1. Analytic Approach

- The goal is to build a **predictive model** that can detect whether a credit card transaction is fraudulent or not. We'll use **classification techniques**, which basically means sorting things into two groups: "fraudulent" or "legitimate."
- This model will help predict fraud in real time so banks can act quickly and avoid financial losses.

2. Data Requirements

- To build the model, we need data from past transactions, such as:
    - **Transaction amount**, **location** (where the purchase was made), **time of day**, **card type** (credit or debit), and **merchant** (where the card was used).
    - We also need information about the **user**, like how often they use their card and their credit limits.
- Plus, each transaction must be labeled as either "fraudulent" or "legitimate" so the model can learn what fraud looks like.

3. Data Collection

- The **data collection** will come from the historical transaction records that the bank or card processor already has.
- We could also gather extra information, like details about **location** or **merchant**, if available, to help the model find fraud patterns.

4. Data Understanding and Preparation

- First, we need to **understand the data** by looking at simple charts and statistics to see how fraudulent transactions behave compared to legitimate ones.
- Then, we prepare the data:
    - **Clean** the data by removing duplicates, handling missing values (if some information is incomplete), and fixing any errors.
    - Convert some data into formats the model can understand, like turning words into numbers (for example, merchant type).
    - We also need to deal with the fact that there are **many more legitimate transactions** than fraudulent ones, which can make it hard for the model to learn properly. For this, we can use techniques to balance the data better.

5. Modeling and Evaluation

- This is where we create the **model**. Basically, we are teaching a computer program to identify fraud by using the data we already have.
- We try different approaches to “train” the model, like making it learn patterns from the historical data. Common models here are **Logistic Regression** (which tries to predict the probability of fraud) and **Random Forest** (which makes decisions based on different transaction features).

- Once we have a model, we test it on new data to see **how well** it works. Instead of focusing on technical terms:
    - **Accuracy**: How often does the model correctly identify fraud?
    - **Sensitivity**: How well can it catch fraudulent transactions without missing them?
    - **Overall evaluation**: We make sure the model doesn't block too many legitimate transactions (false positives) or miss fraud (false negatives).