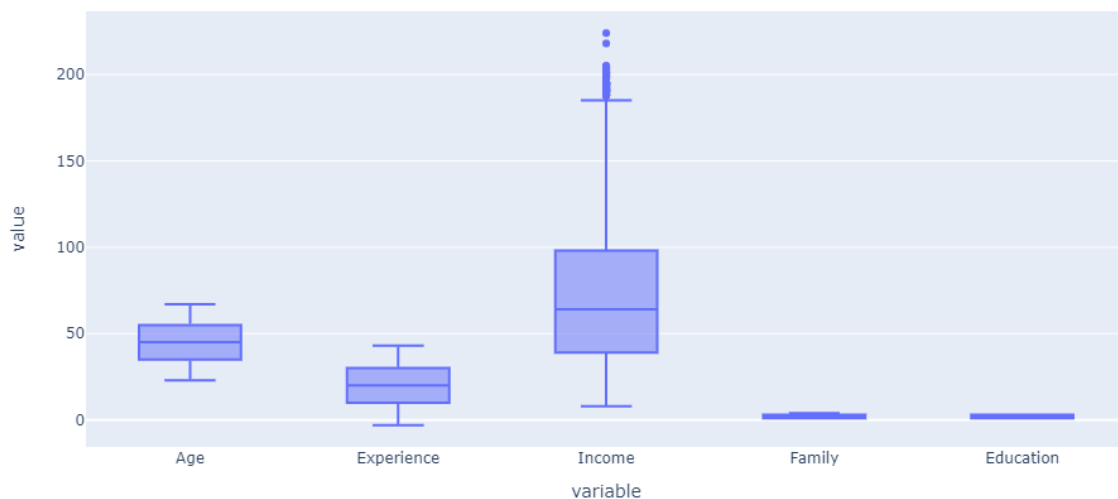


Report Analysis of Bank Loan Dataset

Applying five-point summary concept on the data. It provides a concise summary of the distributions of the observations. It gives the information about the location (median), the spread (from the quartiles) and range (minimum and maximum of the observation). These five points are:

- 1) Sample Minimum
- 2) Lower or first quartile (greater than 25% of data less than 75% of data)
- 3) The middle value
- 4) Upper or third quartile (greater than 75% of data less than the remaining 25% of data)
- 5) Sample maximum

Five-point summary of Age, Experience, Income, Family, Education



- Some Experience values are negative which is not possible.
- We can observe some outliers in data in the income bracket as it is rightly skewed.
- Age and Experience are equally distributed.

Age	-0.029341
Experience	-0.026325
Income	0.841339
Family	0.155221
CCAvg	1.598457
Education	0.227093
Mortgage	2.104002
Personal Loan	2.743607
Securities Account	2.588268
CD Account	3.691714
Online	-0.394785
CreditCard	0.904589
dtype: float64	

Figure 1 Skewness of Different Features of Data

- When the value of the skewness is negative, the tail of the distribution is longer towards the left-hand side of the curve.
- When the value of the skewness is positive, the tail of the distribution is longer towards the right-hand side of the curve.
- Conclusion: large number of positive outliers in income and a smaller number of low outliers in experience.

Histogram Plots of Various Features in Data:

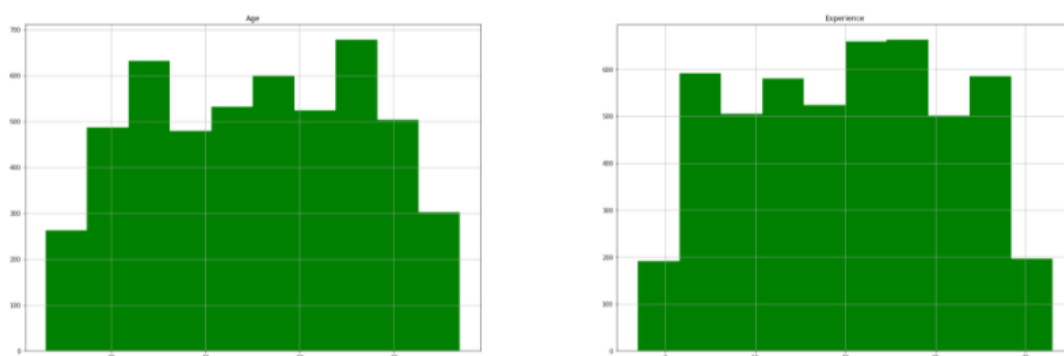


Figure 2 Histogram Plot of Experience & Age

- Age and Experience are equally distributed.

