BRAZILIAN E-COMMERCE ANALYSIS

Assignment Report

Submitted To

Mr. Winston
Mr. Mikio

Khalid Almalki, Hilah Alharthi and Afrah Alharbi

30/12/2020

Table of Contents

Introduction

Recently, using data science in e-commerce has been getting popular. Top e-commerce companies such as Amazon uses data science to detect customer's behavior and track the business performance.

In SDA boot-camp we were given this assignment to practice on data science methods such as Exploratory Data Analysis.

Assignment Objectives

- Explore a Brazilian ecommerce dataset.
- Exploratory data analysis using Tableau.
- Data munging using R.

Plan

Day	Time	Task
Sunday	4:00 p.m. - 4:30 p.m.	Searching for a Dataset
Sunday	4:40 p.m. - 5:40 p.m.	Making an ERD diagram
Sunday	5:50 p.m. - 7:30 p.m.	Making three visualizations in tableau
Sunday	7:40 p.m. - 8:00 p.m.	Writing notes about datasets
Monday	4:00 p.m. - 4:14 p.m.	Checking NULL value in dataset
Monday	4:15 p.m. - 4:20 p.m.	Conducting brainstorming for Data Munging
Monday	4:30 p.m. - 10:30 p.m.	Starting coding using R
Tuesday	4:00 p.m. - 10:30 p.m.	Completing coding
Wednesday	4:30 p.m. - 8:30 p.m.	Writing the assignment report
Thursday	4:00 a.m.- 6:00 a.m.	Submitting assignment

Table1 : Assignment Plan

Datasets Definition

This is a Brazilian ecommerce public dataset of orders made at Olist Store. Olist store which connect with small businesses from all over Brazil to channels with single contracts. Those tradesmen are able to sell their products through the Olist Store and ship them directly to the customers using Olist logistics partners. The dataset has information of 100k orders from 2016 to 2018. It shows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.

Datasets Source

The datasets were downloaded from Kaggle

Data Schema

The data is divided in multiple datasets for better understanding and organization. In the next figure, it shows how these datasets are connecting to each other.
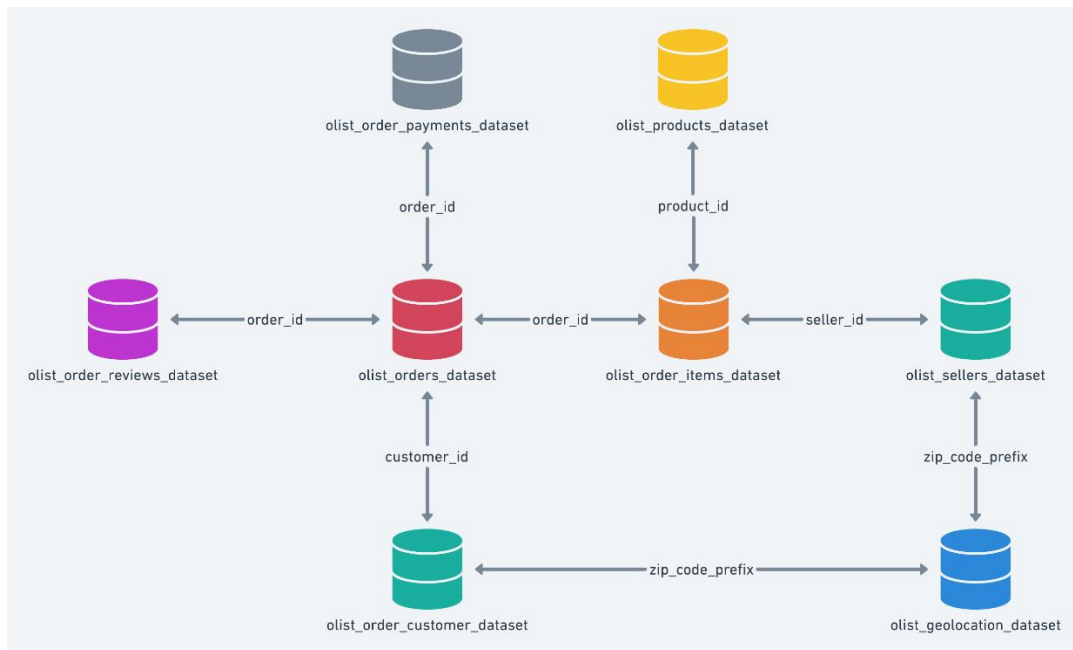

