A Multi-factor model for forecasting crypto returns over time

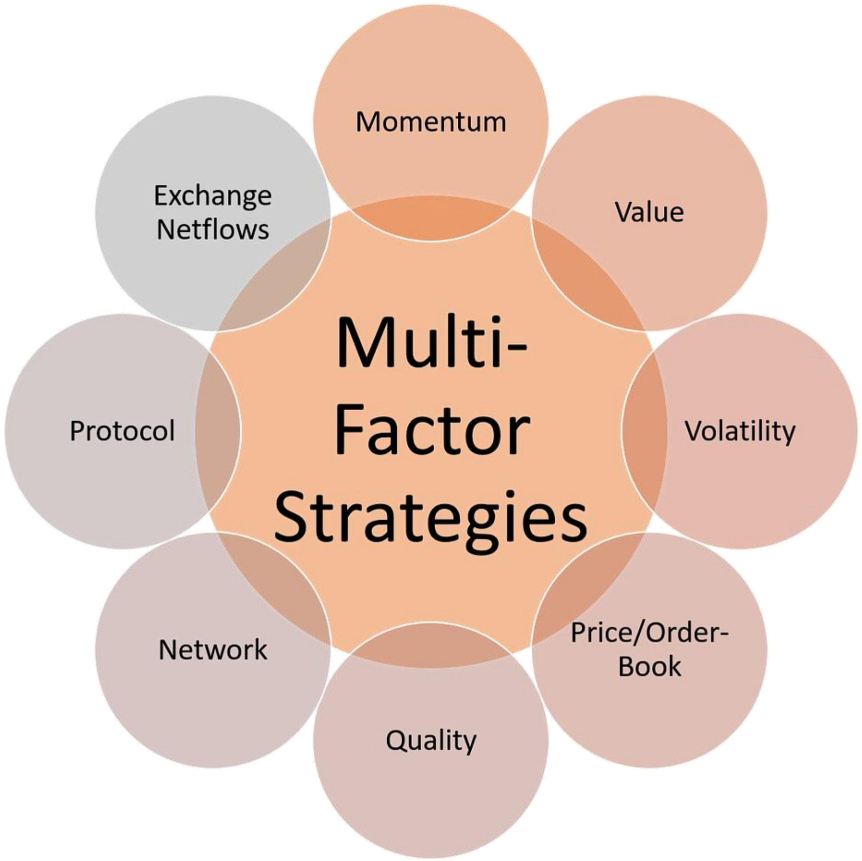
Malihe Alavi

27 Aug 2024

Introduction

Crypto factor modeling is an approach used to analyze and predict the return of cryptocurrencies by identifying and assessing various factors that influence their returns. These factors can include market trends, liquidity, volatility, momentum, and other economic indicators. By modeling these factors, investors and analysts can gain insights into the potential risks and opportunities within the crypto market, helping to inform investment strategies and portfolio management decisions.

In this work , using market factors like Momentum, Value_Factor , size_factors , … and economic factors like inflation rate, unemployment_rate , … and also environment factors like google trends we tried to investigate and analyze the behavior of nearly 120 cryptos over time.



Data Collection

We collected the data from different sources. Unfortunately most of the accurate sources of crypto historical data need premium subscription or are restricted in some specific countries. So, finding a comprehensive and accurate data was really challenging.

Thanks to some free sources like CCXT library (to collect limited historical data), Fred (to collect some economic data), google trends API and Coingecko API, we could gather sufficient amount of historical data for nearly 120 cryptos which are suggested in numerai website by considering the following circumstances:

link

- Stablecoins, wrapped tokens, liquid staking tokens (e.g. stETH, rpETH)
- Tokens less than two years old
- Tokens with less than $1,000,000 trading volume in the last 24 hours
- Lower market cap tokens for duplicate symbols
- After the above are removed, tokens that are either too stable or highly correlated:
- Stable: Average 6 month daily returns less than 0.00001
- Correlated: Removal of tokens with a correlation in daily returns >= 0.95 over the past 6 months, keeping the one with the highest market cap

Data Collection

OHLCV data : this data collected from CCXT library for 128 tokens. Although CCXT has also limitation rate for accessing the full historical data and it allows users download 1000 records per tokens without API key. you can see the complete list in here

Circulating supply : In order to calculate market cap for each crypto, we first gather their corresponding circulating supply from coingecko API.

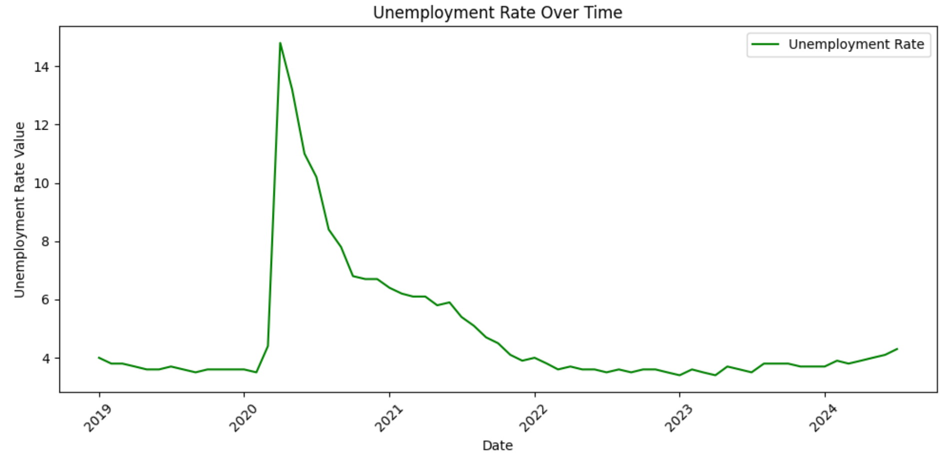
Google trend : to collect the google trend data for each 128 tokens, we used pytrend library in python and fetch 5 years data of each tokens.

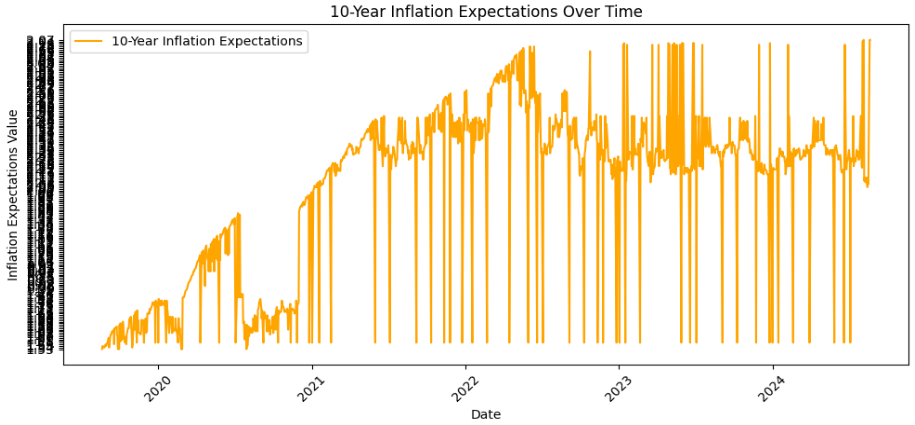
FRED : this website holds all economic data of US and worldwide. We could download 10-years inflation rate , unemployment rate , Gross domestic product (GDP) and interest rate from it.

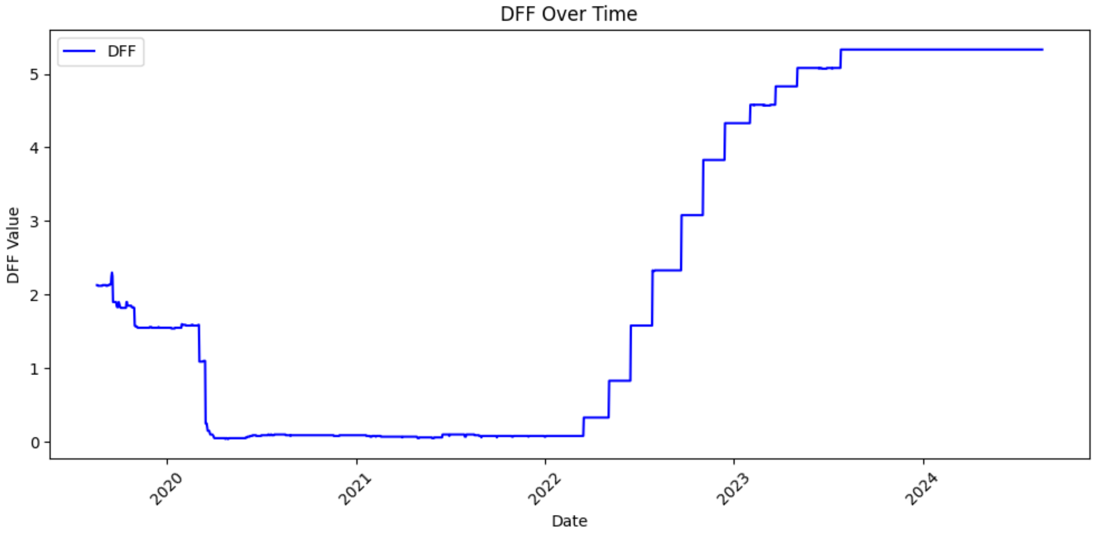
Factors

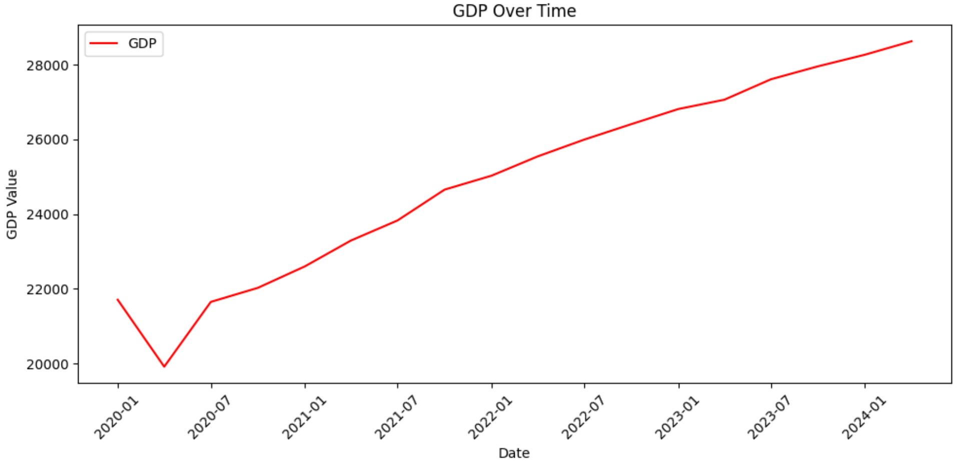
We used 3 kind of factor for our analysis :

- **Market Factors** : Momentum , market cap , size factor, liquidity , volatility , value factor and amihud_illiquidity
- **Economic Factors** : Federal Funds Effective Rate (DFF) , inflation rate. GDP
- **Social Media** : Google Trends









Key findings

Correlation Analysis: The research identified strong correlations among market factors, particularly between returns and HML (High Minus Low) factors, indicating that these factors significantly influence cryptocurrency performance.

Impact of Economic Factors: Economic indicators like the Federal Funds Effective Rate and inflation expectations showed weak correlations with cryptocurrency returns, suggesting limited predictive power when analyzed in isolation.

Modeling Results: The OLS regression models indicated that HML and momentum are significant predictors of returns, while other factors like market cap and volatility did not show significant effects.

Residual Analysis: The residuals from the regression models displayed patterns suggesting potential heteroscedasticity, indicating that the variance of errors may change with different levels of predicted values.

Google Trends Influence: Although Google Trends data was included in the analysis, it showed weak correlations with cryptocurrency returns, indicating it may not be a strong standalone predictor.

Liquidity Metrics: The analysis of liquidity across tokens revealed that higher liquidity is generally associated with better market performance, although this relationship was not uniformly strong.

Market Dynamics: The findings suggest that market dynamics, particularly momentum and value factors, play a critical role in forecasting cryptocurrency returns, which could aid investors in making informed decisions.

