BI SPECIALIST ROLE TEST

Pugacz Natasha Catalina

1 Introduction:

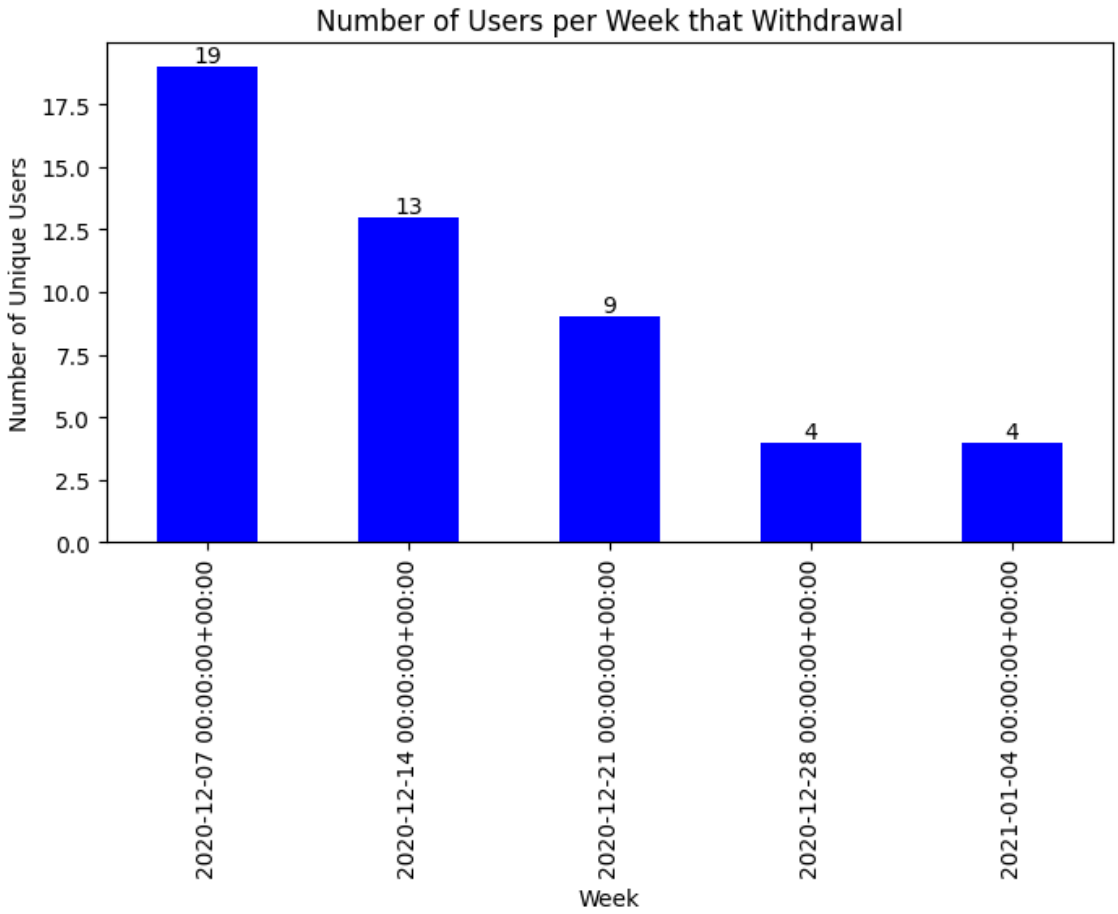
I'm sharing this document that includes the responses I provided for the challenge presented during the application process for the BI Specialist role at Bitso. The challenge was an opportunity for me to showcase my skills and abilities in the field of business intelligence. In the document, I'll be explaining the methods and strategies I used, along with the Python programming, to meet the challenge's requirements.

