**Libraries that we used:**

Pandas, Numpy, Matplotlib and some other Libraries but the main ones are these.

import pandas as pd

import numpy as np

from pandas import Series,DataFrame

import matplotlib.pyplot as plt

**Read our .csv file in to DataFrame df:**

df = pd.read\_csv('New\_ds.csv') # Initially we Uploaded our original file as the df.

**Visualize and summary of Dataset:**

print(df.shape)

df.head(20)

df.describe() # it shows all the numeric values with all aggregation functions.

Df.dtypes()

df['loan\_amnt'].describe()

df['loan\_amnt'].isnull().sum()

print(df[df['loan\_amnt'].isnull()])

df['loan\_amnt'].value\_counts().plot.line()

df.loc[:,'loan\_amnt':'term'].plot.box()

df['home\_ownership'].value\_counts().sort\_index().plot.bar()

df['loan\_status'].value\_counts().sort\_index().plot.bar()

df['term'].value\_counts().sort\_index().plot.bar()

pd.scatter\_matrix(df.loc[:,'loan\_amnt':'home\_ownership'], alpha=0.2)

**converting different features for using them in project:**

removing “%“ sign from ‘Int\_rate’ column:

df['int\_rate'].astype(str).str.rstrip('%')

removing “months” from ‘term’ column:

df['term'].astype(str).str.rstrip('months')

**Dropping columns having ‘NaN’ values in some columns:**

df = df.dropna(how='all', axis=1)

df = df.drop(['debt\_settlement\_flag\_date','settlement\_status','settlement\_date','settlement\_amount',

'settlement\_percentage','settlement\_term'], axis=1)