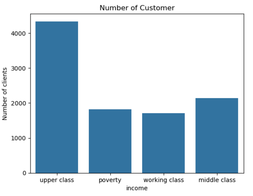
In [1]:
```
import pandas as pd
import numpy as np
import  matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df=pd.read_csv("customer_data.csv")
df.head()
```

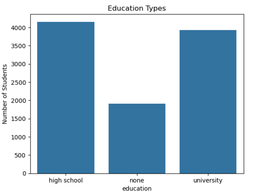
Out[2]:

	id	age	gender	race	driving_experience	education	income	credit_score	vehicle_ownership	vehicle_year	married	children	postal_code	annual_mileage	vehicle_type	speeding_violations	DUIs	past_accidents	outcome
0	569520	65+	female	majority	0-9y	high school	upper class	0.629027	True	after 2015	False	True	10238	12000.0	sedan	0	0	0	False
1	750365	16-25	male	majority	0-9y	none	poverty	0.357757	False	before 2015	False	False	10238	16000.0	sedan	0	0	0	True
2	199901	16-25	female	majority	0-9y	high school	working class	0.493146	True	before 2015	False	False	10238	11000.0	sedan	0	0	0	False
3	478866	16-25	male	majority	0-9y	university	working class	0.206013	True	before 2015	False	True	32765	11000.0	sedan	0	0	0	False
4	731664	26-39	male	majority	10-19y	none	working class	0.388366	True	before 2015	False	False	32765	12000.0	sedan	2	0	1	True

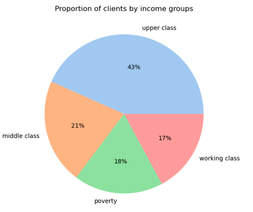
In [9]:
```
sns.countplot(data=df,x="income")
plt.title("Number of Customer")
plt.ylabel("Number of clients")
plt.show()
```



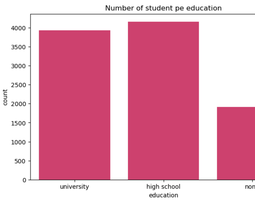
In [10]:
```
sns.countplot(data=df,x="education")
plt.title("Education Types")
plt.ylabel("Number of Students")
plt.show()
```



In [11]:
```
plt.figure(figsize=(6,6))
data=df["income"].value_counts(normalize=True)
labels=["upper class","middle class","poverty","working class"]
colors=sns.color_palette("pastel")
plt.pie(data,labels=labels,colors=colors,autopct="%.0f%%")
plt.title("Proportion of clients by income groups")
plt.show()
```



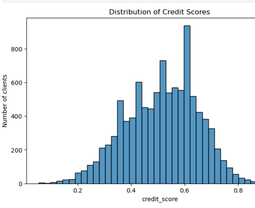
In [12]:
```
plt.figure(figsize=(8,5))
sns.countplot(data=df,x="education",order=["university","high school","none"],color="#E42C6A")
plt.title("Number of student pe education")
plt.show()
```



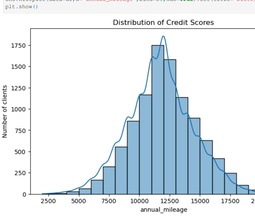
In [13]:
```
df["credit_score"].describe()
```

Out[13]:
```
count    10000.000000
mean         0.515247
std          0.134628
min          0.053358
25%          0.421301
50%          0.523659
75%          0.613827
max          0.960819
Name: credit_score, dtype: float64
```

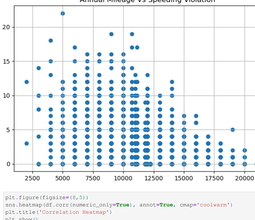
In [14]:
```
plt.figure(figsize=(8,5))
sns.histplot(data=df,x="credit_score",bins=40).set(title="Distribution of Credit Scores",ylabel="Number of clients")
plt.show()
```



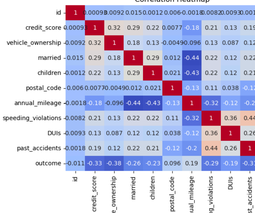
In [15]:
```
plt.figure(figsize=(8,5))
sns.histplot(data=df,x="annual_mileage",bins=20,kde=True).set(title="Distribution of Credit Scores",ylabel="Number of clients")
plt.show()
```



In [16]:
```
plt.figure(figsize=(8,5))
plt.scatter(data=df,x="annual_mileage",y="speeding_violations")
plt.title("Annual Mileage Vs Speeding Violation")
plt.grid(True)
plt.show()
```



In [17]:
```
plt.figure(figsize=(8,5))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



In [20]:
```
plt.figure(figsize=(8,5))
sns.boxplot(x='income', y='age', data=df, hue='income', palette='Set2', legend=False)
plt.title('Age Distribution by Gender')
plt.show()
```

In [ ]: