Olist Ecommerce Data Analysis

Colm Yeh

Agenda

1. Problem
2. Data
3. Customer Segmentation
4. Geospatial Analysis
5. Time Series Forecasting of Demand
6. Recommendations
7. Next Steps

Managers at Olist need to understand future demand in number and geography in order to **properly allocate capital across their operations.**

My method will consist of 3 objectives:

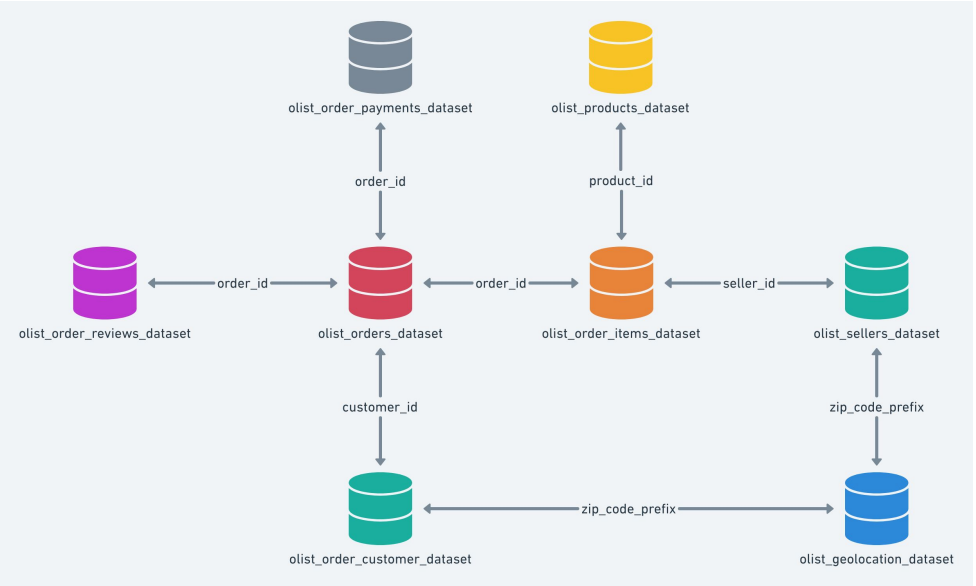
Geospatial Analysis

Segmentation Analysis

Time Series Forecasting of Sales

Data

- Olist Ecommerce Dataset: 9 tables
  - Customers
  - Geolocation
  - Order Items
  - Payments
  - Products
  - Sellers
  - Orders
  - Translations (Portuguese to English)



Recency, Frequency, Monetary: Clustering

RFM analysis examines customers across 3 characteristics: recency, frequency, and monetary value.

Recency: how *recent* a customer purchased

Frequency: how *frequent* a customer purchases

Monetary: how *valuable* the customer purchases are

Modeling: Kmeans

Implemented Kmeans modeling to iteratively cluster our customers according to RFM characteristics. Returned 5 customer segments:

	Recency	frequency	payment_value	
	mean	mean	mean	count
clusters				
0	247.0	1.0	1129.0	4747
1	242.0	1.0	143.0	88367
2	198.0	1.0	22067.0	19
3	339.0	1.0	109313.0	1
4	238.0	1.0	5342.0	262

Customer segment 0:

- Middle of the road monetary value
- Similar to cluster 1, but have much higher average monetary value.
- Second largest cluster by customer count, and second most valuable customer segment as a percentage of sales revenue.

	Recency	frequency	payment_value	
	mean	mean	mean	count
clusters				
0	247.0	1.0	1129.0	4747
1	242.0	1.0	143.0	88367
2	198.0	1.0	22067.0	19
3	339.0	1.0	109313.0	1
4	238.0	1.0	5342.0	262

Customer segment 1:

- Low monetary value with relatively high recency.
- Purchase about once a year with relatively low order value.
- They are the largest customer segment by count and as a percentage of sales revenue.

	Recency	frequency	payment_value	
	mean	mean	mean	count
clusters				
0	247.0	1.0	1129.0	4747
1	242.0	1.0	143.0	88367
2	198.0	1.0	22067.0	19
3	339.0	1.0	109313.0	1
4	238.0	1.0	5342.0	262

Customer Segment 2:

- High monetary value with with relatively low recency.
- On an individual basis, our most valuable customers.
- Highest average order value and have purchased the most recently.

