Who is the audience for this project?

The New York City Taxi and Limousine Commission (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?

Be able to predict taxi ride fares before each ride based on different factors like distance, time, day etc. It will be able to provide reasonable fare to customers which will attract more customers and that increase will be profitable for TLC too.

What questions need to be asked or answered?

- How soon do they need this prediction model?
- What was the method used for data collection?
- How long was the data collected for and is there any known bias in the data?

What resources are required to complete this project?

- A computer with decent computational power than has python installed.
- Access to a data visualization software like Tableau.
- Good quality data or data that can be cleaned/processed to obtain good data.
- Input from stakeholders and other team members.

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

- Cleaned and Processed Data for Descriptive Analysis and Model Building.
- A dashboard to show our analysis of data or relations between data/trends.
- A Machine Learning Model that does the required prediction of fares.
- A dashboard to show our insights based on the model and earlier analysis.

Following are a group of tasks your company's data team has determined need to be completed within this project. The data analysis manager has asked you to organize these tasks in preparation for the project proposal document. First, identify which stage of the PACE workflow each task would best fit under using the drop-down menu. Next, explain why you selected the stage for each task.

1. **Evaluating the model:** Execute

   We need to check accuracy and performance of our model and make improvements to it before finalizing on our model.

2. **Conduct hypothesis testing:** Analyze **and** Construct

   We should check the relationship between dependent and independent variables using statistical tests. This step is done before data fitting to model.

3. **Begin exploring the data:** Analyze

We explore the data and make sure there are no errors and duplicates. We also process the data current variables into other variables if needed.

4. **Data exploration and cleaning:** Plan **and** Analyze

   We plan at the start what tools and resources are available for the project. We explore the background of data and how it was acquired. Then, the steps involve cleaning data and making sure it is good for statistical analysis and model development.

5. **Establish structure for project workflow (PACE):** Plan

   We need to establish structure for the project workflow at the very start with clear goals and deadlines.

6. **Communicate final insights with stakeholders:** Execute

   After finishing up model build, we can share the insights from the model with stakeholders.

7. **Compute descriptive statistics:** Construct

   We compute descriptive statistics right before we model fitting to understand the relationships between data and their distribution. This helps us in choosing the correct and right factors and discard any redundant information if they are present in the data. This can also help us in determining which model would be a better option.

8. **Visualization building:** Analyze **and** Execute

   After analyzing our data, we should plot any interesting trends and information we can find in the data. Additionally, after our model is finalized, we should plot our insights such as how good the model is, what parameters were the most important and their weightage, etc. These things are important for the stakeholders to know.

9. **Write a project proposal:** Plan

   We should write our project proposal at the start phase, Plan. This way we can work towards a clear goal in mind.

10. **Build a regression model:** Analyze **and** Construct

    After properly analyzing the data, we can start working on building a regression model. However, it is a step that involves going back and forth with the aim to improve its performance but not overfit.

11. **Compile summary information about the data:** Analyze

    After analyzing our data, we should compile a summary of what we see in the data and share it with stakeholders along with visualizations of the trends and useful information found during the analysis.

12. **Build machine learning model:** Construct

After doing complete statistical analysis of the data, we can start working on model development.