Sales Data Analysis for e-commerce

Anastasios Vitsios

March 2021

Table of Contents

Executive Summary

In this analysis we will examine the sales and marketing performance of the company Olist.

Olist has a web platform from where any Brazilian company can sell their products in Brazil. They started their activity in September 2017 and the data that we possess are for 24 months (until August 2018).

In this analysis we will see that there was an overall increase in sales in these 2 year with some spikes and some drops.
The big spike was in November 2017 and it was not related to the marketing of Olist. After this month there was an increase in marketing qualified leads which affected the sales positively.

After May 2018 there was a decline in sales in almost all States in Brazil and in the majority of product categories. Even though the sellers (companies which sell products through Olist) were increased after May 2018, the increase in sales was not possible. We don't have marketing data for the months after May 2018 to compare but it is true that in 2018 there was a 9% decline in GDP per capita which may affect the sales performance of Olist.

We will also realize that there are many States in Brazil where the sales are too low compared to the population, so we can expect increase in sales in those States with the proper strategic plan.

Introduction

Olist is the largest department store in Brazilian marketplaces. Olist has a web platform where small businesses from all over Brazil are able to sell their products through the Olist Store and ship them directly to the customers using Olist logistics partners.

The data that we have is from kaggle.com. They are for 24 months, from September 2016 to August 2018. The first 12 months consider as the Year 1 (Sep 2016 - Aug 2017) and the rest of months as the Year 2 (Sep 2017 to Aug 2018)

In total, in those 24 months Olist had:
- 99,441 orders
- 112,649 items sold, from
- 71 product categories, in
- 29 States, from
- 2383 sellers and
- 13,591,498 BR total sales

The analysis was made in files olist_all.xlsx, olist_groups1.xlsx, mrkt_leads.xlsx, .IPYNB

Methodology

I used 7 files from Kaggle.com which are: olist_custome_dataset.csv, olist_order_items_dataset.csv, olist_orders_dataset.csv, olist_products_dataset.csv, olist_sellers_dataset.csv, product_category_name_translation.csv

For the purposes of this analysis I downloaded 3 more tables from the web with the state population, state names/abbreviations and GDP per capita.

The analysis was made:
in the excel files olist_all.xlsx, olist_groups1.xlsx, mrkt_leads.xlsx,
in python file Olist_groupby.IPYNB and
in Power BI file olist.pbix

I created tables in excel by joining the tables from Kaggle.com, with the appropriate format in order to be analysed in Power BI.

I grouped by year, month, product category, seller some tables with python in order to aggregate the sales for further analysis.

The first month for each period is September. So, the first year (from now on Year_1) starts in September 2016 and ends in August 2017. The second year (from now on Year_2) starts in September 2017 and ends in August 2018.

Initial View – Sales for the 2 first years

The total revenue from Sep 2016 to Aug 2018 in Brazilian real (BRL) ➡ **13.59M**
Sales in BRL

Sales in BR and Units sold by Date

● Sales in BRL ● Units sold

Sales in BRL by Period

Period ● Year_2 2017-18 ● Year_1 2016-17

3.16M (23.27%)

10.43M (76.73%)

| Period	Sales	Units	Sales growth	Unit growth
Year_1 (09/2016-08/2017)	3,162,786.24	26108		
Year_2 (09/2017-08/2018)	10,428,712.46	86541	2.30	2.31
Total	**13,591,498.70**	**112649**	**2.30**	**2.31**
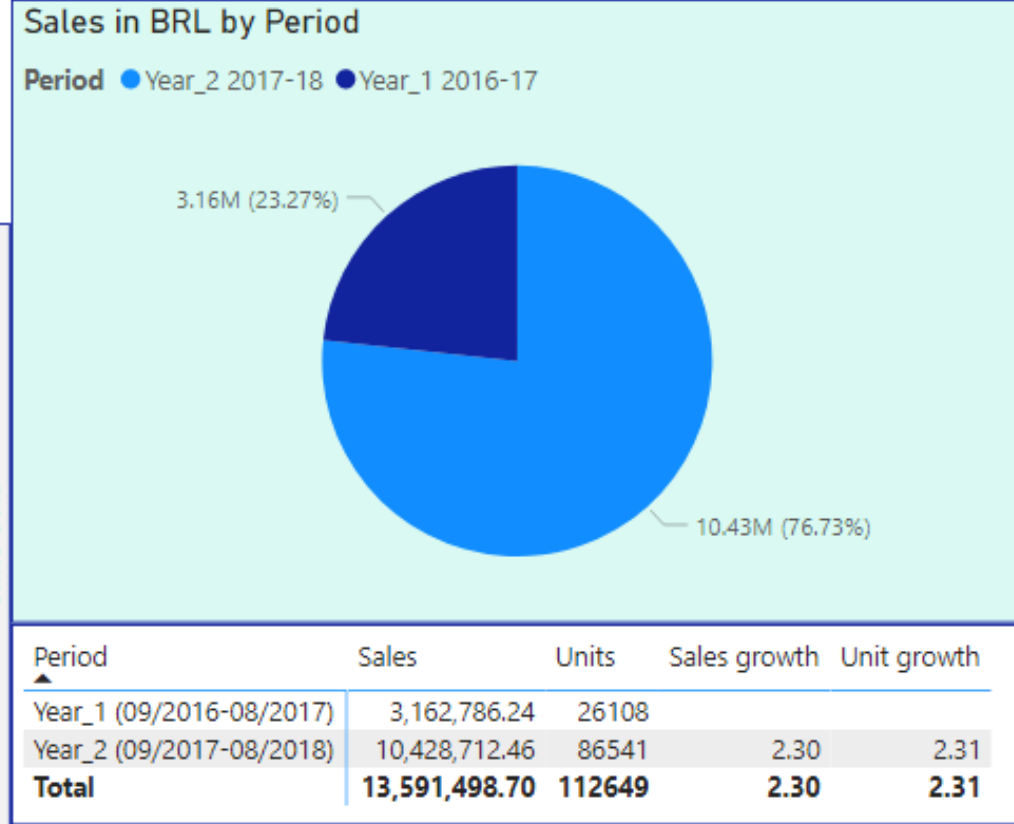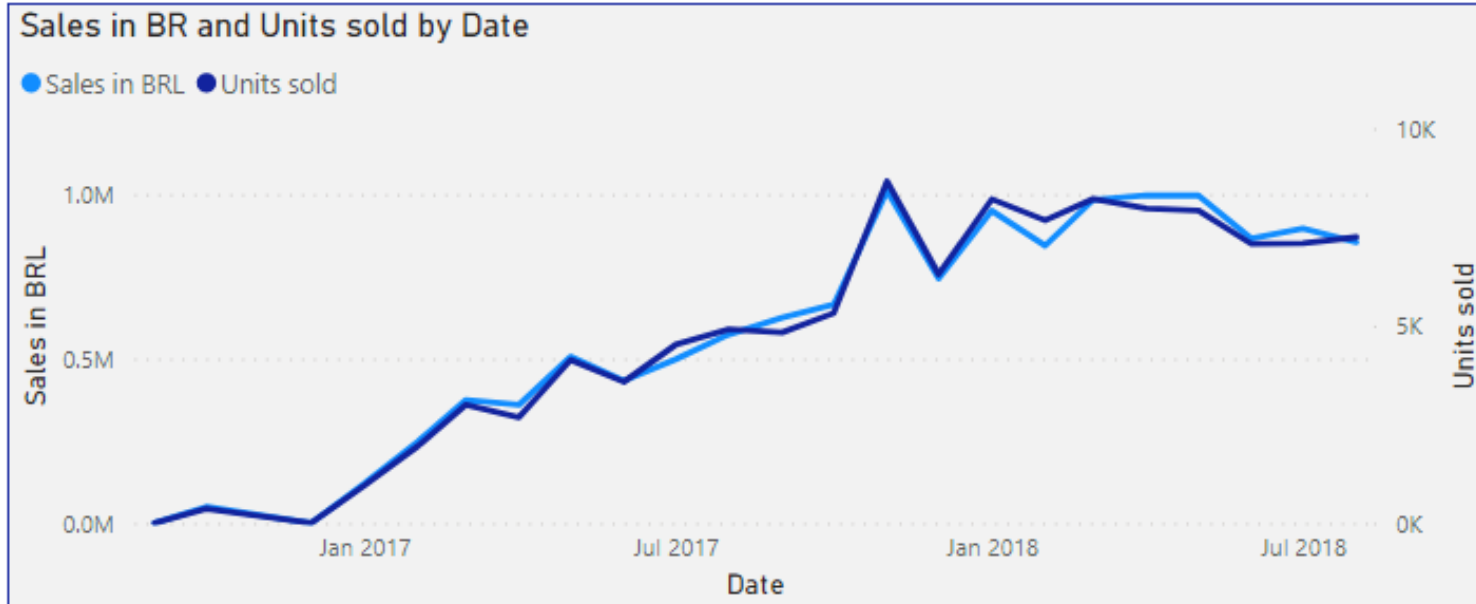
Olist started to sell on September 1st 2016. The data that we have is until August 31st 2018. I split the 24 months sales period in 2 annual periods. The 1st is from 9/1/2016 to 8/31/2017 (named "Year_1") and the 2nd is from The 1st is from 9/1/2017 to 8/31/2018 (named "Year_2") .
In Year_2 there is an increase or 230% in revenue and 231% in units and as we see in the graph "Sales in BR and Units sold" the growth between revenue and units is identically.

