Alpha Platform

Nguyen Duc Khanh Trinh

November 9, 2025

Contents

1 Introduction

The Alpha Platform is designed to support the identification, evaluation, and implementation of investment strategies in a systematic manner. The platform's architecture, as illustrated in Figure 1, provides a comprehensive workflow that includes data collection, signal generation, portfolio construction, and performance evaluation.

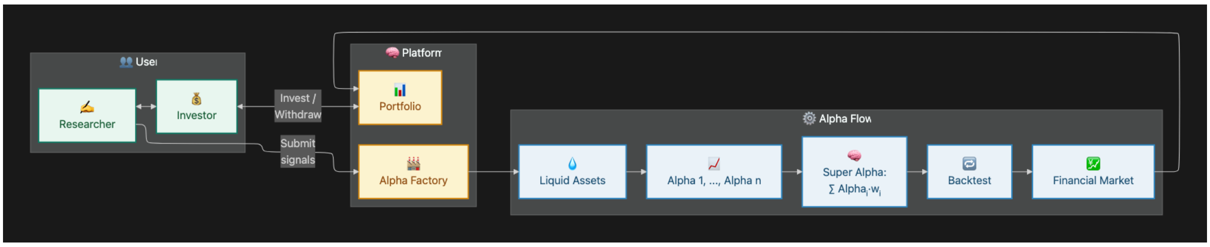

Figure 1: Alpha Platform.

2 Alpha Factory

The Alpha Factory is a structured process for generating and implementing investment signals, as shown in Figure 2. This process ensures that investment decisions are based on rigorous quantitative analysis and align with the platform's objectives of maximizing returns while minimizing risk.

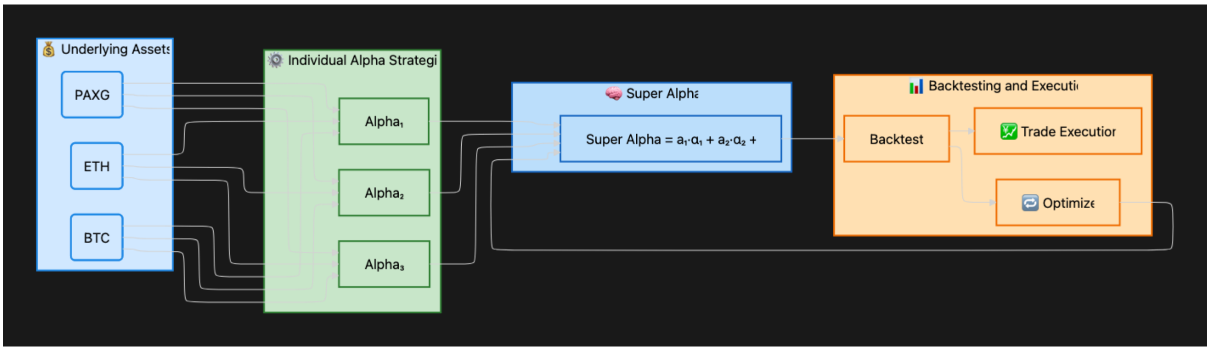

Figure 2: Overview of Alpha Factory.

2.1 Liquidity Asset Selection

The first stage of the Alpha Factory involves the careful selection of liquid assets, specifically coins, for portfolio construction. The selection criteria prioritize assets with high liquidity to minimize liquidity risk and reduce transaction costs, such as bid-ask spreads. Highly liquid assets are typically associated with strong market participation, ensuring stability and reliability in portfolio construction. The selection process uses quantitative metrics, such as trading volume and market capitalization, to identify assets that meet predefined liquidity thresholds.

2.2 Regular Alpha

2.2.1 Alpha Space

At the core of the Alpha Factory is the generation of Regular Alphas, which represent individual investment perspectives or hypotheses about asset performance. These alpha signals are developed in an open and flexible environment, where diverse investment ideas can be explored independently. Each alpha represents a unique perspective, and the collection of alphas forms a multidimensional idea space, analogous to an infinite-dimensional vector space, as illustrated in Figure 3.

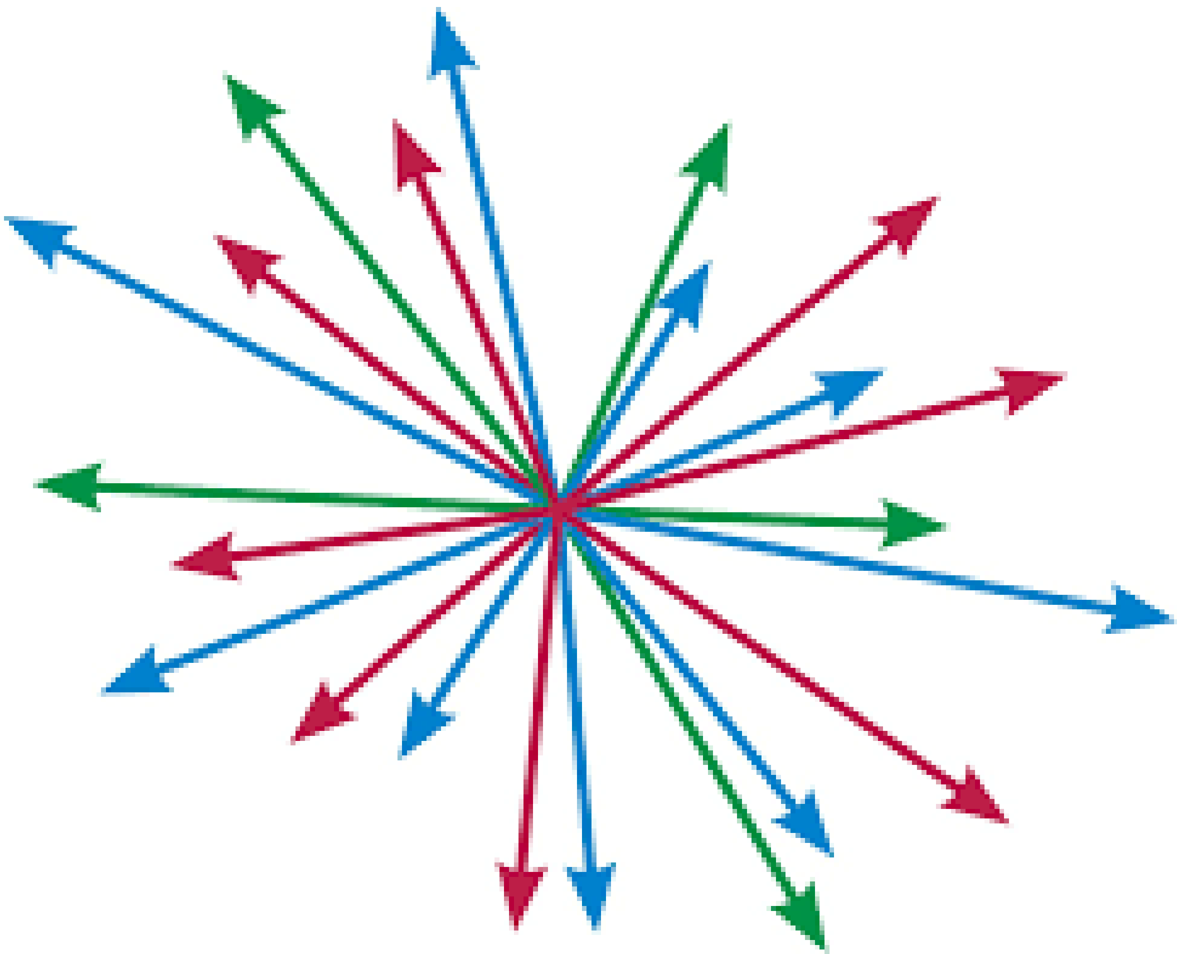

Figure 3: Conceptual Illustration of the Multidimensional Alpha Idea Space.

