

Section #1: Data Exploration

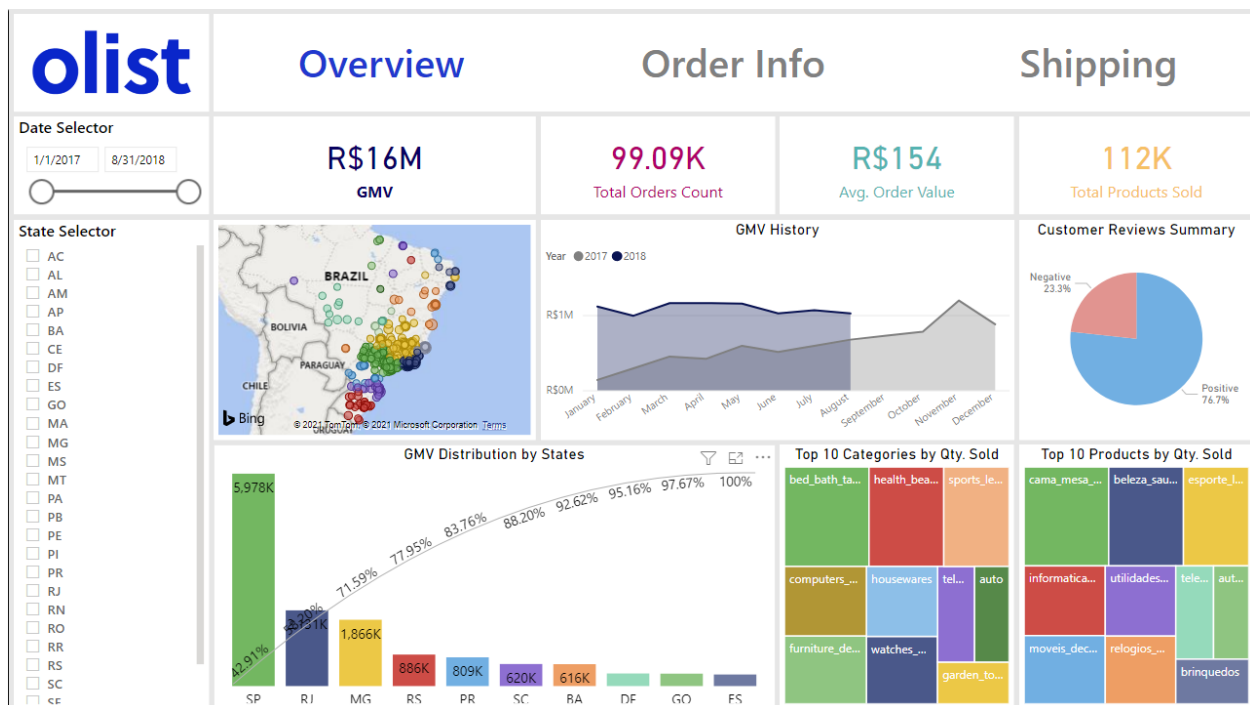
1. Understanding the dataset:

- Olist is a Brazilian e-commerce platform that operates as a SaaS (Software as a Service) since 2015.
- Olist offers a marketplace solution to retailers of all sizes.
- Olist itself does not buy any products nor keep products in stock.
- Participant shops is responsible for the logistics aspect of customers' purchases.

