Data Scientist – Recruitment Exercise

By: Narjes Rohani

Contact me: Nasim.Rohani74@gmail.com

The management of sell-your-stuff.com wants to set up marketing strategies for the next fiscal year based on the summary tables. Your job is to analyze the performance of past campaigns, coupons, profit by countries, etc. and make a report showing information to support your insights.

Solution:

In the first step, I used python language with panda, numpy, seaborn, matplot and other libraries. I read data from sqlite database and created description from its 3 tables. A brief description of each table is presented.

Table 1: Description of Clients Table

	account	indication_coupon	first_deposit_amount	first_transaction_amount	balance_amount
count	5000	5000	5000	5000	5000
mean	3499.5	280.059	6312.718	32.47043	-286586
std	1443.52	72.95865	6166.56	216.0412	1620363
min	1000	0	0	0	-2.9E+07
25%	2249.75	308	104.975	0	-90909.1
50%	3499.5	308	9090.91	0.111216	0
75%	4749.25	308	9090.91	1.821667	9089.72
max	5999	308	100000	6887.185	355184.2

There are 5000 users in the clients table. As it can be seen from table 1, mean of first deposit, first transaction , and balance amounts are 6312.718, 32.47043, and -286586, respectively. The negative mean of balance amount denotes that the average amount of users' debits are more than their credits, furthermore the 50% quartile of balance amount is 0, which means that half of users have negative balance amount. Moreover, the high values of standard deviation yields that there are high variety between users in regard to their first deposit, and balance amounts. By looking at indication coupon column, one can see that indication coupon 308 is the most popular. Other cells of table 1 are obvious.

Table 2: Description of Transactions Table

	account	total_buy	total_sell	total_deposits	total_withdrawals	count_contracts	count_deposits	count_withdrawals
count	36747	36747	36747	36747	36747	36747	36747	36747
mean	3098.361	2100.561	1619.568	1538.843	40052.37	255.9855	0.292786	0.08485
std	1361.246	10336.14	8095.118	5216.912	48857.17	4529.95	0.707798	0.524256
min	1001	0	0	0	0	0	0	0
25%	1873.5	2.06	1.68	0	0	3	0	0
50%	3153	19.57	15.82	0	0	15	0	0

75%	4052.5	200	157.885	0	100000	68	0	0
max	5999	100000	100000	100000	100000	248789	17	33

As it can be seen from table 2, mean of total buy and sell are 2100.561, and 1619.568, respectively. High values of standard deviation in total buy, total sell shows the users' diversity in amount of selling and buying. Furthermore, high standard deviation of total deposits and total withdrawals beside the low values of standard deviation of count deposit and count withdrawals denotes that almost all users have the same number of deposits and withdrawals but with diverse amount. Most of users have no deposits, and withdrawals according to the values of 25%, 50%, and 75% quartiles. Even the maximum values of deposits and withdrawals are very low. However, the maximum values of total deposits and total withdrawals are so high, which means that there a few users that have very high amounts of deposits and withdrawals. Furthermore, the maximum values of total buy and total sell are 10000; so there is no user that has a transaction higher than 10000 dollars.

Table 3: **Description of Campaigns Table**

	total_spend
count	43
mean	12004.85
std	6862.693
min	745.89
25%	7499.46
50%	12463.88
75%	16249.62
max	29866.08

It can be seen that the campaigns usually have about 12000 dollars expenses, according to the mean and 50% quartile values. However, the best campaign had 29866.08 dollars expenses. Moreover, the high value of standard deviation shows that different campaign yields different amount of expenses.

Analyzing clients and transaction tables data

Where are our users from? Lets get familiar with our clients!

41% of accounts are from **Indonesia** (id), 10% from **Brazil** and 0.09 from **Russian** Federation. I used **Tableau** to create dashboard report.
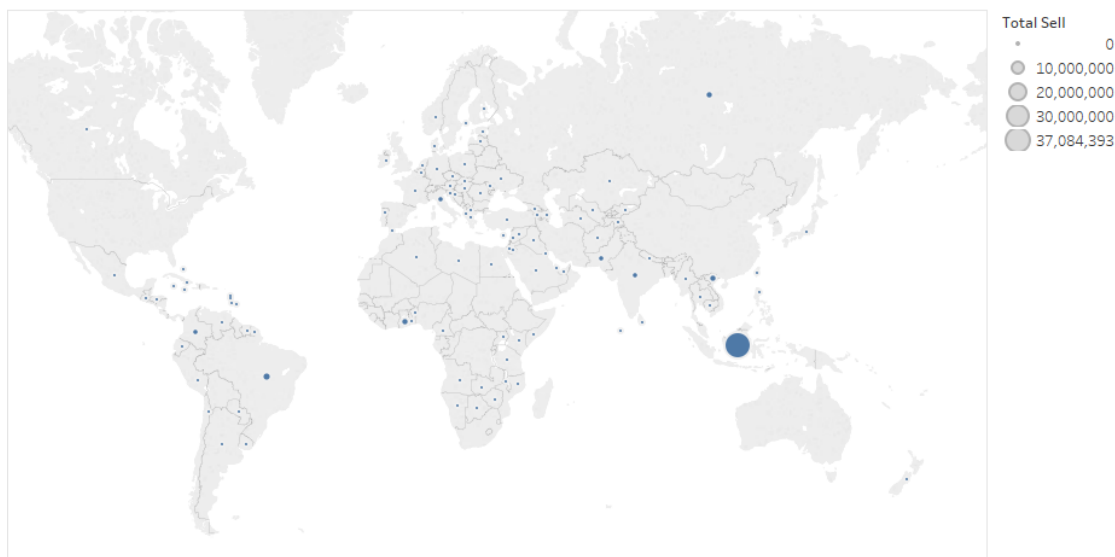
Number of accounts in each country

Map based on Longitude (generated) and Latitude (generated). Size shows COUNT([Account]) / TOTAL(count([Account])). Details are shown for Residence. The view is filtered on Latitude (generated) and Longitude (generated). The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only.

Which Country has the most amount of total sell? Where should we focus on?

The following figure shows the amount of total sell in each country. It is not surprising that most of total sells were in the countries with more users. Such as Indonesia with about 37 billion total sells and Brazil with about 2 billion total sells.

Total sell in each country

Map based on Longitude (generated) and Latitude (generated). Size shows sum of Total Sell. Details are shown for Residence.

When did most of the accounts join?

As you can see in the picture, by 2017, more people have registered on the site.

Number of joined accounts on each month, quarter, and month

The trend of count of Account for Date Joined Month broken down by Date Joined Year and Date Joined Quarter.
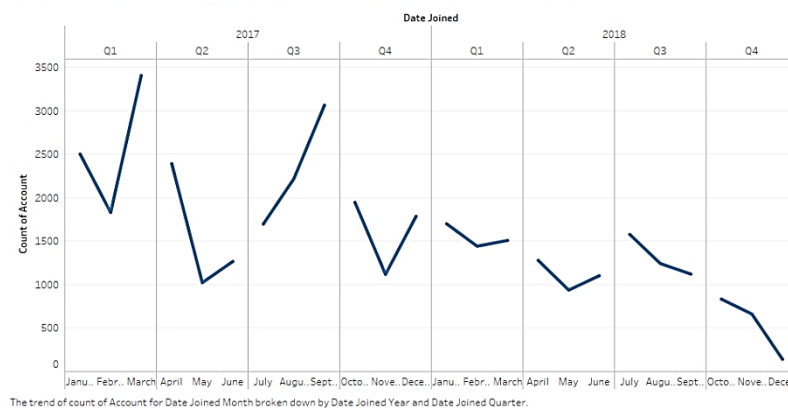
The best month regarding to the number of accounts joining us, is the March 2017 that 3405 users joined. The trend of number of joined accounts is decreasing. This may be due to the novelty of website, curiosity of users or the advertisement and marketing policies in the beginning, but this trend can't be ascending or even constant through the future months and year. In 2018, fewer users joined our website. Even, there is reduction in number of joined accounts in each quarter of 2018. As it can be seen, the worst month is the last month that 141 users joined. It should be noted that this plot is drew monthly, and can be plotted daily, yearly, and etc.

What are the types of client accounts?

Most of accounts are type 6 and type 1. About 62% of users are type 6 and about 0.35% of them are type 1; thus, it is rational that most of buys and sells are done by these two types as it can be seen from two nest figures. And 0% of accounts are type 0.

Percentage of each user type
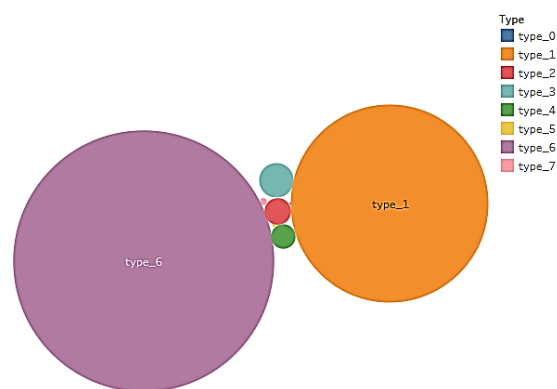
Type. Color shows details about Type. Size shows COUNT([Account]) / TOTAL(COUNT([Account])). The marks are labeled by Type.

Total buy on each type
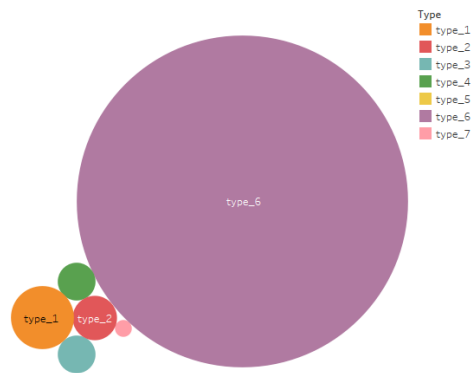
Type. Color shows details about Type. Size shows sum of Total Buy. The marks are labeled by Type. The view is filtered on sum of Total Buy, which keeps non-Null values only.

## Relation of types with total buy, withdrawl, sell, and deposits

Sum of Total Sell, sum of Total Withdrawals, sum of Total Deposits and sum of Total Buy for each Type. The view is filtered on sum of Total Buy, which keeps non-Null values only.

Since almost most of the users are of type 6, most amount of total buy, sell, deposit, and withdrawals are correspond to users of type 6.

What are the most used indication coupon?

Below figure shows the number of each indication coupons used by each type of users. It can be seen that 308 is the most popular indication coupon through all users.

Indication coupon of each type

Type

Indication Coupon for each Type. Size shows count of Indication Coupon.
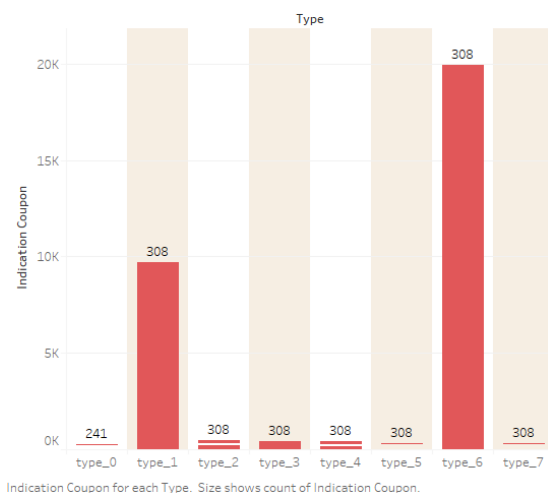
How is the trend of total sell and buy in previous months?

These two subsequent figures shows the total amount of sell and buy on each month, which are the same since selling and buying are each other's complement. It is so surprising that the general trend of total buy and sell in 2017 is ascending but in 2018 it fluctuates so much. The most amount of buy and sell on September which faces a great decrease in two next months. It would be so helpful to investigate the cause of great decrease in last month.

Total buy on each month, quarter, and year

The trend of sum of Total Buy for Transaction Date Month broken down by Transaction Date Year and Transaction Date Quarter. The view is filtered on Transaction Date Month, which excludes Null.
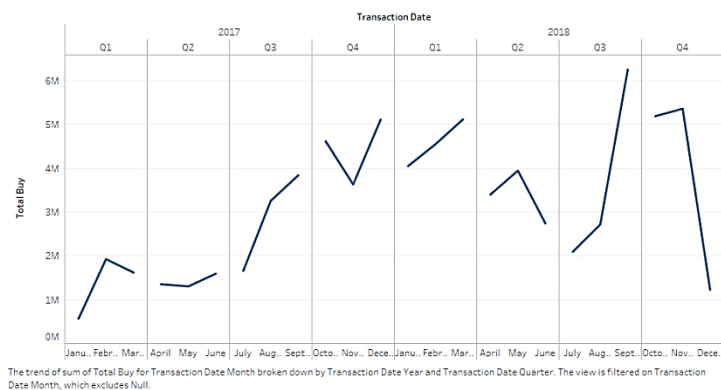
Total sell on each month, quarter, and year

The trend of sum of Total Sell for Transaction Date Month broken down by Transaction Date Year and Transaction Date Quarter. The view is filtered on Transaction Date Year, which keeps 2017 and 2018.
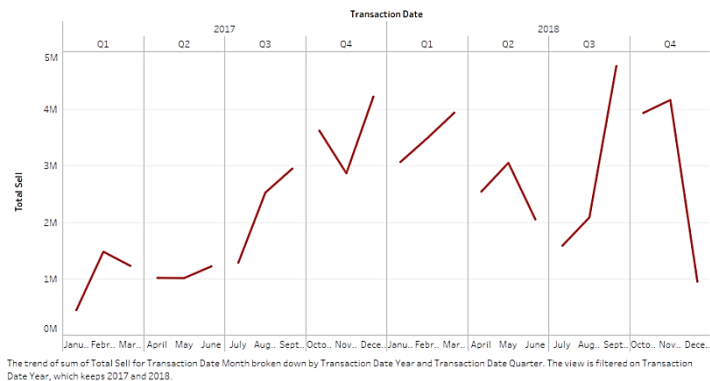
What about the trend of withdrawals?

The amount of withdrawals was very low at the first months which is rational due to the low amount of sell and buy on these months. By growing the amount of sell and buy, the amount of withdrawals increased in next months. It is not surprising that the amount of withdrawals decreased in 2018 since this year brings less money for the users.

Total withdrawl on each month, quarter, and year

The trend of sum of Total Withdrawals for Transaction Date Month broken down by Transaction Date Year and Transaction Date Quarter. The view is filtered on sum of Total Withdrawals, which keeps non-Null values only.
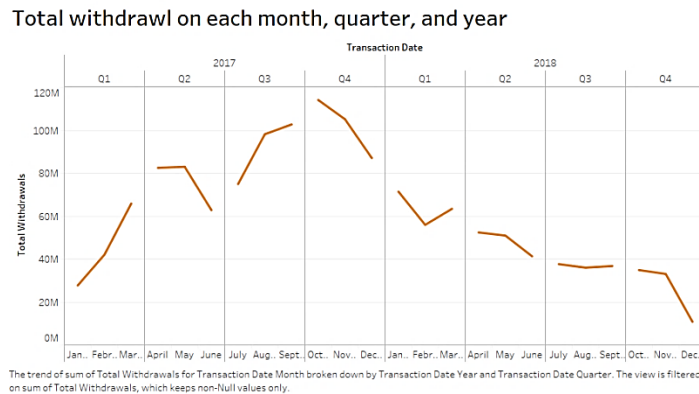
Does the number of transactions grow or fall?

It is interesting that is spite of the decrease in amount of total buy and sell, the number of transactions have an ascending trend. Putting all these figures into account, we can conclude that the number of transactions in 2018 increased but most of these transactions had low prices.

Number of Transactions on each day

The trend of sum of Number of Records for Transaction Date Day. The view is filtered on Transaction Date Day, which keeps non-Null values only.
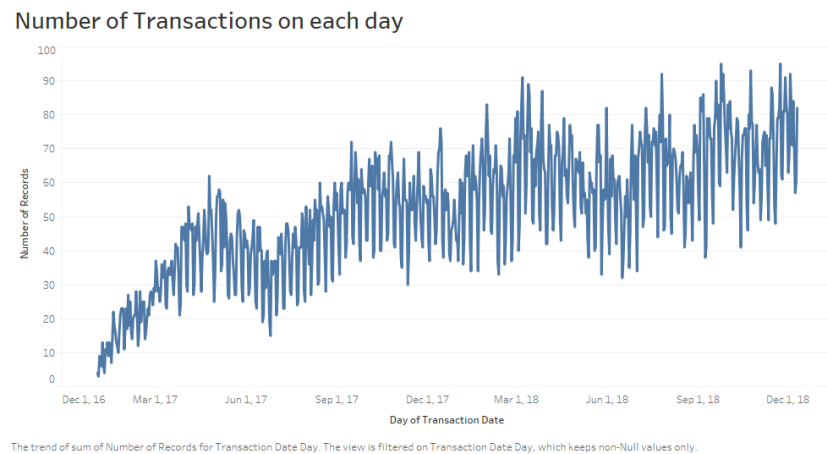
What is the indication coupon of the users with high amount of total sell, buy, withdrawal, and deposit?

The subsequent figure shows that most amount of total buy, sell, withdrawal, and deposit are done by the users with coupon indication 308. This is evident since the majority of users have indication coupon 308. This figure proves that account with id 308 is not fake and really invites so many users to the website and all its invitees really buy and sell in website.

