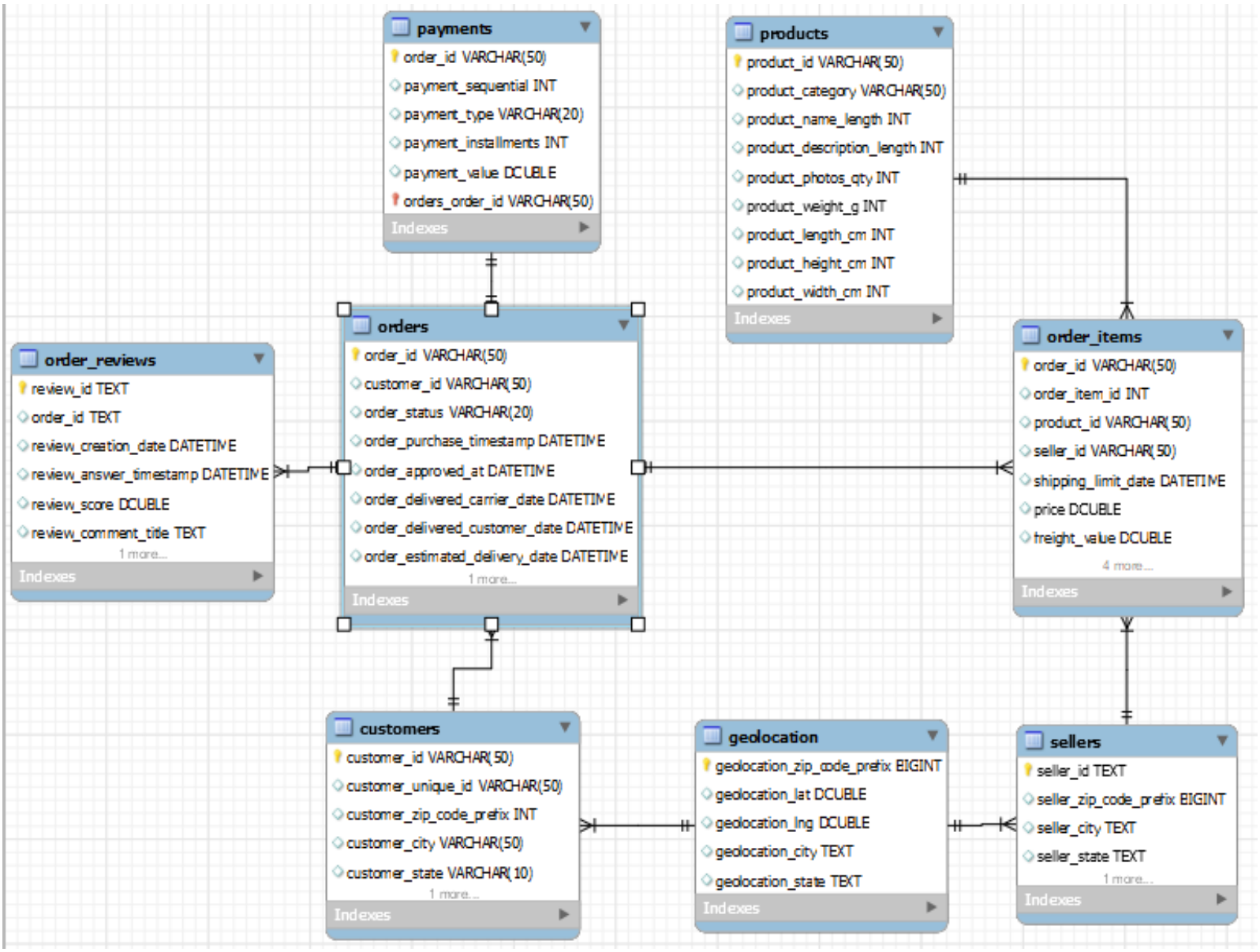
TARGET

- **E-R Diagram:**

  It is showing the Primary Key in each table, data type of entities and the relation between the tables.



- **Time period for which the data is given:**

  - The given data is between the dates from **2016-09-04** and **2018-10-17** i.e. for 2 Years 1 month and 13 days.