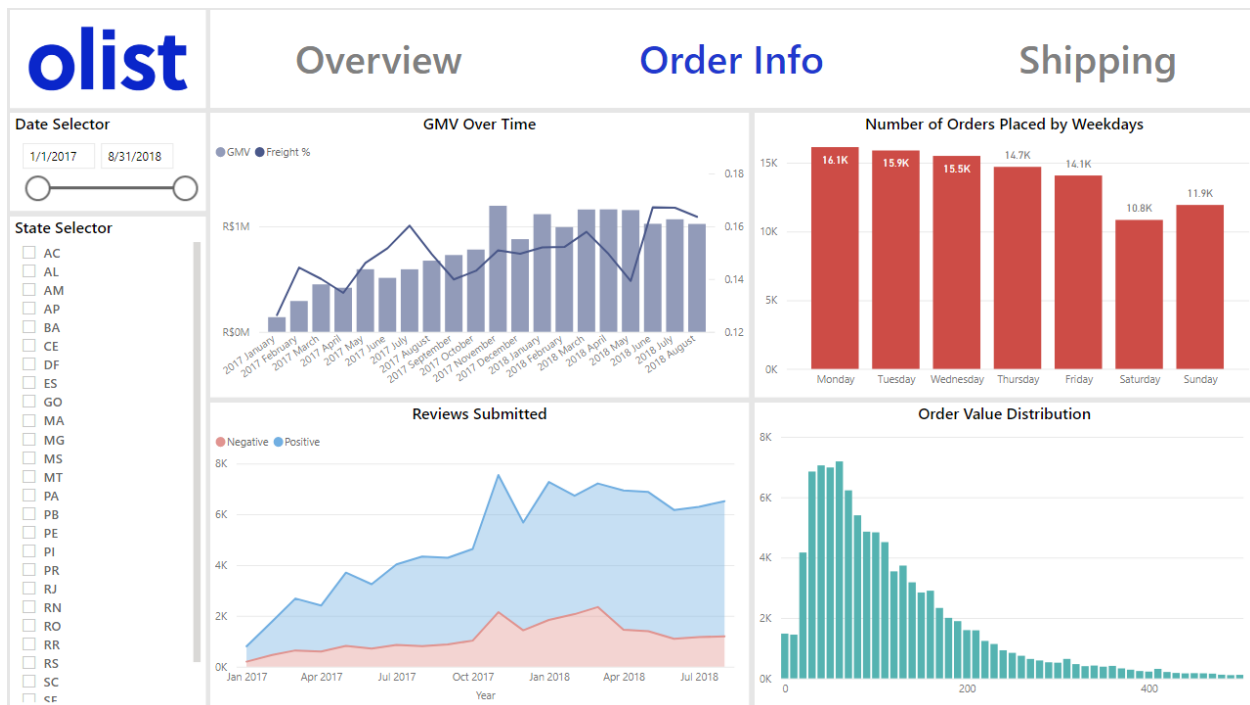
2. Explore and clean the dataset:

- The process has been documented in the Jupyter Notebook file.
- Currency is assumed to be Brazil Real (R\$)
- Data before January 2017 and after August 2018 has been filtered out due to inconsistency.

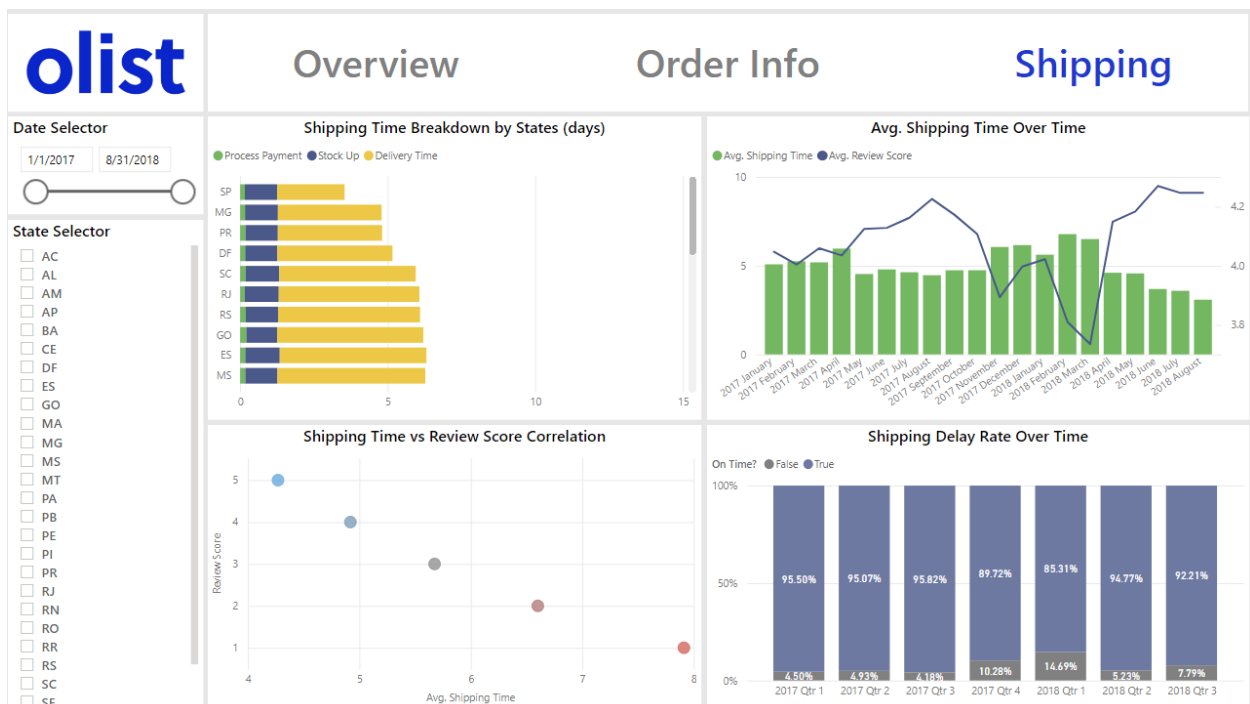
3. Describe the dataset:



