OList eCommerce

Delivery delays

Impact analysis

Remediations

Presentation structure

Outline

About OList

A Brazilian eCommerce...

15.2 M$ accrued
Sep. 2016 to Sep. 2018

... offering a wide range of products...

71 categories
Mainly home, clothing, and electronics

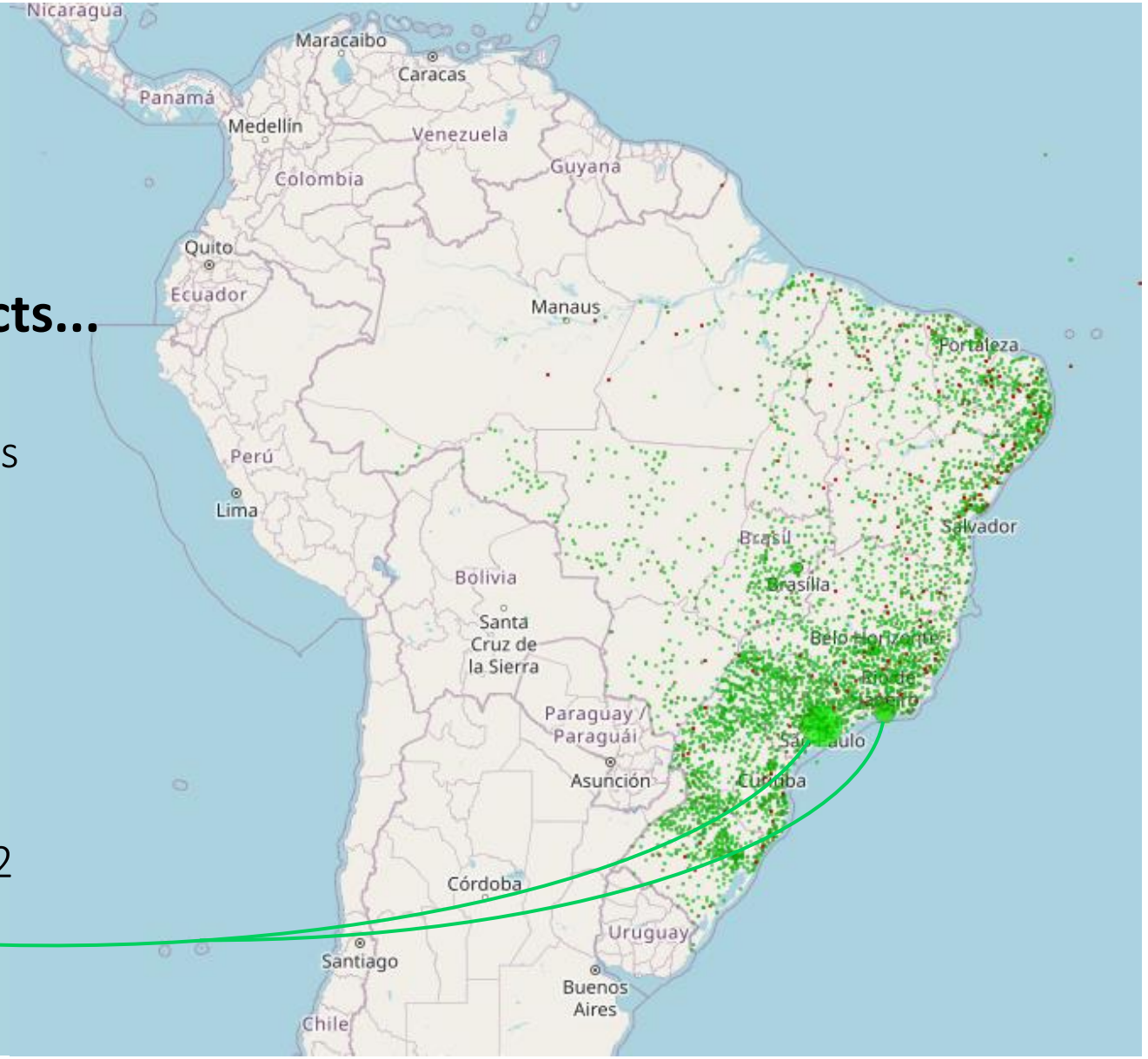
32,000 items sold

3,000 sellers

... and delivering mainly locally.

4,000 cities
Rio De Janeiro and Sao Paulo the top 2

4.2 review average
Based on 90,000+ reviews



Summary

Status, actions, and future directions

Overall, financial standing is sound (low urgency, low risk).

Short-term:

- **Avoid 0% reliable sellers, products, and cities.**
  Negligible losses, but these likely affect data management and logistics.
- **Deliver product “FD00...” reliably (full potential of 200 k$).**
  Despite demand, delivery reliability for this product is abnormally low.

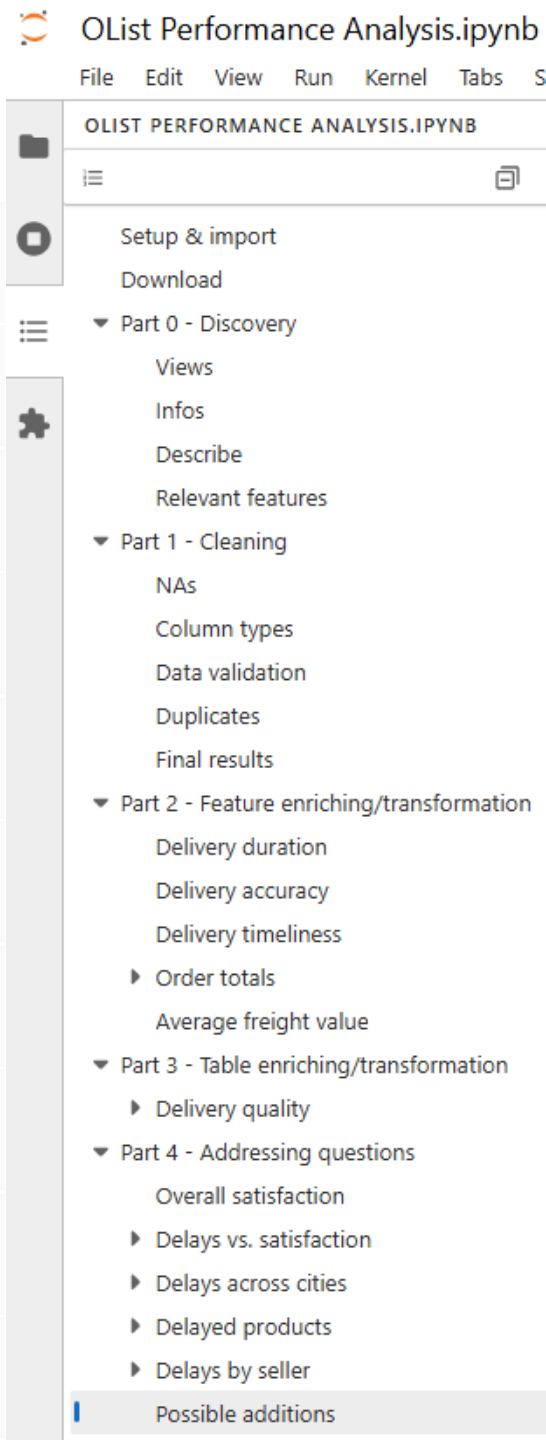
Long-term:

General	Create alerts for abnormal losses. Optimize underpromising to improve customer sentiment.
Sellers	Default policy improvements should be impartial. Monitor exceptions.
Products	Improvements should be targeted; monitor low performance. Deploy product reviews.
Cities	Rio de Janeiro a top candidate for profitable reliability improvement programs.
Database	Synchronize order data across tables. Reduce repeated entries (track item quantity).

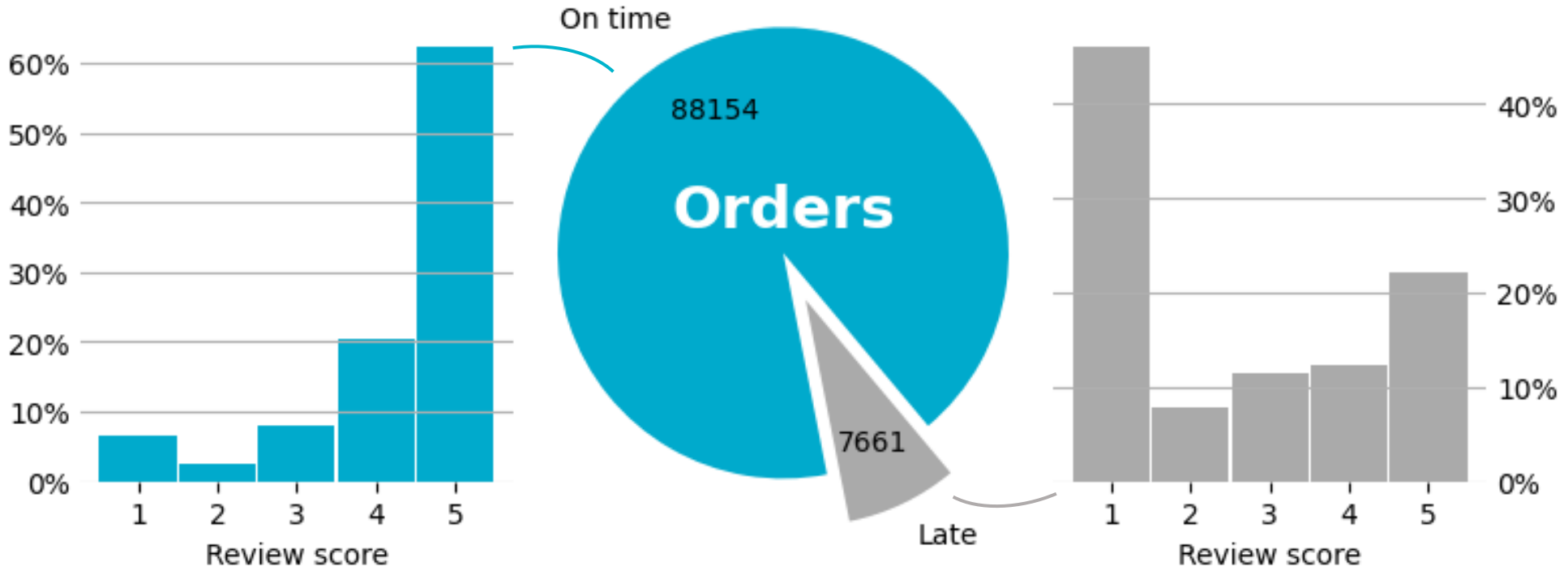
Workflow overview

Methods

Scope	BI analysis, questions from an NPower workshop
Approach choice	Analysis: Python. Platform: Jupyter notebook. Data: as a batch
Data import	9 tables, from Kaggle olistbr/brazilian-ecommerce (.csv)
Discovery	Data intended use, summary statistics, types, table sizes...
Feature selection	Columns from questions (reviews, costs, delays, cities...)
Cleaning	Removing NAs and duplicates, datatype-based formats...
Enriching	Adding features, joining tables (order totals, delays, quality...)
Insight extraction	Reliability (cities, sellers...). Satisfaction (overall, after delays...)
Visualization	Bar charts, folio map, histograms/pareto charts... (next slides)
Recommendations	Made from insights observed (for training purposes only)