2 Questions

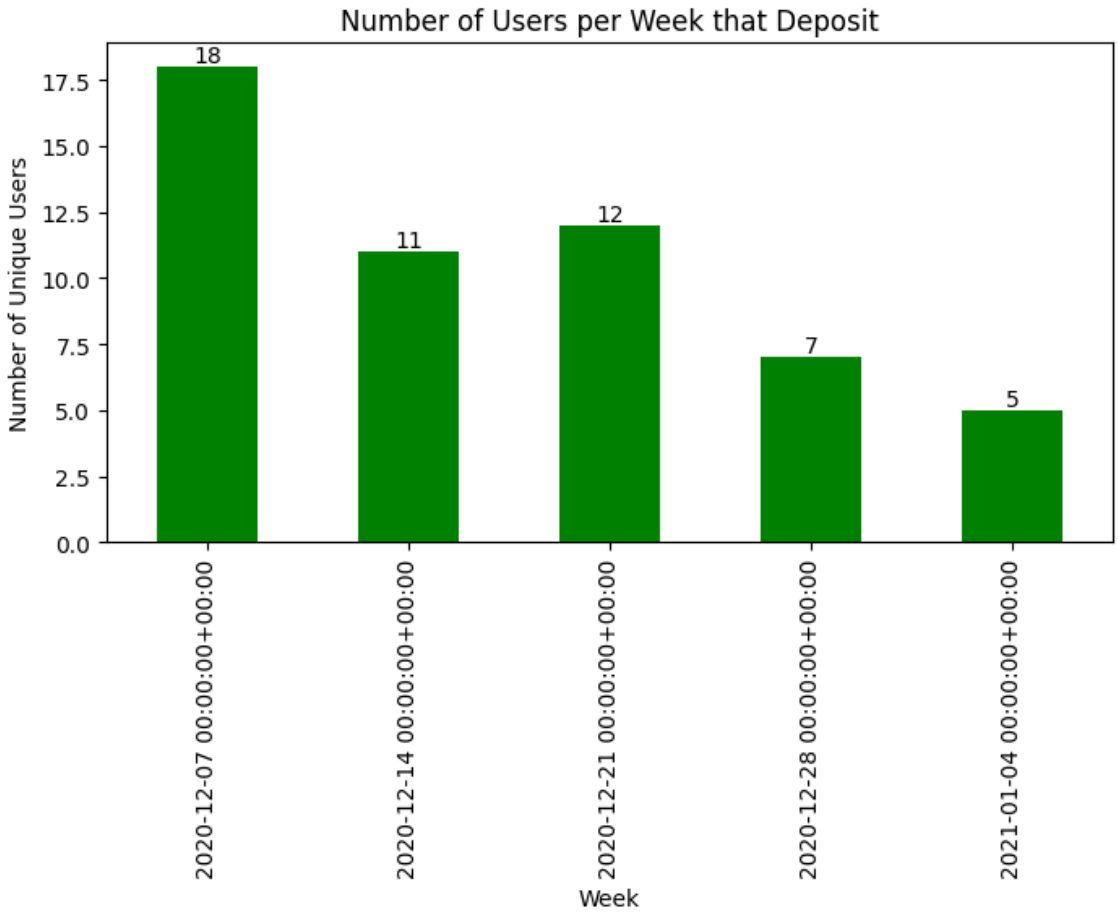
2.1 Deposits & withdrawals matters.

2.1.1 Get the number of users that deposit and withdraw on a weekly basis.

The number of users engaging in weekly deposits and withdrawals is depicted in the following graphs. In both instances, a discernible trend is evident—user numbers progressively decline as time unfolds. This phenomenon could potentially align with the year-end period and the associated celebrations. Notably, considering that the bulk of our user's hail from Mexico, a nation renowned for its diverse festivities, this decline might correlate with various celebratory occasions.



catalinapugacz@gmail.com



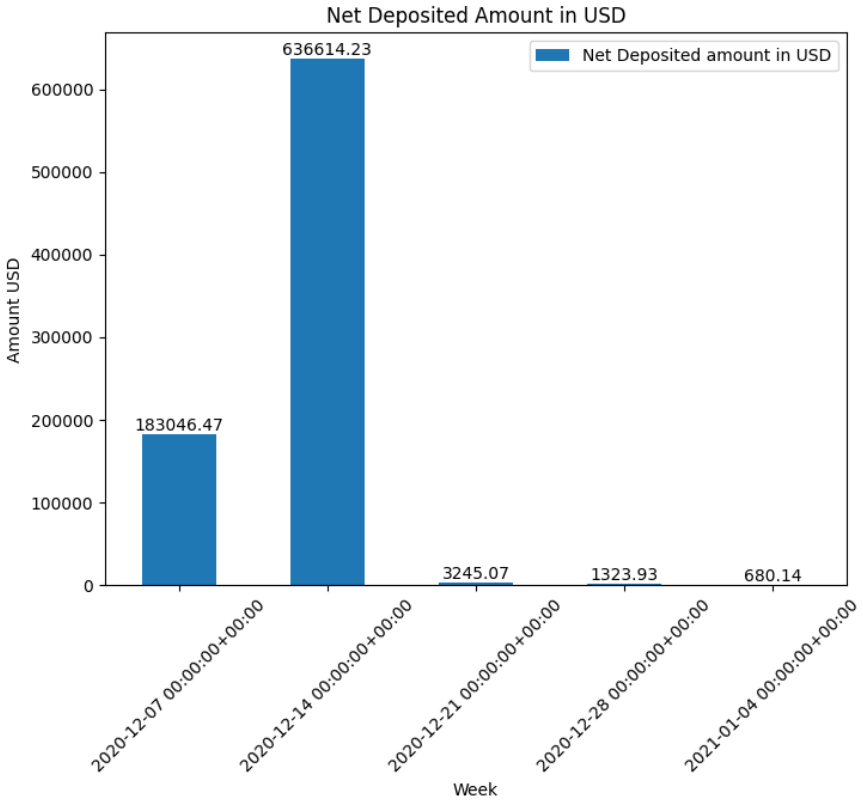
2.1.2 Get the weekly Net Deposited amount in USD.