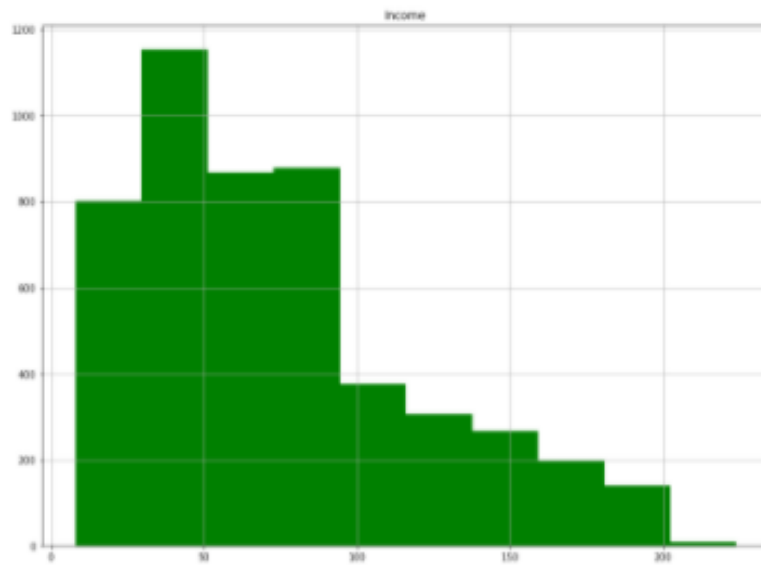


Figure 3 Histogram Plot of Income of Customers

- Income graph is left skewed with positive outliers.

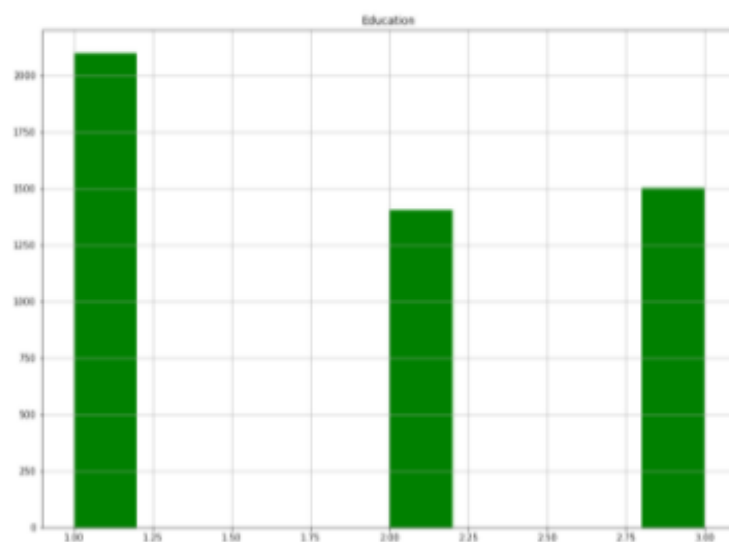


Figure 4 Histogram Plot of Education Status of Customers

- We have more undergraduates than graduates and professionals.

