**Case Study:**

The risk manager from a consultancy wants to analyze Workers Compensation risk in the United States. He has hired you as a consultant to identify the key factors of high cost workers compensation claim across the US. He provides historical loss experience data.

**Background:**

Workers Compensation policy is mandatory for most of the companies in US for the benefit of employees. The experience of claims in workers compensation drives the premium for next year’s renewal. Hence each company would like to understand the key reasons/ factors for the high cost claims.

**Challenge 1:**

Risk manager from his primary analysis has identified that if a case has entered into litigation, this will drive the cost of claims. The challenge here is when a claimant files his claim neither the insurer nor the employer is aware if a case will go for litigation. Hence primary objective becomes to predict whether a claim will go for litigation or not.

**Objective 1.1:**

Explore the claims data (using graphs/plots) using PowerBI or Python to tell a story with actionable business insight.

**Objective 1.2:**

Build a model to predict high cost workers compensation claim and highlight the key drivers of high-cost claims.

**Objective 1.3:**

Prove that litigation is the key driver of high-cost claims using statistical analysis and tests. Predict if the case will go for litigation.

**Challenge 2:**

Imagine you are designing a database schema for a system that manages workers' compensation claims and uses machine learning models for prediction and analysis.

**Objective 2.1:**

Design the database schema with appropriate tables to store the following information:

1. Workers' details
2. Claims data
3. Machine learning model predictions (Use Model from Challenge 1)

Provide the SQL schema for the database, including table definitions, primary keys, foreign keys, and any necessary relationships or constraints.

**Objective 2.2:**

Provide SQL queries to:

1. Retrieve workers' details along with their total claimed costs.
2. Calculate average claim costs based on industry types.

**Challenge 3:**

You are tasked with developing a web-based tool that predicts the cost of workers' compensation claims based on machine learning model. Your goal is to create a user-friendly interface where users can input the required data and get an instant prediction of the claim cost.

**Objective 3.1:**

1. UI Components: Create a simple HTML page with input fields. Include a button to get the prediction.
2. JavaScript Functionality: Implement a JavaScript function *predictClaimCost()* that takes input values and returns the predicted claim cost using a pre-trained machine learning model.
3. Visualization: Display the predicted claim cost on the UI, either as text or in a visually appealing format (e.g., a chart).
4. Error Handling: Handle input validation to ensure that valid data is submitted. Display an error message if any input is missing or invalid.

**Challenge 4:**

Design a data pipeline to integrate and transform worker compensation data from multiple sources, including CSV files, a relational database, and a RESTful API. Briefly discuss the technologies, tools, and steps involved in extracting, transforming, and loading this data into a centralized data warehouse for analysis.

**Hint:**

Engineer features from description such as cause of injury, body parts damaged, and any which helps predict the Litigation cases/high-cost claims. Also, you are free to engineer new features with relevant logic behind them

**Important:**

For Challenges 1 & 3 please share the outputs as Python working files and the key outcomes in a Word file.

For Challenges 2 & 4 please share your work in a Word file.