

colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH38#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▶ Run all

RAM Disk

[1] 0s

#Predicting Home Loan Approval

[2] 1s

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

[3] 1s

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.naive_bayes import GaussianNB
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, precision_score, confusion_matrix
```

[4] 0s

```
df = pd.read_csv('/content/loan_sanction_train.csv')
df.head()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	1.0	
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	1.0	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	
3	LP001006	Male	Yes	0	Not	No	2583	2358.0	120.0	360.0	1.0	

Variables Terminal

5:43 PM Python 3

Top Stories
Leaders across p...

Search

ENG IN 7:12 PM 12/13/2025

Commands + Code + Text Run all

RAM Disk

3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	1.0

Next steps: [Generate code with df](#) [New interactive sheet](#)

```
df.sample(3)
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Prope
534	LP002731	Female	No	0	Not Graduate	Yes	18165	0.0	125.0	360.0	1.0	
6	LP001013	Male	Yes	0	Not Graduate	No	2333	1516.0	95.0	360.0	1.0	
382	LP002231	Female	No	0	Graduate	No	6000	0.0	156.0	360.0	1.0	

```
df.shape
```

```
(614, 13)
```

```
df.isnull().sum()
```

Variables Terminal

5:43 PM Python 3

	0
Loan_ID	0
Gender	13
Married	3
Dependents	15
Education	0
Self_Employed	32
ApplicantIncome	0
CoapplicantIncome	0
LoanAmount	22
Loan_Amount_Term	14
Credit_History	50
Property_Area	0
Loan_Status	0
dtype: int64	
[0] df.duplicated().sum()	
np.int64(0)	

colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH38#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▶ Run all

RAM Disk

[9] ✓ Os

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Loan_ID                614 non-null    object
1   Gender                 601 non-null    object
2   Married                611 non-null    object
3   Dependents             599 non-null    object
4   Education              614 non-null    object
5   Self_Employed          582 non-null    object
6   ApplicantIncome        614 non-null    int64
7   CoapplicantIncome      614 non-null    float64
8   LoanAmount             592 non-null    float64
9   Loan_Amount_Term       600 non-null    float64
10  Credit_History          564 non-null    float64
11  Property_Area          614 non-null    object
12  Loan_Status            614 non-null    object
dtypes: float64(4), int64(1), object(8)
memory usage: 62.5+ KB
```

[10] ✓ Os

df.describe()

1 to 8 of 8 entries Filter ?

index	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
count	614.0	614.0	592.0	600.0	564.0
mean	5403.459283387622	1621.2457980271008	146.41216216216216	342.0	0.8421985815602837
std	6109.041673387174	2926.2483692241917	85.58732523570545	65.12040985461256	0.3648783192364049
min	150.0	0.0	9.0	12.0	0.0

Variables Terminal

5:43 PM Python 3

Top Stories Leaders across p...

Search

ENG IN 7:13 PM 12/13/2025

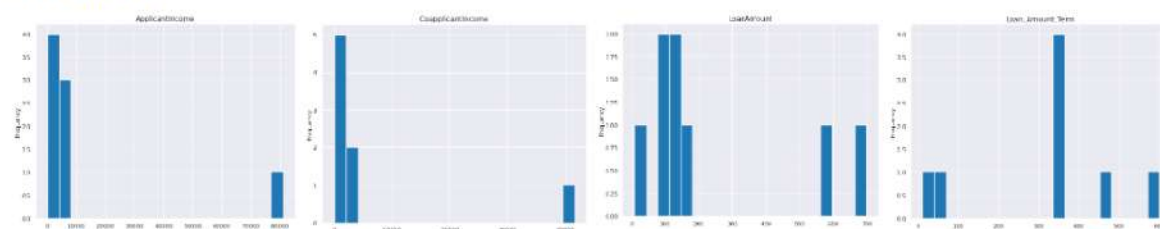
25%	2877.5	0.0	100.0	360.0	1.0
50%	3812.5	1188.5	128.0	360.0	1.0
75%	5795.0	2297.25	168.0	360.0	1.0
max	81000.0	41667.0	700.0	480.0	1.0