Ratings vs. delivery timeliness



Average (on time)
4.3

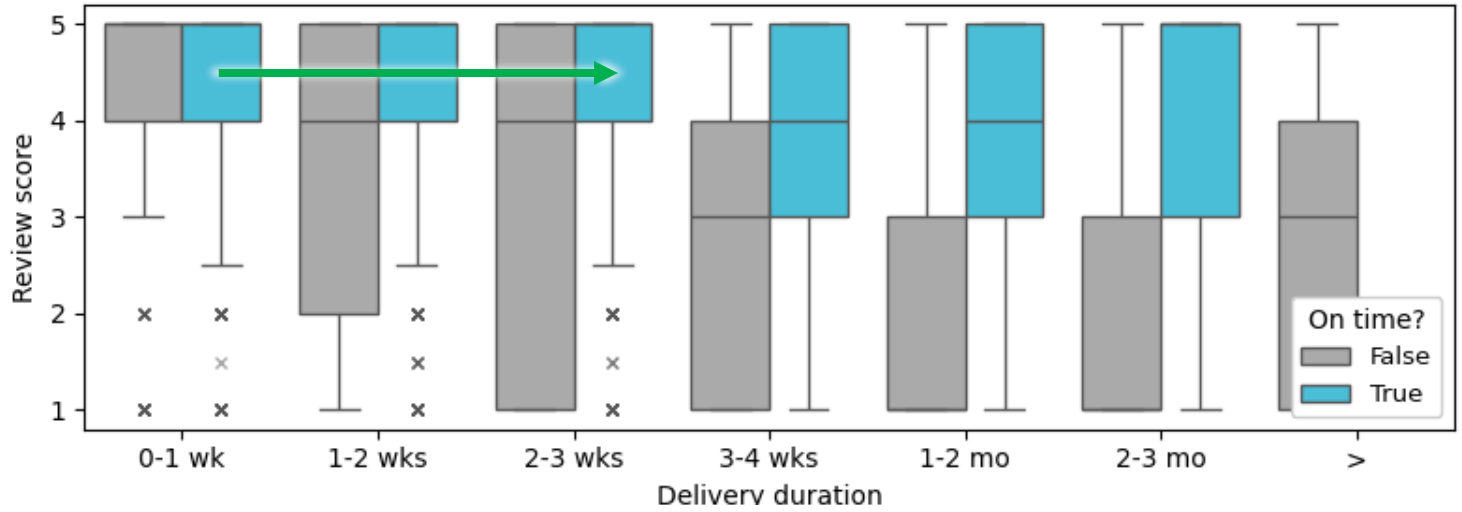
Average (global)
4.2

Average (late)
2.6

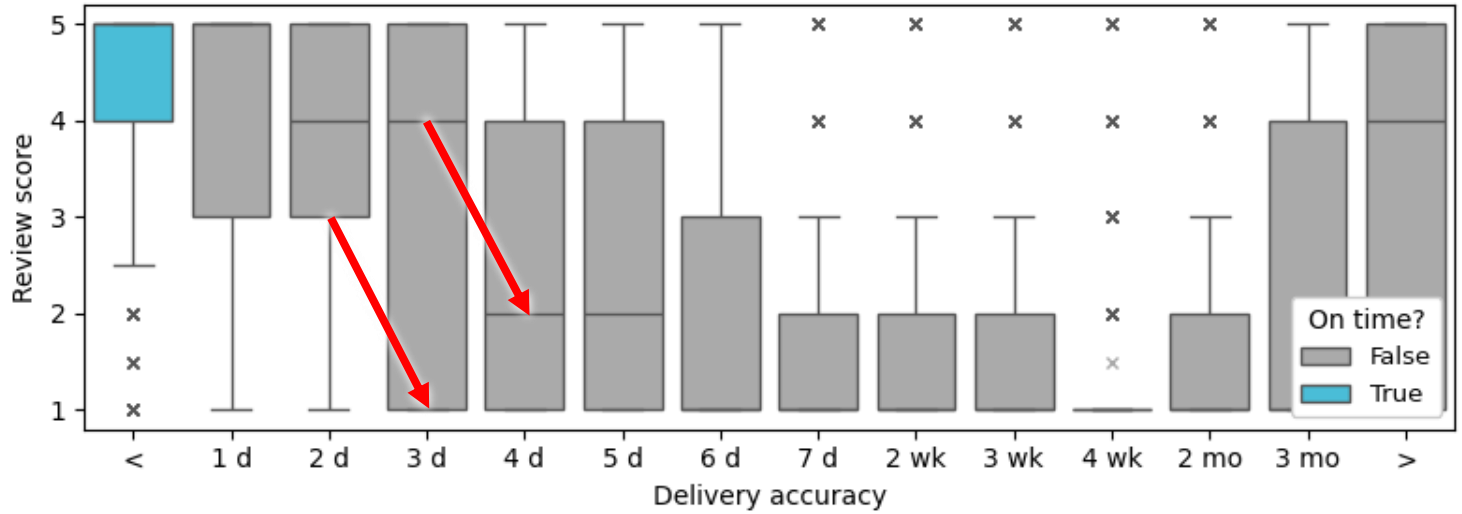
Delayed order ratings are lower by 1.7 points, but rare.

Ratings vs. delivery delays

Customers **don't** mind deliveries taking **3 weeks**.