An alpha represents the weight allocated to each asset in the portfolio based on predictive trading signals. To ensure that the portfolio adheres to the fully invested constraint, alpha signals are normalized so that the sum of the absolute weights equals one. The normalization process is mathematically expressed as follows:

$$\text{weight}_{i,t} = \frac{\alpha_{i,t}}{\sum_{i=0}^{N} |\alpha_{i,t}|}$$

$$\sum_{i=0}^{N} |\text{weight}_{i,t}| = 1$$

where:

- i: Index representing the i-th asset in the portfolio.
- t: Time of the investment decision.
- N: Total number of assets in the portfolio.
- $\alpha_{i,t}$: Alpha signal for asset i at time t.
- $\text{weight}_{i,t}$: Normalized weight of asset i at time t.

This normalization process ensures that the portfolio is fully allocated while maintaining diversification across assets. A visual representation of the alpha signal generation and normalization process is provided in Figure 4.

```
sim = Simulate("crypto", "TOP7", booksize=1, compound=False)
alpha= sim.regular("ts_quantile(close*volume,300)/ts_std(returns,300)")
alpha.weights
```

✓ 0.0s

	BTCUSDT	ETHUSDT	BNBUSDT	XRPUSDT	SOLUSDT	PAXGUSDT	TRXUSDT
TradingDate							
2022-07-22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2022-07-23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2022-07-24	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2022-07-25	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2022-07-26	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
...	...	...	...	...	...	...	...
2025-10-07	0.117440	0.070119	0.119682	0.074227	0.069234	0.394841	0.154455
2025-10-08	0.126682	0.064176	0.118977	0.073804	0.068720	0.392856	0.154786
2025-10-09	0.117275	0.069946	0.119246	0.074209	0.069011	0.394597	0.155715
2025-10-10	0.116391	0.069612	0.116009	0.073599	0.068639	0.399107	0.156644
2025-10-11	0.116315	0.069407	0.122837	0.073269	0.064302	0.397718	0.156152

1178 rows × 7 columns

Figure 4: Illustration of the Alpha Signal Generation and Normalization Process.

2.2.2 Performance of Regular Alpha Signals

After completing the simulation of alpha signals, performance evaluation is a critical step to determine the quality and effectiveness of these signals. Key performance metrics, including returns, Sharpe ratio, maximum drawdown, and portfolio turnover, provide a comprehensive view of the alpha's effectiveness in generating value and managing risk. These metrics not only reflect profitability but also assess the stability and sustainability of the investment strategy.

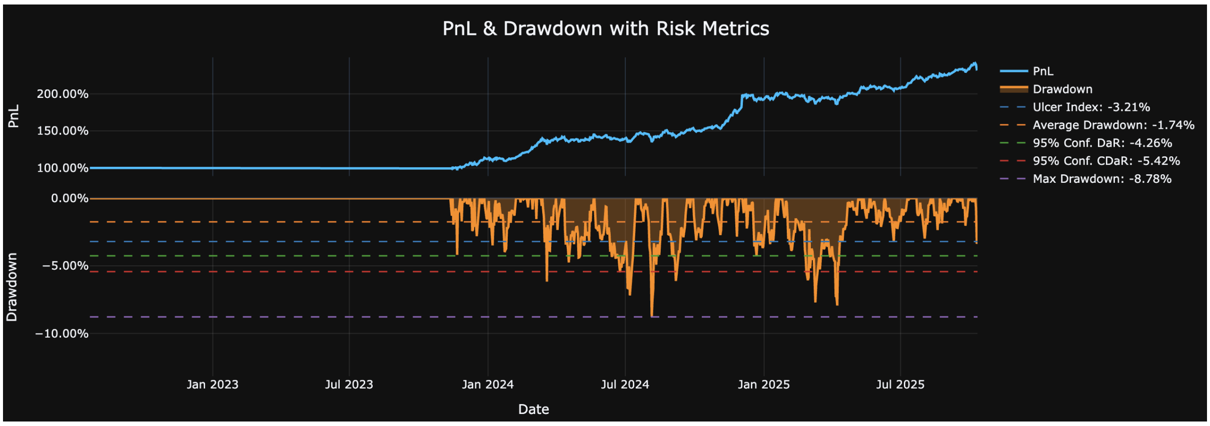

Figure 5: Cumulative profit and loss chart of regular alpha signals.

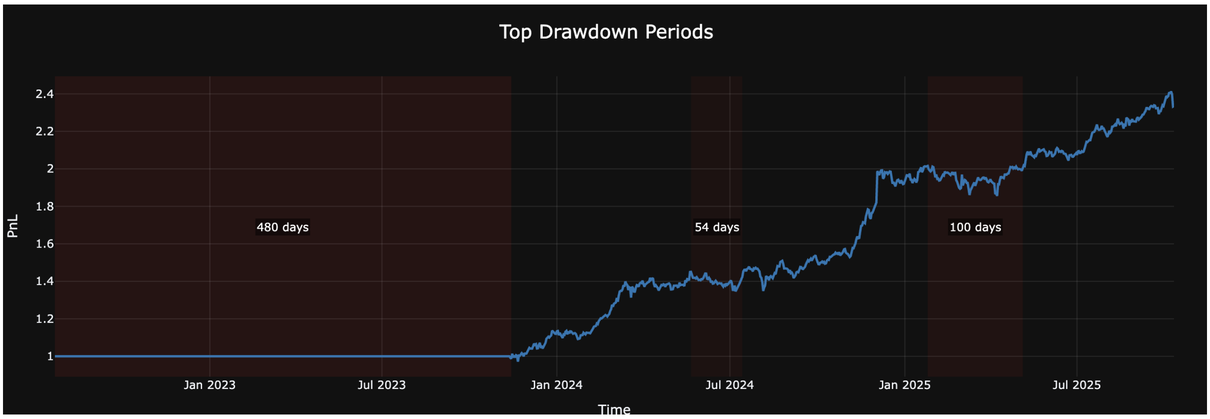

Figure 6: Maximum drawdown over time of the portfolio.

