Business Understanding phase is the First phase in our project. As per CRISP DM methodology Business understanding phase consists of following Tasks.

1. Determine Business Objectives
2. Assess the situation
3. Determine Data Science Goals
4. Produce Project plan
5. **Task1->Determining Business Objectives:** (What our client Santander wants to accomplish?)

Santander , a financial firm or Bank wants to make use of Data Science to improve Customer satisfaction , better classification of Customers to know whether a Customer can afford a product or not and also to classify or to predict the Customers who can repay loan amount and find out who will default the payment prior to lending a loan.

The Problem statement is that We want to identify which customers will make a specific transaction in the future irrespective of the amount of money transacted.

Outputs of the Task:

1.Background:

Santander , a financial firm wants to make use of Data Science to optimize Revenues and Customer relationship.

2.Business Objectives:

Improve Business by increasing Revenues , help Customer maintain financial health and help them to achieve their Monetary goals.

3.Business Success criteria:

Santander is a reputed bank and has been doing well in Banking Business since the day of its Inception. Business of it has been doing well and it aims at improving itself far more using Data Science.

b.**Task2->Assessing the situation:** This task aims at getting into the crux of situation.

As the problem statement is limited and assigned to me as part of my project at Edwisor , could not make much into this task . But this task comprises of following outputs aka

1.Inventory of Resources : As it is a student level project , the only resources available are Data Science personnel(1no i.e me) and Laptop. If it is a real time project We need to list out all the available resources available to the project including a schedule of completion and also need to list the personnel in data science team and the Gadgets or tools or software being used as a part of the project.

2.Requirements, Assumptions and Constraints: Requirements are the Data which is provided by Edwisor. Assumptions and Constraints to be listed in real time project. The major assumption I did make is completion of project in the stipulated time provided by Edwisor. Did not found much of constraints as of now(In real time which could be insufficient amount of Data or may be related to computing capacity of tools provided or any other software related issue.)

3.Risks and Contingencies:

Here We list the risks and contingencies which may occur to delay the project or cause it to fail.

4.Terminology: We list out a glossary of Data science Terminology and We list out a Glossary of Business terminology.

5.Costs and Benefits: In real time, We make thorough cost benefit analysis to know potential benefits to our Client and also sometimes the cost of the project may be much more than the results or insights obtained which will make Project to be listed as a failure.

c.**Task3-> Determine Data Science Goals:** A Business Goal objectives in Business terminology and a Data Science Goal states project objectives in technical terms.

Outputs:

1.Describe Data Science Goals: Data Science goals for the problem statement is to predict or classify which Customer is going to make a Transaction.

2.Define Data Science success criteria: Data Science success criteria for the real time project to be described . For this it is predicting which customer is going to transact is our success criteria.

d.**Task4->Produce Project Plan:**  Describing the intended plan for achieving the data mining goals and thereby achieving the business goals.

Outputs:

1.Project plan with project stages ,duration , resources etc..,

Project is intended to complete by stipulated time provided by Edwisor and tools being used for project are R and Python ( Anaconda Wrapper for packages like Scikit-Learn etc..)

2.Initial Assessment of Tools and Techniques:

Tools being used for project are R and Python ( Anaconda Wrapper for packages like Scikit-Learn etc..).

Techniques for our analysis include Statistical Supervised Learning Methods like Linear Regression , Decision Trees and Random Forest.