

## Setting CIBIL to walk the Ramp

CIBIL is a leading credit card provider that gets thousands of credit card applicants every year. But in the past few years, it has experienced an increase in credit loss. The CEO believes that the best strategy to mitigate credit risk is to 'acquire the right customers'.

In this case, you will help CIBIL setup data for advanced statistical analysis.

There are two data sets viz. demographic and credit bureau data.

- Demographic data contains customer-level information on age, gender, income, marital status, etc.
- Credit bureau data contains credit data such as 'number of times 30 DPD', 'outstanding balance' etc.

Both files contain a performance tag which represents whether the applicant has gone 90 DPD in the past 12-months after getting a credit card.

In some cases, you will find that all the variables in the credit bureau data are zero and credit card utilization is missing. These represent cases in which there is a no-hit in the credit bureau. You will also find cases with credit card utilization missing. These are the cases in which the applicant does not have any other credit card.

Data cleaning and preparation

Create a master file with all the relevant variables and conduct the necessary data quality checks and cleaning. You'll note that some variables contain a significant number of missing values.