Figure 1: Data Schema

EDA Using Tableau
ERD

At the beginning, we had to create the relationship between all of the nine datasets to visualize them later.
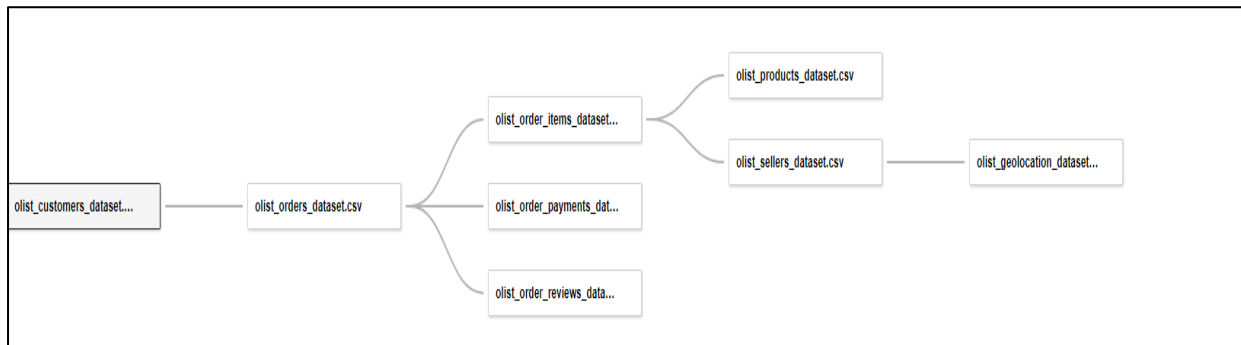

Figure2 : ERD

Tableau Visualizations

In this section we will show three visualizations.

First: Here we show customers who did repurchase the same order many times

most repurchased

customer_unique_id	
8d50f5eadf50201ccdcedf..	17
3e43e6105506432c953e..	9
1b6c7548a2a1f9037c1fd..	7
6469f99c1f9dfae7733b2..	7
ca77025e7201e3b30c44b..	7
12f5d6e1cbf93dafd9dcc1..	6
47c1a3033b8b77b3ab6e..	6
63cfc61cee11cbe306bff5..	6
dc813062e0fc23409cd25..	6
de34b16117594161a6a8..	6
f0e310a6839dce9de1638..	6
35ecdf6858edc6427223b..	5
394ac4de8f3acb14253c1..	5
4e65032f1f574189fb793..	5
56c8638e7c058b98aae6d..	5
5e8f38a9a1c023f3db718..	5
74cb1ad7e6d5674325c1f..	5
b4e4f24de1e8725b74e4a..	5
fe81bb32c243a86b2f86f..	5
083ca1aa470c280236380..	4
08e5b38d7948d37fbb2a5..	4
12d8b5ed661190a3a081..	4
25a560b9a6006157838a..	4
2ddc001b620bd90d0f437..	4
397b44d5bb99eabf54ea9..	4
4facc2e6fbc2bffab2fea92..	4
6358b15ef7fa9b1210f1a..	4
7305430719d715992b00..	4
738ffcf1017b584e9d268..	4
795c1622cf7a53d63d324..	4
83e7958a94bd7f74a9414..	4
9cc5a07f169a1606fd347..	4
a0abd26659c9914690cbd..	4
a1874c5550d2f0bc14cc1..	4
a239b8e2fbce33780f1f1..	4
a73c5464692d3bc09d64e..	4
a7657330b1c135f3acd42..	4
acea6bd29b8c1e3c6a8b2..	4

Figure3 : Most repurchased customers

Second: Here we show the most sold products and we found that the "movies decorate" was the highest desired product