Starting Date:

date(order_purchase_timestamp)	date(order_approved_at)	date(order_delivered_carrier_date)	date(order_delivered_customer_date)	date(order_estimated_delivery_date)
2016-09-04	2016-10-07	2016-10-18	0000-00-00	2016-10-20

Ending Date:

date(order_purchase_timestamp)	date(order_approved_at)	date(order_delivered_carrier_date)	date(order_delivered_customer_date)	date(order_estimated_delivery_date)
2018-10-17	0000-00-00	0000-00-00	0000-00-00	2018-10-30

- The range of shipping limit dates is between **2016-09-19** and **2020-04-09** i.e. the date for which the particular order item should be shipped.
- The range for the dates when the reviews has been given by the customers is between **2016-10-07** and **2018-10-29.**

• About cities, states and zip-codes

1. **Zip-code:**
   - There are total of **19015** unique zip code in geolocation.
   - The no. of zip code for the customer who order something is **14994**.
   - The sellers are from the **2246** unique zip codes.
   - The zip codes that are common for customers, sellers and the geolocations given is **2160**.
   - There are **2162** sellers and customers zip code which have the same value.

2. **Cities:**
   - There are **4119** cities from where the customers are.
   - The top 10 cities and their state from where the most of the customers are:

customer_city	customer_state	No_of_customers
sao paulo	SP	15540
rio de janeiro	RJ	6882
belo horizonte	MG	2773
brasilia	DF	2131
curitiba	PR	1521
campinas	SP	1444
porto alegre	RS	1379
salvador	BA	1245
guarulhos	SP	1189
sao bernardo do campo	SP	938

- The sellers are from **635 different cities**. Top 10 cities with their state from where the most of the sellers are:

seller_city	seller_state	No_of_sellers
sao paulo	SP	695
curitiba	PR	124
rio de janeiro	RJ	93
belo horizonte	MG	66
ribeirao preto	SP	52
guarulhos	SP	50
ibitinga	SP	49
santo andre	SP	45
campinas	SP	41
maringa	PR	40

3. State:

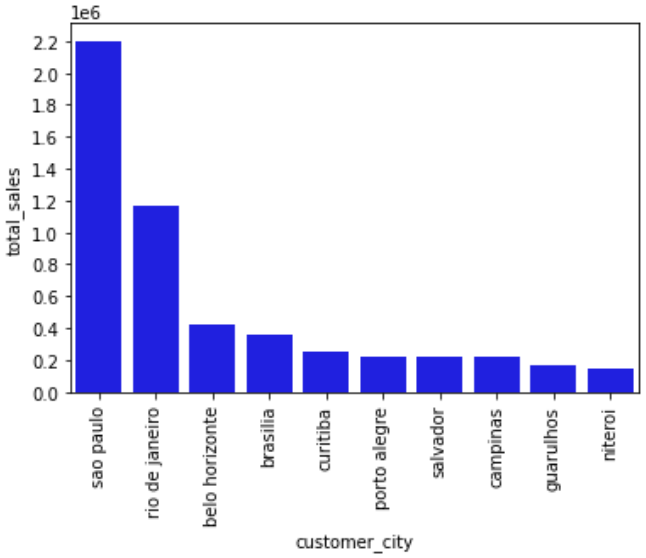
- There are total of **27 States** from where the customers have placed their order. and top 10 states with the most no. of customers are:

customer_state	No_of_customers
SP	41746
RJ	12852
MG	11635
RS	5466
PR	5045
SC	3637
BA	3380
DF	2140
ES	2033
GO	2020

- There are total of *23 States from where the sellers come*. Top 10 states from where the most of the sellers come are:

seller_state	No_of_sellers
SP	1849
PR	349
MG	244
SC	190
RJ	171
RS	129
GO	40
DF	30
ES	23
BA	19