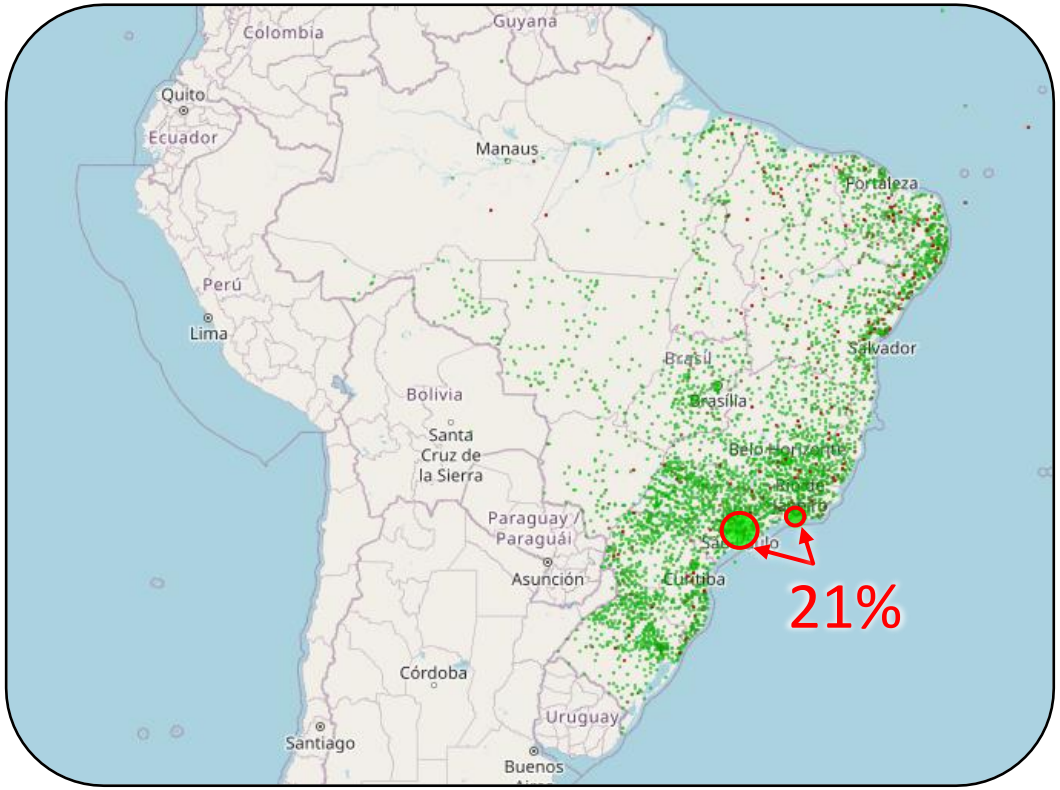


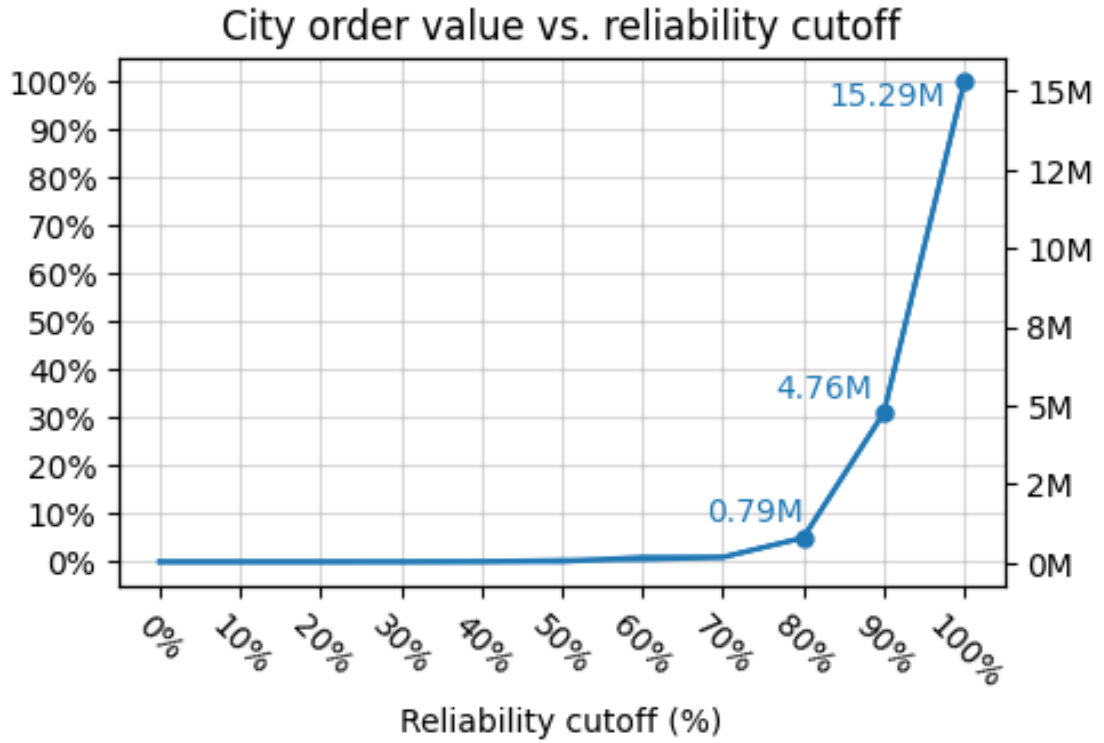
However, they **do** mind deliveries delayed by **3 days**.



These metrics can support an underpromising policy.

City impact by reliability





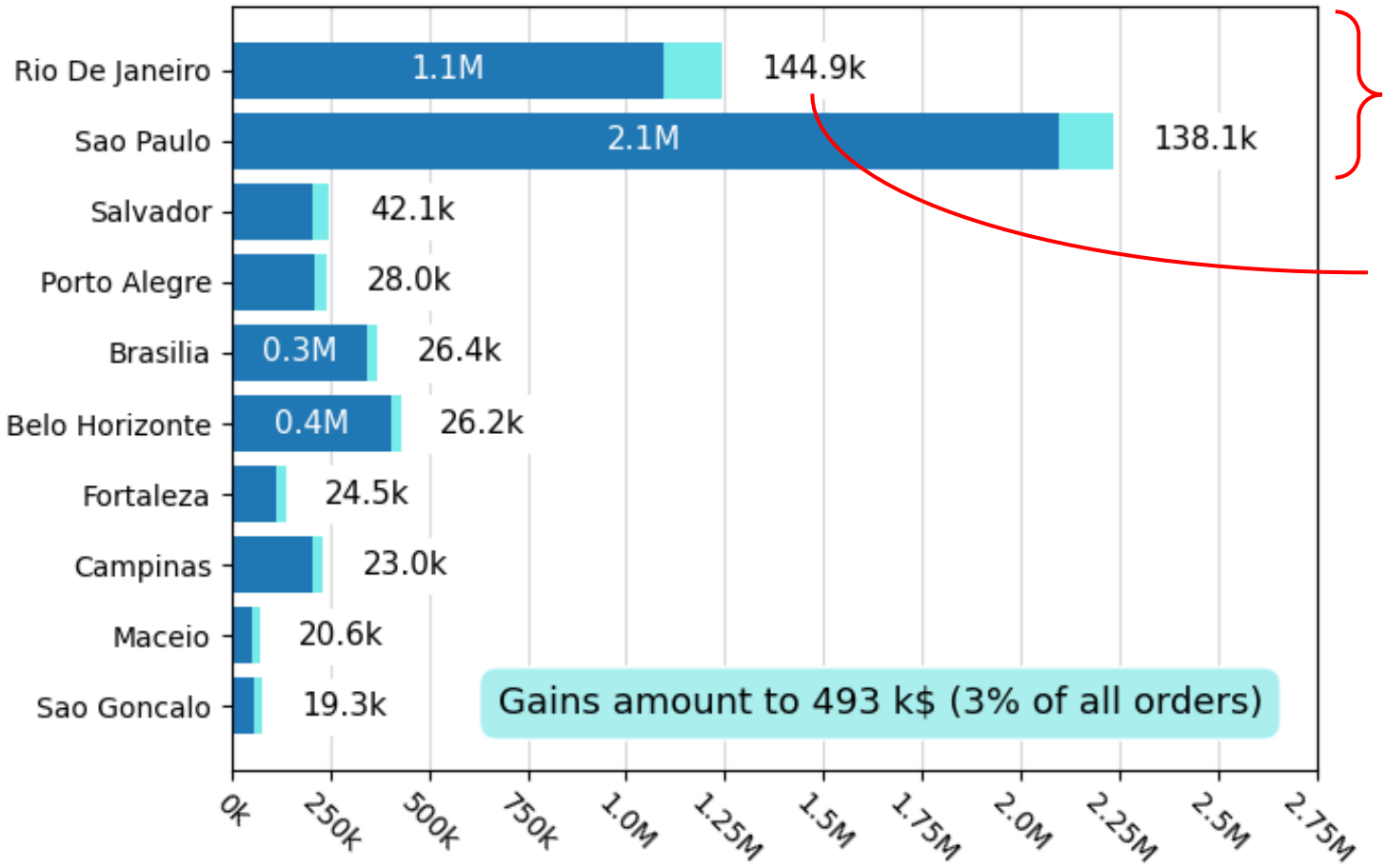
Rio de Janeiro and Sao Paulo share 3.2M$ (21%) of value.

