

**Take Home Assignment**

**Context**

The assignment required us to analyze customers trading activity over the period Jan-2020 to March-2020. For this, 4 files were given namely: ledger\_entries.csv, accounts.csv, rates.csv and trades.csv.

Using this were required to link each **trade** to its user, timestamp, market pair and USD volume for further analysis and to draw insights on customer trading behavior.

**Summary of Business Question Answers**

1. **Customer Status Distribution**

My Analysis showed that the number of UNIQUE clients trading for each of the months Jan, Feb and March on exchange / broker were 15, 17 and 18 respectively. The 15 clients for Jan were all considered to be returning clients (this was the first month’s data we had and so could not determine if they belonged to any of the other segments. Using this as our starting point, our analysis revealed the followed client status distribution for each of the months considered:

**Jan-2020 Clients (15 Active Customers)**

* 15 Returning

**Feb 2020 Clients (17 Active Customers)**

* 13 returning
* 2 Churned
* 4 New

**March 2020 Clients (18 Active Customers)**

* 14 returning
* 3 Churned
* 2 New
* 2 Reactivated

1. **Churn Rate for February and March**

* **Feb Churn Rate** was calculated as **11.76%** (2 churned / 17 active)
* **March Churn Rate** was calculated as **16.67%** (3 churned / 18 active)

1. **Customer Transacting Segment Trends**

Our analysis revealed that for new customers the preferred market-pair was XBT/MYR. As will be shown below, just over $11k of this pair were bought by new clients over Feb and March. This was somewhat surprising given that the XBT/ZAR market-pair was by far the most traded currency (over $1,9m traded) over the three months investigated. It was still the most popular pair for returning clients but not for new clients who preferred XBT/MYR – this would suggest that we had more new clients from Malaysia than South Africa.

1. **Average Trade Volume by Transacting Segment**

Our findings show that for returning customers only, the average volume traded for each of the months were $926, $242 and $530 for Jan, Feb and March respectively. On average about 70% of returning customers traded below this average each month – indicating that there were some customers (about 30%) that traded significantly higher than the mean monthly average.

We also note the drop in average USD volume traded for returning customers in Feb ($242) compared to Jan ($926) and March ($530). I would say that once reason for this could be the price of XBT/ZAR in Feb. Bitcoin rallied hard in Jan, which would have contributed to the higher volume seen for this month. However, in February the Bitcoin Price seem to stagnate and sometime during that month began to decline with the trend continuing down in March.

Returning customers would have done most of their purchasing in Jan during the bull run and would appear to have stood back in Feb particularly, they re-entered again in March when prices were at more attractive levels to buy.

The rest of the report briefly describes the methodology used to produce (join/merge) our files into dataframes for further analysis. Using this we then provide further insights int customer trading activities that may be useful and noteworthy.

**Methodology**

The first step was to prepare the data. The exact procedure followed is documented in the code but on a high level (after importing all received files as pandas dataframes) the following steps were taken:

* Merged the ledger (on ‘account\_id’) and accounts (on ‘id’) files.
* Merged resulting dataframe (on ‘foreign\_id’) with trades (on ‘id) file.
* From the resulting dataframe, a new “hourly” column was created from the ‘timestamp\_at’ column.
* Thereafter for the rates file, I calculated the hourly average price per currency.
* Then using this I merged the rates file (on hourly ‘reference\_at\_date’ AND ‘currency’) with the previous dataframe (on ‘hourly’ AND ‘currency’)

