Ethereum Begging Activity Analysis

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.**

Lets load the data and resample it based on 15 minute time interval using python.
Our Data looks like this:

	tagged_begging
0	2018-01-01 20:13:18
1	2018-01-01 20:16:10
2	2018-01-01 20:16:37
3	2018-01-01 20:16:36
4	2018-01-01 20:26:21
5	2018-01-01 20:21:41
6	2018-01-01 20:12:16
7	2018-01-01 20:35:47

tagged_begging	Count
2018-01-01 20:00:00	2
2018-01-01 20:15:00	6
2018-01-01 20:30:00	9
2018-01-01 20:45:00	7
2018-01-01 21:00:00	1
2018-01-01 21:15:00	4
2018-01-01 21:30:00	0
2018-01-01 21:45:00	4

Lets transfer the dataframe in excel file and perform the following analysis in Tableau.
Overall Observed Data:

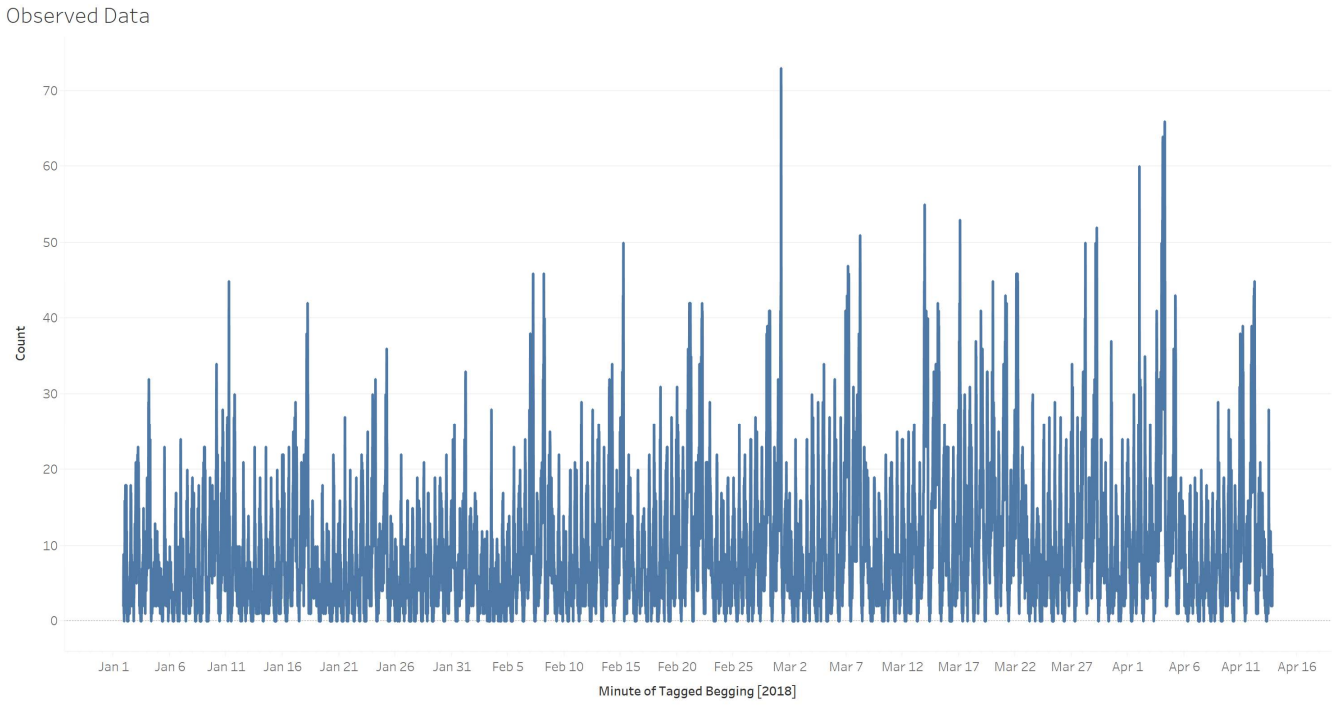

Fig.1 Observed Data per Minute

Lets deep dive into the data to understand the underlying pattern of the activity. The following diagram shows the trend analysis of the begging activity. As you can see, there is downward trend for monthly and daily activity.But if we look at the weekly pattern and hourly pattern, the trend goes upward showing that number of requests are increasing weekly and hourly. Based on the data, the first quarter has upward trend. We cannot conclude that this is the case for next quarters as we do not have the data for all the months in a year.

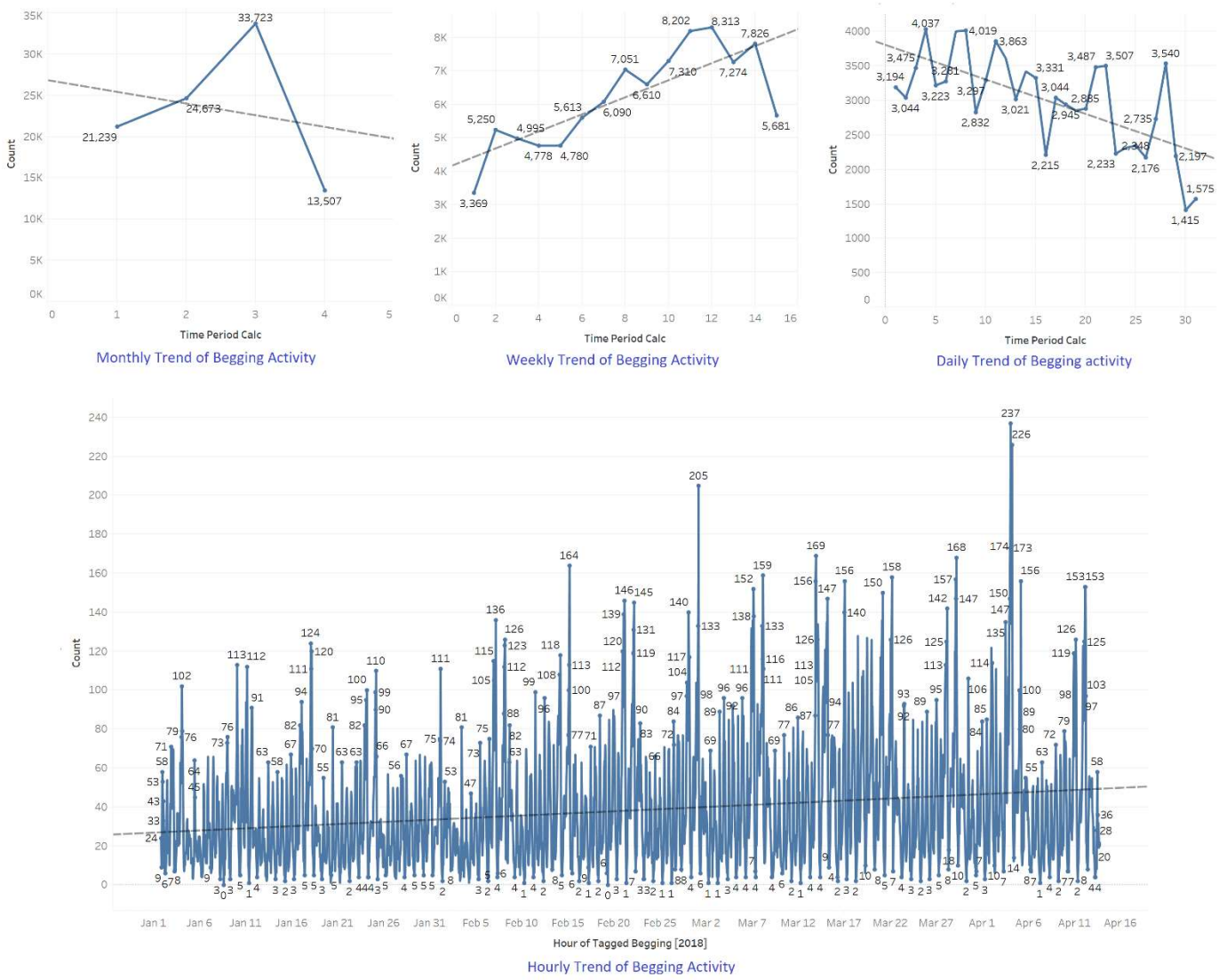

Fig.2 Trend Analysis per Time Period

In figure 3, it is clearly seen in the plot that the change in the request activity is periodic. In each month, the number of requests increases when transitioning from first week to second week. And when approaching last week of each month, number of requests are declining.

Fig.3 Weekly Periodicity

In the figure 4, the pattern repeats itself after every 15 mins. There is periodicity in 15 mins time interval.