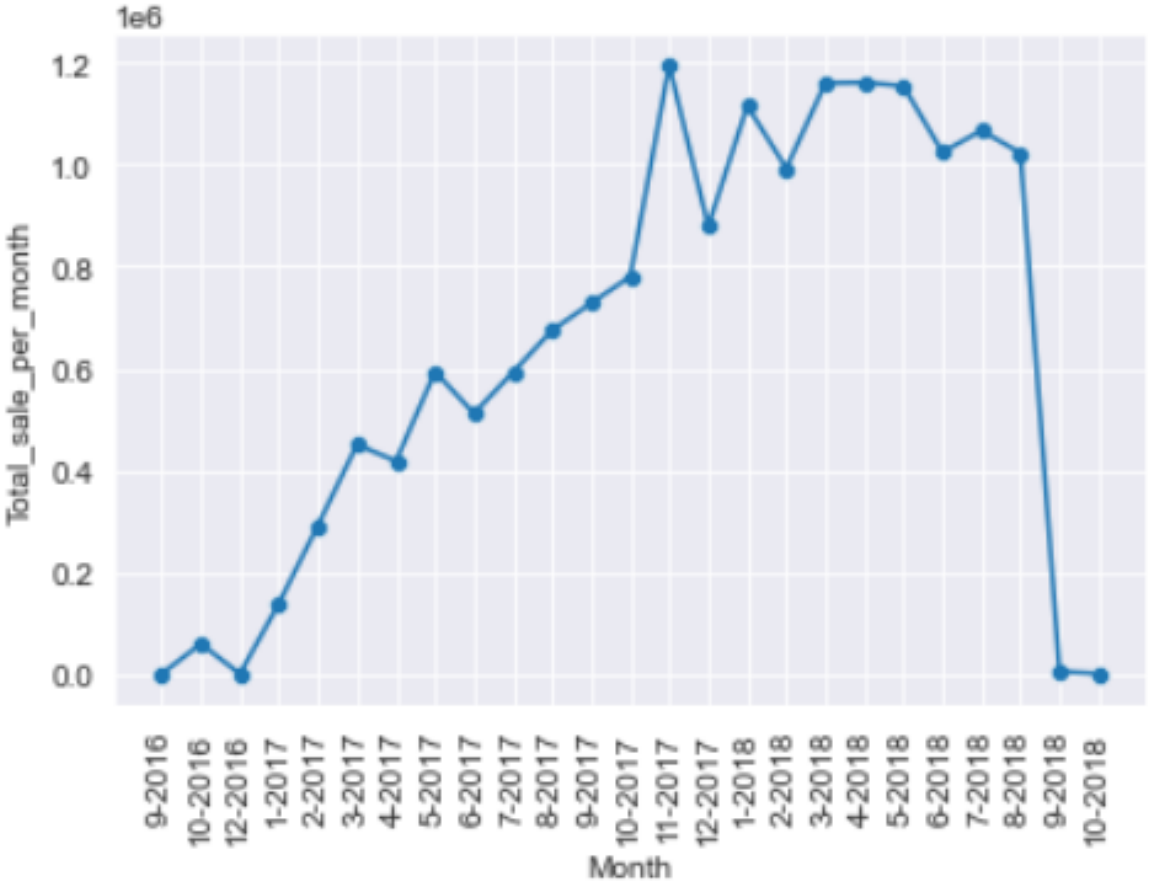
➢ The max sales in the top 10 city are:



customer_city	total_sales
sao paulo	2203373.09
rio de janeiro	1161927.36
belo horizonte	421765.12
brasilia	354216.78
curitiba	247392.48
porto alegre	224731.42
salvador	218071.5
campinas	216248.43
guarulhos	165121.99
niteroi	139996.99

From the above data, we can easily identify that where the Target should set up there warehouses so that delivery will be fast and it can directly impact on the customer review which significantly increases the customer satisfaction and brand value of the company.

- **E-Commerce Trend in Brazil**



	Month	Total_sale_per_month
0	9-2016	252.24
1	10-2016	59090.48
2	12-2016	19.62
3	1-2017	138488.04
4	2-2017	291908.01
5	3-2017	449863.60
6	4-2017	417788.03
7	5-2017	592918.82
8	6-2017	511276.38
9	7-2017	592382.92
10	8-2017	674396.32
11	9-2017	727762.45
12	10-2017	779677.88
13	11-2017	1194882.80
14	12-2017	878401.48
15	1-2018	1115004.18
16	2-2018	992463.34
17	3-2018	1159652.12
18	4-2018	1160785.48
19	5-2018	1153982.15
20	6-2018	1023880.50
21	7-2018	1066540.75
22	8-2018	1022425.32
23	9-2018	4439.54
24	10-2018	589.67

The increasing trend can be defined when the sales increase by every passing month.