Show 25 per page

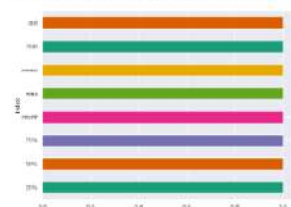


Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

Distributions

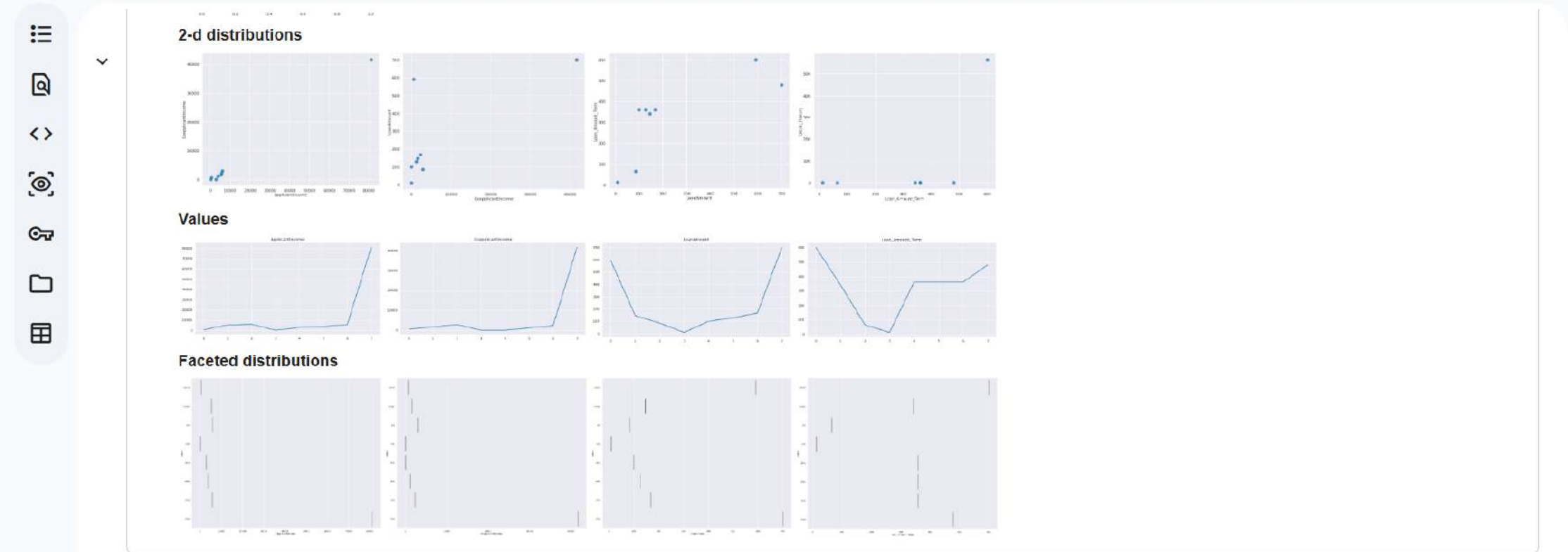


Categorical distributions



2-d distributions





[11]

✓ 0s

#EDA

[12]

✓ 0s



#Data Visualization

df.head()



colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH38#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▶ Run all

RAM Disk

#EDA

[12] #Data Visualization
df.head()

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	1.0	
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	1.0	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0	
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	1.0	

Next steps:

Generate code with df

New interactive sheet

[13]

sns.set_style('darkgrid')
plt.figure(figsize=(10,5))
sns.countplot(x=df['Loan_Status'], palette='Reds')
plt.title('Loan Status Distribution', fontsize=16)
plt.xlabel('Loan_Status', fontsize=13)
plt.ylabel('Count', fontsize=13)
plt.show()

