**Hawk Financial Anomaly Detection Project**

**Project Structure**

| **Folder/File** | **Description** |
| --- | --- |
| HawkFinancial\_AnomalyDetection.ipynb | Main Jupyter Notebook with full end-to-end anomaly detection code for businesses and transactions. |
| Hawk Financial Take Home Assignment.pptx | Summary of key results/findings in a very short presentation. |
| plots/ | Folder containing all generated plots (for Q1, Q2, and Q3 analysis). |
| Analysis/ | Folder containing output CSV files summarizing detected anomalies, scores, and SHAP explanations for Q1, Q2, and Q3. |
| html\_reports/ | Full detailed HTML reports for each business and overall summary. Start with index.html. |
| challenge\_txs.csv | Original provided transaction dataset. |
| Data Challenge.docx | Problem description file as originally provided. |
| Hawk.zip | Zipped backup of all relevant files. |

**How to Use**

1. **Explore Results Quickly**:
   * Open html\_reports/index.html for an interactive report.
2. **View Code**:
   * Open HawkFinancial\_AnomalyDetection.ipynb to review the complete pipeline, from data prep to anomaly detection and SHAP explanation.
3. **View Summary**:
   * Check Hawk Financial Take Home Assignment.pptx for a quick visual presentation of findings.
4. **Review Output Data**:
   * Look inside Analysis/ for CSV files summarizing:
     + Anomalous businesses
     + Anomalous transactions
     + Top 3 reasons (SHAP) per anomaly
5. **See Visual Insights**:
   * Navigate to plots/ to view global feature importance, radar charts, bar charts, etc.

**Python Libraries Required**

Please make sure you have the following libraries installed:

pip install pandas numpy scipy scikit-learn matplotlib seaborn shap plotly pycountry

Specifically used:

* pandas, numpy: Data manipulation
* scipy.stats: Z-score calculation
* sklearn.ensemble.IsolationForest: Anomaly detection
* matplotlib, seaborn, plotly.express: Visualizations
* pycountry: Country code mapping (optional but installed)
* shap: SHAP values for feature impact explanation

**Notes**

* **Anomaly Detection** was applied both at business-level (Q1) and transaction-level (Q2, Q3) using Isolation Forest.
* **SHAP Explanations** provide human-understandable reasons for anomalies.
* **Thresholds and filters** (such as minimum transaction amounts) were carefully set to ensure focus on meaningful anomalies.
* **Cyclic Pattern Checks** were also added to detect if holding parties appeared as counterparties.