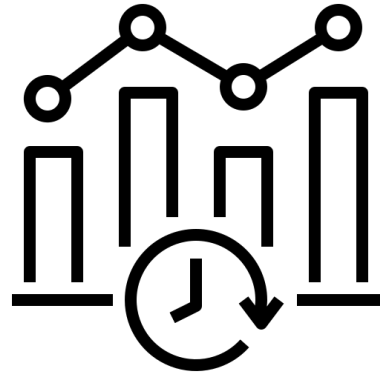


# Understanding Real-time dataset: Back Correction Analysis

---

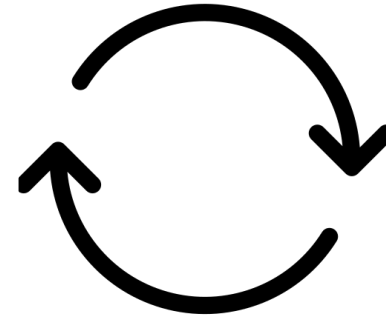
Nguyen Huong, Tran  
13/05/2024  
MAIS, ERSD, WTO

# Contents



## 1. Understanding real-time data:

- Data preparation
- Data overview
- Back correction analysis



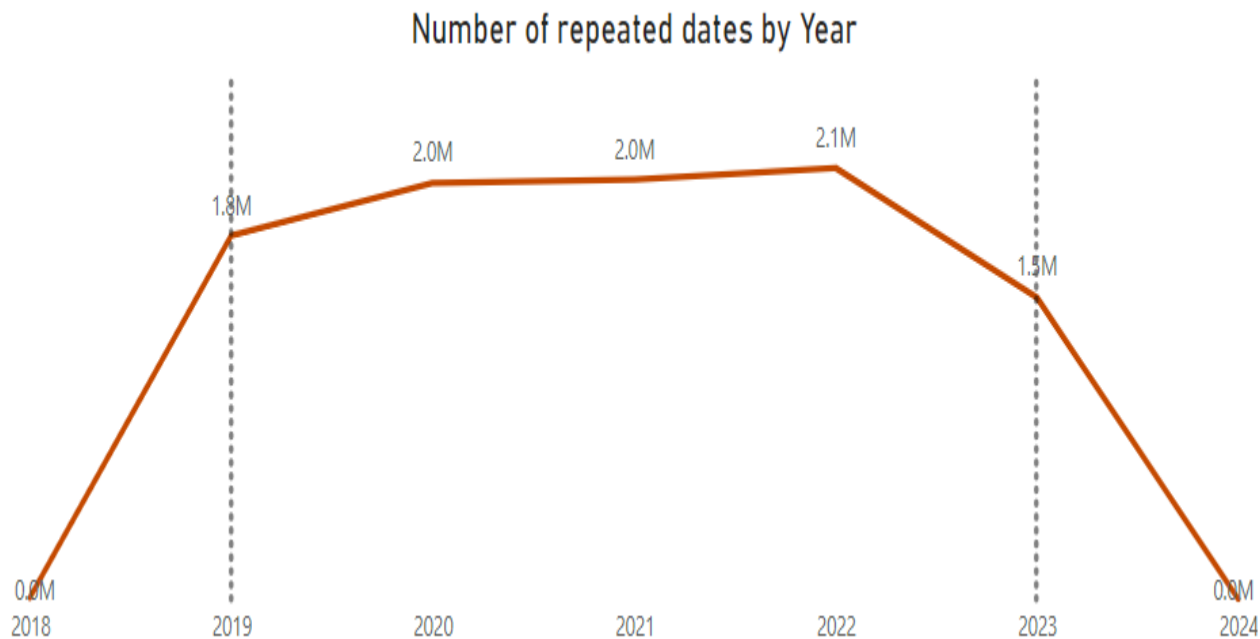
## 2. Daily back correction dashboard

- Architecture Diagram
- Real-time dashboard

# Data Collection

- Source: Azure data lake, [wtomais| Containers| kpler-data](#)
- Snapshot data collected from 12/06/2023 to 29/01/2024 (232 days)
  - Captures historical data from 01/09/2018 to 29/01/2024 across all data
    - For example: The snapshot taken on 12/06/2023 captures cargo shipments from 01/09/2018 to 12/06/2023. However, it may not cover all dates within this range, as there could be missing dates with no transactions recorded.
  - Annual calculations will be restricted from 2019 to end 2022

# Data Collection



- The number of repeated dates from 2019 to 2022 shows a consistent total of repetitions. Notably, there are fewer recorded values in 2018, 2023, and 2024.

→ This analysis covers the period from 2019 to end 2022 (Including replicated values in order to further examine the window where back corrections occur.)

# Selected Variables

Variables	Description
Product	Type of a product
Id (Trade)	Recorded trading transactions
Trade status	Status of a transaction: <ul style="list-style-type: none"><li>• Scheduled</li><li>• Loading</li><li>• In transit</li><li>• Delivered</li></ul>
Cargo (tons)	Shipment Capacity (tons) loaded/discharged at the terminal
Start (origin)	Start date of the trip
End (origin)	End of the loading process at the origin

Variables	Description
Installation origin	Name of the original installation
Installation Destination	Name of the destination installation
Country (origin)	Name of the original country
Country (destination)	Name of the destination country

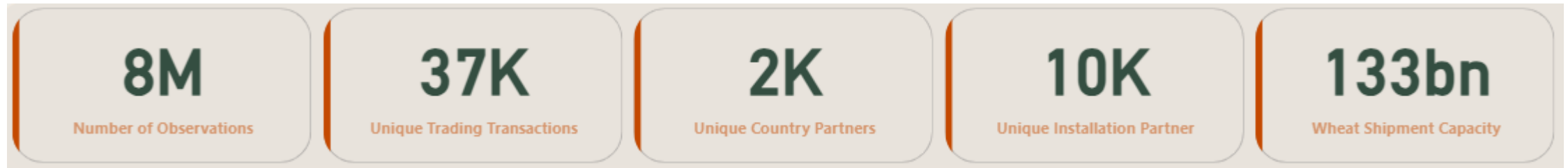
# Operationalising Variables

Variables	Description
Tracking date	Snapshot date
Time lag tracking date	This variable calculates the time difference (in days) between each subsequent occurrence (tracking date) of the same trade ID and the <b>first</b> occurrence of that trade ID.
Back correction	The percentage difference in cargo volume relative to the initial cargo volume observed in each subsequent instance of the trade ID.
Back correction groups	Depending on the back correction value, we categorize them into various groups, including: <ul style="list-style-type: none"><li>• Decrease by more than 75%,</li><li>• Decrease by 50% to 75%,</li><li>• Decrease by up to 50%,</li><li>• Increase by up to 50%,</li><li>• Increase by 50% to 75%,</li><li>• Increase by 75% to 100%,</li><li>• Increase by 100% to 500%,</li><li>• Increase by more than 500%</li></ul>

Variables	Description
Installation partner	Merging the names of the original and destination installations.
Country partner	Merging the names of the original and destination countries.

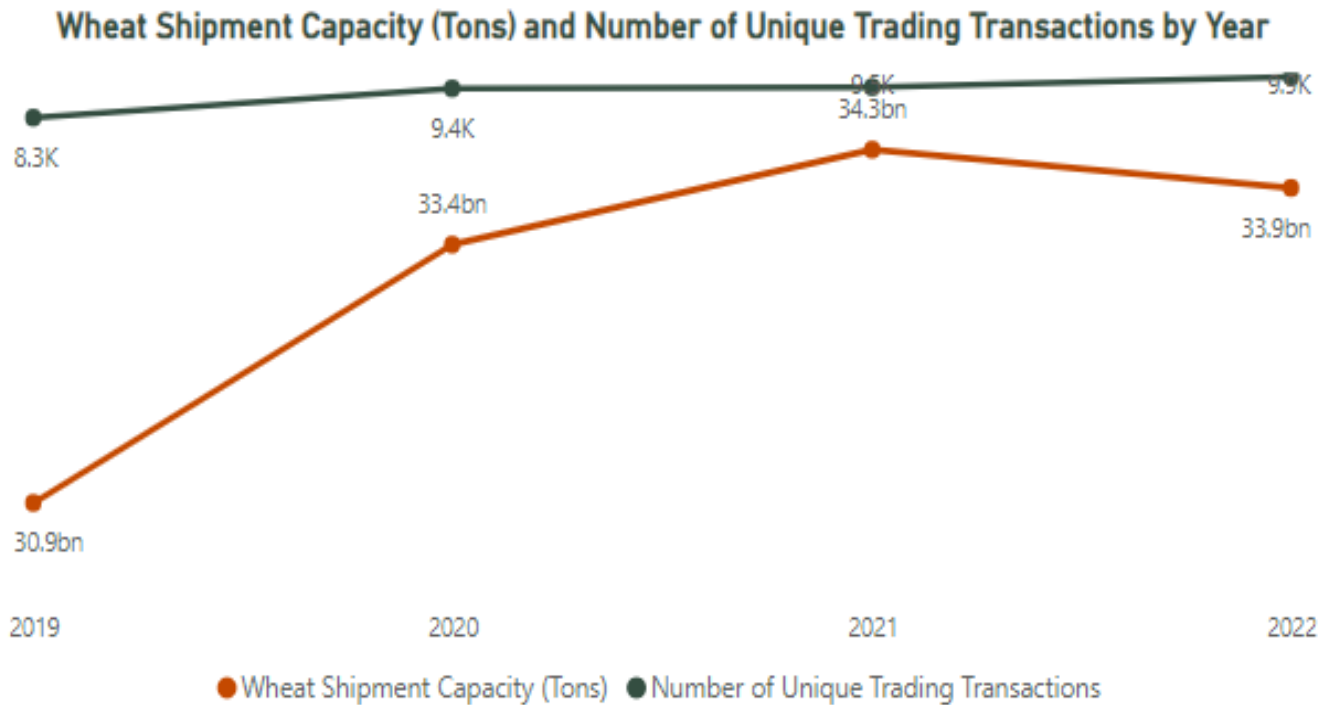
- When discussing trading, it's natural to consider trading partners. Therefore, it makes sense to merge the data of loading and discharging countries and ports