Relation of indication coupons with total buy, withdrawl, sell, and deposits
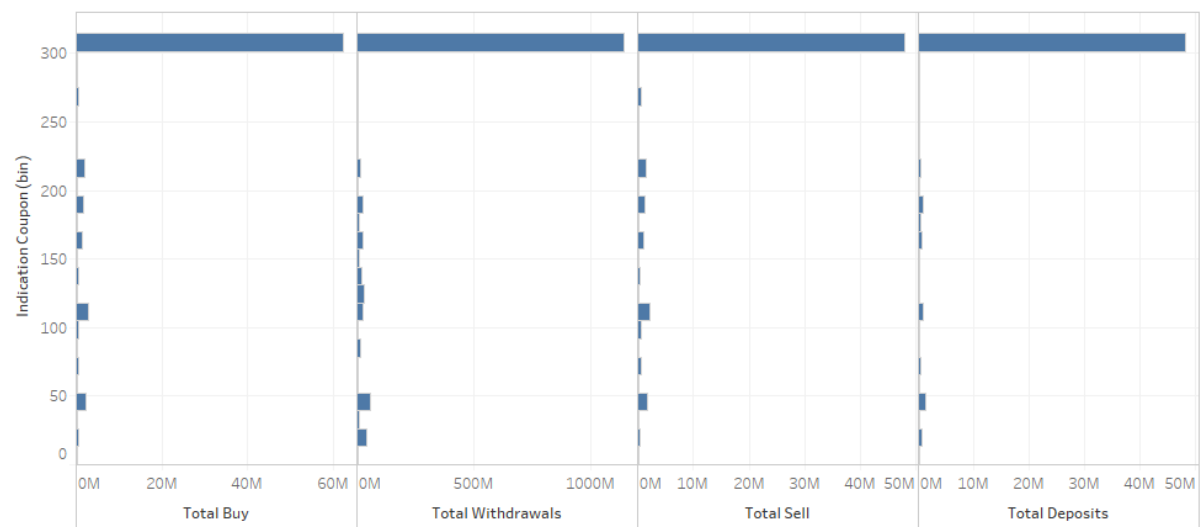

The trends of sum of Total Buy, sum of Total Withdrawals, sum of Total Sell and sum of Total Deposits for Indication Coupon (bin).

When was the first day of deposit of accounts?

The following figure shows the date of first deposit of users. It can be seen that by growing the number of joined account in second and third quarter of 2017, users did their deposits and by decreasing the number of joined accounts in 2018, fewer people did their first deposit in 2018. It is somehow interesting that in both years, the third quarter is the most popular quarter for doing the first deposit.

First deposit day of accounts
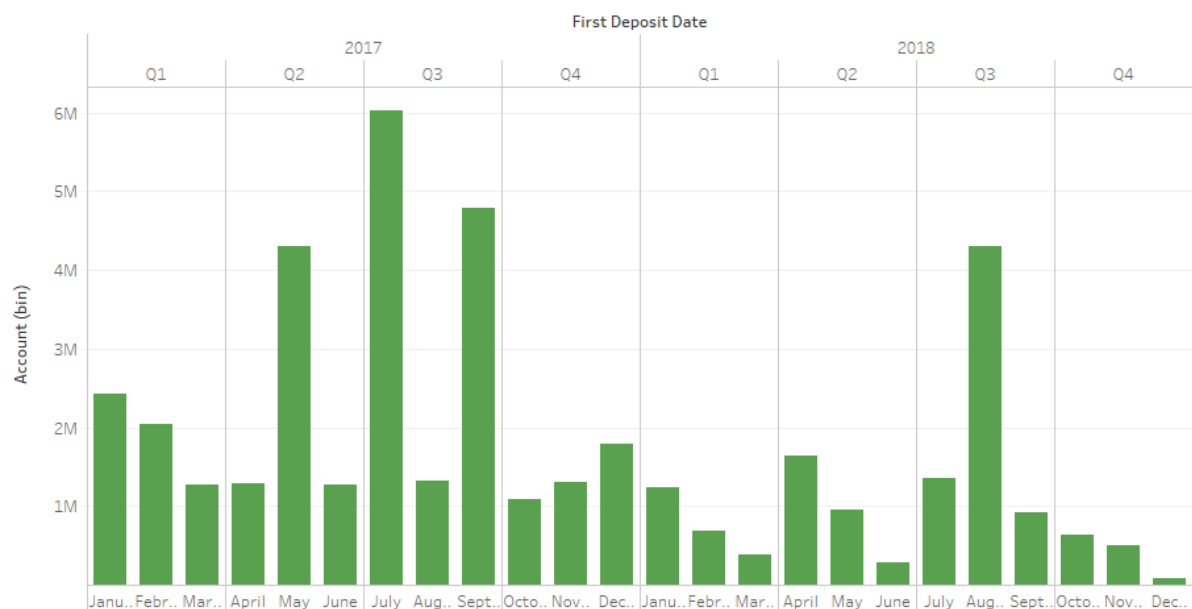
Account (bin) for each First Deposit Date Month broken down by First Deposit Date Year and First Deposit Date Quarter. Size shows sum of Total Buy. Details are shown for Residence. The view is filtered on First Deposit Date Month, which excludes Null.

Who bought the most?

The following figure is a brief analysis of users. It can be seen that the account with id 3479, had