Most value is in top cities

95% of value is in cities with >80% reliability.

This is acceptable, and can be improved further.

Gain potential by city

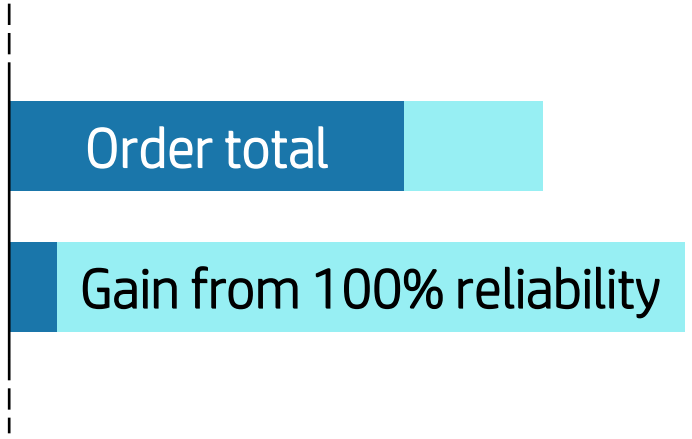


Clear top 2

20% share of all potential gains

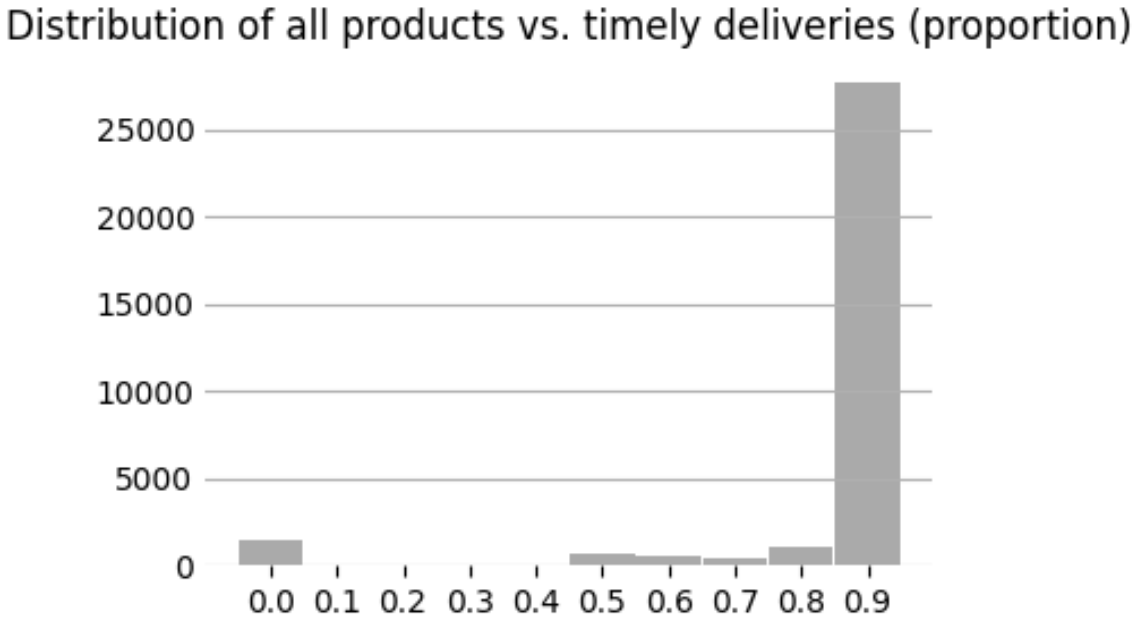
Top gain, lesser effort

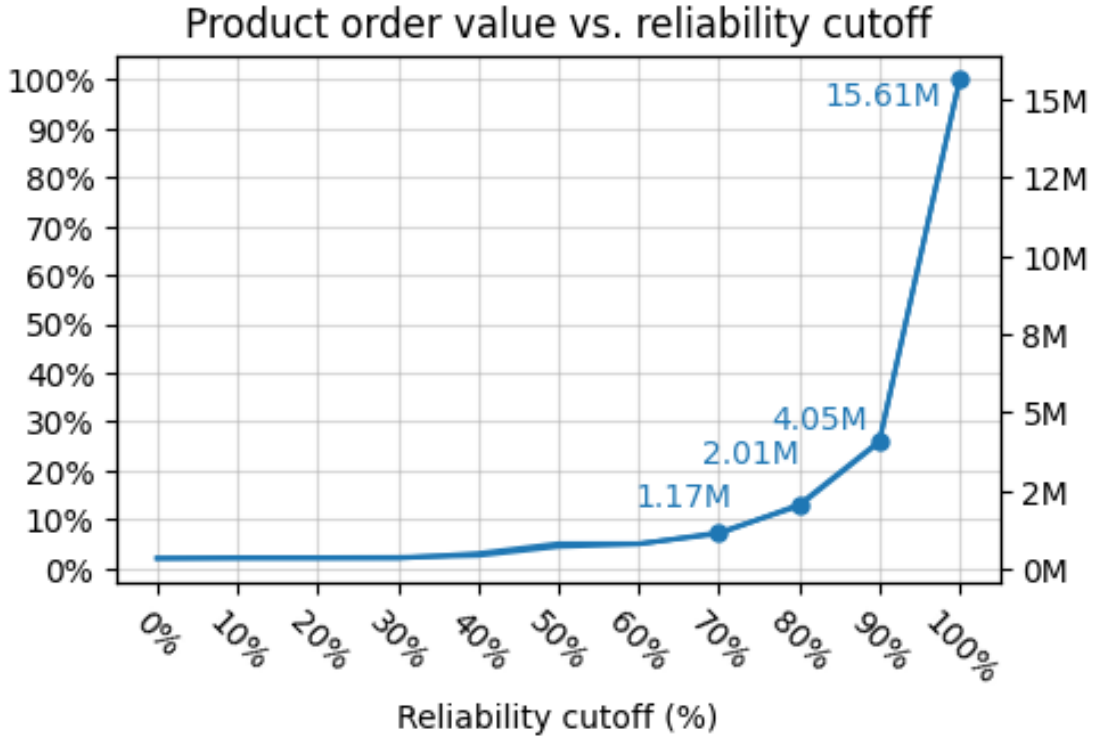
(88% reliability in RDJ, 93% in Sao Paulo)



