Case study on Cloudwalk dataset for the logistics team

Created and performed by Victor Reichenbach Requião – Data Analyst

Started 29/05/2023 18:36

Problem #1 - Identifying the problem:

1. What questions should start your analysis?

- How many records are there in the database?

42710 records ranging from 2023-02-28 to 2021-09-08

- What is the overall distribution of statuses in the database?

42220 orders delivered, 98.85% of the overall orders were successfully delivered.

445 orders returned, 1.04% of the overall orders were returned for some reason.

45 orders cancelled, 0.11% of the overall orders were cancelled for some reason.

- How many unique providers are there?

There are two unique providers; Provider 1 and Provider 2

Provider 1 made 8917 transactions as deliverer

- 8611 transactions were successfully delivered (96.56 %)
- 306 transactions were returned (3.44%)

Provider 2 made 33793 transactions as deliverer

- 33609 transactions were successfully delivered (99.45%)
- 139 transactions were returned (0.41%)
- 45 transactions were cancelled (0.14%)

- What is the geographical distribution of orders (states and cities)?

Sudeste (Southeast) region: (45.21% of all orders)

SP (São Paulo): 9,824 orders

MG (Minas Gerais): 5,834 orders

RJ (Rio de Janeiro): 2,747 orders

ES (Espírito Santo): 903 orders

Nordeste (Northeast) region: (22.68% of all orders)

BA (Bahia): 3,133 orders

CE (Ceará): 1,733 orders

PE (Pernambuco): 1,480 orders

PB (Paraíba): 794 orders

MA (Maranhão): 711 orders

AL (Alagoas): 532 orders

PI (Piaui): 482 orders

RN (Rio Grande do Norte): 452 orders

SE (Sergipe): 371 orders

Sul (South) region: (14.93% of all orders)

PR (Paraná): 3,101 orders

SC (Santa Catarina): 1,789 orders

RS (Rio Grande do Sul): 1,485 orders

Centro-Oeste (Central-West) region: (11.25% of all orders)

GO (Goiás): 2,426 orders

MT (Mato Grosso): 858 orders

DF (Distrito Federal): 800 orders

MS (Mato Grosso do Sul): 723 orders

Norte region: (5.93% of all orders)

PA (Pará): 1,229 orders

RO (Rondônia): 436 orders

AM (Amazonas): 328 orders

TO (Tocantins): 304 orders

AC (Acre): 105 orders

AP (Amapá): 73 orders

RR (Roraima): 57 orders

- What is the average shipment cost for each provider?

Average shipment cost – R$21,54

Average shipment cost for Provider 1 – R$23,40

Average shipment cost for Provider 2 – R$21,05

- How do the delivery times vary across regions?

Overall average delivery time considering sales order and delivery is 9.8 days

Overall average delivery time considering shipping order and delivery is 5.92 days

Average delivery time by region, considering shipping order and delivery

Norte - 11.36 days

Nordeste - 8.12 days

Centro-oeste - 6.12 days

Sul - 6.18 days

Sudeste 3.98 days

Average delivery time by region, considering sales order and delivery

Norte - 16 days

Nordeste - 13 days

Centro-Oeste - 9,6 days

Sul - 9,5 days

Sudeste - 7,5 days

- How does the On-Time Delivery (OTD) performance differ between providers?

Provider 1

- Delivered 8917 products
- 7388 were on time (82.85%)

Provider 2

- Delivered 33793 products
- 32840 were on time (97.18%)

OTD (Brazil)%

- On time 95%
- Delayed 5%

- Perhaps this delay value is a little lower, consequently 'On-time' deliveries are a higher percentage, since there were some deliveries returned or canceled, which generated a null value in the 'delivered_at' column

OTD per region is also excellent

North has 97.5% of On-time deliveries

Northeast has 98.1% of On-time deliveries

Center-East has 94.01% of On-time deliveries

South has the lowest percentage with 89.26% of On-time deliveries

Southeast has 95.88% of On-time deliveries

- Are there any noticeable trends or patterns in the data?

Provider #2 is cheaper, faster, and more precise courier for the products (in general).

Shipping costs significantly more and take significantly more time to the North(Norte) than any other region in Brazil.

Provider #1 is cheaper than Provider #2 for shipping products to North and Northeast (Norte e Nordeste) on an average.

Provider #2 is cheaper than Provider #1 in all other regions.

Provider #1 is faster on deliveries to North (Norte) on an average.

Provider #2 is faster on deliveries to any other region.

On average, provider #2 is cheaper, faster and have a OTD rate way higher than provider #1 but also, provider #2 made almost 4 times more deliveries than provider #1

OTD for provider #1 is 82.85%

OTD for provider #2 is 97.18%

The successfulness of deliveries is over 95% for both providers

South east (Sudeste) and North east (Nordeste) are the regions with more orders followed by south (Sul), center-east (Centro-Oeste) and north (Norte).

- Transit time: The average time taken for orders to move through the different stages (in-transit, local distribution, etc.).

The average time difference between a sale (sales_order_created_at) and a delivery order (device_order_created_at) is **3.39 days**

The average time difference between ordering delivery (device_order_created_at) and processing the delivery (processing_at) is **15 minutes and 25 seconds**

The average time difference between processing the delivery (processing_at) and transiting to local distribution (in_transit_to_local_distribution_at) is **5 hours and 7 minutes**

The average time difference between transit to local distribution (in_transit_to_local_distribution_at) and arrival at the local distributor (local_distribution_at) **is 3.18 days**

The average time difference between arriving at the local distributor (local_distribution_at) and leaving for delivery (in_transit_to_deliver_at) is **1.76 days**

The average difference between the time from departure to delivery (in_transit_to_deliver_at) and delivery time (delivered_at) is **1.3 days**

- Customer satisfaction: Collect feedback from customers to measure their satisfaction with the delivery process.

