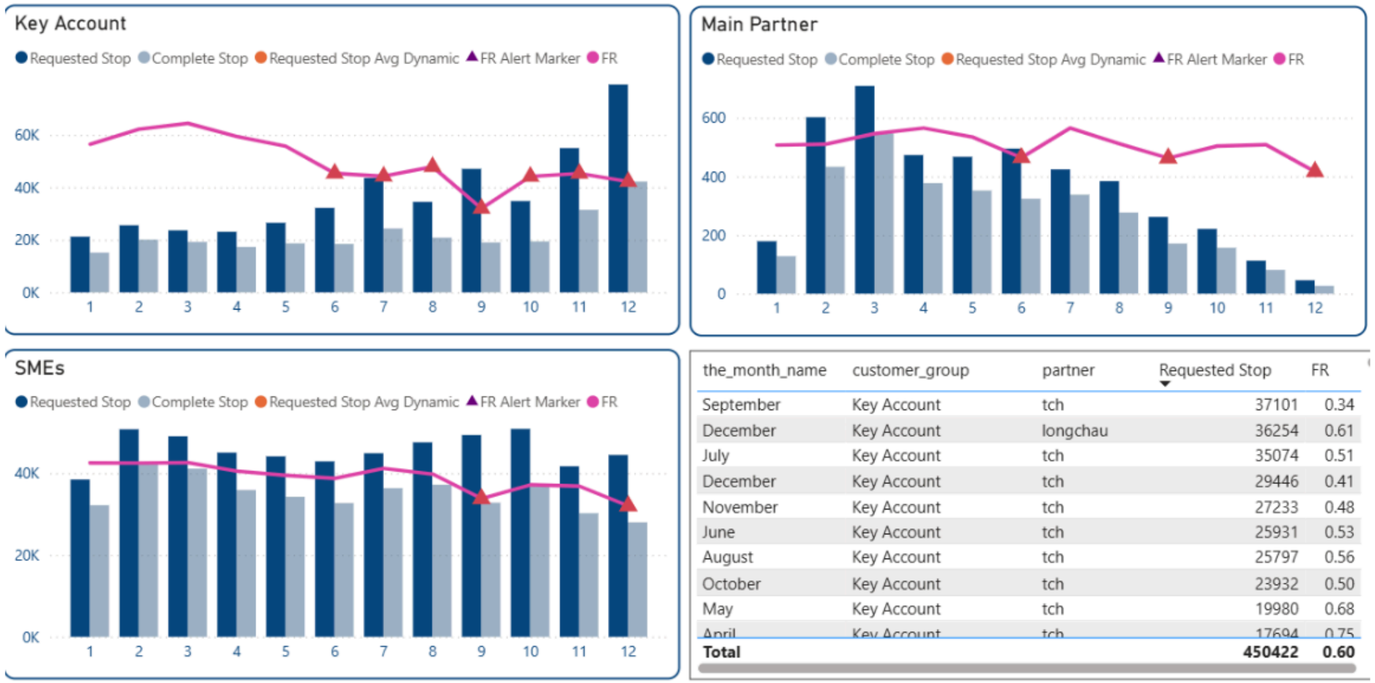


Figure 5.5: Average requests per supplier

3.4. Root Cause Analysis: Problem 1

Method: 5 Whys to uncover the true root cause beyond surface symptoms

Step	Original Question / Analysis Focus	Finding & Data	Conclusion
3.	Which customer group is driving the surge?	Primarily the KEY ACCOUNT (KA) group (TCH, Long Chau). KA volume ↑ from 20K (Q2) to 40K (Dec).	Demand surge is Highly Concentrated in the KA segment.



the_month_name	customer_group	partner	Requested Stop	FR
September	Key Account	tch	37101	0.34
December	Key Account	longchau	36254	0.61
July	Key Account	tch	35074	0.51
December	Key Account	tch	29446	0.41
November	Key Account	tch	27233	0.48
June	Key Account	tch	25931	0.53
August	Key Account	tch	25797	0.56
October	Key Account	tch	23932	0.50
May	Key Account	tch	19980	0.68
April	Key Account	tch	17694	0.75
Total			**450422**	**0.60**

Figure 5.6: Monthly Delivery Performance by Key account

ROOT CAUSE
Why does KA demand cause the bottleneck?
KA FR plummeted severely (0.75 in Apr → 0.34 in Sep).
CRITICAL OPERATIONAL BOTTLENECK due to unsustainable KA demand surpassing fulfillment capacity.