Factor Construction-Market Factors

Return is calculated as the percentage change in the closing price from one period to the next. This measures the asset's performance over time, indicating gains or losses.

```
df['return'] = df['Close'].pct_change()
```

Momentum measures the tendency of an asset's price to continue moving in its current direction. it is calculated as the percentage change in the closing price over a 14-day period. This measures the speed and magnitude of recent price changes, indicating whether the asset's price is increasing or decreasing.

```
df['Momentum'] = df['Close'].pct_change(periods=14)
```

Volatility is measured as the standard deviation of returns over a 30-day window. This captures the degree of variation in returns, reflecting the asset's risk.

```
df['Volatility'] = df['Return'].rolling(window=30).std()
```

Factor Construction-Market Factors

Market Cap Market capitalization is the total value of a cryptocurrency's circulating supply. It is calculated by multiplying the current price by the circulating supply. This metric provides an estimate of the cryptocurrency's market value and size.

```
df['market_cap'] = df['Close'] * circulating_supply
```

Liquidity is calculated as the average trading volume over a 30-day period. Higher average volume suggests better liquidity, indicating ease of buying or selling the asset without significant price impact.

```
df['Liquidity'] = df['Volume'].rolling(window=30).mean()
```

Amihud Illiquidity This measure captures the price impact per unit of trading volume. It is the average of the absolute return divided by volume over 30 days. Higher values indicate less liquidity, as small trades cause larger price changes.

```
df['amihud_illiquidity'] = (df['return'].abs() / df['Volume']).rolling(window=30).mean()
```

Factor Construction-Market Factors

market factor is calculated as the asset's return minus the risk-free rate which is considered `.0425`. This represents the excess return over a risk-free investment, capturing the asset's sensitivity to market movements.

```
df['Market_Factor'] = df['return'] - risk_free_rate
```

The size factor often refers to the market cap of a company or asset, used to categorize it into small, mid, or large-cap. In crypto, it can be used to differentiate between larger, more established coins and smaller, emerging ones. The size factor can influence the risk and return profile of the asset, with smaller caps generally being more volatile.

```
df['Size'] = pd.cut(df['Market_Cap'], bins=[0,  1e6, 1e8, float('inf')], labels=['Small','medium', 'Large'])
df['size_factor'] = df.apply(lambda row: -1 if row['Size'] == 'Large' else 1 if row['Size'] == 'Small' else 0, axis=1)
```

Factor Construction-Market Factors

Value factor The HML (High Minus Low) factor is constructed by first calculating the NVT (Network Value to Transactions) ratio, which is the market cap divided by volume. This ratio is ranked over a rolling window to assess relative value. High and low thresholds are set using quantiles to identify overvalued and undervalued assets. The HML factor assigns negative returns to overvalued assets (above the high threshold) and positive returns to undervalued assets (below the low threshold). Any missing values are filled with zero to ensure completeness. This process captures the value premium by differentiating between high and low NVT-ranked assets.

```
df['NVT'] = df['Market_Cap'] / df['Volume']
window_size = 15
df['NVT_rank'] = df['NVT'].rolling(window=window_size).apply(lambda x: x.rank().iloc[-1])
high_threshold = df['NVT_rank'].quantile(0.7)
low_threshold = df['NVT_rank'].quantile(0.3)
df['HML'] = np.nan
for i in range(window_size, len(df)):
    current_rank = df['NVT_rank'].iloc[i]
    if current_rank > high_threshold:
        df['HML'].iloc[i] = -df['return'].iloc[i]
    elif current_rank <= low_threshold:
        df['HML'].iloc[i] = df['return'].iloc[i]
# print(df['HML'].isna().sum())
df['HML'].fillna(0, inplace=True)
```

EDA and Analysis – correlations

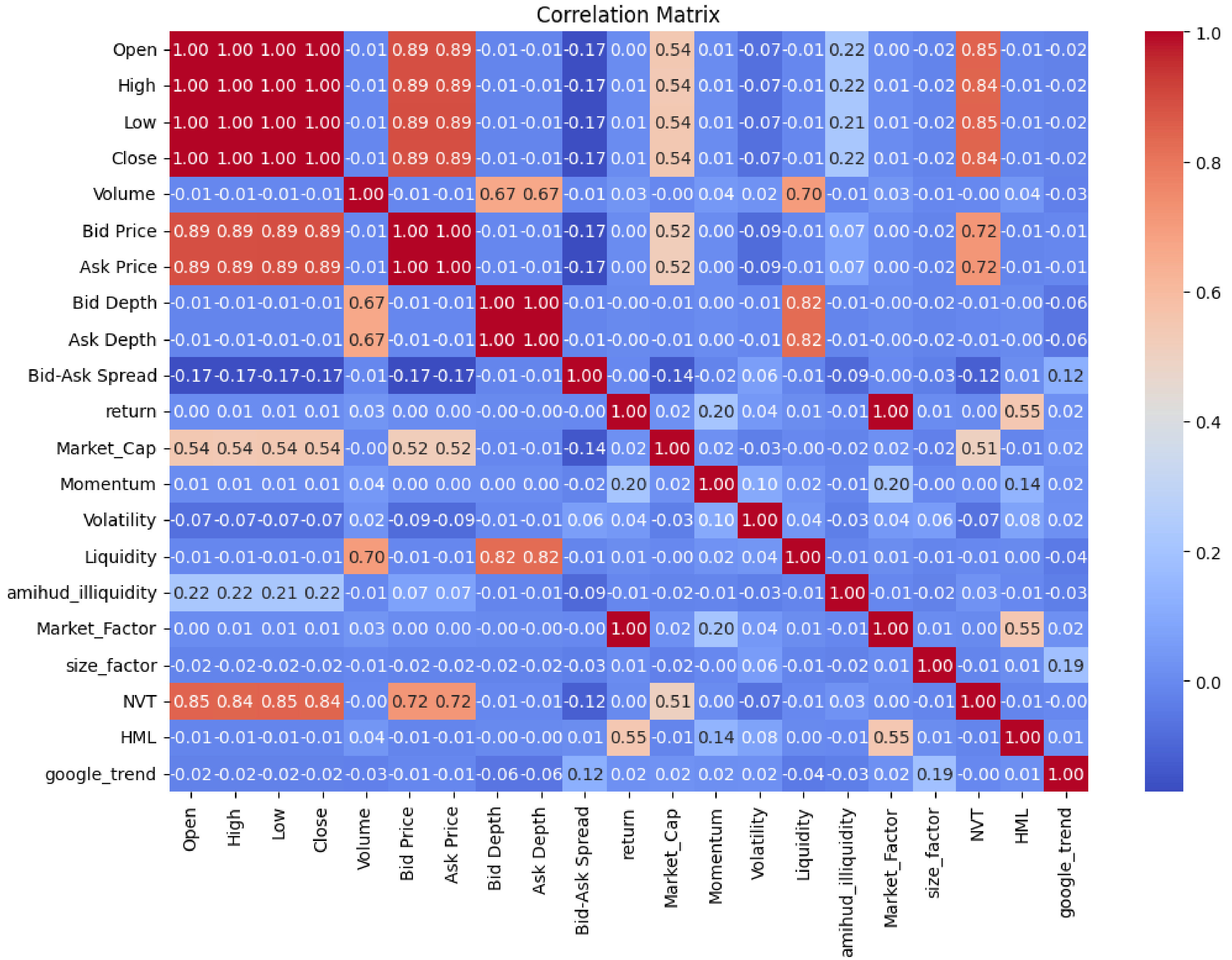
```
High_corr {
('HML', 'XVS_USDT'): 0.7821237261591748,
('HML', 'ADA_USDT'): 0.7707710129059268,
('HML', 'AXS_USDT'): 0.7092297450556192,
('HML', 'BADGER_USDT'): 0.7456920709075722,
('HML', 'BAND_USDT'): 0.7907051721895819,
...
('HML', 'ZIL_USDT'): 0.7714801688957789,
('HML', 'YGG_USDT'): 0.7051464392158925,
('HML', 'YFI_USDT'): 0.7677148973203305,
('HML', 'NMR_USDT'): 0.8171057089908815,
}
moderate_corr {
('Momentum', 'BAND_USDT'): 0.30264481027012635,
('Momentum', 'CTXC_USDT'): 0.3978617805347398,
('Momentum', 'CVC_USDT'): 0.32050631219154335,
('Momentum', 'DEXE_USDT'): 0.30166308731418573,
('Momentum', 'FUN_USDT'): 0.3281178617860198,
('Momentum', 'GRT_USDT'): 0.3056983957607997,
('Momentum', 'NMR_USDT'): 0.3032882422480108,
('Momentum', 'NULS_USDT'): 0.321347864041776,
('Momentum', 'RAY_USDT'): 0.3362683687231586,
('Momentum', 'REQ_USDT'): 0.31928480661675285,
('Momentum', 'TRB_USDT'): 0.305495166828514 5,
('HML', 'AGIX_USDT'): 0.5629940645692844,
...
('HML', 'OCEAN_USDT'): 0.6455779842630939
}
```

We examined the correlation of each factors with return for each crypto-currency. We categorized the correlation into three categories : low, moderate and high. As it was expected market_factor are highly correlated with return since it was built from return by subtracting a constant value. HML is also highly correlated in some tokens.
Among other factors, Momentum , Volume, HML and google trends are moderately correlated with return in some tokens.
Volume for most tokens were reported to have moderate correlation with return (about 0.3) so we haven’t report it here and just mentioned the other 3 factors.
Other factors have low correlation (less than 0.3) with return.

```
moderate_corr {

('google_trend', '1INCH_USDT'): 0.313547307462867 9,
('google_trend', 'BLZ_USDT'): 0.3185770397804029,
('google_trend', 'DGB_USDT'): 0.3149872481742219,
('google_trend', 'FIL_USDT'): 0.4294516538325876 5,
('google_trend', 'OGN_USDT'): 0.338282474979288 8}
}
```

EDA and Analysis – Pearson Correlations



EDA and Analysis – correlations

From the left correlation matrix

1.**Strong Correlations**:

Open, High, Low, Close, Bid Price, Ask Price: These have perfect or near-perfect correlations with each other, indicating they move together closely, which is typical for price data.

Bid Depth and Ask Depth: Also highly correlated, suggesting a relationship in market depth dynamics.

2. **Moderate Correlations**:

Volume and **Liquidity**: A moderate correlation exists, indicating that higher trading volumes might be associated with better liquidity.

Market Cap and NVT: Shows moderate correlation, suggesting that larger market cap cryptocurrencies might have more consistent network value to transaction ratios.

Momentum and Return: A moderate correlation, indicating that momentum could be a significant factor affecting returns.

Factors of Interest:

HML and Return: A moderate correlation suggests that high minus low (HML) factor could be relevant for explaining returns.

Google Trend: Shows weak correlations with most factors, indicating it might not be a strong predictor of any particular aspect of the market dynamics.

From the DFF and T10YIE

1. **DFF (Federal Fund Effective Rate):**

Return: Shows a weak negative correlation, suggesting that changes in the federal funds rate might have a slight inverse impact on returns.

Volatility: Has a weak negative correlation, indicating that higher DFF might slightly decrease volatility.

HML: Displays a weak positive correlation.

2.**T10YIE (10-Year Inflation Expectation)**:

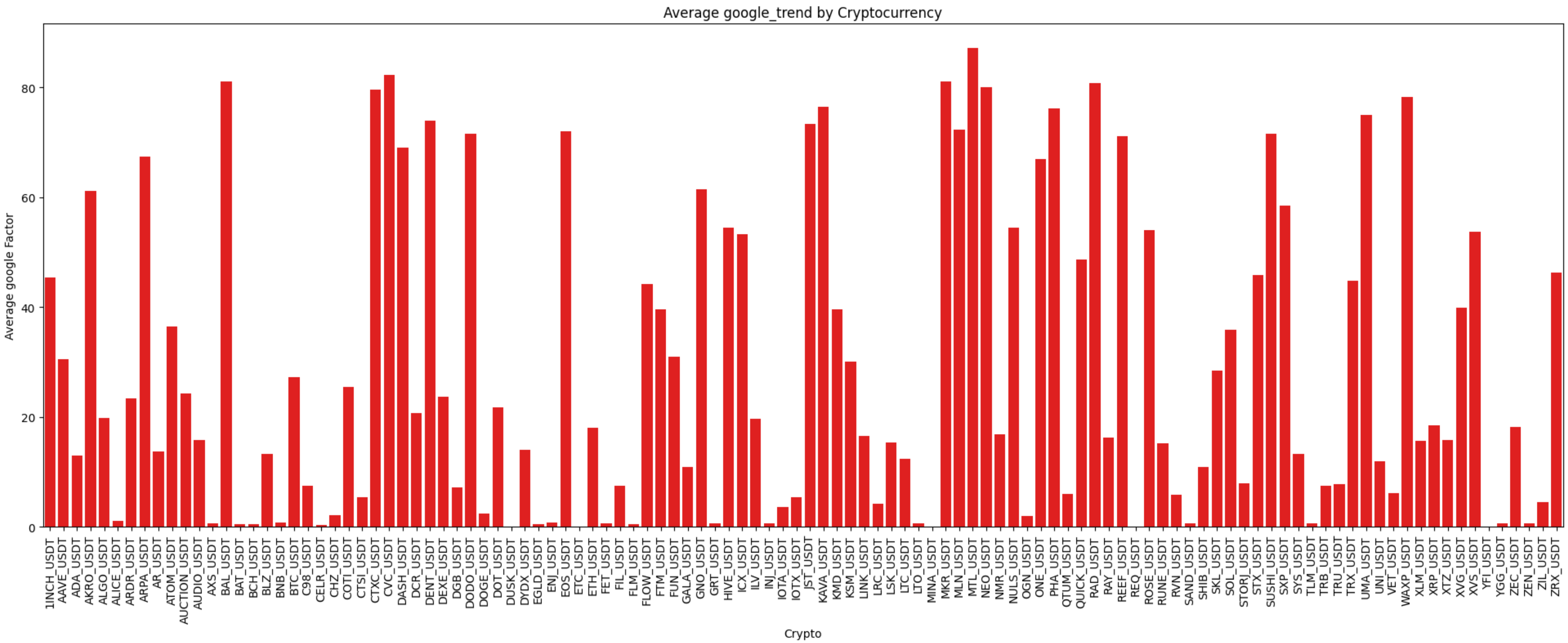
Market Cap: Shows a weak positive correlation, suggesting that inflation expectations might slightly influence market cap.

Return: Displays a weak negative correlation, indicating that higher inflation expectations might slightly decrease returns.

Volatility: Exhibits a weak positive correlation.

DFF and T10YIE have minor correlations with most variables, suggesting they might not be strong standalone predictors but could still provide insights when combined with other factors.

