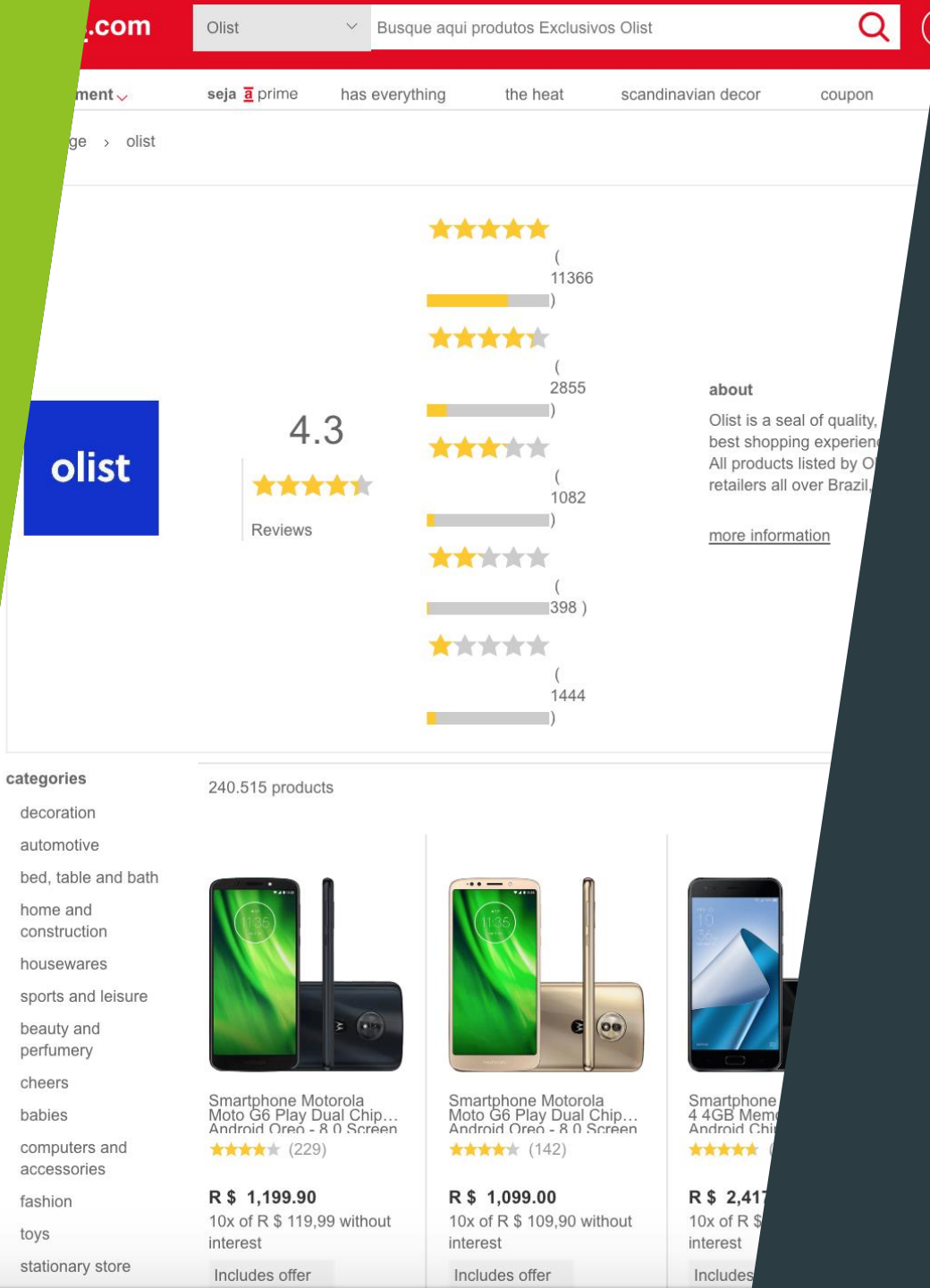


Entering The Online Marketplace via Olist

MAR653 Final Project

Kristen Cheung, Cindy Khuu,
Kenneth Kelly, Kyle
Wojtaszek



Executive Summary

- ▶ Olist is the largest department store in Brazilian marketplaces. Allows merchants from small businesses throughout Brazil to sell their products and ship them directly to the customers through Olist's logistics partners.
- ▶ Identified six customer clusters based on basket size, freight, and review score with an opportunity to tailor more focus towards two of the largest segments DIY/ Tech and Home/Decor Customer
- ▶ Identified four seller clusters based on the number of products listed, number of categories listed in, and average price of items. Used these clusters in tandem with customer clusters to determine best opportunities
- ▶ Delivery experience overall has a large impact on review score with more negative experience likely results in lower scores and positive experienced likely results in positive scores

Business Overview

- Olist

- ▶ Olist is an e-commerce company based in Curitiba, Brazil
- ▶ Brazilian version of Etsy's marketplace for products
- ▶ Founder Tiago Dalvi created Olist in 2015 with the goal of helping shopkeepers reach larger and the best marketplaces nationally and internationally.
- ▶ Olist is among the Top 3 largest department store inside Brazil's largest marketplaces and has generated \$1.8 Million in revenue annually.

olist

olist store pricing para quem é o olist? faq enterprise more

olist store

A maior loja de departamentos dos marketplaces formada unicamente por outros lojistas

Quero vender mais

- ✓ Um poderoso canal para aumentar suas vendas
- ✓ Built-in reputation from day 1
- ✓ Single contract and less bureaucracy
- ✓ Novidade - Ferramenta competitiva de sugestão de preço!



