

# 1 Spare some Ethereum Comrad

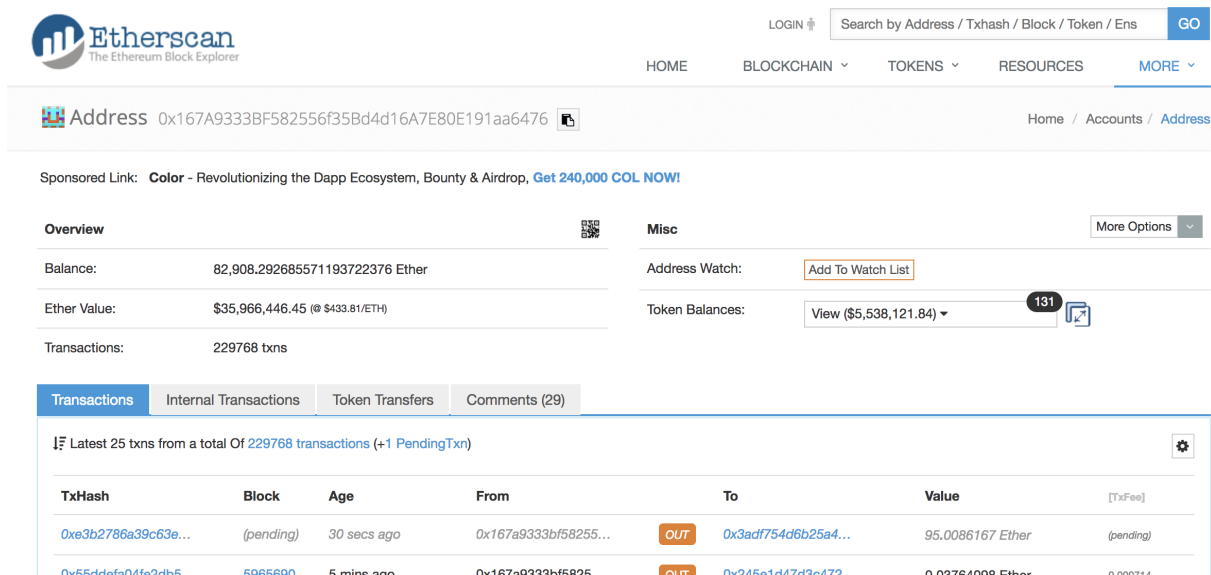


Figure 1: Etherscan page for a particular high value address.

A common method for analyzing the ethereum blockchain is the online tool *Etherscan*. The tool allows users to search, comment and tag individual entities on the blockchain. Figure 1, shows a sample transactions page where various transactions are occurring.



Figure 2: A typical eth-donation request on etherscan page.

Beyond the transactions which are of value in understanding the decentralized cryptocurrency economy, a very poorly understood part of *Etherscan* is the comments section for various addresses. The comments section of *Etherscan* is highly entertaining, with most comments either claiming they were scammed by the associated address, or begging for ETH in intentionally broken english. We suspect the ETH's are going to fund the decentralized revolution.

Figure 2 shows a typical ETH requester asking for ETHs giving their public key where the ETHs are to be deposited. ETH, short for Ethereum, being the fuel that runs all smart contracts on the network. Suppose we have an NLP system that can scan all pages of *Etherscan* in real-time tagging all such ETH donation requests, over all addresses. Then understanding the begging activity of decentralized economy can be of value for cocktail party conversation.

## 1.1 Analysis Instructions

1. The attached file contains the timestamps of tagged ETH donation request activity over the entire ethereum address space. Aggregate the counts based on a 15 minute time interval, visualize and describe the resulting time series of the counts in ways that best characterize the underlying pattern of activity. Please report/illustrate important features of the activity, such as any periodicities. If there

are data quality issues, please report them. Try your best to tell a data story supported by sound statistical analysis.

2. Suppose you're tasked with predicting the amount of begging activity. Based on the insights you developed above describe a method to forecast the expected future activity at the 15 minute granularity over the following hour (the 4 following time periods). Apply your method to the dataset and evaluate it's predictive accuracy in a reasonable way. Discuss your method and it's accuracy.
3. Not only are accurate mean predictions important, but so are appropriate characterizations of the stochastic nature of the request activity. Describe how you would evaluate the accuracy of these stochastic predictions.

There is no hard deadline, but we will expect some response within 1 - 1.5 weeks. There is no language constraint, however given the nature of the problem python or R are preferred. Additionally, there is no "right" answer so we want to see how well you can tell a story with the data and develop methodologies to answer the general problem statements. Please provide your source & write up.