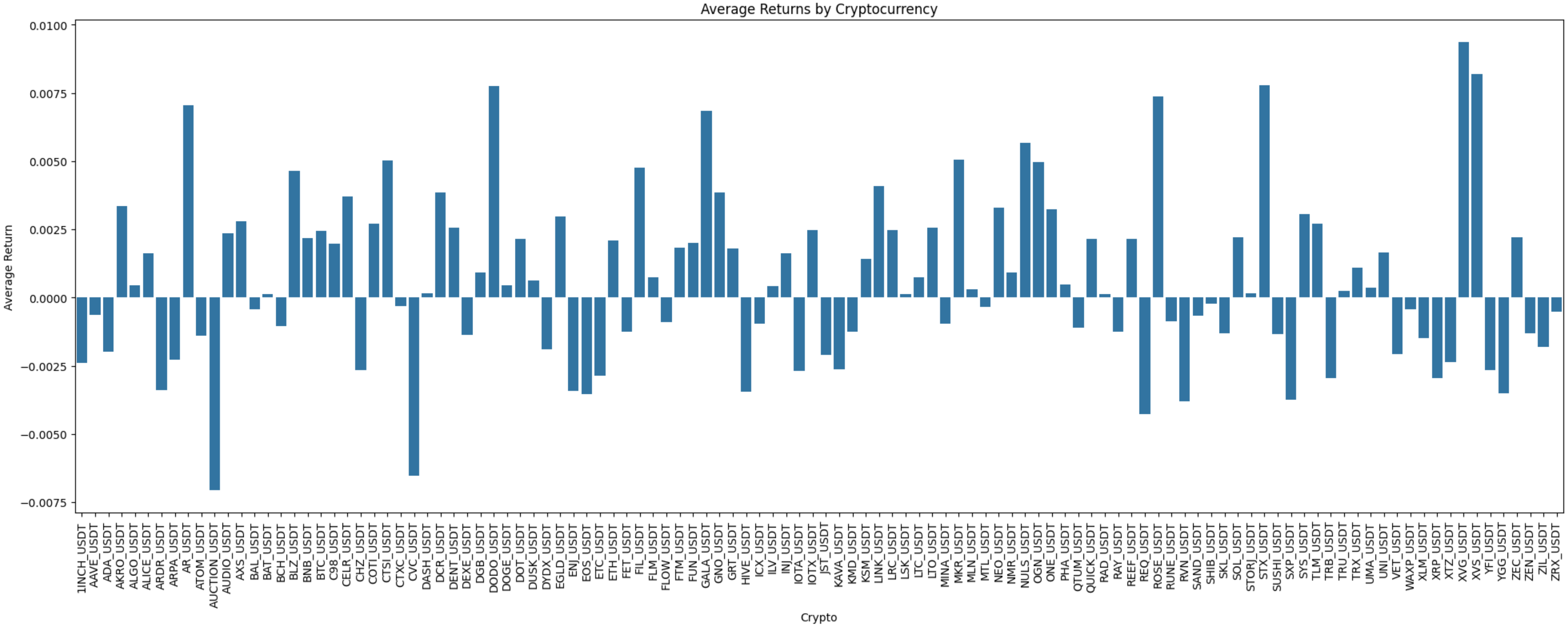
EDA and Analysis – Google Trend



This image shows the average google trend per token. Each token has a corresponding bar indicating its average Google trend score. Based on the image, here are the top five tokens with the highest average Google trend **from 2021 till now** :

1-MTL_USDT 2- BAL_USDT 3-CVC_USDT 4-PAD_USDT 5- CTXC_USDT

EDA and Analysis – Returns



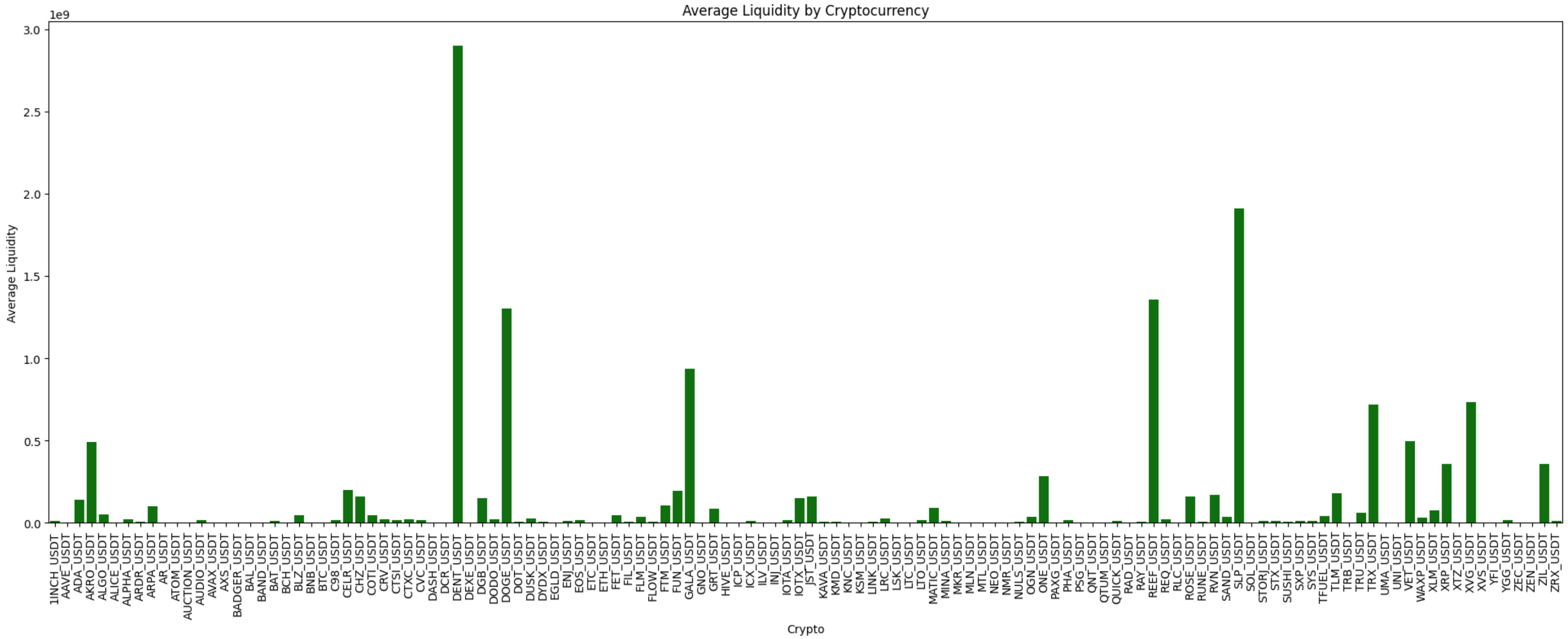
This image shows the average return per token. Based on the image, here are the top 10 tokens with the highest average return scores **from 2021 till now** :

:

1.GALA_USDT , 2-XVG_USDT , 3-XVS_USDT, 4-FIL_USDT , 5- DOGE_USDT

The variation in returns among the tokens indicates differing levels of volatility and market performance. Investors might consider this when making decisions, balancing between high-return, high-risk tokens and more stable options.

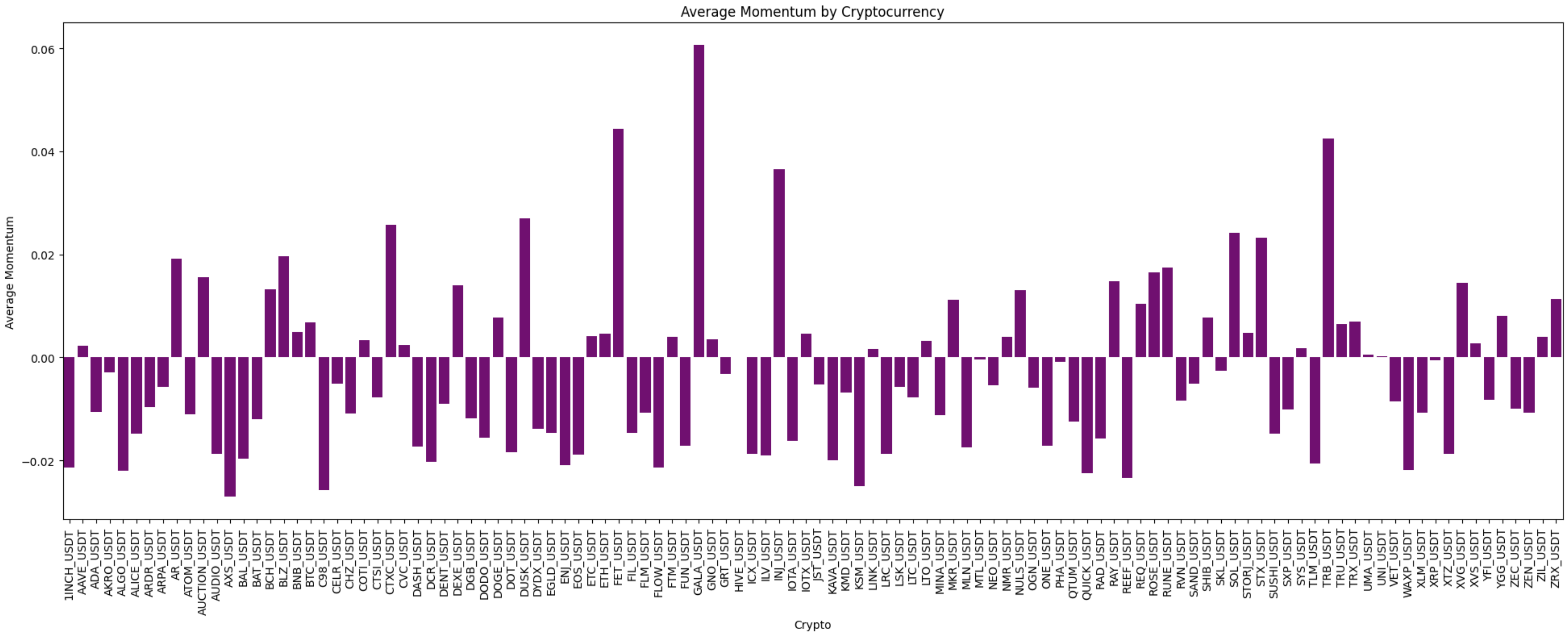
EDA and Analysis – Returns



This image shows the average Liquidity per token. Based on the image, here are the top 10 tokens with the highest average Liquidity scores **from 2021 till now** :

1.DENT_USDT , 2-SLP_USDT , 3-REEO_USDT, 4-DOGE_USDT , 5-GALA_USDT

EDA and Analysis – Momentum



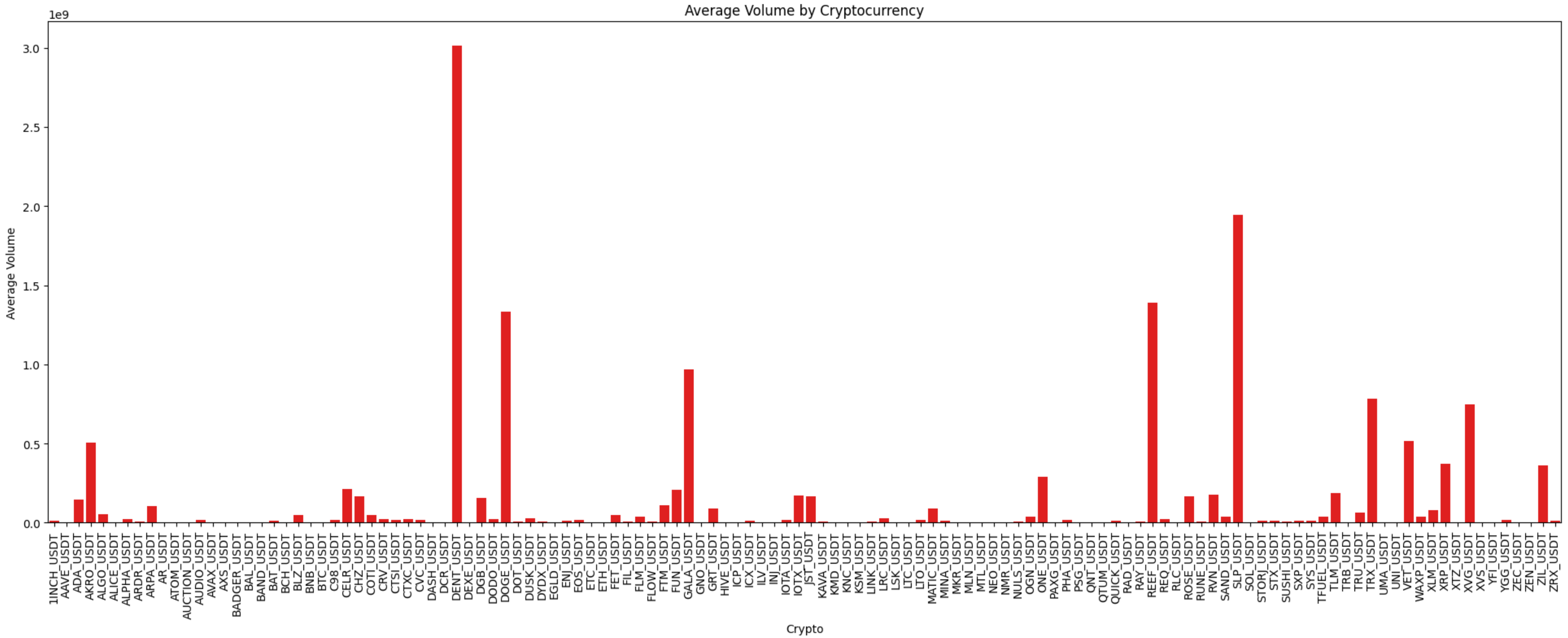
This image shows the average Momentum per token. Based on the image, here are the top five tokens with the highest average Momentum scores **from 2021 till now** :

1.GALA_USDT 2-AUCTION_USDT 3-FLOW_USDT 4-MINA_USDT 5-MKR_USDT

•**Positive Momentum**: The tokens listed above have the highest positive momentum, indicating they are currently experiencing upward trends in their market performance.

•**Negative Momentum:** Some tokens, such as BTC_USDT and ETH_USDT, show negative momentum, suggesting they are currently in a downward trend.