Be part of a community of more than 4.500 lojistas que vendem seus produtos no



americanas.com



Entenda como o Olist funciona

Research Objective

To better understand how to sell on platform in order to maximize success of a small businesses by:

- Researching what types products are sold the most and who are buying them
- Identifying opportunities via customer reviews and customer purchase history

► Questions:

- How to improve revenue with a new seller?
- How to give a new vendor the best chance for success?
- Who are our customers?
- What makes a business successful today?
- What drives a good customer review score?

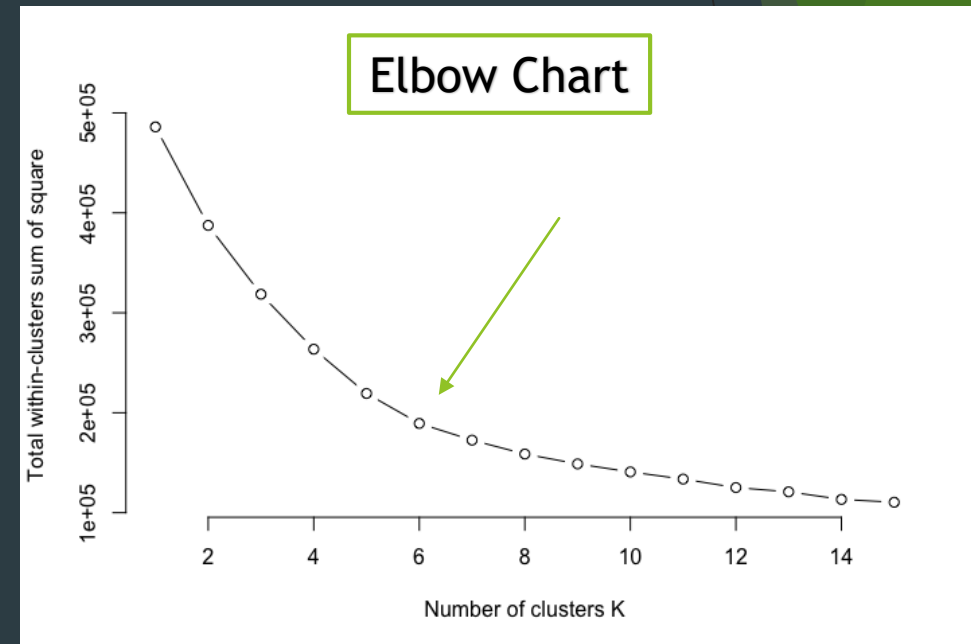
Research Plan and Methodology

- ▶ Data usage:
 - ▶ Customer reviews and score tied to product, order, and seller
 - ▶ Customer transaction level purchase history tied to product and seller
 - ▶ Product details: Units sold, unit price, product photo representation online
 - ▶ Sales and order fulfillment data from each seller
- ▶ Difficulties with the data set:
 - ▶ Data set does not contain many repeat customers, so CLV calculations and determining what brings customers back was not possible
 - ▶ Review score text analysis did not provide novel insights
 - ▶ Bad reviews had negative words, good reviews had positive words
 - ▶ Finding an automated way to search through the linear regression results for the significant ones
 - ▶ Finding factors other than price that significantly affected sales, and outside of a few small pockets, there aren't really any good trends. Overall, as you would expect, for a given item, the lowest price generally sells the most.

Clustering & Identifying Types of Customers

Used K-Means to identify 6 clusters based key traits to consider when a customer shops and the elbow plot:

- **Avg Sales \$ Basket**
 - The average dollar spent on each order can tell us how much they are willing to spend on each order. Are they higher basket dollar values or smaller dollar basket values?
- **Avg Unit Basket**
 - The average units purchased can tell us how many units do they usually purchase. Do they purchase only one item, or do they purchase a lot of items? Do they purchase in bulk?
- **Avg Review Score**
 - The average review score can indicate whether customers are satisfied or dissatisfied based on their previous shopping experience.
- **Number of Photos Available on Ordered Products**
 - The number of photos tied to the products they purchase can tell us if some customers prefer to visually see photos of a products before making a purchasing decision.
- **Freight \$ % of Total Ordered \$**
 - The percentage of freights from the total dollars spend on orders can help us indicate whether or not freight cost is a consideration when making a purchase since freight cost can vary at large with the distance between the seller and customer.