- The 76.73% of total sales became in Year_2.
- **There is a spike in November 2017 and a decline in sales after May 2018 which we will examine in the following slides**
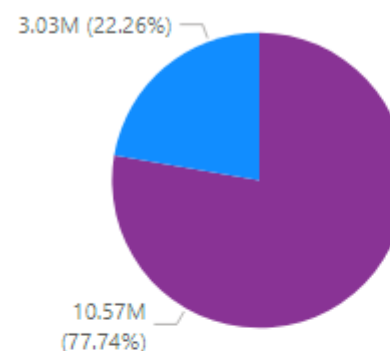
Sales in Top 6 States and Top 15 Products Categories
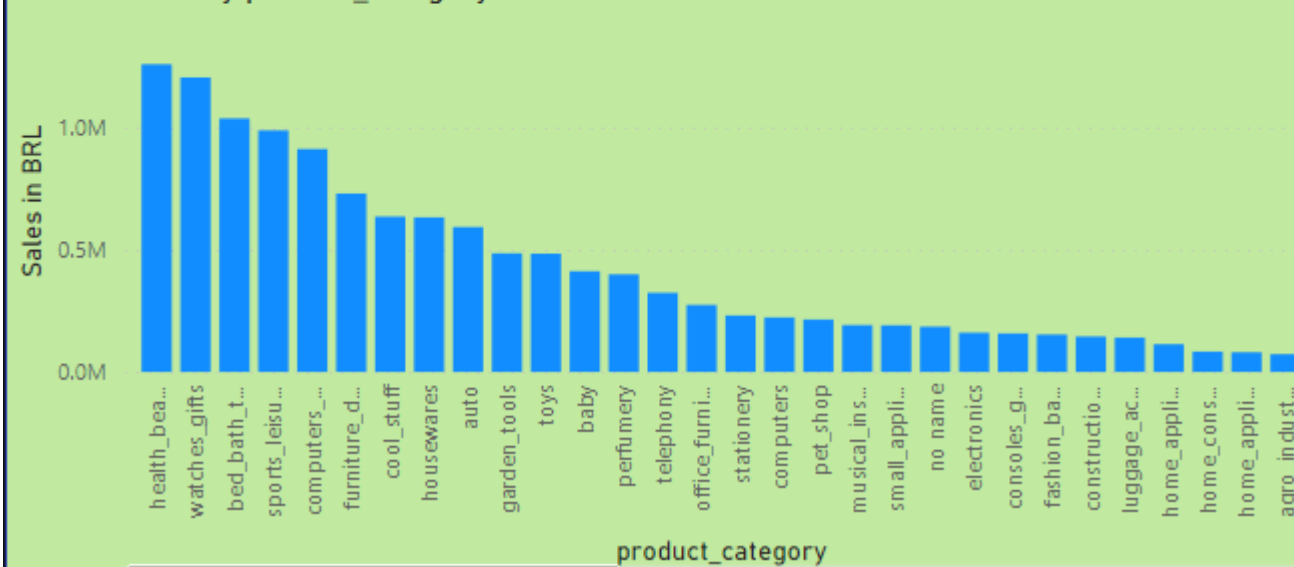

Sales in BRL by State
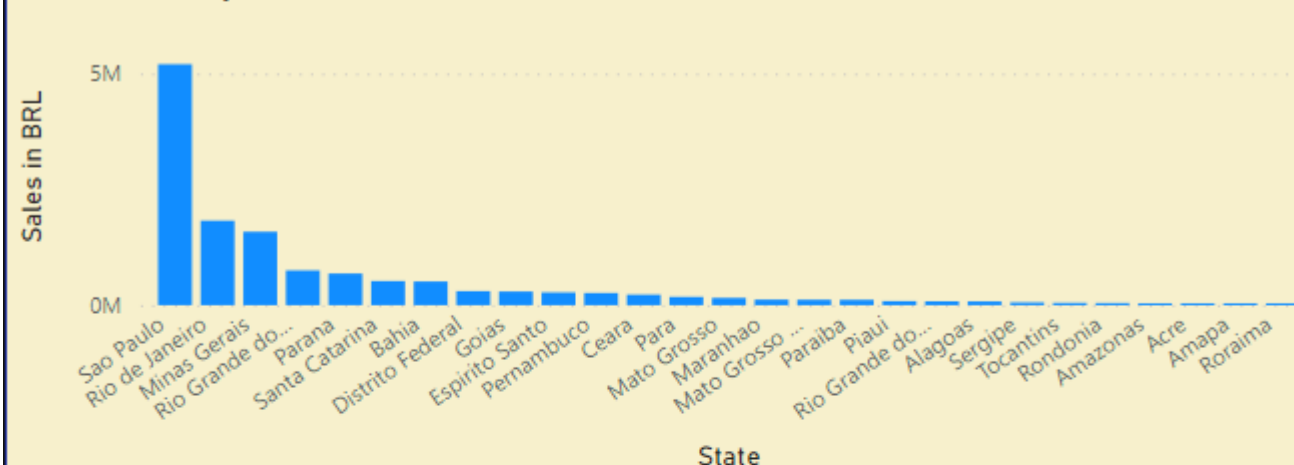
Sales by State_groups 27 States
State_gr... ● Top 6 States ● Other 21 States

3.03M (22.26%)

10.57M (77.74%)

Top 6 States - Sales

State	Sales in BRL
Sao Paulo	5,202,810.05
Rio de Janeiro	1,824,092.67
Minas Gerais	1,585,308.03
Rio Grande do Sul	750,304.02
Parana	683,083.76
Santa Catarina	520,553.34
Total	**10,566,151.87**

Sales in BRL by product_category

Sales by product_groups 71 Categories
product... ● Top 15 Categories ● Other 56 Cat...

3.22M (23.71%)

10.37M (76.29%)
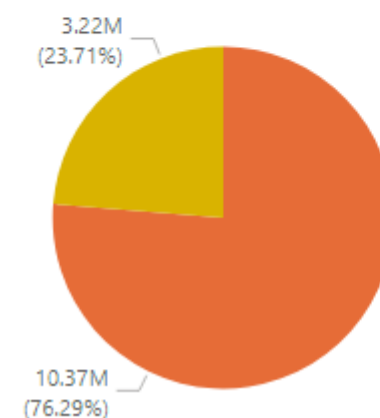
Top 15 Product Categories - Sales

product_category	Sales in BRL
health_beauty	1,258,681.34
watches_gifts	1,205,005.68
bed_bath_table	1,036,988.68
sports_leisure	988,048.97
computers_accessories	911,954.32
furniture_decor	729,762.49
cool_stuff	635,290.85
housewares	632,248.66
auto	592,720.11
garden_tools	485,256.46
toys	483,946.60
baby	411,764.89
perfumery	399,124.87
telephony	323,667.53
office_furniture	273,960.70
Total	**10,368,422.15**

The sales became in 27 States in Brazil and in 71 different product categories. Here we will analyse the sales performance in the **top 6 States** and in the top **15 product categories** where Olist has **77.74%** and **76.29%** of total sales respectively.