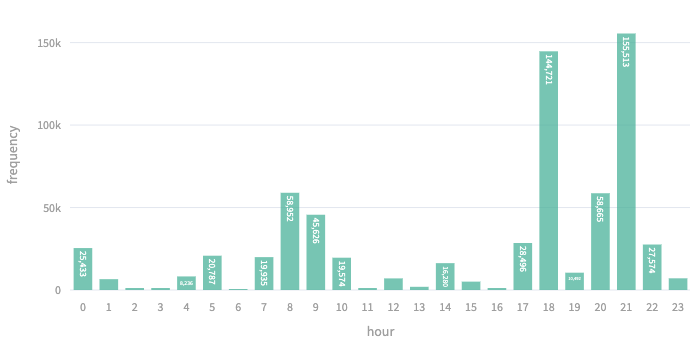
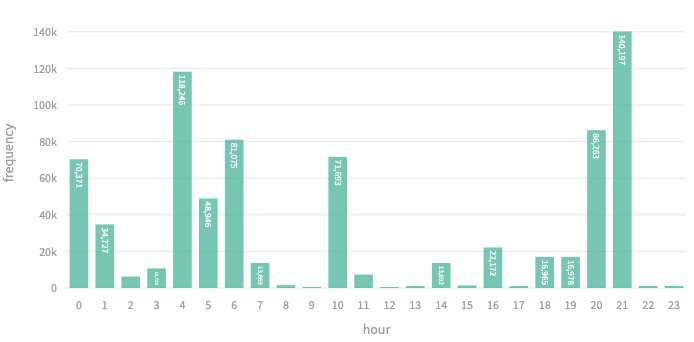
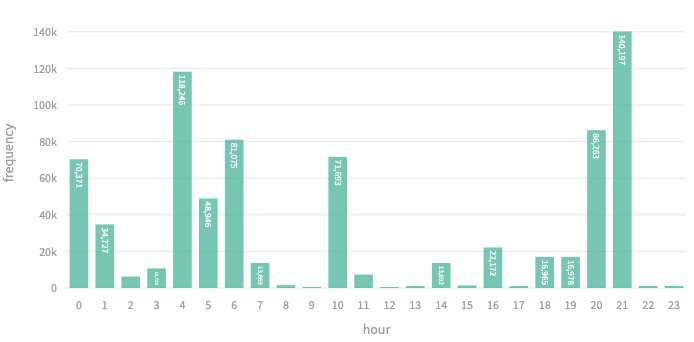
The above outlines the high-level steps taken to merge all dataframes together – again exact details are given in the code.

**Key Insights & Analysis**

**Hourly Distribution**

As an exploratory analysis I first looked at the number of trades coming through the exchange / broker each hour for the three months considered. Although I couldn’t find any clear patterns for the hourly number of trades (i.e. hourly trade count) , I did notice that when considering the total USD Volume (so the equivalent USD amount that was traded) – I noticed that there high volumes coming through towards the late part of the day (from 18h00 onwards) – perhaps this could be from individuals trading bigger volumes after work when they have more available time and had time to process the days news.

**Jan-2020 Feb-2020 March-2020**

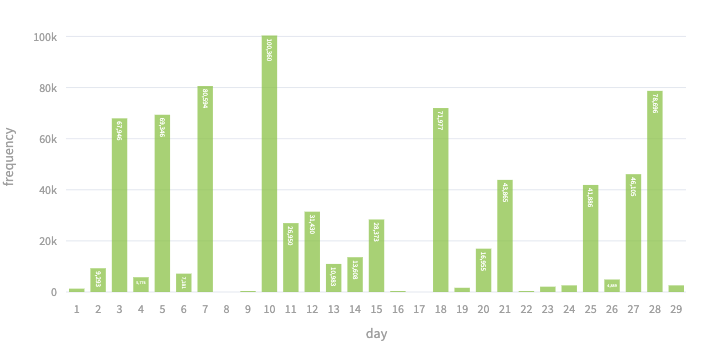
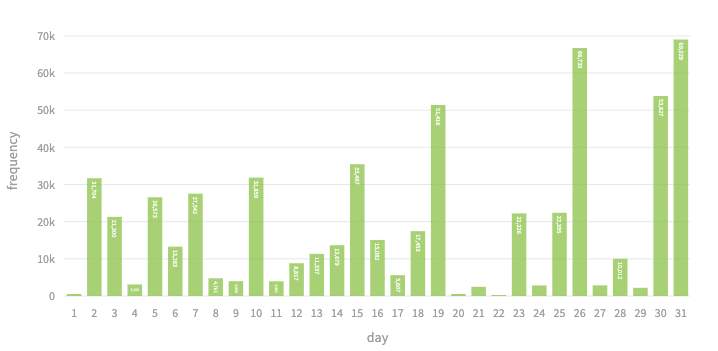
 

**Daily Distribution**

I did the same exploratory analysis by considering the daily trades distributed over the three months. Again the number of trades (daily trade count) didn’t reveal anything that stood out but when looking at the total USD Volume, it can be observed that there are some bigger volumes coming through towards the later part (after day 20) of the months. This makes some sense if you consider that people get paid from from around this time and may have extra cash available to trade on the exchange/broker.