There are two ways to consider conversions to USD. In this report, we'll illustrate both approaches.

The first approach involves considering that all cryptocurrency funds are held without conversion until the moment they need to be valued. At that point, the conversion rate for that day would be used. This is the approach taken for this particular chart.

The second option, which will be used for "Revenue" later in this document, takes into account that various taxes/fees generated in the trade's currency are immediately converted to USD. To do this, we seek the closest exchange rate to the trade's date.

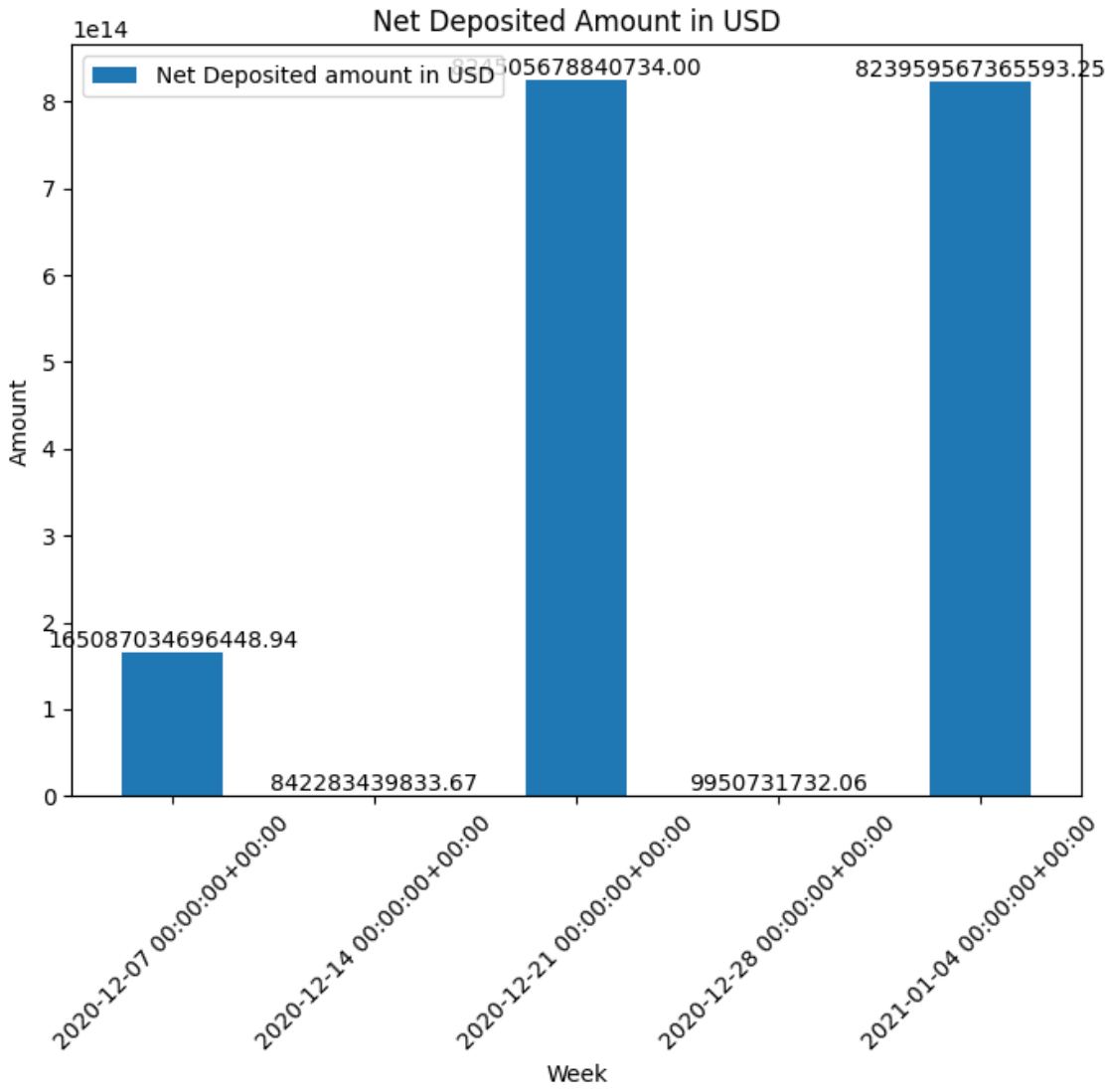
For this analysis, all fiat and cryptocurrency values were converted to USD using the provided price data, taking into consideration the most recent conversion rates within the dataset. The outcomes reveal that the second week records the highest amount of deposited funds.



2.1.3 Divide the number of deposits and withdrawals by fiat and crypto.

For this purpose, all fiat and cryptocurrency values were converted into USD using the provided price data, considering the most up-to-date conversion factors from the dataset. The outcomes demonstrate significant deposit values. In order to discern trends for the given week, a comprehensive analysis should juxtapose data from previous years, concurrent economic events, as well as recent marketing and promotional efforts undertaken by the app that might have incentivized user engagement.

I perceive a promising opportunity to implement a machine learning algorithm for forecasting deposits.



2.1.4 Divide the number of deposits and withdrawals by fiat and crypto.

To achieve this, all deposits and withdrawals were aggregated, and subsequently filtered by fiat and cryptocurrency. Transactions conducted in Mexican pesos far outweighed those in cryptocurrency. Given that the wallet's country of origin is Mexico, this pattern aligns logically. The chosen approach hinges on the strategic objective, whether it is to sustain transactions in local currencies or harness cryptocurrencies for the company's benefit.

2.1.4.1 Fiat

```
Table of row counts per currency in ds_withdrawal:
  Currency  Withdrawal Count
0      ars                2
1      mxn             1411
2      usd               42

========================================

Table of row counts per currency in ds_deposit:
  Currency  Deposit Count
0      ars            18
1      brl            15
2      mxn          1216
3      usd            19
```

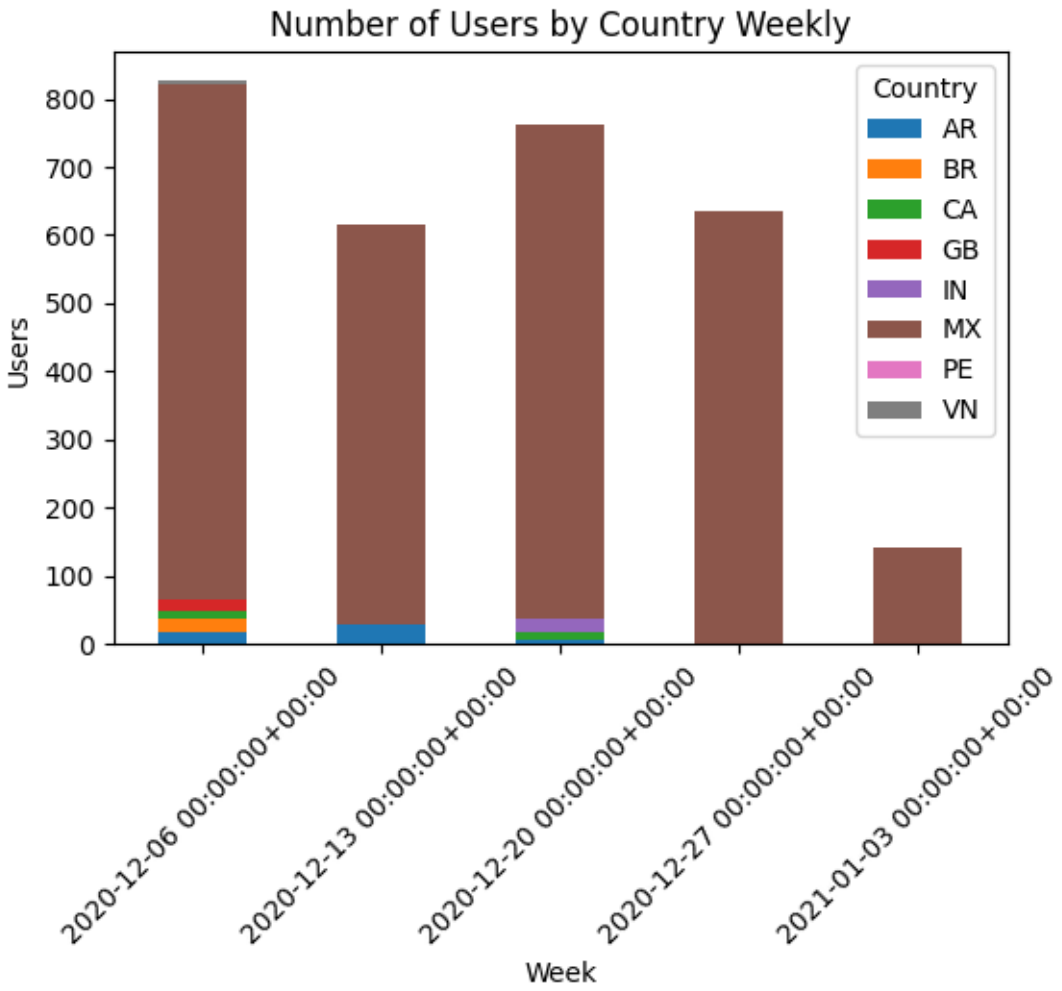
2.1.4.2 Crypto

```
Table of row counts per currency excluding 'mxn', 'ars', 'usd', and 'brl' in ds_withdrawal:
  Currency  Withdrawal Count
0      bat                5
1      bch                3
2      btc               21
3      dai                3
4      eth               16
5      gnt                1
6      ltc                1
7     mana                1
8     tusd                8
9      xrp                5

====================================

Table of row counts per currency excluding 'mxn', 'ars', 'usd', and 'brl' in ds_deposit:
   Currency  Deposit Count
0       bat            17
1       bch            13
2       bsv            12
3       btc            28
4       dai            17
5       etc            12
6       eth            16
7       gnt            13
8       ltc            12
9      mana            13
10     rbtc            11
11     tusd            17
12      xrp            13
```

2.1.5 What country has more active users on a weekly basis?

This question posed quite a challenge, as it required the maintenance of each user independently regardless of the number of transactions, while avoiding repetition. Moreover, it necessitated the amalgamation of deposits and withdrawals, all while integrating the user database.