7

Sales between Year_1 and Year_2 - Spike in November 2017

Year_1 2016-17 and Year_2 2017-18 by Month
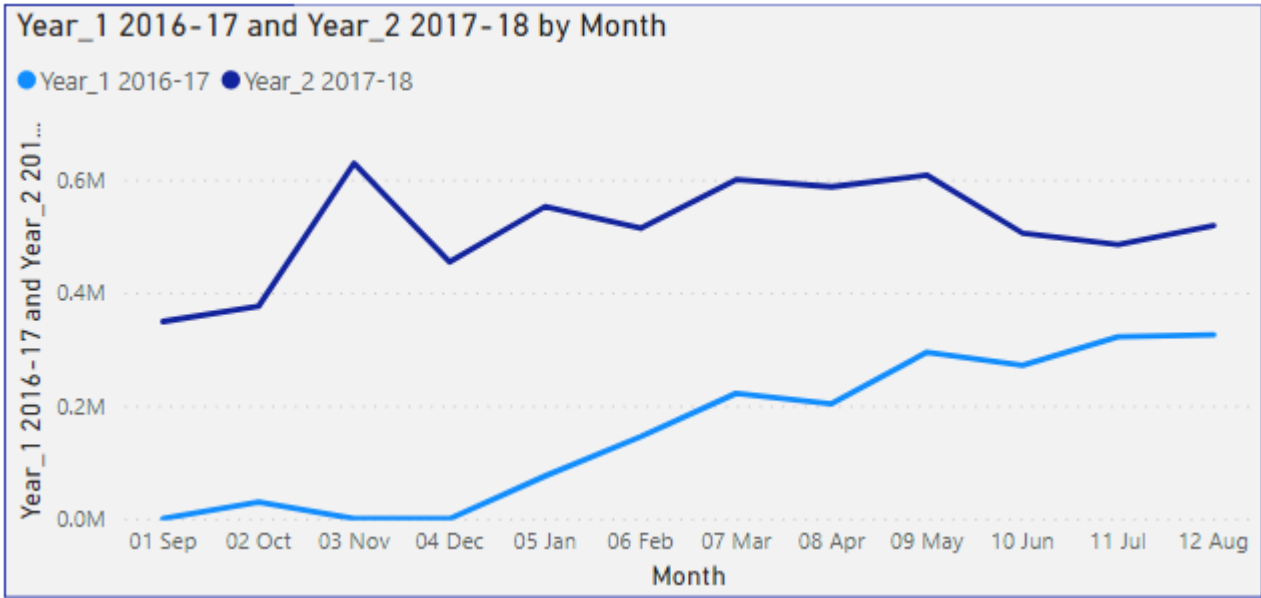
In graph "*Year_1 2016-17 and Year_2 2017-18 by Month*" we see the sales in the annual periods.

We notice a spike in November 2017 when there is not something similar in November 2016. In graphs "*Year_2 2017-18 by State and Month*" and "*Year_2 2017-18 by product_category and Month*" we see that there is an increase in sales from October to November 2017 in all States and in all Product categories. So the spike is not related to some specific States or products.

Also, we notice a decline after May 2018, which we will examine later.

Year_2 2017-18 by State and Month

Year_2 2017-18 by product_category and Month
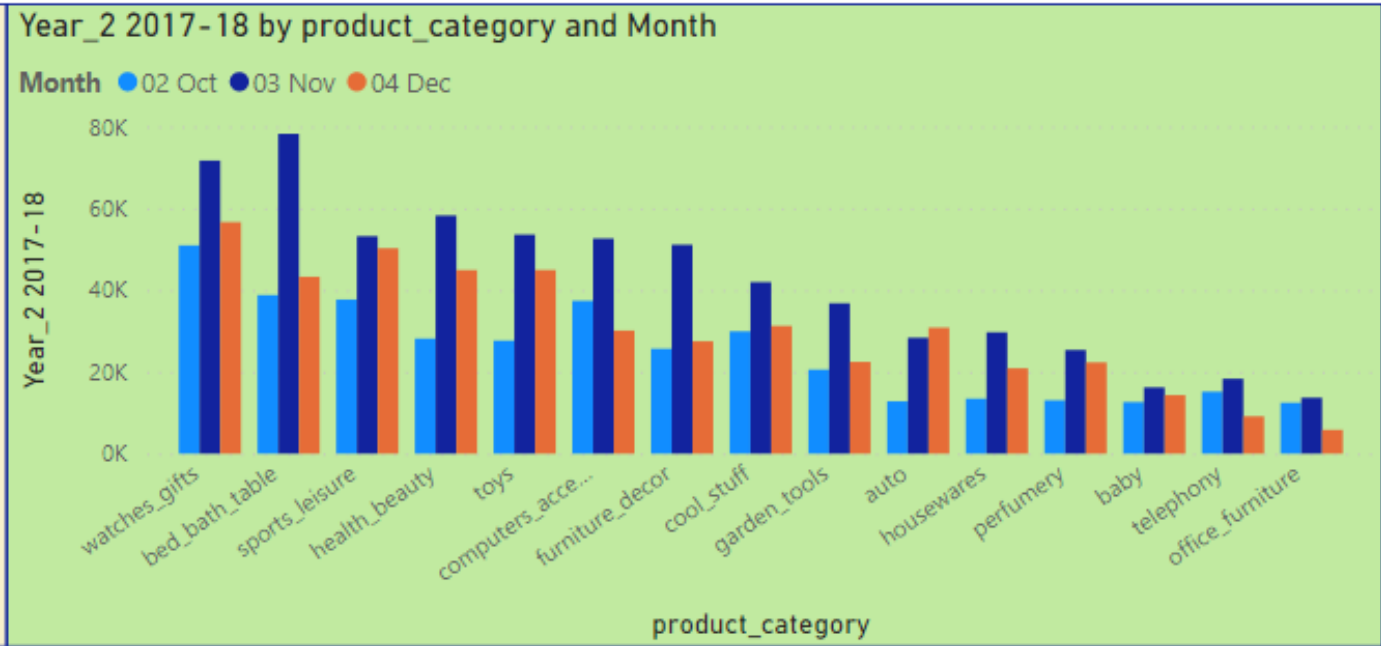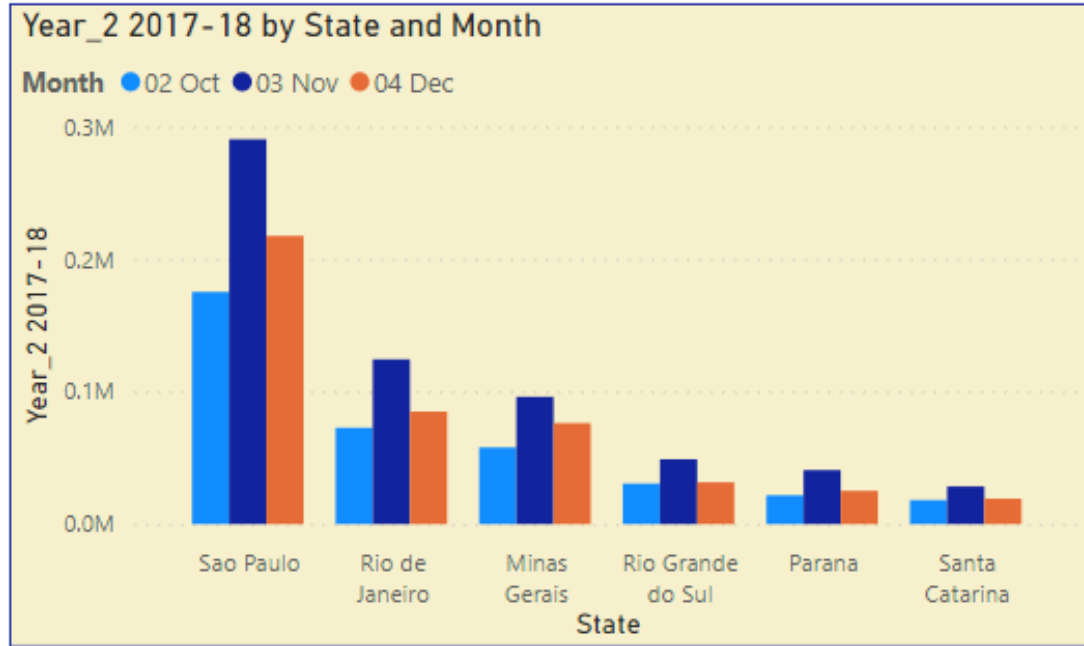
Spike in November 2017 – Sales vs MRKT qualified leads

Sales & Marketing leads for Jun 2017 - May 2018

Legend: ● Sales in BRL ● Num_leads
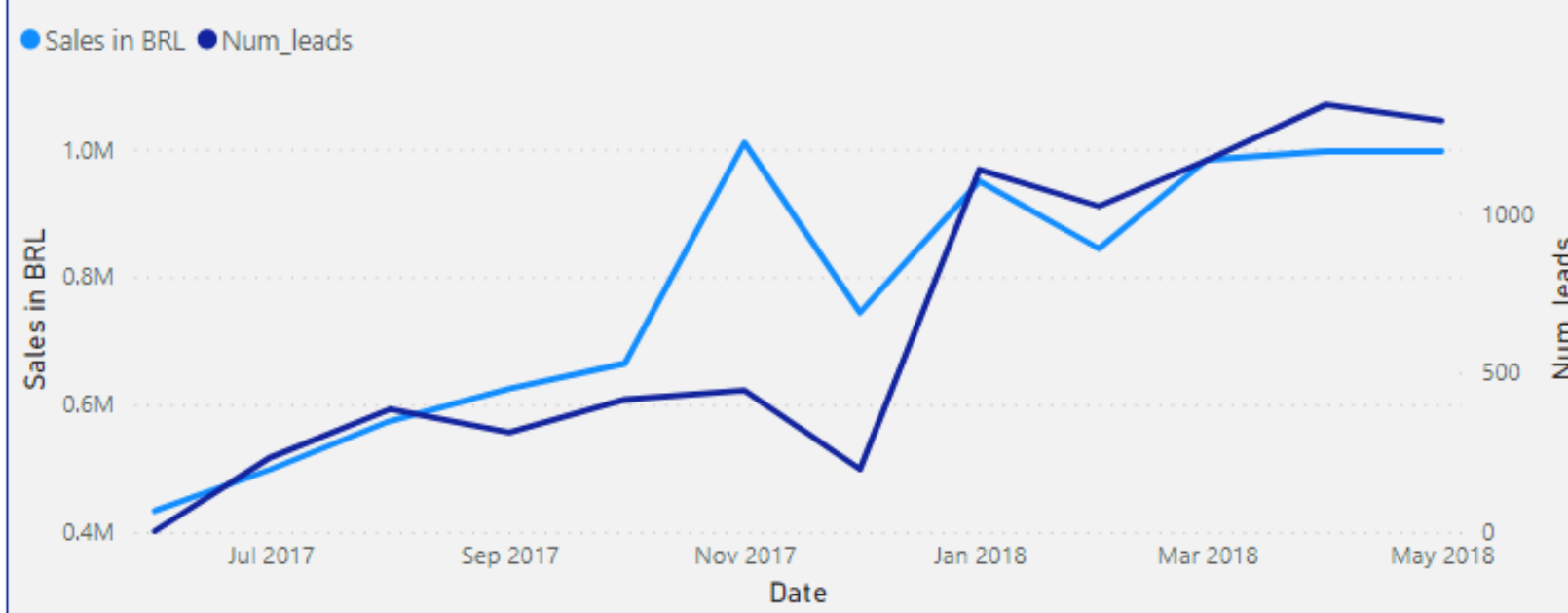
Sales & Marketing leads for Jun 2017 - May 2018

Year	Month	Day	Sales in BRL	Num_leads	Units sold
2017	June	1	433,038.60	3	3583
2017	July	1	498,031.48	234	4519
2017	August	1	573,971.68	385	4910
2017	September	1	624,401.69	312	4831
2017	October	1	664,219.43	415	5322
2017	November	1	1,010,271.37	444	8665
2017	December	1	743,914.17	197	6308
2018	January	1	950,030.36	1136	8208
2018	February	1	844,178.71	1021	7672
2018	March	1	983,213.44	1165	8217
2018	April	1	996,647.75	1339	7975
2018	May	1	996,517.68	1289	7925
Total			**9,318,436.36**	**7940**	**78135**

The data for marketing qualified leads are for only 12 months, from June to 2017 to May 2018

The increase from October 2017 to November 2017 is:
52% in revenue
62% in units sold
7% in marketing qualified leads

From this we can assume that either the same number of people bought more items on average or more qualified leads were converted to customers (or of course both). This could happened because of a discount campaign (more units were sold than the revenue which was created) or another marketing campaign.
Another assumption is that the big event of Gran Prix which took place in Brazil in November 2017, could affect positively the growth in sales if it was combined with a marketing campaign.

Decline in Sales after May 2018

Year_2 2017-18 by State and Month

Month ●09 May ●10 Jun ●11 Jul ●12 Aug
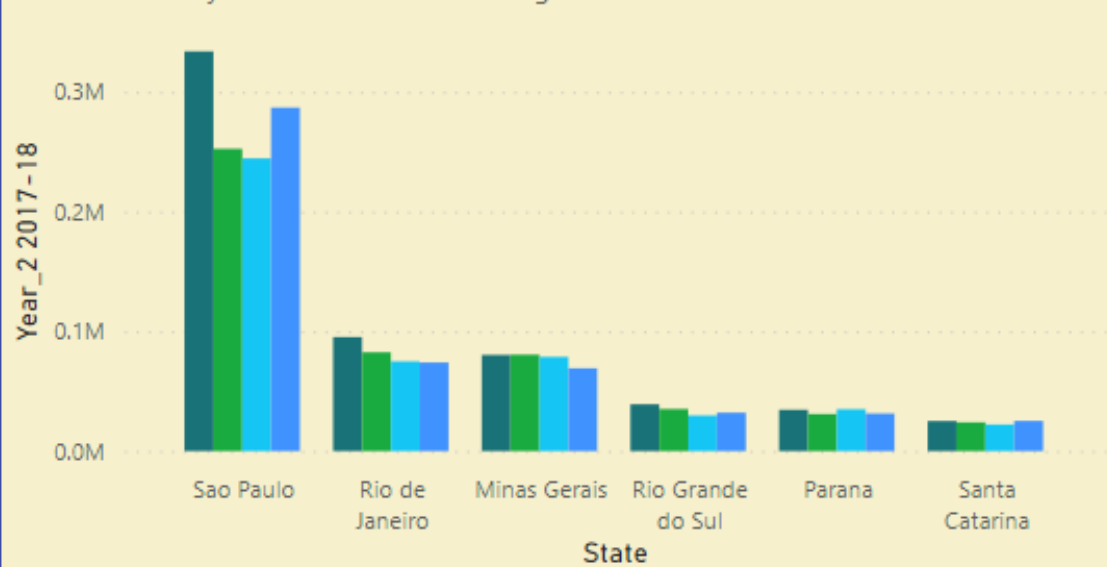
Year_2 2017-18 by product_category and Month

Month ●09 May ●10 Jun ●11 Jul ●12 Aug

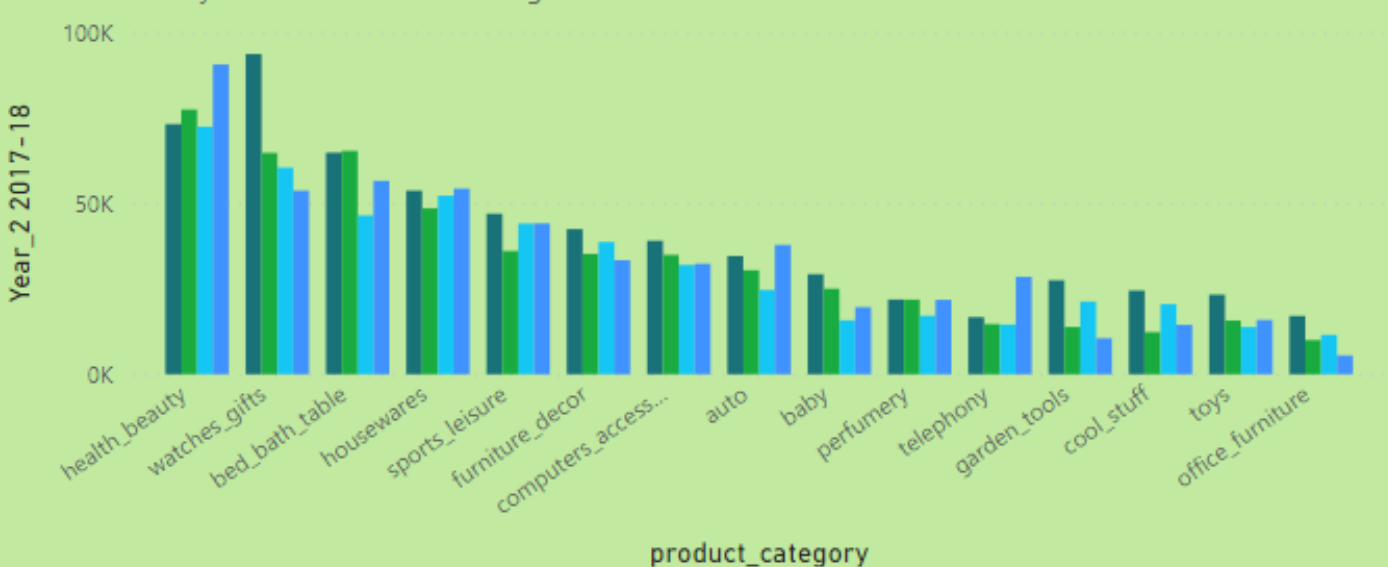The decline in sales after May 2018 happened in almost all States in almost for all product categories as we see in the 2 graphs above.
Here we can consider as the main reason the 9% drop in GDP per capita in 2018, which could be a reason for this drop (see table below "GDP per capita 2017 & 2018")

GDP per capita 2017 & 2018

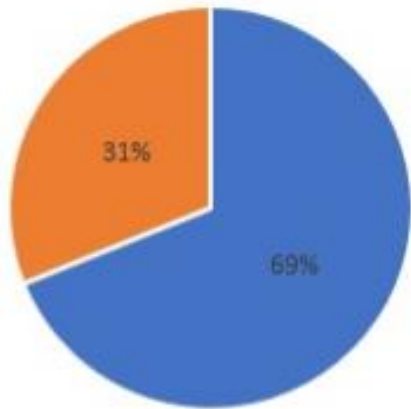| Year	Sum of GDP per capita BR $	GDP Growth	Period	Sum of Sales	Sales growth
2017	56,076.25	0.14	Year_1 (09/2016-08/2017)	3,162,786.24	
2018	50,855.65	-0.09	Year_2 (09/2017-08/2018)	10,428,712.46	2.30
	106,931.90			**13,591,498.70**	

The GDP per capita in 2018 was 9% less than 2017 but in 2018 Olist had increase in sales of 230%.
The high increase in sales is normal since is the 2nd year that the company is active but by the fact the GDP per capita is less, we understand that Olist increased their proportion in the Brazilian market versus their competitors.
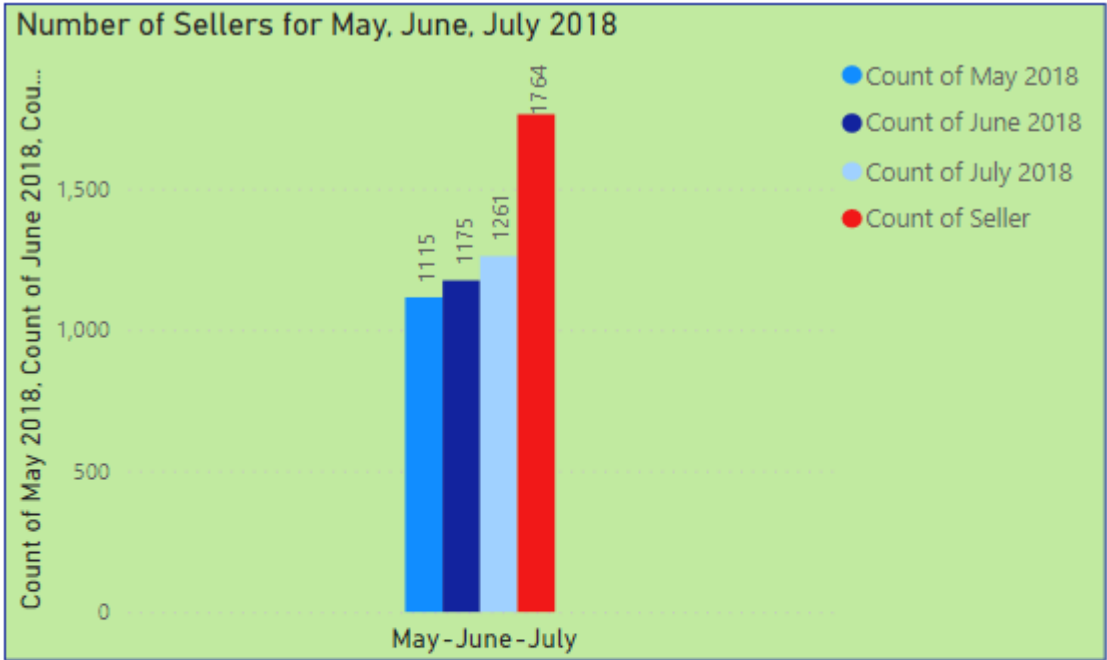
Sellers for period May - July 2018

	Sellers	Proportion
Drop after May	768	68.9%
No drop after May	347	31.1%
Only in June & July	649	
Total Sellers	1764	

Sellers May to July 2018

■ Drop after May   ■ No drop after May

31%

69%

Graph from file olist_all.xlsx sheet=Sellers-drop

Number of Sellers for May, June, July 2018

- Count of May 2018
- Count of June 2018
- Count of July 2018
- Count of Seller

1115 1175 1261 1764

May-June-July

Sales in Year_2 for May to July 2018

● Total Sales of all States & Products ● Top 6 States & Top 15 Product_categories

1.0M

0.5M

09 May   10 Jun   11 Jul
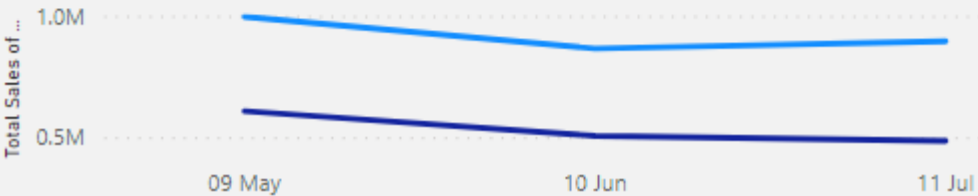
We see that the 69% (768) of all 1097 Sellers (who sold in May 2018) had drop in sales from May 2018 to July 2018.
Although in June and in July 2018 the sellers increased 58% (649 more) from May 2018, there was a 10% decline in sales from May to July for all States and Product_categories.
So, the drop is not related to some specific sellers.
However, in July 2018 there was a 3.5% raise in total, but there was a 4% decline in the top 6 States and in top 15 Product categories as we see in graph "Sales in Year_2 for May to July 2018".

11

Top 5 Product_categories - Growth in 2018

Top 5 Product_categories in Year_2

product_category ● baby ● health_beauty ● housewares ● telephony ● watches_gifts
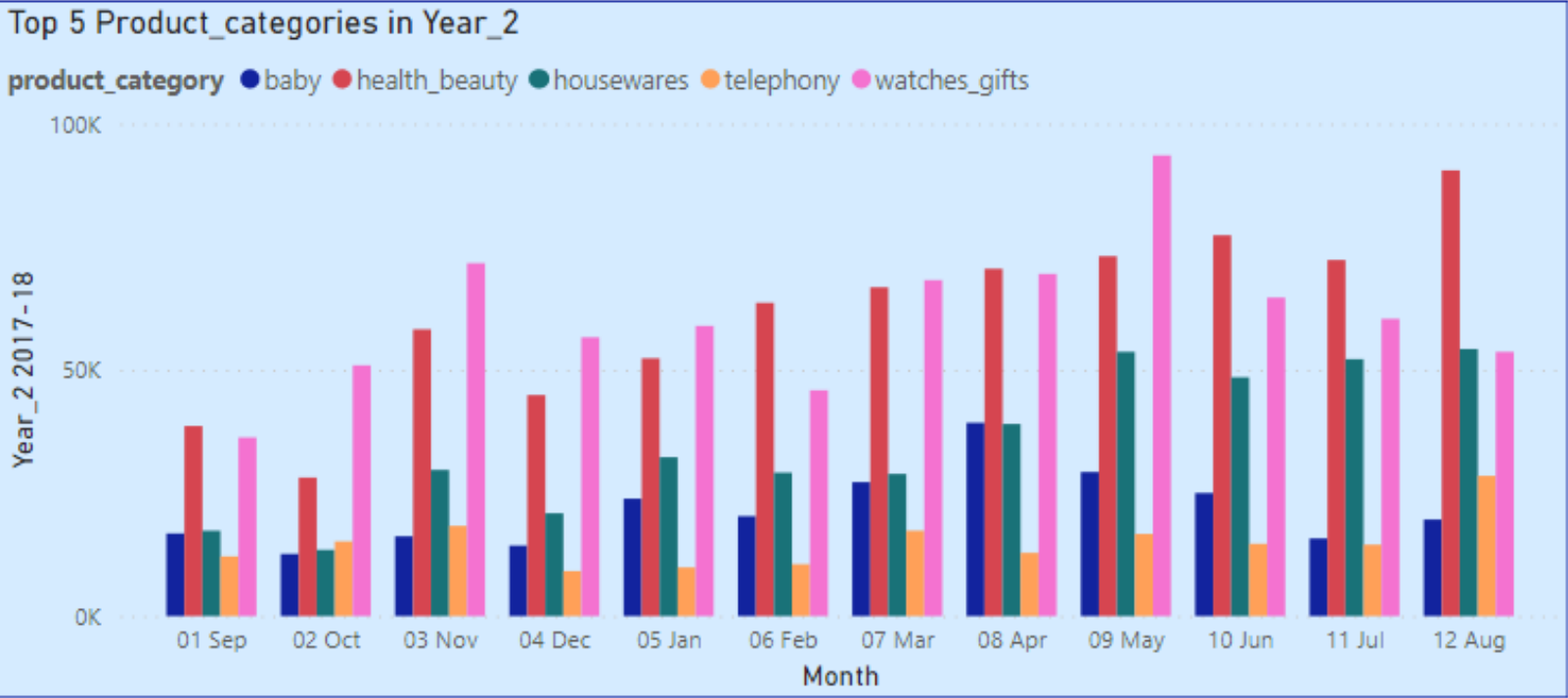
Growth by Product_category for Year_2