Loan Status Distribution

Variables Terminal

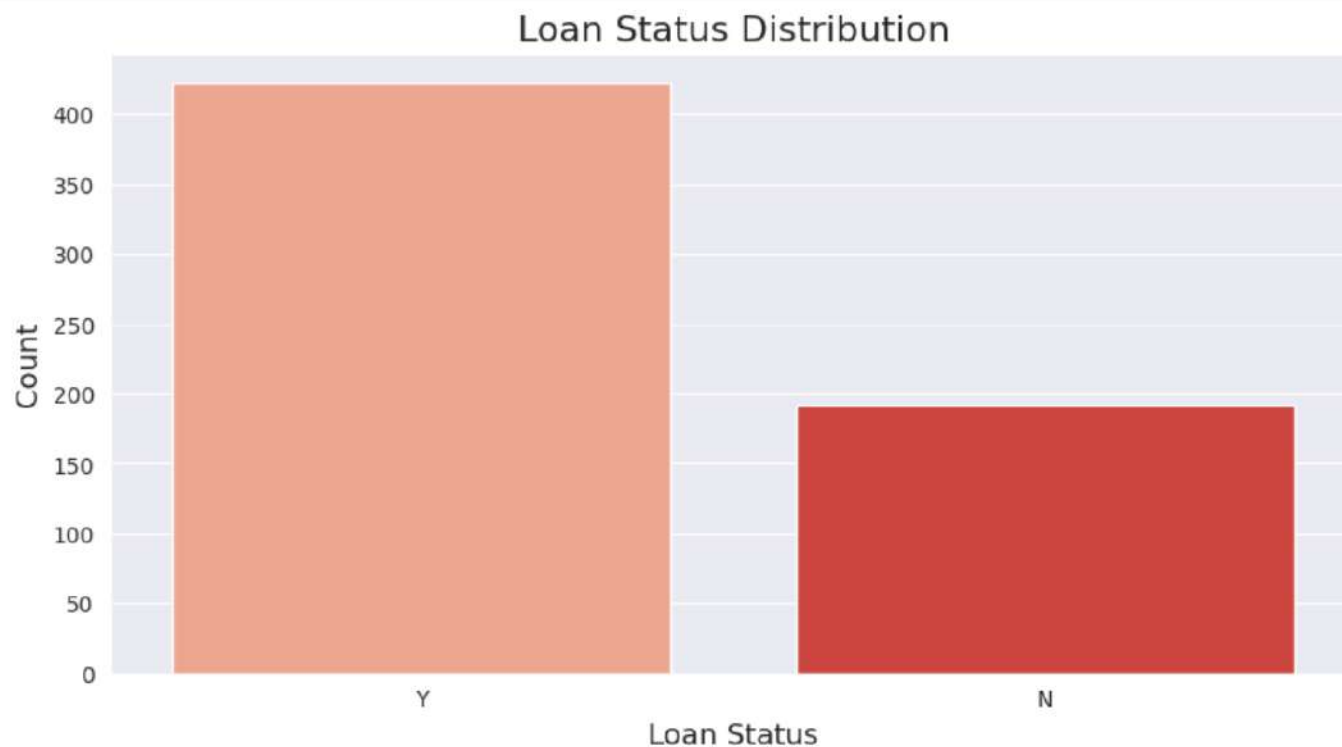
5:43 PM Python 3

28°C Partly cloudy

Search

ENG IN

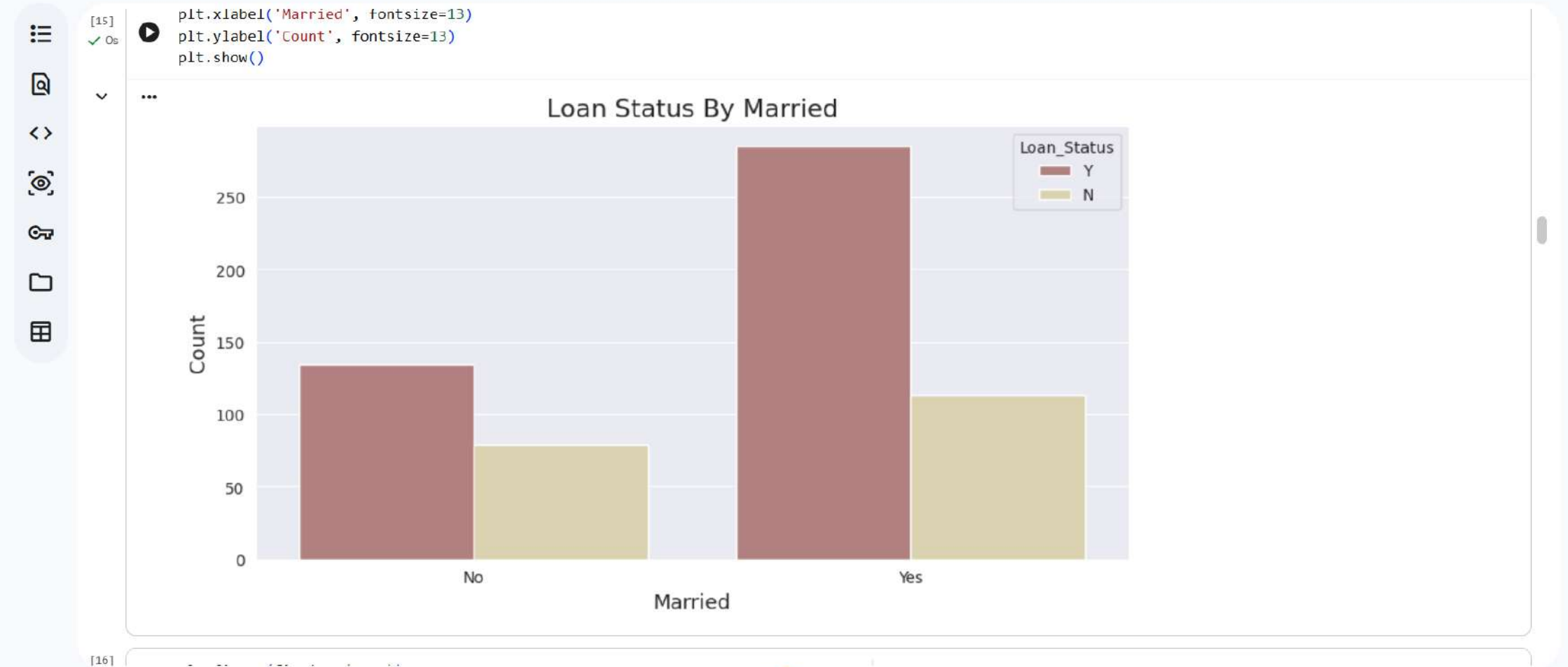
7:13 PM 12/13/2025



```
[14]  
✓ 0s  
plt.figure(figsize=(10,5))  
sns.countplot(x=df['Gender'], hue=df['Loan_Status'], palette='crest')  
plt.title('Loan Status By Gender', fontsize=16)  
plt.xlabel('Gender', fontsize=13)
```




```
[15]  
✓ Os  
plt.figure(figsize=(10,5))  
sns.countplot(x=df['Married'], hue=df['Loan_Status'], palette='pink')  
plt.title('Loan Status By Married', fontsize=16)  
plt.xlabel('Married', fontsize=13)
```



colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH38#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▶ Run all

RAM Disk

[16] ✓ 0s

```
plt.figure(figsize=(10,5))
sns.countplot(x=df['Property_Area'], hue=df['Loan_Status'], palette='magma')
plt.title('Loan Status By Property Area', fontsize=16)
plt.xlabel('Property Area', fontsize=13)
plt.ylabel('Count', fontsize=13)
plt.show()
```

Loan Status By Property Area

Property Area	Loan_Status Y	Loan_Status N
Urban	135	70
Rural	110	70
Semiurban	180	55