most sold products

product_id (olist_product..	product_category_name	
aca2eb7d00ea1a7b8ebd4..	moveis_decoracao	527
99a4788cb24856965c36a..	cama_mesa_banho	488
422879e10f46682990de2..	ferramentas_jardim	484
389d119b48cf3043d3113..	ferramentas_jardim	392
368c6c730842d78016ad8..	ferramentas_jardim	388
53759a2ecddad2bb87a07..	ferramentas_jardim	373
d1c427060a0f73f6b889a..	informatica_acessorios	343
53b36df67ebb7c41585e8..	relogios_presentes	323
154e7e31ebfa092203795..	beleza_saude	281
3dd2a17168ec895c781a9..	informatica_acessorios	274
2b4609f8948be1887449..	beleza_saude	260
7c1bd920dbdf22470b68b..	beleza_saude	231
a62e25e09e05e6faf31d9..	relogios_presentes	226
5a848e4ab52fd5445cdc0..	Null	197
bb50f2e236e5eea010068..	beleza_saude	195
e0d64dcfaa3b6db5c54ca..	relogios_presentes	194
e53e557d5a159f5aa2c5e..	informatica_acessorios	183
42a2c92a0979a949ca4ea..	utilidades_domesticas	183
b532349fe46b38fbc7bb3..	moveis_decoracao	169
35afc973633aaeb6b877ff..	casa_conforto	165
a92930c327948861c015c..	relogios_presentes	160
6cdd53843498f92890544..	beleza_saude	156
19c91ef95d509ea33eda9..	beleza_saude	156
437c05a395e9e47f9762e..	beleza_saude	155
f1c7f353075ce59d8a6f3c..	cama_mesa_banho	154
3fbc0ef745950c7932d5f2..	beleza_saude	150
c4baedd846ed09b85f78a..	ferramentas_jardim	149
461f43be3bdf8844e65b6..	relogios_presentes	146
7a10781637204d8d1048..	relogios_presentes	143
06edb72f1e0c64b14c5b7..	cama_mesa_banho	143
0aabfb375647d9738ad0f..	consoles_games	142
8c591ab0ca519558779df..	ferramentas_jardim	141
656e0eca68dcecf6a31b8..	informatica_acessorios	141
d017a2151d543a988560..	fashion_bolsas_e_acesso..	140
2028bf1b01cafb2d2b190..	perfumaria	138
e0cf79767c5b016251fe1..	beleza_saude	137
ec2d43cc59763ec916945..	cama_mesa_banho	131
b0961721fd839e998242	ferramentas_jardim	127

Figure 4: Most desired products

Third: Here we show the amount of payment over years to see how the Brazilian E-commerce performed, and we obtain that we had either disaster or missing data in the end of 2018.
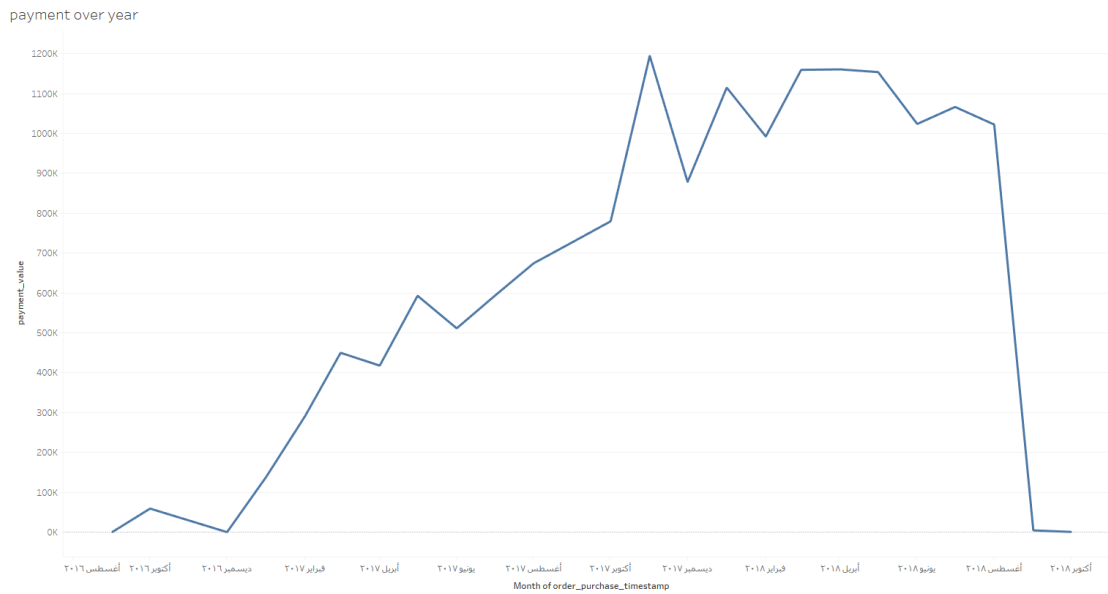
payment over year

Figure 5 : Payment value over years

Problems:

We faced a problem in linking a dataset with two different datasets at the same time. Unfortunately, we do not know how to solve it and we tried our best but we have not reached an answer.

Data Munging Using R

As we noticed that a lot of data are not cleaned and also there are numbers of Null values, we decided to munge the data using R.

Data Munging Processes

Importing Libraries: At the first we imported all related libraries so we can manipulate the date in later phasis. We imported Dplyr, Tidyverse and Lubridate libraries.

Importing Datasets: The Brazilian E-Commerce database has nine datasets. Each of them represents a table and has it's own primary key. In our case we imported all of the datasets but actually we only needed three of them to be cleaned.
They are:

- olist_orders_dataset
- olist_order_reviews_dataset
- olist_products_dataset

olist_orders_dataset: This dataset was the one that we spent most of the time working on it. We faced some challenges and difficulties in solving it but finally we succeed.