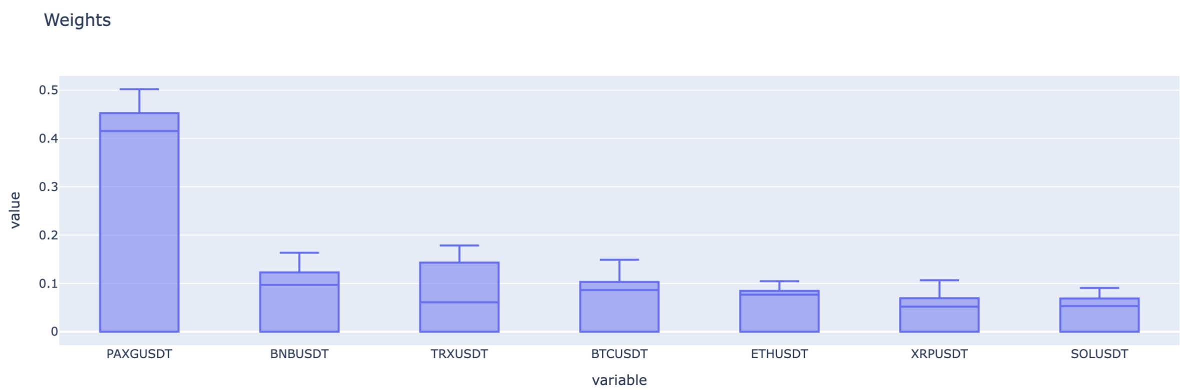

Figure 7: Weight distribution of assets in the portfolio.

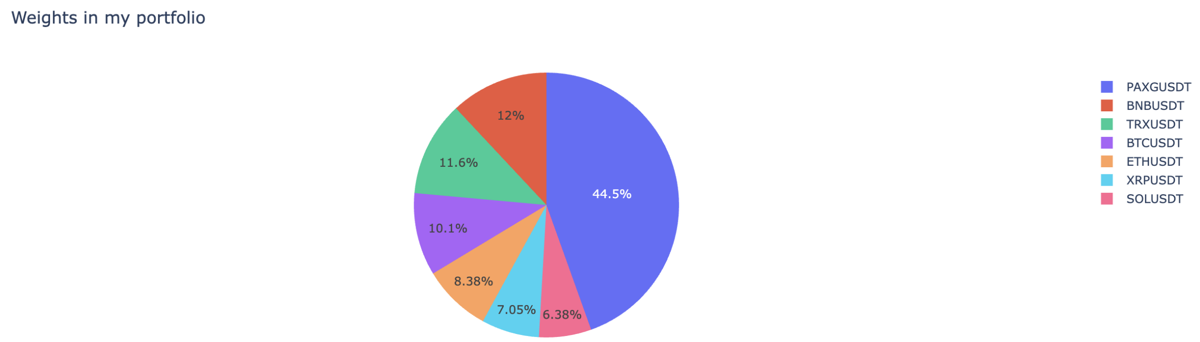

Figure 8: Weight allocation by industry in the portfolio.

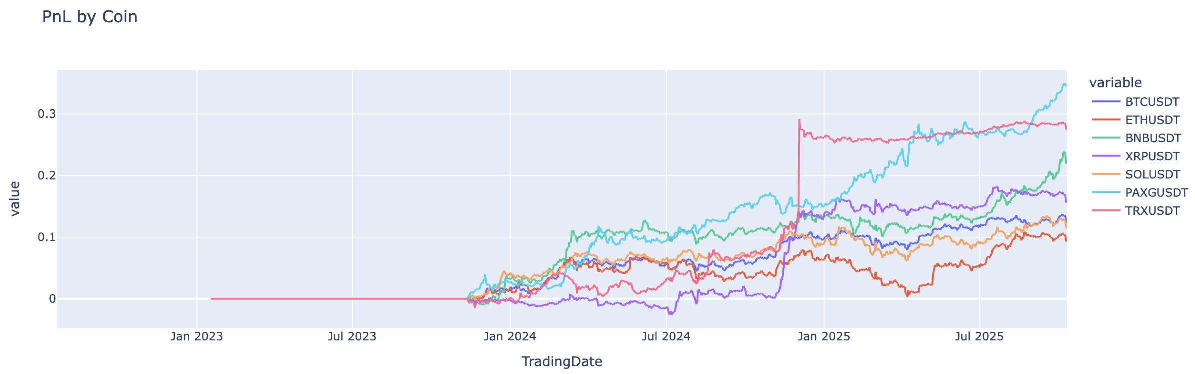

Figure 9: Profit allocation by industry in the portfolio.

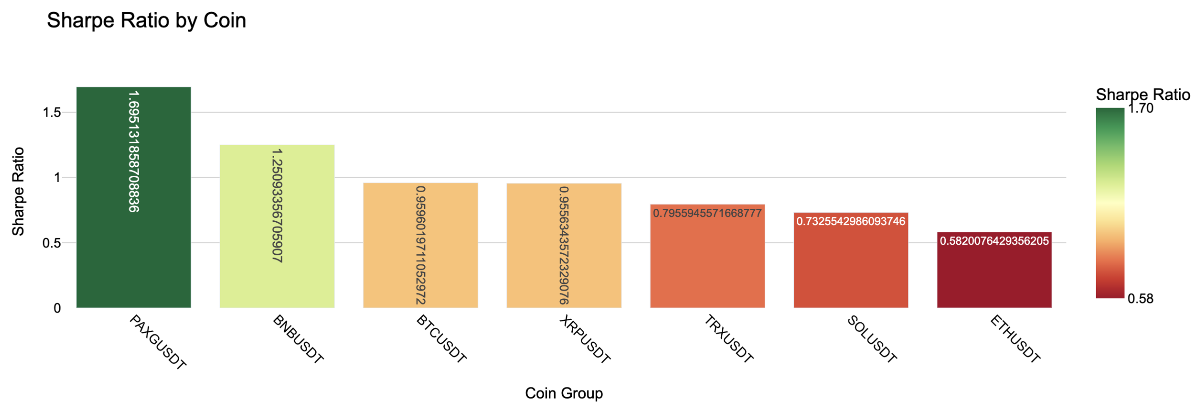

Figure 10: Sharpe ratio of individual coins in the portfolio.

The charts from Figure 5 to Figure 10 provide detailed insights into the performance of alpha signals. Specifically, Figure 5 shows cumulative returns over time, evaluating the overall profitability of the strategy. Figure 6 illustrates maximum drawdown, reflecting the risk level during adverse market conditions. Figures 7 and 8 provide information on the weight allocation across assets and industries, ensuring portfolio diversification. Finally, Figures 9 and 10 analyze profit allocation by industry and the Sharpe ratio of individual coins, assessing the strategy's efficiency in optimizing risk-adjusted returns.

2.3 Super Alphas

Super Alpha signals are linear combinations of Regular Alpha signals, representing a spanned space of regular alphas. The primary goal of Super Alpha is to optimize investment performance by combining Regular Alphas in a way that achieves the highest returns while effectively controlling risk. The construction of Super Alpha involves two main steps: (1) selecting high-quality Regular Alphas based on performance metrics, and (2) determining the method to combine the selected alphas to optimize the portfolio. This process is detailed in Figures 11 and 12.