- Most of Olist customers live in the east coast, which is in line with the population distribution of Brazil (<https://worldpopulationreview.com/countries/brazil-population#popDensityMap>)
- Olist GMV has increased a lot since 2017. Their total for the first 8 months of 2018 is R\$9M compared to R\$7M from their 2017 tally.
- Each order consists of around 1.1 items on average.
- Approximately 1 out of 4 reviews made by their customers is a negative one.
- 3 out of 27 states make up for 71.6% of total GMV, with SP ranked 1st by a huge margin.
- The most popular categories are bed_bath_table, health_beauty, and sports_leisure.

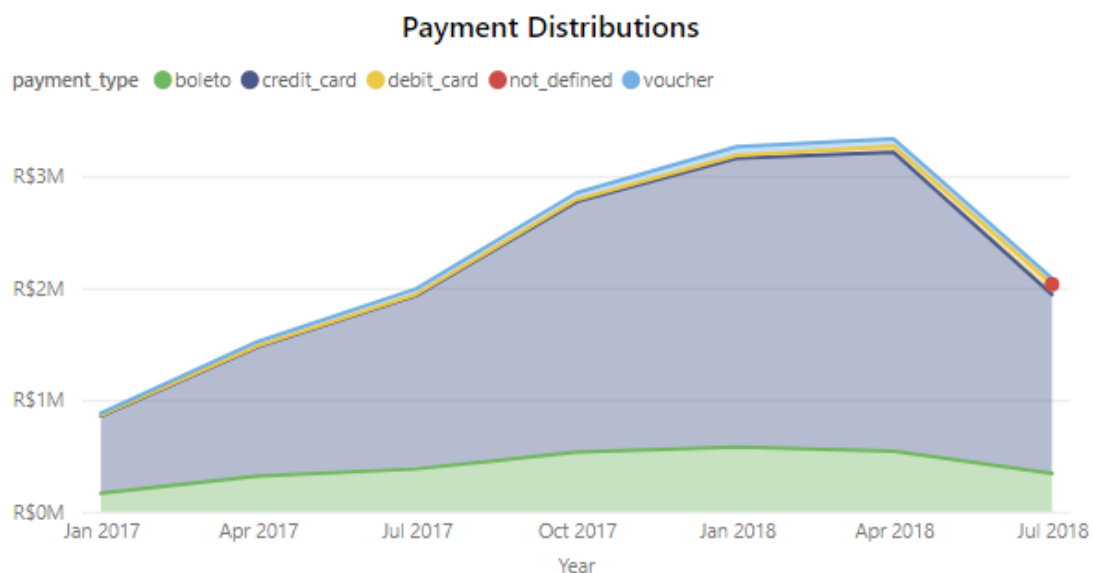


- For each R\$1 worth of products, Olist customer pay around R\$0.15 worth of shipping fee.
- Number of orders placed are skewed toward the beginning of the week and decrease gradually until Saturday.
- Number of reviews submitted by customers also increased as more orders are placed.
- A lot of orders fall in the range between R\$30 – R\$70.