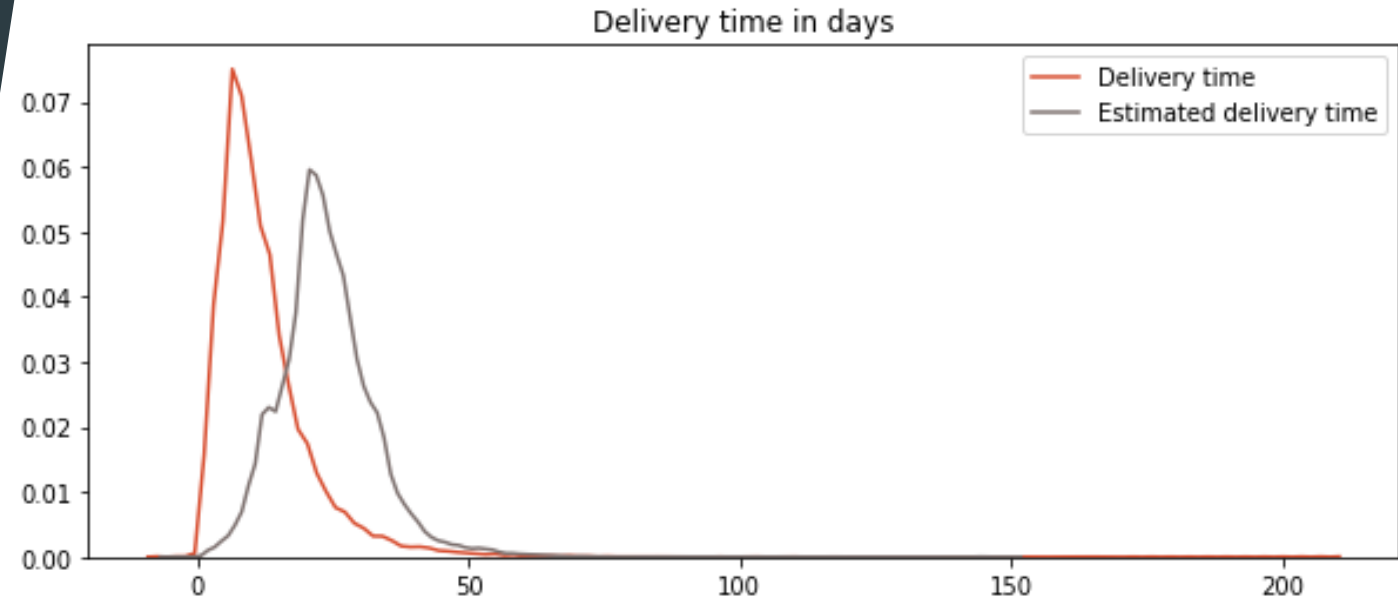


Customer Segmentations & Profiles

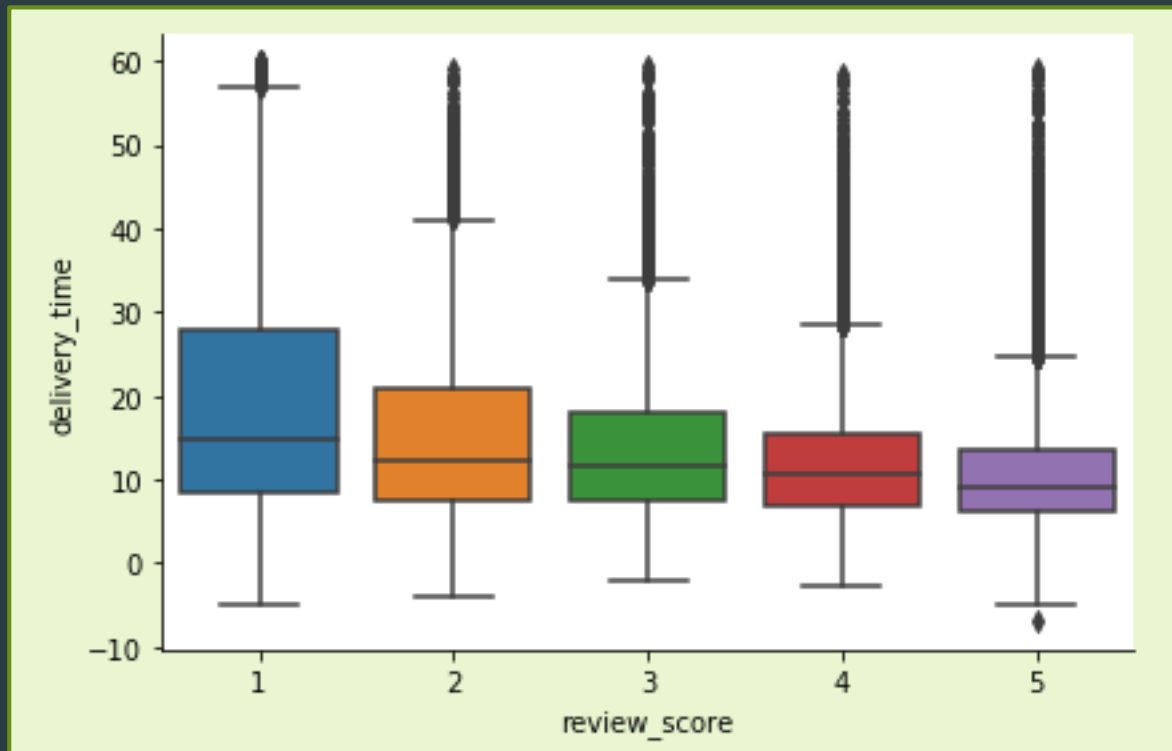
Segments	Profiling Descriptions	# Customers	% of total customers	avg_sales_basket	avg_basket_units	orders_tot_product_photos_qty	avg_review_score	avg_num_install_payment	freight_percent_order	orders_tot_product_dimension_volume_cubic_cm	orders_tot_product_weight_g	auto	bed_bath_table	computers_accessories	furniture_decor	furniture_living_room	garden_tools	health_beauty	home_construction	housewares	toys
1	Home / Décor Customer: -Likes to upkeep home with living room furniture, houseware, garden tools -Likes photos to help with purchase decision -Largest customer segment at 33% with \$146 avg basket	32,441	33%	\$ 146.34	0.99	3.50	4.61	3.03	15.0%	16,494	2,194	0.0000	0.0080	0.0012	0.0000	0.5468	0.3457	0.0000	0.0004	0.0694	0.0375
2	On A Budget Buyers: -Limited wallet with smallest avg basket of \$56.78 -Pays higher freight % of total order due to distance from seller based on freight cost on volume and weight of delivered items -Considers total out of pocket expense when purchasing versus using installment plan	18,770	19%	\$ 56.78	1.02	2.53	4.49	1.93	39.3%	9,746	1,186	0.0001	0.3093	0.1251	0.0000	0.3261	0.1873	0.0004	0.0244	0.0399	0.0251