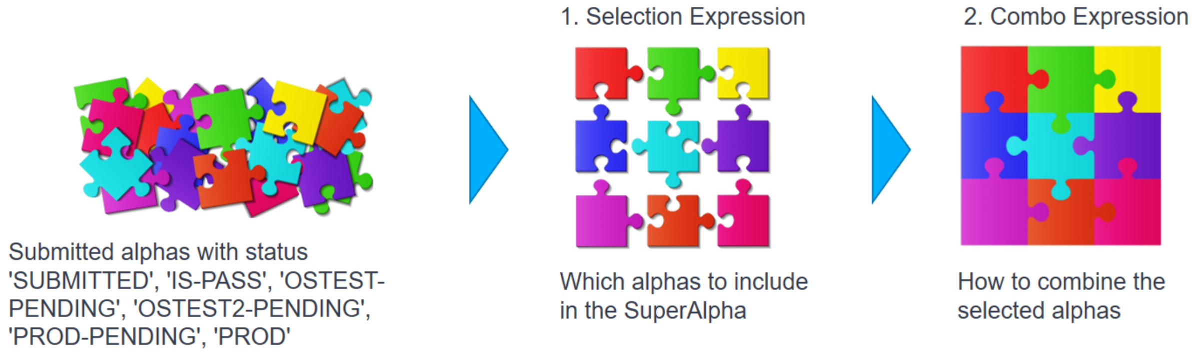

Figure 11: Overview of the Super Alpha construction process.

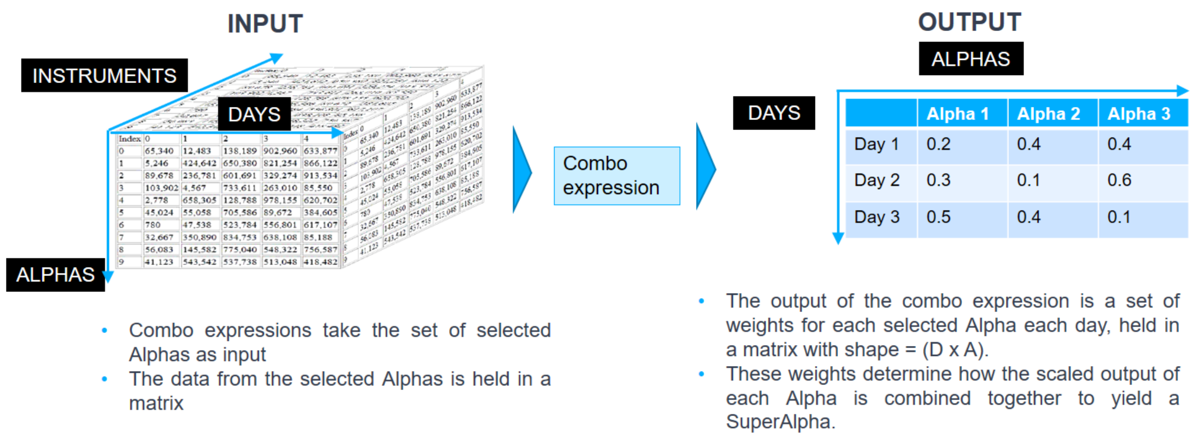

Figure 12: Detailed steps in the Super Alpha construction process.

2.3.1 Alpha Signal Selection

The first step in constructing Super Alpha is selecting Regular Alphas with superior performance based on quantitative metrics such as returns, Sharpe ratio, maximum drawdown, and other indicators. This process aims to filter out alphas with the potential to positively contribute to the overall portfolio. For example, an SQL query for alpha selection might be expressed as follows:

```
(sharpe > 1.58) AND (returns > 0.15)
```

This query selects Regular Alphas with a Sharpe ratio greater than 1.58 and annualized returns exceeding 15%. These criteria ensure that only high-performing and stable alphas are included in the combination phase.

2.3.2 Alpha Signal Combination

The second step is determining the method to combine the selected Regular Alphas to create a Super Alpha. A common approach is to allocate higher weights to alphas with superior performance, such as those based on annualized returns. For example, a method to calculate the weighted sum based on returns might be expressed as follows:

```
ts_sum(returns, 252)
```

This query assigns higher weights to alphas with higher annualized returns, using a time window of 252 trading days (equivalent to one year). This method ensures that high-performing alphas contribute more significantly to the Super Alpha portfolio, optimizing overall returns.

2.3.3 Super Alpha Performance

The performance of Super Alpha is evaluated through metrics such as cumulative returns, Sharpe ratio, and profit allocation over time. The charts from Figure 13 to Figure 15 provide detailed insights into the effectiveness of the Super Alpha strategy.

	return	sharpe	turnover	fitness	margin	drawdown
TradingDate						
2013	0.000000	NaN	0.004000	NaN	0.000000	NaN
2014	0.178507	0.828755	0.589956	0.455873	302.576291	-inf
2015	0.302151	1.287313	0.454871	1.049185	664.256711	-0.489490
2016	0.045129	0.241505	0.495461	0.072886	91.083948	-0.318458
2017	0.532569	3.466950	0.516306	3.521129	1031.498858	-0.228735
2018	0.455111	1.272542	0.499619	1.214538	910.916111	-0.192037
2019	0.127720	0.980491	0.510605	0.490377	250.133896	-0.059163
2020	0.765356	2.708446	0.450911	3.528634	1697.356111	-0.141890
2021	0.469726	1.972102	0.441595	2.033948	1063.704608	-0.054731
2022	-0.120408	-0.415444	0.440378	-0.217234	-273.420849	-0.125144
2023	0.025620	0.136910	0.512186	0.030621	50.021807	-0.080541
2024	0.223900	1.462835	0.469293	1.010417	477.101064	-0.063262
2025	0.014707	0.115074	0.526680	0.019229	27.923712	-0.017328

Figure 13: Annual performance of Super Alpha.

