

XYZ CONSORTIUM INVESTED LIMITED

DATA INTAKE

Data Source: The data was obtained from the taxi company (Yellow and Pink Cab) based in the United States. It was made accessible by Data Glacier

Data Extraction: Data was downloaded from the Data Glacier Git Hub.

Data Validation & cleaning: Validating the data was necessary in making sure all the related data were combined for the relationship to be established. Also, the data was cleaned and checked for null, errors, and handling of any missing data including the renaming of columns.

Data Transformation: The Data was transformed and saved into a format (CSV) suitable for storage and analysis.

Data Loading: The data was loaded into the Jupyter Notebook for further analysis. All the preliminary checks were done from checking the data loaded, to missing values, columns, shape, and null columns.

Exploratory Data Analysis: EDA was performed, and tested along with the hypothesis and graph plotting, giving pictorial representation and a clear understanding of the data insights.

Data Governance: Data governance policies and procedures were followed to ensure that the data intake process was consistent, compliant with regulations, and aligned with organisational goals and objectives.

NOTE: This is continuous work, and more progress and additional documentation will be made later in the future.