Exploring to find problems:

Column Name	Count NULL	Delete / Replace	Reason	How to fix it
order_approved_at	1	Delete	no effect	----
order_delivered_carrier_date	12	Delete	no effect	----
order_delivered_customer_date	1780	Replace	we can fix it .	median (delivered customer date - delivered carrier date)

Table 2: Null values in Order dataset

First, we found numbers of Null values in some features that could not be replaced as they were given from the system. Those features were:

- order_approved_at
- order_delivered_carrier_date

Second, we assumed that the date of the order_delivered_carrier_date (The date when the shipment is being shipped) are always greater than the order_delivered_customer_date (The date when the customer receive the shipment). We found interesting results that there are some records where the date did not match the assumption.

Third, we found numbers of Null values in order_delivered_customer_date and we assumed it could be fixed and replaced with meaning values.

How did we handle the mentioned issues?

In the first case, we decided to delete them as we could not fill them with any value.

In the second case, as there were only 23 invalid records, we deleted them.

The third case was the most interesting case. We handled it in three steps:

- We calculated the median of the difference time between order_delivered_customer_date and order_delivered_carrier_date, without including Null values.
- We converted the median time from seconds to days.

```
> diffirent_time_in_second_after_correction=
+   median(as.POSIXct(df$order_delivered_customer_date)-
+           as.POSIXct(df$order_delivered_carrier_date))
> diffirent_time_in_second_after_correction
Time difference of 613453.5 secs
```

Figure 6: Median value of the shipping time in Seconds

```
> d_after_correction=(round(h_after_correction/24))
```

order_delivered_carrier_date	order_delivered_customer_date
2018-06-05 14:32:00	

Figure 8: Order_delivered_customer_date with Null value

- For every Null value in order_delivered_customer_date, we calculated it by adding the median time to the date of order_delivered_carrier_date of the same record.

  In Figure 8, it shows how the order_delivered_customer_date is Null. Then we will fill it with the aggregation of the value of order_delivered_carrier_date plus the median value which is 7 days, so the

order_delivered_carrier_date	order_delivered_customer_date
2018-06-05 14:32:00	2018-06-12 14:32:00

Figure 9: Order_delivered_customer_date after being calculated

  date will be: 2018-06-12 14:32:00 as shown in the next figure.

Once we had completed these steps, we had a cleaned dataset.

olist_order_reviews_dataset:

Exploring to find problems:

Column Name	Count Null	Delete / Replace	Reason	How to fix it
review_comment_title	80% of the data	Replace	we can fix it	replace it with (" comment without title")
review_comment_message	80% of the data	Replace	we can fix it	replace it with (" comment without message")

Table 3: Null values in Reviews dataset

In this dataset we found only two problems:

- Null values in review_comment_title column
- Null values in review_comment_message column

How did we handle the mentioned issues?
For both features, we fill comments in every Null value. The reason of not deleting them is that, Null values cover %80 of the dataset and we would loss a lot of useful data.

olist_products_dataset:
Exploring to find problems:

Column Name	Count Null	Delete / Replace	Reason
product_category_name	30	Delete	we can't replace it with a product name
product_name_lenght	30	Delete	we don't have product _name
product_description_lenght	30	Delete	we don't have product _name
product_photos_qty	30	Delete	we don't have photos

Table 4: Null values in Product dataset

In this dataset we found four problems:
- Null values in product_category_name column
- Null values in product_name_length column
- Null values in product_description_lenght column
- Null values in product_photos_qty column

How did we handle the mentioned issues?
For all mentioned features, we deleted them as the main issue is with the column product_category_name which is the primary key and the core features.

Outcome Samples

- olist_orders_dataset:

	order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date	order_delivered_customer_date	order_estimated_
45	ee64d42b8cf066f35eac1cf57de1aa85	caded193e8e47b8362864762a83db3c5	shipped	2018-06-04 16:44:48	2018-06-05 04:31:18	2018-06-05 14:32:00	2018-06-12 14:32:00	2018-06-28 00
46	6ebaec694d7025e2ad4a05dba887c032	4f28355e5c17a4a42d3ce2439a1d4501	delivered	2017-05-18 13:55:47	2017-05-18 14:05:17	2017-05-19 12:01:38	2017-05-29 12:47:20	2017-06-09 00
47	d17dc4a904426827ca80f2ccb3a6be56	569cf68214806a39acc0f39344aea67f	delivered	2017-05-14 20:28:25	2017-05-14 20:42:45	2017-05-16 08:17:46	2017-05-25 09:14:31	2017-06-12 00
48	25f4376934e13d3508486352e11a5db0	12fd2740039676063a874b9567dfa651	delivered	2018-05-17 16:59:11	2018-05-18 01:17:39	2018-05-18 13:02:00	2018-05-21 15:22:11	2018-05-25 00
49	5820a1100976432c7968a52da59e9364	2b56e94c2f66f2d97cfa63356f69cee8	delivered	2018-07-29 11:24:17	2018-07-29 11:44:19	2018-07-30 13:47:00	2018-08-02 22:09:11	2018-08-13 00
50	2ce1ad82022c1ba30c2079502ac725aa	7f2178c5d771e17f507d3c1637339298	delivered	2017-08-09 20:19:05	2017-08-11 04:15:29	2017-08-11 17:52:32	2017-08-16 17:16:44	2017-08-31 00
51	138849fd84dff2fb4ca70a0a34c4aa1c	9b18f3fc296990b97854e351334a32f6	delivered	2018-02-01 14:02:19	2018-02-03 02:53:07	2018-02-06 19:13:26	2018-02-14 13:41:59	2018-02-23 00
52	47aa4816b27ba60ec948cd019cc1afc1	148348ff65384b4249b762579532e248	delivered	2018-06-26 13:42:52	2018-06-27 08:35:32	2018-06-27 13:20:00	2018-07-03 18:37:46	2018-07-20 00
53	9faeb9b2746b9d7526aef5acb08e2aa0	79183cd650e2bb0d475b0067d45946ac	delivered	2018-07-26 14:39:59	2018-07-26 14:55:10	2018-07-27 12:04:00	2018-07-31 22:26:55	2018-08-16 00
54	641fb0752bf5b5940c376b3a8bb9dc52	f5afca14dfa9dc64251cf2b45c54c363	delivered	2017-12-15 00:06:10	2017-12-15 00:14:55	2017-12-19 01:58:00	2018-01-03 15:09:32	2018-01-16 00
55	e425680f760cbc130be3e53a9773c584	f178c1827f67a8467b0385b7378d951a	delivered	2017-08-31 08:15:24	2017-08-31 08:30:17	2017-08-31 20:06:14	2017-09-04 20:59:55	2017-09-20 00
56	40c5e18f7d112b59b3e5113a59a905b3	67407057a7d5ee17d1cd09523f484d13	delivered	2018-06-11 10:25:52	2018-06-11 10:58:32	2018-06-14 13:03:00	2018-06-19 00:31:13	2018-07-16 00
57	734e7d1bbaeb2ff82521ca0fe6fb6f79	2932d241d1f31e6df6c701d52370ae02	delivered	2018-06-11 08:18:19	2018-06-11 08:31:50	2018-06-11 14:54:00	2018-06-14 21:32:21	2018-07-05 00
58	66e4624ae69e7dc89bd50222b59f581f	684fa6da5134b9e4dab731e00011712d	delivered	2018-03-09 14:50:15	2018-03-09 15:40:39	2018-03-15 00:31:19	2018-04-03 13:28:46	2018-04-02 00
59	a685d016c8a26f71a0bb67821070e398	911e4c37f5cafe1604fe6767034bf1ae	delivered	2017-03-13 18:14:36	2017-03-13 18:14:36	2017-03-22 14:03:09	2017-04-06 13:37:16	2017-03-30 00
60	2edfd6d1f0b4cd0db4bf37b1b224d855	241e78de29b3090cfa1b5d73a8130c72	delivered	2017-06-13 21:11:26	2017-06-15 03:05:45	2017-06-16 14:55:37	2017-06-19 18:51:28	2017-07-06 00
61	68873cf91053cd11e6b49a766db5af1a	4632eb5a8f175f6fe020520ae0c678f3	delivered	2017-11-30 22:02:15	2017-12-02 02:51:18	2017-12-04 22:07:01	2017-12-05 20:28:40	2017-12-18 00
62	f346ad4ee8f630e5e4ddaf862a34e6dd	dd5095632e3953fc0947b8ab5176b0be	delivered	2018-08-05 13:09:48	2018-08-05 13:24:34	2018-08-06 13:41:00	2018-08-10 18:35:40	2018-08-15 00
63	8f06cc6465925031568537b815f1198d	9916715c2ab6ee1710c9c32f0a534ad2	delivered	2017-11-15 11:31:41	2017-11-15 11:46:42	2017-11-16 22:03:00	2017-11-22 22:41:07	2017-12-05 00
64	ccbabeb0b02433bd0fcbac46e70339f2	c77ee2d8ba1614a4d489a44166894938	delivered	2018-02-19 20:31:09	2018-02-21 06:15:25	2018-02-22 21:04:23	2018-03-09 22:22:25	2018-03-13 00

