**ABSTRACT**

**TEAM SIZE: 2**

**TEAM MEMBERS**:

1. Shruti Bhadoriya-141100007
2. Pedapati Gnanadeep-141100041

**DOMAIN:**  Data Analysis, Automation, Machine Learning

**TETANTIVE PROBLEM STATEMENT**:

Indian Railways owing to its cost-efficiency remains the most convenient and sought-after mode of transportation. One of the major problems encountered by the users is to make a decision about the train they would opt for, based on the delay, seat availability and other factors. We have observed that the data that is present is scattered and reaching to the conclusion would often require intensive search thus consuming a lot of time. Our aim is to collect relevant information, analyse the data and create visual based report which are easier to comprehend and would expedite the decision making process. Along with the visual reports, with the data that we have obtained, we aim to develop an algorithm which would predict the seat availability using machine learning.

**TECHNOLOGY:** Java, Python, Selenium Web Driver, Power BI, SQL Server, SSIS

**METHODOLOGY:**

* Pulling data from various websites using Selenium Web Driver and dumping the data into database and process the data to get task relevant data.
* The multidimensional data cube would then be built on the top of database.
* Implementing visual reports for various aspects of data using Power BI by fetching required data from cube.
* We aim to achieve end to end automation of the process, right from the data pull to refreshing Power BI reports.
* The data procured would also be used to train a model to predict the seat availability.

**CURRENT STATUS:**

Collected prerequisite needed to develop POWER BI reports. We are at the initial stage of Database Design and working on details to develop a user-friendly application.