Course 1: Gathering Information from the Last Executive Meeting of Automatidata

In the last executive meeting at Automatidata, key points were discussed regarding the New York City Taxi & Limousine Commission (TLC) project. Here are the relevant notes:

1. Data Sources:

- Identify and gather data sources related to NYC taxi fares, considering historical and real-time data.

- Explore possibilities for integrating external factors affecting fares, such as traffic, weather, and special events.

2. Data Cleaning and Preprocessing:

- Develop a strategy for cleaning and preprocessing the gathered data to ensure accuracy in fare estimates.

- Implement data validation techniques to handle outliers and missing values.

3. \*\*Feature Engineering:\*\*

- Identify relevant features that can contribute to accurate fare estimates (e.g., distance traveled, time of day, traffic conditions).

- Explore creating new features that may enhance the model's predictive power.

4. \*\*Machine Learning Model Selection:\*\*

- Evaluate various machine learning models suitable for predicting taxi fares.

- Consider regression models that can handle continuous fare values.

5. \*\*Model Training and Testing:\*\*

- Split the dataset into training and testing sets.

- Train the selected machine learning model using historical data and evaluate its performance.

Course 2: Assigning PACE Stages to the Requested Tasks

P - Planning: Tasks 1, 2

- Planning involves identifying data sources and developing a strategy for data cleaning and preprocessing.

A - Acquisition: Task 1

- Acquisition involves gathering relevant data sources for taxi fares.

C - Cleaning: Task 2

- Cleaning involves implementing data cleaning and preprocessing techniques to ensure data accuracy.

E - Exploration: Task 3

- Exploration involves exploring potential features and creating new ones for enhancing the model.

P - Prediction: Tasks 4, 5

- Prediction involves selecting and training machine learning models for fare estimation.

Course 3: Organizing Tasks into Milestones

\*\*Milestone 1: Data Preparation\*\*

- Task 1: Identify and gather relevant data sources.

- Task 2: Develop and implement a strategy for data cleaning and preprocessing.

\*\*Milestone 2: Feature Engineering\*\*

- Task 3: Identify relevant features and explore creating new ones.

\*\*Milestone 3: Model Development\*\*

- Task 4: Evaluate and select machine learning models suitable for fare prediction.

- Task 5: Split the dataset, train the selected model, and evaluate its performance.

Course 4: Creating a Project Proposal for Executive Team Approval

\*\*Project Title:\*\* NYC TLC Fare Estimation App Development

\*\*Objective:\*\* Develop an app that enables TLC riders to estimate taxi fares in advance.

\*\*Milestones:\*\*

1. Data Preparation

- Tasks: Identify and gather data sources, implement data cleaning and preprocessing.

- Timeline: 4 weeks

2. Feature Engineering

- Task: Identify relevant features and explore creating new ones.

- Timeline: 3 weeks

3. Model Development

- Tasks: Evaluate and select machine learning models, split the dataset, train the model, and evaluate performance.

- Timeline: 6 weeks

\*\*Overall Timeline:\*\* 13 weeks

\*\*Resources Needed:\*\*

- Data Scientists: 2

- Data Engineers: 1

- Machine Learning Experts: 2

- Project Manager: 1

\*\*Budget Estimate:\*\* $500,000

\*\*Risk Management:\*\*

- Potential delays in data acquisition.

- Model performance may require iterative improvements.

\*\*Approval:\*\*

This proposal is submitted for executive team approval. Upon approval, the project will proceed with the outlined milestones and timelines.

[Signature]

[Date]

Background on the Automatidata Scenario:

Automatidata is a data consulting firm specializing in transforming unused and stored data into practical solutions for clients. The firm collaborates with clients to address business needs through performance dashboards, customer-facing tools, and strategic business insights. One of Automatidata's ongoing projects involves consulting for the New York City Taxi and Limousine Commission (TLC). The TLC has partnered with Automatidata to develop a regression model that estimates taxi fares based on the data gathered from over 200,000 taxi and limousine licensees, conducting approximately one million combined trips per day.

Project Background:

Automatidata is at the initial stages of the TLC project, and the following tasks are required before commencing data analysis:

1. Develop a project proposal outlining goals and milestones.

2. Provide visuals for TLC's executives in collaboration with Titus Nelson.

3. Inspect the TLC dataset through exploratory data analysis (EDA) to understand the information it provides.

4. Test the model for consistent results.

5. Ensure the model meets project requirements before presenting insights to TLC.

6. Identify main talking points for the TLC presentation.

7. Set up Python for the TLC project.

8. Establish relationships between variables in the TLC data.

9. Consider A/B testing to analyze relationships between the most useful variables.

Stakeholder Information:

\*\*Automatidata Team Members:\*\*

- Udo Bankole, Director of Data Analysis

- Deshawn Washington, Data Analysis Manager

- Luana Rodriquez, 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

Meeting Notes:

Meeting notes from the Automatidata leadership team provide insights into the project requirements and expectations from key team members. Notable points include the need for a global-level project document, visual materials for TLC's executives, dataset inspection through EDA, testing for model consistency, and Python as the preferred language for the project.

Project Deliverables:

1. \*\*Course 1 PACE Strategy Document:\*\*

- Develop a document using the PACE strategy to plan the project, considering audience members, teammates, milestones, and the overall project goal.

2. \*\*Project Proposal for the Data Team:\*\*

- Create a comprehensive project proposal that outlines goals, milestones, tasks, timelines, and resource requirements.

Key Takeaways:

- The end-of-course project is designed for practical application of skills in a fictional workplace scenario.

- The project involves planning tasks, creating milestones, considering PACE workflow, and identifying relevant stakeholders.

- Communication style, problem-solving, and efficient communication are essential skills for successful completion.

- The project provides an opportunity to enhance the portfolio and showcase skills to future employers.