3	DIY /Tech Customer: -Likes DIY person who likes house projects by purchasing home construction tools and bed bath table products -Likes to personally update and fix own computer by purchasing computer accessories - Second largest customer segment at 26% with avg basket of \$147	25,748	26%	\$ 147.15	1.02	1.12	4.61	3.00	15.7%	15,250	2,261	0.0001	0.6997	0.2824	0.0000	0.0000	0.0000	0.0001	0.0558	0.0000	0.0000
4	Unsatisfied Customers: Low review scores of 1.3	12,449	13%	\$ 143.66	1.07	2.54	1.30	3.00	20.1%	17,408	2,462	0.0001	0.3504	0.1421	0.0000	0.3323	0.1819	0.0001	0.0267	0.0357	0.0206
5	Convenience Seekers: -Likes convenience of choosing all products from one place -Have larger baskets sizes compared to other segments from a wide assortment -Like a lot of photos to help with purchasing decision	6,429	7%	\$ 244.90	2.62	7.81	3.62	3.55	23.8%	40,602	5,213	0.0000	0.4302	0.1462	0.0005	1.0726	0.7292	0.0005	0.0347	0.1289	0.0946
6	High Spenders: -buys higher price items based on the UPT and avg sales basket -Prefers longer installment payment plans	2,139	2%	\$ 1,179.25	1.09	2.66	4.08	6.48	5.3%	54,708	8,741	0.0000	0.3604	0.0991	0.0000	0.3403	0.1987	0.0000	0.0266	0.0542	0.0182

Delivery Expectations

- ▶ The delivery time is based on when the product arrives to the customer's home after placing an order.
- ▶ The estimated delivery time was typically longer than the actual delivery time.