	Recency	frequency	payment_value	
	mean	mean	mean	count
clusters				
0	247.0	1.0	1129.0	4747
1	242.0	1.0	143.0	88367
2	198.0	1.0	22067.0	19
3	339.0	1.0	109313.0	1
4	238.0	1.0	5342.0	262

Customer Segment 3:

- Outlier customer who purchased a significantly large amount a long time ago.
- Probably some kind of business.
- Does not fit into the other segments, so they stand alone.

	Recency	frequency	payment_value	
	mean	mean	mean	count
clusters				
0	247.0	1.0	1129.0	4747
1	242.0	1.0	143.0	88367
2	198.0	1.0	22067.0	19
3	339.0	1.0	109313.0	1
4	238.0	1.0	5342.0	262

Customer Segment 4:

- Middle-high monetary value with lower recency.
- Monetary values sit between clusters 0 and 2
- Second lowest recency out of all our clusters.
- Higher payment values, and also ordered more recently than our other customer segments (excluding segment 2).

	Recency	frequency	payment_value	
	mean	mean	mean	count
clusters				
0	247.0	1.0	1129.0	4747
1	242.0	1.0	143.0	88367
2	198.0	1.0	22067.0	19
3	339.0	1.0	109313.0	1
4	238.0	1.0	5342.0	262