Variables Terminal

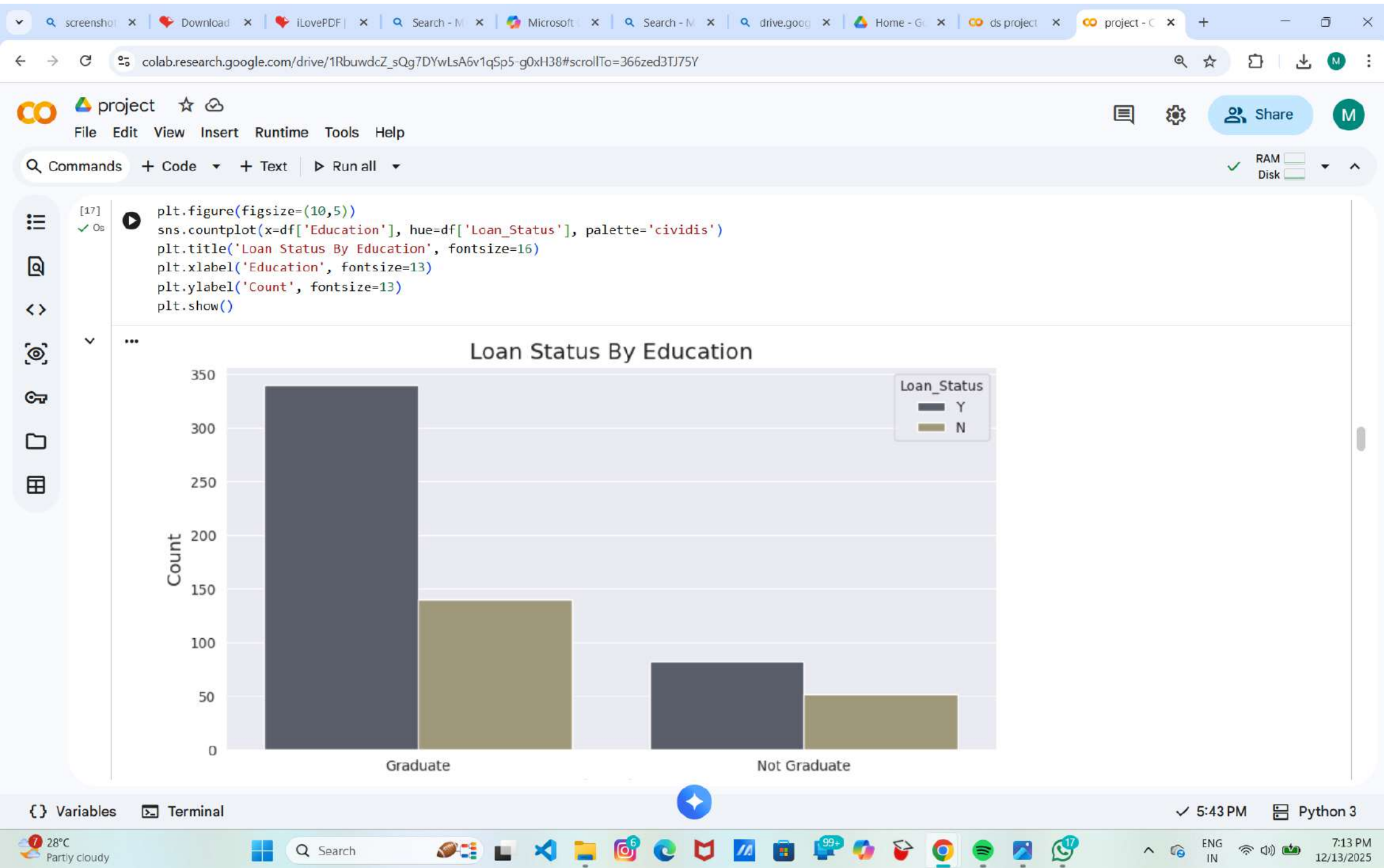
5:43 PM Python 3

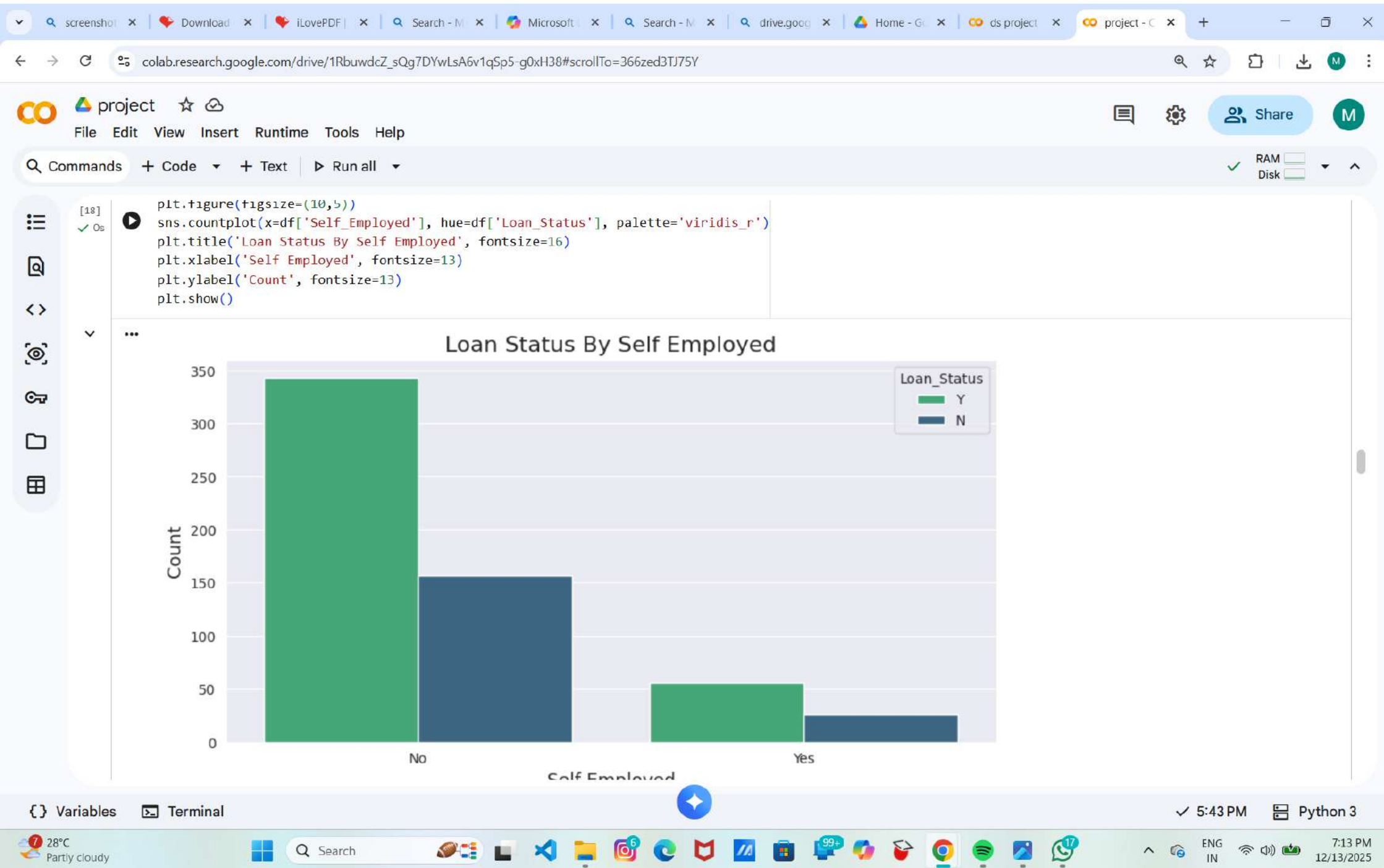
28°C Partly cloudy

Search

99+

7:13 PM 12/13/2025





colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH30#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▶ Run all

RAM Disk

```
[19] plt.figure(figsize=(10,5))
sns.scatterplot(x=df['ApplicantIncome'], y=df['Loan_Amount_Term'], hue=df['Loan_Status'], s=100)
plt.title('Applicant Income Vs Loan Amount Term', fontsize=16)
plt.xlabel('ApplicantIncome', fontsize=13)
plt.ylabel('Loan_Amount_Term', fontsize=13)
plt.show()
```

