**2025 FinHack Challenge Cases**

JSOM Finance Lab ([JSOMFinLab@utdallas.edu](mailto:JSOMFinLab@utdallas.edu))

**Case 1 Title: Robo-Advisory Portfolio Optimizer: Building Intelligent Investment Portfolios**

**Case Objective:**

To design and implement a robo-advisory tool that recommends an optimized investment portfolio based on an individual's risk tolerance and financial goals. The solution should consider risk and return tradeoff, leveraging historical market data and financial theories.

**Problem Statement:**

You are part of a fintech startup tasked with building a next-generation robo-advisory platform. The platform should recommend investment portfolios tailored to users' risk tolerance, investment horizon, and financial goals. The objective is to maximize risk-adjusted returns while maintaining user-specific constraints.

**Deliverables:**

1. **Code Notebook or Application:** Well-documented code implementing the robo-advisor, including:
   * User input collection, portfolio optimization, and backtesting.
2. **Dashboard or User Interface:**
   * Interactive dashboard displaying portfolio recommendations and insights.
   * Simple, intuitive UI for user interaction and feedback.
3. **Presentation:**
   * Overview of the robo-advisory tool and its features.
   * Explanation of the optimization model and risk management techniques.
   * Demonstration of results with visualizations and performance metrics.

**Evaluation Criteria:**

1. **Optimization Accuracy:** Effectiveness of the portfolio optimization model.
2. **User Experience:** Quality and usability of the user interface and dashboard.
3. **Risk Management:** Robustness of risk management rules and constraints.
4. **Innovation and Complexity:** Novelty of features (e.g., predictive modeling, ESG integration).
5. **Presentation Quality:** Clarity, logical flow, and ability to justify decisions.

**Case 2: Predictive Healthcare Cost Analysis**

**Case Objective**: Healthcare costs are difficult to predict, driven by complex factors including medical history, demographic attributes, regional variations, and evolving economic conditions. This case challenges participants to design a predictive model that estimates individual healthcare costs with greater accuracy and transparency. Using publicly available datasets, participants will explore how AI and data science can be applied to forecast medical expenses, identify cost drivers, and support risk assessment for insurers.

**Statement:** You are a Data Scientist at a health insurance company, responsible for developing an AI-driven healthcare cost prediction tool. Your goal is to build a model that helps insurers estimate future medical expenses by analyzing historical healthcare data, patient demographics, and macroeconomic indicators.

1. Identify and analyze the key drivers of healthcare cost variations.
2. Engineer features from patient records, diagnosis codes, and expenditures.
3. Develop an ML model to predict costs using medical and economic indicators.
4. Optimize model accuracy to reduce cost estimation errors.
5. Visualize insights through interactive dashboards and data visualizations.

**Data Sources & Guidelines:**

* Historical insurance claims (CMS Medicare, HCCI datasets).
* Hospital charges (CMS public datasets).
* Socioeconomic & demographic data (U.S. Census, BEA, FRED).
* Macroeconomic indicators (healthcare inflation, employment trends, GDP impact).
* Public health data (CDC disease indicators, NIH medical expenditures).

**Deliverables:** At the end of the challenge, teams need to submit:

* A working model implementation (Python, SQL, Streamlit, Jupyter Notebook).
* A summary report in ppt or Jupyter Notebook includes as follows:
  + Features used in the model.
  + Methodology applied.
  + Key insights from the analysis.
  + Recommendations for optimizing predictions.
* Data visualizations showcasing cost trends, model predictions, and key drivers.

**Case 3: The $100K Trading Strategy Challenge**

**Case Objective:** Participants will develop an innovative stock trading strategy starting with $100,000 in capital. The goal is to design a model that balances creativity, profitability, and risk management while ensuring feasibility within a short time frame.