Geospatial Analysis

Customer Count

SP: 48.8K

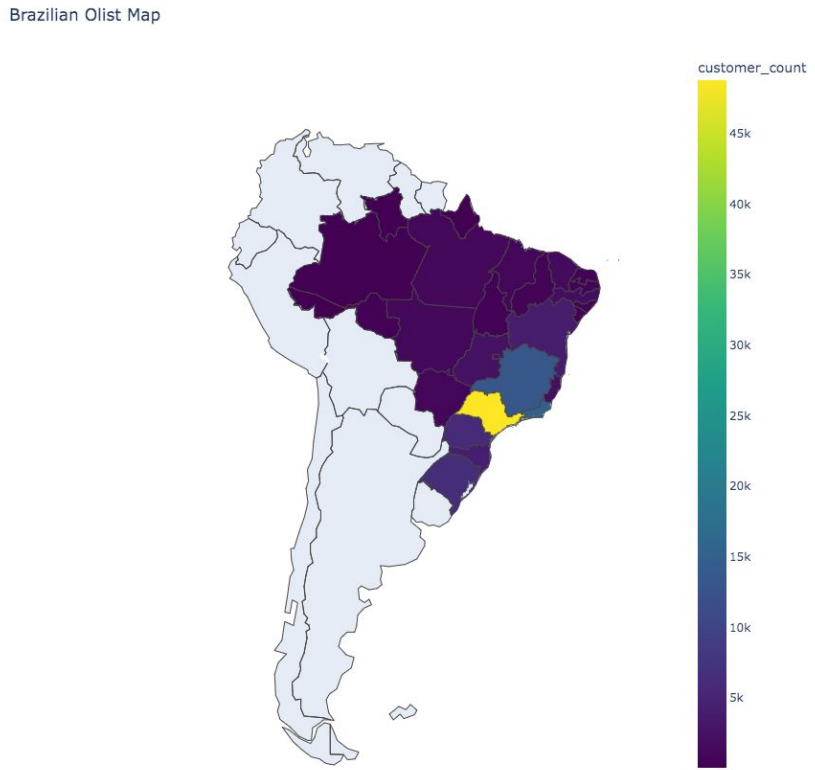
RJ: 14.9K

MG: 13.4K

RS: 6.4K

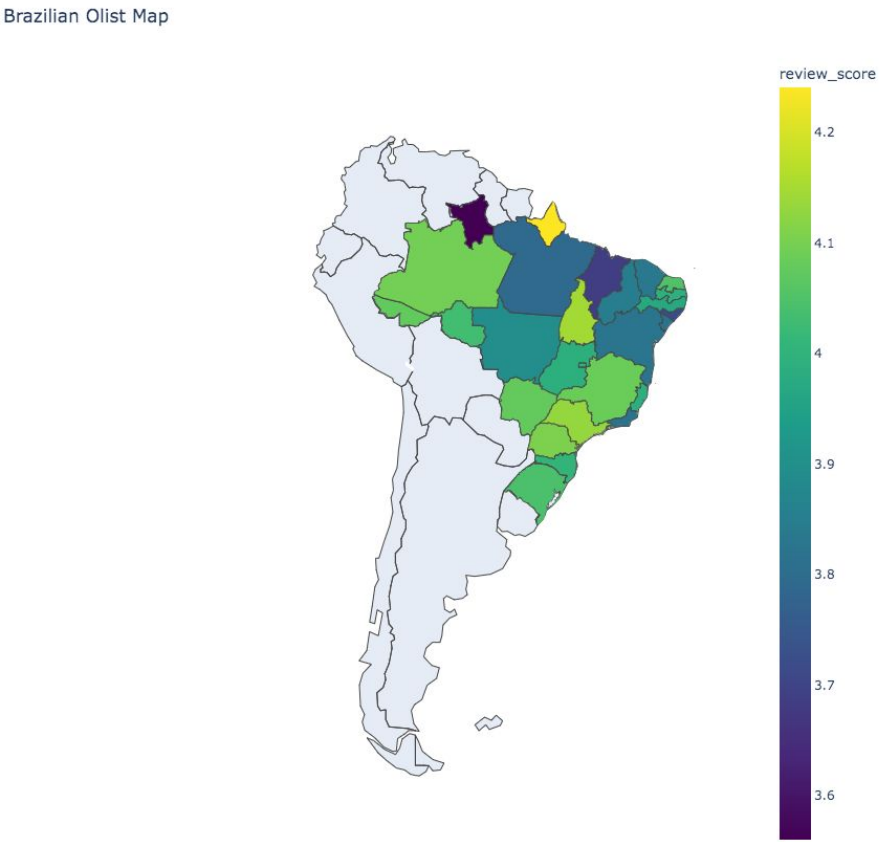
PR: 5.8K

Urban areas of the south have more customers than more rural north



Review Score

Mostly 4+ in southern Brazil, with scores generally declining northward. This is likely due to more rural areas experiencing longer delivery times and lower delivery rates.



Freight Value

SP: 742.95K

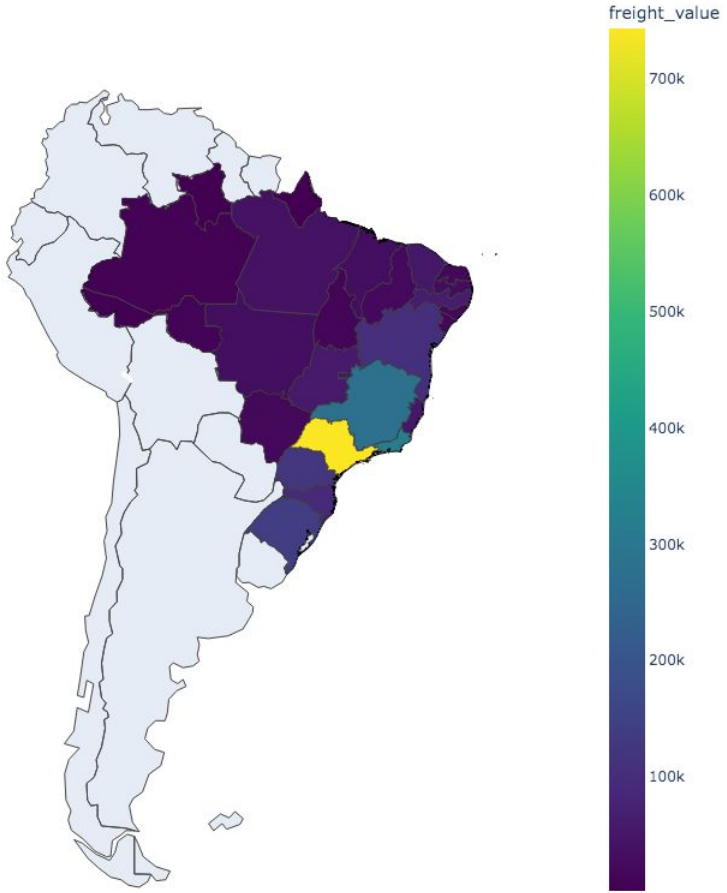
RJ: 316.86K

MG: 277K

PR: 121.1K

RS: 139K

High freight values are concentrated in southern Brazil. FV declines in the northern regions.