Applicant Income Vs Loan Amount Term

Loan_Status

- Y
- N

Variables Terminal

5:43 PM Python 3

28°C Partly cloudy

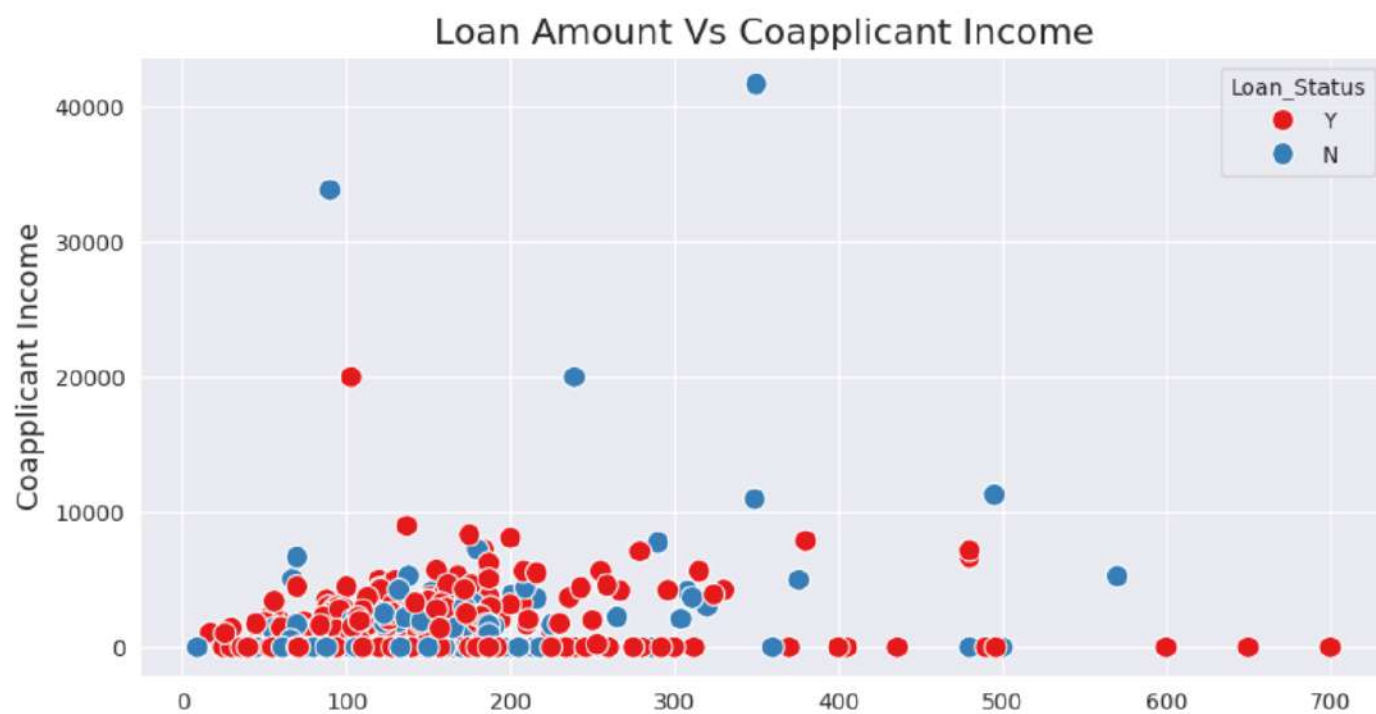
Search

99+

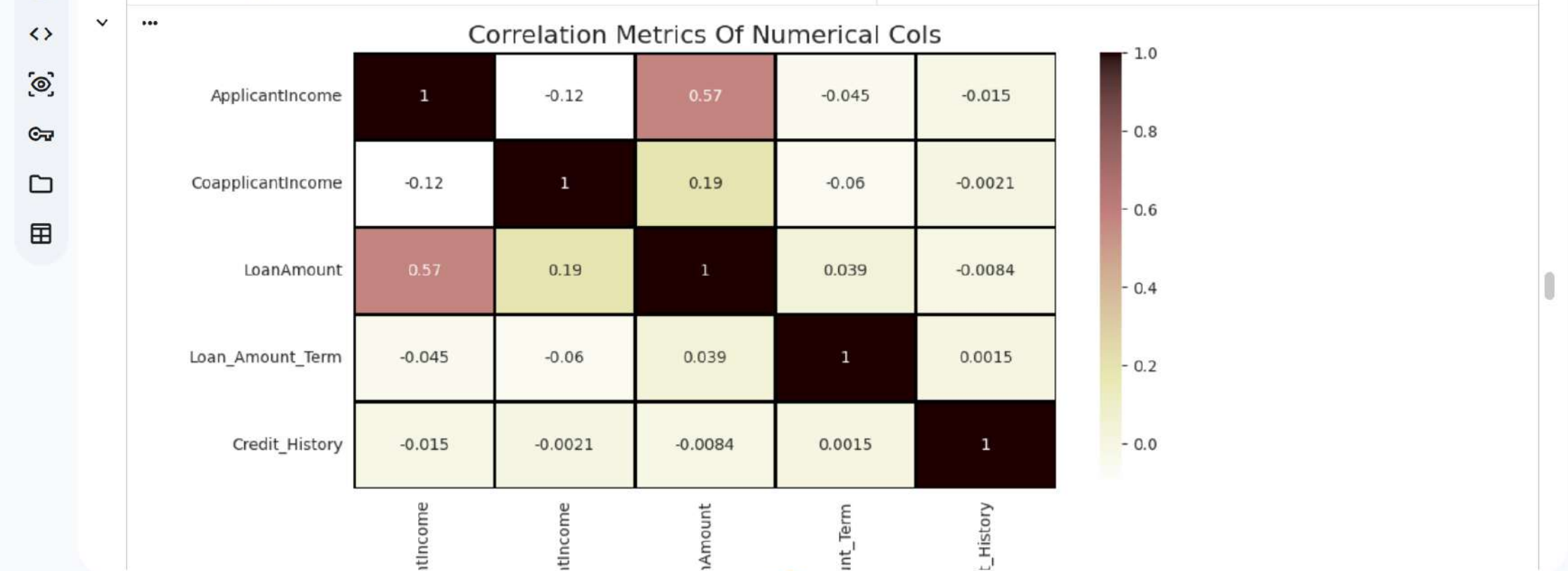
ENG IN

7:13 PM 12/13/2025

```
[20] plt.figure(figsize=(10,5))
sns.scatterplot(x=df['LoanAmount'], y=df['CoapplicantIncome'], hue=df['Loan_Status'], s=100, palette='Set1')
plt.title('Loan Amount Vs Coapplicant Income', fontsize=16)
plt.xlabel('Loan Amount', fontsize=13)
plt.ylabel('Coapplicant Income', fontsize=13)
plt.show()
```



```
[21] plt.figure(figsize=(10,5))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='pink_r', linewidths='2', linecolor='black')
plt.title('Correlation Metrics Of Numerical Cols', fontsize=16)
plt.show()
```



colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH38#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▶ Run all

Share

RAM Disk

ApplicantIncome

CoapplicantIncome

LoanAmount

Loan_Amount_Term

Credit_History

[22] ✓ 1s

```
plt.figure(figsize=(13,15))
for c, cols in enumerate(['ApplicantIncome', 'CoapplicantIncome', 'LoanAmount', 'Loan_Amount_Term',
                          'Credit_History']):
    plt.subplot(3, 2, c+1)
    sns.histplot(x=cols, data=df, color='blue', kde=True)
    plt.title(f"Distribution Of {cols}")
    plt.tight_layout(pad=4.0)

plt.show()
```