Accounts transactions number
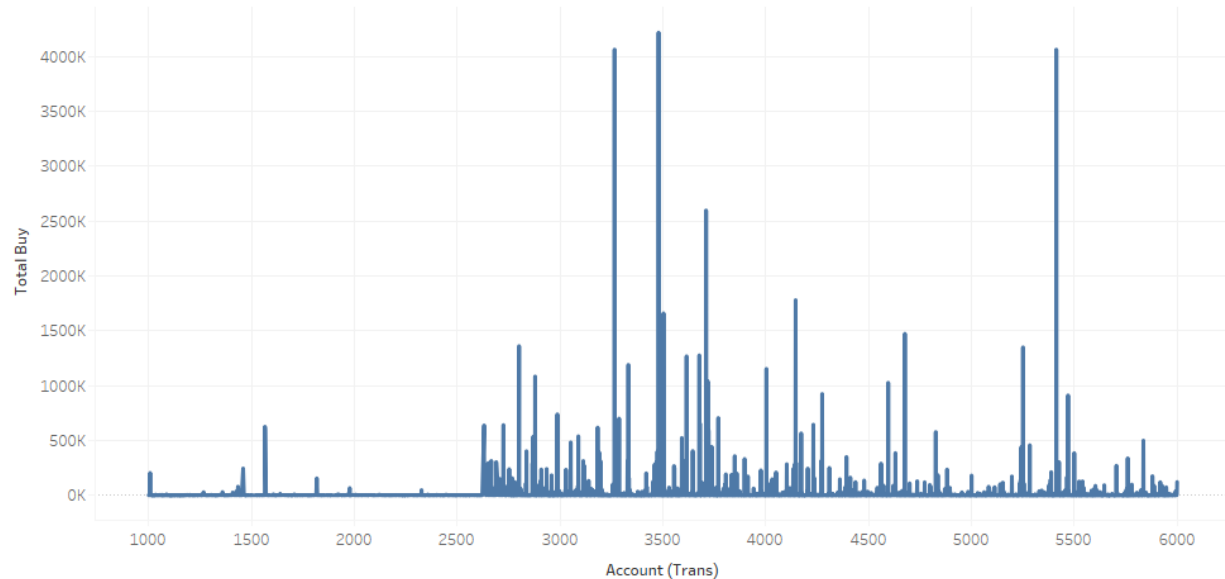
The trend of sum of Total Buy for Account (Trans). The view is filtered on Account (Trans), which keeps non-Null values only.

4,220,270 total buy. Two other accounts with id 3265 and 5412 had total buy 4,065,269 and 4,064,805 dollars. These three users contribute the most to the total amount of buy.

Tell me about the number of deposits, withdrawals and contracts in each month and year.

The following chart helps us to analyze the number of deposits, withdrawals and contracts in each month and year. Blue bars correspond to 2018 and yellow bars are related to 2017. By taking a glance at this chart, one concludes that 2018 is less fruitful, because less number of deposits, withdrawals and contracts has been done in 2018 in comparison to 2017. The third and fourth quarters of both years are the best quarters. The last blue bar of number of contract in December 2018 shows that the number of contracts grows exponentially in this month but most of these contracts had low prices which are in accordance with the above figures.

Number of contracts, deposits, and withdrawls on each month and quarter
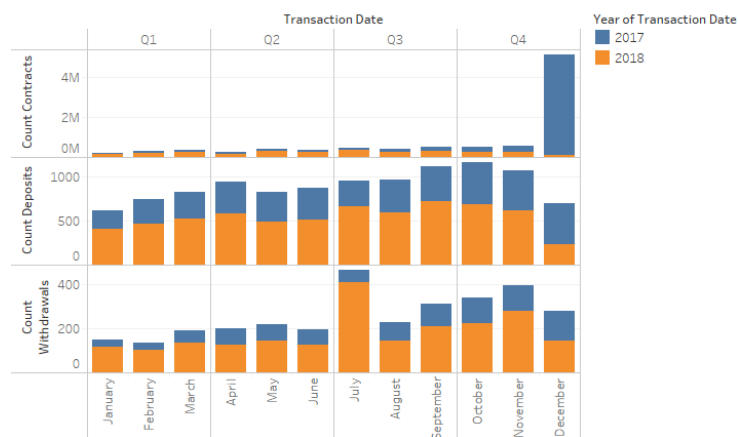
Sum of Count Contracts, sum of Count Deposits and sum of Count Withdrawals for each Transaction Date Month broken down by Transaction Date Quarter. Color shows details about Transaction Date Year. The view is filtered on sum of Count Contracts, which keeps non-Null values only.

On which weekday users had highest amount of buy or sell?

As it can be seen,Sundays and Mondays are the most popular days and the best days of website, since more amount of buy and sell are earned or spent on these two days. Other weekdays are in the second stage and the same in regard to the mean of total sell and total buy.
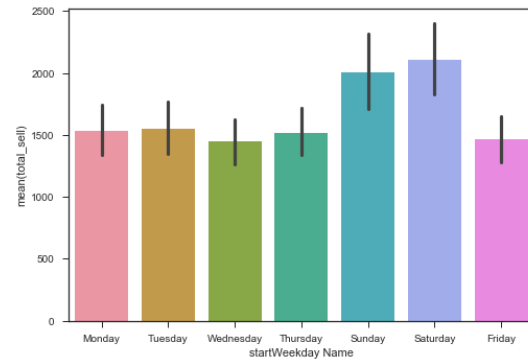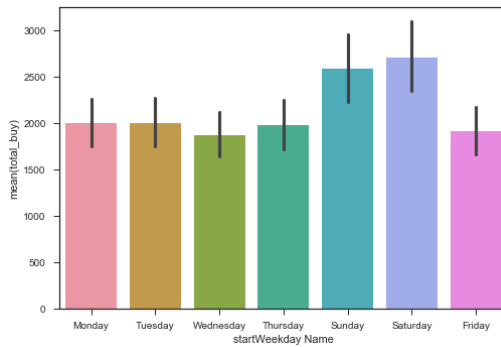
Analyzing campaign data

Where were the campaigns hold?

Below figure help to understand which countries spend more money in campaigns. As it can be seen, Indonesia with about 185 thousand dollars has the most amount of spend in campaigns. This is obvious due to the number of users in this country. Brazil and Russia with more than 40 thousand dollars are the next two countries with high spend monies in campaigns.

Total spend on each country

Map based on Longitude (generated) and Latitude (generated). Size shows sum of Total Spend. Details are shown for Country. The view is filtered on Latitude (generated), Longitude (generated) and Country. The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only. The Country filter keeps 19 of 19 members.

What is the relation of date and total spend?

