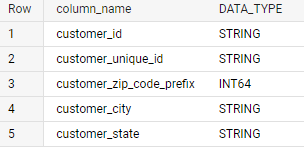
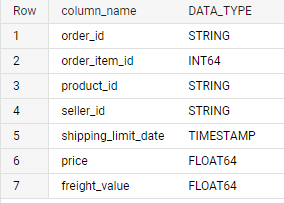
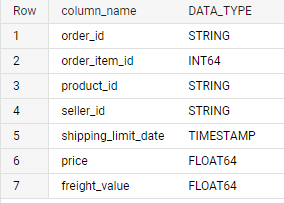
1. **STRUCTURE & CHARACTERISTICS OF THE DATASET**
   * There are 8 tables in the dataset:
     1. Customers: 5 columns and 99441 rows
     2. Geolocation : 5 columns and 1000163 rows
     3. order\_items : 7 columns and 112650 rows
     4. payments : 5 columns and 103886 rows
     5. reviews: 6 columns and 99224 rows
     6. orders: 8 columns and 99441 rows
     7. products: 9 columns and 32951 rows
     8. sellers: 4 columns and 3095 rows
   * There are no Null values in any column of Customers, Geolocation, Order items, payments and sellers.
     1. Order\_reviews have 87675 nulls in review\_comment\_title
     2. Orders have 160 nulls in order\_approved\_at, 1783 nulls in order\_delivered\_carrier\_date and 2965 nulls in order\_delivered\_customer\_date.
     3. Products have 610 nulls in product\_category, name\_length , description\_length and photos\_qty each and 2 nulls in weight, length, height width each.
   * Data types:
     1. Customers



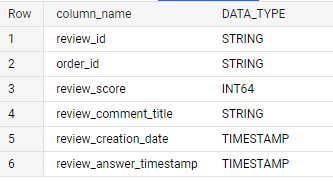
* + 1. Geolocation



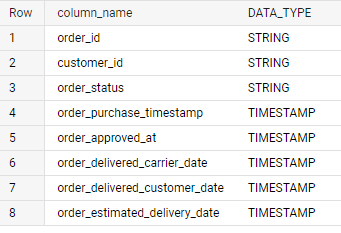
* + 1. Order items



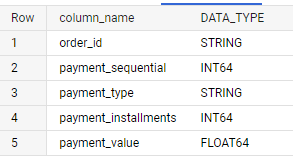
* + 1. Order reviews



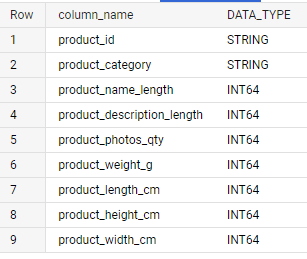
* + 1. Orders



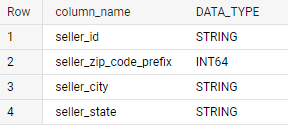
* + 1. Payments



* + 1. Products

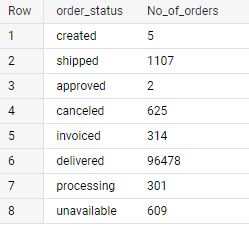


* + 1. Sellers



* + The data is available from 2016-09-04 till 2018-10-17
  + There are customers from 4119 cities in the dataset.
  + There are customers from 27 states in the dataset.

1. **In-depth exploration**
   * **How many orders do we have for each order status?**



* + **Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario?**

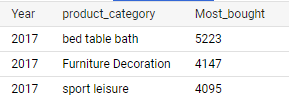
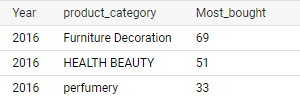
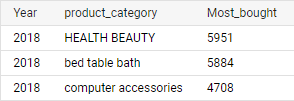
We can observe a substantial increase in the number of order per year and also per month of every year except for June and December for 2016.

Number of orders per month in 2018 are less than the previous months of 2017 but total orders are greater.

Therefore we can conclude that there is a growing trend on e-commerce.

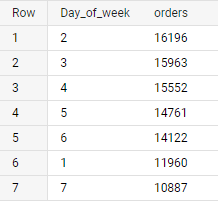
When observing top 3 most bought product\_categories every year we see that there is a trend of buying from Furniture Decoations, Health Beauty and bed table bath categories as compared to other categories.

* + **On what day of week brazilians customers tend to do online purchasing?**

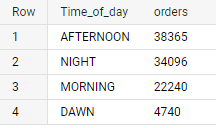
The Brazilian customers tend to do most online shopping on Monday (1), Tuesday (2) and Wednesday (3) and least on Saturday (7).

The sales on weekends, Saturday and Sundays are the least.



* + **What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?**

Customers tend to buy most during Afternoon followed by Night. They buy least during the Dawn, which is expected.



