Marketing Funnel Analysis - Onlist

In this analysis, we look at **Olist** a small business e-commerce enabler ecosystem that specializes in the field of logistics and capital. Olist connects small businesses from all over Brazil to channels without hassle and with only a single contract. Those merchants are able to sell their products through the Olist Stores and ship them directly to the customers using Olist's logistics partners. Olist is leading the way as the #1 e-commerce enabler for small and medium businesses in Brazil, now expanding globally.

The dataset we used had 8,000 Marketing Qualified Leads (MQLs) that requested contact between June 1st 2017 and June 1st 2018. They were randomly sampled from the total of all MQLs. Its features allow viewing a sales process from multiple dimensions: lead category, catalog size, behavior profile, etc. This is real data that has been anonymized.

Marketing/Sales Data: https://www.kaggle.com/datasets/olistbr/marketing-funnel-olist
Seller's Data: https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce

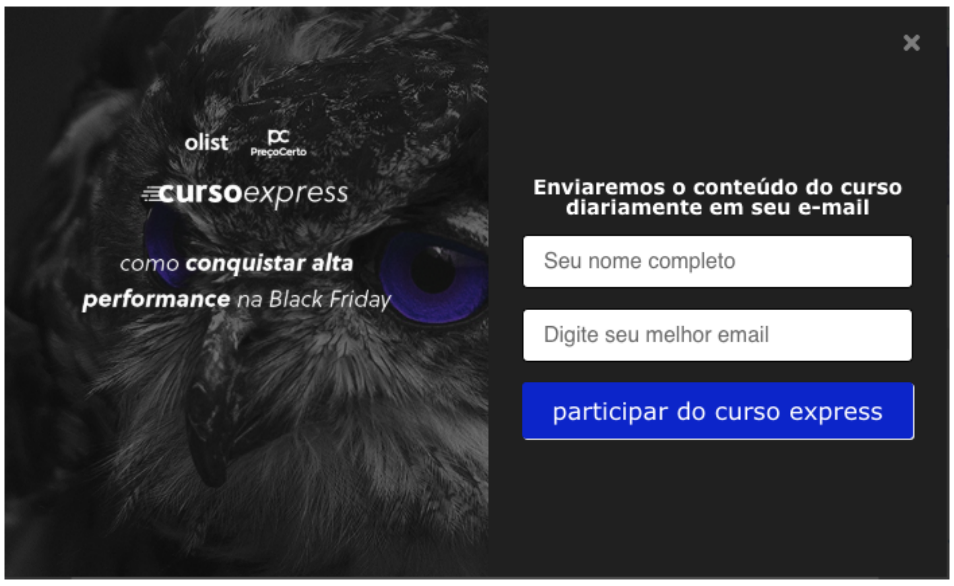
For this analysis, I will be performing some SQL queries using MySQL. I was able to gather this data from Kaggle and have already loaded in the different datasets. Below you will see each query and the code that was used.

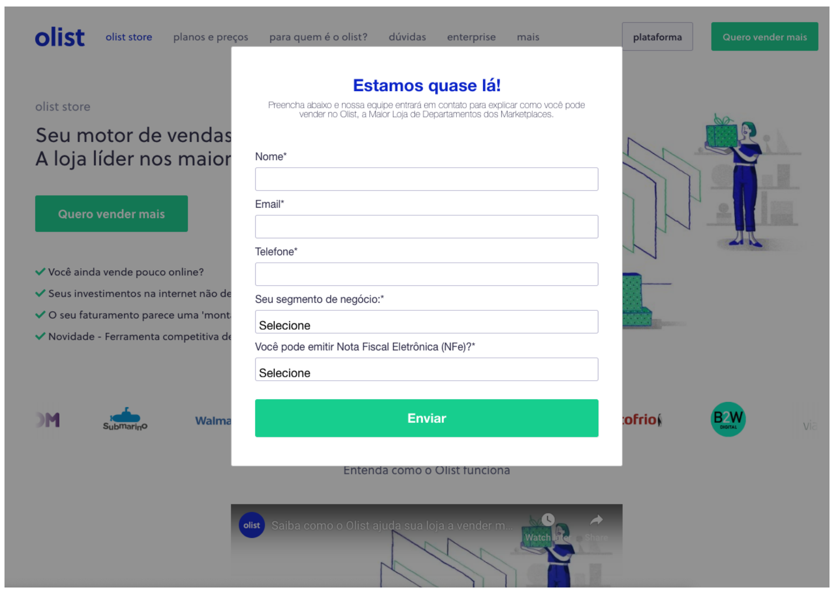
Sales/Marketing Cycle

A potential seller joins Olist through a marketing and sales funnel through the process below:

1. Sign-up on an Onlist landing page.
2. Get contacted by a Sales Development Representative (SDR), to confirm some information and schedule a consultancy.
3. Then a consultancy is made by a Sales Representative (SR).
4. The SR will try to close the deal (lead sign up) or lose the deal (lead leaves without sign up).
5. Lead becomes a seller and starts building their catalog on Olist.
6. Products are published on marketplaces and ready to sell.

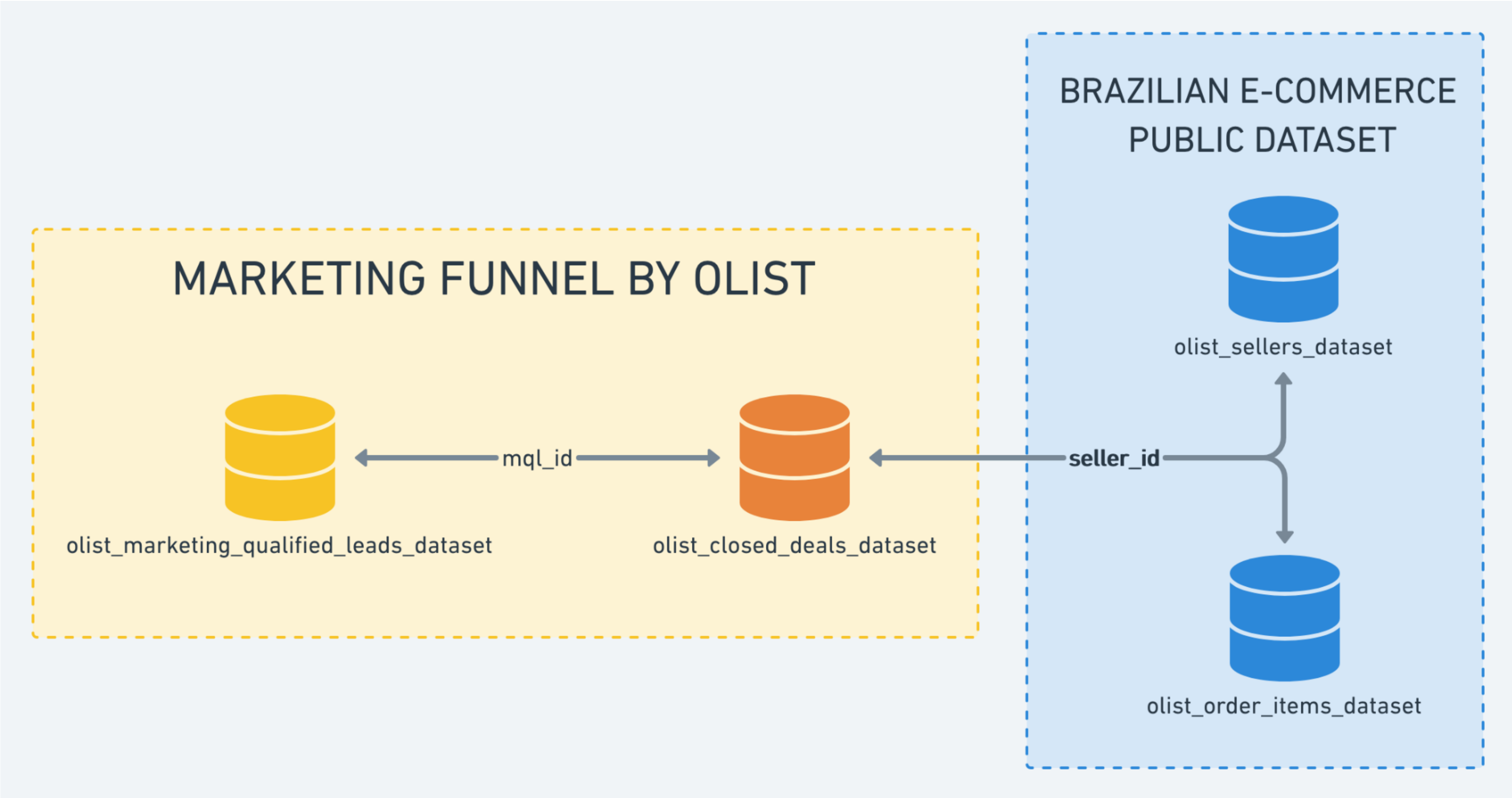
Note: A seller MQL might come from multiple sources (They might subscribe on two different landing pages, for instance).



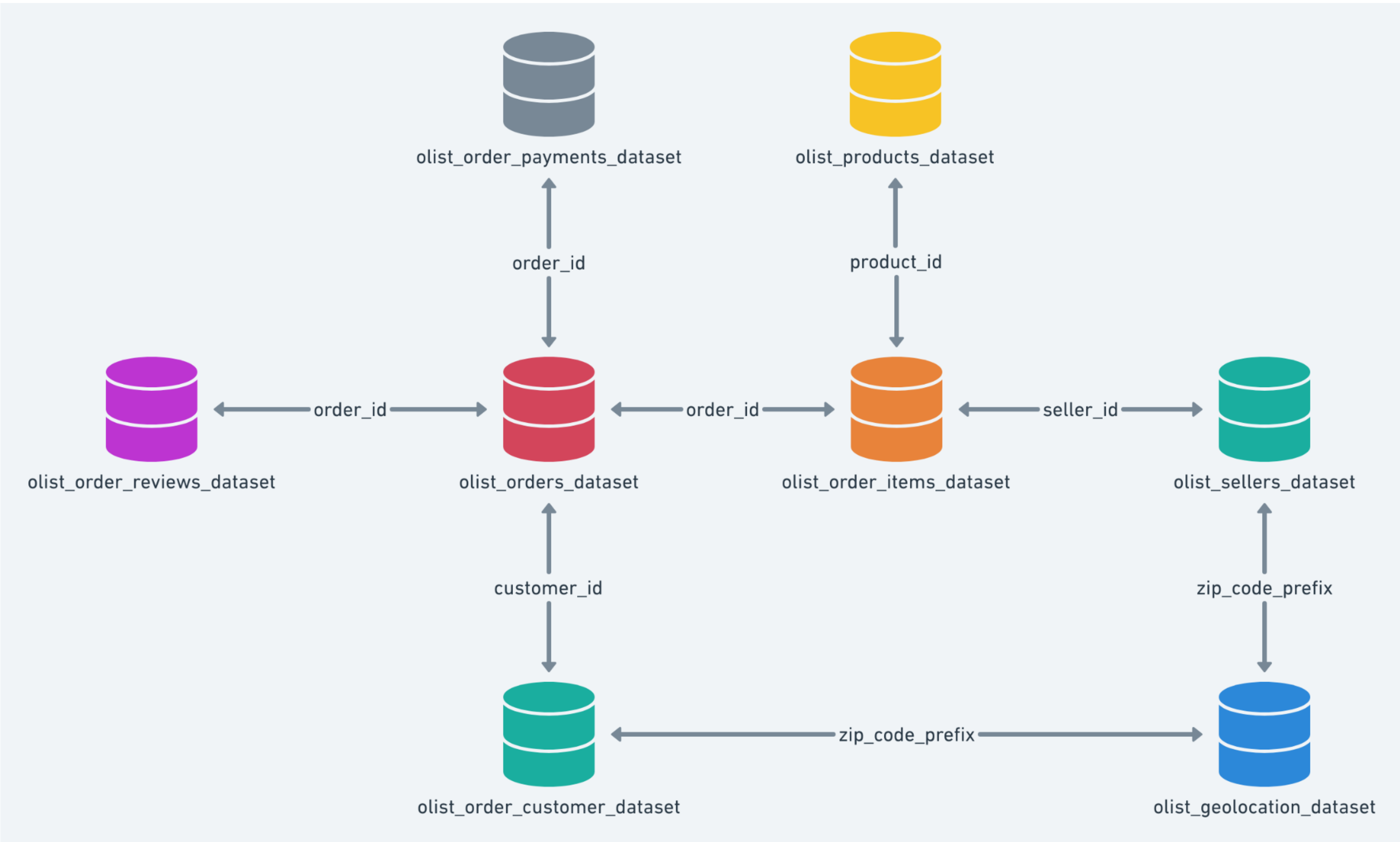


Data Schema

The data is divided in multiple datasets for better understanding and organization. By using both **mql_id** and **seller_id**, these datasets can be joined together.



In this project, we also queried results from our Onlist seller database to get a better understanding of our customers. There are a number of different tables that we can work with and join together to find insights as to what products are being sold, shipping/delivery times, and product review scores etc.



Marketing/Sales Funnel Analysis

Let's first get a baseline of how successful our current marketing/sales funnel was overall. We can do this by combining both the leads and sales tables.

MySQL Code

```
SELECT
COUNT(distinct a.mql_id) as total_leads,
SUM(CASE WHEN b.mql_id IS NOT NULL THEN 1 ELSE 0 END) as total_sales,
ROUND(SUM(CASE WHEN b.mql_id IS NOT NULL THEN 1 ELSE 0 END)/COUNT(*)*100,2) as perc_sales
FROM
olist_marketing_qualified_leads_dataset a
LEFT JOIN
olist_closed_deals_dataset b
ON
a.mql_id = b.mql_id;
```

Results

total_leads	total_sales	perc_sales
8000	842	10.53

We can see that based on the result, we have a total of 8000 leads and 842 sales with a conversion rate of 10.5%. The conversion rate doesn't seem too bad but has room for improvement as the average range varies from 2% to 20%. **Note:** These two tables are sampled data and don't show the entirety of the Onlist data source.

Now let's dig deeper by looking into our marketing leads and seeing how they are signing up to learn more about the Onlist's seller program. These are customers who signed up on Onlist's site to potentially become a seller on Onlist.

MySQL Code

```
SELECT
    origin,
    COUNT(distinct a.mql_id) as leads,
    ROUND(COUNT(distinct a.mql_id)/(SELECT COUNT(distinct mql_id) FROM olist_marketing_qualified_leads_dataset)*100,2) as perc_total_leads,
    COUNT(b.mql_id) AS total_sales,
    ROUND(COUNT(b.mql_id)*100/COUNT(a.mql_id),2) AS conversion_rate
FROM
olist_marketing_qualified_leads_dataset a
LEFT JOIN
olist_closed_deals_dataset b
ON
a.mql_id = b.mql_id
GROUP BY 1
HAVING origin NOT LIKE ''
ORDER BY 2 DESC;
```

Results

origin	leads	perc_total_leads	total_sales	conversion_rate
organic_search	2296	28.70	271	11.80
paid_search	1586	19.83	195	12.30
social	1350	16.88	75	5.56
unknown	1099	13.74	179	16.29
direct_traffic	499	6.24	56	11.22
email	493	6.16	15	3.04
referral	284	3.55	24	8.45
other	150	1.88	4	2.67
display	118	1.48	6	5.08
other_publicities	65	0.81	3	4.62

The highest percentage of leads are coming from our organic search, paid search and social channels. Onlist is investing their marketing budget in paid per click campaigns as well as social ads in order to increase the amount of qualified leads. Unfortunately, social media has very low conversion rates, which means that leads are not maturing. I would recommend changing the marketing strategy by either changing the content of the social ads or even cutting social out as a viable channel as most likely costs are adding up to even acquire these leads. Organic search also had a significant amount of traffic, which means SEO efforts seemed to be doing very well as customers are finding Onlist on search engines. About 13.74% of users are coming from unknown traffic. We would have liked to know what actually is "unknown" traffic but without proper categorization and the company's insider knowledge, this is hard to figure out. It had the highest conversion rate so something like this should be asked about to another member on the Onlist team. **Note: There was also missing data of 60 users that I didn't include in this table due to improper categorization.**

We should also see which industry these sellers are from in order to try and target similar users in the future. Understanding what their interests are can help us change the copy or visuals for social ads, which are currently struggling to convert new users. We can even dig deeper into the data to see if they are either resellers or manufacturers of a product. This information can be useful in trying to figure out how to approach conversations when sales representatives speak with them.

MySQL Code

```
SELECT
business_segment,
COUNT(mql_id) as total_sales,
ROUND(COUNT(distinct mql_id)/(SELECT COUNT(distinct mql_id) FROM olist_closed_deals_dataset)*100,2) as perc_sales,
SUM(CASE WHEN business_type = 'reseller' THEN 1 ELSE 0 END) as reseller,
SUM(CASE WHEN business_type = 'manufacturer' THEN 1 ELSE 0 END) as manufacturer
FROM
olist_closed_deals_dataset
GROUP BY 1
HAVING business_segment NOT LIKE ''
ORDER BY total_sales DESC;
```

Results

business_segment	total_sales	perc_sales	reseller	manufacturer
home_decor	105	12.47	45	59
health_beauty	93	11.05	67	25
car_accessories	77	9.14	67	10
household_utilities	71	8.43	48	23
construction_tools_house_garden	69	8.19	51	18
audio_video_electronics	64	7.60	58	6
computers	34	4.04	34	0
pet	30	3.56	24	6
food_supplement	28	3.33	18	10
food_drink	26	3.09	12	14
sports_leisure	25	2.97	18	5
bags_backpacks	22	2.61	12	10
bed_bath_table	22	2.61	15	7
toys	20	2.38	16	4
fashion_accessories	19	2.26	10	8
home_office_furniture	14	1.66	6	8
phone_mobile	13	1.54	12	0
stationery	13	1.54	9	4
handcrafted	12	1.43	4	6

We can see that most of our sellers are interested in these verticals: home decor, health beauty, car accessories and household utilities. These companies are mostly reselling their goods and are not manufacturers. Olist's business model is very similar to Amazon in which a lot of small businesses will tend to resell manufactured goods from China and place their labels on them. Amazon will then store the products and ship them to their customers. This is known as "dropshipping" and I theorize that this similar business model applies to our customers on Onlist. Targeting specific users that are interested in selling these products through paid ads, social and SEO efforts can help to increase sales overall.

Landing Page Analysis

How are our landing pages performing for the Onlist marketing/sales funnel? Are they successful in converting new users? We see that there are a lot of pages that Onlist uses to gather leads. Some of these pages have low traffic and conversions, which could mean that A/B tests were performed to figure out the highest converting pages. The top 2 results show a lot of leads being generated as well as some of the highest sales conversion rates. These 2 LPs might be what Onlist is currently using based on the number of total leads and sales being the highest. Hopefully, Onlist is not using the 3rd and 4th LPs on this table as conversion rates are low for both sales and leads. These pages might mean that they are not optimized, which is causing conversion rates to lower. This costs Onlist more money as it takes a higher marketing budget to achieve similar results as the 1st and 2nd LPs. In these situations, I recommend checking the copy on the website as well as buttons and organizing the site so that users aren't getting confused when first arriving on the page.

MySQL Code

```
SELECT
distinct landing_page_id,
COUNT(a.mql_id) as total_leads,
COUNT(b.mql_id) as total_sales,
ROUND(COUNT(b.mql_id)/COUNT(a.mql_id)*100,2) as total_conversion_rate
FROM
olist_marketing_qualified_leads_dataset a
LEFT JOIN
olist_closed_deals_dataset b
ON
a.mql_id = b.mql_id
GROUP BY 1
ORDER BY 2 DESC;
```

Results

landing_page_id	total_leads	total_sales	total_conversion_rate
b76ef37428e6799c421989521c0e5077	912	171	18.75
22c29808c4f815213303f8933030604c	883	174	19.71
58326e62183c14b0c03085c33b9fdc44	495	27	5.45
88740e65d5d6b056e0cda098e1ea6313	445	31	6.97
ce1a65abd0973638f1c887a6efcfa82d	394	59	14.97
40dec9f3d5259a3d2dbcdab2114fae47	330	67	20.30
f017be4dbf86243af5c1ebed0cff36a2	310	21	6.77
e492ee5eaf1697716985cc6f33f9cd9b	291	10	3.44
a7982125ff7aa3b2054c6e44f9d28522	156	18	11.54
73f31a40697cc90a86c1273563ac230e	115	4	3.48
241f79c7a8fe0270f4fb79fcbbcd17ad	109	14	12.84
65d9f9d71e562365e8b44037c2888d98	95	5	5.26
87732658ba41d8775e8577df347a64db	92	2	2.17
1722481ac9e5371e5099dea226b5421d	86	2	2.33
1ceb590cd1e00c7ee95220971f82693d	71	9	12.68
7fa6214d82e911d070f51ef79381b956	68	11	16.18
b6885f18d203a61176418c1fb3764815	67	7	10.45

Sales Representative Analysis

Sales representatives are in charge of talking to qualified leads in order to get them to sign up onto Onlist as a seller. We want to see if we can try to find ways in order to improve sales conversion efficiency so that we can onboard more sellers. Who was the most successful in getting the most amount of sales as well as doing it the quickest?

MySQL Code

```
SELECT
sr_id,
COUNT(*) as total_sales,
AVG(datediff(won_date, contact_date)) as avg_sale_time
FROM
olist_marketing_qualified_leads_dataset a
INNER JOIN
olist_closed_deals_dataset b
ON
a.mql_id = b.mql_id
GROUP BY 1
ORDER BY 2 DESC;
```

Results

sr_id	total_sales	avg_sale_time
4ef15afb4b2723d8f3d81e51ec7afefe	133	30.4887
d3d1e91a157ea7f90548eef82f1955e3	82	46.6463
6565aa9ce3178a5caf6171827af3a9ba	74	25.7838
85fc447d336637ba1df43e793199fbc8	64	33.7031
495d4e95a8cf8bbf8b432b612a2aa328	63	48.5873
2695de1affa7750089c0455f8ce27021	59	48.3898
fbf4aef3f6915dc0c3c97d6812522f6a	59	21.9661
de63de0d10a6012430098db33c679b0b	53	61.7358
9ae085775a198122c5586fa830ff7f2b	51	41.7059
c638112b43f1d1b86dcabb0da720c901	36	22.3889
060c0a26f19f4d66b42e0d8796688490	32	25.2188
068066e24f0c643eb1d089c7dd20cd73	27	122.3333
a8387c01a09e99ce014107505b92388c	26	96.6538
9e4d1098a3b0f5da39b0bc48f9876645	24	17.0000
56bf83c4bb35763a51c2baab501b4c67	24	55.1250
34d40cdaf94010a1d05b0d6212f9e909	10	172.0000
4b339f9567d060bcea4f5136b9f5949e	9	180.5556
9749123c950bf8363ace42cb1c2d0815	7	234.0000
9d12ef1a7eca3ec58c545c678af7869c	6	214.3333
0a0fb2b07d841f84fb6714e35c723075	1	306.0000
b90f87164b5f8c2cfa5c8572834dbe3f	1	175.0000
6aa3b86a83d784b05f0e37e26b20860d	1	321.0000

The first user '4ef15afb4b2723d8f3d81e51ec7afefe' had 133 total sales and it took on average 30 days for a person who signed up on the website to turn into a sale. This avg_sale_time is based on each of the sales representatives. Some of these sales reps are more skilled at closing than others based on average sale time. This analysis would have been more clearer if we also had data on who was assigned to each lead. This would have given us details on the success rate of each representative and would have been a better metric overall to compare each of their performances.

Seller's Analysis

Let's now take a look at how our sellers on Onlist are performing and what they are selling. We can aggregate total orders, total revenue, average price and freight costs.

MySQL Code

```
SELECT
business_segment,
b.seller_id,
COUNT(order_id) as total_orders,
ROUND(SUM(price),2) as total_price,
ROUND(AVG(price),2) as avg_price,
ROUND(AVG(freight_value),2) as avg_freight
FROM
olist_order_items_dataset a
INNER JOIN
olist_closed_deals_dataset b
ON
a.seller_id = b.seller_id
GROUP BY 1,2
ORDER BY 3 DESC;
```

Result

business_segment	seller_id	total_orders	total_price	avg_price	avg_freight
watches	7d13fca15225358621be4086e1eb0964	578	113628.97	196.59	14.93
health_beauty	c70c1b0d8ca86052f45a432a38b73958	338	36537.37	108.10	11.65
home_appliances	612170e34b97004b3ba37eae81836b4c	110	23065.02	209.68	20.97
bed_bath_table	8a432f4e5b471f8da497d7dc517666e2	106	7555.00	71.27	18.54
pet	70c27847eca8195c983ed7e798c56743	98	12077.52	123.24	14.81
health_beauty	f46490624488d3ff7ce78613913a7711	91	5802.85	63.77	19.21
books	ba143b05f0110f0dc71ad71b4466ce92	86	6110.58	71.05	15.62
home_decor	dfa0c4c6229ab200a4a1336b4d7128ff	86	4777.10	55.55	20.55
health_beauty	516e7738bd8f735ac19a010ee5450d8d	84	6643.09	79.08	17.46
sports_leisure	d566c37fa119d5e66c4e9052e83ee4ea	73	5145.70	70.49	13.41
household_utilities	4c18691b6037662be2df78a765d98ab5	67	3195.52	47.69	16.73
home_office_furn...	cc63f0dd2acba93ffed4fe9f8e0321fa	63	7096.71	112.65	22.22
toys	d9a84e1403de8da0c3aa531d6d108ba6	59	1425.20	24.16	12.41
household_utilities	138dbe45fc62f1e244378131a6801526	58	1125.56	19.41	11.88

Seller id '7d13fca15225358621be4086e1eb0964' is the most successful business in terms of total revenue and total orders. Even the average price for each item is higher than most of the other sellers. Watches tend to be a high priced item so being able to produce the top amount of sales is impressive since they cost more money to buy than other items. In this instance, I would try to investigate more into this seller's marketing strategy as well as their inventory to see what specific watches they are selling. Using this information could be beneficial as Onlist might want to partner up with them in order to try and sell more of their watches. Watches, health and beauty, and home appliances are the most successful businesses and from an earlier analysis, these types of sellers are who Onlist should be targeting or even trying to convert. As those verticals tend to produce the most amount of sales per item.

Payment Analysis

How do customer's pay for seller's items? This analysis might be useful for sellers who only accept certain payment types rather than others. As well as seeing how much they spend on average per payment type.

