

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

```
In [2]: data = pd.read_csv('https://raw.githubusercontent.com/d0r1h/Statistical-Learn:
```

```
In [3]: data.head()
```

```
Out[3]:
```

	ID	Income	Limit	Rating	Cards	Age	Education	Gender	Student	Married	Ethnicity	Balance
0	1	14.891	3606	283	2	34	11	Male	No	Yes	Caucasian	333
1	2	106.025	6645	483	3	82	15	Female	Yes	Yes	Asian	903
2	3	104.593	7075	514	4	71	11	Male	No	No	Asian	580
3	4	148.924	9504	681	3	36	11	Female	No	No	Asian	964
4	5	55.882	4897	357	2	68	16	Male	No	Yes	Caucasian	331

Quantitative and Quantitative

```
In [4]: data_num = data.select_dtypes(np.number)
data_num.head()
```

```
Out[4]:
```

	ID	Income	Limit	Rating	Cards	Age	Education	Balance
0	1	14.891	3606	283	2	34	11	333
1	2	106.025	6645	483	3	82	15	903
2	3	104.593	7075	514	4	71	11	580
3	4	148.924	9504	681	3	36	11	964
4	5	55.882	4897	357	2	68	16	331

1. Scatter plot

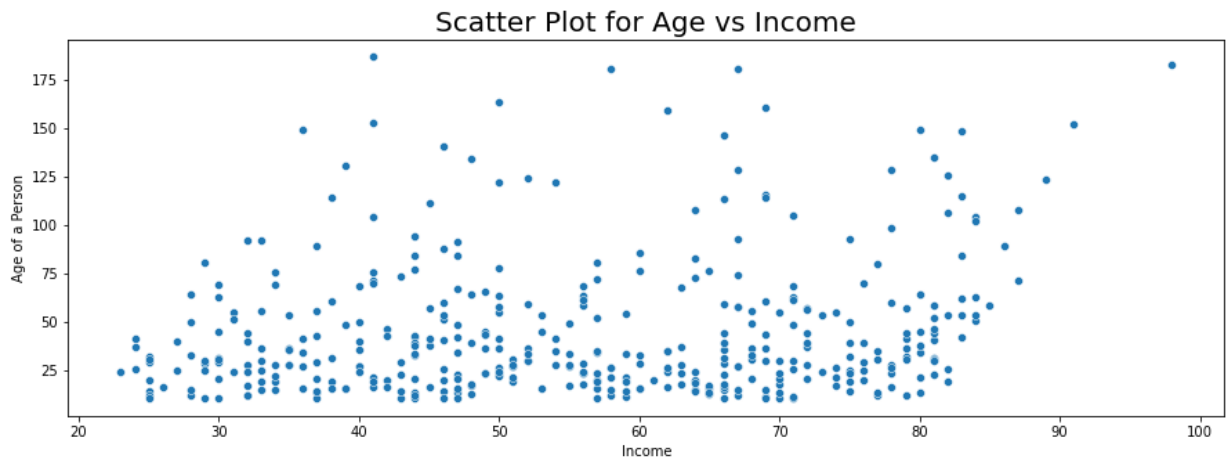
Let's start the exploring the relationship between Age and Income.

```
In [5]: plt.rcParams['figure.figsize'] = [15,5]

ax = sns.scatterplot(x='Age', y='Income', data=data_num)

ax.set_ylabel('Age of a Person', fontsize=10)
ax.set_xlabel('Income', fontsize=10)
plt.title("Scatter Plot for Age vs Income", fontsize=20)

plt.show()
```