# Data Overview (1)



- The wheat Shipment Capacity does not accurately reflect the true volume as it includes replicated loading wheat volumes. However, it still serves as a valuable metric, providing useful insights into overall wheat trade dynamics.
- On average,
  - At least 159 unique trading transactions per day,
  - Every trading country partner maintains a minimum of 5 installation pairs,
  - Each installation pair entails a minimum of 3 transactions.

## Data Overview (2)

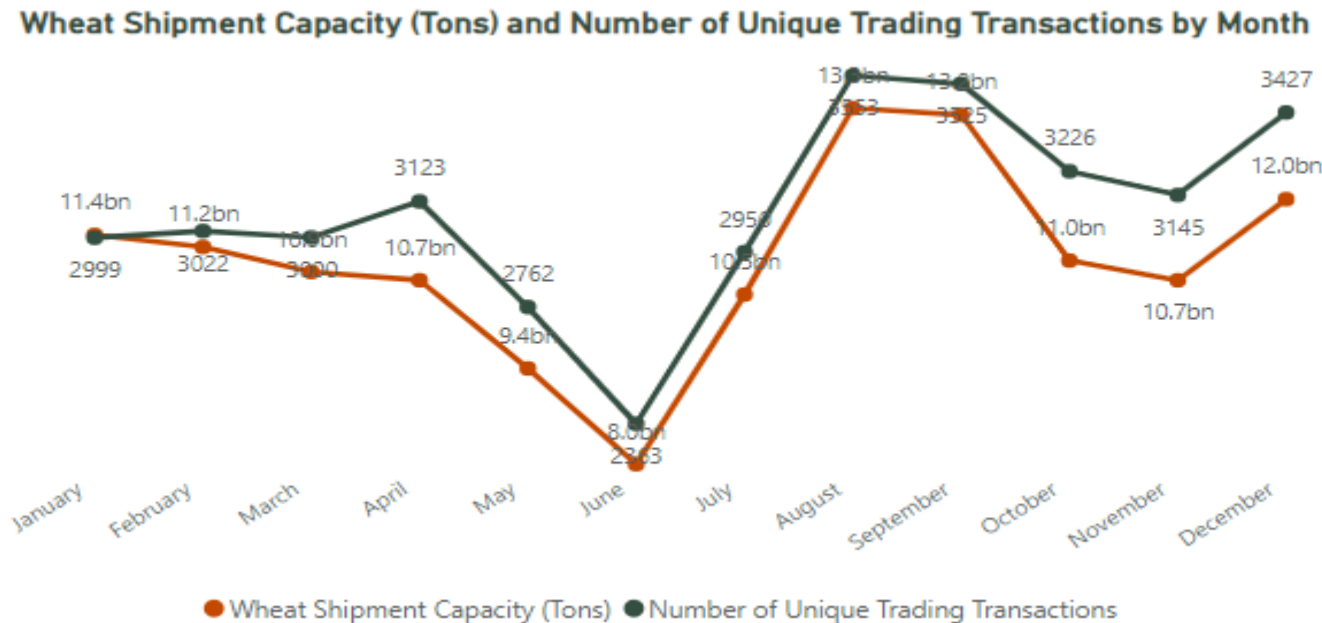


- The Wheat Shipment Capacity, measured in tons, is expected to show an upward trend as it accumulates all replicated values. However, there was a decline from 34.3 billion to 33.9 billion (by 1.18%), suggesting that a downward adjustments were made by the end of 2022

→ Suggesting the variable 'Number of Unique Trading Transactions' which excludes duplicate values for further analysis



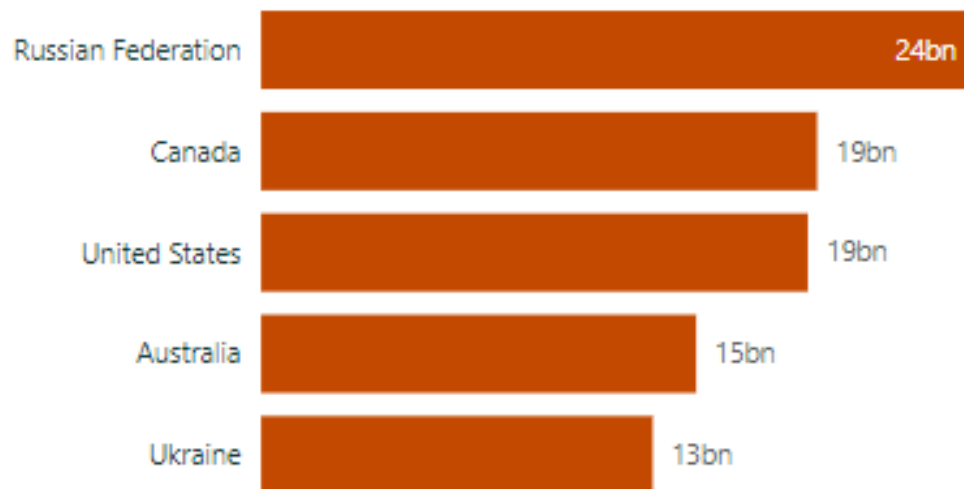
# Data Overview (3)



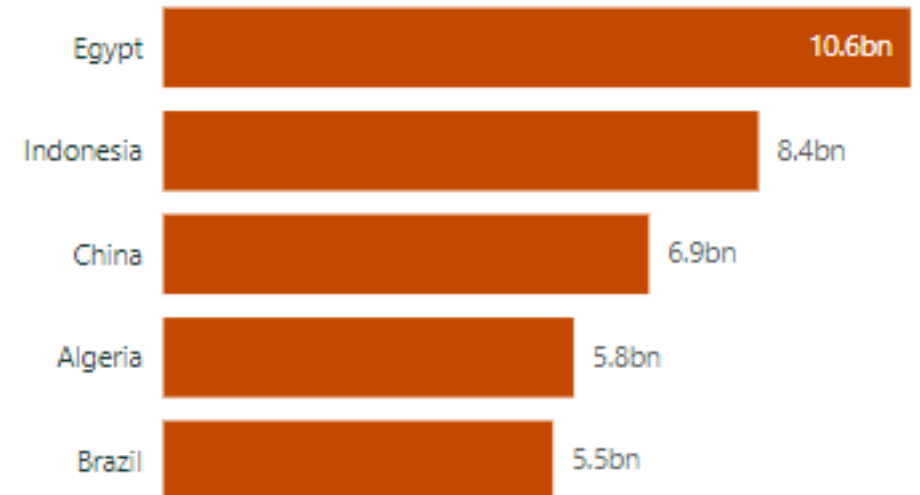
- Both economic indicators: the Wheat Shipment Capacity (in tons) and the Number of Unique Trading Transactions— demonstrate a sharp decline in wheat loading volumes in June, reaching their peak in August. This seasonal pattern has consistently been observed.