product_category	Year_1 2016-17	Year_2 2017-18	Growth Sales %
watches_gifts	155,321.72	730368	370.23
baby	55,944.98	260083	364.89
health_beauty	174,684.07	736515	321.63
telephony	42,616.24	179607	321.45
housewares	109,226.01	419163	283.76
computers_accessories	162,874.17	553872	240.06
auto	100,962.49	334004	230.82
sports_leisure	181,765.56	598992	229.54
bed_bath_table	216,017.31	676963	213.38
toys	99,513.72	288145	189.55
furniture_decor	154,364.98	446407	189.19
garden_tools	104,926.16	263550	151.18
office_furniture	63,952.48	155803	143.62
perfumery	99,575.03	220387	121.33
cool_stuff	167,001.34	315585	88.97
Total	**1,888,746.26**	**6179444**	**227.17**

In the graph "*Top 5 Product_categories in Year_2*" we see the performance by month of each 5 product categories which had the highest growth in Year_2.

The products "watches_gifts" had the highest growth in whole 2nd Year (370%).

Only the watches_gifts and baby had a decline after May with the first to have the highest drop.

In the next page we see the sales growth of watches_gifts by State, in order to examine deeper if the drop was related to some specific States

Decline in 4 Product Categories in 2018 (of top 15)

Sales by product_category and Year

Year ● 2016 ● 2017 ● 2018

(Sales axis: 0.0M – 0.6M; product_category axis: auto, baby, bed_bath_table, computers_acce..., cool_stuff, furniture_decor, garden_tools, health_beauty, housewares, office_furniture, perfumery, sports_leisure, telephony, toys, watches_gifts)

Product_categories with biggest decline in Year_2 & Product_categories sales average

● cool_stuff ● garden_tools ● perfumery ● toys ● Product Sales AVRG

(Y axis: cool_stuff, garden_tools, perfumery, to...; 20K – 60K; Month axis: 01 Sep, 02 Oct, 03 Nov, 04 Dec, 05 Jan, 06 Feb, 07 Mar, 08 Apr, 09 May, 10 Jun, 11 Jul, 12 Aug)
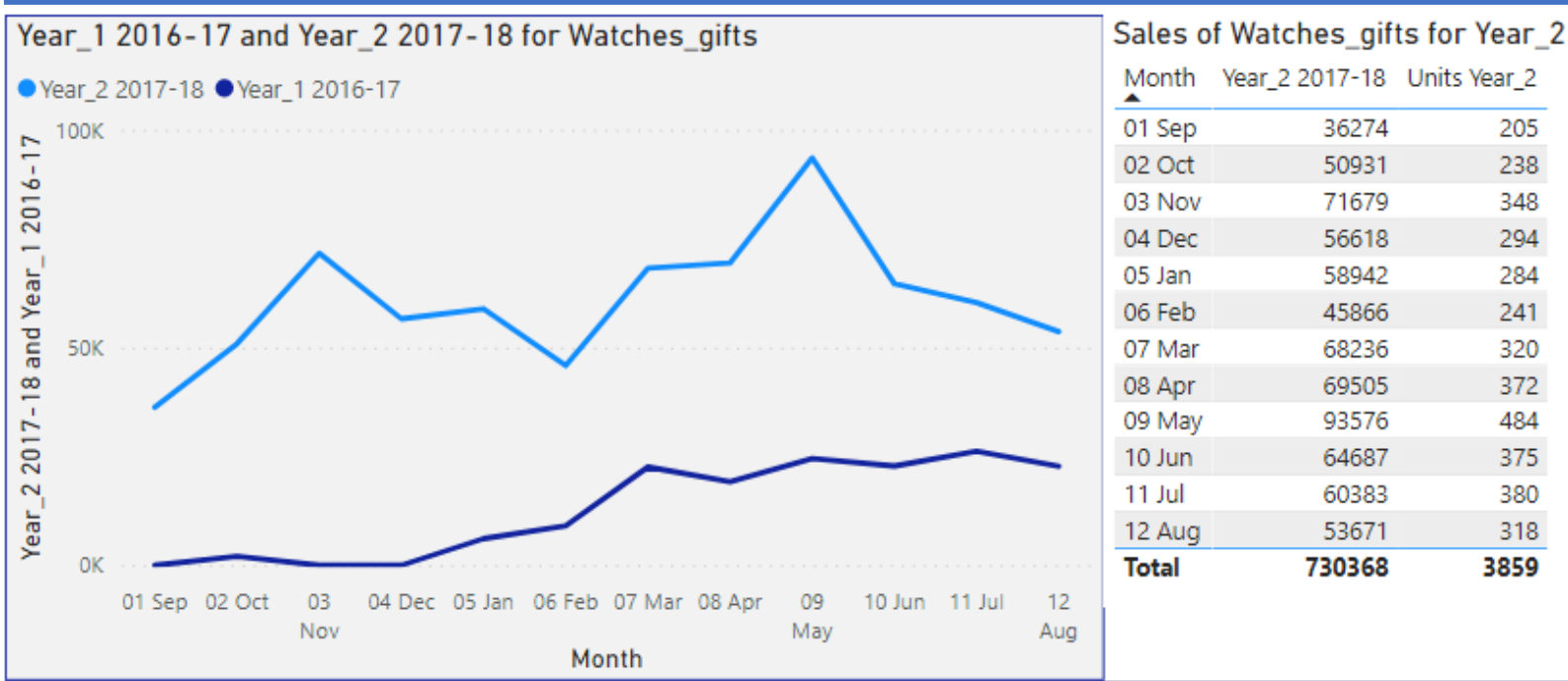
In graph "*Sales by Year and product_category*" we see the performance of all top 15 product categories. The worst performance in 2018 belongs to product categories **cool_stuff, graden_tools, perfumery** and **toys**. I see separately these 4 categories in the next graph.

