DOWNLOADING THE DATASET FROM KAGGLE

```python
from google.colab import files
uploaded = files.upload()
```

```
<IPython.core.display.HTML object>

Saving train_u6lujuX_CVtuZ9i.csv to train_u6lujuX_CVtuZ9i.csv
```

LOADING THE DATA

```python
import pandas as pd
df = pd.read_csv('train_u6lujuX_CVtuZ9i.csv')
print(df.head())
```

```
    Loan_ID Gender Married Dependents     Education Self_Employed  \
0  LP001002   Male      No          0      Graduate            No
1  LP001003   Male     Yes          1      Graduate            No
2  LP001005   Male     Yes          0      Graduate           Yes
3  LP001006   Male     Yes          0  Not Graduate            No
4  LP001008   Male      No          0      Graduate            No

   ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0             5849                0.0         NaN             360.0
1             4583             1508.0       128.0             360.0
2             3000                0.0        66.0             360.0
3             2583             2358.0       120.0             360.0
4             6000                0.0       141.0             360.0

   Credit_History Property_Area Loan_Status
0             1.0         Urban           Y
1             1.0         Rural           N
2             1.0         Urban           Y
3             1.0         Urban           Y
4             1.0         Urban           Y
```

INSPECTING THE DATA

```python
print(df.head())
print(df.tail())
```

```
    Loan_ID Gender Married Dependents     Education Self_Employed  \
0  LP001002   Male      No          0      Graduate            No
1  LP001003   Male     Yes          1      Graduate            No
2  LP001005   Male     Yes          0      Graduate           Yes
3  LP001006   Male     Yes          0  Not Graduate            No
```

```
4  LP001008   Male      No          0      Graduate           No

   ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0             5849                0.0         NaN             360.0
1             4583             1508.0       128.0             360.0
2             3000                0.0        66.0             360.0
3             2583             2358.0       120.0             360.0
4             6000                0.0       141.0             360.0

   Credit_History Property_Area Loan_Status
0             1.0         Urban           Y
1             1.0         Rural           N
2             1.0         Urban           Y
3             1.0         Urban           Y
4             1.0         Urban           Y
      Loan_ID  Gender Married Dependents Education Self_Employed  \
609  LP002978  Female      No          0  Graduate            No
610  LP002979    Male     Yes         3+  Graduate            No
611  LP002983    Male     Yes          1  Graduate            No
612  LP002984    Male     Yes          2  Graduate            No
613  LP002990  Female      No          0  Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term
\
609             2900                0.0        71.0             360.0

610             4106                0.0        40.0             180.0

611             8072              240.0       253.0             360.0

612             7583                0.0       187.0             360.0

613             4583                0.0       133.0             360.0


     Credit_History Property_Area Loan_Status
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N
```

HANDLING MISSING VALUES

```
missing_values = df.isnull().sum()
columns_with_missing = missing_values[missing_values>0].index.tolist()
print("Columns with missing values:", columns_with_missing)
```

```
print("Number of missing values in each column:\n",
missing_values[missing_values > 0])
```

```
Columns with missing values: ['Gender', 'Married', 'Dependents',
'Self_Employed', 'LoanAmount', 'Loan_Amount_Term', 'Credit_History']
Number of missing values in each column:
 Gender              13
Married              3
Dependents          15
Self_Employed       32
LoanAmount          22
Loan_Amount_Term    14
Credit_History      50
dtype: int64
```

FINDING NON NUMERICAL VALUES

```
non_numerical_cols =
df.select_dtypes(exclude=['number']).columns.tolist()
print("Non-numerical columns:", non_numerical_cols)
```

```
Non-numerical columns: ['Loan_ID', 'Gender', 'Married', 'Dependents',
'Education', 'Self_Employed', 'Property_Area', 'Loan_Status']
```