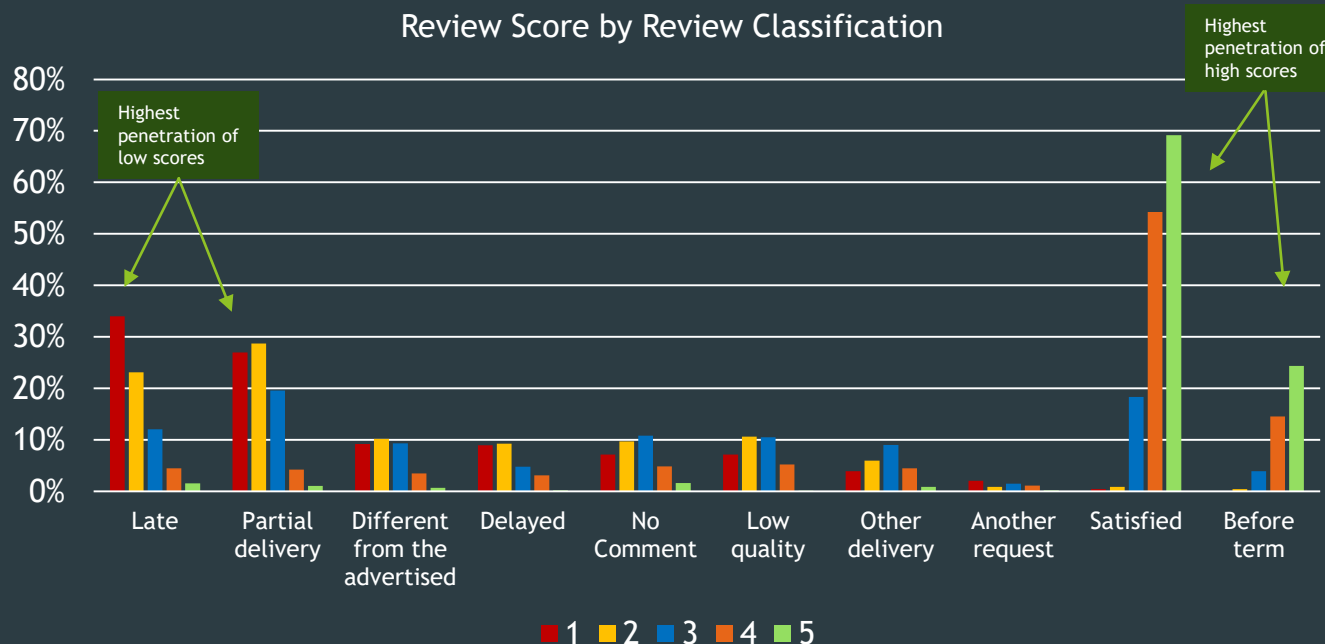
Delivery Time Effect on Customer Review



- The longer the delivery time, the lower the review score.
- Typically customers are better satisfied when receiving their product at a quicker rate

Understanding Customer Review Scores

- Below shows the percent of each review score across the different shipment outcome classifications
- For low review scores 1 & 2, **Late Deliveries** and **Partial Deliveries** have the highest penetration percent of these low scores.
- For high review scores 4 & 5, **Satisfied** and **Before Term** have the highest penetration of these scores.
- Overall, delivery outcome seems to have the largest impact on review scores depending if it's a positive or negative experience.



Definition Key
Late = Late delivery
Partial Delivery = Partial order was delivered
Delayed= Order was delayed
No Comment = No review as written
Low quality= customer thought product was low quality
Other delivery= order arrived with different delivery method
Another request= customer had to place another requested
Satisfied = satisfied customer
Before term= order came before expected delivery date

What factors cause someone to choose a product listing?

- ▶ Possible influences: Rating, Number of Photos, Price, Vendor ID
- ▶ 5,870 items from 1,442 vendors sold with more than one price point/photos/vendor ID combination
- ▶ For each product, measure the percentage sold per listing for each combination
- ▶ Linear regression to see what influences this percentage
 - ▶ Items sales percentage - Dependent variable
 - ▶ Rating, Number of Photos, and Price - Independent variables

What factors cause someone to choose a product listing?

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.892e-01	7.440e-03	52.321	< 2e-16	***
price	-4.691e-05	9.327e-06	-5.029	4.97e-07	***
rating	-3.955e-03	1.641e-03	-2.411	0.0159	*
numPics	6.852e-03	1.097e-03	6.246	4.31e-10	***



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.731e-01	3.270e-03	114.100	< 2e-16	***
price	-4.654e-05	9.327e-06	-4.990	6.12e-07	***
numPics	6.797e-03	1.097e-03	6.196	5.92e-10	***

- ▶ Price and Number of Pictures are significant contributors
 - ▶ As Number of Pictures goes up, sales go up
 - ▶ As Price goes down, sales go up
- ▶ Rating has a much weaker significance
 - ▶ Still Important!
 - ▶ Overall Olist reputation
 - ▶ Access to features like Amazon's Buy Box
 - ▶ Best avenue for customer feedback

What about vendor?

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.000	6.602	0.606	0.564
seller_id0b90b6df587eb83608a64ea8b390cf07	-2.000	13.204	-0.151	0.884
seller_id229c3efbfb0ea2058de4ccdfbc3d784a	-3.000	13.204	-0.227	0.827
seller_id6973a06f484aacf400ece213dbf3d946	14.000	13.204	1.060	0.324
seller_id88460e8ebdecfbecb5f9601833981930	7.333	9.337	0.785	0.458
seller_id8e6d7754bc7e0f22c96d255ebda59eba	8.000	10.439	0.766	0.469
seller_idf457c46070d02cadd8a68551231220dd	11.333	9.337	1.214	0.264

- ▶ For the products which had multiple vendors, we can look at each one individually and see if the vendor ID influences sales, again using linear regression
- ▶ Among products sampled, vendor ID rarely had any effect.
 - ▶ When it did, this effect was cancelled out when including price, meaning that the reason it seemed effective was because that vendor consistently posted lower prices, not due to brand recognition.

What are properties of successful vendors on Olist?

- ▶ 3,033 unique vendors, generating 13.4 million Brazilian Reals in revenue
 - ▶ \$3.4 million USD
- ▶ Top 100 vendors generate approximately half of that revenue!
- ▶ What makes them successful?
- ▶ Can we connect their success to customer profiles to determine the right approach for a new entry into the market?

