Automatidata Project Overview (Courses 1-6)

Project Goal:

In this fictional scenario, the New York City Taxi and Limousine Commission (TLC) has approached the data consulting firm Automatidata to develop an app that enables TLC riders to estimate the taxi fares in advance of their ride.

Background:

Since 1971, the TLC has been responsible for licensing and regulating New York City's taxi cabs, for-hire vehicles, commuter vans, and paratransit vehicles. Automatidata is a data consulting firm that works with clients like the TLC to transform their unused and stored data into useful solutions.

Scenario:

Automatidata has been hired as the data consulting firm. You are the newest member of their data analytics team.

Team Members:

Automatidata Team Members:

- Udo Bankole, Director of Data Analysis
- Deshawn Washington, Data Analysis Manager
- Luana Rodriguez, Senior Data Analyst
- Uli King, Senior Project Manager

New York City TLC Team Members:

- Juliana Soto, Finance and Administration Department Head
- Titus Nelson, Operations Manager

Note:

- The story, all names, characters, and incidents portrayed in this project are fictitious.
- No identification with actual persons (living or deceased) is intended or should be inferred.
- The data shared in this project has been altered for pedagogical purposes.

Course-Specific Details:

Course 1: Project Planning and Proposal

- **Project Background:**
    - Automatidata is in the earliest stages of the TLC project.
    - The following tasks are needed before the data analysis process:
        - A project proposal identifying the following:
            - Organize project tasks into milestones
            - Classify tasks using the PACE workflow
            - Identify relevant stakeholders
- **Your Assignment:**
    - Create a project proposal that will create milestones for the tasks within the TLC project.
    - Consider your audience, team, project goal, and PACE stages of each task.
- **Specific Project Deliverables:**
    - Course 1 PACE Strategy Document
    - Project proposal for the data team

Course 2: Data Understanding and Preparation

- **Project Background:**
    - Automatidata is in the earliest stages of the TLC project.
    - The following tasks are needed before the team can begin the data analysis process:
        - Build a dataframe for the TLC dataset
        - Examine data type of each column
        - Gather descriptive statistics

- **Your Assignment:**
    - Build a dataframe for the TLC data.
    - Organize the data for the process of exploratory data analysis.
    - Update the team on your progress and insights.

- **Specific Project Deliverables:**
    - Course 2 PACE strategy document
    - Answers to questions in the Jupyter notebook project file
    - Coding prep work on project's Jupyter notebook
    - Summary of the column Dtypes
    - Executive summary

Course 3: Exploratory Data Analysis (EDA) and Visualization

- **Project Background:**
    - The New York City TLC data is ready for exploratory data analysis (EDA) in Python.
    - Tasks:
        - Load data, explore, and extract the New York City TLC data with Python
        - Use custom functions to organize the information within the New York City TLC dataset
        - Build a dataframe for the New York City TLC project
        - Create an executive summary for Automatidata for a general audience of internal professionals
- **Your Assignment:**
    - Conduct exploratory data analysis on data for the TLC project.
    - Use Tableau to create visuals for an executive summary.
- **Specific Project Deliverables:**
    - Course 3 PACE Strategy Document
    - Answers to the questions in the Jupyter notebook project file
    - Jupyter notebook of full EDA
    - Tableau visualization showing two important variables
    - Executive summary of results including a visualization

Course 4: Statistical Analysis and Hypothesis Testing

- **Project Background:**
    - Exploratory data analysis is complete for the project.
    - The New York City TLC would like the data team at Automatidata to analyze the relationship between fare amounts and payment type.
    - The team agrees that the next step is to perform a hypothesis test using the data.

- **Your Assignment:**
    - You will conduct hypothesis testing on the data for the TLC data.
    - You've been asked to investigate TLC's dataset to determine which hypothesis testing method best serves the data and the TLC project.
- **Tasks:**
    - Compute descriptive statistics
    - Conduct a hypothesis test using the New York City TLC dataset
    - Create an executive summary for the Automatidata data team before sharing the results with the client
- **Specific Project Deliverables:**
    - Course 4 PACE Strategy Document
    - Answers to the questions in the Jupyter notebook project file
    - Statistical testing
    - Report results in executive summary
    - Hypothesis test prepared with Python

Course 5: Regression Analysis

- **Project Background:**
    - The relationship between fare amounts and payment type has been analyzed.
    - The operations manager with New York City TLC is seeking more insight through regression modeling.
    - The team's next milestone is to run a regression model for taxi fares based on variables in the dataset.
- **Your Assignment:**
    - You will create a regression model.
    - Determine the type of regression model that is needed and develop one using the TLC data.
- **Tasks:**
    - Compute descriptive statistics
    - Create a regression model from the New York City TLC dataset
    - Create an executive summary for the Automatidata data team before sharing the results with the client
- **Specific Project Deliverables:**
    - Course 5 PACE Strategy Document
    - Answers to the questions in the Jupyter notebook project file
    - Build a regression model in Python
    - Report the results in an executive summary

Course 6: Machine Learning

- **Project Background:**
    - New York City TLC stakeholders have been impressed with the data analytical work completed by the Automatidata team in this project.
    - As a result, they have reached out once again for assistance in creating a machine learning model that can help predict whether or not a rider will be a generous tipper.
- **Your Assignment:**
    - You will create a machine learning model for the TLC data.
    - You will be responsible for leading these tasks, which include feature engineering, model development, and evaluation.
- **Tasks:**
    - Build a random forest model from the New York City TLC dataset
    - Create an executive summary for the Automatidata data team before sharing the results with the client
- **Specific Project Deliverables:**
    - Course 6 PACE Strategy Document
    - Answers to the questions in the Jupyter notebook project file
    - Design and implement a machine learning model
    - Draft an executive summary of your results

Key Takeaways (All Courses):

- The end-of-course project is designed for you to practice and apply course skills in a fictional workplace scenario.
- By completing each course’s end-of-course project, you will have work examples that will enhance your portfolio and showcase your skills for future employers.