FILTERING

```
non_graduates = df[df['Education'] != 'Graduate']['Loan_ID']
print("Loan IDs of non-graduates:\n", non_graduates)
```

```
Loan IDs of non-graduates:
 3      LP001006
6      LP001013
16     LP001034
18     LP001038
20     LP001043
         ...
595    LP002940
596    LP002941
601    LP002950
605    LP002960
607    LP002964
Name: Loan_ID, Length: 134, dtype: object
```

```
low_income = df[df['ApplicantIncome'] < 5000]['Loan_ID']
print("Loan IDs of people with income < 5000:\n", low_income)
```

```
Loan IDs of people with income < 5000:
 1      LP001003
2      LP001005
3      LP001006
6      LP001013
7      LP001014
         ...
607    LP002964
608    LP002974
609    LP002978
610    LP002979
613    LP002990
Name: Loan_ID, Length: 418, dtype: object
```

```
unmarried_dependents = df[(df['Married'] == 'No') & (df['Dependents']
== '3+')]['Loan_ID']
print("Loan IDs of unmarried people with '3+' dependents:\n",
unmarried_dependents)
```

```
Loan IDs of unmarried people with '3+' dependents:
 34     LP001100
255    LP001846
338    LP002113
390    LP002255
442    LP002418
592    LP002933
600    LP002949
Name: Loan_ID, dtype: object
```

```
rows_dependents_3_plus = df[df['Dependents'] == '3+']
print("Rows with '3+' dependents:\n", rows_dependents_3_plus)
```

```
Rows with '3+' dependents:
       Loan_ID  Gender Married Dependents     Education Self_Employed
\
7    LP001014    Male     Yes         3+      Graduate            No

34   LP001100    Male      No         3+      Graduate            No

61   LP001206    Male     Yes         3+      Graduate            No

68   LP001238    Male     Yes         3+  Not Graduate           Yes

73   LP001250    Male     Yes         3+  Not Graduate            No

74   LP001253    Male     Yes         3+      Graduate           Yes

78   LP001263    Male     Yes         3+      Graduate            No

79   LP001264    Male     Yes         3+  Not Graduate           Yes
```

109	LP001384	Male	Yes	3+	Not	Graduate	No
126	LP001448	NaN	Yes	3+		Graduate	No
135	LP001488	Male	Yes	3+		Graduate	No
155	LP001536	Male	Yes	3+		Graduate	No
171	LP001585	NaN	Yes	3+		Graduate	No
172	LP001586	Male	Yes	3+	Not	Graduate	No
177	LP001610	Male	Yes	3+		Graduate	No
202	LP001682	Male	Yes	3+	Not	Graduate	No
211	LP001711	Male	Yes	3+		Graduate	No
213	LP001715	Male	Yes	3+	Not	Graduate	Yes
215	LP001720	Male	Yes	3+	Not	Graduate	No
255	LP001846	Female	No	3+		Graduate	No
257	LP001854	Male	Yes	3+		Graduate	No
259	LP001864	Male	Yes	3+	Not	Graduate	No
267	LP001882	Male	Yes	3+		Graduate	No
295	LP001949	Male	Yes	3+		Graduate	NaN
321	LP002053	Male	Yes	3+		Graduate	No
324	LP002065	Male	Yes	3+		Graduate	No
338	LP002113	Female	No	3+	Not	Graduate	No
340	LP002115	Male	Yes	3+	Not	Graduate	No
343	LP002126	Male	Yes	3+	Not	Graduate	No
352	LP002141	Male	Yes	3+		Graduate	No
359	LP002160	Male	Yes	3+		Graduate	No
376	LP002219	Male	Yes	3+		Graduate	No
390	LP002255	Male	No	3+		Graduate	No
391	LP002262	Male	Yes	3+		Graduate	No

409	LP002317	Male	Yes	3+	Graduate	No
442	LP002418	Male	No	3+	Not Graduate	No
461	LP002484	Male	Yes	3+	Graduate	No
466	LP002500	Male	Yes	3+	Not Graduate	No
472	LP002519	Male	Yes	3+	Graduate	No
481	LP002536	Male	Yes	3+	Not Graduate	No
515	LP002659	Male	Yes	3+	Graduate	No
522	LP002692	Male	Yes	3+	Graduate	Yes
531	LP002720	Male	Yes	3+	Graduate	No
539	LP002740	Male	Yes	3+	Graduate	No
557	LP002795	Male	Yes	3+	Graduate	Yes
567	LP002837	Male	Yes	3+	Graduate	No
574	LP002863	Male	Yes	3+	Graduate	No
592	LP002933	NaN	No	3+	Graduate	Yes
600	LP002949	Female	No	3+	Graduate	NaN
602	LP002953	Male	Yes	3+	Graduate	No
610	LP002979	Male	Yes	3+	Graduate	No

