* **Project Overview: Detecting Data Anomalies**
  + **The goal of this project is to detect anomalies in a dataset consisting of approximately 150K daily trade data entries.**
* **Scope:**
  + **To improve data quality and decision-making, focus on identifying and mitigating potential outliers, errors, or unusual patterns in trading data.**

1. **Objectives**

* **Primary Goal:**
  + **Develop a machine-learning model to accurately detect anomalies in the dataset.**
* **Secondary Goals:**
  + **Improve data preprocessing techniques.**
  + **Establish robust evaluation metrics to measure the performance of anomaly detection.**
  + **Implement the model in a scalable and maintainable way for future use.**

1. **Project Milestones**
2. **Data Collection and Exploration:**
   * **Deliverables: dataset collection, initial data exploration, and summary statistics.**
3. **Data Preprocessing:**
   * **Deliverables: data cleaning, normalization, and feature engineering.**
4. **Model selection and training:**
   * **Deliverables: selection of appropriate anomaly detection algorithms, model training, and initial testing.**
5. **Model evaluation and optimization:**
   * **Deliverables: model evaluation, parameter tuning, and performance optimization.**
6. **Implementation and Deployment:**
   * **Deliverables: In a production environment, the model is deployed.**
7. **Final Review and Documentation:**
   * **Deliverables: final report, project documentation, and stakeholder review.**
8. **Data Exploration**

* **Data Source:**
  + **The trading dataset contains approximately 150K daily trade entries.**
* **Exploratory Data Analysis (EDA):**
  + **We utilize summary statistics, visualizations, correlation analysis, and anomaly detection techniques.**
* **Key Insights:**
  + **We have identified patterns, outliers, and correlations that will guide the model development phase.**

1. **Data Preprocessing**

* **Data Cleaning:**
  + **Handle missing values, correct data entry errors, and filter out irrelevant data.**
* **Normalization:**
  + **Normalize data to ensure uniformity across features (based on the data).**
* **Feature Engineering:**
  + **Create new features or transform existing ones to improve model performance, if required.**
* **Tools:**
  + **Python, Pandas, NumPy, and Scikit-learn.**

1. **Model Selection**

* **Algorithms (yet to be decided)**
  + **Take into account techniques like isolation forest, one-class SVM, or autoencoders.**
* **Model Training:**
  + **Split the data into training and testing sets, then train the selected models.**
* **Evaluation Metrics:**
  + **Use metrics to evaluate the model's performance.**

1. **Model Evaluation and Optimization**

* **Performance Metrics:**
  + **Evaluate the model on unseen data using cross-validation techniques.**
* **Optimization Techniques:**
  + **We perform hyperparameter tuning, model ensembling, and feature selection.**
* **Tools:**
  + **Python, Scikit-learn.**

1. **Implementation and Deployment**

* **Environment Setup:**

**Still to discuss**

* **Model Integration:**
  + **Integrate the model into the existing trading platform.**
* **Monitoring:**
  + **Implement logging and monitoring to ensure the model performs well in production.**

1. **Risk Management**

* **Potential Risks:**
  + **There are problems with data quality, model overfitting, and computational resource limitations.**
* **Mitigation Strategies:**
  + **We conduct regular data quality checks, validate models, and plan resource allocation.**