In [41]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [42]:
```
df1 = pd.read_csv("german_credit_data.csv")
```

In [52]:
```
df1.head()
```

Out[52]:

	Unnamed: 0	Age	Sex	Job	Housing	Saving accounts	Checking account	Credit amount	Duration	Purpos
1	1	22	female	2	own	little	moderate	5951	48	radio/T
3	3	45	male	2	free	little	little	7882	42	furniture/equipmer
4	4	53	male	2	free	little	little	4870	24	ca
7	7	35	male	3	rent	little	moderate	6948	36	ca
9	9	28	male	3	own	little	moderate	5234	30	ca

In [36]:
```
df1.isnull().sum()
```

Out[36]:
```
Unnamed: 0            0
Age                   0
Sex                   0
Job                   0
Housing               0
Saving accounts     183
Checking account    394
Credit amount         0
Duration              0
Purpose               0
dtype: int64
```

In [44]:
```
df1.isnull().sum()
```

Out[44]:
```
Unnamed: 0            0
Age                   0
Sex                   0
Job                   0
Housing               0
Saving accounts     183
Checking account    394
Credit amount         0
Duration              0
Purpose               0
dtype: int64
```

In [45]:
```
dataframe1 = df1.dropna(subset=["Saving accounts"], inplace=True)
```

In [47]:
```
dataframe2 = df1.dropna(subset=["Checking account"], inplace = True)
```

In [48]:
```
df1.isnull().sum()
```

Out[48]:
```
Unnamed: 0          0
Age                 0
Sex                 0
Job                 0
Housing             0
Saving accounts     0
Checking account    0
Credit amount       0
Duration            0
Purpose             0
dtype: int64
```

In [50]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 522 entries, 1 to 999
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Unnamed: 0        522 non-null    int64 
 1   Age               522 non-null    int64 
 2   Sex               522 non-null    object
 3   Job               522 non-null    int64 
 4   Housing           522 non-null    object
 5   Saving accounts   522 non-null    object
 6   Checking account  522 non-null    object
 7   Credit amount     522 non-null    int64 
 8   Duration          522 non-null    int64 
 9   Purpose           522 non-null    object
dtypes: int64(5), object(5)
memory usage: 44.9+ KB
```

In [54]:
```
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
X1 = df1[['Sex','Housing','Job','Purpose']].values
ohe = ColumnTransformer([('anyname',OneHotEncoder(), [0])],remainder='passthrough
print(ohe.fit_transform(X1))
```

```
[[1.0 0.0 'own' 2 'radio/TV']
 [0.0 1.0 'free' 2 'furniture/equipment']
 [0.0 1.0 'free' 2 'car']
 ...
 [0.0 1.0 'own' 3 'car']
 [0.0 1.0 'free' 2 'radio/TV']
 [0.0 1.0 'own' 2 'car']]
```

In [55]:
```
from sklearn.model_selection import train_test_split
X1 = df.drop(columns=['Purpose'])
Y1 = df['Purpose']
X1_train, X1_test, y_train, y_test = train_test_split(X1,Y1, test_size=0.3)
```

In [56]:
```
print(X1_train.shape)
print(X1_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(365, 9)
(157, 9)
(365,)
(157,)
```