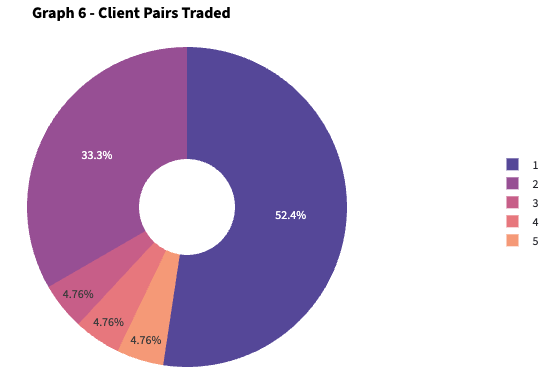
**Jan-2020 Feb-2020 March-2020**

A graph of green bars

AI-generated content may be incorrect.  

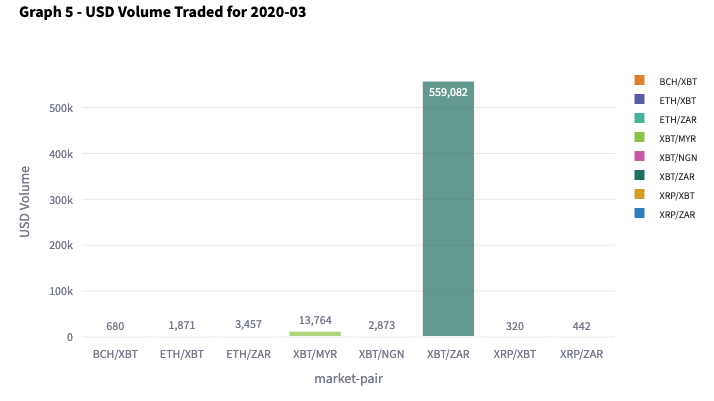
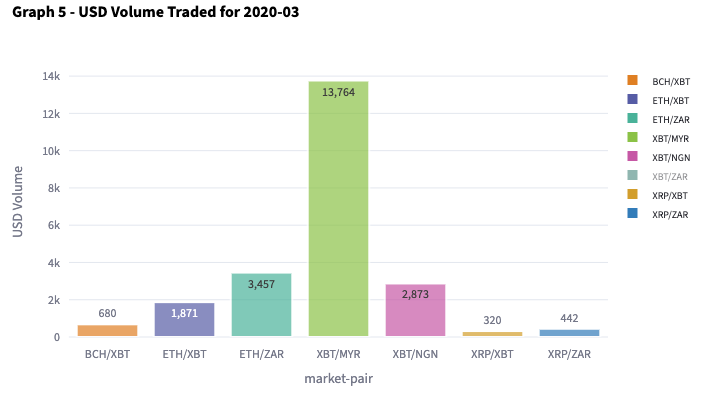
**Market-Pair USD Volume Traded**

Next, I looked at the distribution or split for the number of market-pairs traded by clients on exchange / broker. By this I mean that from the identified clients, how many have only traded in one pair, how many traded two pairs and so on… The results from my merged dataframe showed that there were 21 unique customers and from this 11 (52.4%) traded only one market-pair with a further 7 Clients (33.3%) trading 2 currency pairs. The highest we had was 1 client (4.76%) who traded a notable 5 pairs over the 3 months investigated.

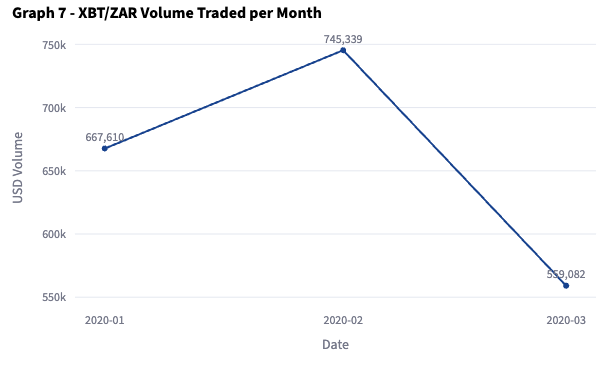
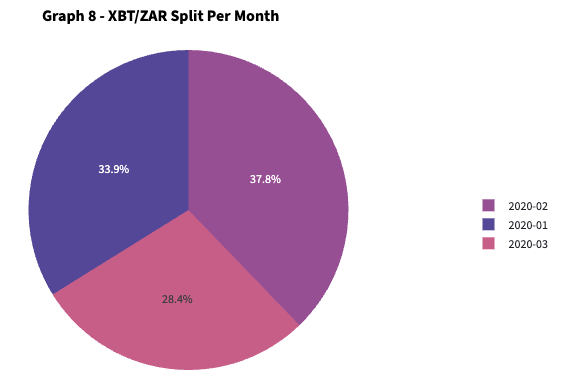
****

**Market-Pair USD Volume Traded**

By far the highest market-pair traded on exchange / broker was the BTC/ZAR pair. Graph 5 (Left) shows the USD volume of XBT/ZAR traded ($559,082) in March which was significantly higher than the second highest pair traded which was the XBT/MYR pair ($13,764). Graph 5 (Right) shows the same USD Volume by each currency pair in March but with XBT/ZAR removed.

