**Mock Project Batch-2**

* Data Dictionary and problem statement

| **application\_record.csv** |  |  |
| --- | --- | --- |
| Feature name | Explanation | Remarks |
| ID | Client number |  |
| CODE\_GENDER | Gender |  |
| FLAG\_OWN\_CAR | Is there a car |  |
| FLAG\_OWN\_REALTY | Is there a property |  |
| CNT\_CHILDREN | Number of children |  |
| AMT\_INCOME\_TOTAL | Annual income |  |
| NAME\_INCOME\_TYPE | Income category |  |
| NAME\_EDUCATION\_TYPE | Education level |  |
| NAME\_FAMILY\_STATUS | Marital status |  |
| NAME\_HOUSING\_TYPE | Way of living |  |
| DAYS\_BIRTH | Birthday | Count backwards from current day (0), -1 means yesterday |
| DAYS\_EMPLOYED | Start date of employment | Count backwards from current day(0). If positive, it means the person currently unemployed. |
| FLAG\_MOBIL | Is there a mobile phone |  |
| FLAG\_WORK\_PHONE | Is there a work phone |  |
| FLAG\_PHONE | Is there a phone |  |
| FLAG\_EMAIL | Is there an email |  |
| OCCUPATION\_TYPE | Occupation |  |
| CNT\_FAM\_MEMBERS | Family size |  |
|  |  |  |
| credit\_record.csv |  |  |
| Feature name | Explanation | Remarks |
| Credit\_record.csv  ID | Client number |  |
| MONTHS\_BALANCE | Record month | The month of the extracted data is the starting point, backwards, 0 is the current month, -1 is the previous month, and so on |
| STATUS | Status | 0: 1-29 days past due 1: 30-59 days past due 2: 60-89 days overdue 3: 90-119 days overdue 4: 120-149 days overdue 5: Overdue or bad debts, write-offs for more than 150 days C: paid off that month X: No loan for the month |

**Background and objective:**

XYZ client is a leading credit card issuing bank, we are partnering with risk team to develop “credit Default identification”.

Task

1. Explore the data.
2. Treat the missing data and outliers.
3. What is the percentage of delinquent customers?
4. Team wants to understand leading indicators of credit default so that potential defaulters can be identified to mitigate risk. Build a machine learning model to predict if an applicant is 'good' or 'bad' client.
5. What is the probability of a customer doing fraud?
6. Percentage of customers and accounts roll forwarded, roll backward and remained in same bucket.