Here we first find the cost of item that each customer purchases and then we grouped it by month to get the total purchase month-wise

From the graph, we can see that the trend for E-Commerce in Brazil *is increasing from Sept.-2016 to June-2018.*

After then it decreases for 3 continuous months and then there is a sudden drop in **September-2018**

The peak months are: **May-2017, November-2017, December-2017.**

- There is a *sharp peak in Nov.-2017*. The top product category sales for that month are shown below:

product_category	count
bed table bath	975
Furniture Decoration	783
sport leisure	612
HEALTH BEAUTY	584
Garden tools	550
computer accessories	532
toys	495
Watches present	469
housewares	423
telephony	383

- Brazilian customers tend to buy at Evening i.e. between 4:00 pm to 9:00pm

Timings	No_of_Orders
Night	21589
Morning	22932
Afternoon	26632
Evening	32123
Dawn	610

- **Analyzing data by month on month orders by region, states:**

On analyzing data month by month, we found the state which has the maximum amount of order per month.

The query used to compute this is:

```
select  month_name, customer_state, No_of_order
from (
        select
                concat(month(order_purchase_timestamp),'-', year(order_purchase_timestamp)) month_name,
                customer_state,
                first_value(count(customer_state))
                over (partition by year(order_purchase_timestamp), month(order_purchase_timestamp) order by count(customer_state) desc) 'No_of_order' ,
                rank() over (partition by year(order_purchase_timestamp), month(order_purchase_timestamp) order by count(customer_state) desc) 'Ranks'
        from
                city_pay
        group by
                month(order_purchase_timestamp), year(order_purchase_timestamp), customer_state
        order by
                date(order_purchase_timestamp), count(*) desc) p
where p.Ranks=1
group by p.month_name;
```

The top 10 states with their month that has max no. of orders are:

month_name	customer_state	No_of_order
8-2018	SP	3337
5-2018	SP	3310
1-2018	SP	3181
4-2018	SP	3178
11-2017	SP	3166
3-2018	SP	3150
6-2018	SP	2867
7-2018	SP	2843
2-2018	SP	2793
12-2017	SP	2470

We find out that the state SP has the most no. of orders placed by customers except for the month of Sept. and Dec. 2016.

Following table shows the no. of customers according to city of the customer are

customer_city	customer_state	No_of_customers
sao paulo	SP	15540
rio de janeiro	RJ	6882
belo horizonte	MG	2773
brasilia	DF	2131
curitiba	PR	1521
campinas	SP	1444
porto alegre	RS	1379
salvador	BA	1245
guarulhos	SP	1189
sao bernardo do campo	SP	938
niteroi	RJ	849
santo andre	SP	797
osasco	SP	746
santos	SP	713
goiania	GO	692

Max. customers are from **sao paulo** which has **15540** customers.

In-depth Exploration:

- **Percentage change in the cost of orders from 2017 to 2018 for month Jan.-Aug:**

The query used to compute the percentage change is:

```
with year_2017 as (select monthname(order_purchase_timestamp) 'Month_name', round(sum(payment_value),2) 'cost_of_orders_Year_2017'
from city_pay
where year(order_purchase_timestamp)=2017 and month(order_purchase_timestamp) between 1 and 8
group by month(order_purchase_timestamp)
order by date(order_purchase_timestamp)),
year_2018 as (select monthname(order_purchase_timestamp) 'Month_names', round(sum(payment_value),2) 'cost_of_orders_Year_2018'
from city_pay
where year(order_purchase_timestamp)=2018 and month(order_purchase_timestamp) between 1 and 8
group by month(order_purchase_timestamp)
order by date(order_purchase_timestamp))

select Month_name, cost_of_orders_Year_2017, cost_of_orders_Year_2018,round((cost_of_orders_Year_2018-cost_of_orders_Year_2017),2) difference,
       concat(round(((cost_of_orders_Year_2018-cost_of_orders_Year_2017)/cost_of_orders_Year_2017)*100,2),' %') '%_change'
from year_2017 e join year_2018 i on e.Month_name=i.Month_names;
```

Below table shows the change in percentage:

Month_name	cost_of_orders_Year_2017	cost_of_orders_Year_2018	difference	%_change
January	138488.04	1115004.18	976516.14	705.13 %
February	291908.01	992463.34	700555.33	239.99 %
March	449863.6	1159652.12	709788.52	157.78 %
April	417788.03	1160785.48	742997.45	177.84 %
May	592918.82	1153982.15	561063.33	94.63 %
June	511276.38	1023880.5	512604.12	100.26 %
July	592382.92	1066540.75	474157.83	80.04 %
August	674396.32	1022425.32	348029	51.61 %

The above data shows that the sales increased rate is at-least 50% from the last year.