- Need more information on customer feedback to evaluate.

2. What does it mean to be more "secure"?

Being more "secure" in the context of logistics can mean multiple things:

- Ensuring the safety of the shipped products during transportation.

- Minimizing the risk of damage, loss, or theft during the delivery process.

- Maintaining accurate and reliable tracking and monitoring systems for shipments.

- Maybe considering how many orders were delivered, cancelled and returned and then evaluate why some of the orders were cancelled and returned.

- The data provided is not enough to draw conclusions about the security of the operation.

3. What are good KPIs to check for efficiency?

- OTD, shipment cost, shipment time, transit time and customer feedback.

4. Which data should we examine more closely and why?

- On-Time Delivery (OTD) and difference OTD: Understanding the percentage of orders delivered on time and the difference between the estimated and actual delivery dates can provide insights into delivery performance and potential areas for improvement.

- Delivery time: Analyzing the time taken to deliver orders can help identify bottlenecks in the process and optimize delivery routes/resources.

- Shipment cost: Examining the shipment cost can identify cost-saving opportunities, such as negotiating better rates with providers or optimizing delivery routes.

- Region: Analyzing delivery performance and customer satisfaction across different regions can help identify specific challenges or opportunities in each area.

5. Are we doing good? Are we doing bad?

- After performing extensive analysis on the data provided, the operation is doing very well compared to national competition. By maintaining an OTD performance for over 95%(that is a standard in the industry) as an overall and the shipping costs being on an average of around R$21,54 considering the value of the product and the complexity of it.

- Over 98% of the deliveries were successful and not returned or cancelled, the percentage of cancelled and returned products are very small and could be evaluated for better understanding of why those orders returned or were cancelled to overcome the bottlenecks and turn them into successful deliveries.

- Considering some research performed, in 2015 the average delivery time for e-commerce products were 12 days (https://www.terra.com.br/economia/direitos-do-consumidor/veja-a-media-para-entrega-de-produtos-de-compras-on-line,a112eb9a46b3b5ebbcdc54333534a9c6gpsbRCRD.html) in 2020 the average delivery time went to 21 days, excluding big e-commerce dealers like Amazon and Mercado Livre (https://www1.folha.uol.com.br/mercado/2020/05/volta-do-frete-gratis-nao-ajuda-a-atrair-clientes-para-comercio-online.shtml) knowing that the overall delivery time on an average for Cloudwalk products is around 9 days, that's another indicative that the logistics operation for the company is doing fine.

- Other than that, we could use customer feedback to develop more on the indicatives of good or bad performance considering what the customers expect

6. What are the strong and weak sides of our operation?

- From the logistical and data analysis perspective, the strong side of the operation are the organization e cost effective deliveries and the high rate of on time deliveries.

- We could use customer feedback on this part to evaluate more strong and weak sides of the operation.

- From my point of view the weak side of the operation is the time between the sale order and the delivery order that could take up to 4 days on average for the product to be shipped.

- Also, there are numbers on the dataset that clearly have some inconsistency with the actual data that could be cleaned and organized.

7. What information would improve the decision-making process?

- Detailed tracking data: Access to granular data on each stage of the delivery process can help point out bottlenecks and improve operations.

- Customer feedback: Obtaining more specific feedback from customers can highlight areas for improvement.

- Comparative data: Information against industry standards or competitors' performance can provide context and insights for decision-making.

- Historical data: Analyzing trends and patterns over time can help identify seasonality, areas for improvement, or changes in performance.

8. What information makes your analysis wrong?

Inaccurate or incomplete data can lead to incorrect analysis.

- Accuracy, completeness, and consistency of the data.

- Lack of context: Consider external factors such as holidays, weather conditions, or other events that may impact delivery performance.

- Failure to consider qualitative factors: Quantitative data is important, but it is also essential to consider qualitative factors such as customer feedback and experiences.

Problem Solving

Based on the data already analyzed, what would you do to reduce the average delivery time to 3 days?

- First: Would be to reduce the time between a sale and a delivery order, we would already be able to save more than 3 days in delivery time on average.

- Second: Would be to consider the regions with the highest average delivery times, being North and Northeast and consider a new logistics partner, or keep a larger stock of products at a shipping point within each region, where the delivery order would be outputted within the region, further reducing the time for final delivery.

- Third: An alternative would be to consider a new partner with greater knowledge and experience with deliveries in the mentioned regions, especially the north region which has an average delivery time of almost twice that of the southeast region (the shortest delivery time considered).

In addition to what has been reported, maintaining a stock at a central point for shipping goods, especially in the North and Northeast regions, would possibly also reduce the value of delivery to these regions. Obviously, values such as rent and inventory maintenance must be taken into account in addition to shipping values.

If you were to ask the engineering teams to gather more data, what should it be and why?

- We would need more precise information about the shipment step-by-step, many columns without information make it difficult to understand the bottlenecks that the product encounters until it reaches the consumer, despite a huge and almost total percentage of deliveries being within the estimated period

- I indicate collecting data on consumer feedback, to understand expectations of value and shipping time and compare them with the biggest competitors, to, in the future, even use it as a marketing strategy

What were the problems, how did you find them and what should we do to fix them?

- First, in the data, there was information with little precision that considerably altered the shipping time averages, precisely in the columns with ID 14883, 14838 and 14715, the delivered_at information was repeated and had very imprecise dates, as they represented very little of the general total of orders, I chose to remove them from the table and consider the averages with the values that remained.

- Second, the lack of step-by-step information on shipments makes it difficult to understand why some deliveries were so late

- Finally, no information was made available about the reason for canceling and returning some goods, so we cannot evaluate this metric.