Traders Market Sentiment Report

Introduction:

According to my problem statement I was expected to analyse the relation between traders and market sentiments and share insights for the same.

In the above process I have used some data science concepts for data cleaning modification and predictive models for accurate data analysis which I will be mentioning further

Datasets used:

The dataset that were used include historic_data.csv and fear_greed_index.csv which was share in the drive link that was shared on Internshala.

I have modified the historic_data and fear_greed_index to include add some more insightful columns which were then used for better analysis which were included into another cdv file named as processed_data.csv

Methodology:

Now here comes the technical part different methodologies like data cleaning, Merging, Feature Engineering, Correlation analysis, Predictive models were used for effective analysis of the data.

1. Data cleaning: Some columns like Account and start position were not useful for analysis of finding a relation between market sentiment and a trader so those columns were removed for data cleaning I simply used drop command.
2. Data modifications : Columns like date(in fear_greed_index dataset) and Timestamp IST(in historic_data dataset) cannot be fully merged like in the “Timestamp IST” columns Hour and minutes were also given which was not in the case of “date” column.
3. Feature Engineering : Now after completing all the conversions here comes the important part what I did is I manually analysed datasets and saw that the “Coin” , “PnL” can make a relation with each other and similarly other columns so for confirming that I have added some more important features which were also shown in the “processed_data” dataset those columns include “dayofweek”, ”sentiment_score”, “buy_pressure”, “price_momentum”, ”volume_change”, ”price_volatility” , ”win_rate” etc.

   All these are serving an important in the part of analysis

   a. dayofweek: this come from the Timestamp IST columns which will tell us at what time a trader is interested for buying and selling a stock

b. sentiment_score: Now in Machine Learning we are taught that a machine can only understand the language of numbers so strings like "BUY" "SELL" etc can create an error in the analysis and they are one of the important columns so for that I converted that into 0,1 . If a trader "BUY" then its 1 else it is 0.
c. buy_pressure: Now from here comes the first analysis part relation between "Coin" and the "sentiment score" which tells us whether a trader sells or buys the stock which was mentioned by buy_pressure.
d. Price_momentum: Likewise continuing the analysis relation between "Coin", "Execution Price" which tells us how fast and in what direction a price is moving. Whether it is a positive momentum or a negative momentum or flat neither up nor down
e. Volume_change: This column tells us no of units traded in a given period that will help us comparing the market sentiments

Likewise more columns were used for the analysis that were useful for predicting the market sentiments with traders behvaiour + by adding these columns find the Correlation between all the makes it easy.

4. Correlation analysis : For correlation here the most important columns are "date", "Timestamp IST" , "Closed PnL", "Coin" so for finding a relation between all the columns we use correlation analysis and I have shared heatmaps in which it is clearly visible how they are correlated to each other closer to 1 better the correlation between those columns there are many heatmaps I have shared which also include the values column from "fear_greed_index" dataset and "Classification". Such graphs help us to understand what factors are directly correlated to the market if there is a direct relation then the number comes to be in positive and if it is inversely relatable then it shows negative correlation and if there is very low correlation it shows near to zero value.

5. Predictive Models: So after finding the correlations I tried to make a predictive model which will clarify how market and traders react so I tried to find whether it's a linear relation or I have to use techniques like "Random Forest " which will help find the "hidden patterns" between the dataset and below is my analysis:

    a. Linear Relation: I was checking if there is a Linear relation between the dataset but as I thought obviously it was wrong traders don't only consider only one factor and trade they go through several factors that's why the R2 score comes out to be in Negative which tells that the models is highly inaccurate and RMSE value is very high which tells us that the data highly dependent on several factors the Traders and market sentiment doesn't follow a single pattern they change through sometimes the traders don't even see the "Fee" they just "BUY" Which can be analysed with just a biased analysis. So to find a pattern I gone through Random Forest.

    b. Random Forest: Here comes the main Hero in analysis which not only tells us the hidden patterns but also analyse in many unlikely Linear regression which only dependent on linear factors . After applying the Random Forest Regressor below are some of the analysis part

1. After ap[lying RFR it is found that all the features have corr values between -0.02 and 0.01 with the “Closed PnL” which means there is no strong linear relation exist between features and the target.
2. The only noticeable and considerable relation is between “price_ momentum“ and “price_volatility” which says that current or present market sentiment are not explaining trader profitability well.

So I thought to remove some of the outliers effecting the alanysis and adding columns like win_rate and run again the RFR results came out to be:

1. This time with the heatmap I plotted the bar graph of the features that are effecting the most so it is found that the “Closed PnL” is highly influenced with “Fee” and “hour” in the model’s view.

2. Which effectively means that the model is that the variation in the higher/lower “Fee” directly effects Higher/Lower “PnL” which means the relation between market and the trader is not much stable.

3. “hour” being the second biggest suggests time-of-day effects the trader profitability.

4. While analysing with the other dataset it is noticable that there is is a direct strong relation between “Size USD” and “Fee” which is kind of fair too because Fee usually scale with trade sizes.

Final Insights/ Observation / Results

Observation 1:

a. If we talk about insights than the only strong relationship that is noticeable is between Size USB and Fee this suggests that the trading behaviour is not directly reacting to market sentiments – traders appear unaffected by whether by market is fearful or greedy.

Observation 2:

a. From the all the analysis which includes correlation of all the columns it is visible that the only strong relation is Size-Fee relation that reacts to market sentiments.

Observation 3:

a. Fee is the biggest driver of PnL predictions which means – traders outcomes are more strongly tied to costs than to market movement signals.

b. Hour ranks second, indicating a timing effect – certain hours may be riskier or less profitable.

c. Price momentum, buy pressure, and price volatility have low importance scores, implying traders either don't base decisions on them effectively or they simply aren't predictive of PnL in this dataset.
   In short, most traders' performance is influenced more by structural factors (timing and fees) than by market dynamics.