The outcomes unveil a substantial and consistent positive trend attributed to Mexico, which persists prominently in the subsequent weeks. As previously mentioned, this phenomenon is rationalized by the wallet's origin in that region.

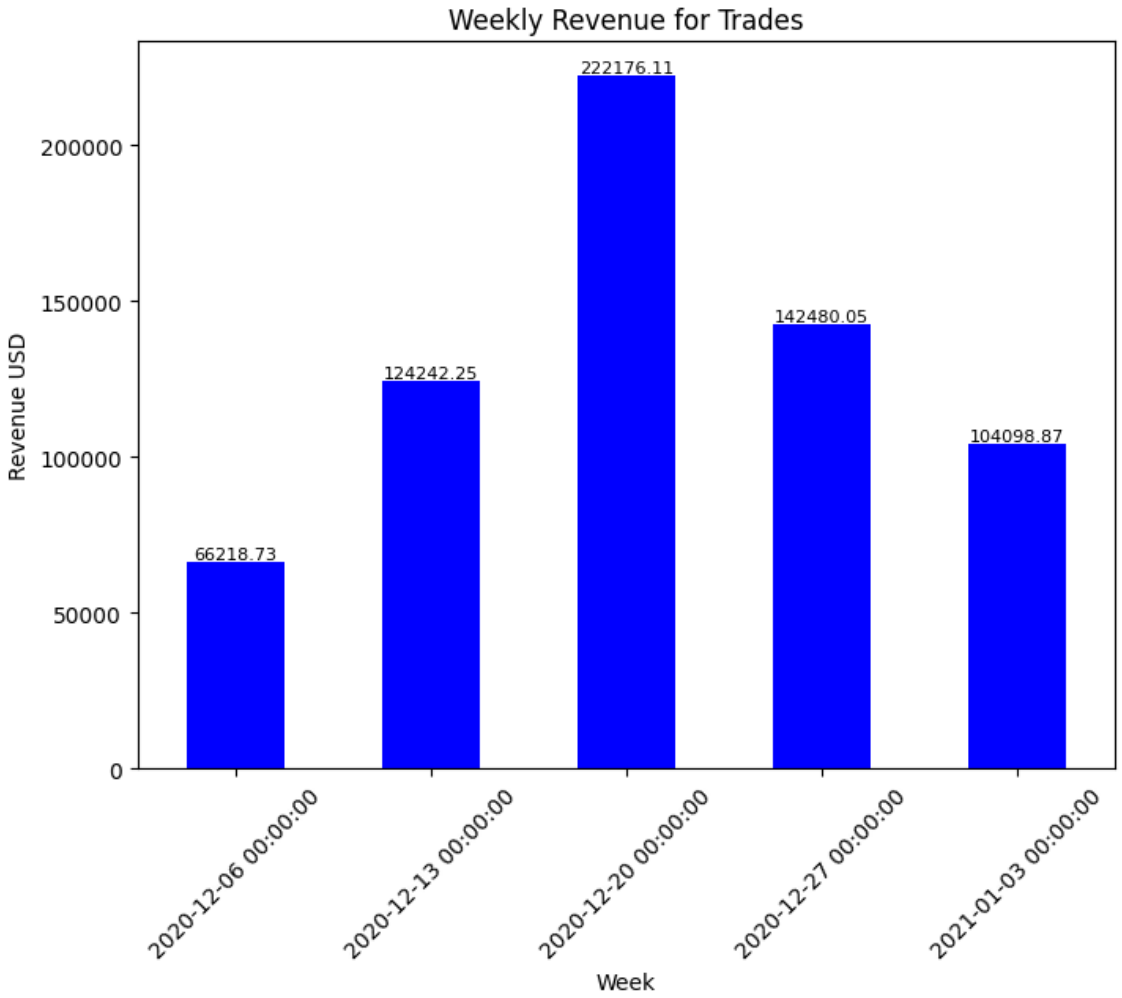


2.2 Revenue

2.2.1 Get the total weekly revenue divided between trades and withdrawals. revenue.

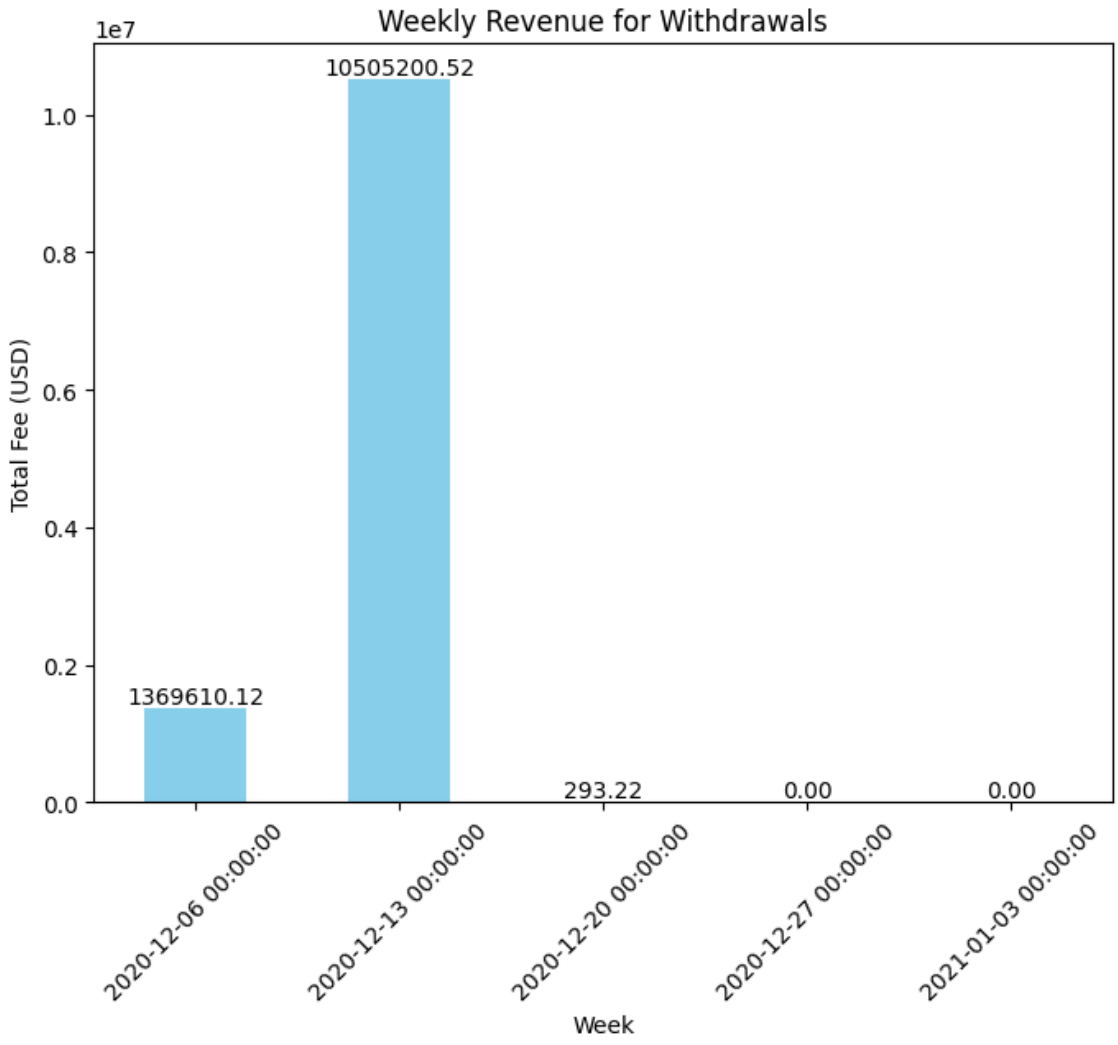
To calculate the revenue derived from trades, both the major currency user and the minor currency user are taken into account. The process involves finding the exchange rates for both cryptocurrencies at the time of the transaction. Then, these values are multiplied by their respective fees. Finally, the two fees are summed to calculate the transaction's revenue in dollars.

These results are grouped on a weekly basis to determine the total weekly revenue from trades.



For withdrawals, an initial step involves creating a table containing information about each fixed fee for every withdrawal currency. This table is subsequently utilized to multiply the appropriate fee based on the currency used for the withdrawal transaction.

Some withdrawal fee data is not found in the list; these values are assumed as NaNs or zeros. This may potentially impact the final results when conducting a revenue study. The options could include considering an arbitrary percentage for withdrawals not listed or assigning a fixed dollar value. For the sake of consistency with other reports, we have chosen to set these values to 0. However, my preferred option would have been to assign a fixed value, approximately 3 USD, for withdrawals in dollars, similar to USDT.



