

Assignment 1: Automatidata EDA Exercise Guide

Learn about the Automatidata workplace scenario!

This assignment focuses on your ability to use exploratory data analysis to organize and understand the data within an assignment. This assignment was designed with you in mind, offering an opportunity for you to practice and apply your data analytic skills.

Background on the Automatidata scenario

Automatidata works with its clients to transform their unused and stored data into useful solutions, such as performance dashboards, customer-facing tools, strategic business insights, and more. They specialize in identifying a client's business needs and utilizing their data to meet those business needs.

Automatidata is consulting for the New York City Taxi and Limousine Commission (TLC). New York City TLC is an agency responsible for licensing and regulating New York City's 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 data that TLC has gathered.

The TLC data comes from over 200,000 taxi and limousine licensees, making approximately one million combined trips per day.

Note: This assignment's dataset was created for pedagogical purposes and may not be indicative of New York City taxi cab riders' behavior.

Assignment background

Automatidata is working on the TLC assignment. The following tasks are needed before the team can begin the data analysis process:

- EDA and cleaning
- Select and build visualization(s) type
 - Create plots to visualize relationships between relevant variables
- Share your results with the Automatidata team

Your assignment

You will conduct exploratory data analysis on data for the TLC assignment. You'll also use Tableau to create visuals for an executive summary to help non-technical stakeholders engage and interact with the data.

The members of Automatidata and the New York City TLC

Automatidata Team Members

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

Your teammates at Automatidata have technical experience with data analysis and data science. However, you should always be sure to keep summaries and messages to these team members concise and to the point.

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 assignment are fictitious. No identification with actual persons (living or deceased) is intended or should be inferred. And, the data shared in this assignment has been created for pedagogical purposes.

The TLC team members are program managers who oversee operations at the organization. Their roles are not highly technical, so be sure to adjust your language and explanation accordingly.

Specific assignment deliverables

With this assignment, you will gain valuable practice of your new skills as you complete the following deliverables:

- Answer the questions in the Jupyter notebook assignment file
- Create a Jupyter notebook of full EDA
- Create a Tableau visualization showing two important variables
- Write an executive summary of results and include a visualization

Good luck in your role! Automatidata looks forward to seeing how you communicate your creative work and approach problem-solving!