Distribution Of ApplicantIncome

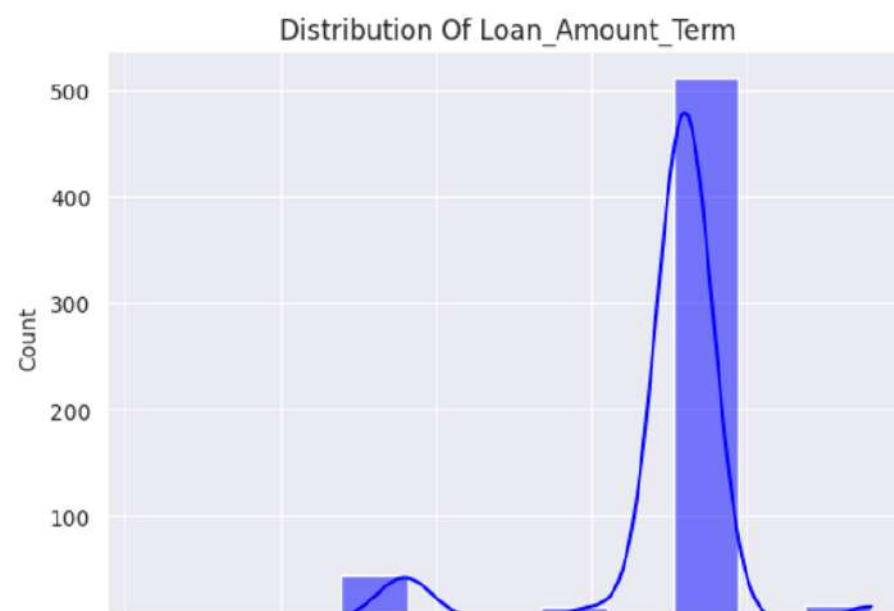
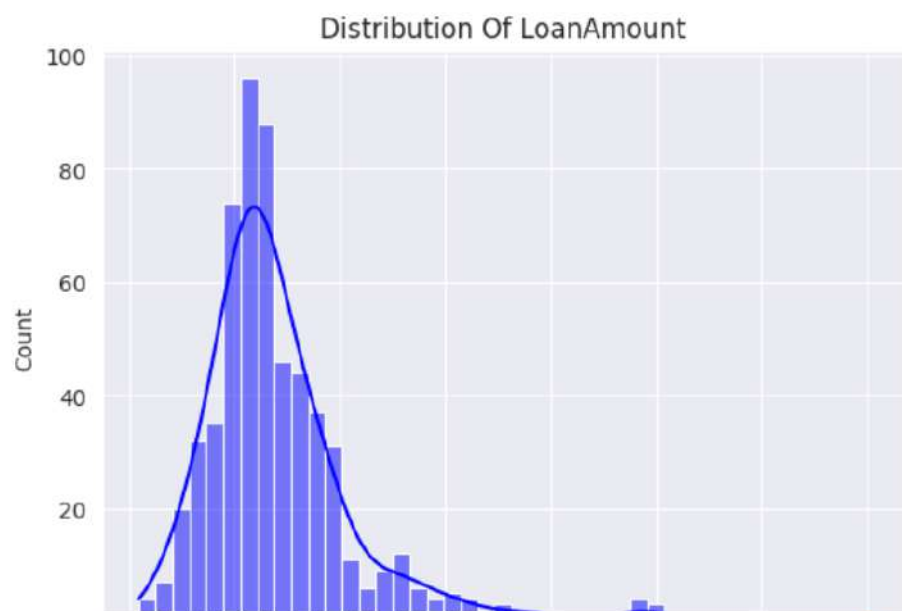
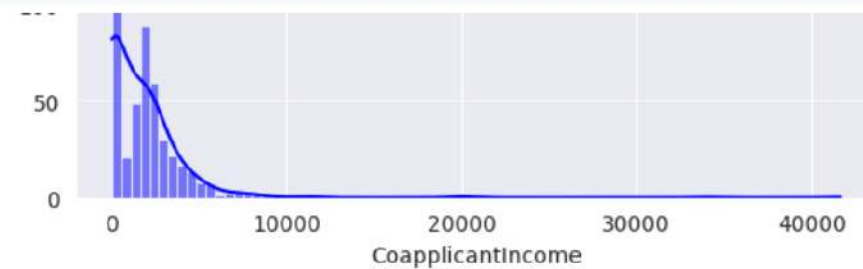
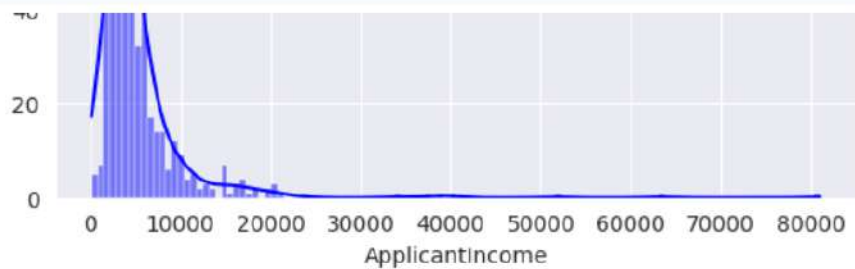
Distribution Of CoapplicantIncome

Variables Terminal

5:43 PM Python 3

28°C Partly cloudy

7:14 PM 12/13/2025



colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH38#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Share

RAM Disk

...

Distribution Of Credit_History



Credit_History	Count
0.0	~90
1.0	~480

[23] ✓ 1s

```
plt.figure(figsize=(13,15))
for c, cols in enumerate(['ApplicantIncome', 'CoapplicantIncome', 'LoanAmount', 'Loan_Amount_Term',
                           'Credit_History']):
    plt.subplot(3, 2, c+1)
    sns.kdeplot(x=cols, data=df, color='blue', shade=True)
    plt.title(f"Distribution Of {cols}")
```

Variables Terminal

5:43 PM Python 3

28°C Partly cloudy

Search

Taskbar icons: File Explorer, VS Code, Edge, Mail, etc.

