

Progress Pre-Processing

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
path = "https://raw.githubusercontent.com/AzizLike29/loan-status-prediction/master/dataset/loan_data.csv"
```

```
df = pd.read_csv(path)
```

```
df.head(10)
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_Score
0	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	601
1	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	588
2	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	596
3	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	600
4	LP001013	Male	Yes	0	Not Graduate	No	2333	1516.0	95.0	360.0	590
5	LP001024	Male	Yes	2	Graduate	No	3200	700.0	70.0	360.0	600
6	LP001027	Male	Yes	2	Graduate	NaN	2500	1840.0	109.0	360.0	588
7	LP001029	Male	No	0	Graduate	No	1853	2840.0	114.0	360.0	590
8	LP001030	Male	Yes	2	Graduate	No	1299	1086.0	17.0	120.0	601
9	LP001032	Male	No	0	Graduate	No	4950	0.0	125.0	360.0	600

Next steps: [View recommended plots](#)

```
df.tail(7)
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_Score
374	LP002940	Male	No	0	Not Graduate	No	3833	0.0	120.0	360.0	588
375	LP002943	Male	No	NaN	Graduate	No	2987	0.0	120.0	360.0	588
376	LP002953	Male	Yes	3+	Graduate	No	5703	0.0	120.0	360.0	600
377	LP002974	Male	Yes	0	Graduate	No	3232	0.0	120.0	360.0	588
378	LP002978	Female	No	0	Graduate	No	2900	0.0	120.0	360.0	588
379	LP002979	Male	Yes	3+	Graduate	No	4106	0.0	120.0	360.0	600
380	LP002990	Female	No	0	Graduate	Yes	4583	0.0	120.0	360.0	601

```
df.sample(10)
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome
118	LP001669	Female	No	0	Not Graduate	No	1907	
361	LP002872	NaN	Yes	0	Graduate	No	3087	
161	LP001864	Male	Yes	3+	Not Graduate	No	4931	
64	LP001322	Male	No	0	Graduate	No	4133	
76	LP001398	Male	No	0	Graduate	NaN	5050	
378	LP002978	Female	No	0	Graduate	No	2900	
303	LP002522	Female	No	0	Graduate	Yes	2500	
147	LP001792	Male	Yes	1	Graduate	No	3315	
119	LP001671	Female	Yes	0	Graduate	No	3416	
197	LP002004	Male	No	0	Not Graduate	No	2927	

Check Missing Value

```
df.isna().sum()
```

```
Loan_ID      0
Gender       5
Married      0
Dependents   8
Education    0
Self_Employed 21
ApplicantIncome 0
CoapplicantIncome 0
LoanAmount   0
Loan_Amount_Term 11
Credit_History 30
Property_Area 0
Loan_Status  0
dtype: int64
```

Imputasi Missing Value

Type Data Numeric

```
df.Credit_History.fillna(df.Credit_History.mean(),inplace=True)
```

```
df.Credit_History.mean()
```

```
0.8376068376068375
```

```
df.Credit_History.fillna(df.Credit_History.median(),inplace=True)
```

```
df.Credit_History.median()
```

```
1.0
```

```
df.Credit_History.fillna(df.Credit_History.mode(),inplace=True)
```

```
df.Credit_History.mode()
```

```
0 1.0
Name: Credit_History, dtype: float64
```

Type Data Category

```
df['Credit_History'].fillna('S', inplace=True)
```

```
df.Credit_History.value_counts()
```

```

Credit_History
1.000000    294
0.000000    57
0.837607    30
Name: count, dtype: int64

```

✓ Exploratory Data Analysis (EDA)

```
data = pd.read_csv('https://raw.githubusercontent.com/AzizLike29/loan-status-prediction/master/dataset/loan_data.csv')
```

```

# Asumsikan 'data' memiliki kolom bernama 'Credit_History'
credit_history = data['Credit_History'].value_counts()

```

```

# Membuat bar plot
plt.figure(figsize=(8, 6))
sns.barplot(x=credit_history.index, y=credit_history.values, palette='viridis')
plt.title('Distribusi Credit History')
plt.xlabel('Credit_History')
plt.ylabel('Jumlah')
plt.xticks(ticks=[0, 1], labels=['Tidak Terlambat', 'Terlambat'])
plt.show()