2.2.2 How many paying users do we have each week?

Defining a paying user as someone whose revenue is greater than 0, the objective with this data is to filter out all those users whose transactions result in a revenue greater than 0. After this initial filtering, unique users are considered to avoid counting the same user multiple times. This process helps in identifying and focusing on users who contribute positively to the revenue.

```
weekly_paying_users table:
     created  users
0 2020-12-06     10
1 2020-12-13      0
2 2020-12-20      5
3 2020-12-27      7
4 2021-01-03      3
```

2.2.3 Get the weekly ARPU (average revenue per user) and ARPPU (average revenue per paying user).

Without delving into the specific definitions of these two indices, for ARPU (Average Revenue Per User), all unique users from both trades and withdrawals are considered, regardless of

whether they generate revenue or not. As for ARPPU (Average Revenue Per Paying User), it only takes into account the paying users who have a revenue greater than 0. This distinction helps provide insights into how revenue is distributed among different user segments.

created	user_id	fee_usd	ARPU
2020-12-06	17	11496.349617	676.255860
2020-12-13	13	0.000000	0.000000
2020-12-20	12	33.436432	2.786369
2020-12-27	9	1.056495	0.117388
2021-01-03	7	0.146604	0.020943

created	user_id	fee_usd	ARPPU
2020-12-06	10	17244.524426	1724.452443
2020-12-13	0	NaN	NaN
2020-12-20	5	60.185578	12.037116
2020-12-27	7	1.358351	0.194050
2021-01-03	3	0.244341	0.081447

2.2.4 Based on what you already know, should Bitso invest more in trading or ramps (deposits & withdrawals) strategies to increase profitability? You can support your answer with any other insights from the data in the challenge.

Based on the provided insights and the nature of Bitso, the company should focus its investment strategies more on trading rather than ramps (deposits and withdrawals) to increase profitability. Here's why:

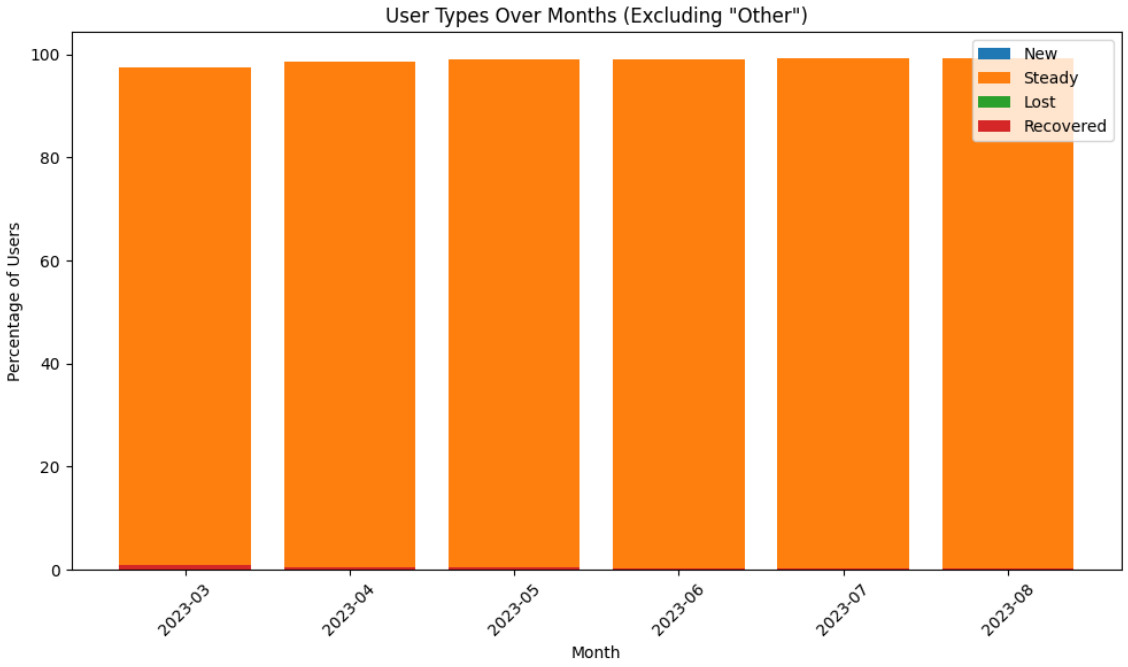
- **ARPPU vs. ARPU Discrepancy**: The fact that the Average Revenue Per Paying User (ARPPU) is significantly higher than the Average Revenue Per User (ARPU) on a weekly basis suggests that there is untapped potential among the existing user base for generating more revenue through trading activities. This implies that promoting and incentivizing other types of trades that generate commissions or introducing fees for currently free services could be an effective strategy.
- **Profitability Source**: The data indicates that the company's profits are predominantly driven by trading activities rather than low withdrawals. Given this, it's prudent to capitalize on this strength by implementing advertising and promotional campaigns that continue to encourage users to engage in trading activities. Enhancing the trading experience, offering educational resources such as free courses, seminars, and trade events with benefits like fee waivers could attract and retain users, thus boosting trading volumes.
- **User Expansion Strategy**: While attracting new users is always a positive strategy, it's equally important to understand the state of the existing user base and how it evolves over time. This goes beyond merely having positive balances. Exploring the user base comprehensively and their changing dynamics is crucial. Efforts should be made to prevent user stagnation by encouraging them to explore the trading features. Offering educational resources and incentives can play a significant role in expanding the pool of active traders.