seller_state	Count of seller_id
SP	1849
PR	349
MG	244
SC	190
RJ	171
RS	129
GO	40
DF	30
ES	23
BA	19
CE	13
PE	9
PB	6
MS	5
Total	3095

- The duration for payment and stock up processes are consistent across all states, with SP having the shortest average shipping time. This is because most sellers registered on Olist are also based in SP.
- The average shipping time has decreased from ~5 days since January 2017 to ~3days in August 2018.
- The average shipping time has a direct correlation with the average review scores. The later the products are delivered, the lower the review score is likely to be.



- Most customers use credit card or boleto as their preferred payment methods.

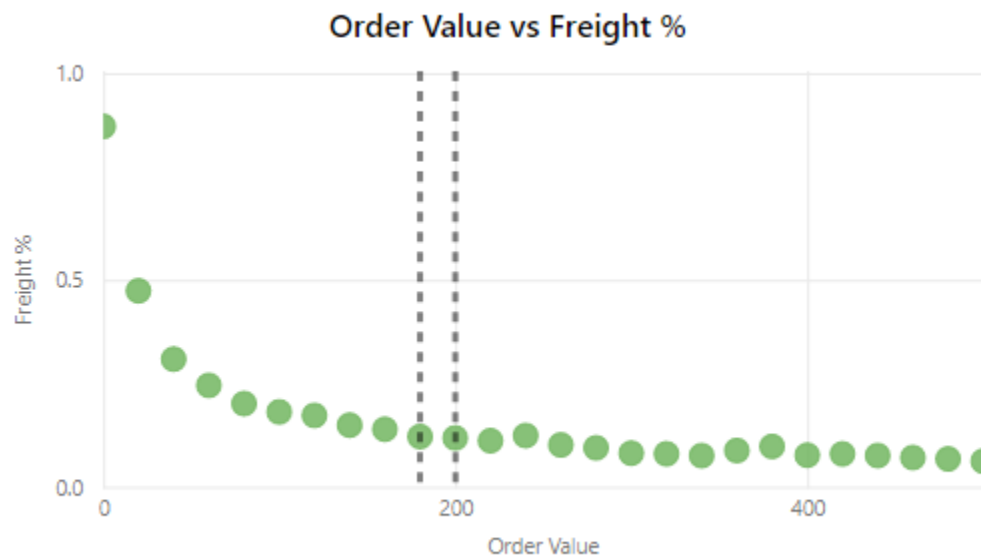
Section #2: Business Acumen

Main strategy of the company is to maximize **GMV** and optimize spending

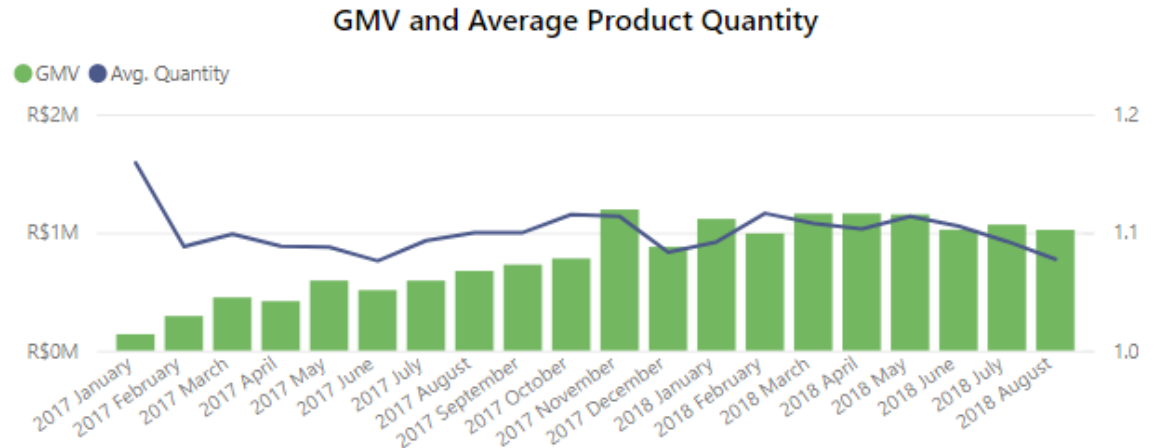
1. **Identify and measure at least three key suitable metrics to analyze the company performance and deliver insights for improvement. Explain rationale behind those metrics**

