- **OBSERVATION**
  - Data Provided **shape** – 5000 * 20
  - **Column's**
    - 'Village', 'branch_id', 'partner_id', 'Group ID', 'Member ID',
    - 'Branch Name', 'Circle Name', 'Loan Amount', 'Product Type',
    - 'Name of Model', 'Product Detail', 'Institution Name',
    - 'Manufacturer Name', 'Loan Account Number', 'Loan Application Date',
    - 'Product Installed Date', 'Document Signature Date', 'loan_id',
    - 'household_id', 'product_id'
  - **Dropped** the meaning less data

'Group ID'	(Simply drop)
'Member ID'	(Simply drop)
'Circle Name'	(Simply drop)
'Institution Name'	(Simply drop)
'Manufacturer Name'	(Simply drop)
'loan_id'	(Simply drop)
'household_id'	(Simply drop)
'Loan Account Number'	(Simply drop)
'branch_id'	(Simply drop)
partner_id'	(Simply drop)
Document Signature Date	Duplicate
Product Installed Date'	(Created a New Feature)
'Loan Application Date	(Created a New Feature)
'Product Type'	Clubbed
'Name of Model'	Clubbed
'Product Detail'	Clubbed

  - **Dropping** a record with wrong details (most of the details are wrong So I dropped the records) – row no 4113
    - full_loan_data.drop([**4113**],inplace=True)
  - **Data-Type Conversion**

Loan Application Date	Date
Document Signature Date	Date
Loan Application Date	Date
Loan Amount	Float

- **Missing Value Treatment**
  - As we can see Village has 0.0014% null value, I replace the null value by new category “Others” as difference b/w top 2 value is very high. (optional for machine learning)
  - Although Product Detail feature has 0.7175% data as null, but we cannot drop the column’s as it is giving the important information while taking loan . IT's a unique identification of the product that is going to be brought after taking loan Replacing null value with “Others”.

- Data **Unique Values**
  - I am not Binning any data (for ML we need to Bin the data)

Village	2543
Branch Name	160
Loan Amount	86
Product Type	3
Name of Model	14
Product Detail	13
Loan Application Date	331
Product Installed Date	226
Document Signature Date	331
product_id	14

- **Statistical Analysis**
  - Loan range is between 90000 to 1240
  - Data seems to be between 2017-06-26 00:00:00+05:30 to 2019-04-24 00:00:00+05:30
  - Villagers from 'Bijapur CMC' are most likely to demand for loan.
  - Branch in 'Chikodi' is most likely to provide the loan to their customer
  - 'Greenway Jumbo Stove GJS1' is the most demanding Product Type
  - 'GJS1' is the most demanding Model in 'Greenway Jumbo Stove GJS1'
  - Product with product_id i.e. combination of (Product Type', 'Name of Model', 'Product Detail),
    - --- 38 is most demanding product.
    - --- 23 & 42 are least demanding product.
  - Document Signature Date & Loan Application Date are same for all the records. There is a probability it may be auto signed.

- **Data Extraction and Adding new Feature to Data**

Loan Application Year	(Self-explanatory)
Loan Application Month	(Self-explanatory)
Loan Application date	(Self explanatory)
Product Installed Year	(Self explanatory)
Product Installed Month	(Self explanatory)
Product Installed Date	(Self explanatory)
Loan Processed Time	(Self explanatory)
Full_Product_Details	('Product Type', 'Name of Model', 'Product Detail')

- **Data Swapping**
  - Product Installed Date must always be grater then Loan Application Date but for some of the case it doesn't seems so, So In idea case we should drop the rows or replace with mean date, but I swap the value.

- **Data Clubbing**
  - Combining the three features to one single feature 'product_id' indicate the same information that is shown by Full_Product_Details -- (created new feature)
- **Graph - Analysis**
  - Most of the loans are demanded b/w 2450 -- 5000 range --- (Graph 2, 3)
  - There's is a high demand for loan b/w 15 to 25 for date for each Year of month --- (Graph 6, 7)
  - Selling seems to be good for Year -- 2018 on 9th month --- (Graph 8, 9)
  - There is more chance that a customer may ask for loan post 8th month for each year (can we seen 2017, 2018) --- (Graph 4, 5, 9)
  - Higher the product id is there is more chance they will be buy at end month of year --- (Graph M_1)
- **Groupby - Analysis**
  - As Village is highly distributed and has very low frequency data in it (2543 unique values) So difficult to identify and pattern in it.
  - Villages from Sirigere took higher loan i.e. 140150.0 is pending loan with Sirigere villages
  - Brach from Arasikere provides higher loan i.e. 806800.0.0 is provided
  - Highest selling product is with product_id as 38.0
    - i.e Greenway Jumbo Stove GJS1, GJS1, others
  - Highest selling product is with product_id as 42.0 at S.M. Krishnagara Village
    - i.e. Solar Light Selco-- S15HLS-- 12 CFL + 2 Fan

- Product 38.0 is at very high demand most of the villages are high count (34,17,16)
    - i.e. Greenway Jumbo Stove GJS1, GJS1, others
- most selling year was 2018 with 4077390.0 loan amount
- sale got increased on the 2018 but again got decreased on 2019 for all the product
- **Processing time for an application is getting reduced each year from 2017 to 2019 which is average 6.08 days to 3.55 days**