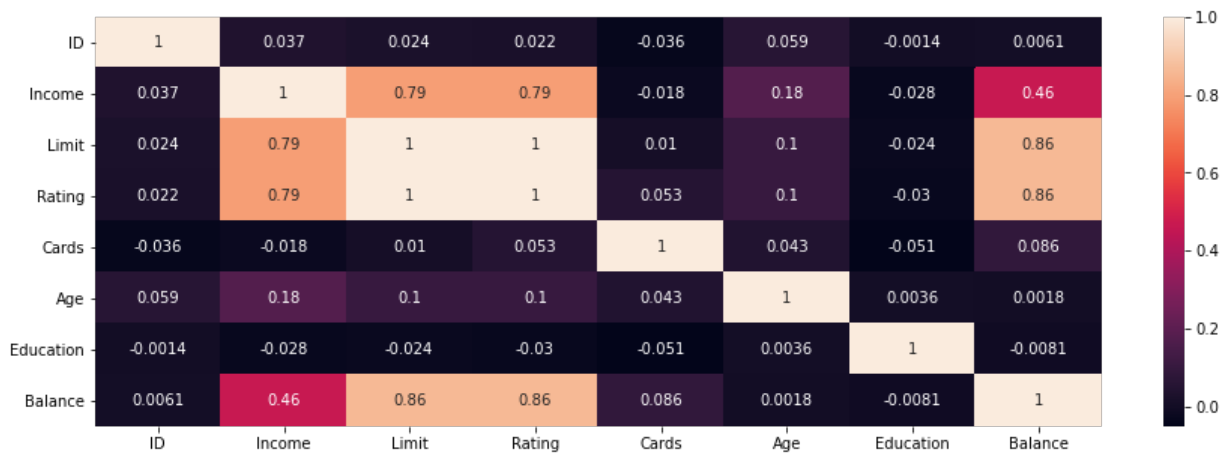


2. Correlation

```
In [6]: plt.rcParams['figure.figsize'] = [15,5]

sns.heatmap(data_num.corr(), annot=True)

plt.show()
```



Here, we can observe that correlation between Age and Income is not strong, same we found through scatter plot.

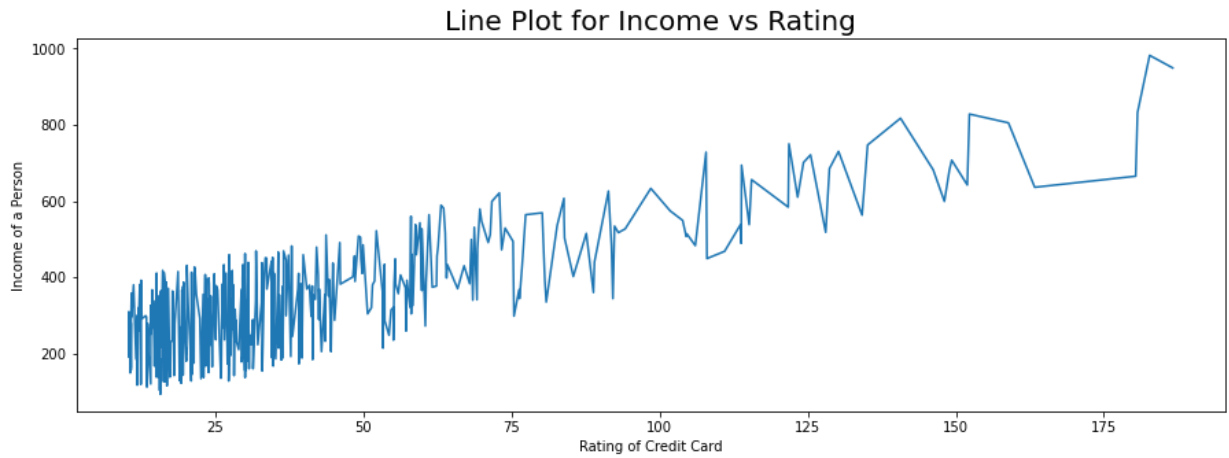
3. Line Plot

```
In [7]: plt.rcParams['figure.figsize'] = [15,5]

ax = sns.lineplot(x='Income', y='Rating', data=data_num)

ax.set_ylabel('Income of a Person', fontsize=10)
ax.set_xlabel('Rating of Credit Card', fontsize=10)
plt.title("Line Plot for Income vs Rating", fontsize=20)

plt.show()
```



Here, we can observe that as the person Income increase their credit card rating also increases.

Categorical and Quantitative

4. Bar Plot

Impact of Marital Status on Credit Rating.

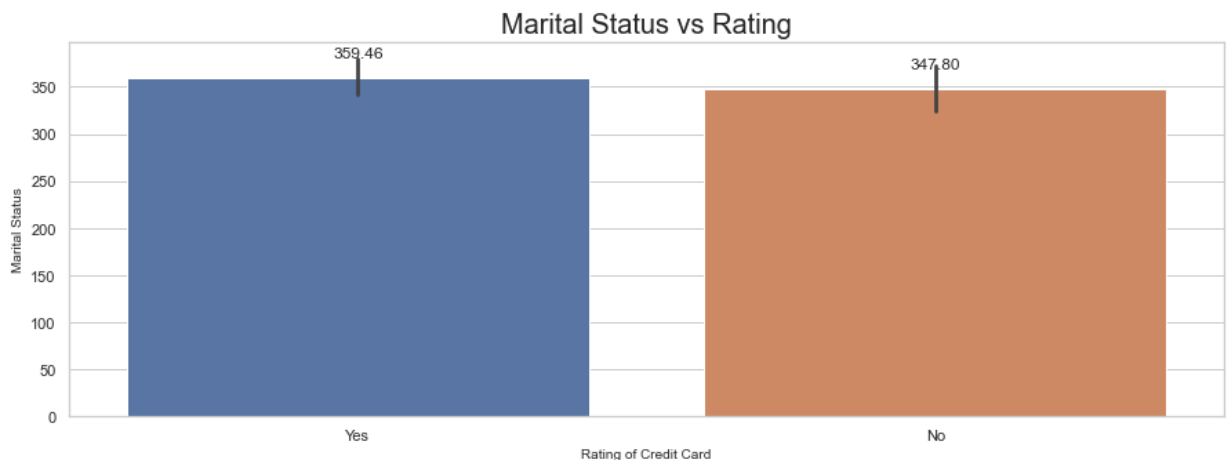
```
In [8]: plt.rcParams['figure.figsize'] = [15,5]
sns.set_theme(style='whitegrid')

ax = sns.barplot(x='Married', y='Rating', data=data)

for p in ax.patches:
    ax.annotate(format(p.get_height(), '.2f'), (p.get_x() + p.get_width() / 2,
        xytext = (0, 15), textcoords = 'offset points')

ax.set_ylabel('Marital Status', fontsize=10)
ax.set_xlabel('Rating of Credit Card', fontsize=10)
plt.title("Marital Status vs Rating", fontsize=20)

plt.show()
```