(1) Average Order Value:

- Keeping track of the Average Order Value is essential for any e-commerce to maximize their GMV. Having the AOV at a reasonably high enough level will both benefit the customer (pay less shipping fee to get more product value), the seller (higher profit margin), and Olist (more revenue going directly into the company, given that Olist does not benefit from the logistic aspects of the purchase)



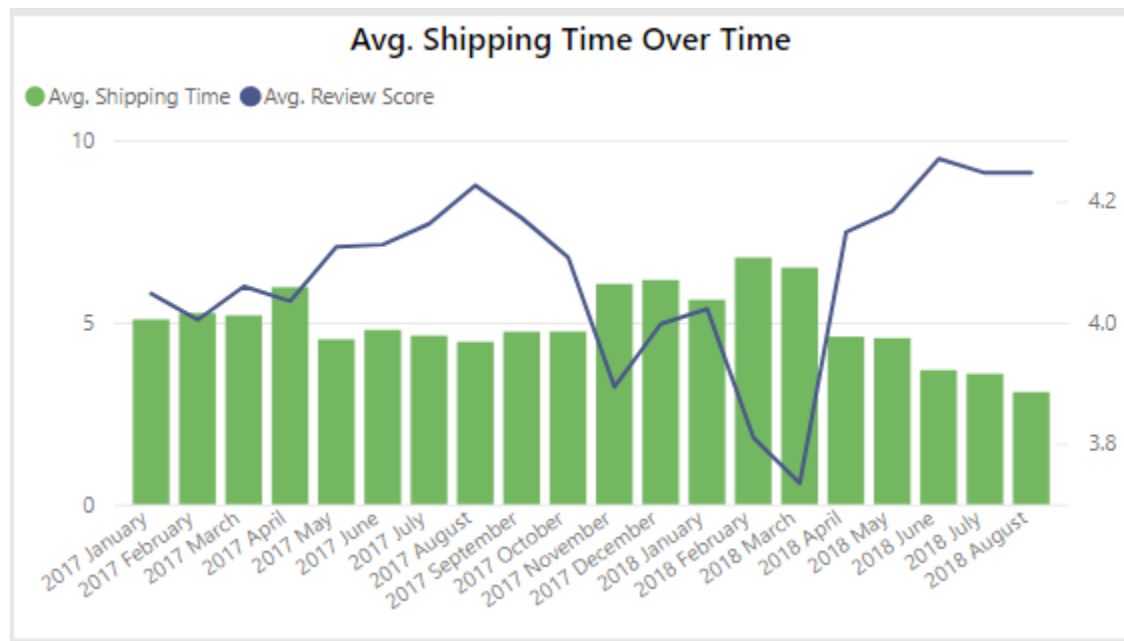
- The reasonably sweet spot to aim for seems to be around R\$180-R\$200. It's where the shipping cost percentage is at the lowest (around 12%) without making the customer having to upsize their purchase too much from their current behavior (~R\$150 – R\$160).