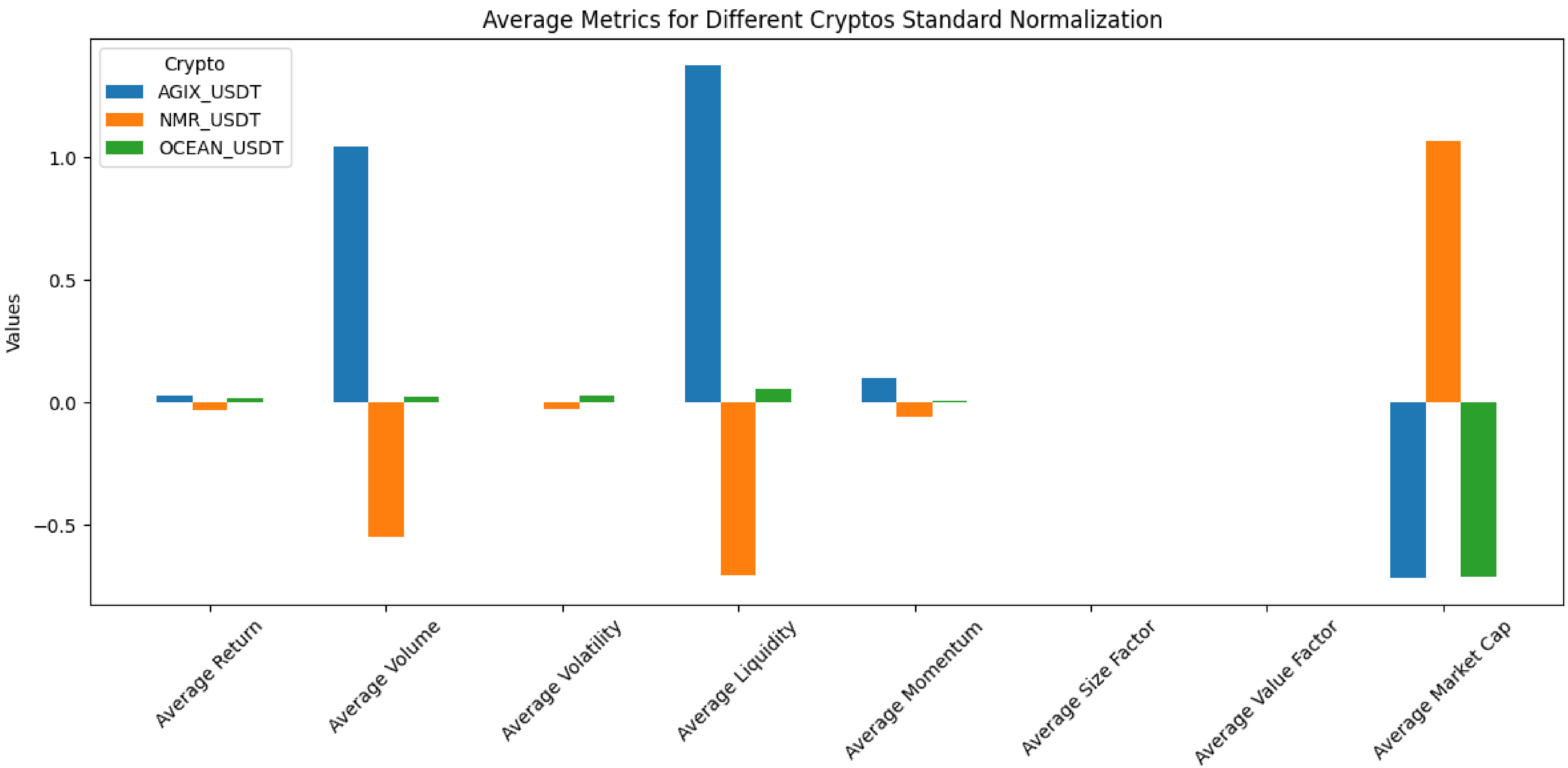
EDA and Analysis – Volume



This image shows the average Momentum per token. Based on the image, here are the top five tokens with the highest average Volume scores **from 2021 till now** :

1-DENT_USDT 2- SLP_USDT 3-REEF_USDT 4-DOGE_USDT 5- GALA_USDT

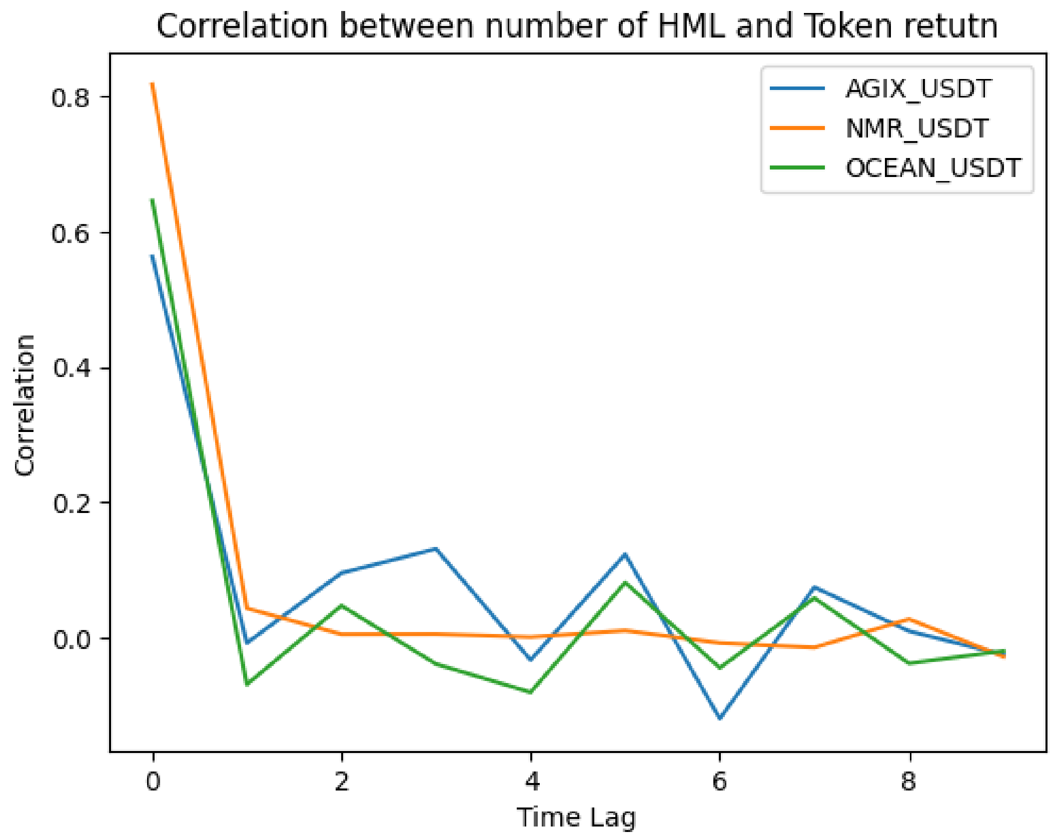
EDA and Analysis – Ocean , Numerai, Singuarity

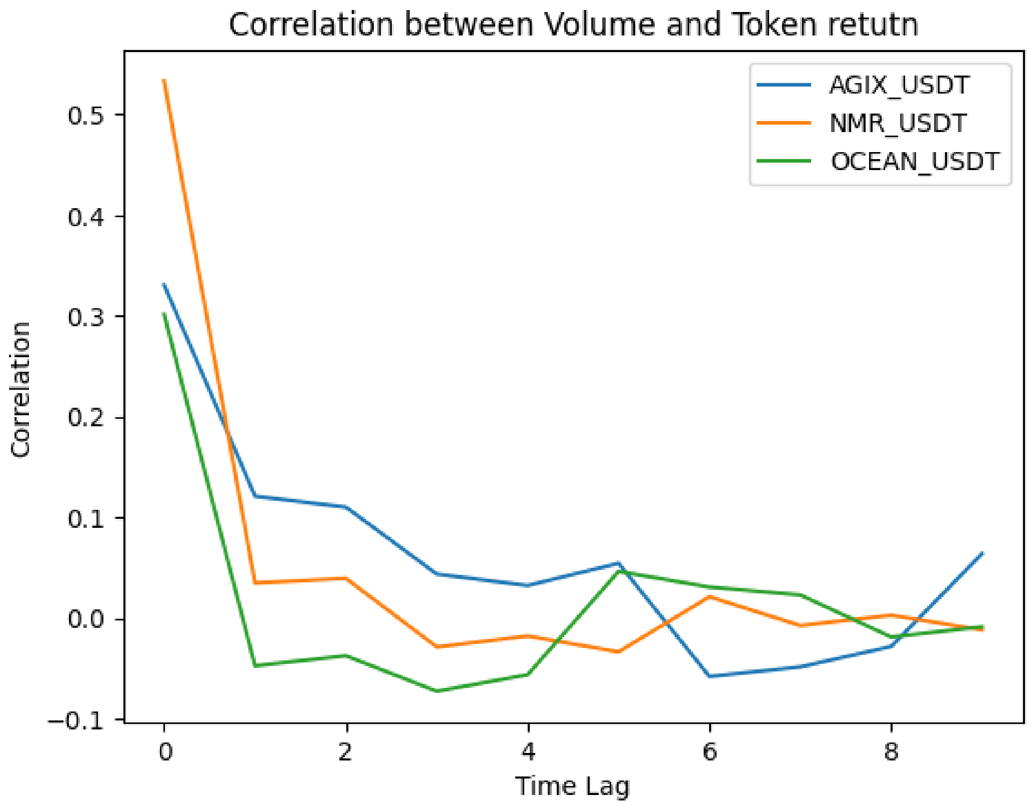


Maximum Time Lag– Ocean , Numerai, Singuarity

```
Maximum correlation of HML and AGIX_USDT_return: 0.562994064569284
Corresponding lag: 0
Maximum correlation of HML and NMR_USDT_return: 0.817105708990881
Corresponding lag: 0
Maximum correlation of HML and OCEAN_USDT_return: 0.6455779842630929
Corresponding lag: 0
```

```
Maximum correlation of Volume and AGIX_USDT_return: 0.33071695892931535
Corresponding lag: 0
Maximum correlation of Volume and NMR_USDT_return: 0.5333360050899352
Corresponding lag: 0
Maximum correlation of Volume and OCEAN_USDT_return: 0.301541655171176
Corresponding lag: 0
```

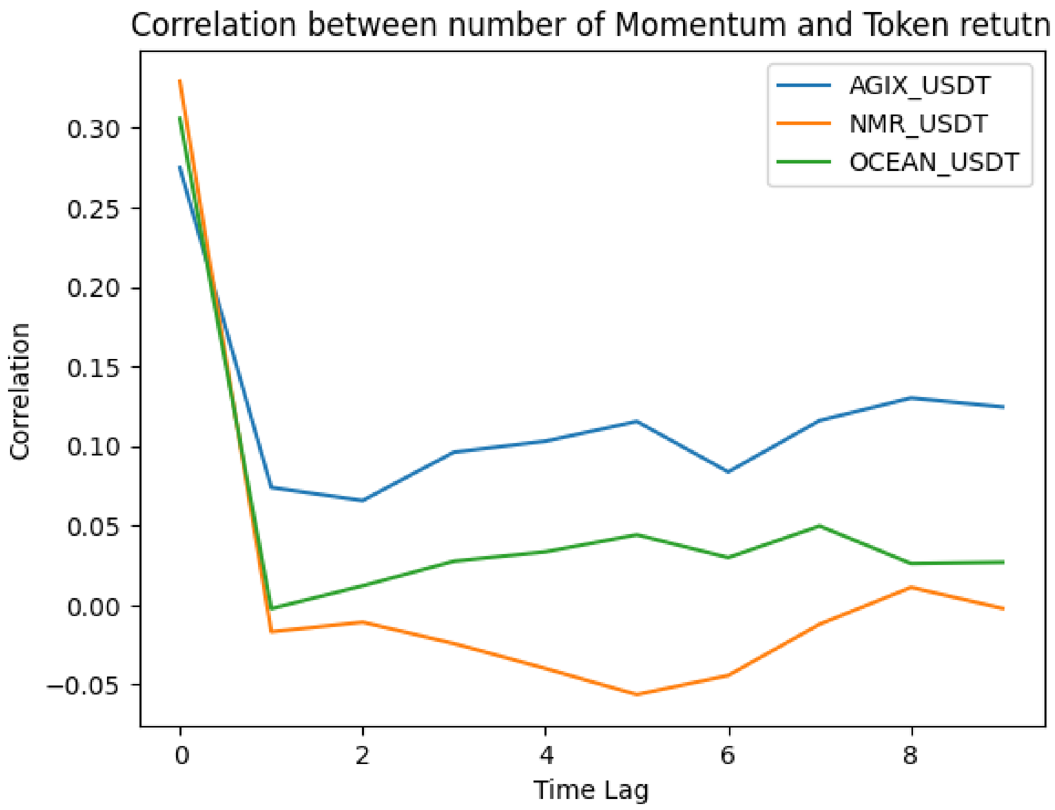


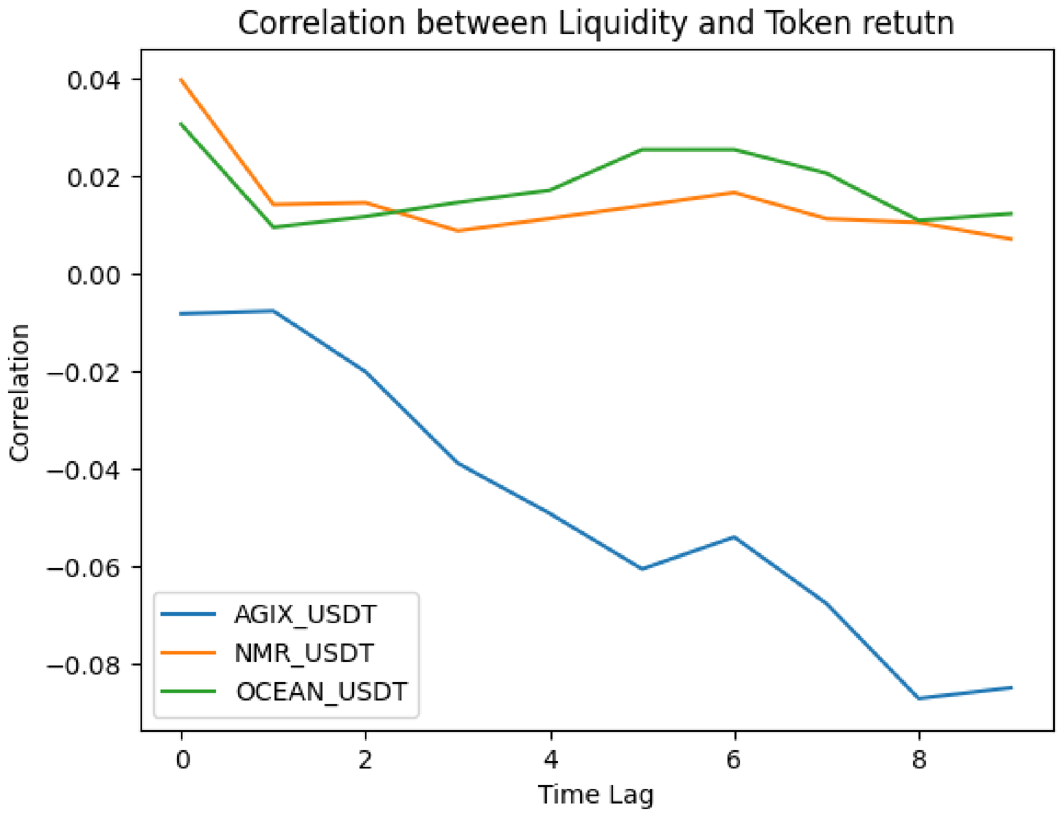


Maximum Time Lag– Ocean , Numerai, Singuarity

```
Maximum correlation of Momentum and AGIX_USDT_return: 0.27494360359945824
Corresponding lag: 0
Maximum correlation of Momentum and NMR_USDT_return: 0.3292125870575445
Corresponding lag: 0
Maximum correlation of Momentum and OCEAN_USDT_return: 0.30580604027862623
Corresponding lag: 0
```

```
Maximum correlation of Liquidity and AGIX_USDT_return: -0.0076343283959940415
Corresponding lag: 1
Maximum correlation of Liquidity and NMR_USDT_return: 0.03975225566457364
Corresponding lag: 0
Maximum correlation of Liquidity and OCEAN_USDT_return: 0.030673710143464223
Corresponding lag: 0
```