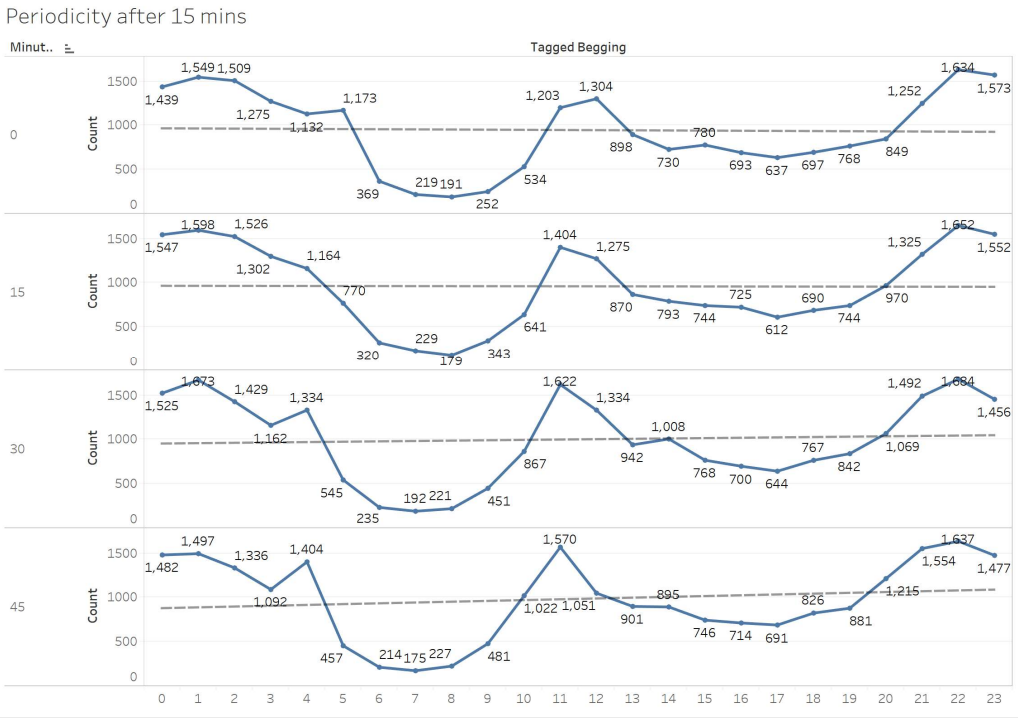

Fig.4 Periodicity after 15 mins

Fig.5 shows the distribution of no. of requests per month. In first quarter, the number of requests were increasing, March being the top month.

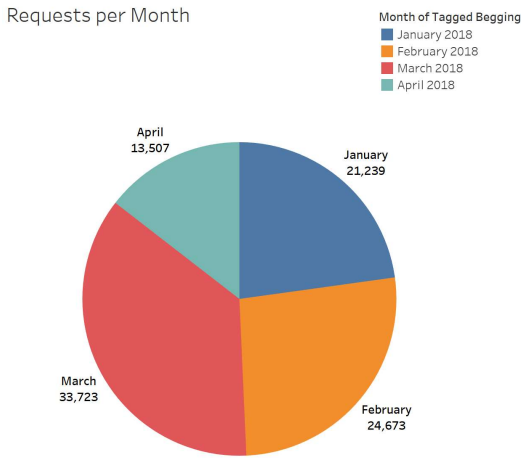

Fig.5 Monthly distribution of requests

In the figure 6, we can see the most active minutes, time of day and days of week.
Active Minutes: Last 15 mins of the hour.
Active time of day: 12am-1am, 10am-12noon, 11pm-12pm.
Active Days: Wed, Thurs.

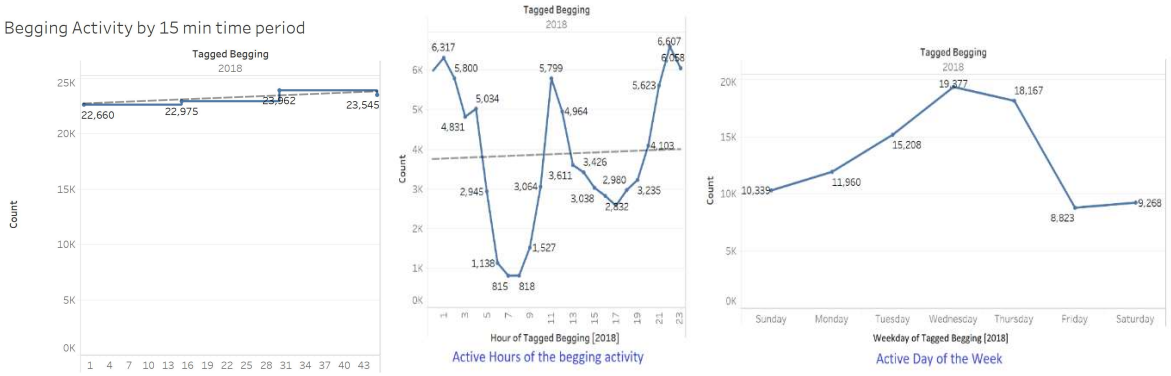

Fig.6 Active minutes, time of day and days of week

Figures 7 & 8 shows the relation between the daily begging activity and daily price changes. We can infer that as the price is declining over the period of time, number of requests are increasing.