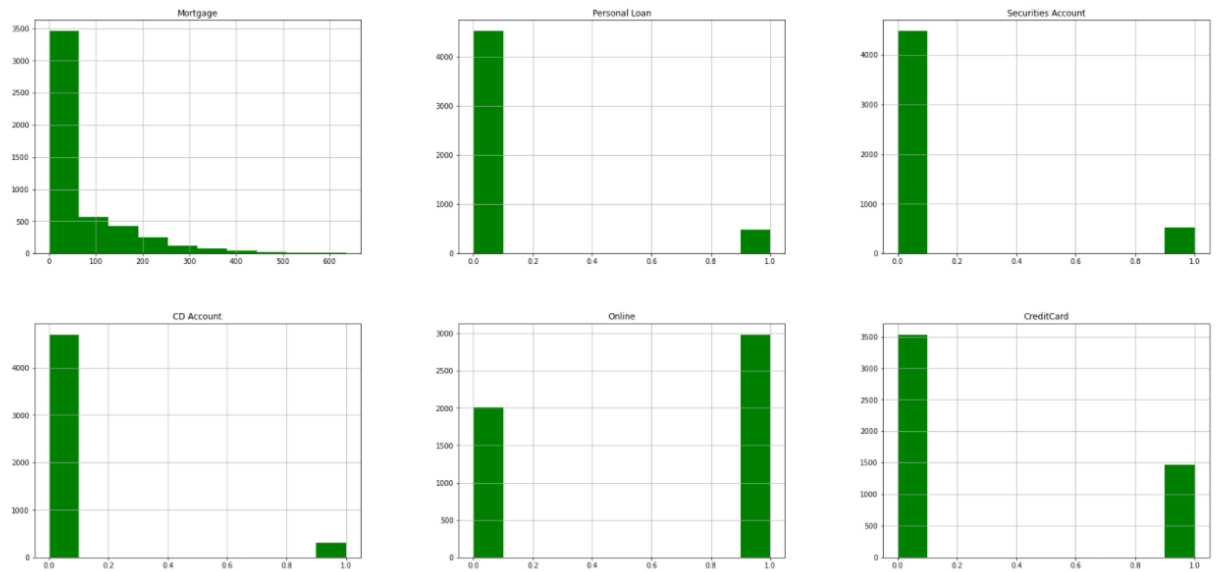
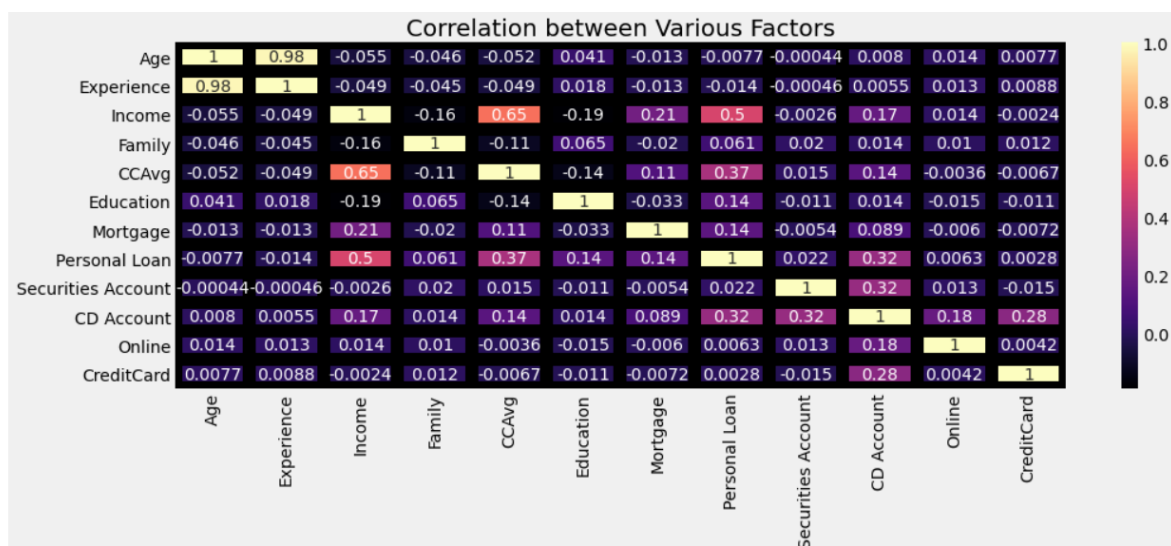