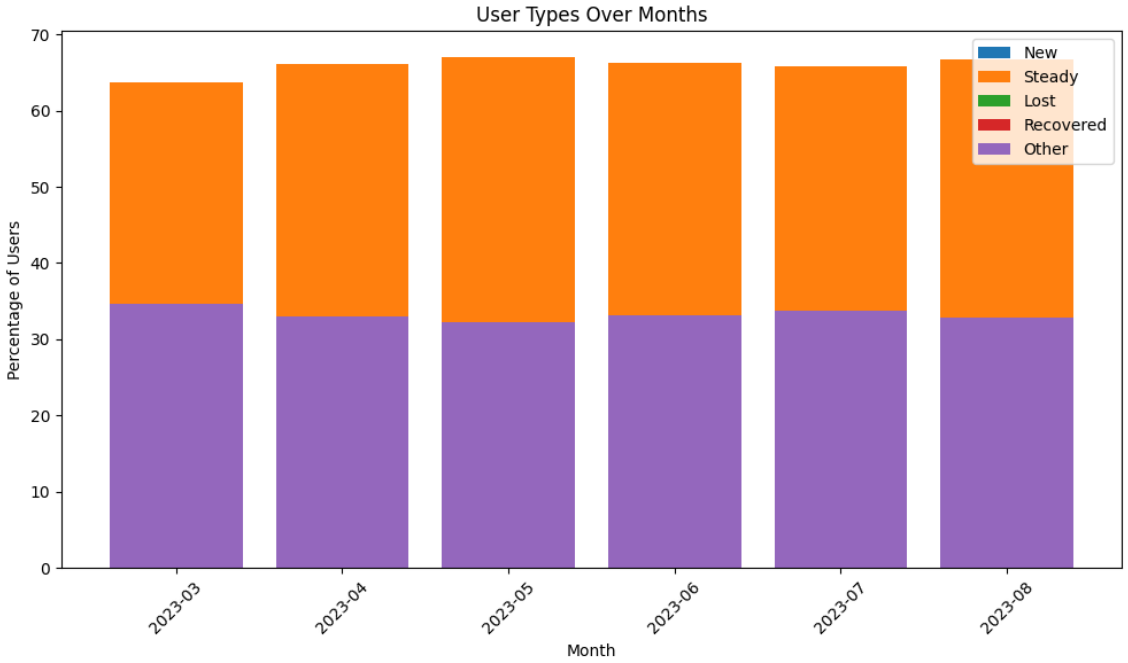
In summary, Bitso should consider investing more in trading strategies to increase profitability. This decision is influenced by the significantly higher ARPPU compared to ARPU, the dominance of trading in generating profits, and the potential for expanding the user base through educational initiatives and incentives.

2.2.5 BONUS

The bonus challenge entails creating a 100% stacked column chart representing different user states based on their account balance. To achieve this, we utilized the balance database, assigning a "true" value to each user type within each row. Surprisingly, there are many users who do not fit into these four segments. To accommodate this, we introduced another segment labeled as "others" to encompass these users.

The majority of users maintain positive balances, which is a positive sign. However, it's crucial to delve deeper into the "others" segment, as it represents a significant portion of the chart. Understanding this segment better is essential for formulating a strategic approach moving forward.





2.3 TOTAL DISCLAIMER

It's important to note that the insights and conclusions drawn from the data analysis in this report are based on the information available within the provided dataset. The accuracy and relevance of these conclusions may be subject to real-world factors, market dynamics, and external influences that extend beyond the scope of this dataset.

The recommendations and strategies proposed here are speculative and should be carefully evaluated in a real-world business context. Additionally, any financial or strategic decisions should be made with consideration of a broader range of data, market conditions, and expert consultation as necessary.

The analysis in this report serves as a preliminary exploration of the data and provides a starting point for further investigation and decision-making. It should not be the sole basis for any critical business decisions without comprehensive validation and external context.

In a professional setting, it's important to acknowledge that assessing the significance of the generated graphs and identifying any potential programming errors often necessitates comparative data, such as data from previous years. In a real-world work environment, the next logical step would involve the validation of this information.

Comparing current data with historical data is essential for conducting trend analysis and detecting anomalies. This comparative approach allows us to determine whether the observed patterns align with past trends or if there are substantial deviations. Such deviations could signal errors or unusual circumstances.

2.4 Conclusion

Completing this challenge was a lot of fun. While researching the topic during this time, I learned a ton and always strive to add extra value to the information I have. Finding that key strategy in the data to tackle challenges is what truly excites me.

I wanted to thank you for the time you took to put together this challenge, along with the datasets. I also greatly appreciate the extra time you provided. I assure you I made the most of every second to complete this report to the best of my ability.

catalinapugacz@gmail.com

I hope you find my work useful, and of course, I look forward to working with you in the near future. Thank you very much!

catalinapugacz@gmail.com