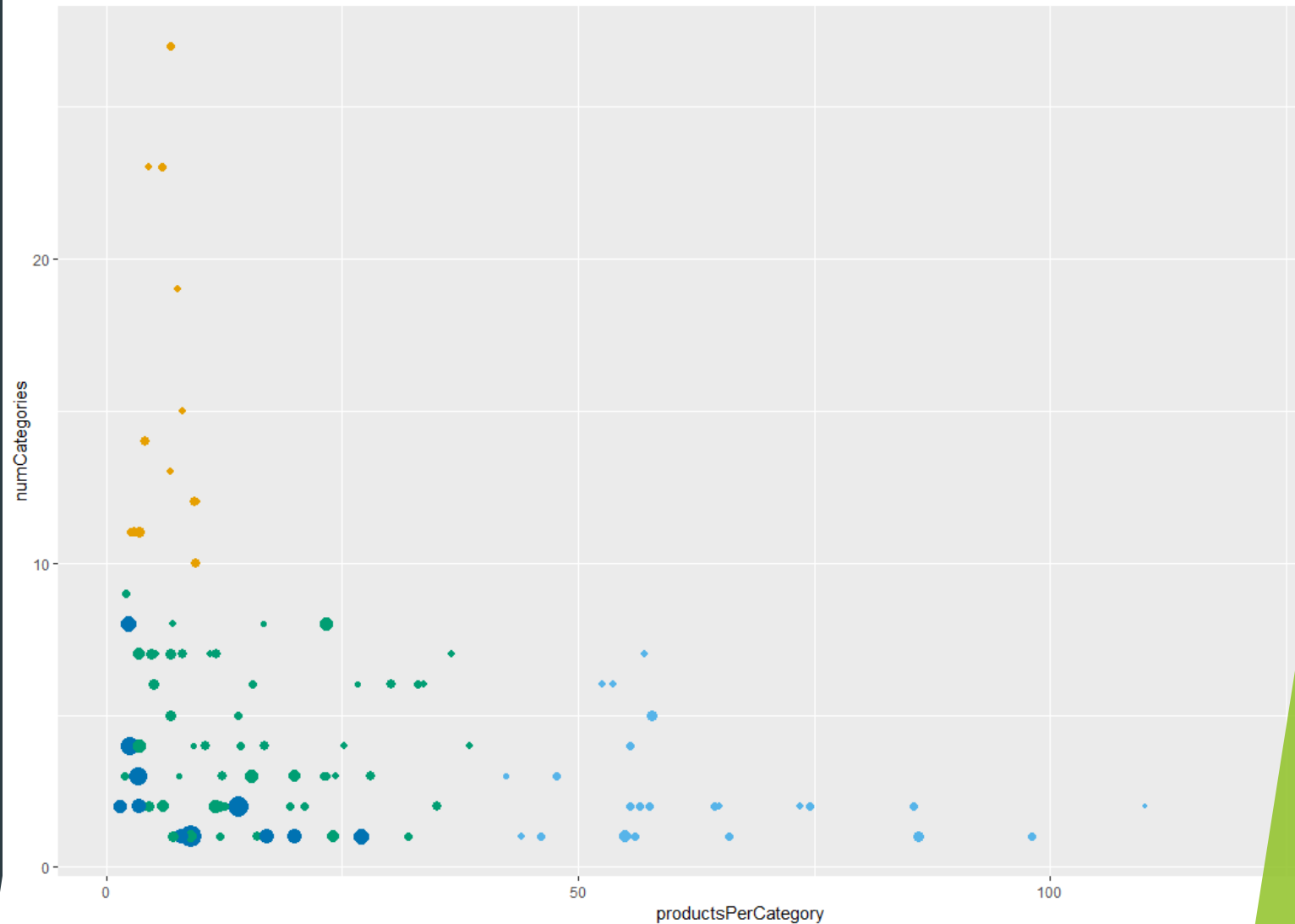
Seller Clustering

- Setup

- ▶ We will look at the top 100 sellers in our data set
 - ▶ As previously mentioned, these top 100 account for almost half the revenue from the entire data set of over 3000 sellers
- ▶ We will cluster them based on three variables:
 - ▶ Average Price - Are they selling high priced items more likely to be low volume? Or are they loading up on many low-priced items?
 - ▶ Number of Categories Listed In - How broad are their offerings? Do they stick to a few specialized categories or go all over the place?
 - ▶ Products per Category - How deep are their offerings? Are they saturating a category, or just dabbling in each one?
- ▶ Other variables were considered, but didn't offer much insight
 - ▶ Overall number of products - Weaker measure than normalizing by number of categories, as a seller with many categories also will have many items. Created collinearity.
 - ▶ Number of Photos and Avg Shipping Cost - Cluttering the data more and ultimately resulted in very similar clusters.