Sales Revenue BR$

SP: $5.36 million

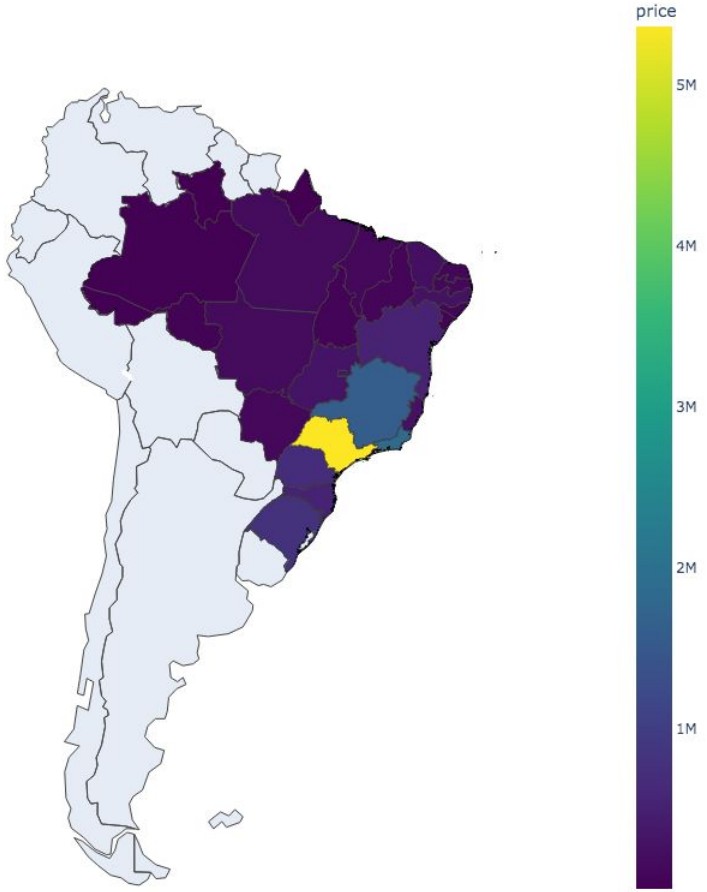
RJ: $1.86 million

MG: $1.61 million

RS: $779.37K

PR: $697.62K

Sales Revenue concentrated in the south, with Sao Paulo leading. Again, the northern, more rural regions show the weakest numbers.



