Freight Insights

Sriya Chakravarthula, Ivy Lu, Theodore Omonode, Jonathan Onwude, Parmindar Singh, Kevin Yi

PURDUE UNIVERSITY | The Data Mine

Roche

1 Background

- Roche tracks global logistics using **third-party freight data** containing over **3 million records**, but data quality issues limit its reliability for **forecasting**, **reporting**, and **operational insight.**

Project Goal

- Developed an automated solution to transform datasets into **contextualized visual insights**.
- Reduced repetitive manual review by creating a repeatable process for **identifying anomalies** and standardizing key fields.
- Designed a repeatable process that can support both **current and future data sources.**

2 Tools & Technologies

Development Tools:
- Jupyter Notebook
- Anvil
- Excel

Python Libraries:
- Pandas
- Matplotlib
- GeoPy

3 Methodology

Exploratory Data Analysis

Found recurring data quality issues: missing values, inconsistent formats, unrealistic entries, and incorrect types.

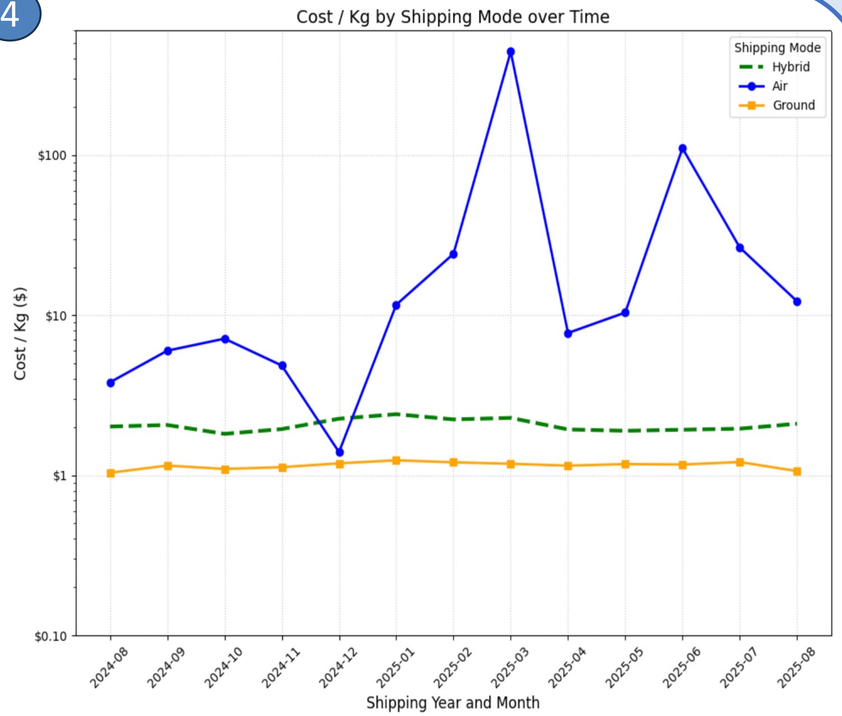
Repeatable Cleaning Workflow

Isolated key columns, ensured consistency, standardized fields like ZIP codes, and flagged suspicious records.

Aggregated Master Dataset

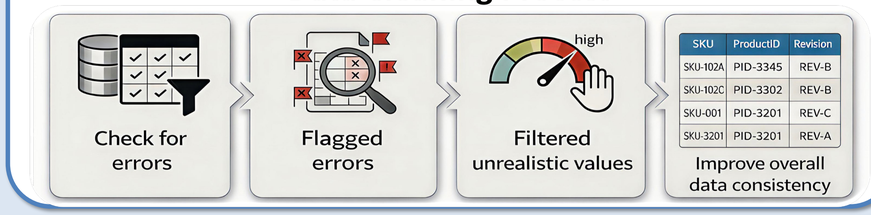
Finalized a revised dataset to support analysis via visualizations and detect changes in operational metrics.

4



- The chart above illustrates the **cost per kilogram across different shipping modes** from August 2024 to August 2025.
- Costs were calculated by **dividing the total amount paid by the billable weight (in kg)** for each shipment.
- The **Air, Ground, and Hybrid** (Air/Ground/Other) lines were calculated by grouping shipments by month (and mode) and dividing the total monthly amounts paid by the total monthly billable weights.
- Grouping by month reveals the **cost trends over time**, while grouping by mode enables a **direct comparison between Air and Ground expenses**.

5 Data Cleaning Process



Disclaimer: This study utilizes pseudonymized data. All identifiers have been removed or masked. The results shown here illustrate data modeling techniques and should not be interpreted as a reflection of real-world activity.

6 Analysis Steps

- Isolated the specific columns
- Identified and addressed missing values
- Detected anomalies and inconsistent records
- Generated visualizations to summarize key patterns

7 Conclusion

- Roche's freight data has **anomalies** that need to be cleansed and harmonized for **reliable analysis.**
- We developed a **repeatable cleaning workflow** to standardize key fields and prepare the data for visualization.
- The resulting **master dataset** supports more **reliable reporting** and **future analysis**.
- Developed a script that produces a work-ready dataset that Roche analysts can use immediately to derive insights.

8 Future Goals

- **Automate**: Data cleansing, aggregated dataset creation, and visualization
- **Adapt**: Flexible code for evolving rules and metrics
- **Enable Insight**: Faster detection and investigation of freight changes
- **Integrate**: Connect freight data with Roche financial datasets for broader analysis.
- **Analyze**: Apply regression analysis and AI driven methods to uncover deeper insights and improve forecasting.

9 Acknowledgements

Mentors: Mike Cassiero, Rich Bottorff, Roche Staff supporting this project

Other Contributors: Atiyah Ellerbee, Andrew Hage, Jayden Loo, Dilveer Singh

The Data Mine Staff: Maggie Betz, Bryce Castle, Jacqui Kane