General Analysis on Rekt Database

The analysis contains the data between (June 2011 and January, 2023)

First of all, the data was web-scraped from the shared link following the pattern with page numbers. Having got the data, the data cleaning process started. Basically, the time/date columns are converted to datetime to make datetime operations without losing the information/dates/etc. In the next stage, the null values are checked, the portion of them per column. Apparently, FundsRecovered, IsVerifiedSourceCode, IsPublicedTeam contains significant number of null values; thus, those columns are dropped.

The analysis is implemented in two different approach. The first one is to analyze the whole dataset whilst another one is the analysis of the whole 2022 year.

In order to find the Net Loss, the fundsReturned is subtracted from fundLost. Then, per TypeOfIssue, the data is grouped and aggregated as sum. Based on the graphics, it is visible that (if we exclude Other) Rugpull is the one that causes significant loss for the entire dataset. When the analysis is done on number of occurred issues, rather than the net loos, it is quite clear that Honeypot is the issue that happens the most. Rugpull is placed as the 2nd in terms of number of frauds. So, based on the analysis, one can drive that RugPull is considered as one of the most negatively affecting issue to the system. To prevent this, it is better to build a machine learning algorithm (XGBoost Scoring could be fine) to score the SmartContracts executing the transactions.

To continue with, while executing the code (GeneralAnalysis), it provides a user option to input desired resampling, such as Monthly, daily, 3daily, weekly, etc. So as the user selects desired input, the graph is auto generated. Here, it shows that despite the recently significant decreases in RugPull, it is still quite dangerous to re-appear. ***Phishing*** appears to be in the ***increasing trend***, so reasonable actions should be implemented to prevent the issue from increasing. Since the end of April, honeypot does not appear so it is quite successful.

In the another approach describing the whole 2022-year analysis, monthly number of attacks are formatted in dictionary. In other words, the number of issues happened each month of 2022 is found out. Based on the least square polynomial fitting, the slope is found for every issue. Based on the results, despite the recent 2 months decreases, Access Control is also considered as possibly increasing issue. However, ***Flash Loan Attack, Phishing*** are the type of issues that are increasing trend in the last two months.

To sum up, while occurrences of the issues such as Access Control, Rugpull, significantly got decreased, *Phishing* is increasing. As a data analyst, I believe this type of issue is implemented through email aiming to get the wallet key information of the users. To solve the problem, the users have to be well-educated that efficiency can be measured by A/B testing on different groups.