3.5. Root Cause Analysis: Problem 2

Method: Fishbone (Ishikawa) Diagram to categorize potential causes

Bone	Cause Group & Key Finding	Indicator
1. Manpower	Severe Supply Shortage (AR ↓ → Order Ignored)	"Driver no accept" (User) & "Auto-cancel, no driver" (System) are top reasons.

Cancellation Reason	Requested Stop
By user	**44530**
User has no item	2887
Không có tài xế nhận đơn	5913
Driver no accept	21286
Cancel by user	2096
Cancel by User	6901
	5447
By system	**16968**
Nhân viên hủy đơn hàng. Lý do: Hệ thống tự động hủy đối tác giao hàng (đơn hàng quá hạn)	2637
Long package waiting time	313
Cancel by admin for user (Unavailabe Package(s))	129

Cancellation Reason	Requested Stop
By supplier	**897**
Vehicle is broken down	36
Driver's not ready	857
Cancel by supplier (Pick-up location is too far from driver)	1
Cancel by supplier (Driver's not ready)	1
Cancel by supplier (Delivery distance is too far)	2
By others	**182**
Other	143
Người gửi yêu cầu hủy đơn	2
Không thuộc những lý do trên	1
Không liên lạc được Người gửi	1
Total	**62577**

Figure 5.8: Breakdown of Cancellation Reasons by Party in September

All | Low 5 | Top 5

Cancellation Reason	Requested Stop
By system	**9417**
Auto cancel, no driver accepted	8217
Nhân viên hủy đơn hàng. Lý do: Hệ thống tự động hủy đối tác giao hàng (đơn hàng quá hạn)	1040
Long package waiting time	91
Auto cancel, user not pay	40
Cancel by admin for user	29
Total	**9417**

Figure 5.9: Breakdown of Cancellation Reasons by Party in December

3.5. Root Cause Analysis: Problem 2

Method: Fishbone (Ishikawa) Diagram to categorize potential causes

Bone	Cause Group & Key Finding	Indicator
2. Milieu	Adverse Weather Catalyst (Heavy Rain/Storms)	Cancellation Rate ≈ 80% on Sep 14th & 28th (Heavy Rain days).