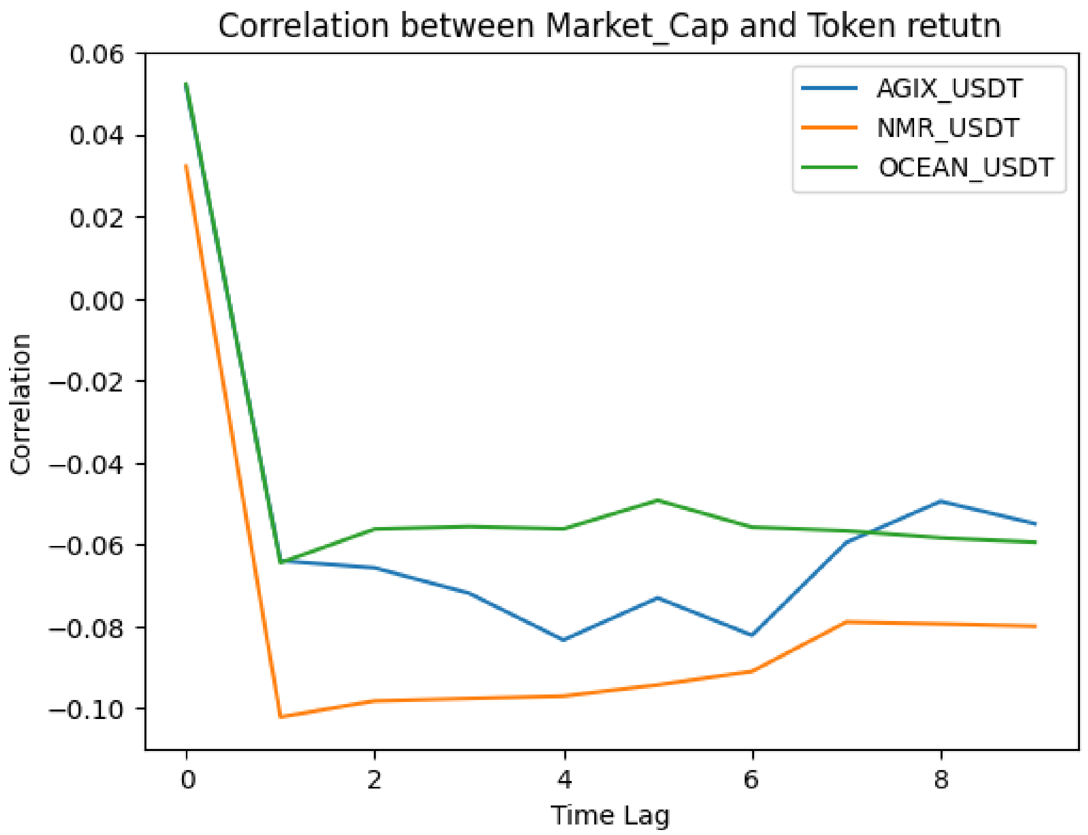


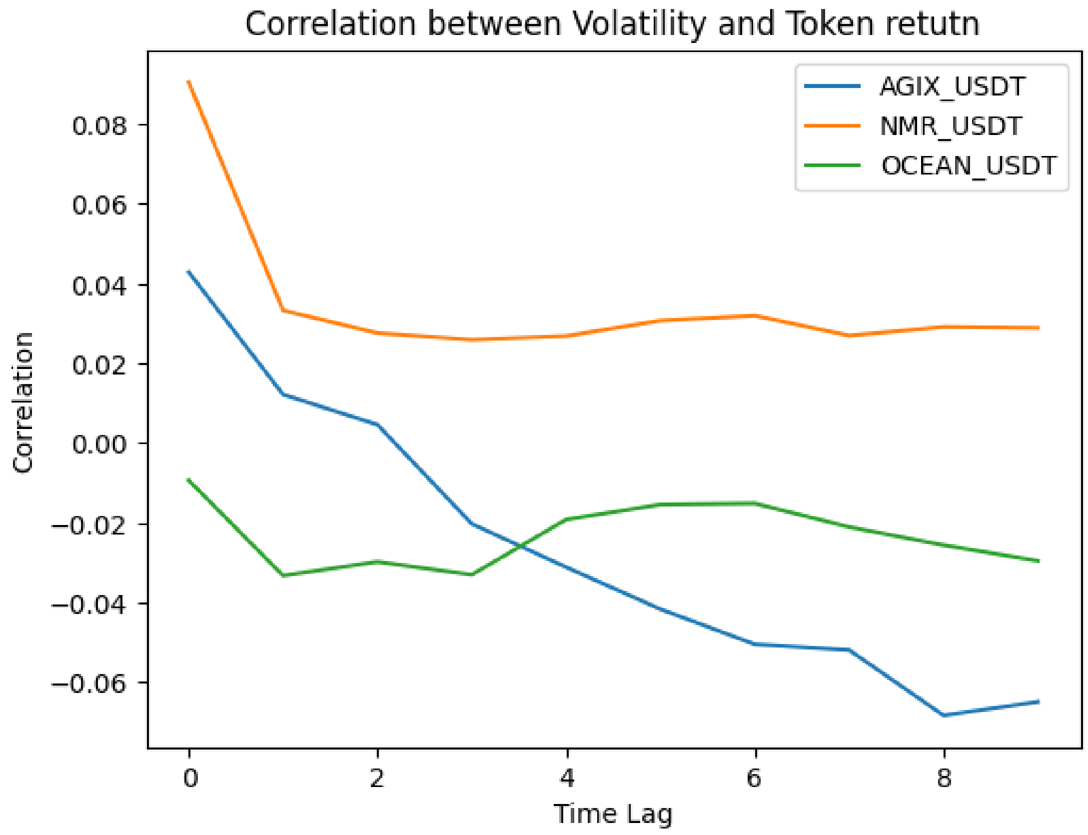


Maximum Time Lag– Ocean , Numerai, Singuarity

```
Maximum correlation of Market_Cap and AGIX_USDT_return: 0.05168432545816474
Corresponding lag: 0
Maximum correlation of Market_Cap and NMR_USDT_return: 0.03229220584946276
Corresponding lag: 0
Maximum correlation of Market_Cap and OCEAN_USDT_return: 0.05231033681913843
Corresponding lag: 0
```

```
Maximum correlation of Volatility and AGIX_USDT_return: 0.04289886172596687
Corresponding lag: 0
Maximum correlation of Volatility and NMR_USDT_return: 0.09051155574484925
Corresponding lag: 0
Maximum correlation of Volatility and OCEAN_USDT_return: -0.009314658148842
Corresponding lag: 0
```





Modelling –
with economic factor

The OLS regression analysis have been performed on a combined dataset of all 128 cryptos. The datasets in this model containing economic factors without google trend since their join was empty without overlap in time. This model explores what factors influence crypto returns. Based on R-squared metric, the model can explain approximately 67.3% of the variability in returns which indicated that these factors couldn’t well predict the return.

Significant Factors:
HML (High Minus Low): This factor has a strong positive impact on returns and is highly significant.
Momentum: This factor is marginally significant, suggesting it may slightly influence returns.

Insignificant Factors: Other variables, such as Market Cap, Volatility, Liquidity, and others, do not show a significant effect on returns.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 return   R-squared:                       0.542
Model:                            OLS   Adj. R-squared:                  0.542
Method:                 Least Squares   F-statistic:                     8196.
Date:                Mon, 26 Aug 2024   Prob (F-statistic):               0.00
Time:                        10:46:01   Log-Likelihood:             1.6224e+05
No. Observations:               89965   AIC:                        -3.244e+05
Df Residuals:                   89951   BIC:                        -3.243e+05
Df Model:                          13
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                   0.0089      0.001     11.503      0.000       0.007       0.010
Market_Cap          -1.929e-15   1.31e-15     -1.469      0.142    -4.5e-15    6.46e-16
Momentum                0.0252      0.001     38.534      0.000       0.024       0.026
Volatility             -0.0549      0.005    -10.802      0.000      -0.065      -0.045
Liquidity           -4.957e-16    3.3e-16     -1.503      0.133   -1.14e-15    1.51e-16
amihud_illiquidity  -3.125e-07   2.73e-08    -11.462      0.000   -3.66e-07   -2.59e-07
Bid Price              -0.0002      0.001     -0.314      0.754      -0.001       0.001
Ask Price               0.0002      0.001      0.314      0.754      -0.001       0.001
Bid Depth           -6.267e-10   5.07e-10     -1.236      0.216   -1.62e-09    3.67e-10
Ask Depth            2.177e-09   1.76e-09      1.237      0.216   -1.27e-09    5.63e-09
Bid-Ask Spread       1.073e-05   1.49e-06      7.203      0.000    7.81e-06    1.37e-05
T10YIE                 -0.0034      0.000    -13.033      0.000      -0.004      -0.003
DFF                     0.0002   7.02e-05      2.210      0.027    1.75e-05       0.000
size_factor            -0.0004      0.000     -1.727      0.084      -0.001    5.34e-05
NVT                 -7.405e-12   7.78e-12     -0.952      0.341   -2.27e-11    7.85e-12
HML                     0.8201      0.003    302.960      0.000       0.815       0.825
==============================================================================
Omnibus:                    28243.230   Durbin-Watson:                   2.127
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1549179.073
Skew:                           0.717   Prob(JB):                         0.00
Kurtosis:                      23.279   Cond. No.                     1.58e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.58e+17. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Modelling – google trend

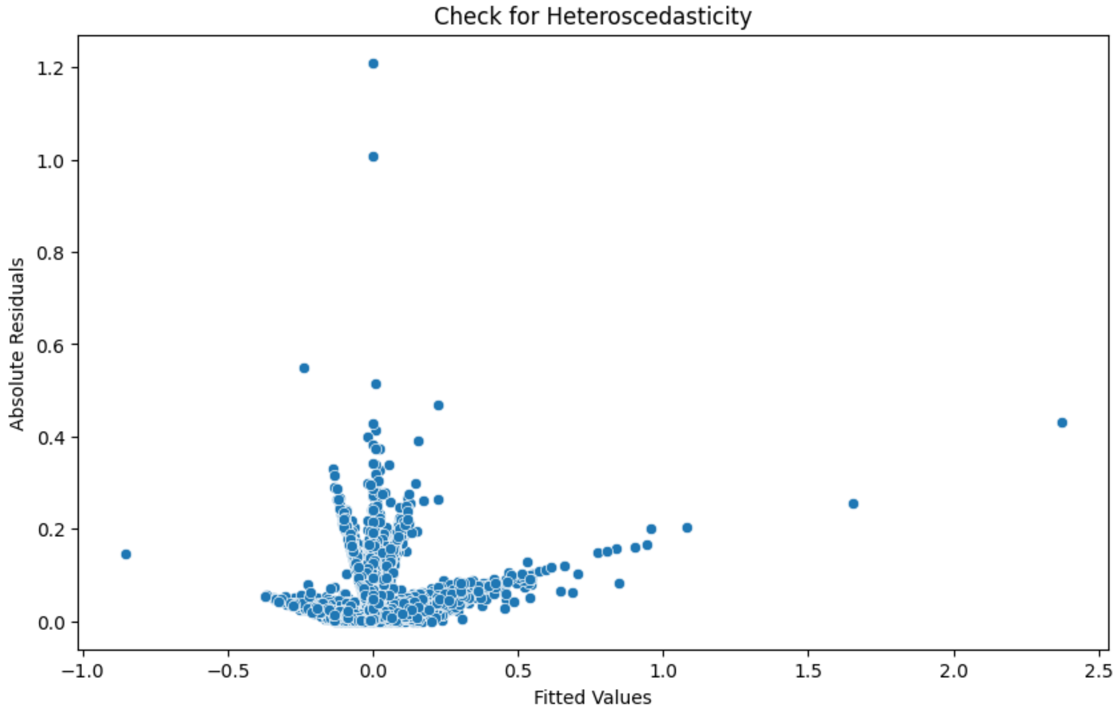
The OLS regression analysis have been performed on a combined dataset of all 128 cryptos. The datasets in this model containing google trends without econo factors . The model shows that about 31.7% of the changes in returns can be explained by the model's factors. Key findings include that higher momentum and HML (high minus low) values are linked to increased returns, while higher volatility tends to decrease returns. The model is statistically significant overall, but some variables, like bid price and google trends, do not significantly affect returns.

