Todd Gavin

Final Project Report

1. **What's the name of your final project? Please describe it as a research question and provide a short description [3 points]**

My final project is titled "Analyzing how the transaction data from major cryptocurrency wallets on the polygon network associate with the price action of the MATIC token."

The MATIC token is a cryptocurrency that operates of the Polygon layer 2 network that sits on top of the Ethereum network. Essentially, the project collects a long list of transaction data and metadata associated with each transaction for each wallet listed in the function. Once the transaction data is collected, it is merged with a price action dataset for the MATIC token. Using the data frames, a series of analysis using line graphs, scatter plots, and linear regression analyses are used to analyze and interpret the data.

2. **How to run your code? [3 points]**

Link to Github Repository: https://github.com/todd-gavin/DSCI510-Final-Project

Run as Python Notebook (Most preferable to elicit best result):

A. Download the file 'masterCode.ipynb' from the '/code' folder.
B. Open sed file in VScode or another python notebook compiler such as Jupyter.
C. Censure that all the dependencies are pip installed on to your device for which you are running the code on.
D. In the same folder for which you store sed file in, create a folder called 'visualizations'. (The plots will be saved to this folder once the code runs.)
E. Click 'run all' at the top of your interpreter and let the magic happen!

Run as Python Script:

A. Download the files 'main.py, collectData.py, organizeAndMergeData.py, analyzeData.py' from the '/code' folder. Make sure to download the 'requirements.txt' file.
B. In the same folder for which you store sed file in, create a folder called 'visualizations'. (The plots will be saved to this folder once the code runs.)
C. Open a terminal at the folder for which the code is stored in. Run command 'pip install r- requirements.txt'.
D. To run the code, run the command 'python main.py'. And let the magic happen.

3. **What data did you collect? How did you collect them? How many data samples did you collect? [3 points]**

I used two different data sources to collect data. I used the Etherscan API to collect the transaction data from the crypto wallet and I used the Python Historic_Crypto library to collect the pricing data for the MATIC token.

The Etherscan API returns the transaction data as a JSON and the Historic_Crypto library returns the data as a dataframe.

For the Etherscan API, I collected about 4,000 transaction from several wallets and for the Historic_Crypto library, I collected about 1000 days (`3 years) worth of daily pricing data from the MATIC token.

Compared to my original plan, I am only analyzing the data from the MATIC token and not the ETH cryptocurrency. I also decided to increase the number of wallets (hence, more transaction) I was analyzing for my project. Additionally, I decided to add linear regression analyses' to my project.

4. **What kind of analyses or visualizations did you do? [3 points]**

I created 7 visualizations for which I performed a linear regression analysis on the first 4 of the 7 visualization variables.

Figure visualizations are as followed:

- Scatter Plots with Linear Regression Models:
  - Fig 1. Along the x-axis is the absolute value of the total amount of MATIC moved (bought or sold) for a give day in USD. Along the y-axis is the price of MATIC in USD for a given corresponding day. The R-value is 0.0078 stating that there is no significant correlation between these two variables.
  - Fig 2. Along the x-axis is the price of MATIC in USD for a given day. Along the y-axis is the gas price cost in wei for the corresponding given day. The R-value for the data is 0.026, therefore, there is to a significant correlation in the data.
  - Fig 3. Along the x-axis is the daily price difference between the high and low for a given day. Along the y-axis is the absolute value of the daily price difference between the open and close for the corresponding given day. The R-value is 0.77 with a P-value < 0.05, therefore, there is reasonable statistical correlation between the two variables.
  - Fig 4. Along the x-axis is the number of buys for a given day and along the y-axis is the number of sells for the corresponding given day. The R-value is 0.75 and the P-value < 0.05, therefore, we can infer that there is a reasonable statistical correlation between the two variables.
- Overlapping Line Graphs

o Fig 5. This line plot depicts the price of MATIC to the magnitude of x100000000 and the volume of token movement for the corresponding day. There seems to be a slight resemblance between the volume moved and the peak price of the MATIC token.

o Fig 6. This line plot depicts the price of MATIC to the magnitude of x100000000 and the absolute value token cost moved for the corresponding day.

o Fig 7. This line plot depicts the number of buys against the number of sells for a corresponding day. There seems to be an association between the number of buys and number of sells for a particular day.

Conclusion:

From the data, regression models, and visualizations, there seems to be very little significant association between the transaction data for the major crypto wallets trading MATIC, and the pricing and price action of the MATIC token. Although there are two figure that seem to have a significant correlation, they do not provide a unique predicter for when to purchase or sell the MATIC token. Overall, it goes to show how unpredictable the cryptocurrency market is and even with the best traded crypto wallets, even they do not have a formula for when to buy and sell the MATIC token so that it is the most profitable.

Future Work:

Given more time, I would search for more granular data on the transaction history of a specific crypto wallet. Preferable, I'd like to gather more transaction while also knowing the current amount of MATIC token that is in a crypto wallet at a specific date and time. I would that be able to analyze the existence of (or lack of a) relationship between the crypto wallet MATIC token balance and the pricing of the MATIC token relative to its previous pricing.