Showing 45 to 66 of 97,621 entries, 8 total columns

Invalid records:

order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date	order_delivered_customer_date	order_estimated_delivery_
a1abeb653a4d4cd1e142ccb8c82cd069	5f50465da00b7fed5dd1239f4ecf6e2c	delivered	2017-07-20 11:20:52	2017-07-21 06:43:14	2017-07-28 16:57:58	2017-07-25 19:32:56	2017-08-14 00:00:00
383aa8b2724fe452d9ccd9934a8c628b	b1cb2f9d7a19480f3749e248db14d58f	delivered	2017-07-02 20:58:43	2017-07-02 21:10:20	2017-07-07 17:22:41	2017-07-06 14:27:51	2017-07-21 00:00:00
cb1134f9010d242e9515ad1c78ec0c39	2fd33ac77677bd214b1882868317eeed	delivered	2017-07-16 12:35:34	2017-07-18 06:03:50	2017-07-20 19:22:02	2017-07-19 14:13:28	2017-08-08 00:00:00
dceb62e8fa94b46006c9554fed743df0	2721900eb4e0f1cc2c836dd7bc1b1e11	delivered	2017-07-20 20:58:05	2017-07-22 11:45:11	2017-08-01 18:23:30	2017-07-26 18:09:10	2017-08-11 00:00:00
5f9d46795c3126674e52becb3a1a517f	79287bcaafdde5c793b996fc40bb7d9f	delivered	2017-07-18 11:48:20	2017-07-18 12:03:29	2017-07-20 23:03:42	2017-07-20 18:52:41	2017-07-31 00:00:00
8c78d01de3a9009e23d6877a7cc9be20	6cd7106899e59a1fbd0622d5f1efedf4	delivered	2016-10-08 15:36:50	2016-10-08 18:13:44	2016-10-26 11:41:53	2016-10-25 17:51:46	2016-11-30 00:00:00
b27af682321527a6349f1761eb3f360c	9859dd92e872dbaa60ca3cd5f0d7ad07	delivered	2017-06-14 20:17:04	2017-06-14 20:30:08	2017-06-27 14:51:54	2017-06-26 15:45:35	2017-07-14 00:00:00
1cc3ae63caffff2d6c3ee3e78e074acf	01c843a2c0600def0b7693dba47af460	delivered	2017-08-07 21:35:22	2017-08-08 21:45:15	2017-08-10 18:28:56	2017-08-10 18:05:38	2017-08-25 00:00:00
e37f11cae9985ca58f0b56f268720537	3947a361301f2ff0f3223159a0f2701c	delivered	2017-07-26 11:46:34	2017-07-27 10:10:16	2017-08-01 18:17:47	2017-07-31 17:49:56	2017-08-24 00:00:00
fa3e37584f4fdb1ded0e0de700dfcb4e	63be4feff10a0b1d85f2cfbf10df9754	delivered	2017-07-30 19:32:23	2017-07-30 19:45:09	2017-08-09 18:18:43	2017-08-01 21:13:01	2017-08-18 00:00:00
c1e2bf2b7dd3309f2f5356c6b63968fa	e37d47e7eec62f08dc5deecc7d5532d6	delivered	2017-02-10 10:19:10	2017-02-10 10:30:13	2017-03-02 17:34:26	2017-02-14 15:15:57	2017-03-15 00:00:00
b866af202be0692766081310cd4085e1	d1800078046ed2e5ae1b0792b695c56e	delivered	2017-01-27 14:59:17	2017-01-27 15:30:46	2017-02-20 02:32:08	2017-02-15 03:53:46	2017-04-17 00:00:00
29941903985f944b0ffc49c479c1547d	b56ee98181afc3948a758d73a08423de	delivered	2017-05-29 16:16:50	2017-05-29 16:25:16	2017-06-09 15:07:29	2017-06-02 11:09:16	2017-06-23 00:00:00
76458889992169d3135b264dc13aec67	999196dca58a3db3d966d8f148532010	delivered	2016-10-07 10:05:16	2016-10-07 11:24:43	2016-10-26 11:43:06	2016-10-20 18:03:17	2016-11-29 00:00:00
0922ee1619de7b995648e5a8407afb91	8dab637615e9eae6d33ef5e48644d6d3	delivered	2017-07-11 10:45:44	2017-07-11 10:55:15	2017-07-14 12:41:34	2017-07-12 20:49:42	2017-08-14 00:00:00
fbffbaaa30f4077e87ec6acdb835b278	a31e61974efff601345e65e5d6b7ff31	delivered	2017-07-30 17:58:26	2017-07-31 07:05:07	2017-08-04 20:04:17	2017-08-03 20:30:00	2017-08-18 00:00:00
3ee6694bb99d5a6db0deceffcd099597	127237a4aa0bd9ecc09015f781c9fc27	delivered	2017-07-07 11:22:13	2017-07-07 12:25:07	2017-07-11 18:15:40	2017-07-11 17:41:26	2017-07-20 00:00:00
19feb5627c41ea1b36a8e50a469b3644	b8097c8f0c1f58ab56a53812a446a898	delivered	2016-10-07 17:09:56	2016-10-07 17:32:09	2016-10-26 11:42:05	2016-10-20 19:07:54	2016-12-01 00:00:00
d5558a097766363b8e76b38c43332e8a	49a4fe701c3d30eac6c180908e665ebb	delivered	2017-02-04 19:01:33	2017-02-04 19:31:12	2017-02-15 08:55:26	2017-02-10 07:58:32	2017-03-01 00:00:00
771c4f1f521f462e4b95619e648aaeab	d424b3ce4c850247e8c84ff8752de868	delivered	2017-03-22 20:15:15	2017-03-22 20:15:15	2017-03-30 13:14:34	2017-03-28 17:28:32	2017-04-12 00:00:00