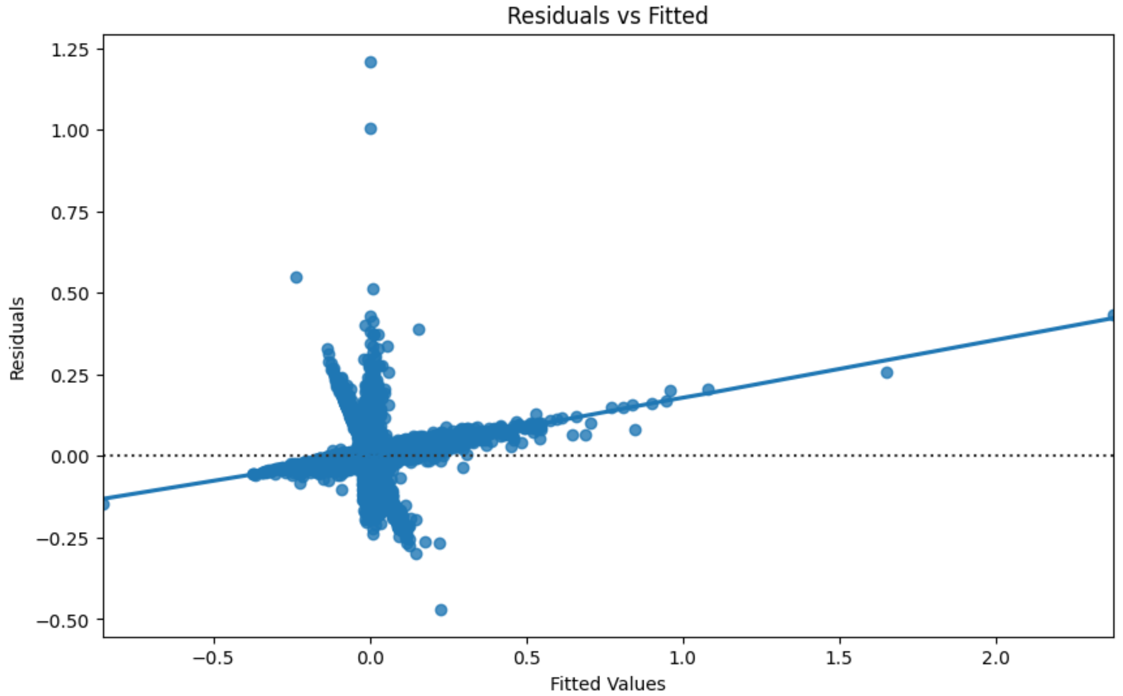
OLS Regression Results

Dep. Variable:	return	R-squared:	0.317
Model:	OLS	Adj. R-squared:	0.317
Method:	Least Squares	F-statistic:	620.2
Date:	Tue, 27 Aug 2024	Prob (F-statistic):	0.00
Time:	13:55:04	Log-Likelihood:	28628.
No. Observations:	16016	AIC:	-5.723e+04
Df Residuals:	16003	BIC:	-5.713e+04
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0016	0.002	-1.037	0.300	-0.005	0.001
Market_Cap	8.304e-15	2.96e-15	2.806	0.005	2.5e-15	1.41e-14
Momentum	0.0295	0.001	19.816	0.000	0.027	0.032
Volatility	-0.0382	0.012	-3.261	0.001	-0.061	-0.015
Liquidity	1.378e-15	7.47e-16	1.845	0.065	-8.59e-17	2.84e-15
amihud_illiquidity	-1.045e-07	8.3e-08	-1.259	0.208	-2.67e-07	5.82e-08
Bid Price	-6.427e-05	0.002	-0.042	0.967	-0.003	0.003
Ask Price	6.426e-05	0.002	0.042	0.967	-0.003	0.003
Bid Depth	-4.769e-10	1.16e-09	-0.410	0.682	-2.76e-09	1.8e-09
Ask Depth	1.642e-09	4.04e-09	0.406	0.684	-6.28e-09	9.56e-09
Bid-Ask Spread	-1.488e-06	1.03e-05	-0.145	0.885	-2.16e-05	1.86e-05
google_trend	1.516e-05	1.13e-05	1.338	0.181	-7.05e-06	3.74e-05
size_factor	0.0002	0.001	0.153	0.878	-0.002	0.003
NVT	-4.736e-11	1.21e-10	-0.392	0.695	-2.84e-10	1.9e-10
HML	0.6212	0.008	80.414	0.000	0.606	0.636

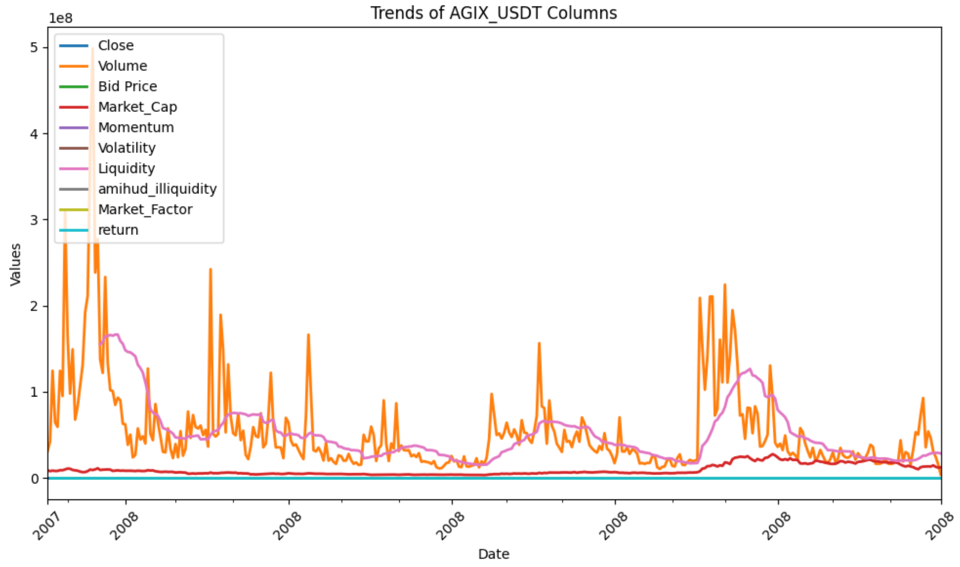
Plot of OLS model on combined dataset





The plot shows that most of the residuals (the differences between the predicted and actual values) are clustered around the zero line, which means the model is doing a good job of capturing the average behavior of the data. However, there are some large differences, or outliers, both above and below this line, indicating that the model may not be perfect and could have issues like varying spread of residuals (heteroscedasticity) or incorrect assumptions about the data. Additionally, there's a slight upward trend in the residuals as the predicted values increase, which might suggest that the relationship between the variables is not simply linear, or that some influential data points are skewing the results.

Modelling -Singularity



the model explains about 33.7% of the changes in returns, which means it captures some, but not all, of the factors affecting AGIX's performance.

HML (High Minus Low): This factor is very significant and positively impacts returns, suggesting that it plays an important role in AGIX's price movements.
Momentum: This variable is approaching significance and may have a slight positive effect on returns.
Other factors like Market Cap, Volatility, Liquidity, and several others do not significantly influence returns based on their high p-values.

```
AGIX_USDT
                            OLS Regression Results
==============================================================================
Dep. Variable:                 return   R-squared:                       0.337
Model:                            OLS   Adj. R-squared:                  0.322
Method:                 Least Squares   F-statistic:                     22.13
Date:                Tue, 27 Aug 2024   Prob (F-statistic):           3.04e-27
Time:                        11:04:50   Log-Likelihood:                 556.66
No. Observations:                 357   AIC:                            -1095.
Df Residuals:                     348   BIC:                            -1060.
Df Model:                           8
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Market_Cap           7.62e-10   8.67e-10      0.879      0.380   -9.44e-10    2.47e-09
Momentum               0.0217      0.012      1.872      0.062      -0.001       0.045
Volatility            -0.2376      0.362     -0.657      0.512      -0.949       0.474
Liquidity          -2.899e-11    2.2e-10     -0.132      0.895   -4.62e-10    4.04e-10
amihud_illiquidity -2.707e-09   4.14e-09     -0.654      0.514   -1.09e-08    5.44e-09
size_factor           -0.0308      0.094     -0.328      0.743      -0.215       0.154
NVT                   -0.0183      0.013     -1.413      0.158      -0.044       0.007
HML                    0.6449      0.058     11.194      0.000       0.532       0.758
DFF                    0.0093      0.018      0.504      0.615      -0.027       0.046
T10YIE                -0.0034      0.006     -0.566      0.572      -0.015       0.009
==============================================================================
Omnibus:                       77.052   Durbin-Watson:                   2.033
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              600.982
Skew:                           0.643   Prob(JB):                    3.15e-131
Kurtosis:                       9.225   Cond. No.                     2.79e+17
```

Modelling -Singularity

Using residual vs fitted we can analyze the data and model from different aspects

Linearity:

The residuals appear to be randomly scattered around the horizontal line (y=0), which is a good sign and suggests that the linearity assumption is reasonable for the model.

1.Homoscedasticity:

There is some spread in the residuals as the fitted values increase, indicating potential heteroscedasticity. Ideally, the spread of residuals should be constant across all levels of fitted values.

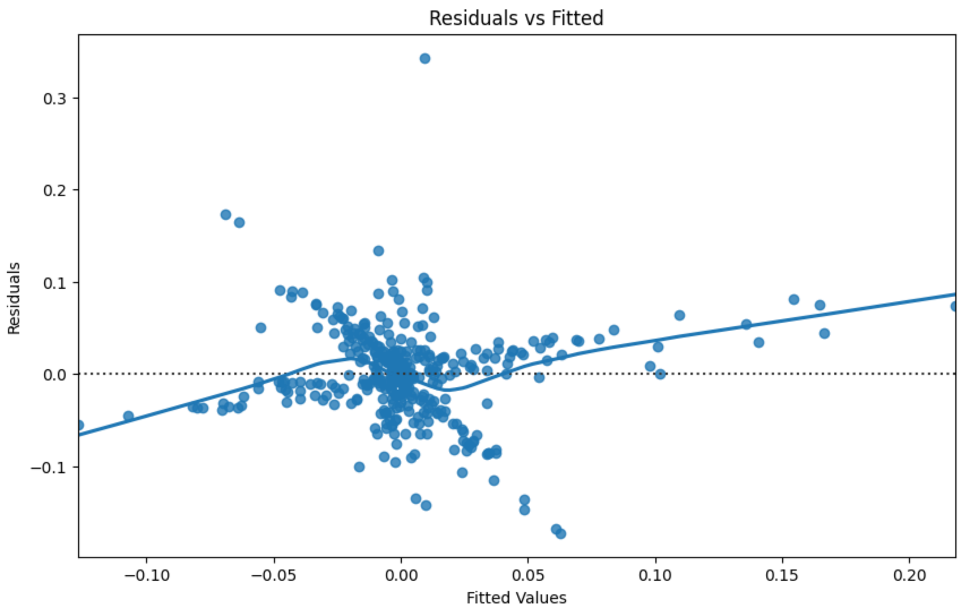
2.Outliers:

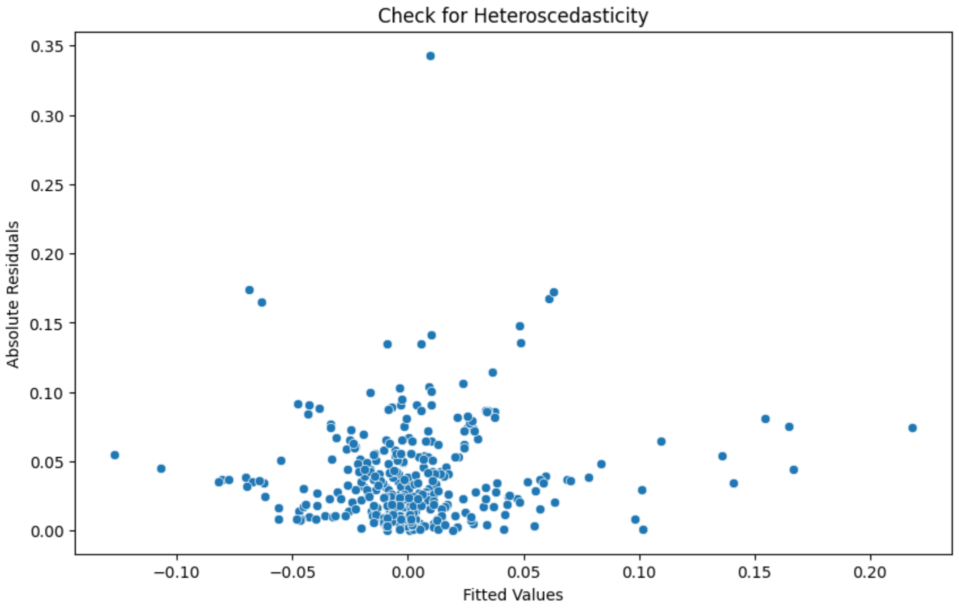
A few residuals are notably far from the rest, particularly on the higher side. These could be outliers or influential points that might affect the model's performance.