RDJ is a top remediation candidate, ahead of Sau Paulo.

Product impact by reliability





25k products are >90% reliable.

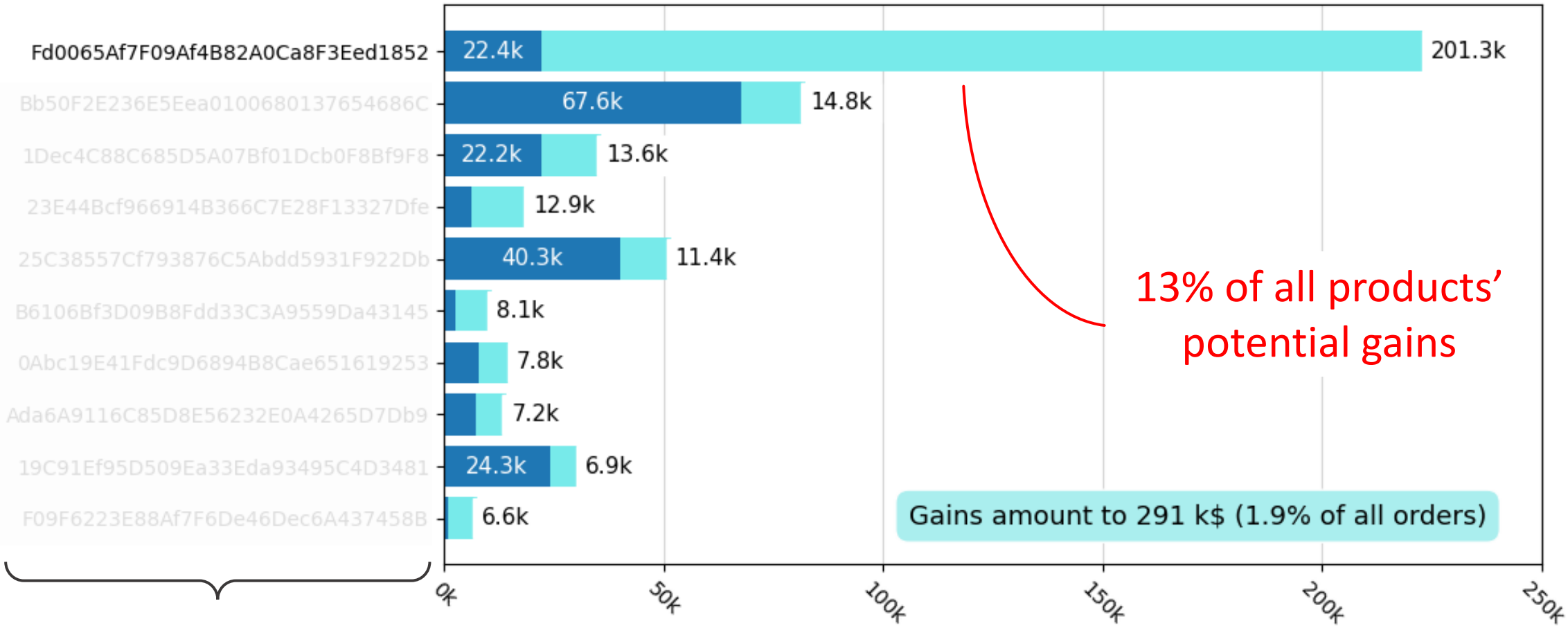
3,2k are <80% reliable, and 1.4k are 0% reliable.

Most value is in top products

88% of value is in items with >80% reliability.

This is acceptable, and should be improved further.

