48 HOUR CHALLENGE
-
Grants Round 3 CLR
-
Data Analysis Bounty!

EXECUTIVE SUMMARY

Prepared By—
Aditya Vyas

GitHub - adivyas99
LinkedIn - adityavyas99

Introduction

✳ The data set as we know was in Json format. It contained various fields including Title, Tags, URL etc.

✳ Firstly we formatted the data into an organized way so that we can perform analysis on it. Then after I extracted those columns which could be useful.

✳ Although in Data Analysis we have to explore all attributes to get the crux of data.

✳ Finally I started to rotate refined data from different angles to get every possible insight.

✳ Tools (or could be) used-

- Python (as language) - Very simple interpreter based programming language capable to perform many tasks like- Backend, ML etc.
- Pandas - Most popular Data management and exploration tool in Python.
- Plotly - An amazing Data Visualization Library capable of making dashboards.
- Matplotlib - Open-source data visualization library
- Seaborn - Uses matplotlib backend and provide APIs to make visualizations

Insights

- All of the links are unique.
- The attributes (estimated_round_3_clr_match_usd and total_amount_received_usd_life) are inversely proportional to each other
- There are 26 unique Tags combination in the dataset
- There are 75 unique admin_profile_name out of 100 entries
- 53% of the Tags combination used only 1 time
- There are 77 unique admin_address in the dataset
- There are 7 data points which don't have any admin_address
- Most of the admin_profile_name are used 1 time except 12 rows
- Most of the estimated_round_3_clr_match_usd are under 2500
- Most of the total_amount_received_usd_life are under 20000
- Highest total_amount_received_usd_life only have 1 contribution which makes it an exception
- Relation between no_of_contributions and estimated_round_3_clr_match_usd is proportional to each other

Thank You