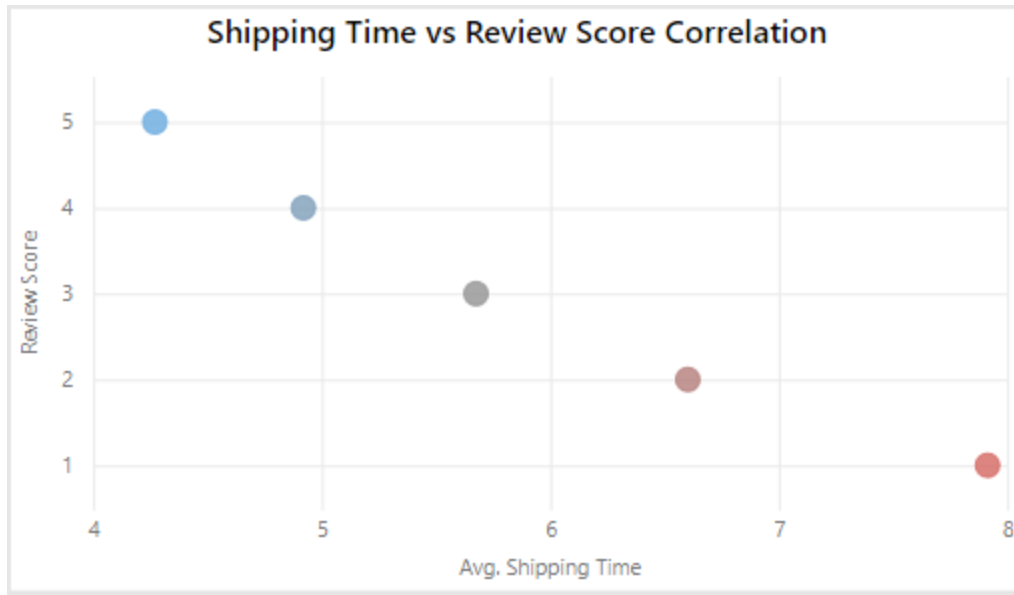


- Another insight that supports this fact is the average product quantity purchased in each order. Disregard January 2017 due to the low number of samples (high variance), it's prominent that all months with the highest GMV are amongst the highest in average quantity sold as well.
- Cross-selling, upselling, volume discounts, or free shipping can be utilized to increase the Average Order Value

(2) Delivery Time:

- In the e-commerce business, delivery time is no less (if not even more) important than the actual product quality. A lot of customers would be willing to change their supplier/e-commerce platform if that means they can receive the product sooner. Same-day delivery is also starting to be the norm in many places as well.





- Without going deep into the actual content of the reviews, a clear pattern can be seen where the longer the delivery time is, the more likely that the customer will leave a bad review.

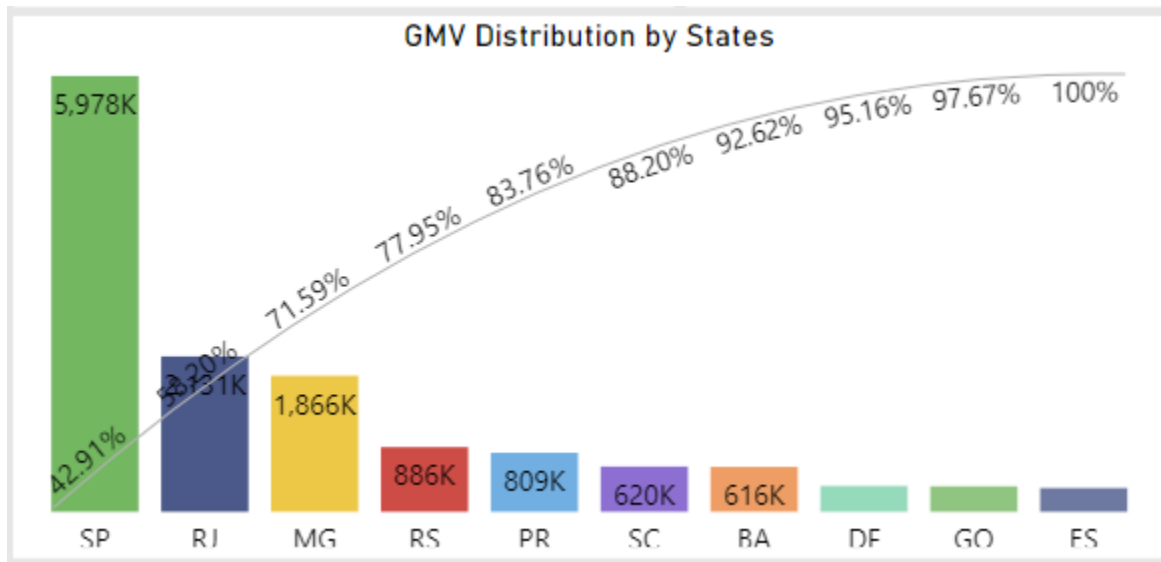
Note: the huge dip in review score and increase in delivery time in 2018 February, March can be related to the trucker strike that happened at the same time in Brazil.



