### **Prop. Trading Analysis - Data Analyst**

### **Task Breakdown:**

Data: [Trading Data](https://docs.google.com/spreadsheets/d/16RfJdqHqj6Q9ETsMlMxpdxUwAXD19eKOZftJWCrpjVI/edit?usp=sharing)

#### **1. Data Handling & Exploration**

* Load the given dataset into a **Jupyter Notebook**.
* Perform an **initial exploratory data analysis (EDA)** to understand the structure and quality of the data.
* Identify and handle **missing values, duplicates, and inconsistencies**.
* Apply **data cleaning and wrangling techniques** to ensure the dataset is ready for further analysis.

#### **2. Profitability Analysis**

* Conduct an in-depth analysis to identify the **most and least profitable logins**.
* Compute **cumulative profits per login** and rank them based on profitability.
* Visualize the **distribution of profits** across different logins.
* Provide insights and interpretations regarding what factors contribute to profitability.

#### **3. Feature Engineering & Predictive Modeling**

* Engineer **new features** that could help distinguish profitable traders from non-profitable ones.
* Apply **data preprocessing** techniques such as scaling, encoding categorical variables, and handling outliers.
* Use **Machine Learning or Statistical Modeling** to predict **profitable traders**.
  + The goal is **not** to achieve the highest accuracy but to ensure that the model’s predicted profitable traders align with cumulative profit calculations.
  + Example: If the model predicts login-123456 as profitable, validate this by checking whether its cumulative profit is actually positive.

#### **4. Clustering & Trader Segmentation**

* Use **Clustering Techniques** to segment traders into meaningful groups:
  + **Losers** (Consistently making losses)
  + **Profitable Traders** (Consistently making profits)
  + **Outliers** (Traders with extreme profits or losses)
* Manually verify that traders classified as **losers** have negative cumulative profits.
* Provide **visual representations** (scatter plots, histograms, heatmaps) to illustrate the clustering results.

### **Deliverables**

1. **Jupyter Notebook (trading\_analysis.ipynb)**
   * All code, explanations, and visualizations should be included in the notebook.
   * Ensure proper **documentation** and **comments** for clarity.
2. **Technical Report (trading\_analysis\_report.pdf)**
   * Summary of **findings, insights, and conclusions** from the analysis.
   * Explanation of **data cleaning, feature engineering, and modeling techniques** used.
   * Discussion of **model evaluation** and whether the predictions align with cumulative profits.
   * Analysis of **clustering results and verification of loser traders**.
3. **Requirements File (requirements.txt)**
   * List of all **necessary Python libraries** required to run the code.
   * The notebook should **run seamlessly** after installing dependencies using:

pip install -r requirements.txt

* + No additional package/module installation should be needed manually.

### **Additional Considerations**

✅ **Data inconsistencies may exist** – Do not assume the data is incorrect; instead, handle inconsistencies appropriately.  
 ✅ **Reproducibility** – The notebook must be executable in any environment after installing dependencies.  
 ✅ **Performance is secondary** – The focus is on **logical validation**, not model accuracy.  
 ✅ **Well-structured code and documentation** – Ensure readability and maintainability.

### **Final Notes**

This task is designed to test **your ability to extract meaningful insights, handle real-world data issues, and apply machine learning techniques effectively**. **Creativity and clarity** in presenting findings will be valued.

Good luck!