Showing 1 to 21 of 23 entries, 8 total columns

- olist_order_reviews_dataset

	review_id	order_id	review_score	review_comment_title	review_comment_message	review_creation_date	review_answer_timestamp
1	7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb	4	comment without title	comment without message	2018-01-18 00:00:00	2018-01-18 21:46:59
2	80e641a11e56f04c1ad469d5645fdfde	a548910a1c6147796b98fdf73dbeba33	5	comment without title	comment without message	2018-03-10 00:00:00	2018-03-11 03:05:13
3	228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b	5	comment without title	comment without message	2018-02-17 00:00:00	2018-02-18 14:36:24
4	e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b	5	comment without title	Recebi bem antes do prazo estipulado.	2017-04-21 00:00:00	2017-04-21 22:02:06
5	f7c4243c7fe1938f181bec41a392bdeb	8e6bfb81e283fa7e4f11123a3fb894f1	5	comment without title	ParabÌ©ns lojas lannister adorei comprar pela Internet segu...	2018-03-01 00:00:00	2018-03-02 10:26:53
6	15197aa66ff4d0650b5434f1b46cda19	b18dcdf73be66366873cd26c5724d1dc	1	comment without title	comment without message	2018-04-13 00:00:00	2018-04-16 00:39:37
7	07f9bee5d1b850860defd761afa7ff16	e48aa0d2dcec3a2e87348811bcfdf22b	5	comment without title	comment without message	2017-07-16 00:00:00	2017-07-18 19:30:34
8	7c6400515c67679fbee952a7525281ef	c31a859e34e3adac22f376954e19b39d	5	comment without title	comment without message	2018-08-14 00:00:00	2018-08-14 21:36:06
9	a3f6f7f6f433de0aefbb97da197c554c	9c214ac970e84273583ab523dfafd09b	5	comment without title	comment without message	2017-05-17 00:00:00	2017-05-18 12:05:37
10	8670d52e15e00043ae7de4c01cc2fe06	b9bf720beb4ab3728760088589c62129	4	recomendo	aparelho eficiente. no site a marca do aparelho esta impress...	2018-05-22 00:00:00	2018-05-23 16:45:47
11	c9cfd2d5ab5911836ababae136c3a10c	cdf9aa68e72324eeb25c7de974696ee2	5	comment without title	comment without message	2017-12-23 00:00:00	2017-12-26 14:36:03
12	96052551d87e5f62e6c9f6974ec392e9	3d374c9e46530bb5ed4a7648915306a6	5	comment without title	comment without message	2017-12-19 00:00:00	2017-12-20 10:25:22
13	4b49719c8a200003f700d3d986ea1a19	9d6f15f95d01e79bd1349cc208361f09	4	comment without title	Mas um pouco ,travando...pelo valor ta Boa.	2018-02-16 00:00:00	2018-02-20 10:52:22
14	23f75a37effc35d9a915b4e1ad483793	2eaf8e099d871cd5c22b83b5ea8f6e0e	4	comment without title	comment without message	2018-03-28 00:00:00	2018-03-30 15:10:55
15	9a0abbb668bafb95a6d2b05db43284c4	d7bd0e4afdf94846eb73642b4e3e75c3	3	comment without title	comment without message	2017-04-30 00:00:00	2017-05-03 00:02:22
16	3948b09f7c818e2d86c9a546758b2335	e51478e7e277a83743b6f9991dbfa3fb	5	Super recomendo	Vendedor confiÌvel, produto ok e entrega antes do prazo.	2018-05-23 00:00:00	2018-05-24 03:00:01
17	9314d6f9799f5bfba510cc7bcd468c01	0dacf04c5ad59fd5a0cc1faa07c34e39	2	comment without title	GOSTARIA DE SABER O QUE HOUVE, SEMPRE RECEBI E ESS...	2018-01-18 00:00:00	2018-01-20 21:25:45
18	8e15a274d95600fa14f8be64e37a0e67	ff1581e08b3011021e7c7de592ddc81e	5	comment without title	comment without message	2018-03-24 00:00:00	2018-03-26 15:58:32
19	fdbdb2629a7cde0f66657acc92084e7f	70a752414a13d09cc1f2b437b914b28e	3	comment without title	comment without message	2017-09-29 00:00:00	2017-10-02 01:12:49
20	373cbeecea8286a2b66c97b1b157ec46	583174fbe37d3d5f0d6661be3aad1786	1	NÌ£o chegou meu produto	PÌ©ssimo	2018-08-15 00:00:00	2018-08-15 04:10:37