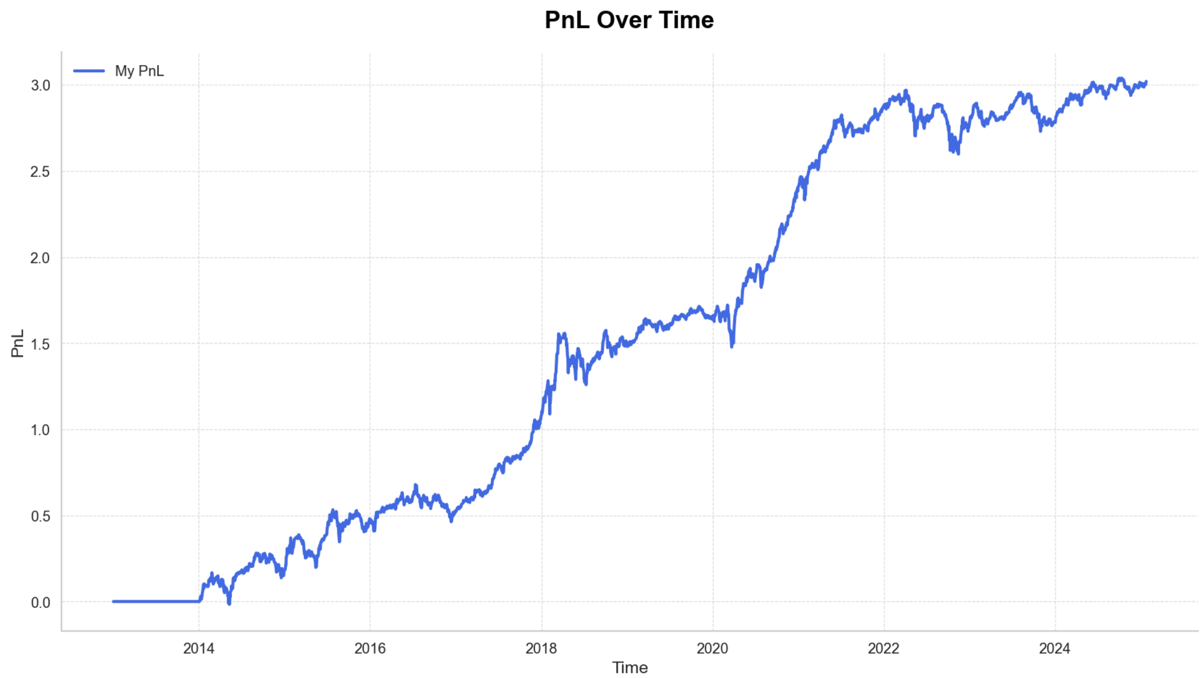

Figure 14: Cumulative profit and loss chart of Super Alpha.

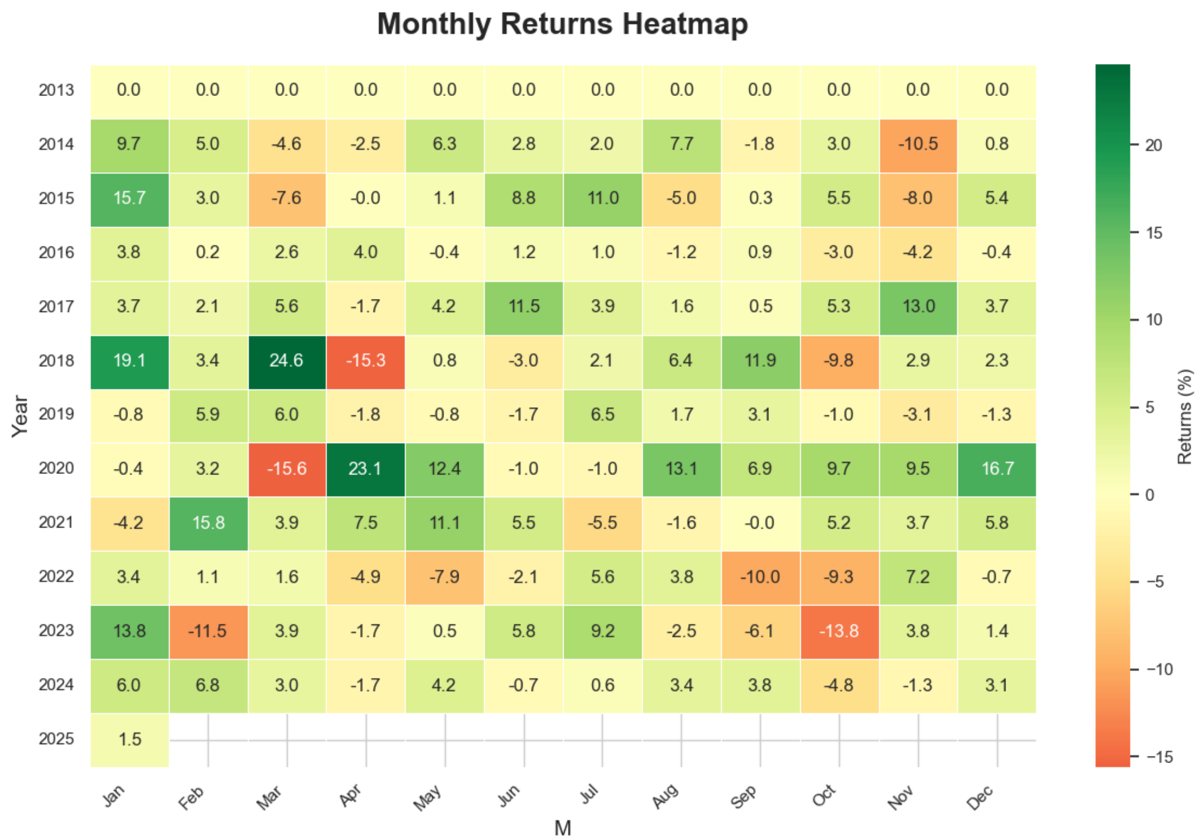

Figure 15: Heatmap showing the profit allocation of Super Alpha.

Figure 13 illustrates the annual performance of Super Alpha, evaluating the strategy's stability and profitability across different market conditions. Figure 14 shows cumulative returns, providing an overview of the strategy's long-term effectiveness. Finally, Figure 15 presents the profit allocation by factors or periods, supporting a detailed analysis of the sources contributing to the overall performance of Super Alpha.

2.4 Risk Management

In addition to developing effective strategies within the investment system, evaluating and managing risks related to market volatility is equally critical. Understanding the nature and extent of these risks allows investors to adjust capital allocation flexibly, particularly during severe market downturns (market crashes). This not only helps preserve capital but also creates opportunities to buy assets at lower prices when the market recovers. This section focuses on analyzing extreme market events, specifically periods when the portfolio's daily returns fall below -2%, to propose effective risk management measures.

2.4.1 Analysis of Daily Returns Data

First, we examine the overview chart of the portfolio's daily returns, as shown in the figure below:

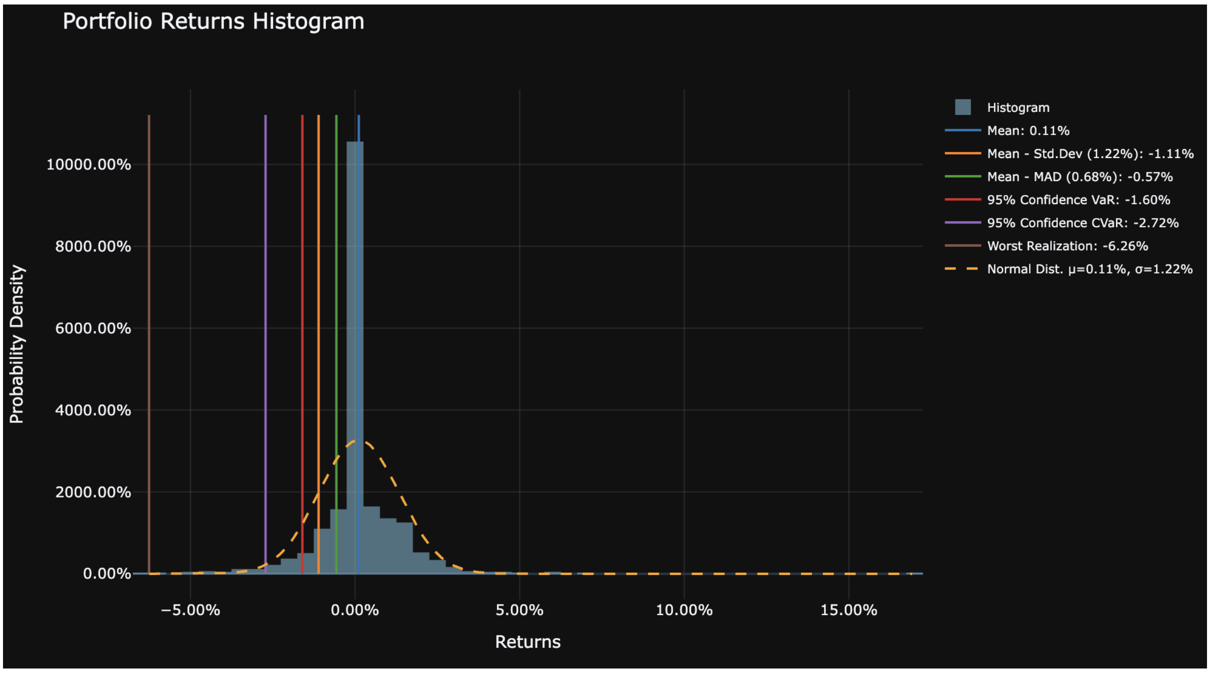

Figure 16: Distribution of the portfolio's daily returns over time.

From Figure 16, it can be observed that approximately 95% of daily return values do not fall below -2.07%, indicating that most market fluctuations are within a safe range. However, to focus on extreme events, we set a threshold of $u = -2\%$ as the criterion for identifying severe declines. These threshold-exceeding events are illustrated in the following figure:

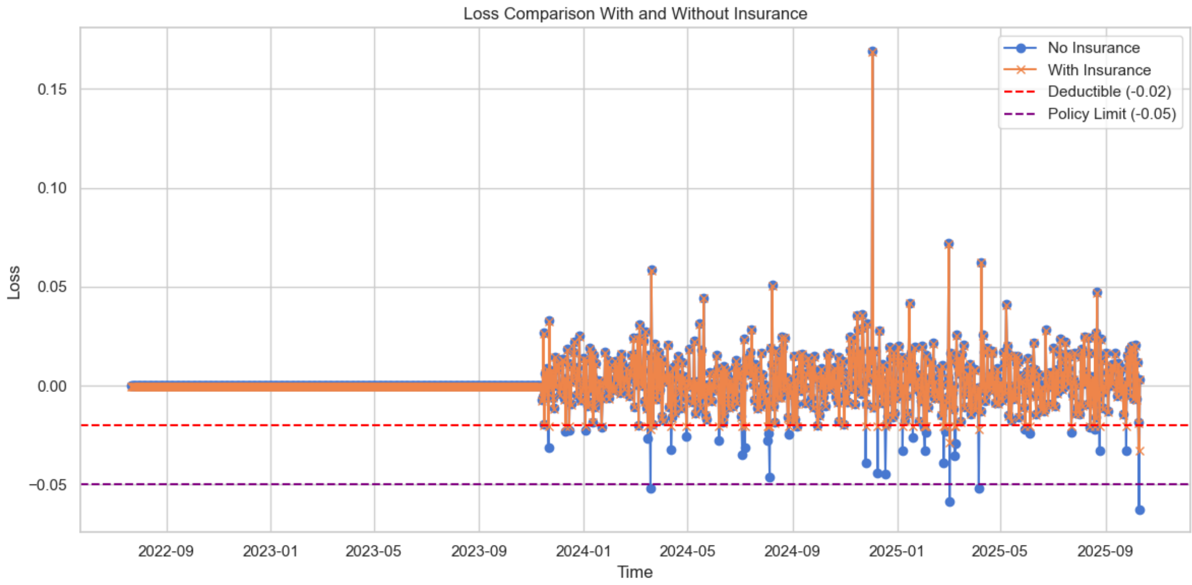

Figure 17: Events where daily returns exceed the -2% threshold, indicating extreme market conditions.

2.4.2 Analysis of Threshold-Exceeding Trends Over Time

Based on the threshold-exceeding events, we plot a chart showing the cumulative number of threshold exceedances over time, as presented in the figure below:

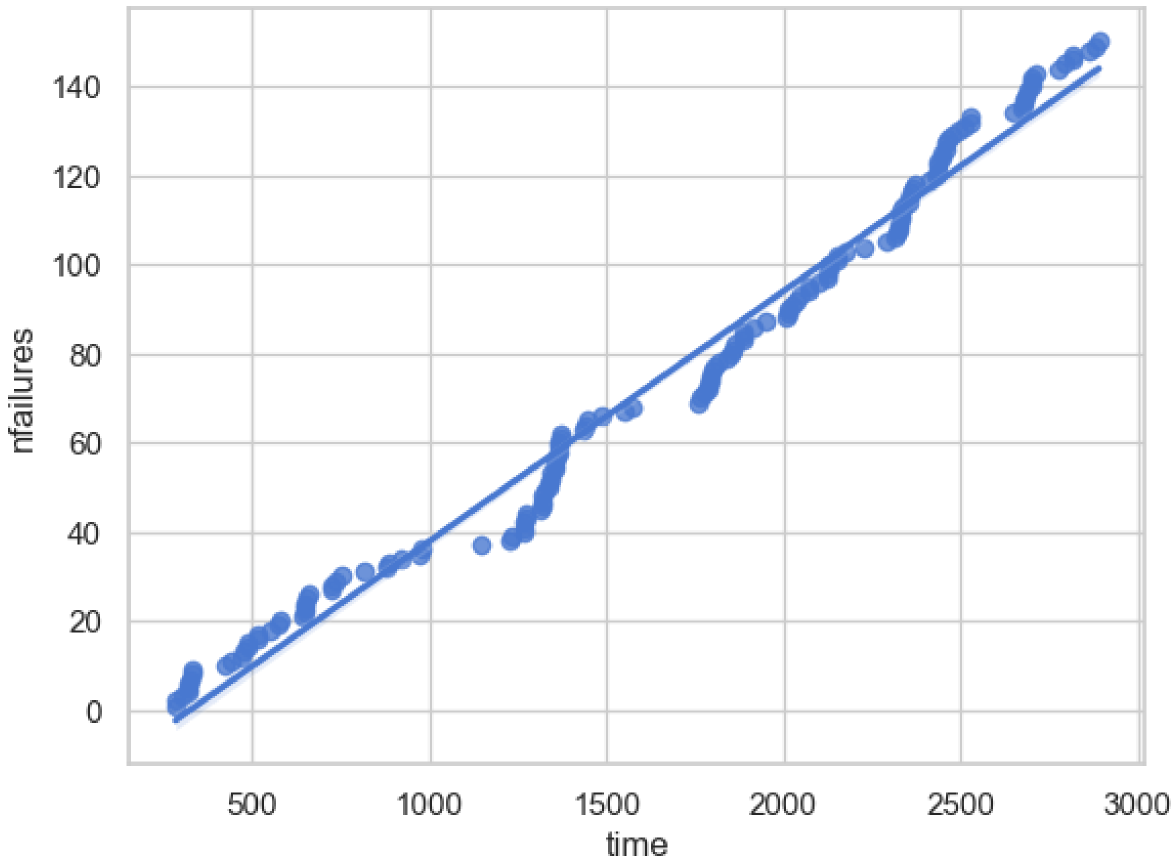

Figure 18: Cumulative number of -2% threshold exceedances over time.