# Data Overview (4)

**Top 5 Leading Loading Countries**

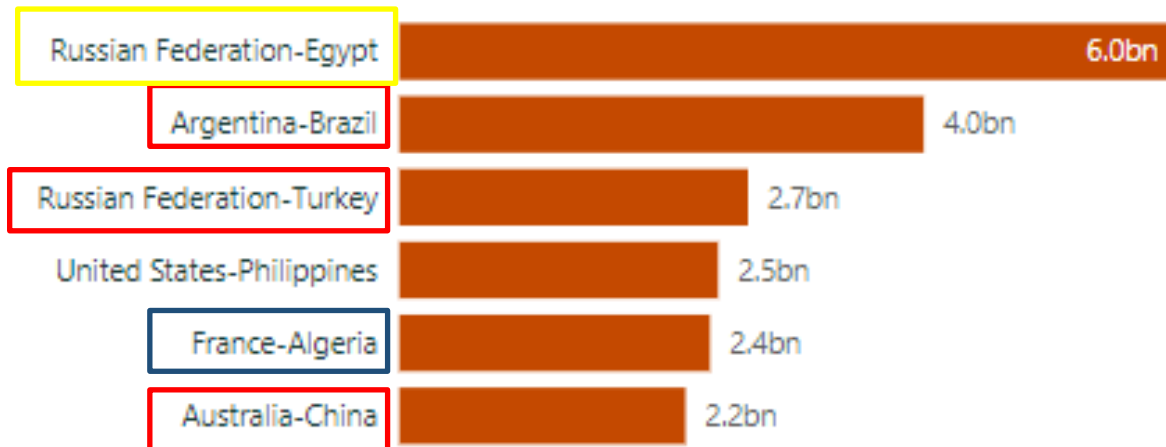


**Top 5 Leading Discharging Countries**

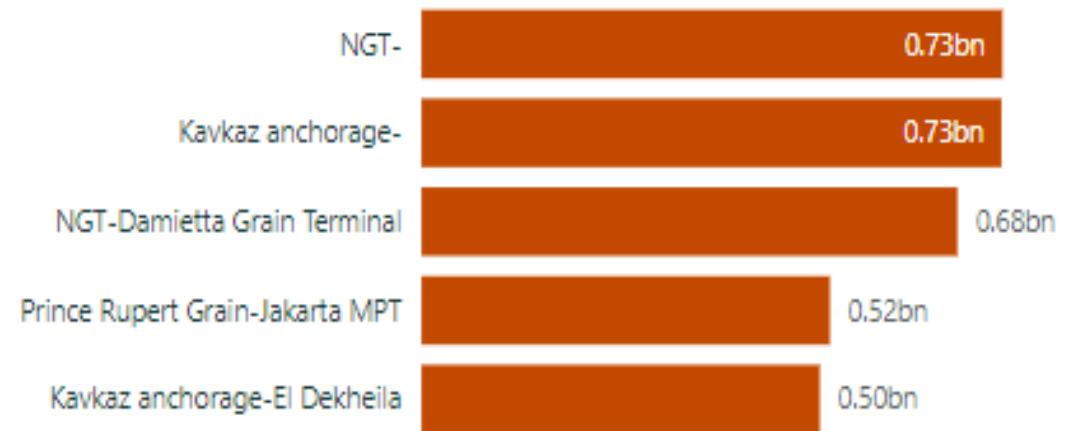


# Data Overview (5)

## Top 5 Trading Country Partners



## Top 5 Installation Partners





## Key takeaways (1)

- Downward Adjustments in 2022: There were reductions made to previously recorded data by the end of 2022.
- Top Wheat-Loading Countries: Russia, Canada, the United States, Australia, and Ukraine
- Top Wheat-Discharging Countries: Egypt, Indonesia, China, Algeria, and Brazil are key importers. High population densities and limited farming land in these developing nations heighten their dependency on imported wheat for food security.



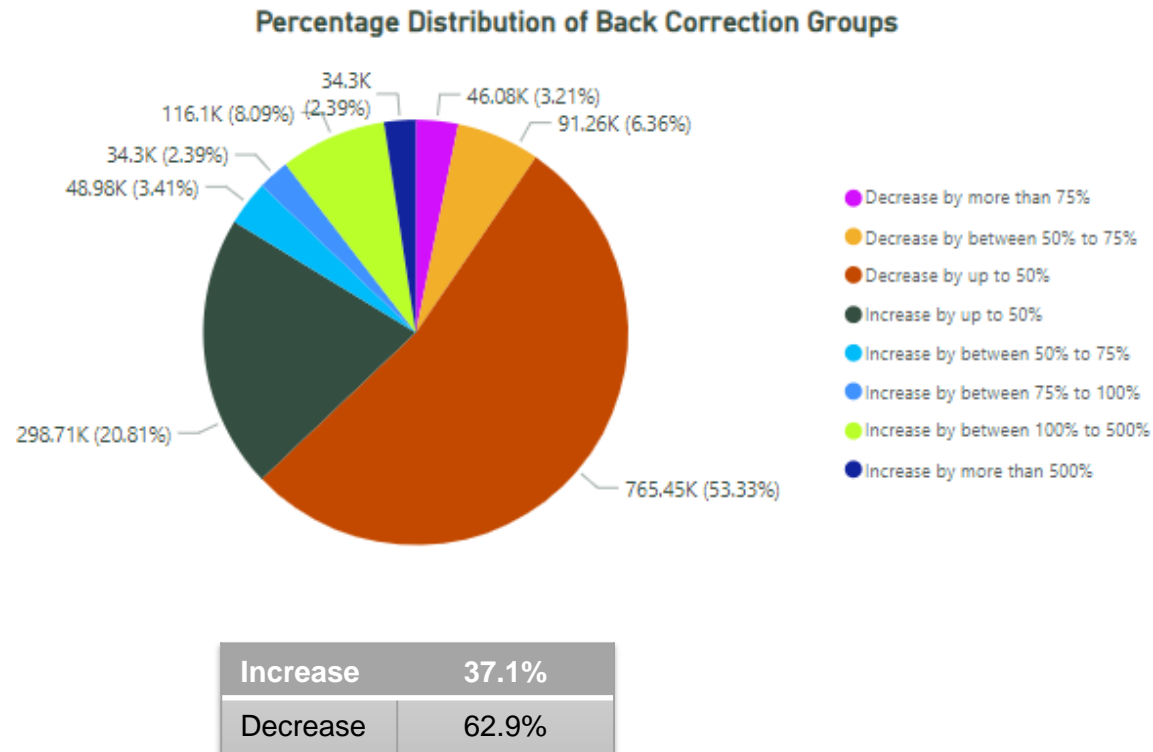
## Key takeaways (2)

- Regional Trading Partnerships: Neighboring countries like Argentina and Brazil, Russia and Turkey, and Australia and China frequently engage in wheat trade due to geographical proximity.
- Algeria's Historical Trade Links: As a former French colony, Algeria continues to receive significant wheat imports from France.
- Shift in Egypt's Wheat Imports: Prior to 2022, Ukraine was a major supplier for Egypt. Post-2022, Egypt has increasingly relied on wheat from Russia, with Russia replacing Ukraine as Egypt's primary wheat supplier. ([see more in slide 27](#))

# Back Correction Analysis (BCA) Definition

- **Back correction** refers to the adjustment made to **initially recorded data** to account for changes or errors discovered after the initial recording. In our study, we talk about **retrospective correction** involving adjusting data after it has been recorded and processed.
  - **Back correction magnitude** refers to the extent or degree of adjustment made during the process of back correction.
  - **Correction window** implies the duration during which revisions or adjustments are made to initially recorded data to ensure accuracy. The length of this window is determined by the **snapshot date**.
  - The "**first recorded value**" refers to the initial wheat shipment volume recorded for each trading transaction (Id trade). In this context, the first recorded values for each transaction are the cargo volume (in tons) documented on **12/06/2023**.