Seller Count

SP: 82.6K

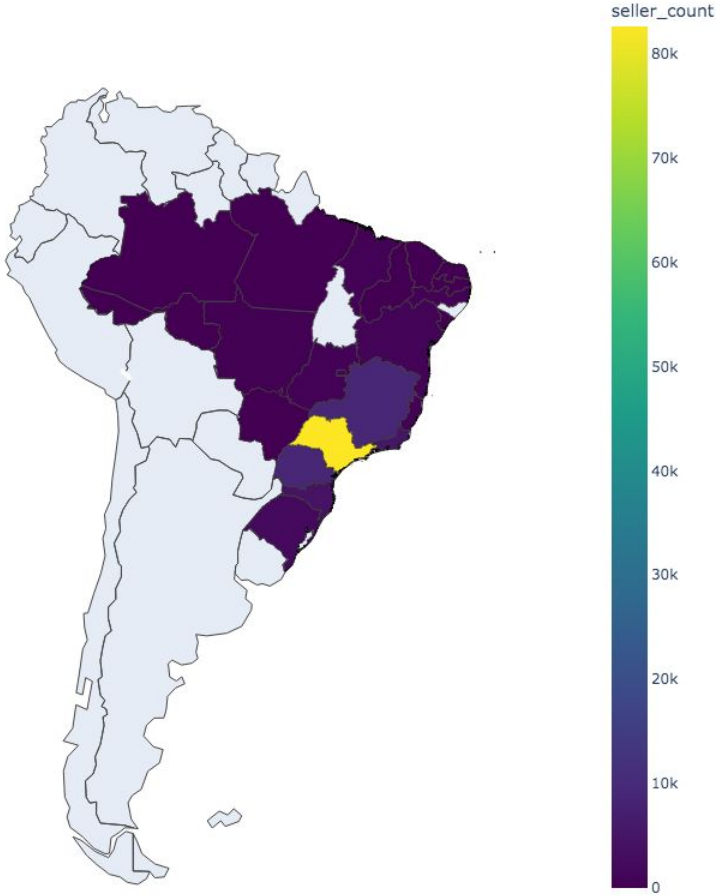
MG: 9046

PR: 8949

RJ: 4930

SC: 4268

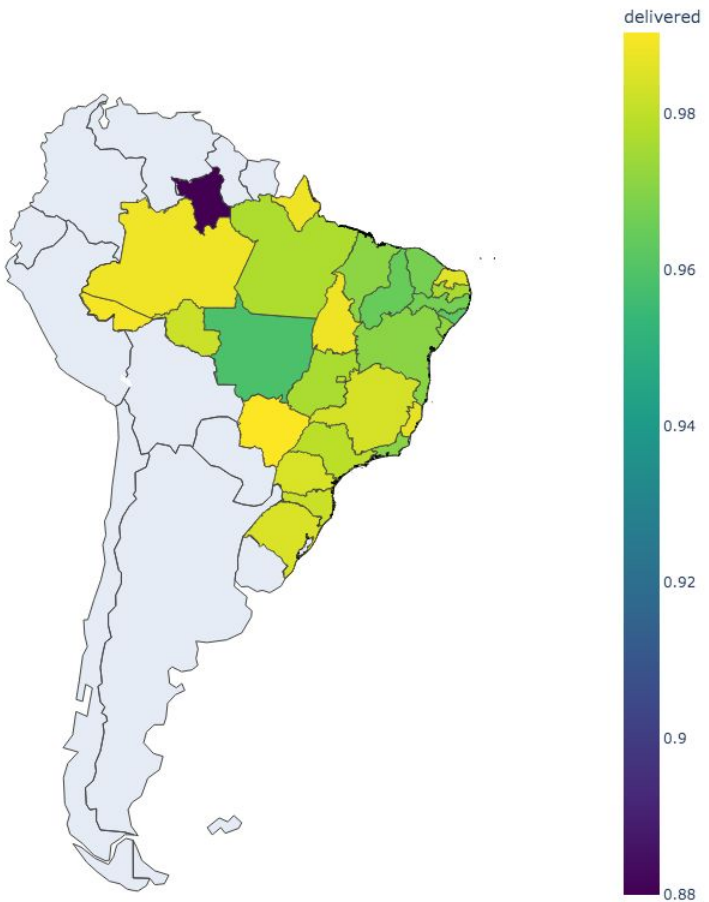
Sellers are extremely concentrated in Sao Paulo. Barely any sellers in the north relative to core geographies in the south.



Delivery Rate

Delivery rates are fairly consistent (>90%) across the country.

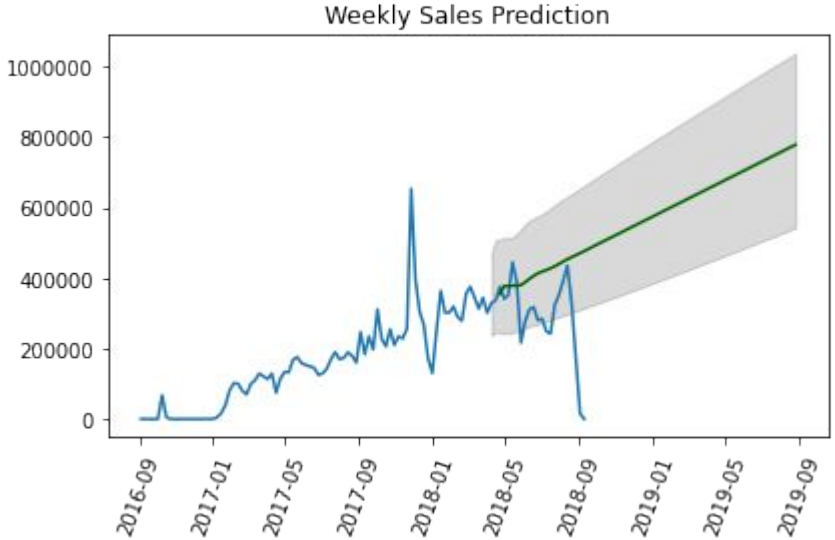
RR stands out at 88%, indicating this northern rural region has delivery issues.