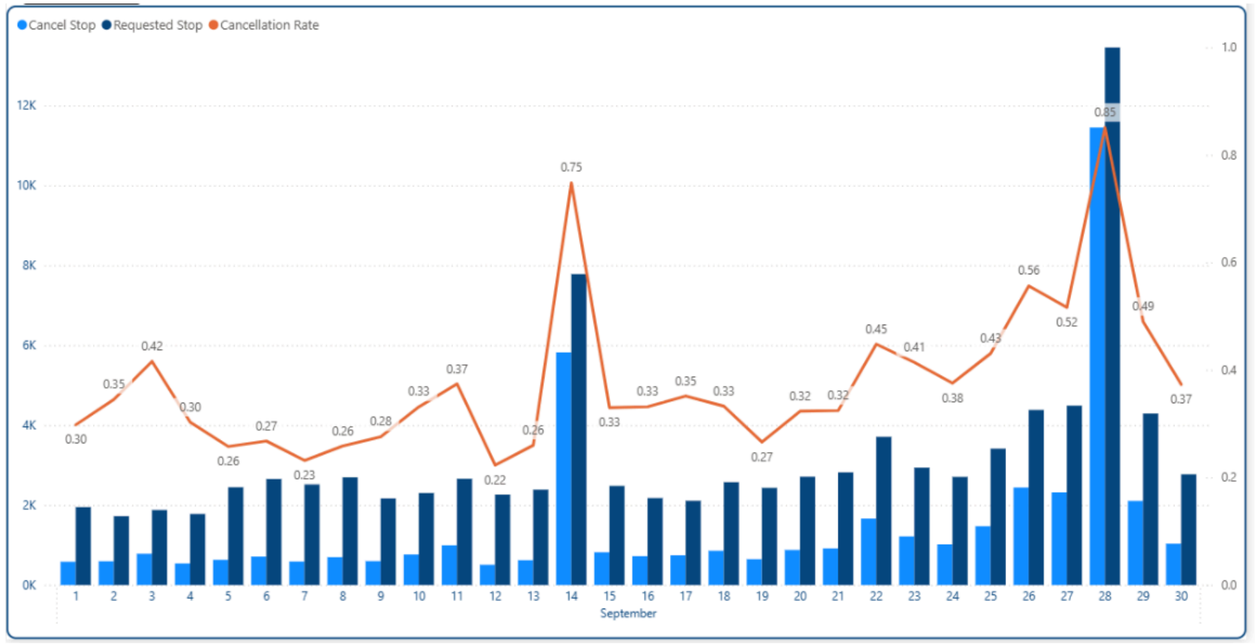


Figure 5.10: Daily Order Volume and Cancellation Rate, September 2023

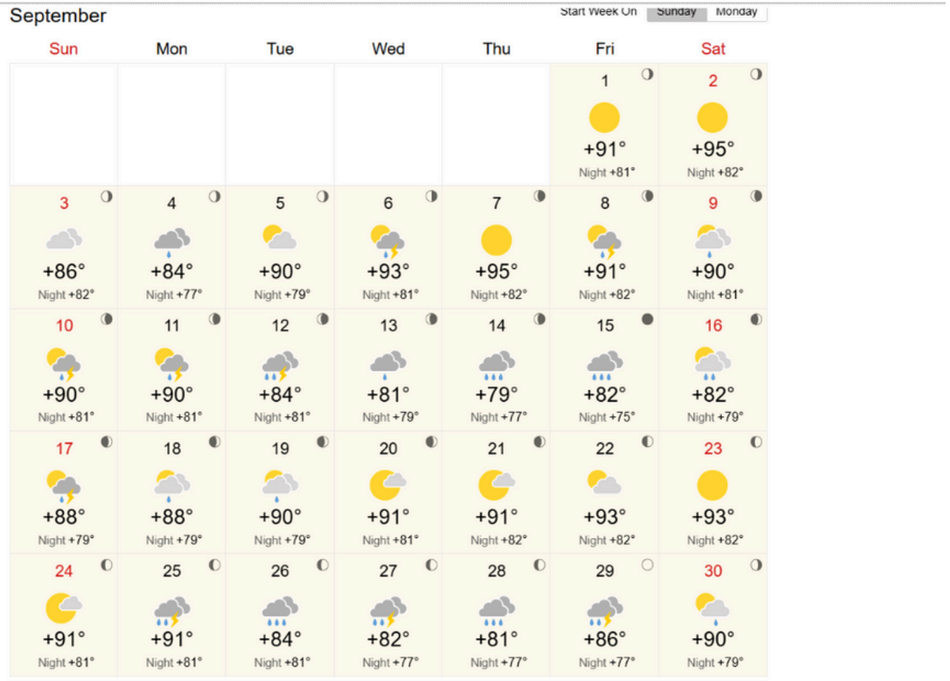


Figure 5.11:Weather History in Hanoi, September 2023. (Source: World Weather Online)

3.5. Root Cause Analysis: Problem 2

Method: Fishbone (Ishikawa) Diagram to categorize potential causes

Bone	Cause Group & Key Finding	Indicator
3. Partners	Underperforming KA Concentration (Low FR/AR, High Volume)	TCH/Long Chau (highest volume) drive disproportionate cancellations.