MySQL Code

```
SELECT
payment_type,
ROUND(AVG(payment_value),2) as avg_payment_amt,
ROUND(AVG(payment_installments),2) as avg_payment_installment,
ROUND(AVG(payment_sequential),2) as avg_payment_methods
FROM
olist_order_payments_dataset
GROUP BY 1
ORDER BY 2 DESC
```

Result

payment_type	avg_payment_amt	avg_payment_installment	avg_payment_methods
credit_card	163.32	3.51	1.00
boleto	145.03	1.00	1.00
debit_card	142.57	1.00	1.03
voucher	65.70	1.00	2.60
not_defined	0.00	1.00	1.00

We can see that the average payment amount is highest for credit cards and even has the most amount of installments. This means that when people use credit cards, they spend more and spread their payments out. Financing often is beneficial for the companies that sell these products as they earn interest depending on how long customer's keep an outstanding debt. Customers tend to spend more on credit cards than other payment types, maybe due to the fact that during the point of sale, they have access to more money than they usually would have. In the table, vouchers are more likely to use another payment method, which makes sense because most likely you can't buy products with a single voucher.

Shipping and Delivery Analysis

Were items being shipped quickly for users that got their items delivered and users that canceled their deliveries? Which part of the shipping process took the longest? The reason why we want to improve shipping times is because we want to limit the amount of canceled items to improve revenue, and customer satisfaction. We can do this by looking at the average time for each step of the delivery process. Here are some questions I wanted to explore.

- How long did it take to ship items?
  - From the initial purchase time to the actual delivered date?
  - From initial purchase time to the shipping carrier?
  - Shipping carrier to customer's house?
  - Was the delivered date close to the estimated delivery date?
- Why did people cancel orders, was it due to expectations not being met?
- Did people cancel orders due to high freight costs or long wait times?

Users with Successful Deliveries

Results

purchase_to_carrier	purchase_to_approved	carrier_to_delivered	purchase_to_delivered	delivered_to_estimated_delivered	total_orders	avg_freight_cost
2.80	0.27	8.73	12.01	11.11	110197	19.95

MySQL Code

```
SELECT
ROUND(AVG(TIMESTAMPDIFF(Day,order_purchase_timestamp,order_delivered_carrier_date)),2) AS purchase_to_carrier,
ROUND(AVG(TIMESTAMPDIFF(Day,order_purchase_timestamp,order_approved_at)),2) AS purchase_to_approved,
ROUND(AVG(TIMESTAMPDIFF(Day,order_delivered_carrier_date,order_delivered_customer_date)),2) AS carrier_to_delivered,
ROUND(AVG(TIMESTAMPDIFF(Day,order_purchase_timestamp,order_delivered_customer_date)),2) AS purchase_to_delivered,
ROUND(AVG(TIMESTAMPDIFF(Day,order_delivered_customer_date,order_estimated_delivery_date)),2) AS delivered_to_estimated_delivered,
COUNT(a.order_id) AS total_orders,
ROUND(AVG(freight_value),2) AS avg_freight_cost
FROM
olist_orders_dataset a
INNER JOIN
olist_order_items_dataset b
ON
a.order_id = b.order_id
WHERE
order_status = 'delivered';
```

For this SQL query, I calculated the average difference in time between each step of the shipping process in order to see how long each of them lasted. The most amount of time spent was having the shipping carrier take the package to the customer's home. On average the whole process takes about 12 days, which seems much longer compared to other companies like Amazon that have similar business practices as Olist. **The estimated time for what Onlist had compared to the actual time was very alarming.** It had a difference of 11 days which means that most users were waiting 11 more days than what Olist told the customer of when they should expect their package.

Users with Canceled Deliveries

Results

purchase_to_carrier	purchase_to_approved	carrier_to_delivered	purchase_to_delivered	delivered_to_estimated_delivered	total_orders	avg_freight_cost
3.08	0.35	11.57	18.00	29.43	542	19.65

For canceled deliveries, we see a very high wait time for customers getting their packages even compared to the delivered items. On average these customers had to wait 5 more days than the average person that received their packages through Onlist. The estimated wait time vs the actual delivery time was also much longer than previously at 29 days. If the estimates were off by this much, we should either not report them to the customers or try to find ways to get a more accurate estimate as well as trying to speed up the whole process. Customers are canceling their orders due to long wait times and not other metrics like freight costs as it didn't appear to change monetarily compared to delivered packages.

Low Reviewed Items vs Top Reviewed Items

Was there a trend for top reviewed items meeting customer's expectations based on how many photos there were on the sales page? As well as average character length for the description of each item? In this analysis, we want to see if providing more details on each item through photos and item summary can meet expectations for each reviewer.

Result

review_score	AVG(product_photos_qty)	AVG(product_description_lenght)	AVG(product_name_lenght)
1	2.1622	704.3784	48.7838
2	1.3750	825.6250	51.3750
3	2.1333	858.2333	48.8000
4	2.3077	953.8974	49.3333
5	2.1043	804.3681	48.1534

MySQL Code

```
SELECT
review_score,
AVG(product_photos_qty),
AVG(product_description_lenght),
AVG(product_name_lenght)
FROM
olist_order_items_dataset a
INNER JOIN
olist_order_reviews_dataset b
ON
a.order_id = b.order_id
INNER JOIN
olist_products_dataset c
ON
a.product_id = c.product_id
WHERE

order_item_id = 1
GROUP BY 1
ORDER BY 1 ASC;
```

Unfortunately, there wasn't a clear trend if more photos or longer description led to higher review scores. Which means that the quality of the items itself are the most significant indicator for higher review scores. On average, most items had at least 2 photos and about 800 characters for each of the descriptions regardless of the review score.

Did the amount of photos on a sales page affect total sales at all?

Result

product_photos_qty	total_orders	avg_review_score	total_unique_products
1	48049	4.0268	15852
2	19176	4.0000	6081
3	11198	3.9722	3772
4	7572	3.9500	2366
5	4975	4.6250	1452

MySQL Code

```
SELECT
DISTINCT product_photos_qty,
COUNT(a.order_id) AS total_orders,
AVG(review_score)
FROM
olist_order_items_dataset a
INNER JOIN
olist_order_reviews_dataset b
ON
a.order_id = b.order_id
INNER JOIN
olist_products_dataset c
ON
a.product_id = c.product_id
WHERE
order_item_id = 1
GROUP BY 1
ORDER BY 2 DESC;
```

Photos didn't tend to influence the amount of sales or review scores for each item sold. We can see that total sales were high for items that had 1 photo but this is because there are a lot of unique products with only 1 photo. The review score didn't change much from items with 1-4 photos but did see a boost with 5 photos. Most likely there is bias here because of the small amount of total purchases with 5 photos that influenced its average score.

Customer Survey Participation

Now let's look at customer surveys. How fast do customer's answer surveys based on the items that they received? Did higher scores indicate if customers were more willing to answer surveys?

MySQL Code

```
SELECT
review_score,
ROUND(AVG(TIMESTAMPDIFF(DAY,review_creation_date,review_answer_timestamp)),2) as avg_diff_hours_for_reviews
FROM
olist_order_reviews_dataset a
INNER JOIN
olist_order_items_dataset b
ON
a.order_id = b.order_id
WHERE
order_item_id = 1
GROUP BY 1
```

Results

review_score	avg_diff_hours_for_reviews
1	1.76
2	1.63
3	5.00
4	2.21
5	2.73

We can see an interesting trend that the lower the item's review score, the faster a customer is willing to fill out a survey. This might mean that customer's that had low satisfaction with the items they purchased were more willing to tell Onlist about it. While products that customers were satisfied with took longer to fill out a survey. The longest time to fill out a survey were products that scored a 3. Which might mean that people who were neutral were most likely to take the longest to fill out a survey

Conclusion

Looking at both the marketing/sales data as well as the seller's data gave me a really good understanding of Onlist's users and business strategy. Onlist had a high overall sales conversion rate for getting new sellers onto the platform but I believe there could be big improvements that would help Onlist grow their company. On the marketing side, they do well with PPC ads and organic traffic but can improve upon things like social and email. Targeting specific groups of people that are interested in selling things like home decor, health beauty, car accessories and household utilities in their ads could help with conversion rates. Another way to help with conversion rates is to look at the copy/images on landing pages and ads in order to help convert more users. Some of the LPs had very low conversions, so figuring out why that is, would help to improve metrics. Shipping/delivery times are one of the most alarming things that need to be fixed. On average, customers are canceling orders due to long wait times. The estimated delivery times are off from the actual delivery date by 11 days, so trying to work with shipping carriers to smooth this process would help to create better customer satisfaction.

Onlist can become successful if they follow their American counterpart, Amazon, by trying to deliver the best customer satisfaction they can. Based on the metrics that I pulled, they still are not hitting the standards that Amazon is hitting but that is understandable due to the size of Amazon. Hopefully by doing more specific targeting to get better qualified leads, they can get more new sellers onto Onlist. As well as fixing shipping times to increase customer satisfaction for products they purchase.