

Automatidata Workplace Scenario

Background:

Been hired as a data analyst at a data consulting firm Automatidata. Automatidata works with its clients to transform their unused and stored data into useful solutions. They specialize in identifying a client's business needs and utilising their data to meet those business needs.

Automatidata is consulting for the New York City Taxi and Limousine Commission (NYC TLC). NYC TLC is an agency responsible for licensing and regulating NYC taxi cabs and for hire vehicles. The agency has partnered with Automatidata to develop a regression model that helps estimate taxi fares before the ride, based on the data that TLC has gathered.

Assignment

Automatidata will need a project proposal that will create milestones for the tasks within the NYC TLC project.

Data Project Questions & Considerations

- Who is your audience for this project?

The New York City Taxi and Limousine Commission (NYC TLC)

- What are you trying to solve or accomplish? And, what do you anticipate the impact of this work will be on the larger needs of the client?

The estimation of taxi fares based on relevant variables that you identify

- What questions need to be asked or answered?

How relevant and up-to date is the data being provided? What variables will be most useful? Are there any trends within the data that will provide useful insights? How can we reduce the impact of bias?

- What resources are required to complete this project?

Relevant up-to date data. Statistical programming language(in this case Python Notebooks). Subject Matter Experts. A team of qualified professionals. A means of rapid communication with the stakeholders in the event of issues arise that need their input.

- What are the deliverables that will need to be created over the course of this project?

The deliverables include a dataset scrubbed for exploratory data analysis, visualizations, statistical model, regression analysis and/or machine learning model.

THE WORKFLOW METHODOLOGY

Project tasks

1. Evaluating the model: Execute

Why did you select this stage for this task?

After the model has been constructed, data is run through to evaluate whether it meets the project's expectations and goals.

2. Conduct hypothesis testing: Analyze and Construct

Why did you select these stages for this task?

During the analyzing stage, it is determined that a statistical test will be used. During the construction phase, the test is carried out.

3. Begin exploring the data: Analyze

Why did you select this stage for this task?

During the analysis phase, you will gain a deeper understanding of the dataset and the information within it.

4. Data exploration and cleaning: Plan and Analyze

Why did you select these stages for this task?

Planning takes place when you first make choices about the methods needed. The cleaning process then takes place in the analyzing stage.

5. Establish structure for project workflow: Plan

Why did you select this stage for this task?

Planning stage. Creating an initial project PACE document outlines the workflow and helps to plan how to best approach a project.

6. Communicate final insights with stakeholders: Execute

Why did you select this stage for this task?

Communication is necessary at various points throughout a project. Final insights are shared with stakeholders in the execute phase of the data project workflow.

7. Compute descriptive statistics: Analyze

Why did you select this stage for this task?

Investigating the statistics within data takes place during analysis.

8. Visualization building: Analyze and Construct

Why did you select these stages for this task?

Visualization begins with data assessment and is created during the construction stage.

9. Write a project proposal: Plan

Why did you select this stage for this task?

Planning stage. A project proposal is the initial document used to define a project.

10. Build a regression model: Analyze and Construct

Why did you select this stage for this task?

During the analyzing stage, the model is examined in detail to be sure it will meet the needs of the task. The building of the regression model will take place in the construction phase.

11. Compile summary information about the data: Analyze

Why did you select this stage for this task?

Inspecting a dataset to compile information would take place in the analysis phase.

12. Build machine learning model: Construct

Why did you select this stage for this task?

The building of a data model would take place in the construct stage.