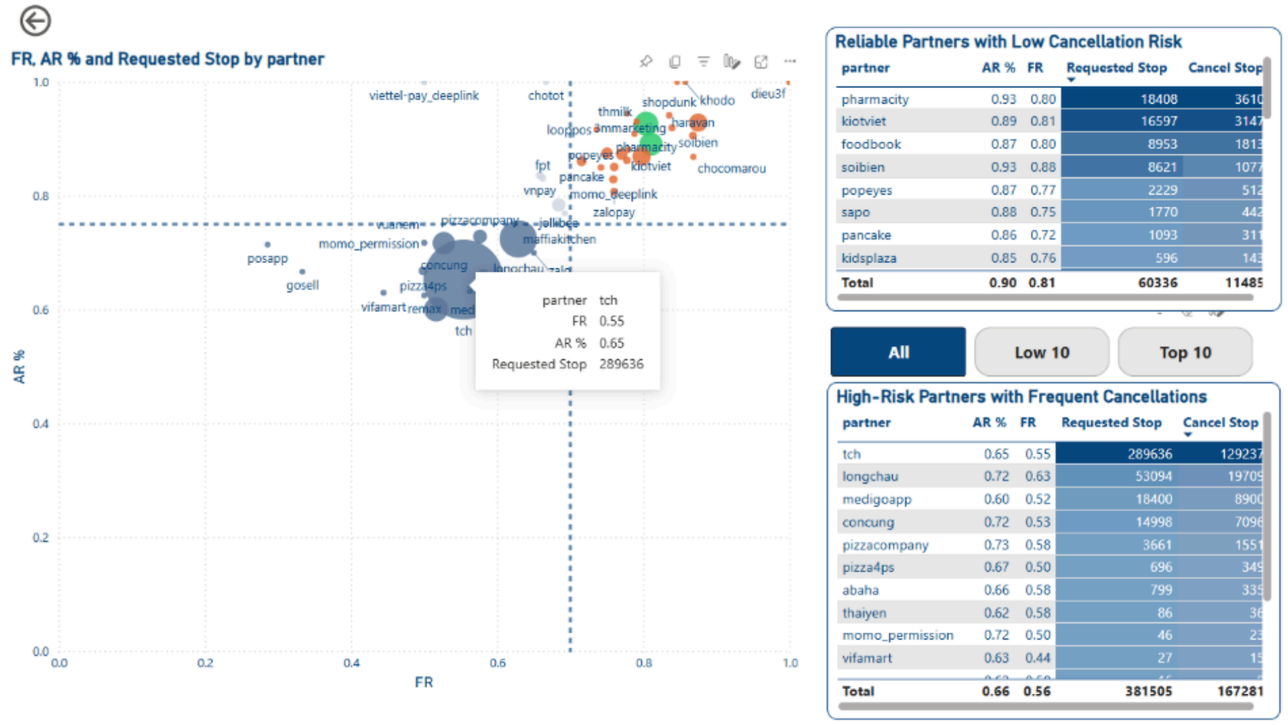
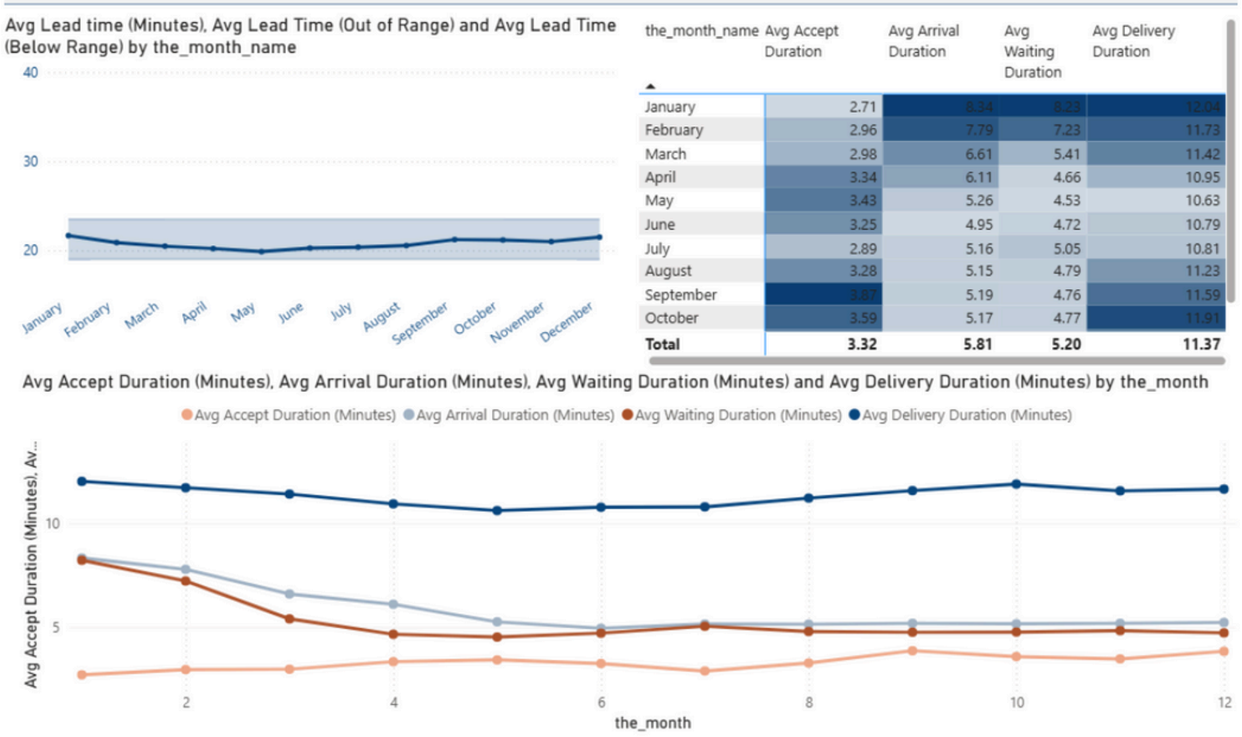