From Figure 18, it can be observed that the cumulative number of threshold exceedances increases over time in a nearly linear pattern. However, notably, these events are not randomly distributed but tend to cluster in specific periods (dense failure clusters), occurring intermittently. This phenomenon suggests that extreme events are not entirely independent but may occur in clusters. To confirm this hypothesis, we apply the DBSCAN algorithm to cluster the threshold-exceeding events, with the results shown in the following figure:

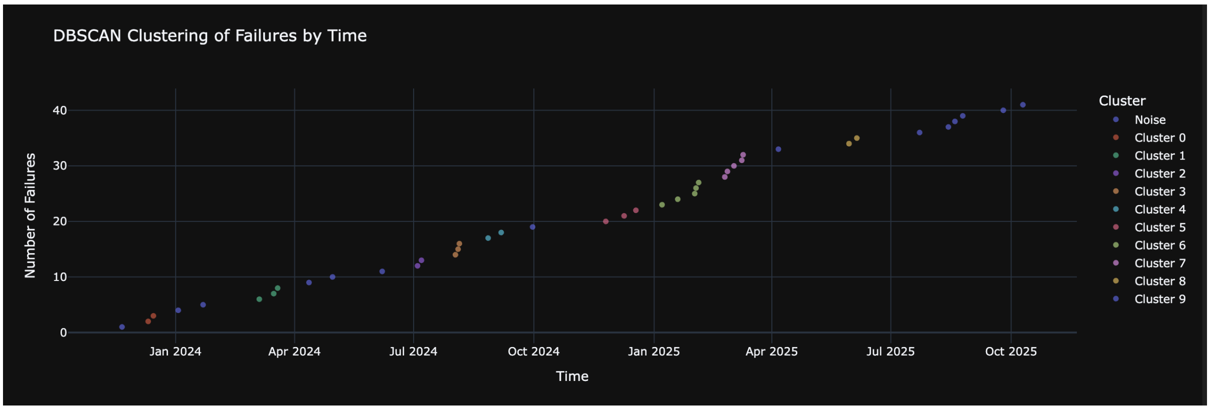

Figure 19: Clustering of -2% threshold-exceeding events using DBSCAN.

2.4.3 Analysis of Cluster Characteristics

Figure 19 shows the clusters formed based on the time intervals between threshold-exceeding events. To further evaluate, we examine the total number of failures in each cluster, as illustrated in the figure below:

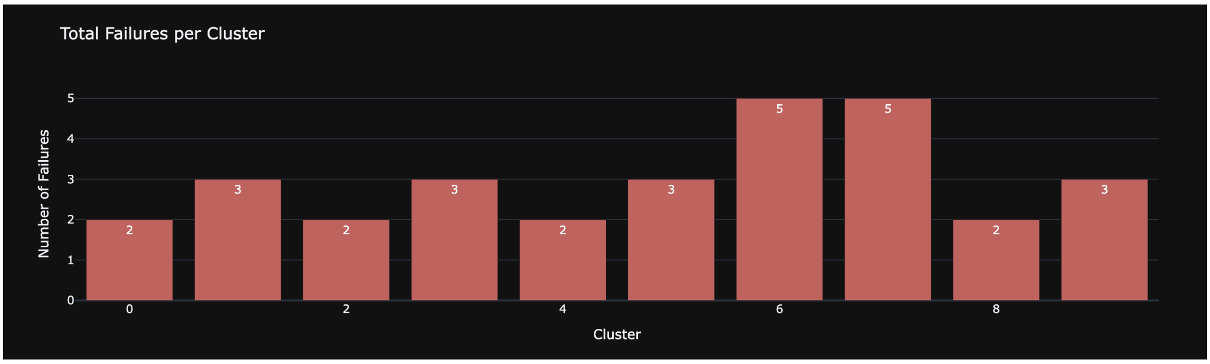

Figure 20: Total number of failures in each cluster, with cluster 10 coinciding with the crypto market crash.

Figure 20 indicates that cluster 10 has a significant number of failures, coinciding with the crypto market crash, reinforcing the accuracy of the clustering. Additionally, we analyze the time intervals between clusters, with the results shown in the figure below:

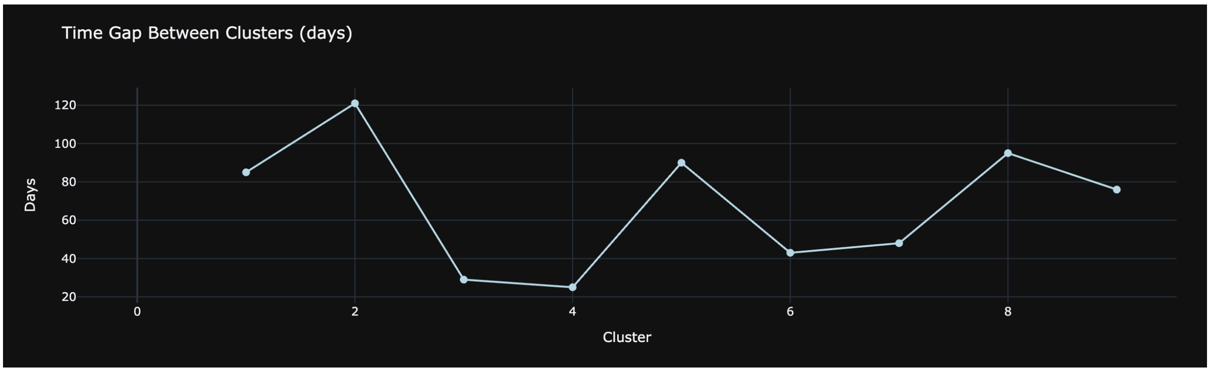

Figure 21: Time intervals (in days) between clusters, averaging approximately 300 days (1 year).

From Figure 21, it can be seen that the average interval between clusters is approximately 300 days (equivalent to 1 year), suggesting that extreme events may follow a certain periodic cycle.