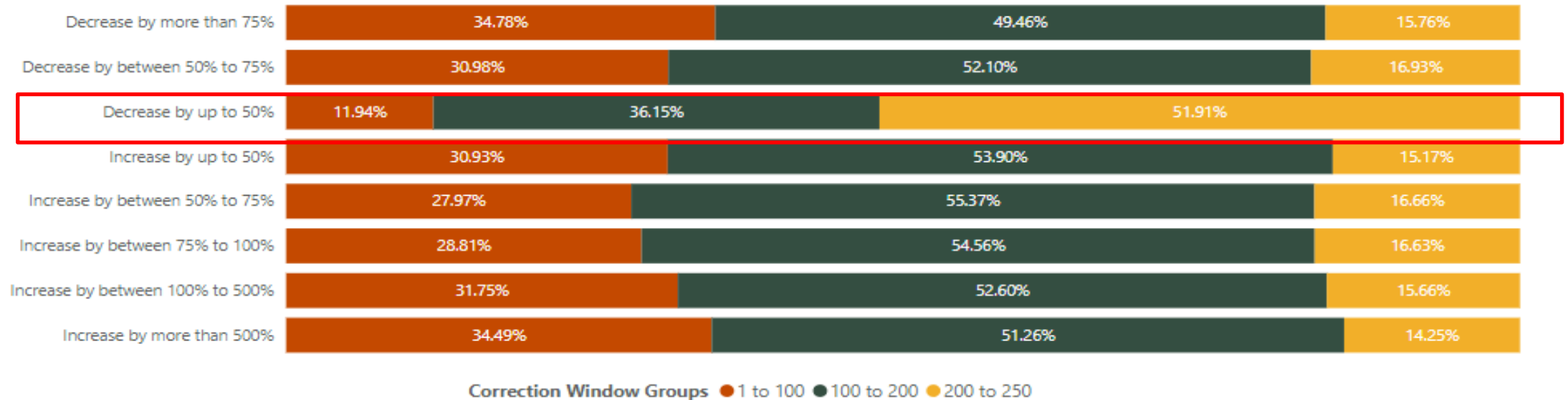
# Back Correction Magnitude



Back Correction Groups	Percentage Distribution (%)
Decrease by more than 75%	3.21
Decrease by between 50% to 75%	6.36
Decrease by up to 50%	53.33
Increase by up to 50%	20.81
Increase by between 50% to 75%	3.41
Increase by between 75% to 100%	2.39
Increase by between 100% to 500%	8.09
Increase by more than 500%	2.39

# Correction Window

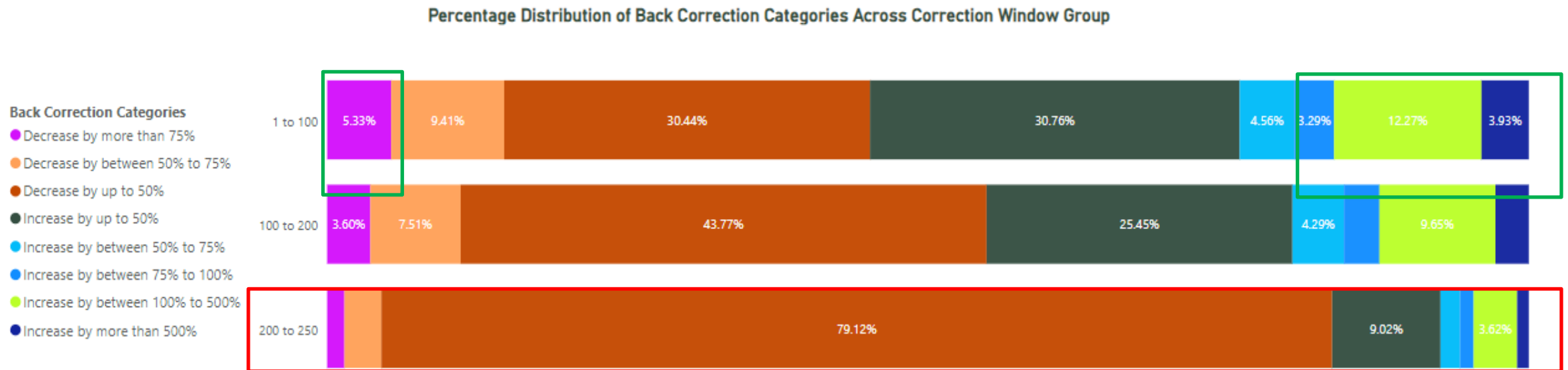
Percentage Distribution of Correction Window Groups Across Back Correction Categories



- In most back correction categories, approximately 50% of adjustments occurred between 100 to 200 days.
- In contrast, within the dominant groups where countries reduced their initial values by up to 50%, the corrections took at least 200 days to be implemented.



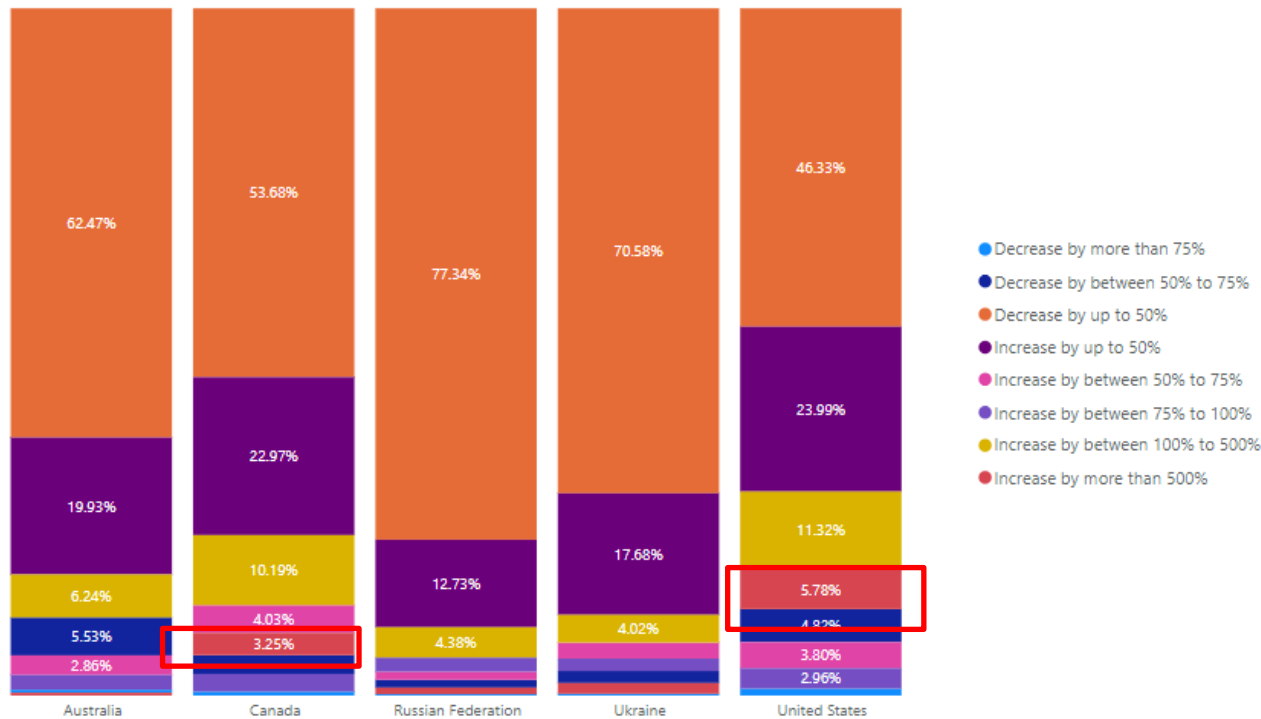
# BCA Overview (2)



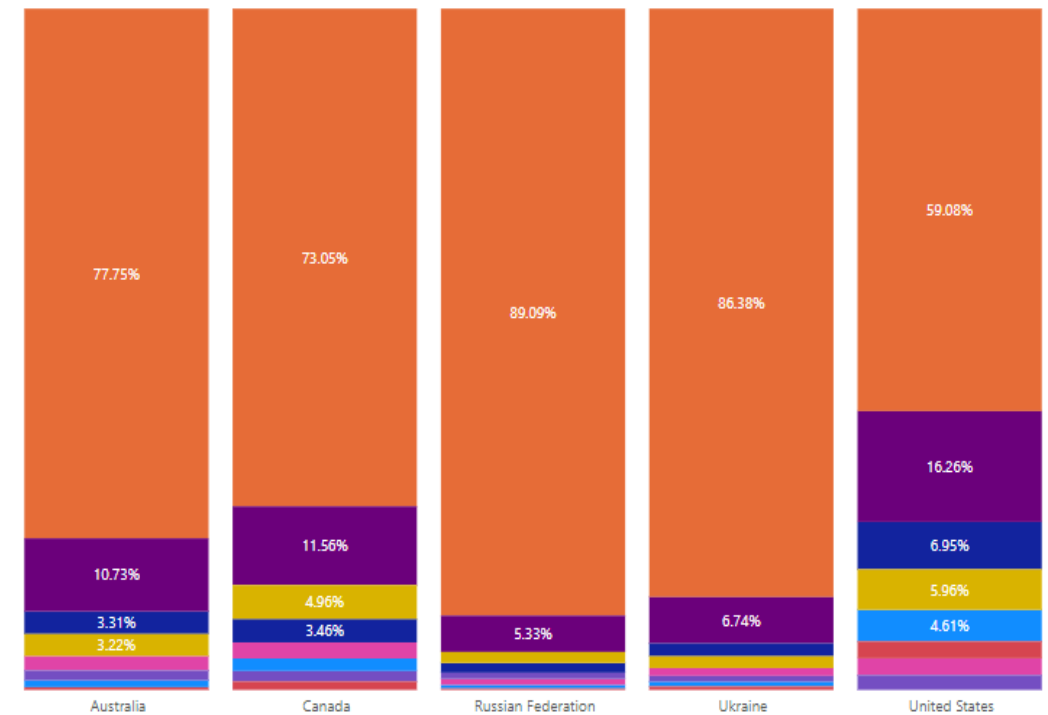
- Adjustments decreasing the original value by up to 50% are dominant in all three correction window categories. Notably, approximately 80% of this group falls within the 200 to 250-day correction window
- The percentage of back corrections exceeding 75% is observed within the 1 to 100 days group, indicating that major adjustments typically take place within the first 100 days.