The max percent change is in **January** and the minimum change percent is in **August**.

……

The state **PB** has the highest mean value of **234.2**

The sum of price and freight value is max for state **SP** whose total is **5921678.12**

customer_state	Total_price (price+freight_value)	count_of_orders	Mean
SP	5921678.12	47449	124.8
RJ	2129681.98	14579	146.08
MG	1856161.49	13129	141.38
RS	885826.76	6235	142.07
PR	800935.44	5740	139.54
SC	610213.6	4176	146.12
BA	611506.67	3799	160.97
DF	353229.44	2406	146.81
GO	347706.93	2333	149.04
ES	324801.91	2256	143.97
PE	322237.69	1806	178.43
CE	275606.3	1478	186.47
PA	217647.11	1080	201.53
MT	186168.96	1055	176.46
MA	151171.99	824	183.46
MS	135956.67	819	166

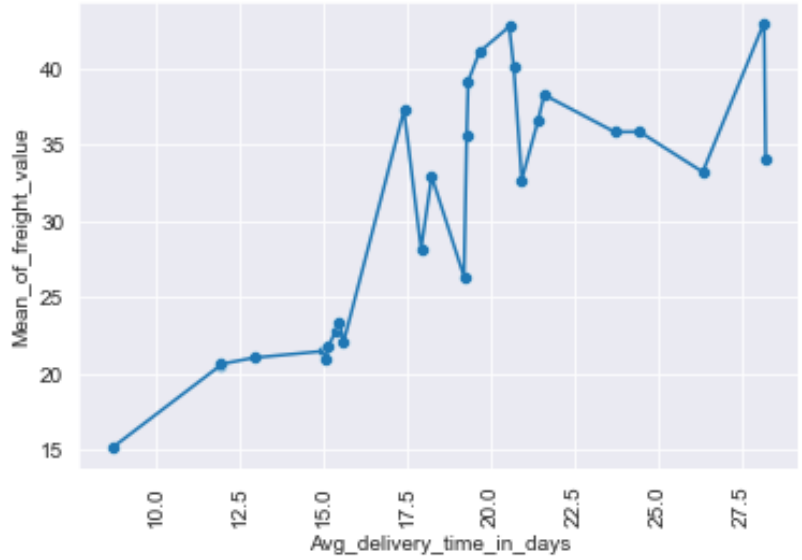
- **Analysis on sales, freight and delivery time:**

  - There are *2965 orders* in which order_delivered_customer_date is not mentioned.
  - The max days for which the order get delivered is *210 days* which is late by *181 days* form the estimated delivery date.
  - No. of orders which are delivered late from their expected delivery date is *6535* i.e. **6.77%** orders are delivered late from their estimated delivery date.
  - There is only 1 order which is delivered on the same day withing 15 hrs.
  - There are *629 orders* which is delivered on the next day.

State	Mean_of_freight_value	Avg_delivery_time_in_days	Avg_estimated_time_in_days
SP	15.15	8.6623	11.2079
PR	20.53	11.8931	13.4861
MG	20.63	11.9207	13.3426
DF	21.04	12.8938	12.2004
SC	21.47	14.9502	11.5727
RJ	20.96	15.0748	12.0148
RS	21.74	15.1345	14.1342
GO	22.77	15.3355	12.2942
MS	23.37	15.4599	11.2293
ES	22.06	15.5874	10.6463
TO	37.25	17.3968	12.3387
MT	28.17	17.9074	14.5718
PE	32.92	18.2245	13.4502
BA	26.36	19.1925	10.9826
RN	35.65	19.2726	13.9520
PI	39.15	19.3174	11.5277
RO	41.07	19.6557	20.0403
PB	42.72	20.5461	13.0375

The above table shows the freight_value, time_to_delivery, diff_estimated_delivery grouped by state.

  - The state SP Take the lowest avg. no. of days in delivering an order.
  - Freight value increases if the delivery time increases. The graph shows the same:



- The avg. delivery date is ahead of estimated delivery date for only 3 states i.e. **SP, PR, MG**.

  For all other states the avg. delivery date took more days than estimated delivery date.

