I suggest checking at what age people often take out loans and for how long. The database of portraits of German bank customers was taken from the Eberly College of Science page (Eberly College of Science): https://newonlinecourses.science.psu.edu/stat508/resource/analysis/gcd

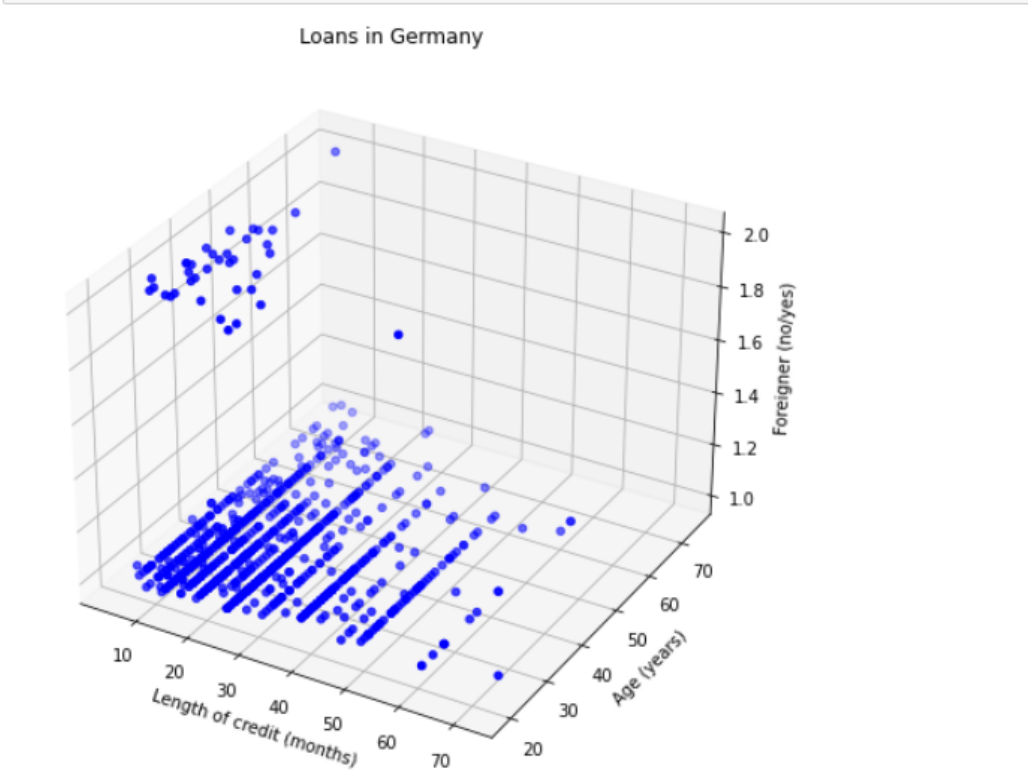
To describe the general portrait, 3 parameters were chosen to test the hypothesis: "working people are more likely to take credit" :

- duration of the loan;
- age of the client;
- whether the client is a foreigner.

We import the necessary libraries. NumPy allows to work with multidimensional arrays, for visualization we need matplotlib and for 3D graphics we need Axes3D module.

```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from matplotlib import pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

trans_set=pd.read_csv('german_credit.csv',
sep=',',header=None)  Data=trans_set.values

x = Data[:,2]
y = Data[:,13]
z = Data[:,20]

def scatter3d(x,y,z,c):
    fig = plt.figure(figsize=(8, 8))
    ax = fig.add_subplot(111, projection='3d')
    ax.scatter(x, y,z,c=c)
    ax.set_title('Loans in Germany')
    ax.set_xlabel('Length of credit (months)')
    ax.set_ylabel('Age (years)')
```

```
    ax.set_zlabel('Foreigner (no/yes)')
    plt.show()
print(type(x))
scatter3d(x,y,z,'blue')
```



From the visualisation we can already make hypotheses about the data. Firstly, most borrowers are German citizens. Secondly, if a foreigner takes out a loan, it is most likely a young person (under 45). We see only one exception in the form of 70+ year olds at the top end. Among citizens, the average age is under 65, with a slight increase in the share between 20-30 years old. Thirdly, people mostly take out loans for up to 48 months, i.e. up to 4 years. Preliminarily we can conclude that in Germany loans are more often taken out by young working citizens for a short term.

We can also divide borrowers into five sections, starting with a person's minimum age and their loan term.