Showing 1 to 21 of 100,000 entries, 7 total columns

- olist_products_dataset

	product_id	product_category_name	product_name_lenght	product_description_lenght	product_photos_qty	product_weight_g	product_length_cm	product_height_cm	produ
1	1e9e8ef04dbcff4541ed26657ea517e5	perfumaria	40	287	1	225	16	10	
2	3aa071139cb16b67ca9e5dea641aaa2f	artes	44	276	1	1000	30	18	
3	96bd76ec8810374ed1b65e291975717f	esporte_lazer	46	250	1	154	18	9	
4	cef67bcfe19066a932b7673e239eb23d	bebes	27	261	1	371	26	4	
5	9dc1a7de274444849c219cff195d0b71	utilidades_domesticas	37	402	4	625	20	17	
6	41d3672d4792049fa1779bb35283ed13	instrumentos_musicais	60	745	1	200	38	5	
7	732bd381ad09e530fe0a5f457d81becb	cool_stuff	56	1272	4	18350	70	24	
8	2548af3e6e77a690cf3eb6368e9ab61e	moveis_decoracao	56	184	2	900	40	8	
9	37cc742be07708b53a98702e77a21a02	eletrodomesticos	57	163	1	400	27	13	
10	8c92109888e8cdf9d66dc7e463025574	brinquedos	36	1156	1	600	17	10	
11	14aa47b7fe5c25522b47b4b29c98dcb9	cama_mesa_banho	54	630	1	1100	16	10	
12	03b63c5fc16691530586ae020c345514	bebes	49	728	4	7150	50	19	
13	cf55509ea8edaaac1d28fdb16e48fc22	instrumentos_musicais	43	1827	3	250	17	7	
14	7bb6f29c2be57716194f96496660c7c2	moveis_decoracao	51	2083	2	600	68	11	
15	eb31436580a610f202c859463d8c7415	construcao_ferramentas_seguranca	59	1602	4	200	17	7	
16	3bb7f144022e6732727d8d838a7b13b3	esporte_lazer	22	3021	1	800	16	2	
17	6a2fb4dd53d2cdb88e0432f1284a004c	perfumaria	39	346	2	400	27	5	
18	a1b71017a84f92fd8da4aeefba108a24	informatica_acessorios	59	636	1	900	40	15	
19	a0736b92e52f6cead290e30b578413b6	moveis_decoracao	56	296	2	1700	100	7	
20	f53103a77d9cf245e579ea37e5ec51f0	cama_mesa_banho	52	206	1	500	16	10	

Showing 1 to 21 of 32,340 entries, 9 total columns

Conclusion

In the conclusion, we would like to highlight some interesting notices about the datasets:

- As shown in Figure 5, there was an incredible decrease in payment values at the end of Year 2018.
- Some of order_delivered_customer_date values were invalid and that implies there are some problems in the system.
- When we tried to import the datasets to R, more than 4000 records were not imported and we found all of these records did not have the complete structure of the data frame so the R did not import them. However, all of these records were imported in Tableau.

References

Assignment code:
- https://rpubs.com/Afrah_H/708744

Datasets reference:
- Kaggle Datasets

Learning Sources:
- R Documentation
- neonscience
- stackoverflow.
- r-bloggers
- unlist
- asDate Function