System tray: ENG IN, 7:14 PM 12/13/2025

colab.research.google.com/drive/1RbuwdcZ_sQg7DYwLsA6v1qSp5-g0xH38#scrollTo=366zed3TJ75Y

project

File Edit View Insert Runtime Tools Help

Commands Code Text Run all

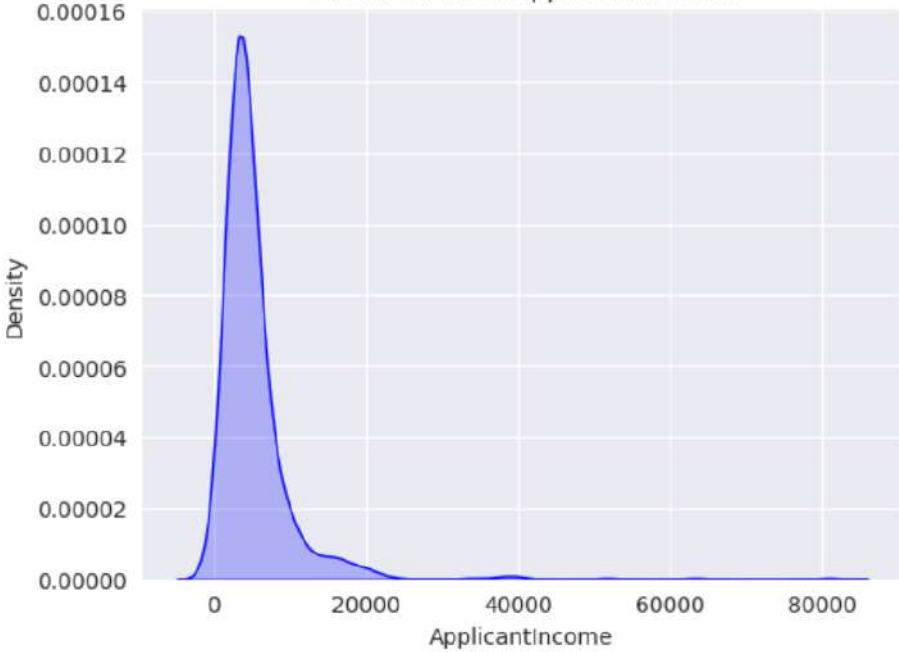
RAM Disk

[23] 1s

plt.tight_layout(pad=4.0)

plt.show()

Distribution Of ApplicantIncome

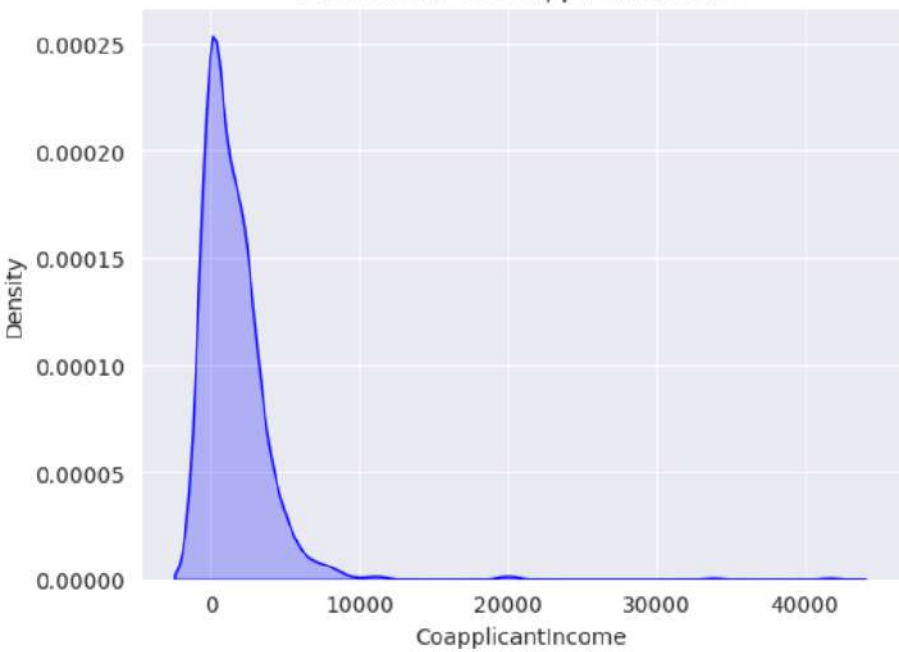


Density

ApplicantIncome

ApplicantIncome	Density
0	0.00000
5000	0.00015
10000	0.00005
20000	0.00000
40000	0.00000
60000	0.00000
80000	0.00000

Distribution Of CoapplicantIncome




Density

CoapplicantIncome


CoapplicantIncome	Density
0	0.00000
2000	0.00025
5000	0.00015
10000	0.00000
20000	0.00000
30000	0.00000
40000	0.00000

Distribution Of LoanAmount



Distribution Of LoanAmount

Distribution Of Loan_Amount_Term



Distribution Of Loan_Amount_Term

Variables Terminal

5:43 PM Python 3

28°C Partly cloudy

Search

7:14 PM 12/13/2025