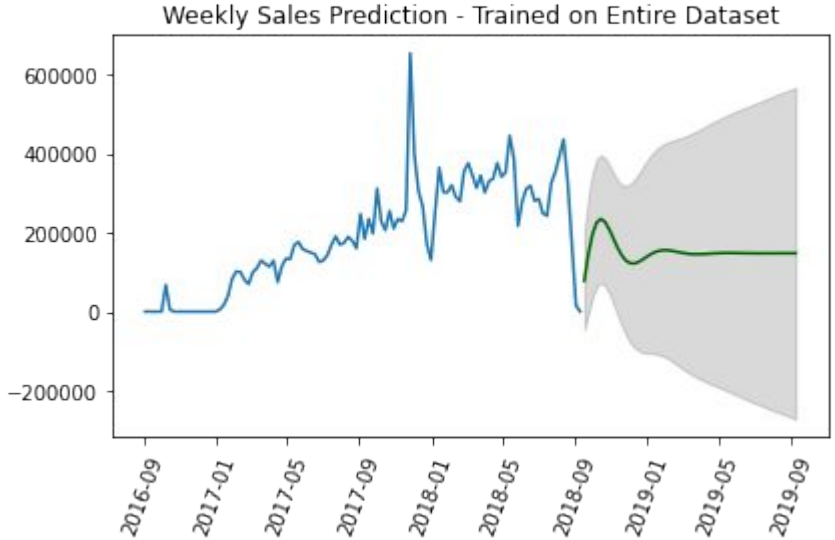
Time Series Analysis and Forecasting

Two SARIMAX models were implemented in order to forecast sales revenue 1 year into the future:

Model A trained with validation set:

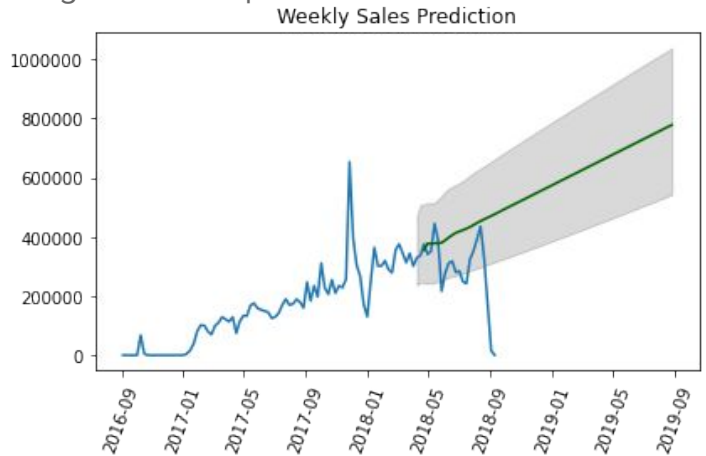


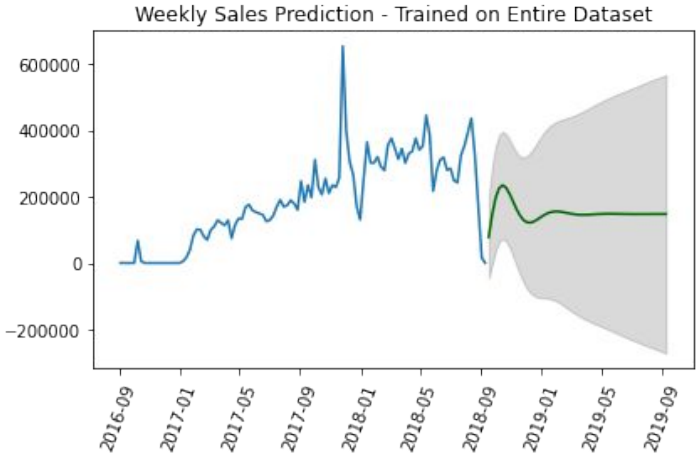
Model B trained without:



Forecasting Issues: data collection

Unusually, our data shows weekly sales dropping off a cliff to zero. This "outlier" event causes some difficulty in our SARIMAX modeling. Model A completely misses the downturn in sales. Model B is trained on the entire dataset, resulting in a better picture.





Model B predicts a bounce back of sales, which stabilize around 147K.

