48-Hour Data Science Take-Home Task

Use the provided nova_pay_transactions.csv dataset to build a small end-to-end fraud detection prototype.

Your work should include:

• Data preparation

Handle missing values, fix data types, clean categorical fields, and extract useful time-based features from the timestamp. Address class imbalance if necessary.

• Exploratory data analysis

Create a brief report summarizing key distributions, patterns, and differences between fraudulent and non-fraudulent transactions. Keep it concise but insightful.

• Feature engineering

Add meaningful features such as time features, transaction behavior indicators, and customer-level patterns. Include short explanations for why they may improve the model.

• Machine learning model

Train and evaluate at least one model (e.g., logistic regression, random forest, XGBoost). Use proper splitting, report performance, and include some interpretation like feature importance or SHAP.

• Working demo

Provide either:

– a REST API (FastAPI/Flask) with a POST /predict endpoint

or

– a simple UI (Streamlit/Gradio) that accepts transaction details and returns a fraud score

The demo must run locally with a single command.

• Submission

Include the EDA report, notebooks, model file, demo code, and a clear README with setup and usage instructions. Deliver via GitHub repo or zip file.