- Top 5 states with highest and lowest average freight value:

State	Mean_of_freight_value
RR	42.98
PB	42.72
RO	41.07
AC	40.07
PI	39.15

Highest average freight value

State	Mean_of_freight_value
SP	15.15
PR	20.53
MG	20.63
RJ	20.96
DF	21.04

Lowest average freight value

- Top 5 states with highest and lowest average time to delivery:

State	Avg_delivery_time_in_days
AP	28.2222
RR	28.1739
AM	26.3374
AL	24.4473
PA	23.7021

Highest Avg. Time To Delivery

State	Avg_delivery_time_in_days
SP	8.6623
PR	11.8931
MG	11.9207
DF	12.8938
SC	14.9502

Lowest Avg. Time To Delivery

- Top 5 states where delivery is really fast/ not so fast compared to estimated date:

State	Avg_delivery_time_in_days	Avg_estimated_time_in_days
SP	8.6623	11.2079
PR	11.8931	13.4861
MG	11.9207	13.3426
RO	19.6557	20.0403
AC	20.6813	20.9780

States with fastest delivery

State	Avg_delivery_time_in_days	Avg_estimated_time_in_days
AL	24.4473	8.7354
MA	21.5900	9.9063
SE	21.4187	10.0027
RR	28.1739	18.3261
AP	28.2222	18.3951

States with Slowest delivery

• *Payment type analysis:*

➢ Month over Month count of orders for different payment types:

The query used to get the data is-

```
SELECT
    concat(month(order_purchase_timestamp),'-',year(order_purchase_timestamp)) Month_year,
    payment_type,
    round(sum(payment_value),2) total,
    count(*) count_of_orders
FROM payments p
join orders o on o.order_id=p.order_id
group by month(order_purchase_timestamp),year(order_purchase_timestamp), payment_type
order by year(order_purchase_timestamp),month(order_purchase_timestamp),payment_type;
```

Result:

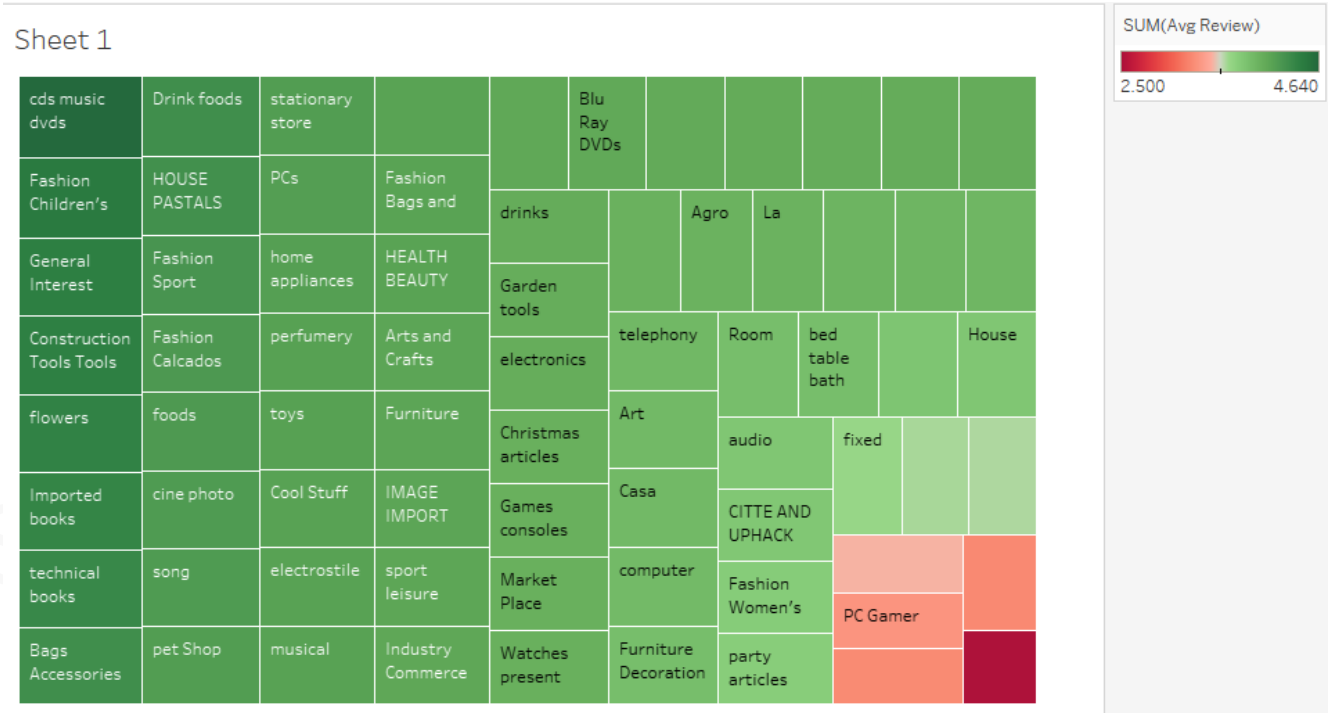
Month_year	payment_type	total	count_of_orders
9-2016	credit_card	252.24	3
10-2016	credit_card	48290.62	254
10-2016	debit_card	241.73	2
10-2016	UPI	9679.06	63
10-2016	voucher	879.07	23
12-2016	credit_card	19.62	1
1-2017	credit_card	109615.68	583
1-2017	debit_card	743.53	9
1-2017	UPI	24074.43	197
1-2017	voucher	4054.4	61
2-2017	credit_card	226753.56	1356
2-2017	debit_card	1510.32	13
2-2017	UPI	57476.74	398
2-2017	voucher	6167.39	119