Business Recommendations: Geospatial

Invest in building out operations in Bahia, a region close to our hub of operations in the south with the third largest population in Brazil (after SP, MG).

- Healthy traction in sales, freight value, and delivery rates
- Close proximity to core geographies of SP, MG, and RJ, allowing for an easier transfer of transportation assets and personnel.
- Managers should focus on increasing sales from the region, along with increasing the number of sellers.

Business Recommendations: Geospatial

Invest in building out operations in Paraná, a region close to our hub of operations in the south with the third largest population in Brazil (after SP, MG).

- Paraná has a large population with close proximity to our hub of operations.
- Sales, freight values, and seller count suggest that Paraná could also become a core geography for Olist.
- Lower total population than Bahia, but nearly twice as dense. (lower average distance traveled for delivery)

Olist's capital allocations should be **concentrated in the south and move north** towards the northeast population centers as the southern markets mature.

- Southern Brazil has multiple important population centers surrounding Olist's hub of operations in Sao Paulo.
- Building out operations in the south should be a first priority.

Business Recommendations: Clustering

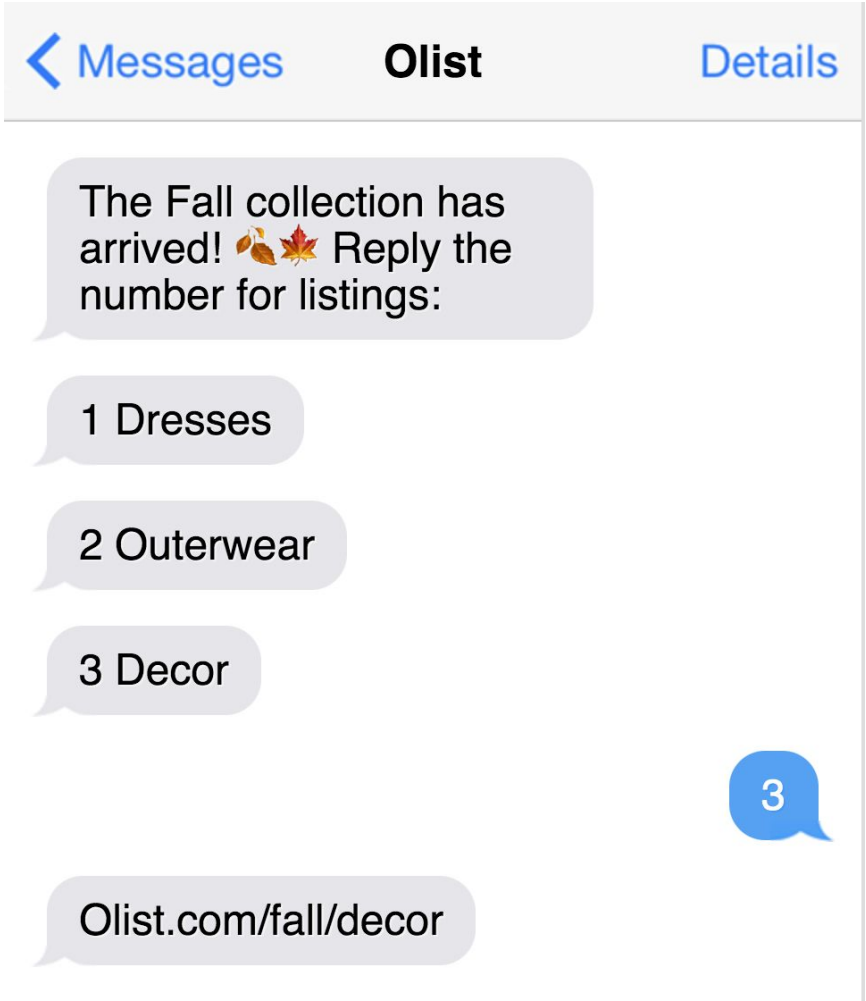
Create targeted upsell recommendations for customers during checkout in order to *increase average order value*.