The following figure investigates the amount of total spend on each year, quarter and month. Here issno special trend and the amount fluctuates highly. The third quarter of 2017 had fruitful campaigns with high amount of spent monies. But the general trend is that the total amount of spend in 2018 is less than 2017. The best month is September 2017 and the worst month is February 2017

Total spend on each year, quarter, month
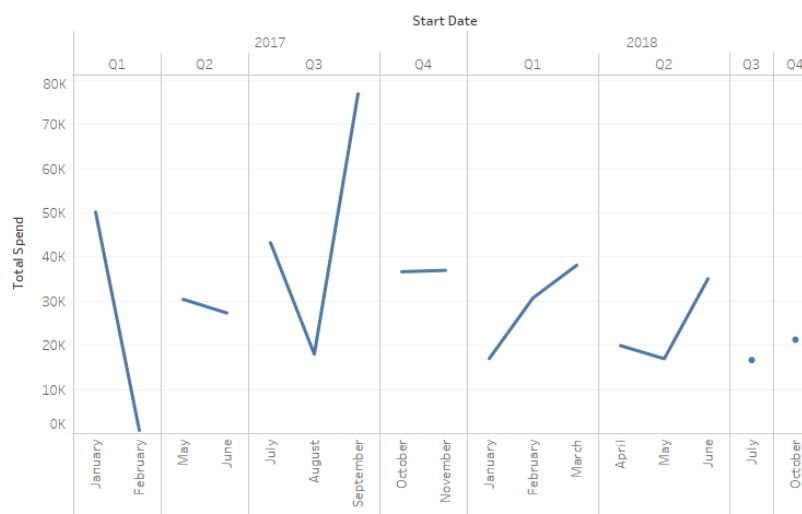
The trend of sum of Total Spend for Start Date Month broken down by Start Date Year and Start Date Quarter.

Which month is the best to start a campaign?

The boxplot of distribution of total spend in each month is presented below. It shows that The highest value of total spend is about campaigns in July; however, this month is so risky to start a campaign, because there were campaigns in July with very low total spend. Campaigns in January, December, March, and May had stable and nearly high amount of total spend. Furthermore, selecting August as the month for starting campaign is not suggested due to the low values of total spend, Nor July is suggested due to its high risk and variability.
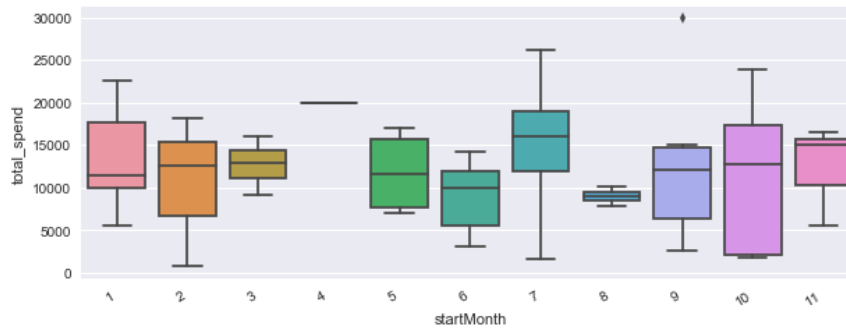
What weekday was the start days of campaigns?

The most frequent day for starting campaigns was Sunday. Other days had equal rate of starting campaign.
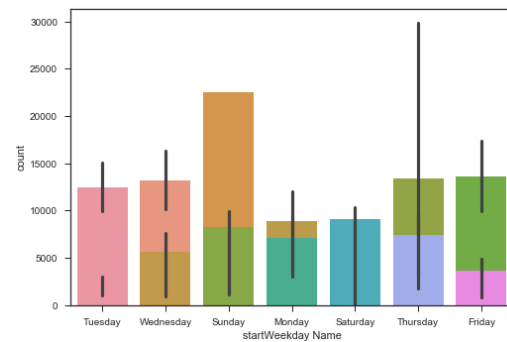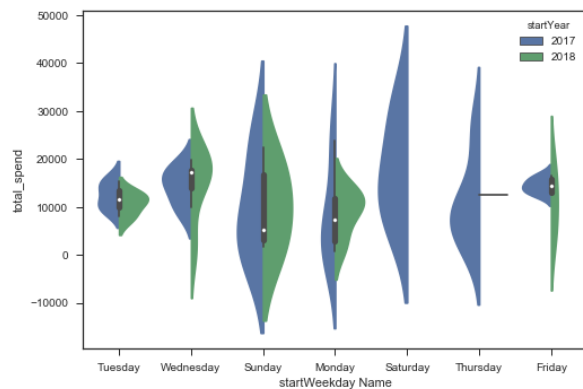

Which weekday is the best to start a campaign?

It is so useful to know the relation between the weekdays of a campaign and its total spend. As it can be seen, The campaigns started on Sundays, had the most mean of total spend. This can help the company for making decision about the best weekday for starting a campaign. Sunday is the best day due to its high value of total spend, Saturday is in the second place.
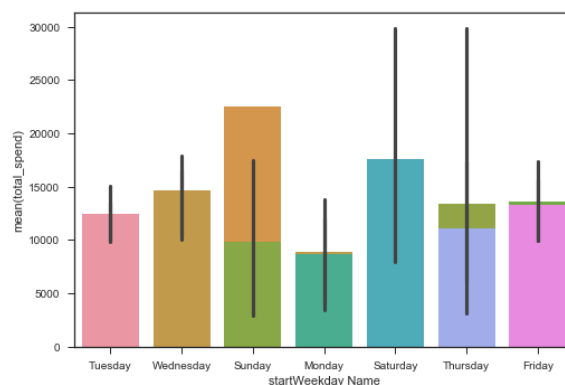
Does the company change its policy for choosing the day of starting campaign in the new year?