Figure 5.13: Partner Performance Matrix (Acceptance Rate vs. Fulfillment Rate)

3.5. Root Cause Analysis: Problem 2

Method: Fishbone (Ishikawa) Diagram to categorize potential causes

Bone	Cause Group & Key Finding	Indicator
4. Method	System Latency → Increased Wait Time	Accept Duration ↑ 15% (Sep–Dec period). This exceeds user patience threshold.



the_month_name	Avg Accept Duration	Avg Arrival Duration	Avg Waiting Duration	Avg Delivery Duration
January	2.71	8.34	8.23	12.04
February	2.96	7.79	7.23	11.73
March	2.98	6.61	5.41	11.42
April	3.34	6.11	4.66	10.95
May	3.43	5.26	4.53	10.63
June	3.25	4.95	4.72	10.79
July	2.89	5.16	5.05	10.81
August	3.28	5.15	4.79	11.23
September	3.87	5.19	4.76	11.59
October	3.59	5.17	4.77	11.91
Total	**3.32**	**5.81**	**5.20**	**11.37**

Figure 5.14: Average Lead Time and its Components by Month

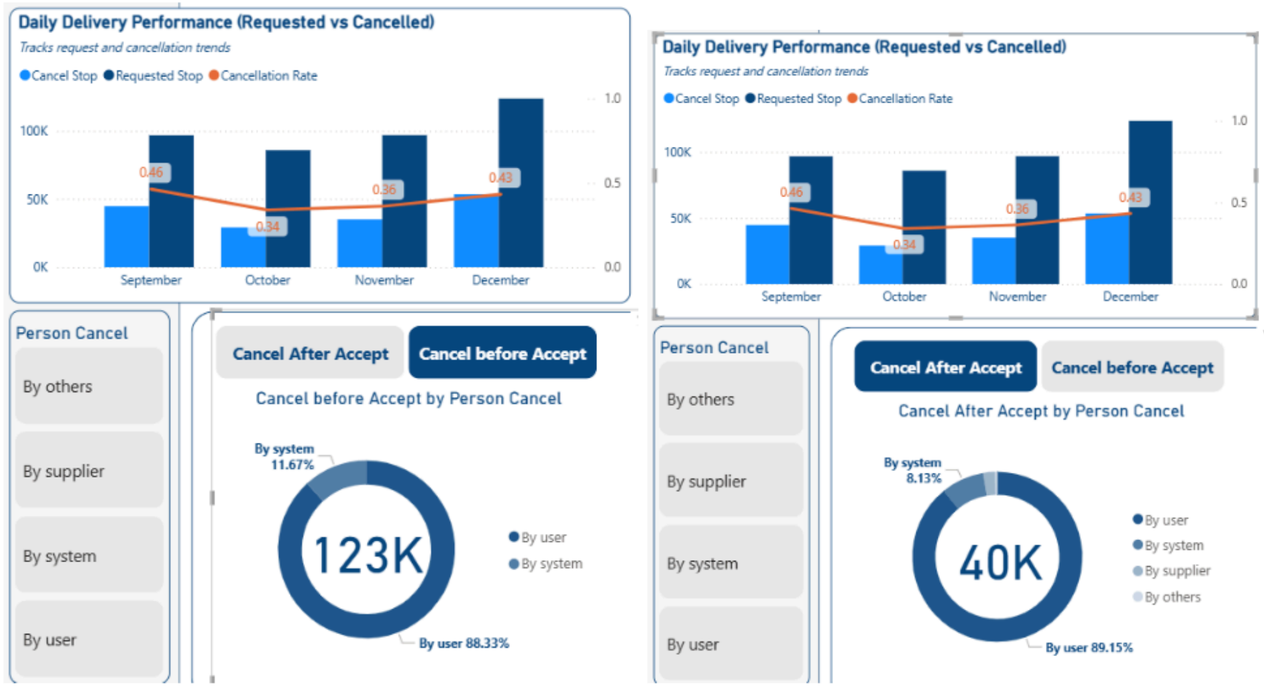
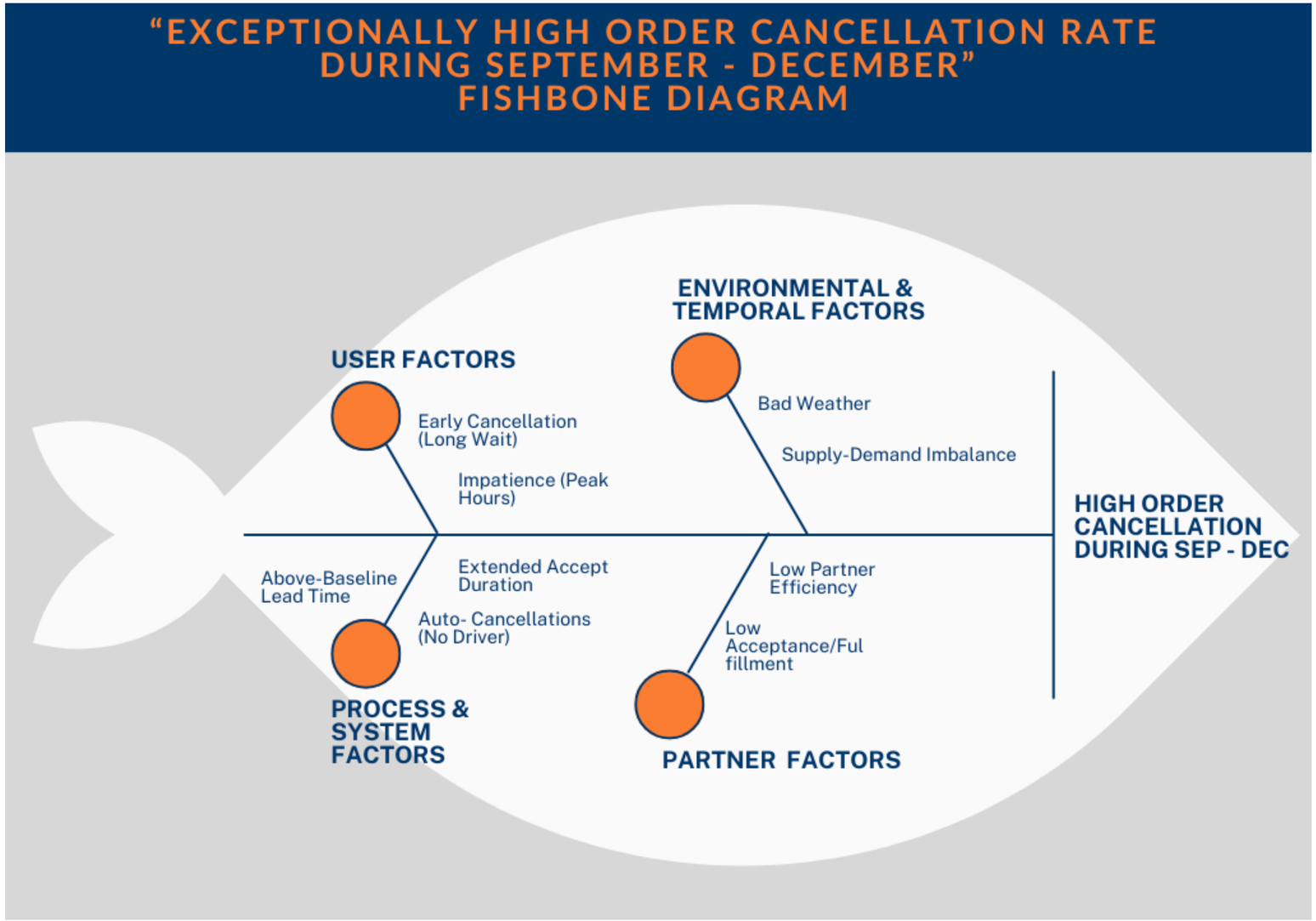