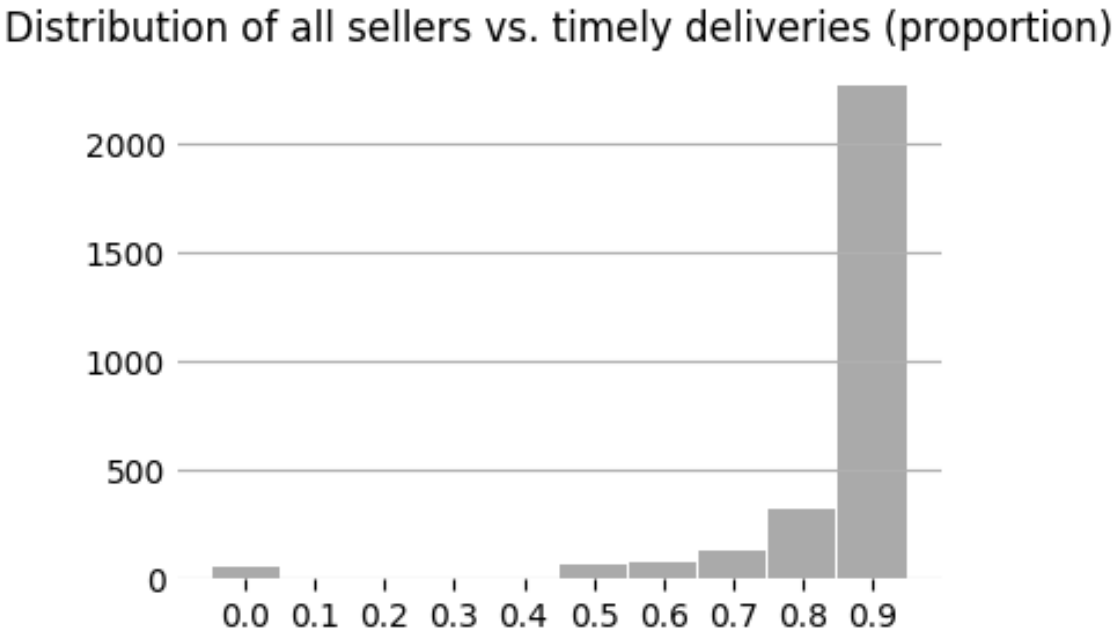
Potential gain by product

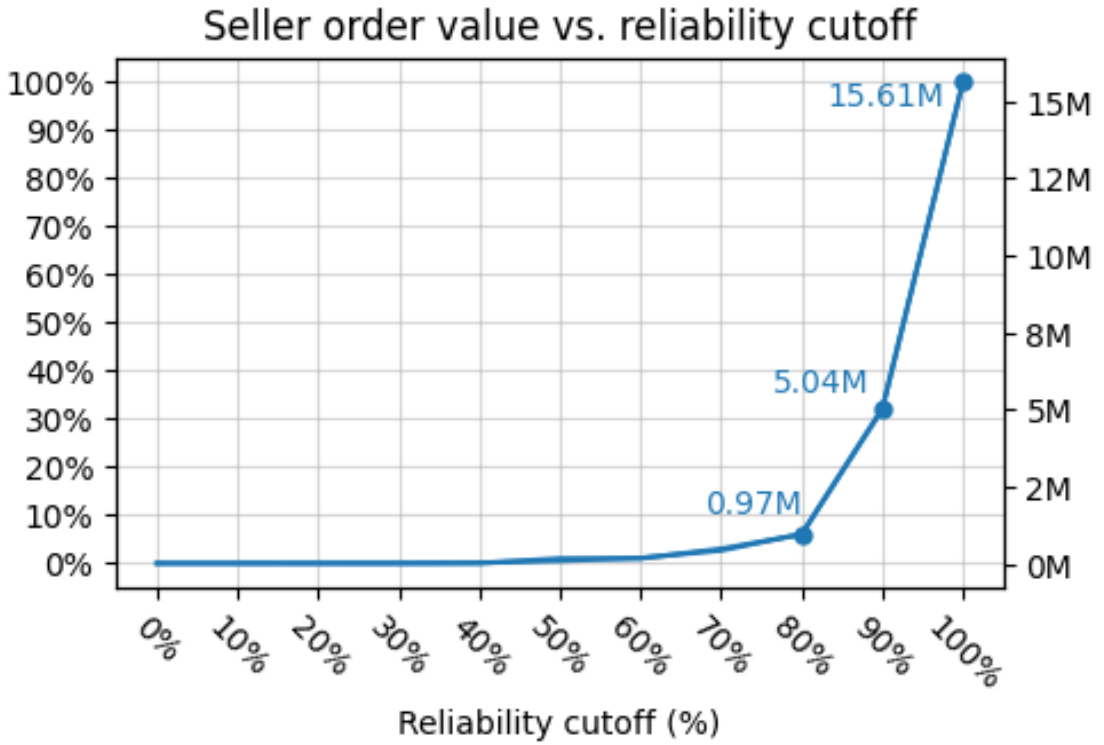


(IDs are obfuscated)

Item "FD00..." stands out as a remediation priority.

Seller impact by reliability





2.5k sellers are >80% reliable.

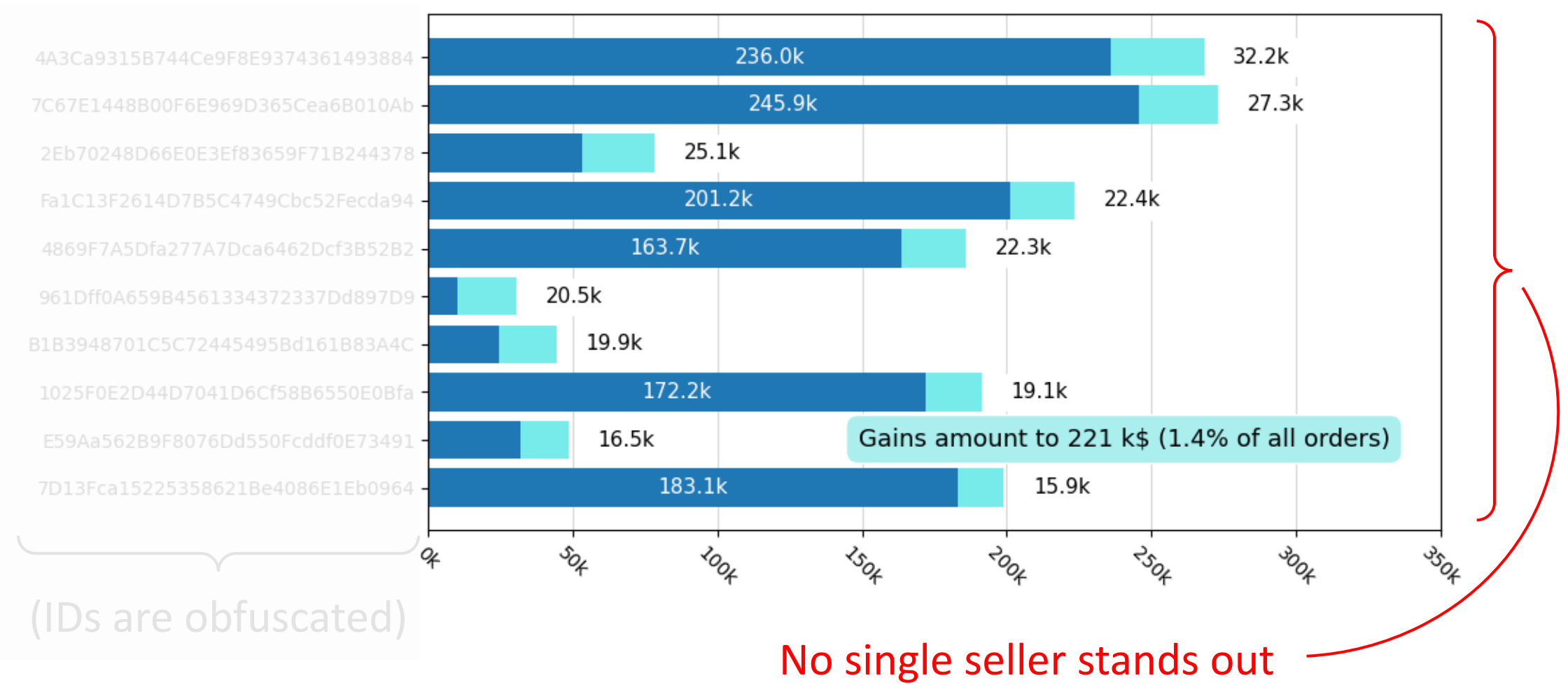
55 sellers are 0% reliable.