As you can see in this figure, none of campaigns in 2018 started on Saturday or Thursday; however, these two days were so casual in 2017. Conversely, the rate of campaigns starting on Fridays increased in 2018. The rate of campaigns starting on other weekdays, remains steadily.
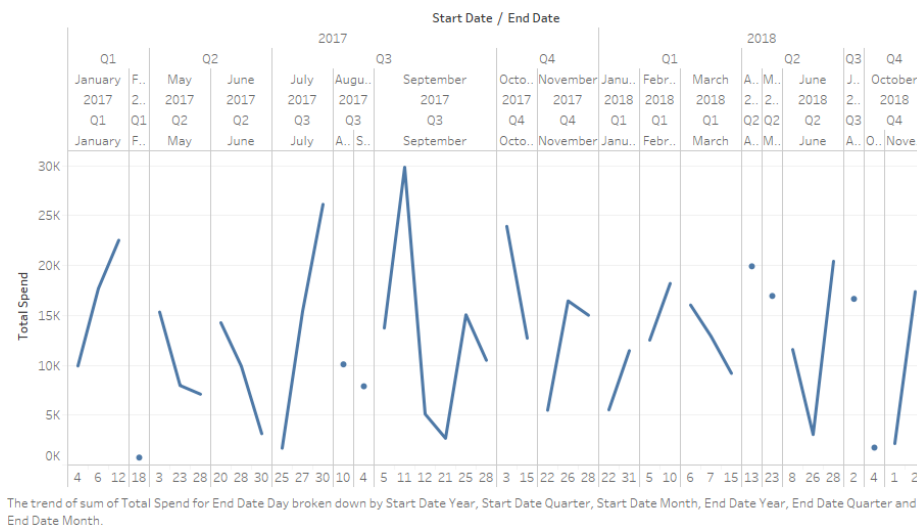

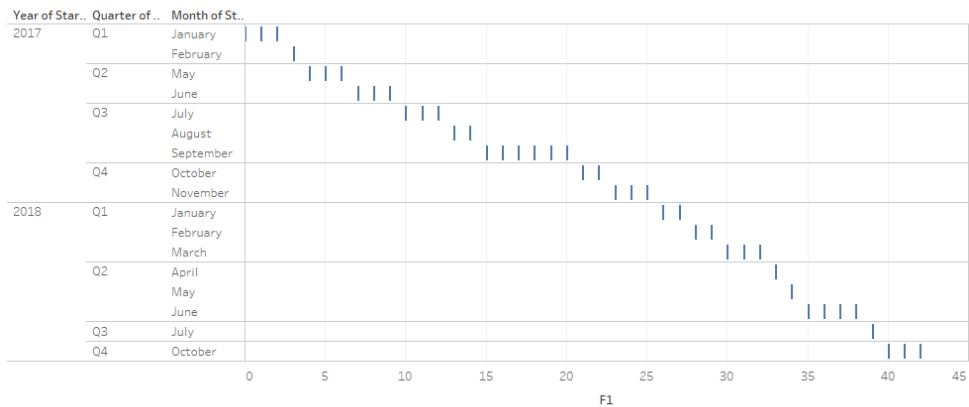The number of campaigns remains the same through these two years.
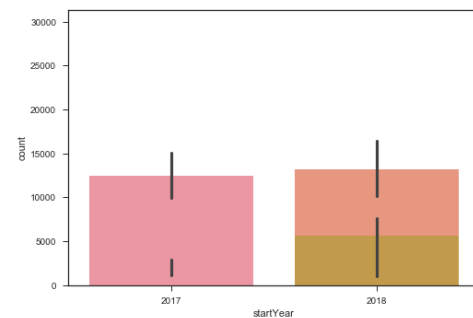
Which day of month is the best to start a campaign?

The subsequent chart shows that there is no obvious relation between the start day of campaign and the amount of spent monies. Some of high values of total spend are related to the first days of months, some are related to the last days of months and there is no general trend. However, the third quarter of year is very fruitful.

Total spends on each start date

The trend of sum of Total Spend for End Date Day broken down by Start Date Year, Start Date Quarter, Start Date Month, End Date Year, End Date Quarter and End Date Month.
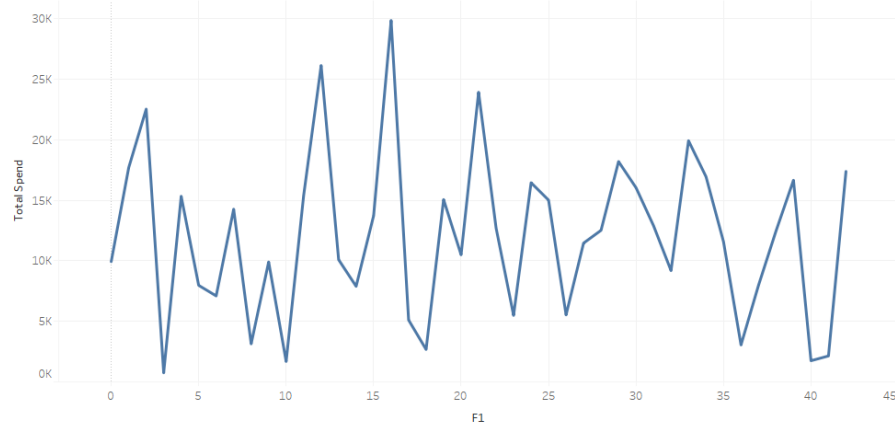
date of each campaigns

F1 for each Start Date Month broken down by Start Date Year and Start Date Quarter.

Above figure is another representation of start days for each of 42 campaigns.

How much monies are spent on each campaign?

Below figure shows the quantity of monies spent for each campaign. There is no general trend, since the amount of monies spent on a campaign depends on many factors such as its start day and month, duration, its coincidence with holidays, its country and the number of users in that country, and etc. It can be seen that the sixteenth campaign is the most fruitful campaign, which 29,866 dollars was spent this campaign.
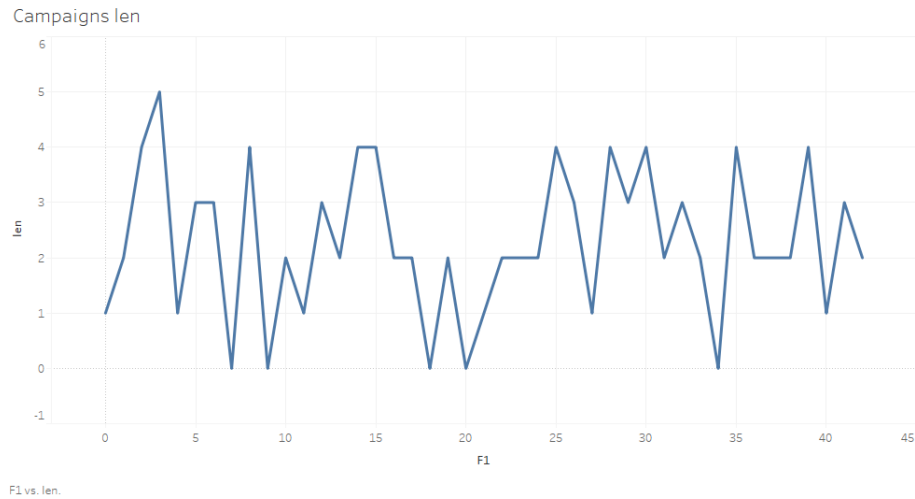

Campaigns total spends

The trend of sum of Total Spend for F1.