\	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term
7	3036	2504.0	158.0	360.0
34	12500	3000.0	320.0	360.0
61	3029	0.0	99.0	360.0
68	7100	0.0	125.0	60.0
73	4755	0.0	95.0	NaN
74	5266	1774.0	187.0	360.0
78	3167	4000.0	180.0	300.0
79	3333	2166.0	130.0	360.0

109	2071	754.0	94.0	480.0
126	23803	0.0	370.0	360.0
135	4000	7750.0	290.0	360.0
155	39999	0.0	600.0	180.0
171	51763	0.0	700.0	300.0
172	3522	0.0	81.0	180.0
177	5516	11300.0	495.0	360.0
202	3992	0.0	NaN	180.0
211	3430	1250.0	128.0	360.0
213	5703	0.0	130.0	360.0
215	3850	983.0	100.0	360.0
255	3083	0.0	255.0	360.0
257	5250	0.0	94.0	360.0
259	4931	0.0	128.0	360.0
267	4333	1811.0	160.0	360.0
295	4416	1250.0	110.0	360.0
321	4342	189.0	124.0	360.0
324	15000	0.0	300.0	360.0
338	1830	0.0	NaN	360.0
340	2647	1587.0	173.0	360.0
343	3173	0.0	74.0	360.0
352	2666	2083.0	95.0	360.0
359	5167	3167.0	200.0	360.0
376	8750	4996.0	130.0	360.0
390	9167	0.0	185.0	360.0
391	9504	0.0	275.0	360.0

409	81000	0.0	360.0	360.0
442	4707	1993.0	148.0	360.0
461	7740	0.0	128.0	180.0
466	2947	1664.0	70.0	180.0
472	4691	0.0	100.0	360.0
481	3095	0.0	113.0	360.0
515	3466	3428.0	150.0	360.0
522	5677	1424.0	100.0	360.0
531	4281	0.0	100.0	360.0
539	6417	0.0	157.0	180.0
557	10139	0.0	260.0	360.0
567	3400	2500.0	123.0	360.0
574	6406	0.0	150.0	360.0
592	9357	0.0	292.0	360.0
600	416	41667.0	350.0	180.0
602	5703	0.0	128.0	360.0
610	4106	0.0	40.0	180.0

	Credit_History	Property_Area	Loan_Status
7	0.0	Semiurban	N
34	1.0	Rural	N
61	1.0	Urban	Y
68	1.0	Urban	Y
73	0.0	Semiurban	N
74	1.0	Semiurban	Y
78	0.0	Semiurban	N
79	NaN	Semiurban	Y
109	1.0	Semiurban	Y
126	1.0	Rural	Y
135	1.0	Semiurban	N
155	0.0	Semiurban	Y
171	1.0	Urban	Y
172	1.0	Rural	N

```
177              0.0     Semiurban           N
202              1.0         Urban           N
211              0.0     Semiurban           N
213              1.0         Rural           Y
215              1.0     Semiurban           Y
255              1.0         Rural           Y
257              1.0         Urban           N
259              NaN     Semiurban           N
267              0.0         Urban           Y
295              1.0         Urban           Y
321              1.0     Semiurban           Y
324              1.0         Rural           Y
338              0.0         Urban           N
340              1.0         Rural           N
343              1.0     Semiurban           Y
352              1.0         Rural           Y
359              1.0     Semiurban           Y
376              1.0         Rural           Y
390              1.0         Rural           Y
391              1.0         Rural           Y
409              0.0         Rural           N
442              1.0     Semiurban           Y
461              1.0         Urban           Y
466              0.0         Urban           N
472              1.0     Semiurban           Y
481              1.0         Rural           Y
515              1.0         Rural           Y
522              1.0         Rural           Y
531              1.0         Urban           Y
539              1.0         Rural           Y
557              1.0     Semiurban           Y
567              0.0         Rural           N
574              1.0     Semiurban           N
592              1.0     Semiurban           Y
600              NaN         Urban           N
602              1.0         Urban           Y
610              1.0         Rural           Y
```

```
df_cleaned = df.dropna()
```

DROP UNWANTED COLUMNS

```
df_cleaned = df_cleaned.drop(columns=['Loan_ID'])
```