Quant Research Assignment:

Assumptions:

1. **Data Quality**:

   The accuracy and reliability of the models heavily depend on the quality of the input data. Assumptions are made that the data used for training and testing the models is accurate, relevant, and free from significant errors or biases.

   - It is assumed that the minute-level implied volatilities (IVs) of Bank Nifty and Nifty are accurate and reliable.
   - Time To Expiry (TTE) data is assumed to be correctly calculated and aligned with the corresponding IVs.
   - The dataset may contain missing values, and strategies need to account for these gaps in the data.

2. **Market Hours:**

   The Indian market trading hours between 09:15 and 15:30 are considered for strategy implementation.

3. **P/L Calculation:**

   The P/L calculation formula specified should be accurate and appropriately modified if there are any changes to the IV or spread series.

4. **Model Parameters:** The assumptions include proper tuning of hyperparameters for each model. The performance of machine learning

models, including Z-score trading, Random Forest, and XGBoost, can be sensitive to parameter settings.

5. **Stationarity:** The models assume that the underlying statistical properties of the data, such as mean and variance, remain relatively constant over time. This is especially crucial for financial data in the Z-score trading model.

6. **Feature Importance:** The assumption is made that the features selected for each model are relevant and have a significant impact on the model's predictive power. This involves feature engineering and selection processes.

7. **Market Conditions:** The models assume that the market conditions during the training period are representative of future conditions. Sudden changes in market dynamics, unforeseen events, or structural shifts could impact model performance.

8. **Assumption of Independence:** For Z-score trading, it assumes that the returns of the financial instruments are independent and identically distributed, which might not always be the case in real-world scenarios.

Summary of Results and Findings:

1. Z-Score Trading Model:

Z-score trading models typically involve assessing the standard score of a financial instrument's returns.

The z-score based trading system utilizes the z-scores of the spread between implied volatilities to identify divergence from the historical mean.

The base model assumes that mean reversion in volatility is a profitable strategy. Assumptions include the normal distribution of returns.

Results may vary depending on the choice of parameters and the assumption of constant statistical properties of returns.

2. Random Forest Model:

Random Forest is an ensemble learning method that builds multiple decision trees and merges them together. It is known for handling complex relationships in data.

Findings could include insights into feature importance, but the interpretability might be a challenge.

Random Forest might perform well on certain types of data but can be sensitive to noisy data or outliers.

3. XGBoost Model:

XGBoost is a powerful gradient boosting algorithm that has gained popularity for its performance and efficiency.

The model likely demonstrates high accuracy due to its ability to capture complex relationships in the data.

It may outperform other models, especially when the dataset is large and there are intricate patterns to be learned.

Conclusion:

In conclusion, while the Z-score trading model, Random Forest, and XGBoost all have their merits, XGBoost stands out with its high accuracy of 0.98 and has a robust performance. It is crucial to interpret these findings in the context of the specific problem domain and the assumptions made during the modelling process. Additionally, the practical implementation of these models should consider real-world factors and potential challenges.