2.4.4 Predicting the Next Crash Using Linear Regression

To predict the timing of the next crash, we apply linear regression to the cumulative number of failures over time. The visualization results are shown in the figure below:

```
Predicted datetime for the 45th failure: 2025-12-11 00:58:21.850508780
Days from now (2025-10-10 00:00:00): 62 days
Predicted datetime for the 50th failure: 2026-03-05 02:22:13.841722480
Days from now (2025-10-10 00:00:00): 146 days
Predicted datetime for the 60th failure: 2026-08-20 05:09:57.824149916
Days from now (2025-10-10 00:00:00): 314 days
```

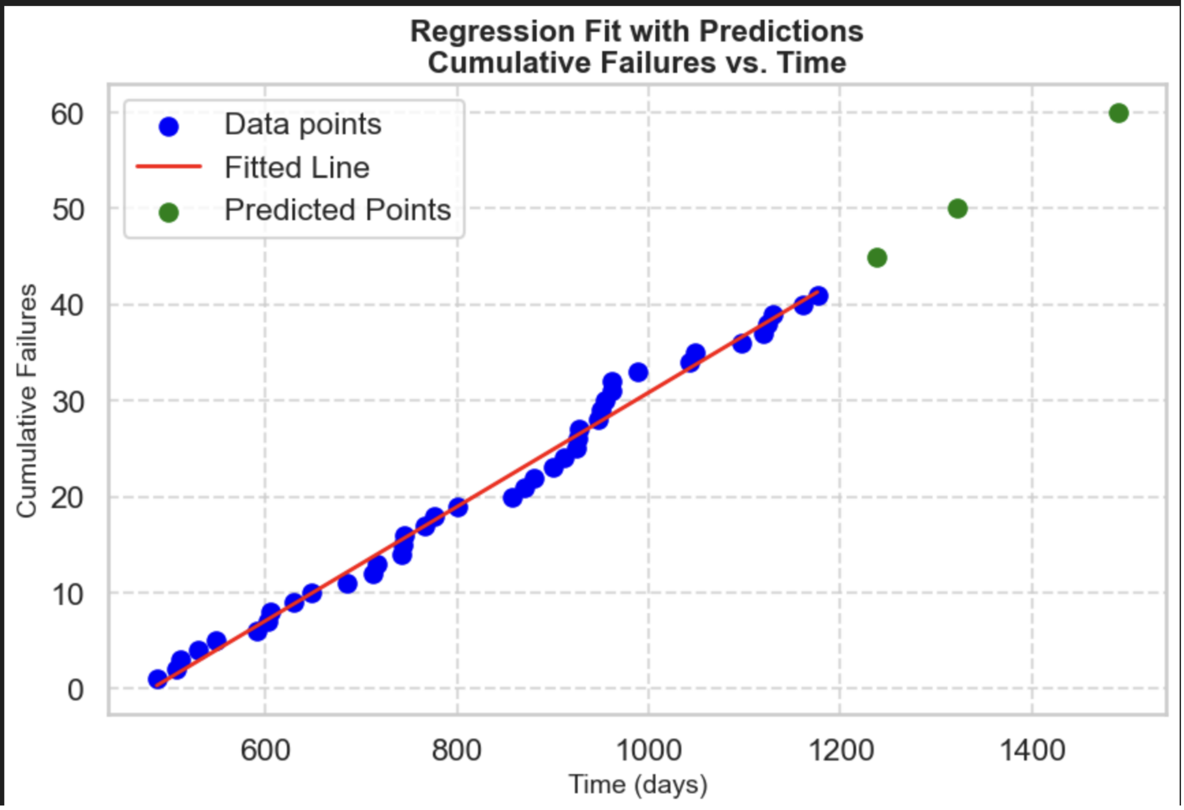

Figure 22: Linear regression analysis predicting the cumulative number of failures over time.

The linear regression results indicate that the model is statistically significant, with the independent variables also achieving significance, as presented in the figure below:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              nfailures   R-squared:                       0.982
Model:                            OLS   Adj. R-squared:                  0.982
Method:                 Least Squares   F-statistic:                     7974.
Date:                Wed, 06 Aug 2025   Prob (F-statistic):         1.27e-130
Time:                        20:07:49   Log-Likelihood:                -477.68
No. Observations:                 150   AIC:                             959.4
Df Residuals:                     148   BIC:                             965.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -18.2206      1.154    -15.785      0.000     -20.502     -15.940
time           0.0561      0.001     89.296      0.000       0.055       0.057
==============================================================================
Omnibus:                       22.587   Durbin-Watson:                   0.069
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                9.035
Skew:                          -0.368   Prob(JB):                       0.0109
Kurtosis:                       2.049   Cond. No.                     4.41e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.41e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 23: Statistical report of the linear regression model, showing the significance of the model and its variables.

Based on this model, we can predict the timing of the n-th crash by extrapolating the time corresponding to the cumulative number of failures.

2.4.5 Additional Analysis Using the Chain Ladder Method

In addition to the above method, we can use the Chain Ladder Method to further study the remaining number of claims (threshold exceedances). The data on the number of claims per month is shown in the figure below:

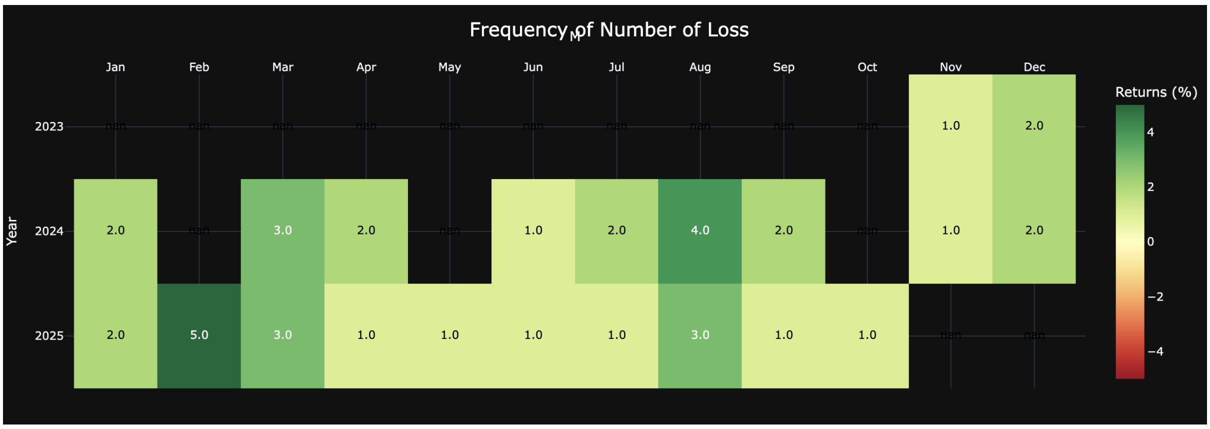

Figure 24: Number of claims per month, used to calculate the total claims per year.

By aggregating the number of claims per year and applying the Chain Ladder Method, we can estimate the average number of remaining extreme value occurrences in the future. This method can also be extended to analyze Severity (degree of severity) and Frequency (frequency), thereby estimating the total potential loss (ultimate loss) by the end of the year. Assuming that the total loss amount per year is nearly constant, this topic can be further expanded by applying a Stochastic Reserve model, such as the Mack Model, to forecast risk reserves more accurately.

2.4.6 Conclusion

The above analysis, combining DBSCAN, linear regression, and the Chain Ladder Method, provides a comprehensive risk management framework, enabling the identification of extreme event clusters, prediction of the next crash, and estimation of potential losses. These methods can be further refined based on real-world data and specific market contexts.