- From the data, we find that customers use credit card more often for the payment, Debit card and UPI has the same no. of usage.
- The max usage of credit card can be seen in the month November 2017 which has been used for 5897 orders that equals to total amount of 942856.67
- The min. no. of installment is 1 and max. no. of installments is 24. Most of the orders bought on 1 installment only.

Review

- The product category with the highest avg. review score is **cds music dvds** having review score of **4.64**

Heatmap for the review score is:



• Insights:

1) The given data is between the dates from *2016-09-04 and 2018-10-17* i.e. for *2 Years 1 month and 13 days.*
2) There are total of *19015 unique zip code* in geolocation.
3) There are *4119 cities* from where the customers are.
4) There are total of *27 States* from where the customers have placed their order.
5) The max sales is for **sao paulo city**.
6) We can see that the trend for E-Commerce in *Brazil* is increasing from *Sept.-2016 to June-2018*. After then it decreases for 3 continuous months and then there is a sudden drop in September-2018
   The peak months are: May-2017, November-2017, December-2017.
7) Brazilian customers tend to buy at **Evening** i.e. between 4:00 pm to 9:00pm
8) The state **SP** has the *most no. of orders* placed by customers except for the month of Sept. and Dec. 2016.

Max. customers are from *sao paulo* which has *15540 customers.*

9) The sales increased rate is **at-least 50%** from the last year.
   The max percent change is in January and the minimum change percent is in August.
10) The max days for which the order get delivered is **210 days** which is late by **181 days** from the estimated delivery date.
11) No. of orders which are delivered late from their expected delivery date is *6535* i.e. **6.77%** orders are delivered late from their estimated delivery date.
12) There is only 1 order which is delivered on the same day within 15 hrs

---

• Recommendations:

1) Target should set up their warehouses in the states and city where the order deliver date is late from estimated delivery date so that delivery will be fast and it can directly impact on the customer review which significantly increases the customer satisfaction and brand value of the company.
2) Target should introduce offers in the month of Aug- Oct. 2018 since in this 3 month, the sales dropped drastically so in order to revive, introducing offers and attracting customers can be a master stroke.
3) Target should maximize their promotion on Social media and TV platforms in the evening as the customers tend to order more in evening time i.e. from 4-9 pm which helps to increase the sales.
4) Since the orders are more in the city sao paulo, more no. of billboards should be installed there to connect with the public and being popular and brand friendly amongst them.
5) Since the sales are more in Jan. to Aug. in decreasing order, the stock of the items should be made accordingly in order to minimize the loss in case of any defectiveness in the product due to various factors.
6) There are 6.77% of orders which are delivered late then their expected delivery date so it should be reduced in the order to increase the customer satisfaction.
7) There are orders which are placed in one day so the model should be made which helps to deliver it on the same day if the order is placed before a particular time of the day. This helps in increasing the customer satisfaction.
8) Target should open their warehouses in the state: AP, RR, AM, AL, PA as this state's avg. delivery time is maximum.
9) Customers uses credit card more often in comparison to other payment types so Target should collaborate with those credit card companies in order to launch offers which will attract more customers which is directly beneficial for the company revenue.
10) The products in the category of insurance and service should be improved as the customer review for the category is 2.5

Thank You

Submitted by: Saksham Singhal