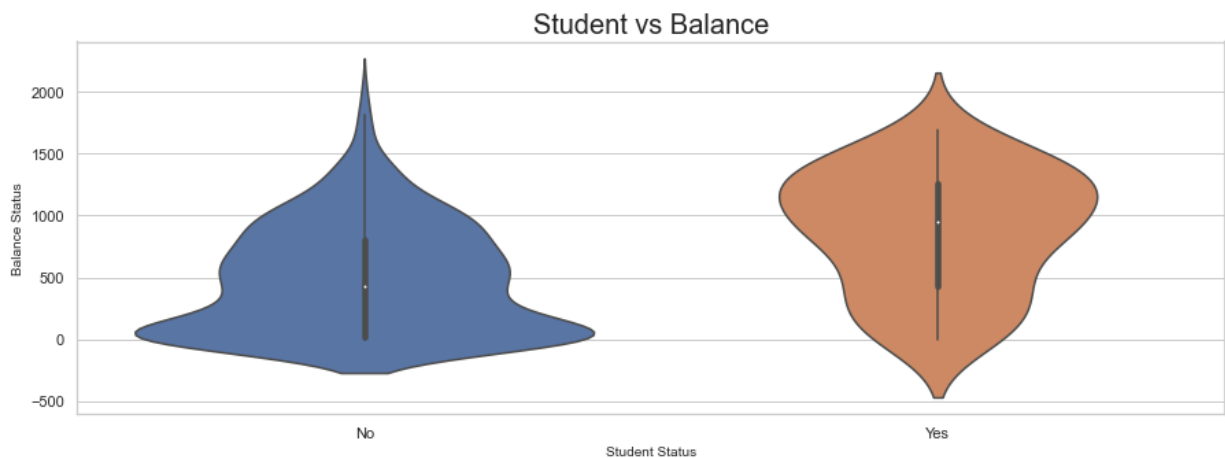
5. Violin Plot

```
In [9]: plt.rcParams['figure.figsize'] = [15,5]

ax = sns.violinplot(x=data.Student, y=data.Balance)

ax.set_xlabel('Student Status', fontsize=10)
ax.set_ylabel('Balance Status', fontsize=10)
plt.title("Student vs Balance", fontsize=20)

plt.show()
```



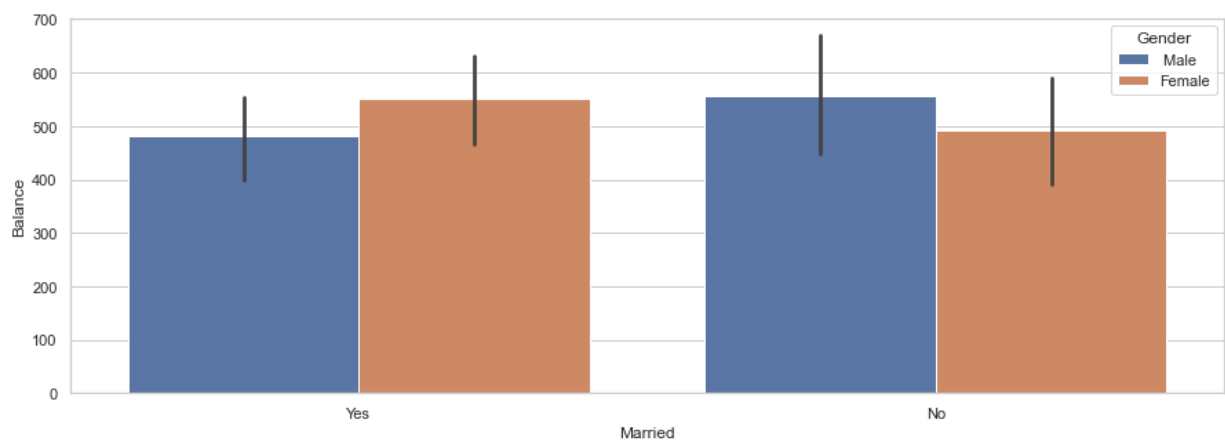
Categorical and Categorical

6. Grouped Bar Plot

We can use grouped bar plot to analyze two different categorical variables.

```
In [10]: plt.rcParams['figure.figsize'] = [15,5]

ax = sns.barplot(x='Married', y='Balance', hue='Gender', data=data)
```



7. Pivot Table

```
In [11]: pd.pivot_table(data, index=['Married', 'Ethnicity'], aggfunc = np.mean)
```

Out [11]:

		Age	Balance	Cards	Education	ID	Income	Limit
Married	Ethnicity							
No	African American	60.510638	514.595745	2.723404	13.468085	209.042553	48.565404	4935.404255
	Asian	56.468750	563.343750	2.812500	13.406250	190.062500	41.703250	4625.250000
	Caucasian	55.565789	511.802632	3.197368	13.065789	192.960526	41.411724	4474.342105
Yes	African American	54.788462	545.826923	3.173077	13.576923	175.865385	46.883731	4832.961538
	Asian	52.742857	488.985714	3.042857	13.700000	187.942857	45.323643	4599.857143
	Caucasian	55.707317	522.634146	2.796748	13.495935	222.170732	46.443707	4885.479675

8. Stacked Bar Plot

In [12]:

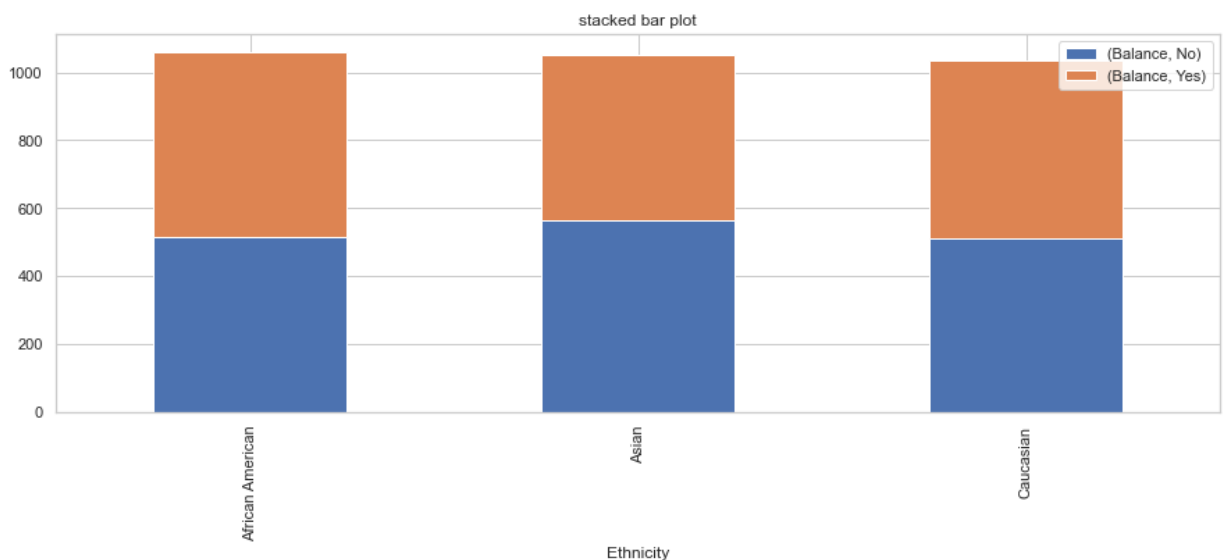
```
stack = pd.pivot_table(data, index=['Ethnicity'], columns=['Married'], values=stack)
```

Out [12]:

	Balance	
Married	No	Yes
Ethnicity		
African American	514.595745	545.826923
Asian	563.343750	488.985714
Caucasian	511.802632	522.634146

In [13]:

```
plt.rcParams['figure.figsize'] = [15,5]
ax = stack.plot.bar(title='stacked bar plot', stacked=True, label='Ethnicity')
ax.legend()
plt.show()
```



9. Crosstab

```
In [14]: pd.crosstab(index=data['Married'], columns= data['Ethnicity'])
```

```
Out[14]: Ethnicity  African American  Asian  Caucasian
Married
No          47      32      76
Yes         52      70     123
```

```
In [15]: pd.crosstab(index=data['Student'], columns= data['Ethnicity'])
```

```
Out[15]: Ethnicity  African American  Asian  Caucasian
Student
No          89      89     182
Yes         10      13      17
```