# BCA Country Level (1)

Top 5 Wheat-Loading Countries by Shipment Capacity, Grouped by Back Correction Categories

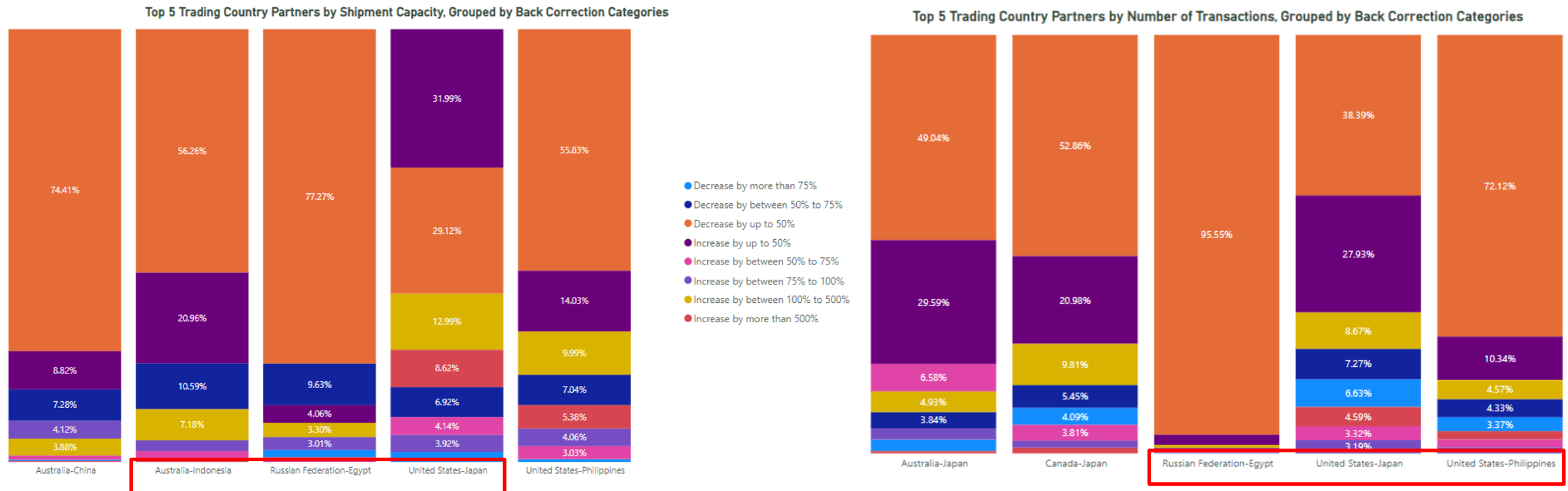


Top 5 Wheat-Loading Countries by Number of Transactions, Grouped by Back Correction Categories



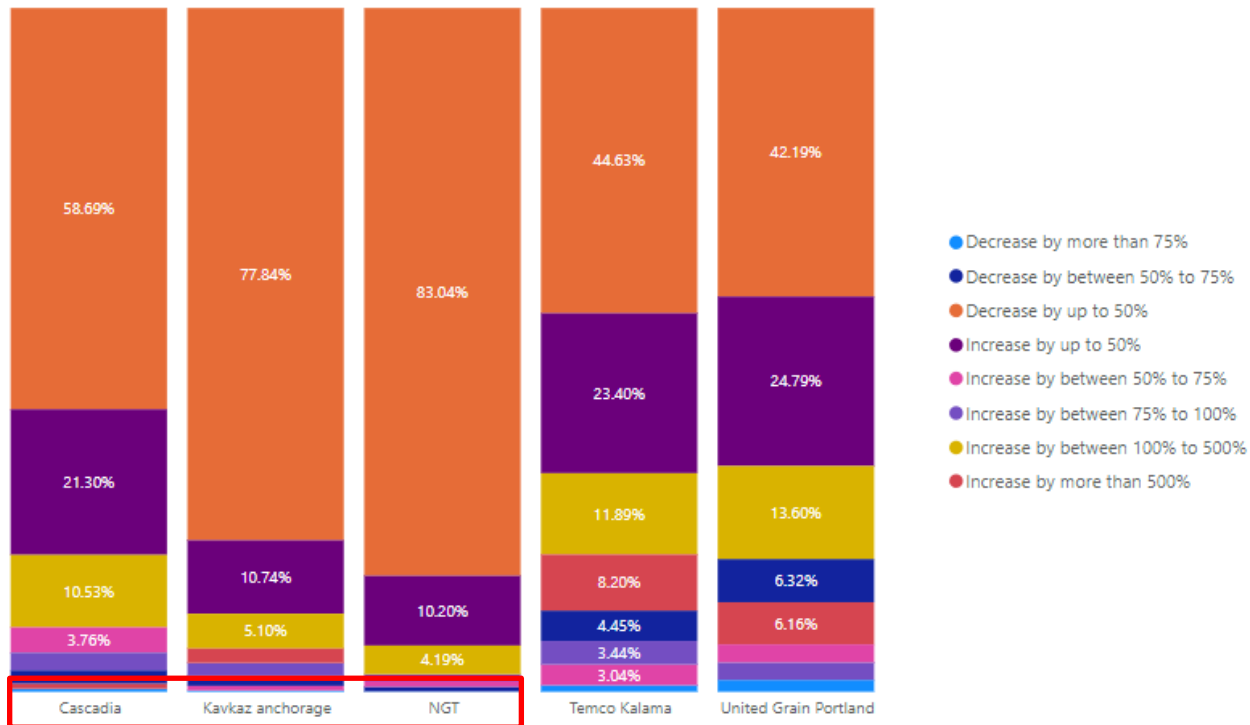
The major players that make adjustments are **the USA, Canada, Russia, Australia, and Ukraine**, which are also the leading wheat-loading countries. In cases where corrections result in a rise by more than 500%, **the USA** (contributing over 50%) and **Canada** (approximately 26%) are the primary contributors.