Most value is from top sellers

95% of value is from sellers with >80% reliability.

This is acceptable, and could be improved further.

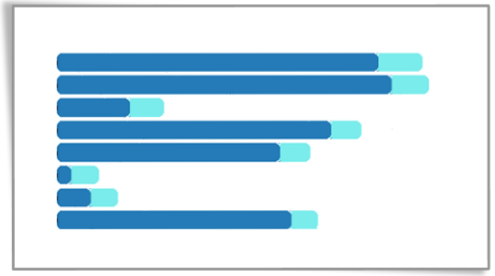
Potential gain by seller



Therefore, improvements to seller policies would be global.

Overall, financial health is positive, and...

Delivery reliability can be improved as follows.

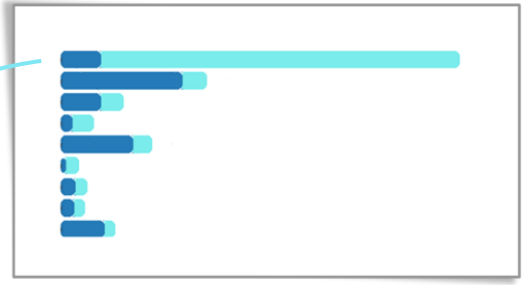


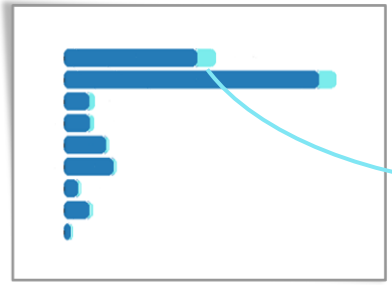
Ensuring seller policy changes are global.

Reason: Potential gains are even across sellers.

Addressing top products on a case-by-case basis.

Reason: Product potential is concentrated (item “FD00...”, 200k$).
Note that considering more reliable vendors is an option.





Prioritizing Rio de Janeiro for pilot programs.

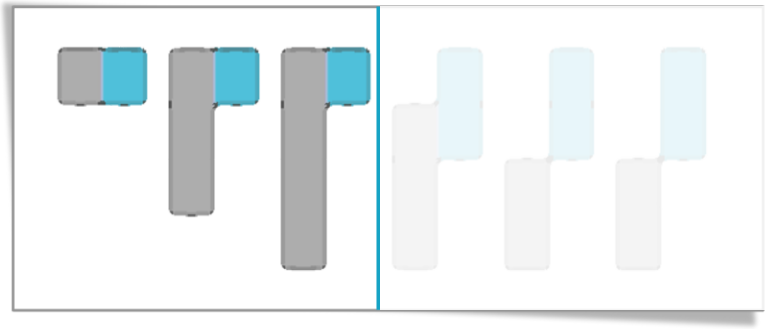
Reason: It is #1 in gain potential (150k$) and “only” 88% reliable.

(Consider automating alerts for items with abnormally high gain potentials.)

Overall, financial health is positive, and...

Data-backed **underpromising policies** could also improve ratings.

Metric 1: **Timely** delivery durations can be capped at **3 weeks**.



Metric 2: Should a delivery be **delayed**, customers seem to have a tolerance of **3 days** before severely lowering ratings.

(Realistically, transparency, competitiveness and urgency should be incorporated first.)

Overall, financial health is positive, and...

Database corrections can save resources and analysis efforts.

		Current state	Issue	Suggestion

Table	`Orders`
	`Order_items`
	`Order_reviews`
		These tables misalign in order information (order ID sets differ)	~800 to 2000 orders may get neglected (~0.5 M$ value)	Propagate updates, apply PRIMARY and FOREIGN convention

	`Order_items`	Each item sold writes a new row	Memory wasted, queries inefficient	Create a quantity column, refactor*
	`Reviews`	Order reviews only (no product data)	Hard to diagnose product quality	Create a table for product reviews**
	`Geolocation`	~900k duplicate locations saved	No added meaning, some memory waste	Remove duplicates, match data in Orders

** This saves memory when >2% of items are in quantities >1 (threshold from a 10-column SQL simulation through Qwen3-Coder-480B-A35B-Instruct).*

*** Should product-based diagnosis be valuable to OList.*

Summary

Status, actions, and future directions

Overall, financial standing is sound (low urgency, low risk).

Short-term:

- **Avoid 0% reliable sellers, products, and cities.**
  Negligible losses, but these likely affect data management and logistics.
- **Deliver product "FD00..." reliably (full potential of 200 k$).**
  Despite demand, delivery reliability for this product is abnormally low.

Long-term:

General	Create alerts for abnormal losses. Optimize underpromising to improve customer sentiment.
Sellers	Default policy improvements should be impartial. Monitor exceptions.
Products	Improvements should be targeted; monitor low performance. Deploy product reviews.
Cities	Rio de Janeiro a top candidate for profitable reliability improvement programs.
Database	Synchronize order data across tables. Reduce repeated entries (track item quantity).

References

Data

- Kaggle (OList eCommerce)
  Data was downloaded from kagglehub.dataset_download("olistbr/brazilian-ecommerce").

Project idea

- NPower Canada (Junior Data Analyst program)
  The analysis was adapted from questions presented during a workshop.