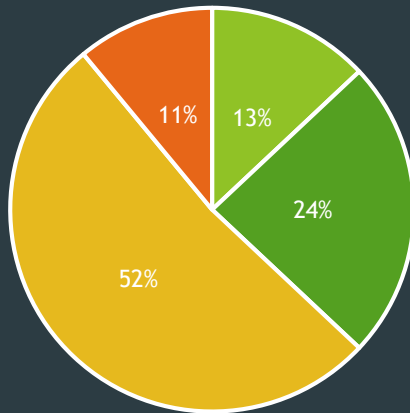
Seller Clustering - Results

- ▶ Size of dot is relative to Average Item Price
- ▶ Four distinct clusters, more details on future slides
- ▶ Cluster 1 - Lots of categories, not many products per category
- ▶ Cluster 2 - Small number of categories, lots of items per category
- ▶ Cluster 3 - Moderate number of products and categories, lower prices. Relies on number of sales.
- ▶ Cluster 4 - Small number of products and categories. Higher priced goods.

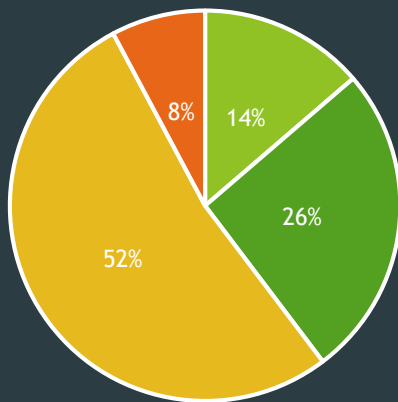


Seller Clusters - Result Verification

Vendor Distribution



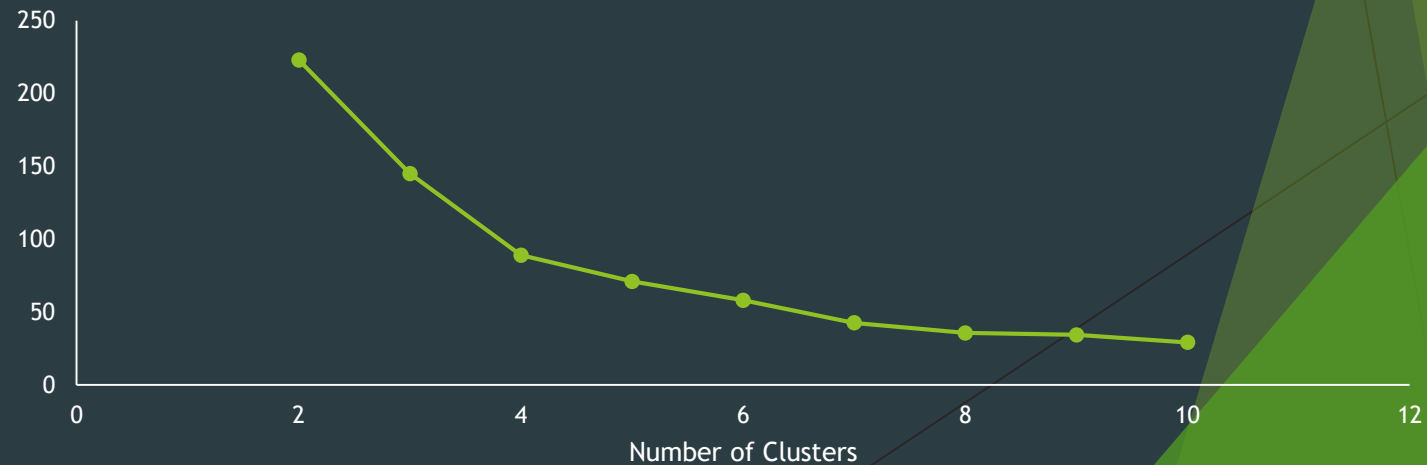
Revenue Distribution



■ 1 ■ 2 ■ 3 ■ 4

- ▶ Vendor and Revenue distributions are very similar to each other
 - ▶ Each cluster is an equal opportunity to make good revenue, no inherit advantages in any cluster
- ▶ Elbow Plot checks out

Elbow Plot



Cluster 1 - Unspecialized Businesses

- ▶ Cluster 1 contains vendors who are listed in 10 or more categories, but on average only have about 5-10 products per category.
- ▶ 13 of the top 100 sellers, representing 13.7% of revenue
- ▶ Lots of categories which lend themselves to cheaper/generic merchandise
 - ▶ Toys, Watches, Computer Accessories, Stationery, Health/Beauty
- ▶ Entertainment categories, more likely for impulse purchases or potentially scalpers of popular gifts
 - ▶ Sports/Leisure, Console Games, Toys
- ▶ Relies on lower prices or impulse purchases to drive volume in order to support revenue stream
- ▶ Easy access, but risky due to potentially small profit margins and large warehousing cost
 - ▶ Better for company with established supply chain

Most popular categories:

1. Toys - 123 products
2. Watches/Gifts - 122
3. Computer Accessories - 121
4. Sports/Leisure - 98
5. Console Games - 94
6. Perfume - 82
7. Telephone - 65
8. Stationery - 59
9. Houseware - 55
10. Health/Beauty - 52

Cluster 2 - Specialized Businesses

- ▶ Cluster 2 contains sellers who have upwards of 40 products in the categories they listed in. Most sellers here only list in 1-2 categories, maximum of 7
- ▶ 24 of the top 100 sellers, representing 26.0% of revenue
- ▶ Interesting mix of categories
 - ▶ Some higher end products like Furniture, Appliances, and Auto
 - ▶ Some category where brand depth can play a role like Bed/Bath, Health/Beauty
- ▶ Large number of products as each company is digging in on specific categories to own it
- ▶ Difficult barrier to entry, as it most likely requires an established brand or an existing portfolio of offerings in a category

Most popular categories:

1. Bed/Bath - 990 products
2. Furniture - 585
3. Sports/Leisure - 348
4. Health/Beauty - 339
5. Home Appliances - 292
6. Watches/Gifts - 284
7. Telephone - 217
8. Auto - 192
9. Cool Stuff - 112
10. Office Furniture - 86

Cluster 3 - Low-Priced Well-Rounded Vendor

- ▶ Cluster 3 contains sellers with a moderate amount of categories and products per category. Their products fall on the lower end of the price spectrum. These vendors make their high revenue by moving a lot of low-priced items.
- ▶ 52 of the top 100 sellers, representing 52.5% of revenue
- ▶ Wider range of categories, less fall off in the popularity of categories
 - ▶ Intuitive mix of potentially lower priced goods like computer accessories, toys, stationery, housewares
- ▶ Reasonable entry point, requires less knowledge/skill in certain category and since goods are low priced, extra stock won't be as costly
 - ▶ Would require some supply chain investment however, key here is to move a lot of product

Most popular categories:

1. Watches/Gifts - 464 products
2. Computer Accessories - 302
3. Sports/Leisure - 255
4. Furniture - 230
5. Toys - 222
6. Stationery - 209
7. Housewares - 194
8. Office Furniture - 146
9. Baby - 128
10. Cool Stuff - 120

Cluster 4 - Highly Specialized Vendors

- ▶ Cluster 4 has a small amount of categories and products per category. This cluster tends to sell a lower volume of higher priced goods to support their high revenue streams.
- ▶ 11 of the top 100 sellers, representing 7.8% of revenue
- ▶ Very small number of products, highly specialized
 - ▶ Luxury Goods: Expensive health/beauty products, watches
 - ▶ Special Use Goods: Appliances, Tools, Instruments
- ▶ Seems like the best place for an established company to come in and sell online, not great for a new company
 - ▶ Brand recognition, low demand items

Most popular categories:

1. Health/Beauty - 48 products
2. Watches/Gifts - 44
3. Small Appliances - 19
4. Home Appliances - 13
5. Construction Tools - 12
6. Musical Instruments - 8
7. Toys - 5
8. Cool Stuff - 4
9. Sports/Leisure - 2
10. Auto - 1

Conclusion - Selling on Olist

Price

- Need to be competitive with pricing in the market

Place

- Shipping - reduce delivery times and provide accurate estimations of delivery time
- Service a large area of Brazil, affecting delivery time
 - A lot of remote areas in Brazil

Promotion

- Marketing Communications - Photos
 - More photos, more products purchased
- Description of the product - meaningful and accurate content

Product

- Amount of products available
 - Wide range from food to electronics
- Low brand awareness

Connecting Customer To Business

- ▶ We can connect these seller clusters to our customer clusters in order to discover our best opportunity for starting a new business selling via Olist
- ▶ High Spenders Customer cluster matches well with the Highly Specialized Business cluster
 - ▶ However, as mentioned, this is not a great space for a new business to enter.
- ▶ Convenience Seeker Customer cluster matches well with the Unspecialized Business cluster
 - ▶ However, this is a small amount of customers (7%)
- ▶ Two largest customers segments combined for 59% of customers
 - ▶ House/Décor Customers: Buys furniture, garden tools, other home products
 - ▶ DIY/Electronics: Bed/Bath, Computer Accessories, Tools
- ▶ These two clusters match well with the Specialized and the Low-Priced Well-Rounded Businesses
 - ▶ Specialized - Often offers bed/bath, furniture, health/beauty, appliances
 - ▶ Low-Price Well-Rounded - Often offers computer accessories, furniture, housewares

Final Recommendation

- ▶ Low-Priced Well-Rounded Vendor is probably the best model to follow to maximize the probability of success of small businesses, as it has a lower barrier to entry than the Specialized Business model.
- ▶ Based on the overlap between our business cluster and our most popular customer clusters, our Low-Price Well-Rounded business should start by selling low-priced computer accessories, housewares, and smaller furniture items since these are good matches between customer demands and the typical items from these type of vendors
- ▶ Our Low-Priced Well-Rounded business could potentially expand into bed/bath and low-priced tools. These are items our target customers demand, but aren't as common from this type of vendor.
- ▶ Over time, evolving a business from Low-Price Well-Rounded to a Specialized Business seems like a reasonable plan as your selling firm finds categories that they excel in.