```

```

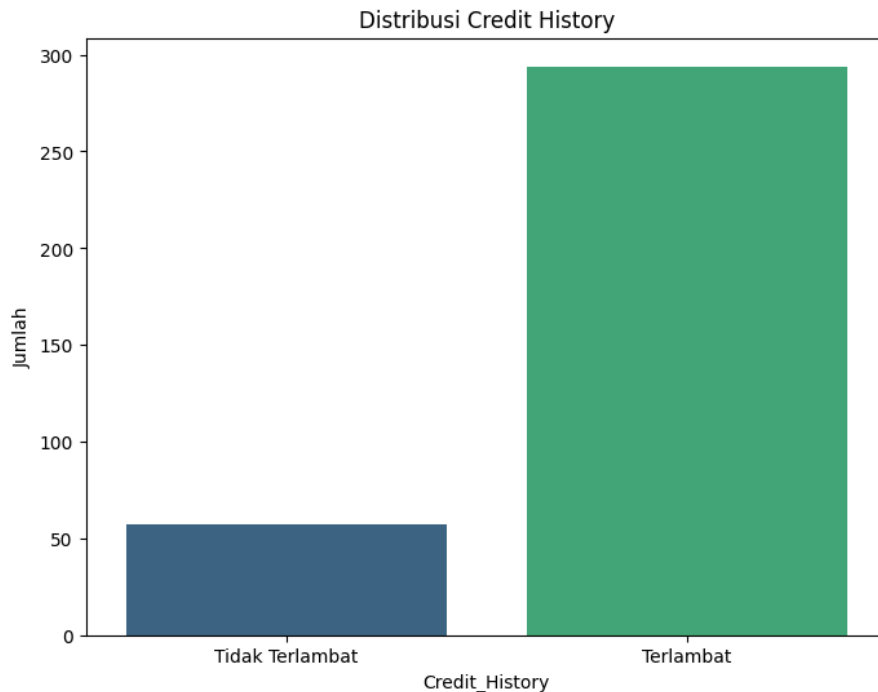
# Deskripsi
print("Deskripsi Grafik:")
print("- Credit History: Variabel kategorikal yang menunjukkan apakah pelamar memiliki riwayat kredit yang baik (1) atau buruk (0).")
print("- Jumlah: Jumlah pengamatan untuk setiap kategori Credit History.")
print("- Grafik menunjukkan distribusi Credit History dari dataset.")

```

```
<ipython-input-16-4b443f0ca5e5>:8: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0.

```
sns.barplot(x=credit_history.index, y=credit_history.values, palette='viridis')
```



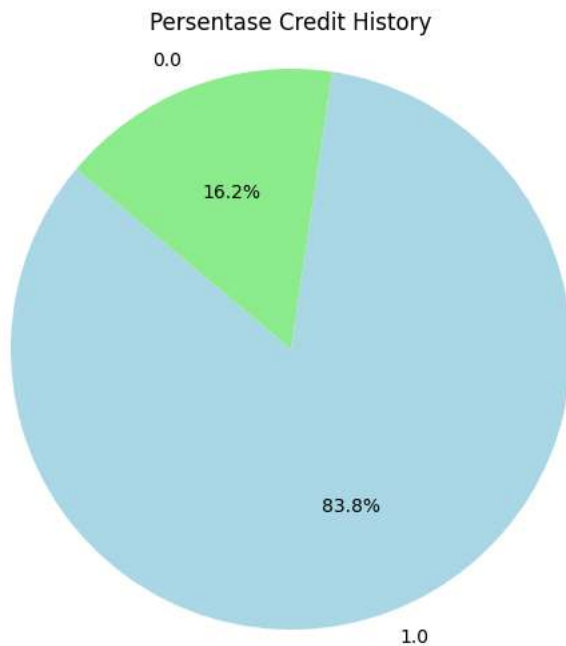
Deskripsi Grafik:

- Credit History: Variabel kategorikal yang menunjukkan apakah pelamar memiliki riwayat
- Jumlah: Jumlah pengamatan untuk setiap kategori Credit History.
- Grafik menunjukkan distribusi Credit History dari dataset.

```
# Menghitung jumlah nilai Credit History
credit_history = data['Credit_History'].value_counts()

# Membuat pie chart
plt.figure(figsize=(8, 6))
plt.pie(credit_history, labels=credit_history.index, autopct='%1.1f%%', colors=['lightblue', 'lightgreen'], startangle=140)
plt.title('Persentase Credit History')
plt.axis('equal') # Memastikan pie chart berbentuk lingkaran
plt.show()

# Deskripsi
print("Deskripsi Grafik:")
print("- Credit History: Variabel kategorikal yang menunjukkan apakah pelamar memiliki riwayat kredit yang baik (1) atau buruk (0).")
print("- Persentase: Persentase dari setiap kategori Credit History dalam dataset.")
print("- Grafik menunjukkan persentase Credit History dari dataset.")
```



Deskripsi Grafik:
 - Credit History: Variabel kategorikal yang menunjukkan apakah pelamar memiliki riwayat
 - Persentase: Persentase dari setiap kategori Credit History dalam dataset.

```
# Membuat cross-tabulation antara Credit History dan Loan Status
cross_tab = pd.crosstab(data['Credit_History'], data['Loan_Status'], margins=True)
```

```
# Plot bar chart
plt.figure(figsize=(10, 6))
cross_tab.drop('All').plot(kind='bar', stacked=True, color=['skyblue', 'lightgreen'])
plt.title('Hubungan antara Credit History dan Loan Status')
plt.xlabel('Credit History')
plt.ylabel('Jumlah')
plt.xticks(ticks=[0, 1], labels=['Tidak Terlambat', 'Terlambat'], rotation=0)
plt.legend(title='Loan Status')
plt.show()
```

```
# Deskripsi
print("Deskripsi Grafik:")
print("- Credit History: Variabel kategorikal yang menunjukkan apakah pelamar memiliki riwayat kredit yang baik (1) atau buruk (0).")
print("- Loan Status: Variabel kategorikal yang menunjukkan apakah pinjaman disetujui (Y) atau tidak (N).")
print("- Grafik menunjukkan hubungan antara Credit History dan Loan Status dari dataset.")
```

<Figure size 1000x600 with 0 Axes>