Figure 5 Histogram plot of Different Features

Inferences:

- 1) Number of Personal Loans taken are less. (Count=480, 9.6%)
- 2) Number of Credit Cards issued are less. (Count=1470, 29.4%)
- 3) Number of CD accounts are very less. (Count=302, 6.04%)
- 4) Number of Security Account (Count=522, 10.44%)
- 5) Number of Bank Users Online (Count=2984, 59.68%)

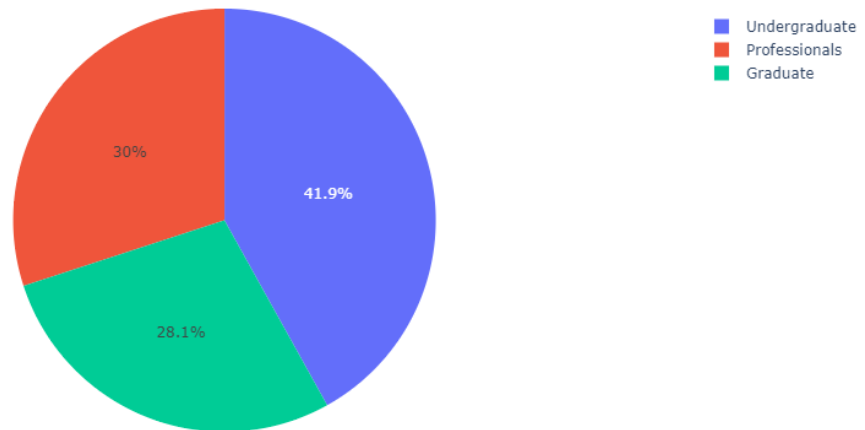
Correlation Heatmap:



- Correlation is the measure of how two or more variables are related to one another.
- Experience and Age are highly correlated, this means that we can disregard one of the features for further analysis.

Analyzing Education Status of Customers:

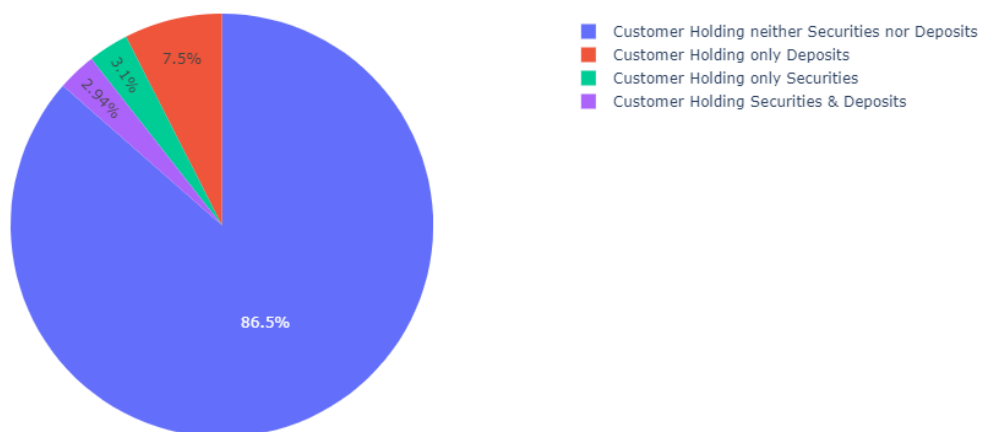
Education



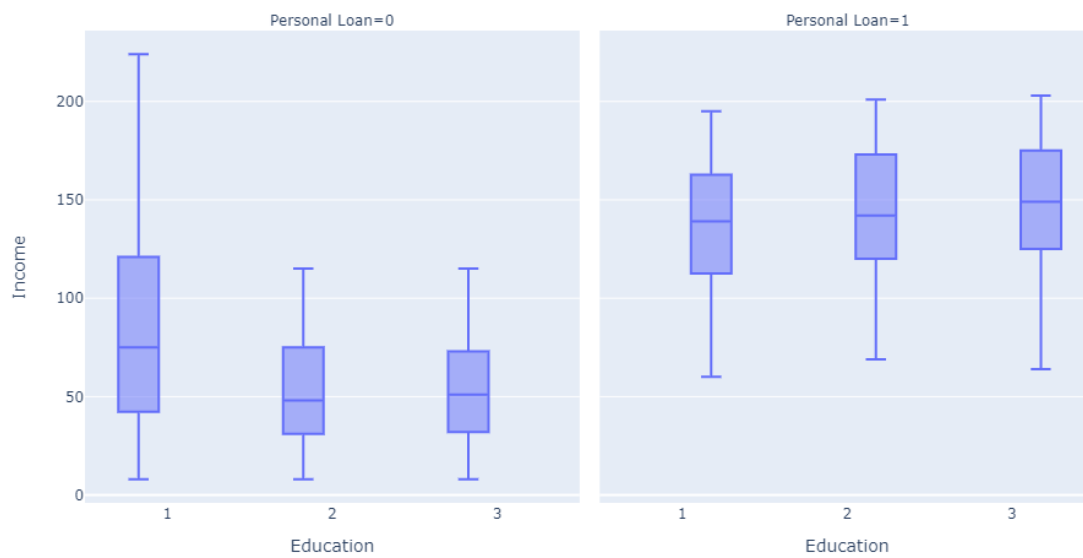
- This pie-chart is created by mapping Education status on the data and grouping it with respect to Age.
- We can observe that maximum number of customers are undergraduates, 41.9%, 30% professionals and 28.1% graduates.

Account Holder Information & Grouping:

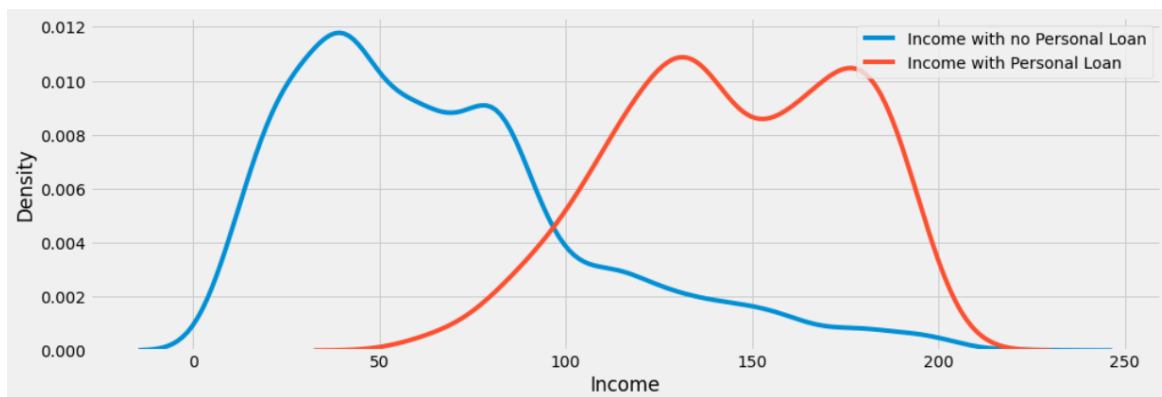
Account Holder Information



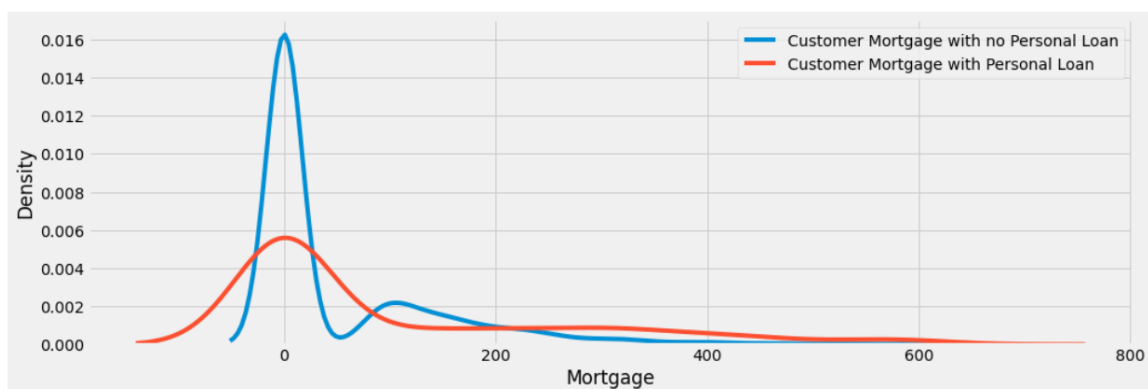
- This pie chart is created by grouping together data from Securities Account & CD account.
- Majority of customers do not have securities and deposits. Banks can encourage opening of such accounts to increase profits.



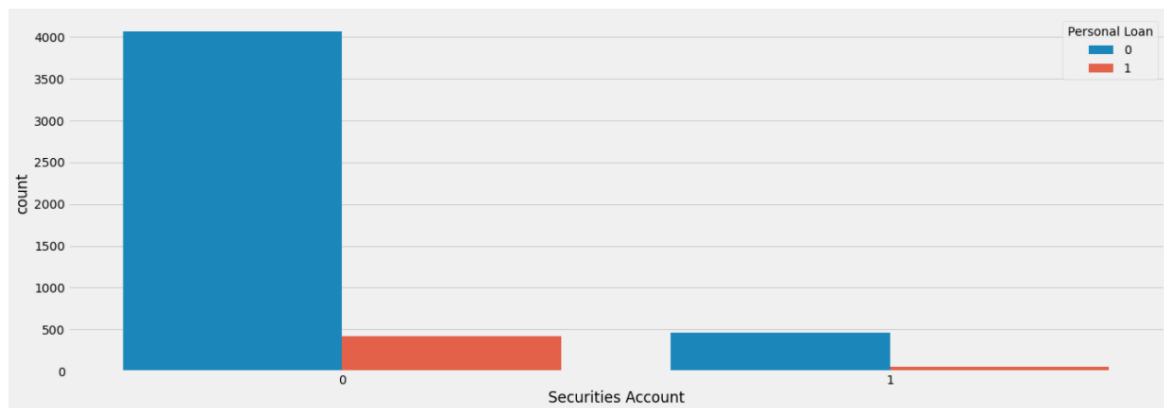
- Customers who have taken Personal Loans fit in to the same income bracket irrespective of their education.



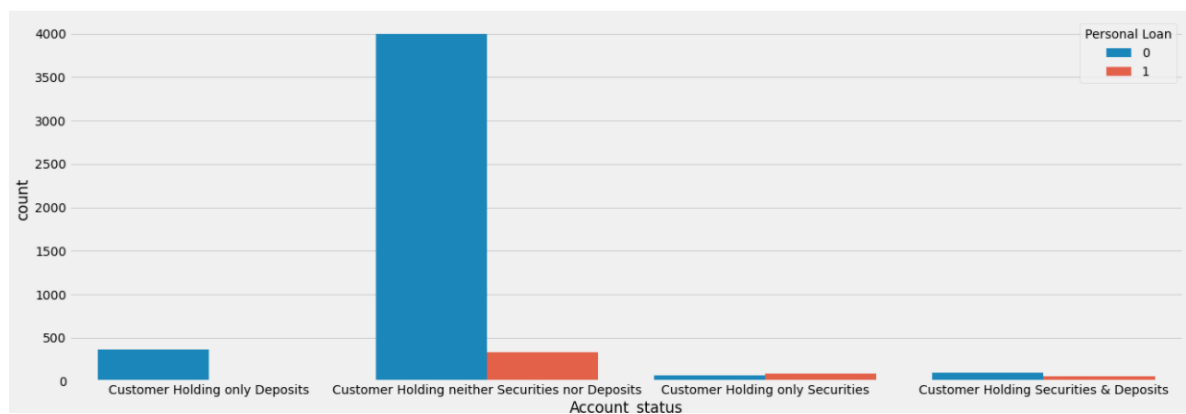
- Customer having high income is more likely to take Personal Loan.