**Scope & Requirements**

* **Backtest Period**: 3 years of historical market data
* **Publicly Available Data Sources**: Stock price data, macroeconomic indicators (e.g., GDP, interest rates), financial fundamental datasets from 10Q/10K SEC filings, or other open online datasets.
* Strategy Logic: Clear explanation of the trading strategy, including entry/exit rules and risk controls
* **Backtest Function**: Basic implementation to simulate trading results over the given period
* **Results & Analysis**: Performance metrics such as total return, Sharpe ratio, drawdowns, and key insights

**Deliverables:**

* **The Report should include as follows**: Strategy explanation, Backtest function and implementation, Key results and performance metrics, Discussion of strengths, weaknesses, and potential improvements
* If you use coding, please attach the coding file

**Case 4: “Black Swan Watch” – Predicting the Next Financial and Geo-Political Crisis**

**Case objectives:** A Black Swan event is an unpredictable financial or economic disaster that causes widespread market panic. However, financial crises do not happen in isolation— global geo-political risks such as wars, trade conflicts, or political instability often act as catalysts or exacerbators.

The crisis in the past of three decades:

* Dot-com Bubble (2000) – Overvaluation of tech companies and subsequent market crash.
* 2008 Global Financial Crisis – Subprime mortgage collapse, leading to a global recession.
* U.S.-China Trade War (2018-2020) – Supply chain disruptions and market volatility.
* COVID-19 Market Crash (2020) – Global pandemic triggered economic lockdowns.
* Russia-Ukraine War (2022) – Increased inflation and energy market instability.

Your team works at a top hedge fund, and your job is to build an early warning system that detects financial and geo-political risks before they trigger a market crisis.

**Task Breakdown**

1. Data Exploration: Analyze past financial recessions and major geo-political conflicts to identify common indicators.

2. Feature Engineering: Extract financial and geo-political risk factors like volatility, bond yield spreads, trade war activity, and military tensions.

3. Machine Learning Modeling: Train a predictive model to classify crisis vs. non-crisis periods.

4. Risk Mitigation Strategy: Recommend portfolio allocation strategies based on model outputs.

**Data Sources** Your team can use publicly available datasets, for example: Macroeconomic Indicators:

* Federal Reserve Economic Data (FRED API): Interest rates, unemployment rates, inflation.
* World Bank & IMF: Global GDP growth, debt ratios. Financial Market Data:
* Yahoo Finance API: S&P 500, Dow Jones, Bitcoin, commodity prices.
* VIX Index (Fear Index): Measures market volatility. Geo-Political Risk Data:
* Global Conflict Tracker: Ongoing wars and military conflicts.
* Political Risk Index (PRS Group): Tracks country-specific political instability.
* UN & World Bank Trade Data: Tariff hikes and trade restrictions as early economic distress signals.

**Case 5: Data Optimization in the Banking Industry**

**Problem Statement:**

Financial databases such as Capital IQ and LSEG aggregate data from company SEC filings, attempting to fit reported data into standardized templates. However, due to variations in how companies report line items, these databases can sometimes misclassify data. One common issue is the categorization of **commercial loans** and **real estate loans**, where inconsistent reporting can lead to inaccuracies in data analysis.

Different banks may categorize similar types of loans differently, leading to significant inconsistencies when calculating the total number of **commercial loans** and **real estate loans**. These discrepancies can hinder accurate financial analysis and decision-making.

**Case Objective:**

Your task is to draft a program that efficiently captures the total number of Commercial & Industrial (**C&I)** and Commercial Real Estate (**CRE)** activity by 10 regional banks (KEY, FITB, HBAN, CMA, CFG, PNC, RF, TFC, USB, ZION, MTB) over the past 5 years. The program should identify inconsistencies in reporting, **correct them**, and present the **accurate and corrected figures**.

**A bonus objective**: could be to pull in other fields relating to these loan types as well: average balances, interest income/rate, criticized/npl loans balance, vintage, reserve for loan and lease losses, charge-offs.

