**Background of the Project**

Credit One business is being affected by the increase of the number of Customers who have failed to pay(defaulted) loans. So, to minimize this problem, the top management of the company is seeking for data scientists support to identify the credit history of the Customers before granting a loan.

**The Goal (Purpose) of the Project**

Based on the Database, to investigate for those customers who have default loans and then count how many default loans they had.

**The Flowchart of the Project**

Identify the Goal and the outcome of the project

1. Clearly identify the business needs

To Learn how the data are sampled and remove duplicates

1. Get and Clean the data

To do Visualization and Statistics for the variable

1. Explore the data

To build a Decision Tree:

1. Model the data

To communicate(present) the result with stakeholders.

1. Communicate and visualize the results

**Data Collection**

The data is collected from the company’s mySQL server data base by using the below connection

db\_connection\_str = 'mysql+pymysql://deepanalytics:Sqltask1234!@34.73.222.197/deepanalytics

some of the common issues that will be with data are duplicates, Sorting and the Column headers.

The duplicates will be removed by using - df.drop\_duplicates(subset=None, keep='first', inplace=False)

The column header will be fixed by using indexing - df.drop(df.index[0])

**Recommendation (Conclusion)**

Make a recommendation based on the findings while exploring the after the model is validated.