# BCA Country Pair Level

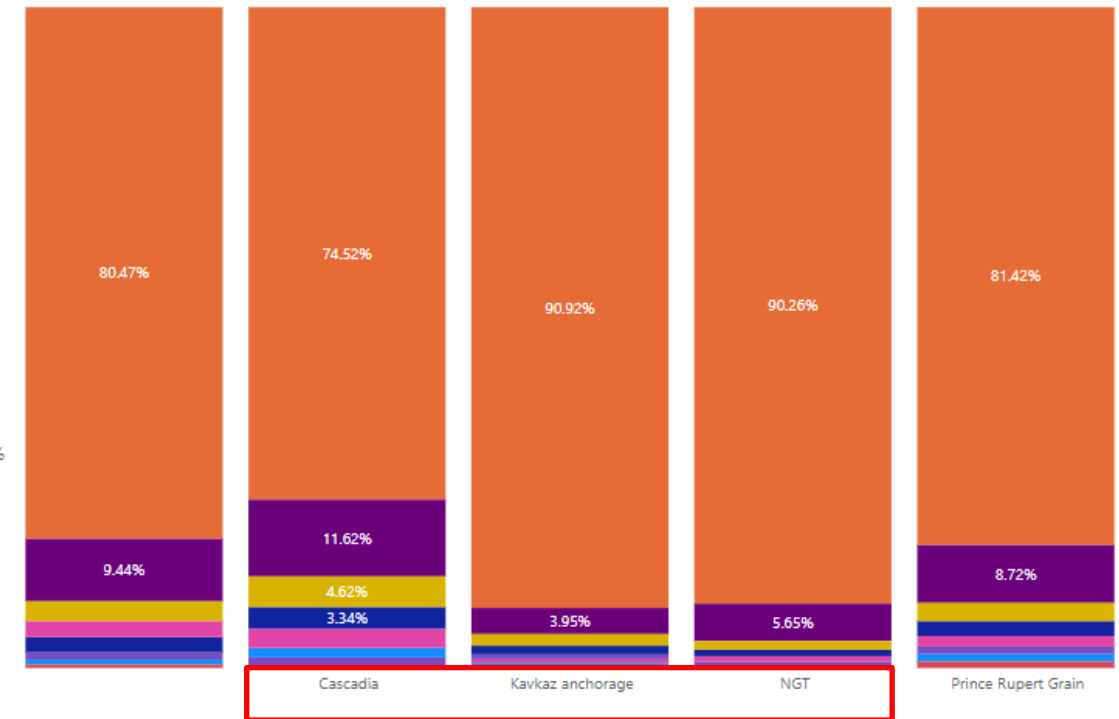


# BCA Port Level

Top 5 Loading Ports by Shipment Capacity, Grouped by Back Correction Categories



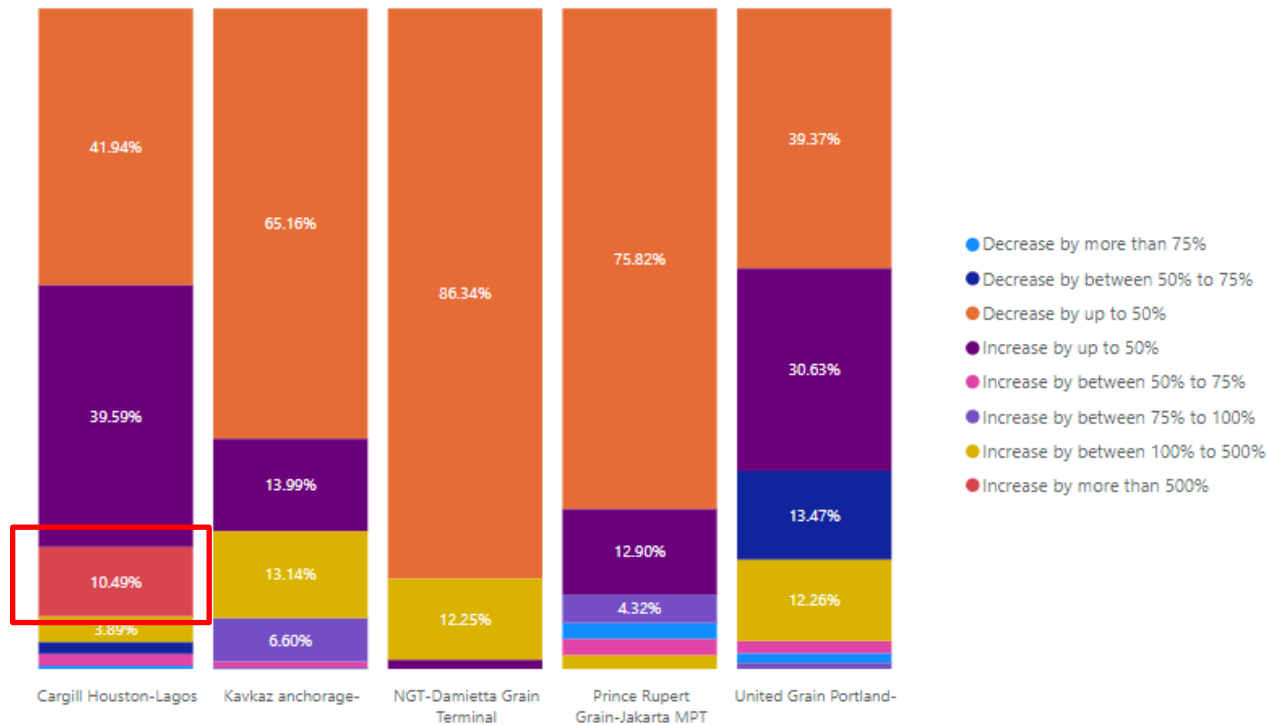
Top 5 Loading Ports by Number of Transactions, Grouped by Back Correction Categories



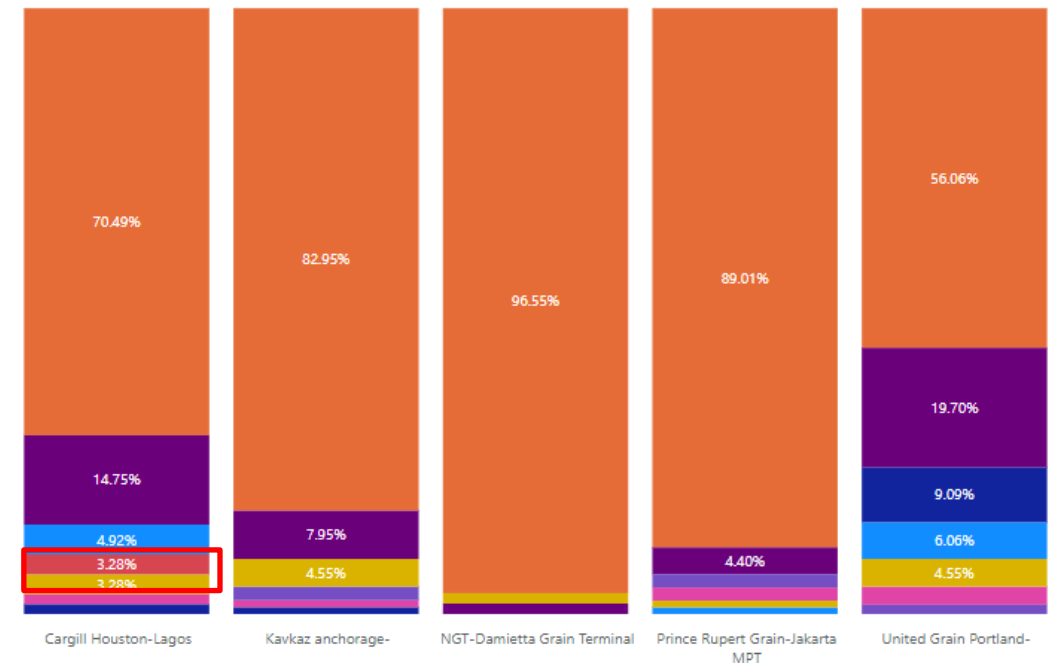
Both economic metrics show that the loading port making adjustments include Cascadia, Kavkaz anchorage and NGT

# BCA Port Pair Level

Top 5 Trading Port Partners by Shipment Capacity, Grouped by Back Correction Categories



Top 5 Trading Port Partners by Number of Transactions, Grouped by Back Correction Categories



- The trading port partners making adjustments include Cargill Houston-Lagos, Kavkaz Anchorage, NGT-Damietta Grain Terminal, Prince Rupert Grain-Jakarta MPT, and United Grain Portland.
- 100% of the adjustments that resulted in an increase of more than 500% originated from Cargill Houston-Lagos.<sup>21</sup>

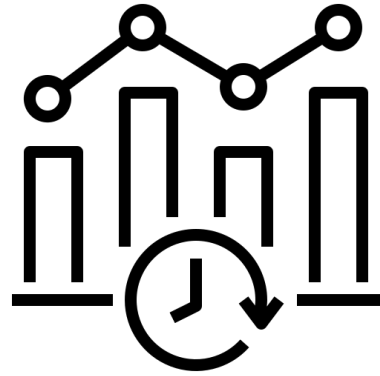
# Summary (1)

- The majority of corrections result in a decrease, with 62.9% of values decreasing from their initially recorded values. When back correction occurs, 53.33% of the corrections result in a decrease of up to 50%  
→ True wheat loading amount tends to vary, with **a tendency for decreases of up to 50%.**
- Within all back correction groups, the process typically occurs within 200 days, with the majority of adjustments taking at least 100 days to complete. However, in the case of decreases by up to 50%, adjustments take at least 200 days to complete.  
→ Waiting for **at least 200 days** before considering recorded values as the most accurate

## Summary (2)

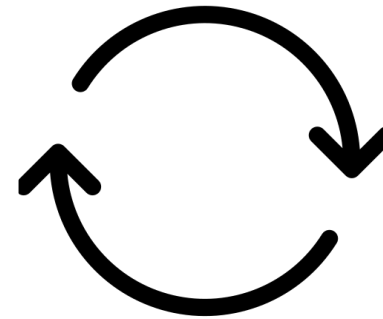
- The major players that make adjustments are **the USA, Canada, Russia, Australia, and Ukraine**, in that order which are also the leading wheat-loading countries. In cases where corrections result in a rise by more than 500%, **the USA** (contributing over 50%) and **Canada** (approximately 26%) are the primary contributors.
- 100% of the adjustments that resulted in an increase of more than 500% originated from **Cargill Houston-Lagos**.
- The major installations that make adjustments are the **Kavkaz anchorage, Cascadia and NGT**.
  - ➔ Conducting further research or analysis to identify the underlying drivers such as transportation routes, storage capacity, and infrastructure capabilities, which can impact trading strategies and market dynamics.