3.Normality:

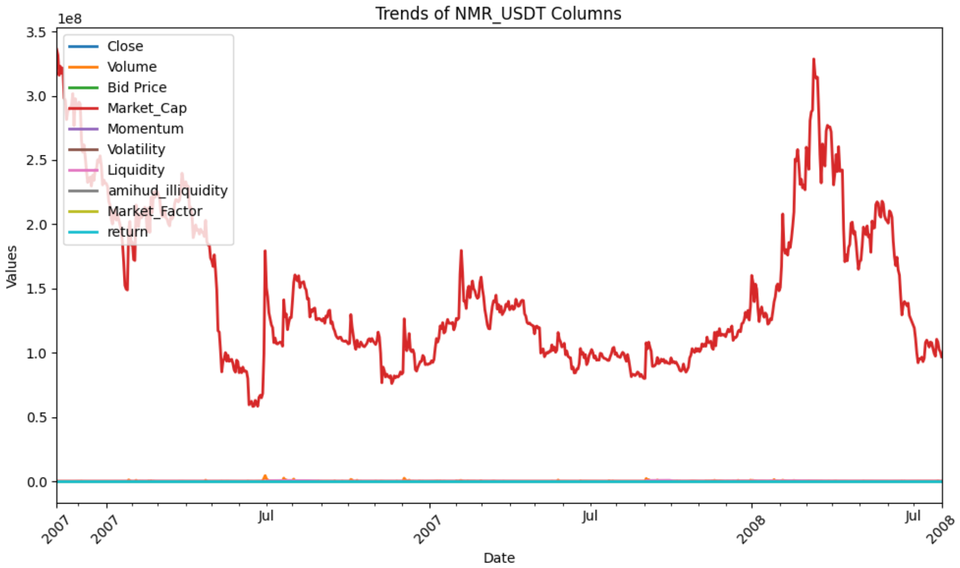
While not directly assessable from this plot, the symmetry of the residuals around the zero line suggests that normality might hold, but this should be confirmed with other plots like a Q-Q plot.

Also Heteroscedasticity plot suggests a relatively constant variance, indicating no strong evidence of heteroscedasticity, which is positive for model assumptions.





Modelling - numerai



NMR_USDT

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 return   R-squared:                       0.673
Model:                            OLS   Adj. R-squared:                  0.669
Method:                 Least Squares   F-statistic:                     160.9
Date:                Tue, 27 Aug 2024   Prob (F-statistic):          2.89e-164
Time:                        11:04:51   Log-Likelihood:                 1337.4
No. Observations:                 714   AIC:                            -2655.
Df Residuals:                     704   BIC:                            -2609.
Df Model:                           9
Covariance Type:            nonrobust
========================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Market_Cap          3.057e-11   2.92e-11      1.048      0.295   -2.67e-11    8.78e-11
Momentum               0.0146      0.008      1.862      0.063      -0.001       0.030
Volatility            -0.0397      0.074     -0.537      0.591      -0.185       0.105
Liquidity           1.051e-08   1.98e-08      0.530      0.596   -2.84e-08    4.94e-08
amihud_illiquidity  5615.7628   4003.021      1.403      0.161   -2243.525    1.35e+04
size_factor           -0.0071      0.011     -0.634      0.527      -0.029       0.015
NVT                -7.168e-08   3.09e-07     -0.232      0.817   -6.79e-07    5.36e-07
HML                    0.8693      0.025     34.392      0.000       0.820       0.919
DFF                    0.0014      0.001      1.366      0.172      -0.001       0.003
T10YIE                -0.0041      0.003     -1.450      0.147      -0.010       0.001
==============================================================================
Omnibus:                       89.144   Durbin-Watson:                   2.150
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              813.216
Skew:                           0.005   Prob(JB):                    2.58e-177
```

The model explains about 67.3% of the variability in NMR returns, indicating a good fit.

Significant Factors:

HML (High Minus Low): This factor has a strong positive impact on returns and is highly significant.

Momentum: This factor is marginally significant, suggesting it may slightly influence returns.

Insignificant Factors: Other variables, such as Market Cap, Volatility, Liquidity, and others, do not show a significant effect on returns.

Modelling - numerai

The "Residuals vs Fitted" plot provides insights into the performance and assumptions of the regression model. Here's a brief analysis:

1.Centering around Zero:

Residuals are mostly centered around zero, which is ideal as it indicates that the model predictions are unbiased.

2.Spread of Residuals:

There is a noticeable spread as fitted values increase, suggesting potential heteroscedasticity. This means the variance of errors may change with different levels of predicted values.

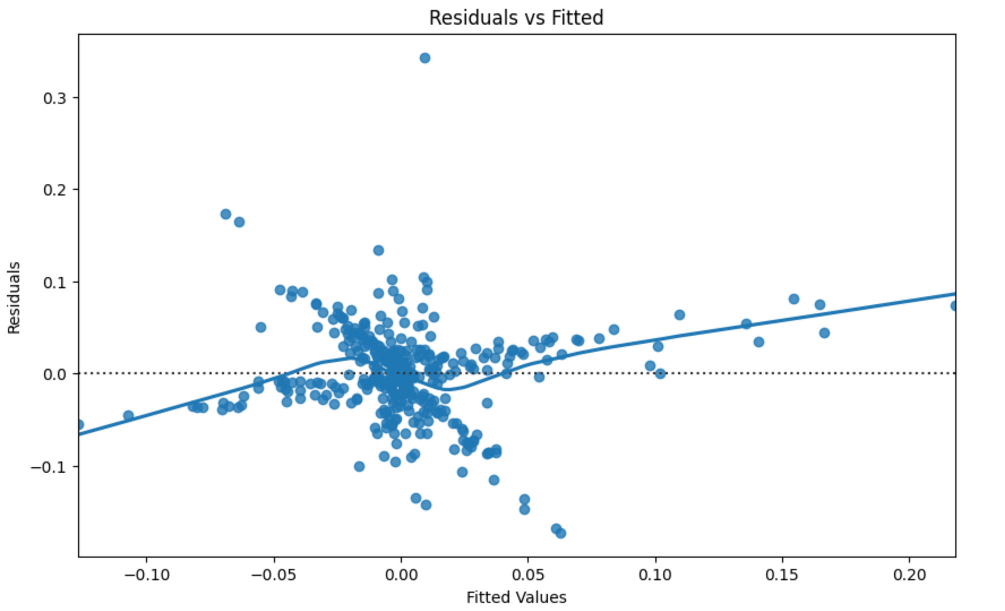
3.Non-linear Pattern:

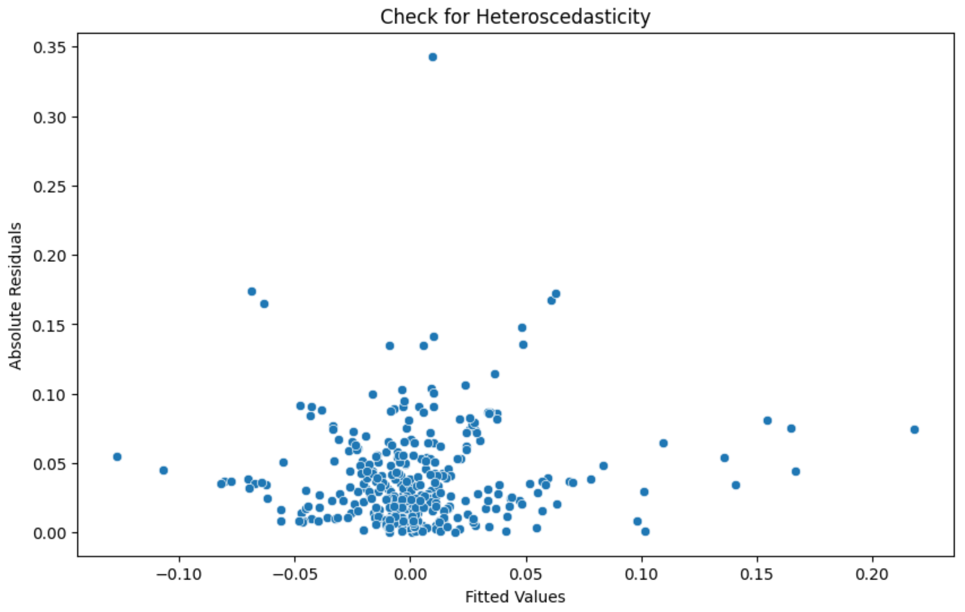
The plot shows a slight curve rather than a random scatter, indicating potential non-linear relationships not captured by the model.

4.Outliers:

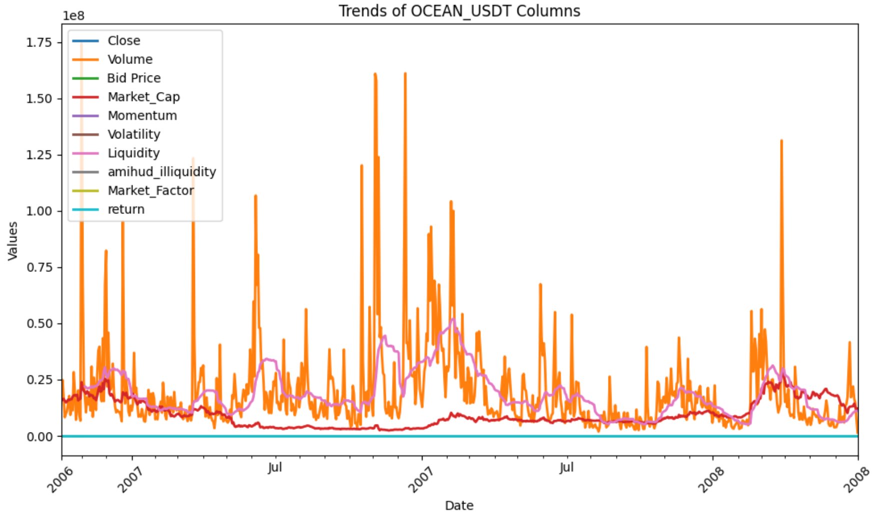
A few points are far from the zero line, suggesting the presence of outliers or influential observations.

The Heteroscedasticity plot suggests that there is no strong evidence of heteroscedasticity, as the spread of residuals appears fairly constant across the range of fitted values.





Modelling – ocean protocol

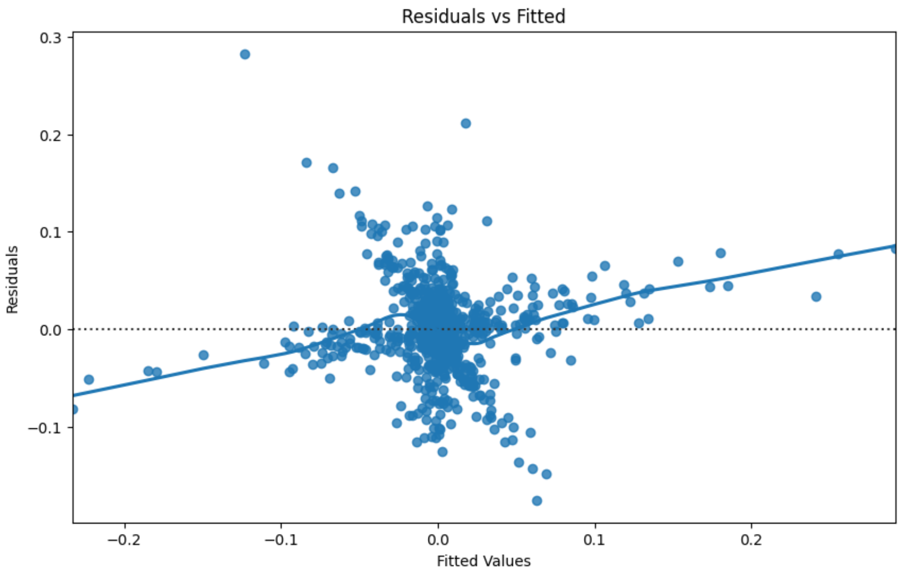


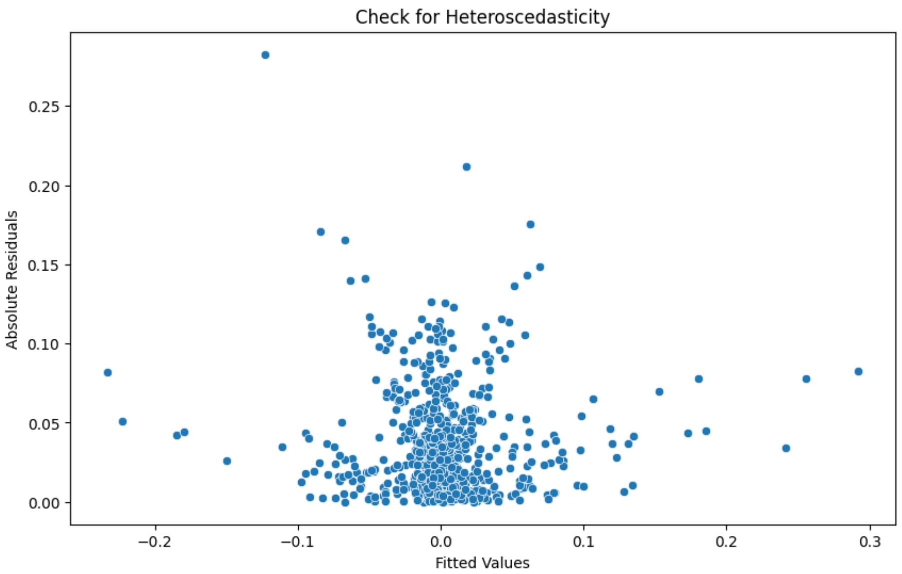
The OLS regression results on Ocean Protocol (OCEAN_USDT) show an R-squared value of 0.443, indicating that approximately 44.3% of the variability in returns is explained by the model. The adjusted R-squared is slightly lower at 0.437, suggesting a reasonable fit but leaving room for other factors or complexities not captured by this model.
Key factors include **momentum**, with a significant positive coefficient (0.0517) and **a p-value** of 0.000, indicating a strong relationship with returns. The **HML factor** also shows a significant positive effect (0.7148) with a p-value of 0.000. Other factors like market cap, volatility, liquidity, and others have high p-values, suggesting they are not statistically significant in this model.

OCEAN_USDT

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 return   R-squared:                       0.443
Model:                            OLS   Adj. R-squared:                  0.437
Method:                 Least Squares   F-statistic:                     70.11
Date:                Tue, 27 Aug 2024   Prob (F-statistic):           1.60e-84
Time:                        11:04:52   Log-Likelihood:                 1166.0
No. Observations:                 714   AIC:                            -2314.
Df Residuals:                     705   BIC:                            -2273.
Df Model:                           8
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Market_Cap          -3.15e-10   4.26e-10     -0.740      0.459   -1.15e-09    5.21e-10
Momentum               0.0517      0.010      5.128      0.000       0.032       0.072
Volatility            -0.0085      0.118     -0.072      0.943      -0.240       0.223
Liquidity          -2.504e-10   2.64e-10     -0.949      0.343   -7.69e-10    2.68e-10
amihud_illiquidity  -1.04e-09   6.16e-09     -0.169      0.866   -1.31e-08    1.11e-08
size_factor            0.0187      0.011      1.673      0.095      -0.003       0.041
NVT                    0.0014      0.003      0.483      0.630      -0.004       0.007
HML                    0.7148      0.034     20.895      0.000       0.648       0.782
DFF                   -0.0006      0.001     -0.620      0.535      -0.002       0.001
T10YIE                -0.0047      0.004     -1.326      0.185      -0.012       0.002
==============================================================================
Omnibus:                       74.436   Durbin-Watson:                   2.269
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              286.941
Skew:                           0.412   Prob(JB):                     4.92e-63
```

Modelling – ocean protocol





The two plots for Ocean Protocol show a slight funnel shape, which suggests that the spread of the errors (residuals) increases as the predicted values go up. Ideally, the errors should be randomly scattered around the horizontal line with no clear pattern, indicating that the model fits well.