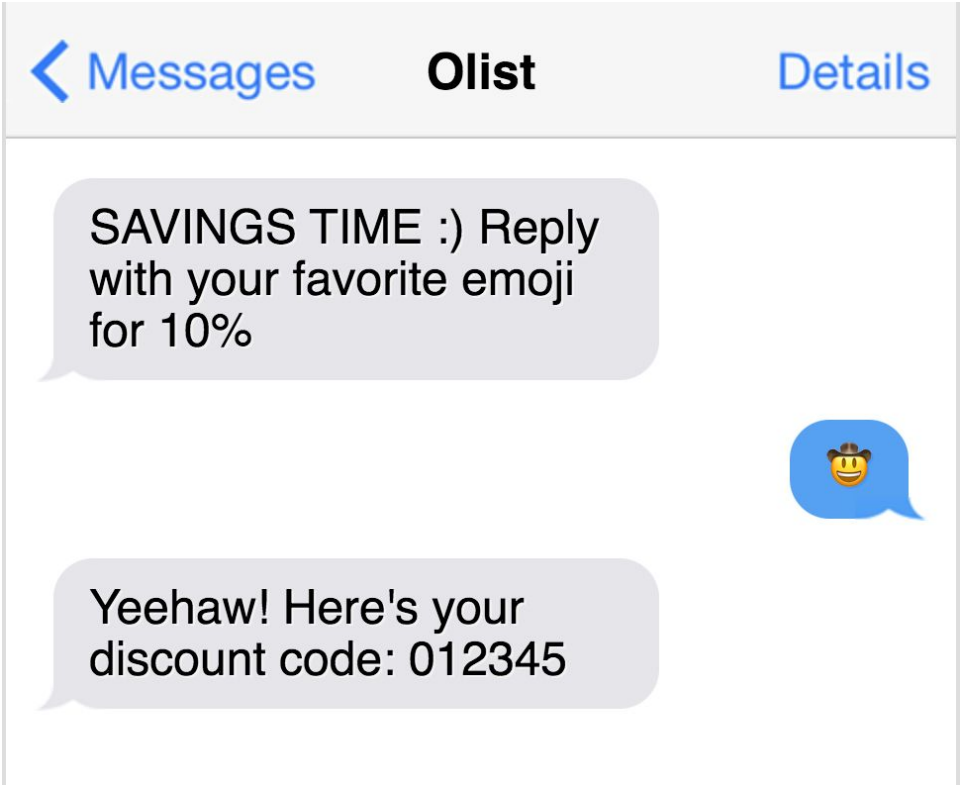
- Our largest customer segment (cluster 0 at 83K customers) has a meager BR $143 average order value (AOV).
- By providing relevant, targeted upsells using a recommendation model, Olist managers will increase AOV and drive more profitability across our customer base.
- Implemented in browser during checkout, as well as emailed in a preprogrammed, post-purchase email flow.

Post Purchase Email Flow

Build out an SMS list to recapture customers and increase customer lifetime value.

- High open rates (up to 90%)
- Collect phone numbers during checkout
- Personal touchpoint with Olist brand
- Periodic SMS campaigns offering discounts, announcing new product offerings, or promoting events (holiday sales, black friday, etc.) should increase the frequency of buyers, resulting in an increase of customer lifetime value.


Messages
Olist
Details
The Fall collection has arrived! 🍂🍁 Reply the number for listings:
1 Dresses
2 Outerwear
3 Decor
3
Olist.com/fall/decor


Messages
Olist
Details
SAVINGS TIME :) Reply with your favorite emoji for 10%
🤠
Yeehaw! Here's your discount code: 012345

Business Recommendations: Time Series

Build out an Olist paid membership program in order to provide more stable, recurring monthly revenue.

- Olist's sales revenue over the examined period had multiple drastic fluctuations, including a sudden drop to zero.
- Olist memberships could provide some predictability and stability in Olist's revenue structure. This will give managers more stable cash flow, allowing them to make more informed decisions on capital allocation.
- Acts as “floor” for monthly revenue even if sales drop to zero.


COSTCO
WHOLESALE
GOLD STAR MEMBER


Walmart+


amazonPrime

Next Steps

Comprehensive review of Bahia and Paraná regions:

Feasibility study using financial and logistics data

- Does Olist have the logistics in place to handle more capacity?
- Do investments in these regions make sense at the current cost of capital?
- Is it better to buy or build the systems needed to increase revenue?

Second, the success of our customer lifetime value campaigns should be monitored and evaluated.

Relevant success metrics include:

- *Average order value,*
- *Frequency and recency*
- *Customer lifetime value.*

Questions