# Contents



## 1. Understanding real-time data:

- Data preparation
- Data overview
- Back correction analysis

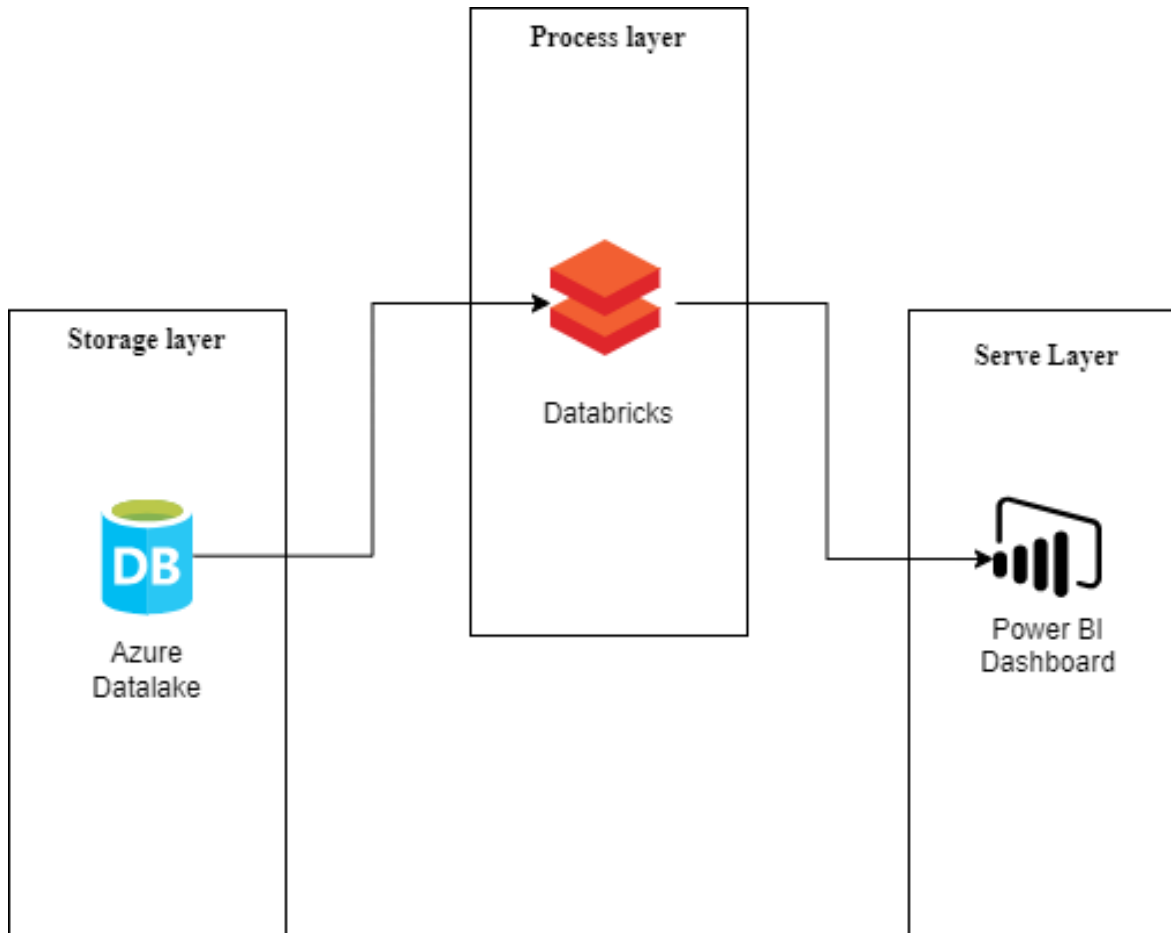


## 2. Daily back correction dashboard

- Architecture Diagram
- Real-time dashboard



# Architecture Diagram



- **Data Storage in Azure Data Lake:** Data is stored under the directory `wtomais|Containers|kpler-data` in Azure Data Lake.
- **Data Processing with Databricks:** Once the data is securely stored, the next step is to process this raw data to make it suitable for analysis. This processing is performed in Databricks.
- **Real-Time Data Visualization with Power BI:** After processing the data in Databricks, the next phase is visualization. We establish a direct connection from Databricks to Power BI.

# Daily back correction dashboard

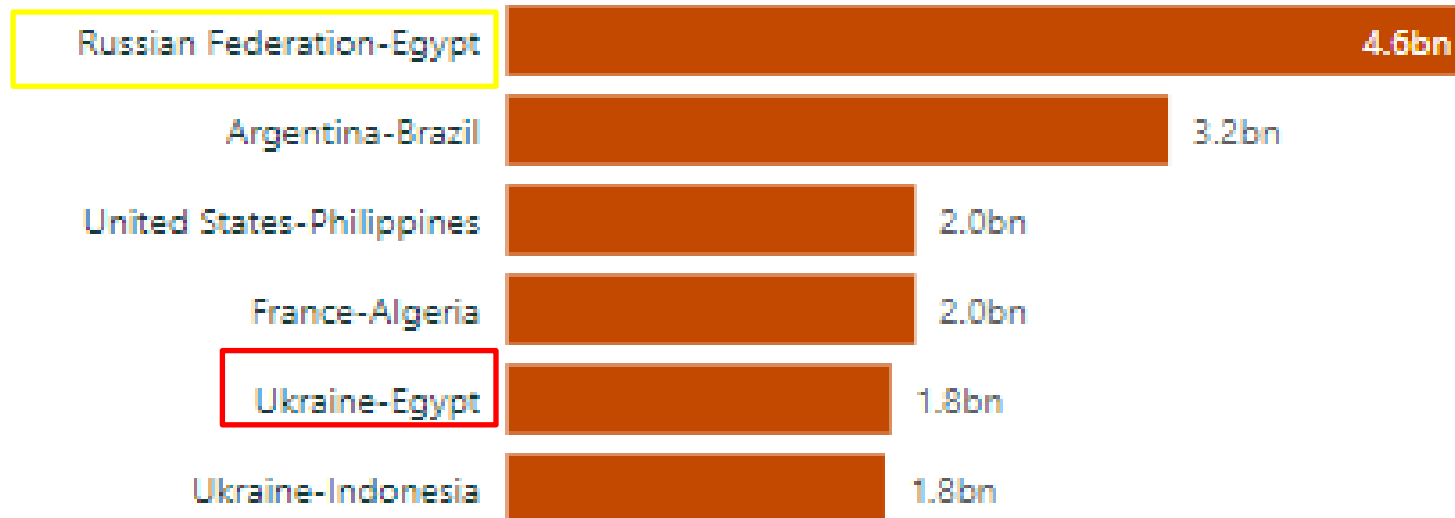
## **Daily back correction** dashboard:

- Displaying the range of the adjustment,
- The percentage distribution of back correction categories,
- The percentage distribution of trade status,
- Identifying countries adjusting,
- Identifying installations adjusting.

[Real time dashboard \(click here\)](#)

# Top 5 Trading Country Partners before 2022

Top 5 Trading Country Partners before 2022



# BCA Country Level (2)

Back Correction Categories	Shipment Capacity	Loading Country
Decrease by more than 75%	49416400	United States
Decrease by more than 75%	21674697	Canada
Decrease by more than 75%	12156970	Australia
Decrease by more than 75%	7015052	Russian Federation
Decrease by more than 75%	4230920	Ukraine
Decrease by more than 75%	3107292	Colombia
Decrease by more than 75%	3091839	Argentina
Decrease by more than 75%	1805141	Lithuania
Decrease by more than 75%	1771480	France
Decrease by more than 75%	1614645	Brazil
Decrease by more than 75%	1148313	Cameroon
Decrease by more than 75%	288337	Romania
Decrease by more than 75%	132660	India
Decrease by more than 75%	49728	Estonia
Decrease by more than 75%	36320	Germany
Decrease by more than 75%	34542	Poland
Decrease by more than 75%	6860	Chile
Decrease by more than 75%	3519	Spain
Decrease by more than 75%	3360	
Decrease by more than 75%	277	Peru

Back Correction Categories	Shipment Capacity	Loading Country
Decrease by between 50% to 75%	236528207	United States
Decrease by between 50% to 75%	153965689	Australia
Decrease by between 50% to 75%	103141300	Canada
Decrease by between 50% to 75%	49776509	Argentina
Decrease by between 50% to 75%	39074730	Russian Federation
Decrease by between 50% to 75%	28901585	Ukraine
Decrease by between 50% to 75%	23743786	France
Decrease by between 50% to 75%	8922785	Poland
Decrease by between 50% to 75%	8165521	Lithuania
Decrease by between 50% to 75%	6639743	Latvia
Decrease by between 50% to 75%	5069456	Estonia
Decrease by between 50% to 75%	3882834	Brazil
Decrease by between 50% to 75%	3791060	Chile
Decrease by between 50% to 75%	3501804	Germany
Decrease by between 50% to 75%	3268350	Uruguay
Decrease by between 50% to 75%	3135630	Cameroon
Decrease by between 50% to 75%	2836417	Romania
Decrease by between 50% to 75%	1111070	Bulgaria
Decrease by between 50% to 75%	457905	Mexico
Decrease by between 50% to 75%	400191	China