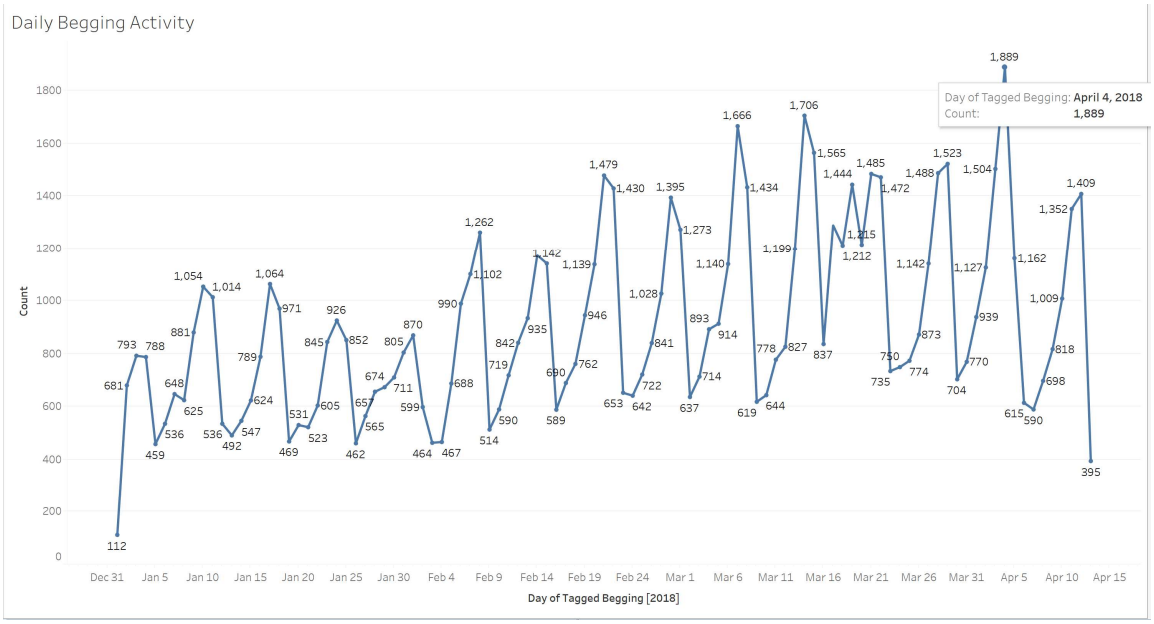

Fig.8 Daily Begging Activity

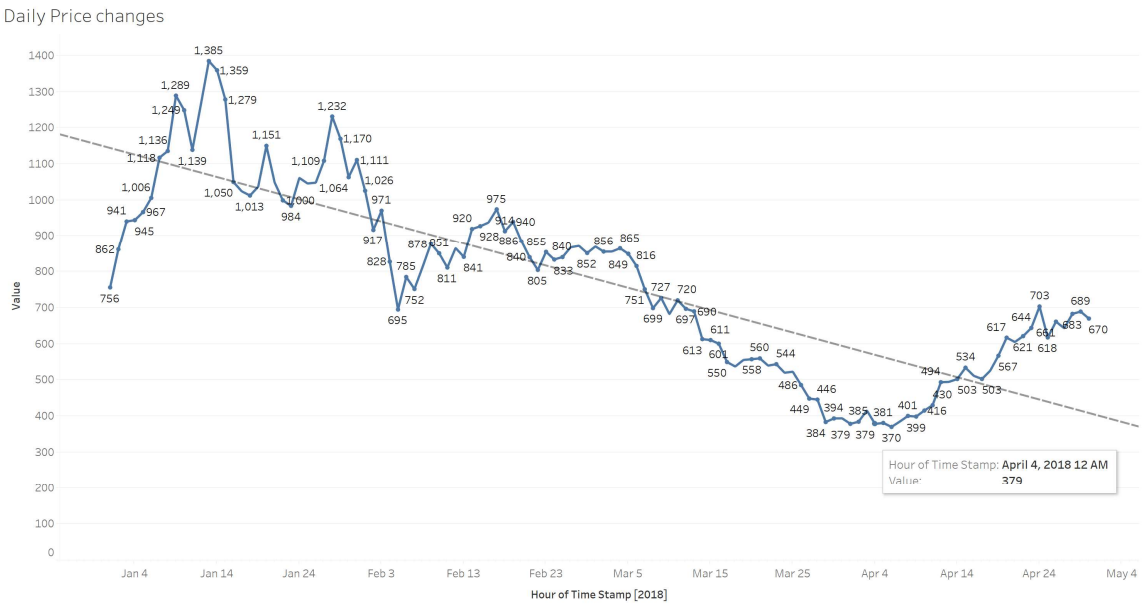

Fig.9 Daily Price changes

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 its predictive accuracy in a reasonable way. Discuss your method and its accuracy.**

   I am using SARIMA model to forecast the expected future activity at the 15-minute granularity over the following hour. Please find the code in Jupyter notebook attached in the email.

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.**

   Please find it in the jupyter notebook. MSE = 4.11.

Sources:
https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b