Although the XBT/ZAR volume was considerably higher than the other pairs – the March volume was somewhat down (Graph 7) when compared to the Jan ($667,610) and Feb ($745,339) volumes. Graph 8 shows the monthly XBT/ZAR percentage split (XBT/ZAR volume for that month divided by total XBT/ZAR volume traded over the three months) and shows that the split wsa relatively even over the three months considered.

**** ****

**Client Status Distribution**

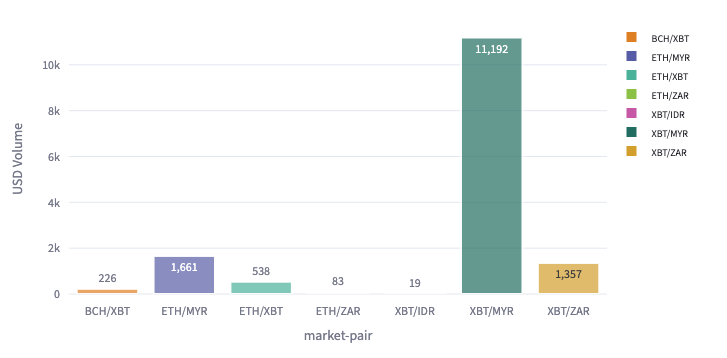
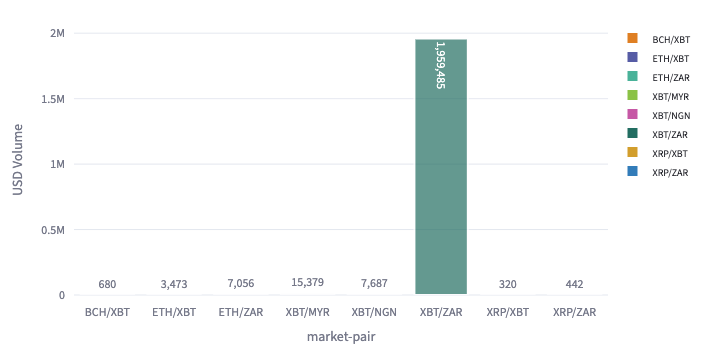
When considering the market-pairs traed by the different client segments (returning, new, reactivated and churned), I found that from the 21 unique clients identified, 2 clients were marked as churned in Feb and a further 3 in March - there were however 2 clients that was marked as reactived in March, meaning that they traded in Jan but not in Feb and re-entered the market again in March.

I made the assumption that all clients trading in Jan-2020 were returning (since this was the earliest month’s data, we could not determine if clients were in any of the other segments and an assumtion ws made that they were all returning. In general we saw positive upward trend in activity with the number of active customers increasing to 15, 17 and 18 for Jan, Feb and Mar respectively.

We had a total of 4 new clients in Feb and a further 2 in March which further supports our positive trend assessment.

Looking at the graphs below, we note that the currency of choice for new customers appears to be XBT/MYR ($11,192 traded) with XBT/ZAR still being the established and prefered currency for returning clients ($1,959,485 traded) over the 3 months.

**New Returning**

**Return Client Monthly Average Distribution**

Lastly, I looked at distribution of monthly average usd volume trade by returning customers each month and compared this to the monthly returning customers mean volume traded. I noted that the monthly usd volume dropped from an average of $926 in Jan to $242 in Feb before recovering somewhat again in March to an average of $530.

One contribution to this could be the XBT/ZAR price as noted earlier, this was the most traded pair for returning customers and when looking the price of Bitcoin (regretably I should have plot the price from the rates file but was running out to submit the assignment so did not do this, but the price is something that should definiitely be looked at) – the price rose substantially in January and peaked sometime in Feb before falling back down again. Returning customers could have been holding on to their stock which would have seen notable unrealised profits which could explain the lower average volume traded in Feb.

As the market price for XBT/ZAR fell off again, participants may have been more will to enter trades again which would have contributed to the higher usd volume in March.

It is also noted that the number of returning customers in each month that traded below the “returning client” mean hovered around 70% meaning that were outliers with some customers trading at notably higher levels – as seen from the graphs below.

**Jan-2020 Feb-2020 March-2020**