* + **Feature Extraction: Through order\_purchase\_timestamp in “orders” dataset extract**
    1. **order\_purchase\_year**
    2. **order\_purchase\_month**
    3. **order\_purchase\_date**
    4. **order\_purchase\_day**
    5. **order\_purchase\_dayofweek**
    6. **order\_purchase\_dayofweek\_name**
    7. **order\_purchase\_hour**
    8. **order\_purchase\_time\_day**

In the file attached

1. **Evolution of E-commerce orders in the Brazil region:**
   * **Get month on month orders by region**

In the file attached

* + **Total of customer orders by state**

In the file attached

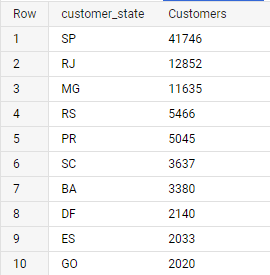
* + **Top 10 Brazilian cities most no. of orders**

In the file attached

* + **How are customers distributed in Brazil?**

We observe a very heavy majority (41746) customers stay in state with code SP.

Next state with most customers are from RJ followed by MG



* + **City wise number of unique customers**

In the file attached

1. **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**

**Step 1: Using CTE**

* 1. **“order\_items” + “order” joined on order id where order\_purchase timestamp is already divided into month & year**
  2. **Group data by year and month, aggregation count(order\_id), sum(price), sum(freight\_value)**
  3. **Create new columns:**

1. **price\_per\_order = sum(price) / count(order\_id)**
2. **freight\_per\_order= sum(freight\_value) / count(order\_id)**



**Step 2: Answer the following questions:**

1. **Total amount sold in 2017 between Jan to August**

3113000.32

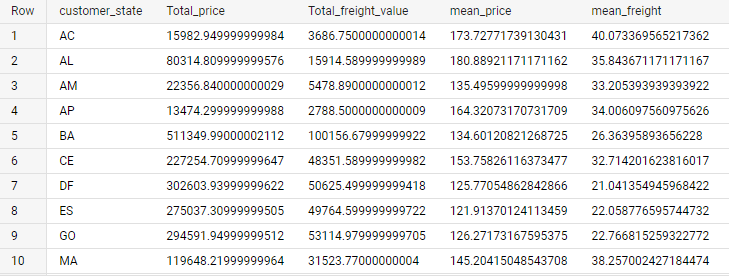
1. **Total amount sold in 2018 between Jan to august**

7385905.80

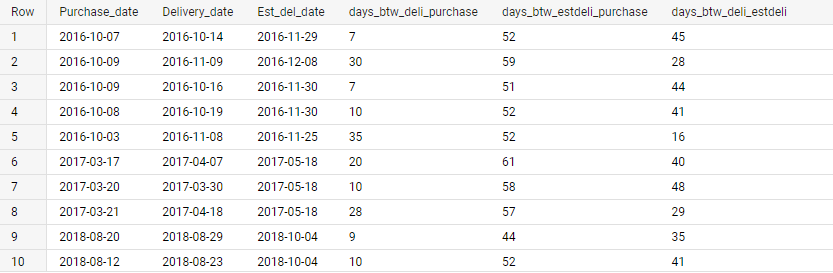
1. **% increase from 2017 to 2018**

**Step 3: Join (orders+order\_items) table from previous step with “customers” table on Customer\_id and find:**

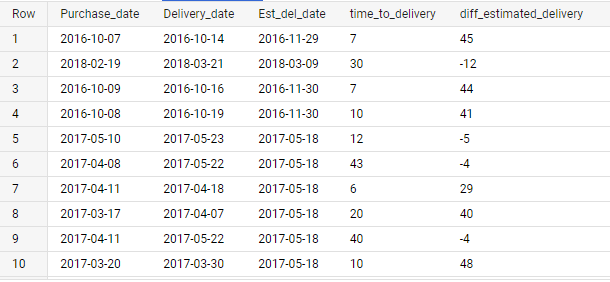
* + **Mean & Sum of price by customer state**
  + **Mean & Sum of freight value by customer state**



1. **Analysis on sales, freight and delivery time**
   * **Calculating days between purchasing, delivering and estimated delivery**

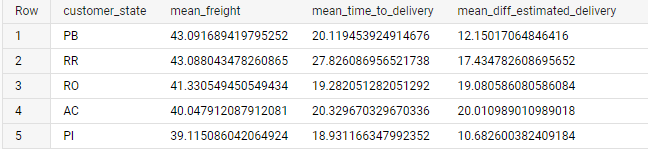


* + **Create columns:**
    1. **time\_to\_delivery = order\_purchase\_timestamp-order\_delivered\_customer\_date**
    2. **diff\_estimated\_delivery = order\_estimated\_delivery\_date-order\_delivered\_customer\_date**

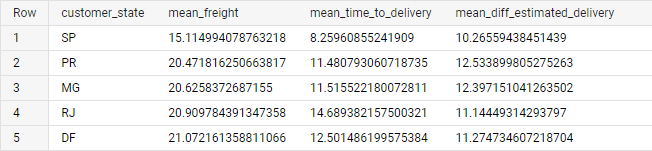


* + **Grouping data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery**
  + **Sort the data to get the following:**
    1. **Top 5 states with highest/lowest average freight value**