In graph "*Product_categories with biggest decline in Year_2 & Product_categories sales average*" we see only these 4 product categories by month compared with the average sales of top 15 categories*.
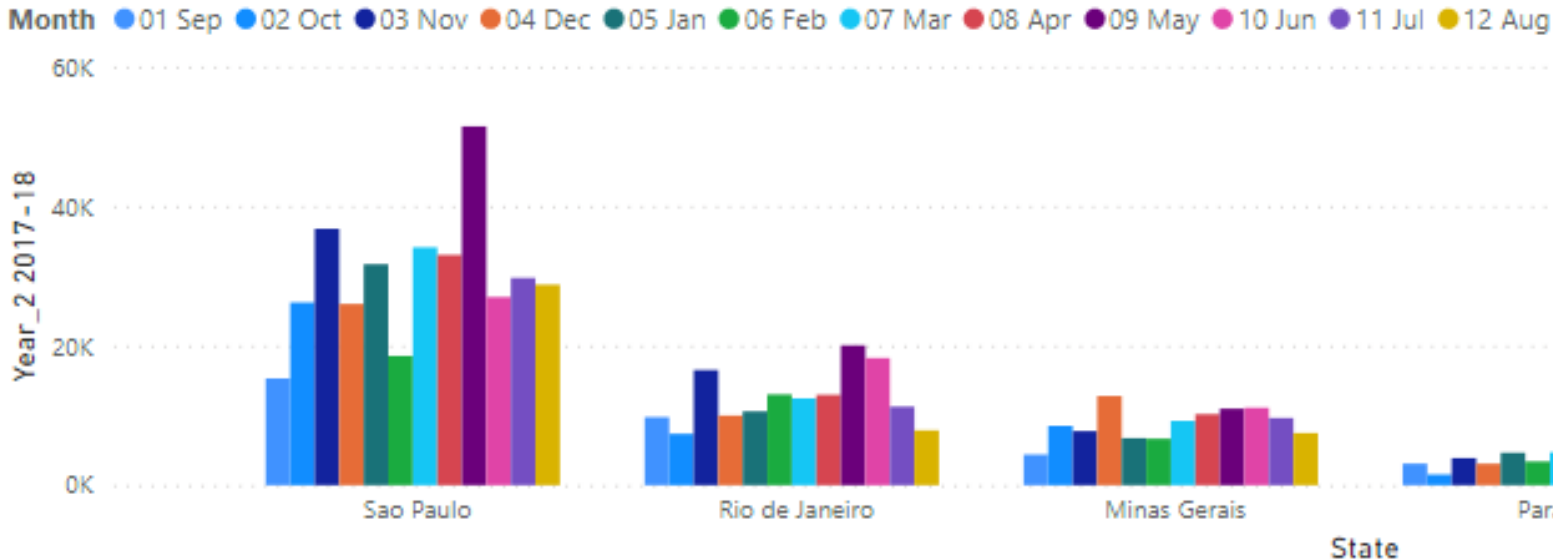**The sales of "toys" and " cool_stuff" were above average until December 2017 but later they declined a lot**

*total sales / 15

13

Watches_gifts - Increase and drop in 2018

Year_1 2016-17 and Year_2 2017-18 for Watches_gifts

● Year_2 2017-18 ● Year_1 2016-17

Sales of Watches_gifts for Year_2

Month	Year_2 2017-18	Units Year_2
01 Sep	36274	205
02 Oct	50931	238
03 Nov	71679	348
04 Dec	56618	294
05 Jan	58942	284
06 Feb	45866	241
07 Mar	68236	320
08 Apr	69505	372
09 May	93576	484
10 Jun	64687	375
11 Jul	60383	380
12 Aug	53671	318
Total	**730368**	**3859**

For the products "watches_gifts" we see 2 spikes, in November 2017 and in May 2018 but not for all States.
In November and in May, the spikes happened only in Sao Paulo and Rio de Janeiro. Parana had a spike only for November.

On the other hand the decline of February happened in Sao Paulo, Parana and Rio Grande do Sul.

So far, we see that whatever happened to Sao Paulo affects the total growth, since in Sao Paulo there is the biggest portion of total sales, as we saw in graph "Sales in BR by State" in the page 7 "Sales in Top 6 States and Top 15 Products Categories".

We need to identify the reasons of increase and decline and react accordingly with a strategic plan.

Sales for Watches_gifts in Year_2 by Month and State

Month ● 01 Sep ● 02 Oct ● 03 Nov ● 04 Dec ● 05 Jan ● 06 Feb ● 07 Mar ● 08 Apr ● 09 May ● 10 Jun ● 11 Jul ● 12 Aug

MARKETING - Origins for MRKT Qualified Leads per Month (data for period June 2017 - May 2018)

Num_leads by origin and month Jun 2017 - May 2018

Months ●01 Jun ●02 Jul ●03 Aug ●04 Sep ●05 Oct ●06 Nov ●07 Dec ●08 Jan ●09 Feb ●10 Mar ●11 Apr ●12 May

Year_2 2017-18 by Month
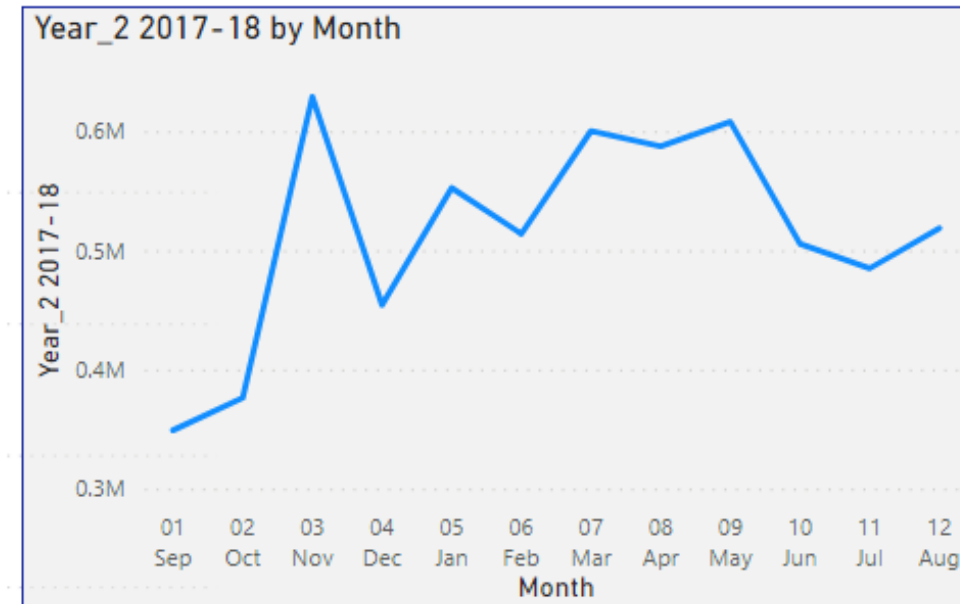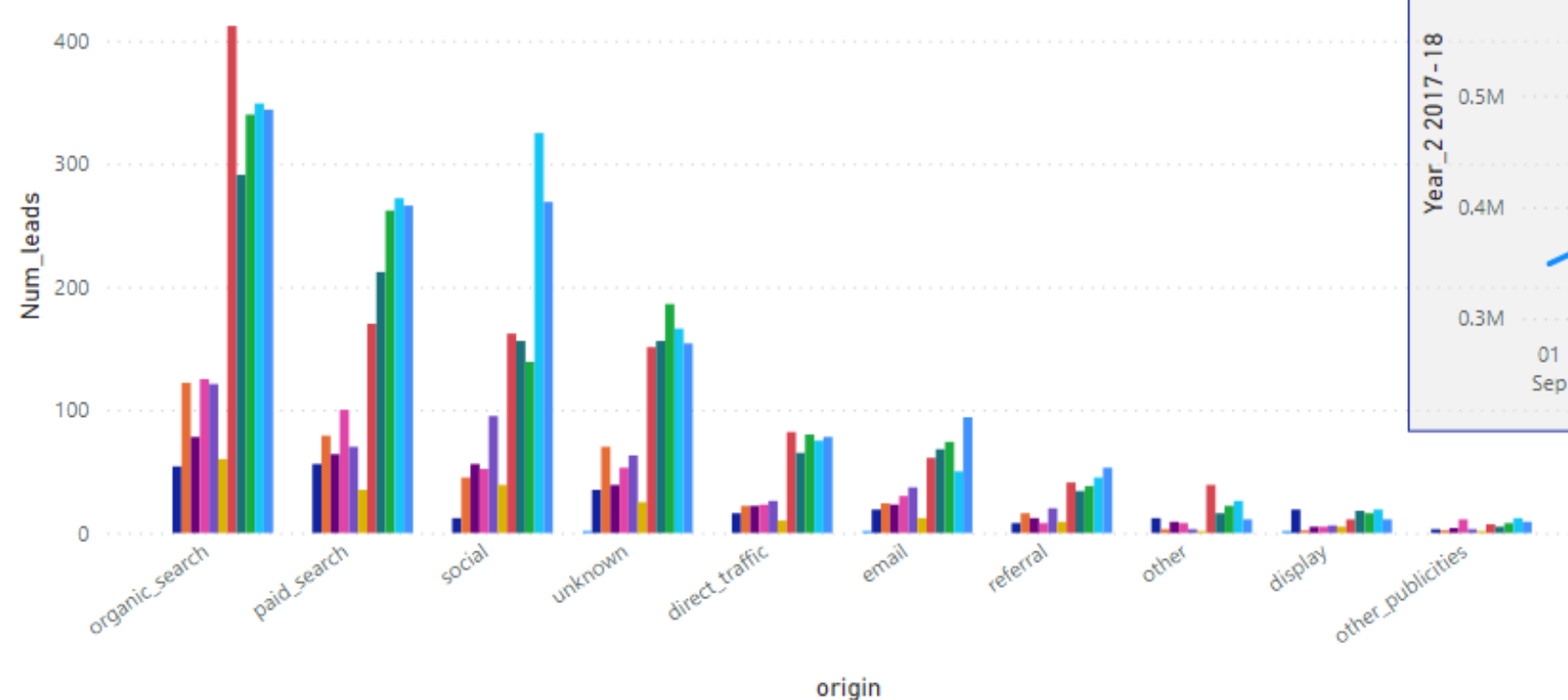
We see in the graph "*Num_leads by origin and month Jun 2017 - May 2018*" that there is a drop in all marketing channels in December 2017 which is reflected also in sales.