How long did the campaigns take long?

Below is the duration length of each campaign. All campaigns last between some hours to 5 days. There was just one campaign that last 5 days. Most of campaigns were four or two days.
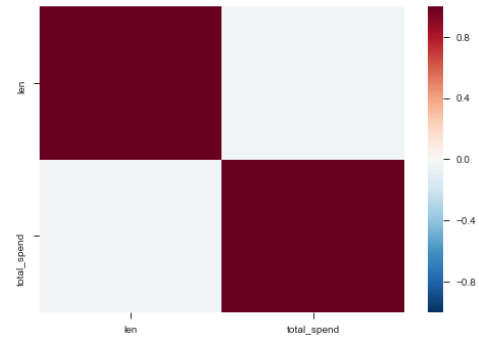
Campaigns len

F1 vs. len.

What is the relation of campaign duration and total spend?

One of the useful analysis to be done is to find out the relation between the time duration of a campaign and total spent monies on the campaigns. Before looking at the following figure, one may guess that the longer the campaign is, the more is the total spent monies. However, the following figure has a surprising reality in it. The best campaigns are those that last 2 days and not more. You can see that the campaigns which last 5 days, yield the least monies even less than campaign that are just for some hours!! It may be due the fact that when users know that the current campaign will last till several next days, they had less enthusiasm to buy immediately and postpone their shopping to the future days and gradually they lose their enthusiasm for the shopping in next days or may forget their decision.
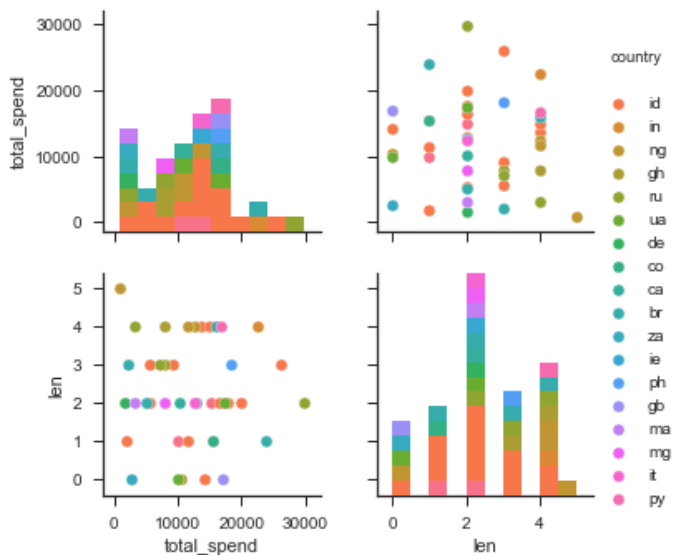

The relation of total spend with duration of campaigns

The trend of sum of Total Spend for len.
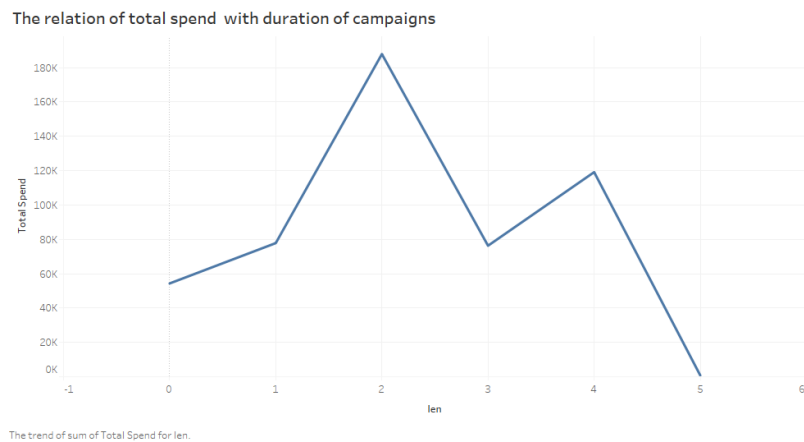
Is there any correlation between total spend and campaign duration?

This figure shows that there is no correlation between them and their correlation is about zero.

A total analysis of campaigns:

This pair-plot contains four plots:

1. Top left plot shows the histogram of total spend. It can be seen that almost most of campaigns have total spend less than 20000 dollars; however, spending 20000 dollars on each campaign is very casual. It also shows that most of campaigns were in Indonesia; although there is no general trend about the total spend in Indonesian campaigns. Furthermore, there is no general trend about the total spend in campaigns of other countries. It can be concluded that there is no relation between the total spend amount and the country of campaign.

2. The top right plot shows the relation between campaign duration and total spend alongside the country. For example, the campaigns in Indonesia had various duration and total spend, but for this country, 2-day and 3-day campaigns were better. In most countries 2-days campaigns are more frequent but with diverse total spend. The highest total spend is a 2-day campaign held in Russia. And the lowest total spend is for a 5-day campaign held in India; however, 4-day campaigns in this country yield more monies.

3. The bottom left plot is same as rotated plot of top right.
4. The bottom right plot shows the histogram of campaign durations alongside their countries. Most of countries had campaigns with various duration. But in Madagascar, Morocco, and Italy all the campaigns lasted 2 days.