However, since the errors spread out at different levels of predicted values, it suggests that the model might not fully capture all the important patterns in the data. This means there is still room to improve the model or include more factors to make it more accurate.

conclusion

The report provides a comprehensive analysis of various factors influencing cryptocurrency returns, emphasizing the importance of market and economic indicators. The significant correlation between HML and momentum with returns suggests that these factors should be prioritized in investment strategies.

However, the limitations in predictive power of certain economic indicators highlight the complexity of the crypto market, necessitating further research and model refinement.

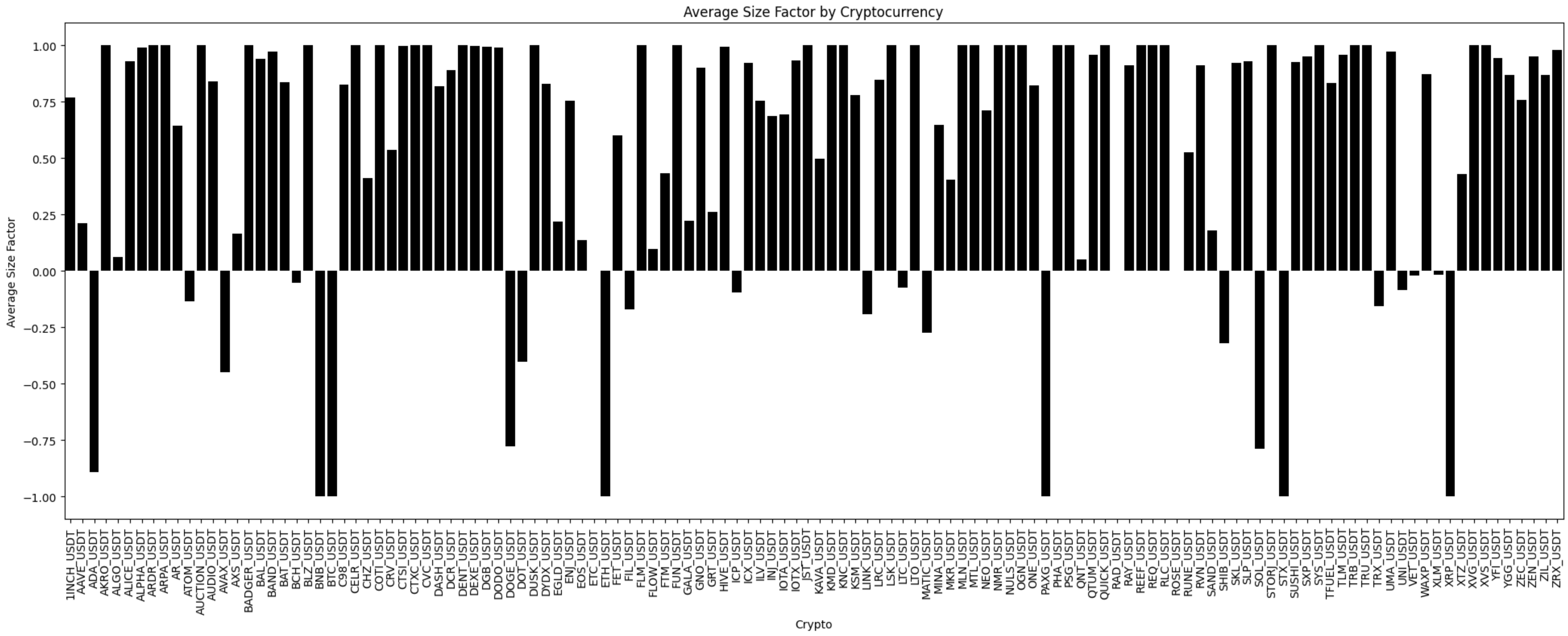
Overall, the insights gained from this study can assist investors and analysts in navigating the volatile cryptocurrency landscape. By leveraging the identified factors and understanding their interactions, stakeholders can enhance their investment strategies and improve portfolio management decisions.

Future work

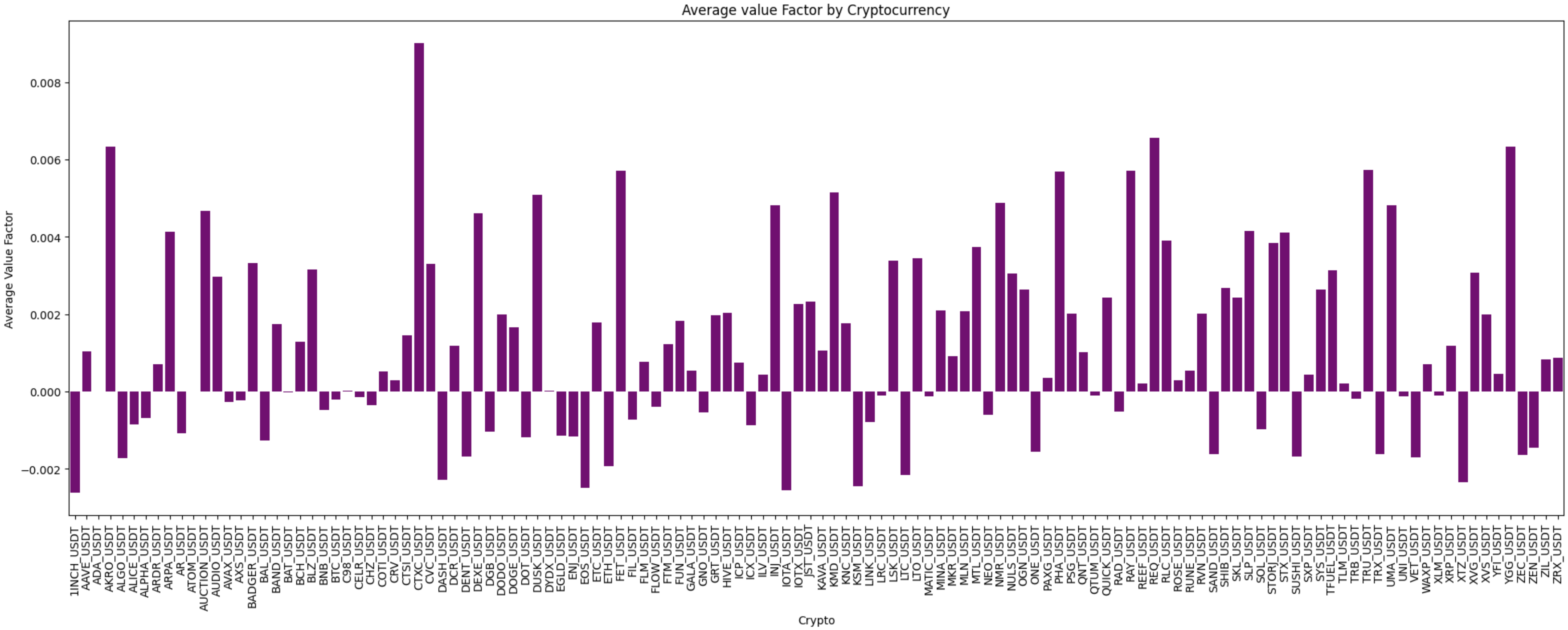
- Given our access to a limited dataset spanning only three years, the analysis lacks comprehensiveness and cannot address every aspect of the subject matter. Expanding the dataset to include a broader time frame would enhance the depth of the investigation

- Factors such as sentiment data and social media activity can significantly influence the returns of individual tokens. By evaluating a wider array of factors, we can better understand their collective impact on returns, leading to more informed insights into cryptocurrency performance.

APPENDIX


Average Size Factor by Cryptocurrency
Average Size Factor
Crypto
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
1INCH_USDT AAVE_USDT ADA_USDT AKRO_USDT ALGO_USDT ALICE_USDT ALPHA_USDT ARDR_USDT ARPA_USDT AR_USDT ATOM_USDT AUCTION_USDT AUDIO_USDT AVAX_USDT AXS_USDT BADGER_USDT BAL_USDT BAND_USDT BAT_USDT BCH_USDT BLZ_USDT BNB_USDT BTC_USDT C98_USDT CELR_USDT CHZ_USDT COTI_USDT CRV_USDT CTSI_USDT CTXC_USDT CVC_USDT DASH_USDT DCR_USDT DENT_USDT DEXE_USDT DGB_USDT DODO_USDT DOGE_USDT DOT_USDT DUSK_USDT DYDX_USDT EGLD_USDT ENJ_USDT EOS_USDT ETC_USDT ETH_USDT FET_USDT FIL_USDT FLM_USDT FLOW_USDT FTM_USDT FUN_USDT GALA_USDT GNO_USDT GRT_USDT HIVE_USDT ICP_USDT ICX_USDT ILV_USDT INJ_USDT IOTA_USDT IOTX_USDT JST_USDT KAVA_USDT KMD_USDT KNC_USDT KSM_USDT LINK_USDT LRC_USDT LSK_USDT LTC_USDT LTO_USDT MATIC_USDT MINA_USDT MKR_USDT MLN_USDT MTL_USDT NEO_USDT NMR_USDT NULS_USDT OGN_USDT ONE_USDT PAXG_USDT PHA_USDT PSG_USDT QNT_USDT QTUM_USDT QUICK_USDT RAD_USDT RAY_USDT REEF_USDT REQ_USDT RLC_USDT ROSE_USDT RUNE_USDT RVN_USDT SAND_USDT SHIB_USDT SKL_USDT SLP_USDT SOL_USDT STORJ_USDT STX_USDT SUSHI_USDT SXP_USDT SYS_USDT TFUEL_USDT TLM_USDT TRB_USDT TRU_USDT TRX_USDT UMA_USDT UNI_USDT VET_USDT WAXP_USDT XLM_USDT XRP_USDT XTZ_USDT XVG_USDT XVS_USDT YFI_USDT YGG_USDT ZEC_USDT ZEN_USDT ZIL_USDT ZRX_USDT

APPENDIX


Average value Factor by Cryptocurrency
Average Value Factor
0.008
0.006
0.004
0.002
0.000
−0.002
Crypto
1INCH_USDT
AAVE_USDT
ADA_USDT
AKRO_USDT
ALGO_USDT
ALICE_USDT
ALPHA_USDT
ARDR_USDT
ARPA_USDT
AR_USDT
ATOM_USDT
AUCTION_USDT
AUDIO_USDT
AVAX_USDT
AXS_USDT
BADGER_USDT
BAL_USDT
BAND_USDT
BAT_USDT
BCH_USDT
BLZ_USDT
BNB_USDT
BTC_USDT
C98_USDT
CELR_USDT
CHZ_USDT
COTI_USDT
CRV_USDT
CTSI_USDT
CTXC_USDT
CVC_USDT
DASH_USDT
DCR_USDT
DENT_USDT
DEXE_USDT
DGB_USDT
DODO_USDT
DOGE_USDT
DOT_USDT
DUSK_USDT
DYDX_USDT
EGLD_USDT
ENJ_USDT
EOS_USDT
ETC_USDT
ETH_USDT
FET_USDT
FIL_USDT
FLM_USDT
FLOW_USDT
FTM_USDT
FUN_USDT
GALA_USDT
GNO_USDT
GRT_USDT
HIVE_USDT
ICP_USDT
ICX_USDT
ILV_USDT
INJ_USDT
IOTA_USDT
IOTX_USDT
JST_USDT
KAVA_USDT
KMD_USDT
KNC_USDT
KSM_USDT
LINK_USDT
LRC_USDT
LSK_USDT
LTC_USDT
LTO_USDT
MATIC_USDT
MINA_USDT
MKR_USDT
MLN_USDT
MTL_USDT
NEO_USDT
NMR_USDT
NULS_USDT
OGN_USDT
ONE_USDT
PAXG_USDT
PHA_USDT
PSG_USDT
QNT_USDT
QTUM_USDT
QUICK_USDT
RAD_USDT
RAY_USDT
REEF_USDT
REQ_USDT
RLC_USDT
ROSE_USDT
RUNE_USDT
RVN_USDT
SAND_USDT
SHIB_USDT
SKL_USDT
SLP_USDT
SOL_USDT
STORJ_USDT
STX_USDT
SUSHI_USDT
SXP_USDT
SYS_USDT
TFUEL_USDT
TLM_USDT
TRB_USDT
TRU_USDT
TRX_USDT
UMA_USDT
UNI_USDT
VET_USDT
WAXP_USDT
XLM_USDT
XRP_USDT
XTZ_USDT
XVG_USDT
XVS_USDT
YFI_USDT
YGG_USDT
ZEC_USDT
ZEN_USDT
ZIL_USDT
ZRX_USDT

Thank you for reading :)