On the other hand there is significant increase in the marketing leads in January 2018 and in sales as well, but the increase in paid_search, email and display of February didn't lead to increase in sales, but the opposite.

From March and on there was an increase in qualified leads from all marketing channels and an increase in sales as well.

Unfortunately, we don't have data regarding the marketing qualified leads after May 2018 to see if the decline in sales was related to marketing activities.

POPULATION IN STATES vs SALES

Population and Year_2 (09/2017-08/2018) by State

● Population ● Year_2 (09/2017-08/2018)
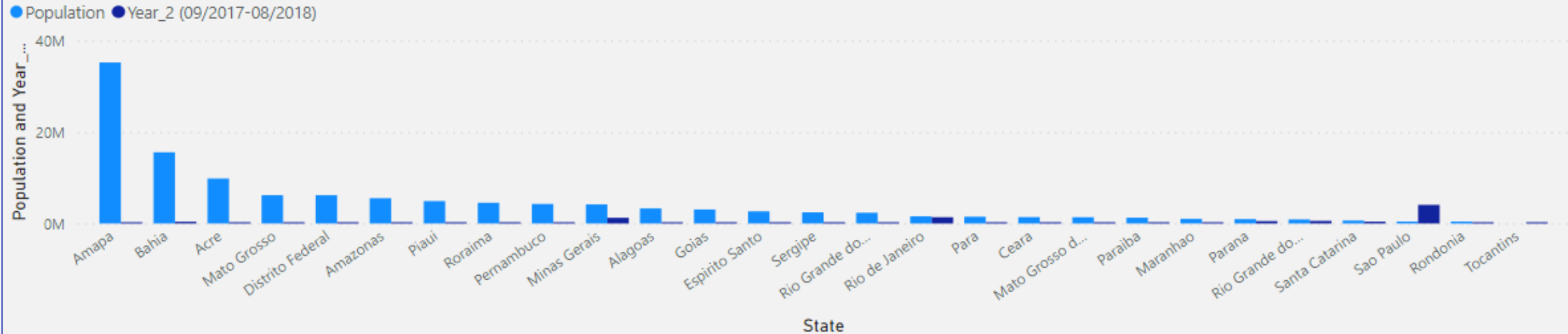
Ratio Sales/Population

State	Population	Year_2 (09/2017-08/2018)	sales/pop Year_2
Sao Paulo	419652	4,062,886.09	9,681.56
Rio de Janeiro	1561806	1,367,744.68	875.75
Rio Grande do Sul	881097	562,742.59	638.68
Santa Catarina	635998	395,809.28	622.34
Parana	955835	518,828.38	542.80
Minas Gerais	4169392	1,239,546.85	297.30
Ceara	1393408	171,762.43	123.27
Maranhao	1021557	94,388.82	92.40
Rondonia	413418	35,220.87	85.19
Para	1474859	121,949.91	82.69
Espirito Santo	2651458	212,864.69	80.28
Goias	3055149	223,511.02	73.16
Paraiba	1258407	89,251.35	70.92
Mato Grosso do Sul	1366926	86,027.92	62.94
Pernambuco	4265247	206,122.13	48.33
Total	**121489598**	**10,428,712.46**	

In Sao Paulo, Olist's ratio sales vs population = 9681*, which means that the revenue is 9,6 times more than the population of the State. Sao Paulo has the greatest sales/population ratio, which means the best performance compared to population

In the other top 5 States (Rio de Janeiro, Minas Gerais, Rio Grande do Sul, Parana and Santa Catarina) the ratio is from 297 to 875, which is not so bad.

But, there are many States, where the population is much higher than the top 6 States and the sales are very low. For example, at Amapa, Bahia, Acre, Mato Grosso the ratio sales/pop is 0,33 to 24. This means that either the habitats there are very poor or (if not) Olist should advertise more in these States because there is a big potential there.

*The ratio sales/pop is : sales/population * 1000

Conclusion

- **Olist has 230% growth in the 2nd year of activity, in which the 76.73% of total sales were occurred**

- **In November 2017 there was a pick**.
Increase 52% in sales from October 2017, in all top 6 States and all top 15 product categories, without the qualified marketing leads to be increased accordingly (only 7% increase). The increase of sales units was about the same (62%). This means that the same number of people bought more items or more qualified leads were converted to clients in November 2017. Maybe it was a discount campaign, more products available in the platform or a deficit from competitors

- **After May 2018 and till July 2018 there was a drop of 20% in top 6 States and in top 15 Product_categories.**
The drop was in almost all States. The highest drop was for States Sao Paulo, Rio Grande do Sul and Rio de Janeiro with 13.8%, 12.7% and 11.3% respectively. The drop was in all products except from health_beauty. The biggest drop was in baby, toys and watches_gifts, 25.7%, 22.3% and 18.8%
respectively. The main reason for this drop could be the 9% decline of GDP per capita in 2018.

- **Less sales with more Sellers after May 2018**
There was a 58% increase in Sellers from May to July 2018 with a 10% decrease in sales for the same period. Also, the 69% of existing sellers from May had a decline in sales after May. We have to find out which was the reason of the drop (apart from GDP per capita) and create a strategic plan.

- **Big decline in watches_gifts**
The product category "watches_gifts" had the highest growth in whole 2nd Year (370%) but the 2nd biggest drop in June 2018 (32.5%) and the drop from May (93576 BR) to August (53671 BR) was 42.6%. The biggest drop from May to August was in Rio de Janeiro (60%) and in Sao Paulo (34%).
We have to investigate the reasons and create a strategy.

Conclusion

- **Big decline in toys and cool_stuff in 2018**

The product categories cool_stuff, graden_tools, perfumery and toys (which belong to the top 15 product categories) had a decline in sales at 2018 and the most worrying are the cool_stuff and toys which even though for the period Sep 2017 - Dec 2017 were above the sales average of the top 15 product categories, in January 2018 both had a decline of 4,3% and 58,6% and in February another 21% and 32.8% respectively. In August both of them were 58.3% and 54% above average sales respectively. A strategic reaction is urgent.

- **The marketing affects the sales**

In December 2017, there was a drop in marketing qualified leads and this affected the sales for the same month. After February 2018 there was an increase in marketing qualified leads with an increase in sales. **On the contrary**, the increase in qualified leads from paid search and email campaigns in February 2018 didn't help the sales which were declined in the same month.

- **In many States the sales/population ratio is very small.**

The fact that in July 2018 there was an overall 3.5% increase in sales but a 4% decrease in top 6 States (graph "Sales in Year_2 for May to July 2018", page Sellers) confirms that there is potential in the rest 23 States.

The Southest Region in Brazil which is composed of the states of Espírito Santo, Minas Gerais, Rio de Janeiro and São Paulo is the richest region of the country*. However Espirito Santo is not in included in the top 6 States and its sales/population ration is low (80,28). This means that there is potential in this State, so we should create a marketing and sales..

*https://en.wikipedia.org/wiki/Southeast_Region,_Brazil

Suggestions

We need to collect more data in order to get better insights. These data should be:

- **GDP per State** in Brazil for the years 2016, 2017, 2018 to examine the growth of sales per State vs the market growth. That way we will identify in which states there is potential so as to advertise more.
Source: https://tradingeconomics.com/ cost: $199 monthly

- **Retail market share** for each product category, in order to compare Olist's sales vs competitors
Source: https://www.reportlinker.com/ cost: € 1500 yearly

- **Marketing cost** for each marketing channel, number of clients who came from each marketing channel in order to calculate the CPC and the cost per customer. Then we can allocate the marketing budget properly.
Source: Olist

Appendix

- **Datasets for Olist**

https://www.kaggle.com/olistbr/brazilian-ecommerce

- **Population for Brazil (2020)**

https://en.wikipedia.org/wiki/List_of_Brazilian_states_by_population

- **Brazil GDP Per Capita**

https://www.macrotrends.net/countries/BRA/brazil/gdp-per-capita

- **Brazilian States - Abbreviations**

https://brazil-help.com/brazilian_states.htm