Note: we’re focused on SEC/Edgar data, NOT call report/FRY-9 regulatory/FFIEC data

**Resources:**  
To access the required filings, use the following website:  
[SEC EDGAR Company Search](https://www.sec.gov/edgar/searchedgar/legacy/companysearch.html)

**Key Requirements:**

1. **Identify Inconsistencies:** Analyse the loan categorization practices of different banks and detect cases where **C&I** and **CRE** may be mislabelled.
2. **Reclassification:** Automatically reclassify loans as needed to ensure the final counts are accurate.
3. **Data Accuracy:** Ensure that the program correctly categorizes loans, minimizing errors.
4. **Output Format:** The final output should be presented in a structured format (e.g., CSV or Excel), summarizing the **corrected counts of commercial and real estate loans** per bank per year.
5. **Efficiency:** Optimize data extraction and cleaning to ensure accuracy and performance.
6. show a sample” view” that is obtainable from the output. For example, banks that have had a massive change in their C&I/CRE mix over time, or banks that are over/under concentrated relative to peers, or banks that earn more yield, or have more losses, or higher reserves on one category or another.

**Guidance:**

1. **Data Extraction:** Scrape data from the SEC EDGAR database using Python libraries such as Beautiful Soup or Selenium.
2. **Data Cleaning and Correction:** Implement logic to detect mislabelled loans based on variations in reporting practices.
3. **Loan Classification:** Develop criteria to determine when a loan should be classified as **commercial** versus **real estate**.
4. **Output Presentation:** Use libraries like pandas to structure and clean the data and export results to a CSV file.
5. **Performance:** Implement efficient data processing techniques to handle large volumes of data.

**Topic 6: Open Topic**

**Objectives:** AI has been developed significantly in last decade. Scientists focus on to build algorithms so machines can identify objects of images and recognize speech. The models were built to discover the patterns of big data. The technology advancement facilitates innovations in the fields of data center (big data), specialized computer chips, Deep Learning, and Internet of Things. Furthermore, the launch of LLM (ChatGPT) and human like text was expected to bring profound impact on the world. The project is to address the opportunities and challenges in financial industry in the next wave of AI.

**Expected Outcomes**: The participants will pitch an idea about the application and deployment of AGI in different financial areas such as financial advising, credit accessibility, banking operations and services, risk assessment and management, decentralization in finance, compliance and reporting, and etc. Be creative and redefine the traditional job roles and skill sets in financial sector.

**General Guideline**

• **Define the deliverable product:** Decide what products are suitable for the project results. You can develop interface solution (API), an AI agent, an application, a program in Python/R/Java, a storybook or data story in Tableau / PowerBI, or Pitched ideas on PPT.

• **Innovation in Data Usage:** Innovative use of data from varieties of resources like government bureaus, Federal Reserve Bank Research Data (FRED), 10K/10Q on SEC, stock/option/commodity market data, social media platforms, financial platforms like Bloomberg / Capital IQ / LSEG Workspace, ESG data, customer data and demographic information.

• **Data Clean-up and Preprocessing:** Clean, preprocess, and integrate data to handle missing values, encode categorical variables, and prepare for analysis.

• **Exploratory Data Analysis (EDA):** Perform EDA to gain insights of the hack topics. Visualize data to uncover patterns and trends.

• **Algorithm and Technology:** The application of advanced machine learning, AI tools, other algorithms. The participant may be asked to test the performance of algorithm by a specified time of period.

• **Quality and Clarity of Code/Documentation:** The organization and thoroughness of the technical documentation.

• **Practical Implementation and Scalability:** The system's applicability and scalability in real-world decision-making process.

• **Originality and Creativity:** Be creative and showcase your unique solutions in the financial industry.

**AI policy**

This AI policy ensures that all participants use AI tools responsibly, ethically, and in a way that fosters creativity, fairness, and learning during the hackathon.

**✅ Permitted Use**

* Use AI tools (e.g., ChatGPT, GitHub Copilot, DALL·E) for ideation, debugging, or inspiration.
* Cite or document AI assistance clearly.
* Use AI to enhance—not replace—original thinking and team collaboration.

**❌ Prohibited Use**

* Submitting entirely AI-generated projects without human contribution.
* Using AI to plagiarize, copy existing work, or generate misleading content.
* Misrepresenting AI-generated work as fully original.

**📖 Disclosure Requirement**

Participants must:

* Disclose which AI tools were used.
* Explain how those tools helped (e.g., “used ChatGPT to draft UI prompts”).
* Include a brief AI usage statement in their final submission.