- A huge gap of review score can be seen between orders which were delivered on time (on or before the estimated delivery time) versus orders that were not.

- An unsatisfied customer is much more likely to spread their negative views to others compared to a satisfied one, which will directly affect the company's GMV moving forward. That is why keeping the delivery time low and consistent (hence higher review score) is very vital.
- Having the customer support team (or the seller) following up unsatisfied customers and provide them with reimbursements/discount voucher is recommended to retain their business. Flipping an unsatisfied customer into a satisfied one would hugely increase the chance that they become a promoter for the company as well.

(3) GMV Distribution by States:

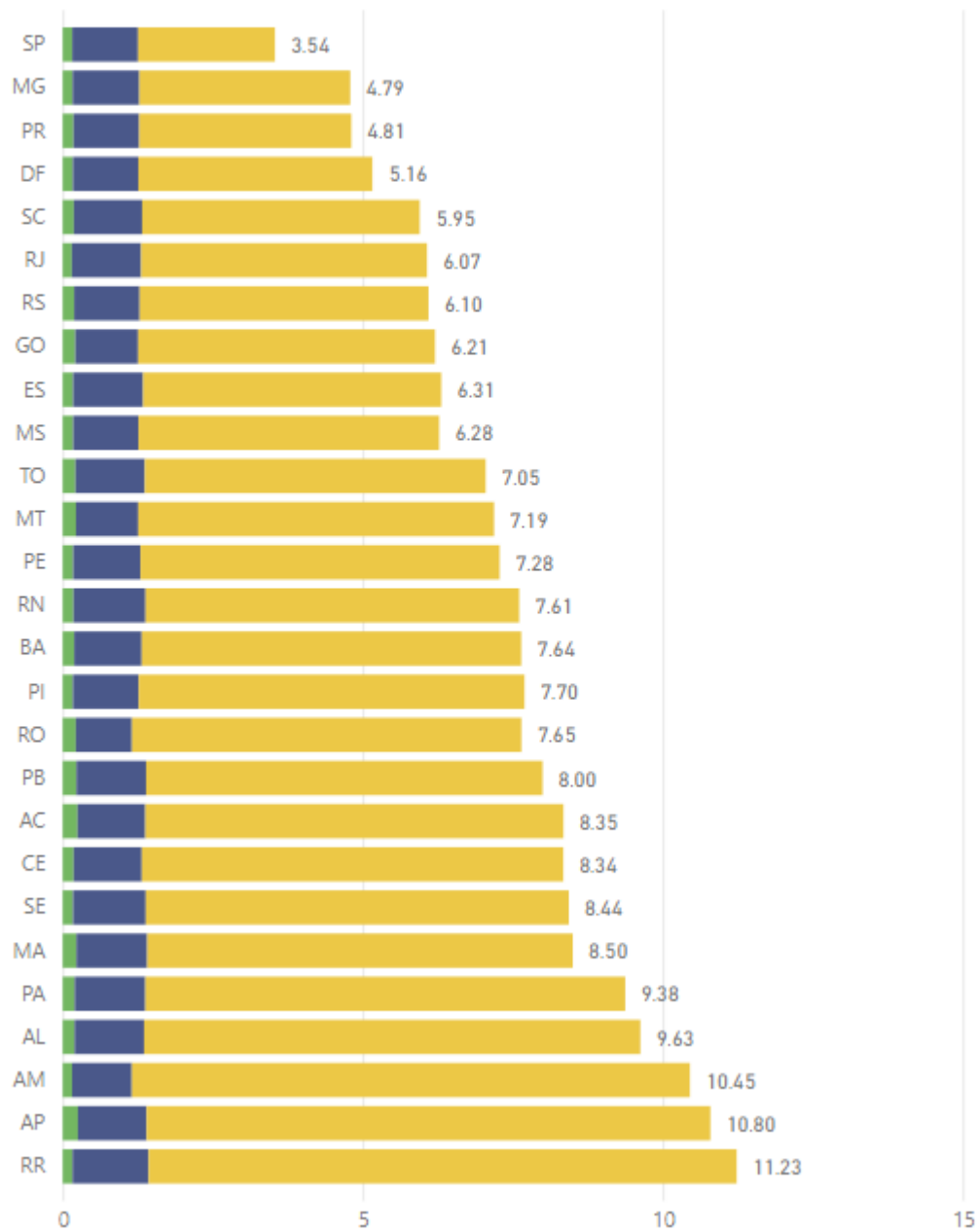


Rank	State	Population (2020) ^[2]	Population (2010) ^[3]	Population (2000) ^[4]	Population approximately equal	Percent of the total Brazil population, 2020 ^[note 1]
1	São Paulo	46,289,333	41,262,199	37,032,403	Argentina	21.9%
2	Minas Gerais	21,292,666	19,597,330	17,891,494	Sri Lanka	10.1%
3	Rio de Janeiro	17,366,189	15,989,929	14,391,282	Netherlands	8.2%
4	Bahia	14,930,634	14,016,906	13,070,250	Cambodia	7.1%
5	Paraná	11,516,840	10,444,526	9,569,458	Belgium	5.4%
6	Rio Grande do Sul	11,422,973	10,693,929	10,187,798	Belgium	5.4%
7	Pernambuco	9,616,621	8,796,448	7,918,344	Hungary	4.5%
8	Ceará	9,187,103	8,452,381	7,430,661	Israel	4.3%
9	Pará	8,690,745	7,581,051	6,192,307	Switzerland	4.1%
10	Santa Catarina	7,252,502	6,248,436	5,356,360	Paraguay	3.4%

- Looking at the difference in population between Sao Paolo and other states, it's understandable why it's the 1st in GMV ranking. However, the main reason why the margin is so high (3 times higher than the 2nd state) can be attributed to the much faster shipping time compared to other states (30% and above), as shown below:

Shipping Time Breakdown by States (days)

● Process Payment ● Stock Up ● Delivery Time



- Like what have been stated in the 2nd metric, delivery time plays a big factor when a customer decides what e-commerce platform to choose. This difference in delivery time indicates that the market share in other states has yet to be fully tapped.
- To bring up the number in other states on par with the population difference, Olist is highly recommended to connect with more sellers from those states to shorten the delivery time, hence increase the market share. Bahia could be a good priority as their average shipping time is much longer than the top 2 states despite ranking 4th in populations.

2. Which additional data/dataset do you think the company should collect and why?

- (1) Very important: seller's kickback percentage. It shows how much actual revenue goes into Olist (not the seller). Only after knowing that number can we plan out business growth strategies appropriately and distribute the correct budget for them (e.g. locating extra marketing effort for sellers that generated high GMV for Olist)
- (2) Marketing analytics data such as avg. session duration, bounce rate/exit rate to find out how well their website is performing in terms of accessibility, user experience quality.
- (3) New customer registration to keep track how well are they expanding the customer base, and how long until they make their first order. It also helps to find out the percentage of new versus returning customers, which determines how good we are at retaining customer
- (4) Cart abandonment rate to send follow up email and retain potentially lost customers.
- (5) Discount used on each order, what type of discount are they (free shipping, shop discount, voucher, loyalty point) to find out the optimized discount value that balances between GMV and market share.