Figure 5.15: Cancellation Breakdown by Party and Type

3.5. Root Cause Analysis: Problem 2

Method: Fishbone (Ishikawa) Diagram to categorize potential causes



Chapter 04: Strategic Recommendations

Action Plan

- **Adjust the Dynamic Incentive system - encourage drivers to focus on high-demand areas.**
- **Strengthen ROI Monitoring - ensure profit per order is optimized.**
- **Communication Plan - Improve customer experience & communication with drivers -> help reduce cancellations due to misunderstandings.**

Dashboards go beyond just data visualization and turn insights into specific action decisions, contributing to improving system operational efficiency.

Chapter 05: Conclusion

5.1 Comparison of Effectiveness

BEFORE VISUALIZATION	AFTER VISUALIZATION
Static reports, discrete data	Automate the entire data process
Difficult to detect trends, correlations	Power BI dashboards display KPIs in real time.
Lack of real-time interaction & tracking	Lack of real-time interaction & tracking
Insight is slow to react	Move from static reporting → dynamic intelligence

5.2 Future Development

FORECASTING

Integrate real-time weather and traffic data to improve forecasting.

PREDICTIVE DASHBOARD

Apply AI/ML to predictive models, helping to proactively coordinate drivers.

COMPREHENSIVE BI SYSTEM

Expand to a national scope instead of just focusing on Hanoi, aiming for a comprehensive BI system for the entire Ahamove.

Thank you