Back Correction Categories	Shipment Capacity	Loading Country
Decrease by up to 50%	2641425077	Russian Federation
Decrease by up to 50%	2275330681	United States
Decrease by up to 50%	2024929965	Canada
Decrease by up to 50%	1740067365	Australia
Decrease by up to 50%	1164753051	Ukraine
Decrease by up to 50%	933685487	Argentina
Decrease by up to 50%	469397626	Romania
Decrease by up to 50%	452926063	France
Decrease by up to 50%	272265455	Lithuania
Decrease by up to 50%	259114378	Germany
Decrease by up to 50%	237242299	Latvia
Decrease by up to 50%	215003482	Poland
Decrease by up to 50%	213205994	Brazil
Decrease by up to 50%	197652407	Bulgaria
Decrease by up to 50%	164495369	India
Decrease by up to 50%	36728580	Spain
Decrease by up to 50%	27481663	Uruguay
Decrease by up to 50%	26417540	Colombia
Decrease by up to 50%	25966752	Estonia
Decrease by up to 50%	22563968	South Korea

# BCA Country Level (3)

Back Correction Categories	Shipment Capacity	Loading Country
Increase by up to 50%	1178323716	United States
Increase by up to 50%	866692231	Canada
Increase by up to 50%	555070905	Australia
Increase by up to 50%	434860939	Russian Federation
Increase by up to 50%	387008001	Argentina
Increase by up to 50%	291801501	Ukraine
Increase by up to 50%	263208130	France
Increase by up to 50%	104678207	Brazil
Increase by up to 50%	98669407	Latvia
Increase by up to 50%	92898758	Lithuania
Increase by up to 50%	69153109	India
Increase by up to 50%	65543701	Poland
Increase by up to 50%	55003124	Germany
Increase by up to 50%	37670676	Romania
Increase by up to 50%	20964646	South Korea
Increase by up to 50%	19816216	Spain
Increase by up to 50%	19563328	Bulgaria
Increase by up to 50%	16802271	Mexico
Increase by up to 50%	16192919	Uruguay
Increase by up to 50%	7018489	Estonia

Back Correction Categories	Shipment Capacity	Loading Country
Increase by between 50% to 75%	186549362	United States
Increase by between 50% to 75%	152208817	Canada
Increase by between 50% to 75%	79556034	Australia
Increase by between 50% to 75%	72791550	Argentina
Increase by between 50% to 75%	40579097	Russian Federation
Increase by between 50% to 75%	38683592	Ukraine
Increase by between 50% to 75%	34309865	France
Increase by between 50% to 75%	19425479	Lithuania
Increase by between 50% to 75%	17516321	Poland
Increase by between 50% to 75%	15692534	Romania
Increase by between 50% to 75%	8957640	Latvia
Increase by between 50% to 75%	8911854	Germany
Increase by between 50% to 75%	6170853	Brazil
Increase by between 50% to 75%	5792172	India
Increase by between 50% to 75%	5414705	Bulgaria
Increase by between 50% to 75%	3826839	Cameroon
Increase by between 50% to 75%	3161116	Venezuela
Increase by between 50% to 75%	2486504	Spain
Increase by between 50% to 75%	2374893	Cape Verde
Increase by between 50% to 75%	1853795	Estonia

Back Correction Categories	Shipment Capacity	Loading Country
Increase by between 75% to 100%	145322191	United States
Increase by between 75% to 100%	96661589	Canada
Increase by between 75% to 100%	70672382	Russian Federation
Increase by between 75% to 100%	59393978	Australia
Increase by between 75% to 100%	29659261	Ukraine
Increase by between 75% to 100%	27906240	Argentina
Increase by between 75% to 100%	26922678	France
Increase by between 75% to 100%	20789681	Latvia
Increase by between 75% to 100%	19304410	Lithuania
Increase by between 75% to 100%	16829232	Poland
Increase by between 75% to 100%	9966191	Uruguay
Increase by between 75% to 100%	9383300	India
Increase by between 75% to 100%	8011698	Brazil
Increase by between 75% to 100%	8005524	Germany
Increase by between 75% to 100%	4253217	Bulgaria
Increase by between 75% to 100%	4067343	Nigeria
Increase by between 75% to 100%	3065043	Mexico
Increase by between 75% to 100%	2699560	Romania
Increase by between 75% to 100%	1929989	Spain
Increase by between 75% to 100%	1751049	Bangladesh

# BCA Country Level (4)

Back Correction Categories	Shipment Capacity	Loading Country
Increase by between 100% to 500%	555852736	United States
Increase by between 100% to 500%	384381794	Canada
Increase by between 100% to 500%	173898426	Australia
Increase by between 100% to 500%	149599318	Russian Federation
Increase by between 100% to 500%	149283006	Argentina
Increase by between 100% to 500%	101189380	France
Increase by between 100% to 500%	66416034	Ukraine
Increase by between 100% to 500%	46525146	Poland
Increase by between 100% to 500%	42218399	Latvia
Increase by between 100% to 500%	28726152	Lithuania
Increase by between 100% to 500%	27581830	Brazil
Increase by between 100% to 500%	18561254	Bulgaria
Increase by between 100% to 500%	13191647	Romania
Increase by between 100% to 500%	10093538	Spain
Increase by between 100% to 500%	8942660	Nigeria
Increase by between 100% to 500%	6719884	Germany
Increase by between 100% to 500%	6419396	India
Increase by between 100% to 500%	6153836	Mexico
Increase by between 100% to 500%	5708864	China
Increase by between 100% to 500%	5093830	Turkey

Back Correction Categories	Shipment Capacity	Loading Country
Increase by more than 500%	284046357	United States
Increase by more than 500%	122689855	Canada
Increase by more than 500%	43913106	Lithuania
Increase by more than 500%	32194394	Russian Federation
Increase by more than 500%	25770508	Ukraine
Increase by more than 500%	17261284	Argentina
Increase by more than 500%	11556765	Australia
Increase by more than 500%	10374427	Latvia
Increase by more than 500%	7476600	Poland
Increase by more than 500%	6894343	Romania
Increase by more than 500%	5636663	France
Increase by more than 500%	2115820	Bulgaria
Increase by more than 500%	2105825	India
Increase by more than 500%	800696	Brazil
Increase by more than 500%	489330	Cameroon

# BCA Country Level (4)

Back Correction Categories	Shipment Capacity	Loading Country
Increase by between 100% to 500%	555852736	United States
Increase by between 100% to 500%	384381794	Canada
Increase by between 100% to 500%	173898426	Australia
Increase by between 100% to 500%	149599318	Russian Federation
Increase by between 100% to 500%	149283006	Argentina
Increase by between 100% to 500%	101189380	France
Increase by between 100% to 500%	66416034	Ukraine
Increase by between 100% to 500%	46525146	Poland
Increase by between 100% to 500%	42218399	Latvia
Increase by between 100% to 500%	28726152	Lithuania
Increase by between 100% to 500%	27581830	Brazil
Increase by between 100% to 500%	18561254	Bulgaria
Increase by between 100% to 500%	13191647	Romania
Increase by between 100% to 500%	10093538	Spain
Increase by between 100% to 500%	8942660	Nigeria
Increase by between 100% to 500%	6719884	Germany
Increase by between 100% to 500%	6419396	India
Increase by between 100% to 500%	6153836	Mexico
Increase by between 100% to 500%	5708864	China
Increase by between 100% to 500%	5093830	Turkey

Back Correction Categories	Shipment Capacity	Loading Country
Increase by more than 500%	284046357	United States
Increase by more than 500%	122689855	Canada
Increase by more than 500%	43913106	Lithuania
Increase by more than 500%	32194394	Russian Federation
Increase by more than 500%	25770508	Ukraine
Increase by more than 500%	17261284	Argentina
Increase by more than 500%	11556765	Australia
Increase by more than 500%	10374427	Latvia
Increase by more than 500%	7476600	Poland
Increase by more than 500%	6894343	Romania
Increase by more than 500%	5636663	France
Increase by more than 500%	2115820	Bulgaria
Increase by more than 500%	2105825	India
Increase by more than 500%	800696	Brazil
Increase by more than 500%	489330	Cameroon