HIGHEST AVERAGE

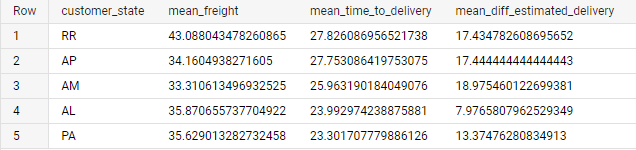


LOWEST AVERAGE

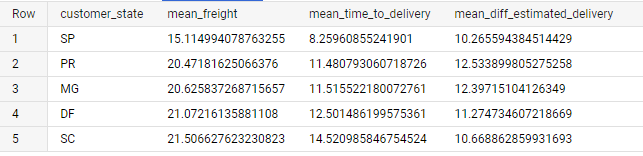


* + 1. **Top 5 states with highest/lowest average time to delivery**

HIGHEST TIME TO DELIVERY

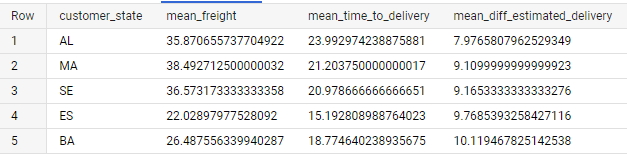


LOWEST TIME TO DELIVERY

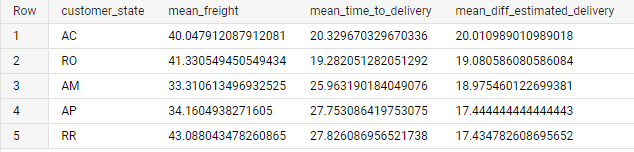


* + 1. **Top 5 states where delivery is really fast/ not so fast compared to estimated date**

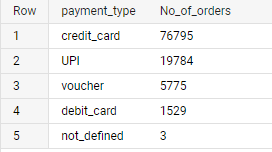
FAST



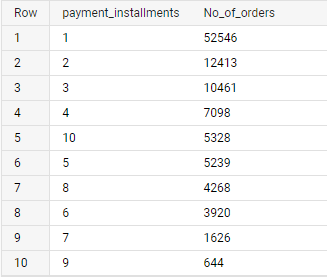
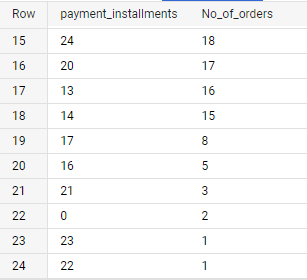
NOT SO FAST



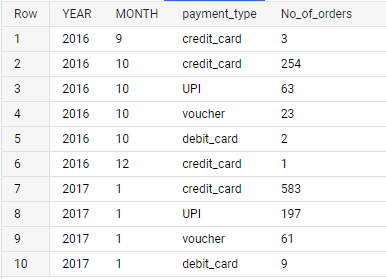
1. **Payment type analysis: Join “payments” dataset with the existing data on order\_id**
   1. **Count of orders for different payment types**



* 1. **Distribution of payment installments and count of orders**

* 1. **Count of orders for different payment types Month over Month**



1. **Actionable Insights**
   * Though the total orders in 2018 are more than 2017, there is a decrease in monthly no. of orders in 2018 against the steady increase observed in the 2017.
   * For 2017 and 2018 we observe after top 7 categories, by sales, the sale is very less for all other categories.   
     Also some categories are redundant with respect to some other, broader categories for example: Furniture- Room furniture, Art-Art and craft etc.
   * The sales on weekends is least compared to the week days.
   * There is a sharp gap between sales for dawn-morning and afternoon-night time.
   * A heavy majority of orders are from SP state. There is a higly significant difference in orders from other states compared to SP.
   * Similar scenario exists for cities with Sao Paulo have very high orders in comparison to other cities.
   * Though most orders being from SP the highest mean price of order is of PB.
   * Highest mean freight is of RR despite having very low orders and total price along with the longest mean time to delivery. These factors might be a reason for low sales in this state.
   * SP has lowest mean freight value along with fastest average time to delivery , which evidently is a factor for SP having highest sales.
   * Customers heavily prefer credit cards for payments compared to other modes.
   * A very high percentage of orders have 1 payment installment and only 2 orders have zero installments implying customers are more inclined towards paying for orders in 1 installment
2. **Recommendations**
   * Find the factors responsible low monthly orders in 2018.
   * Redefine the categories with sub categories for better categorization and tracking.
   * Providing special discounts or offers specifically targeting the low selling categories.
   * Any new launches / schemes / programs must be organized in afternoon/ night due to higher activity during this period.
   * Sale days / offers may start at dawn to increases orders during that time.
   * A deeper study is required for the factors that are resulting in very high sales in specific state/city and very low in others. After that only any valid recommendations can be made.
   * A strategy needs to be developed to reduce time to delivery and the freight values, to attract more customers.
   * Due to higher preference to credit cards, some loyalty programs may be planned in collaborations with card companies to expand the customer base.