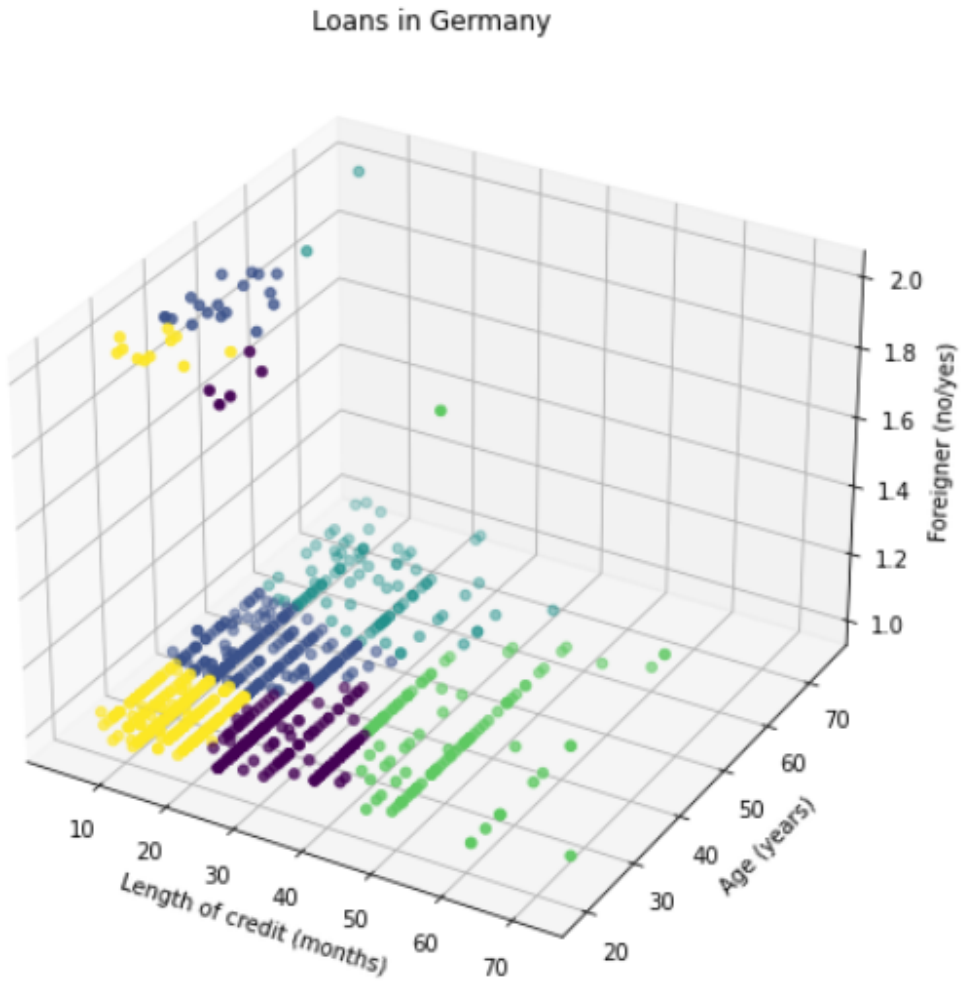
We will start by taking our client base and removing unnecessary information:

```
Clients = Data
Clients = np.delete(Clients, np.s_[0:2], axis=1)
Clients = np.delete(Clients, np.s_[1:11], axis=1)
Clients = np.delete(Clients, np.s_[2:8], axis=1)

kmeans = KMeans(n_clusters=5, random_state=0).fit(Clients) portraits = kmeans.cluster_centers_

for x in range (5):
    if(round(portraits[x][2])==1):
        resident = 'Citizen '
    else:
        resident ='Visitor '
    months = int(round(portraits[x][0]))
    years = round((months/12),1)
      print(resident+'at the age of '+ str(int(round(portraits[x][1])))+' years and a term loan '+str(months)+'months ('+str(years)+' years).')
```

```
Citizen at the age of 29 years and a term loan 27months (2.2 years).
Citizen at the age of 41 years and a term loan 14months (1.2 years).
Citizen at the age of 59 years and a term loan 18months (1.5 years).
Citizen at the age of 37 years and a term loan 45months (3.8 years).
Citizen at the age of 27 years and a term loan 13months (1.1 years).
```

```
labels = kmeans.labels_
color = list(range(len(Data)))

for x in range(len(Data)):
    if labels[x] == 0:
        color[x]= 'red'
    elif labels[x]== 1:
        color[x] ='yellow'
    elif labels[x] == 2:
        color[x] ='green'
    elif labels[x] == 3:
        color[x] ='blue'
    else:
        color[x]='black'
```

```
x = Clients[:,0]
y = Clients[:,1]
z = Clients[:,2]
scatter3d(x,y,z,c=labels)
```



The number of foreigners turned out to be so small that this property was hardly taken into account in the partitioning - they did not form a new group! By the way, we could have changed the dimensionality of the step on the z-axis, because we used only two values, but even now we can clearly see the division into two conditional planes. We have discussed typical representatives, now let us consider the groups themselves.

1. Young people who take short-term loans.

2. people without age division, who take only long-term loans (for more than two years).
3. middle-aged people, who take out loans for one to three years.
4. People of pre-retirement age, who take a loan for up to two and a half years.
5. Pensioners who take out a loan for up to three years. The data is now much easier to understand, isn't it