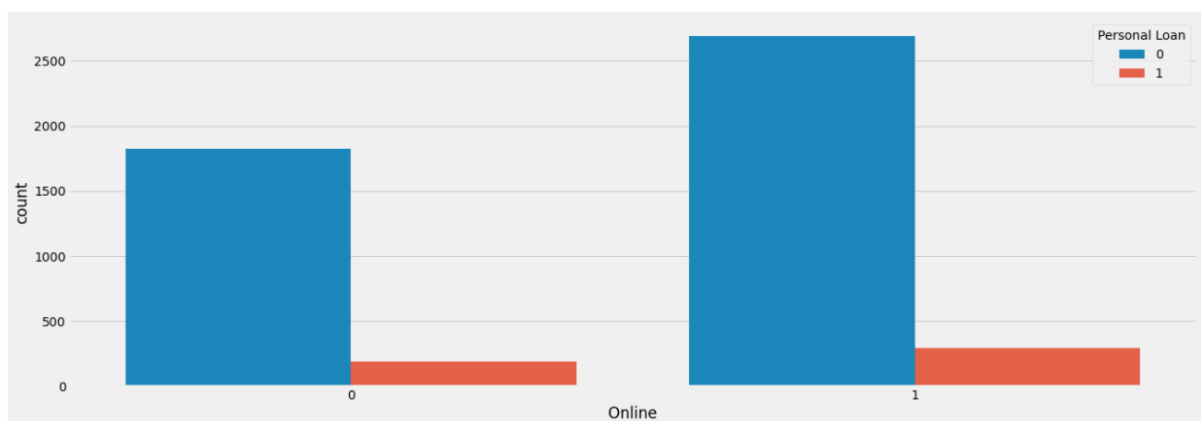
- Customer having Mortgage value greater than 500k is more likely to have a Personal Loan.

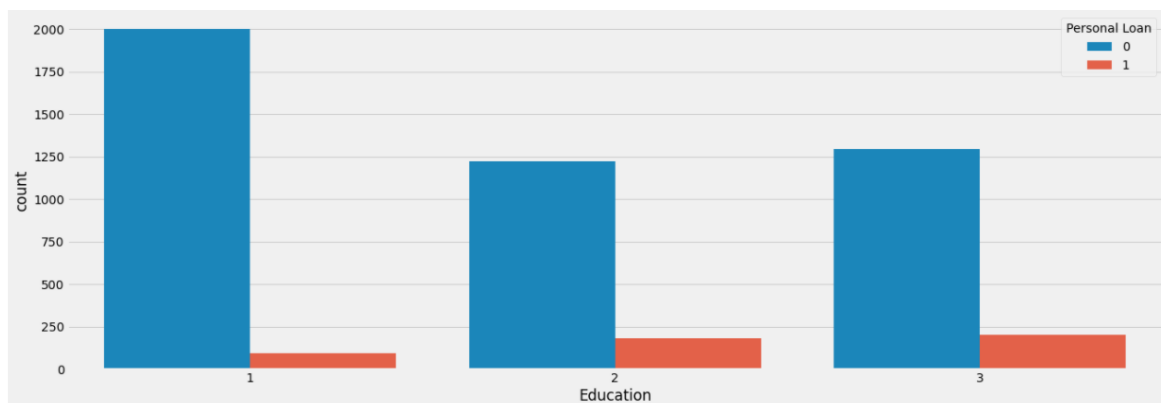
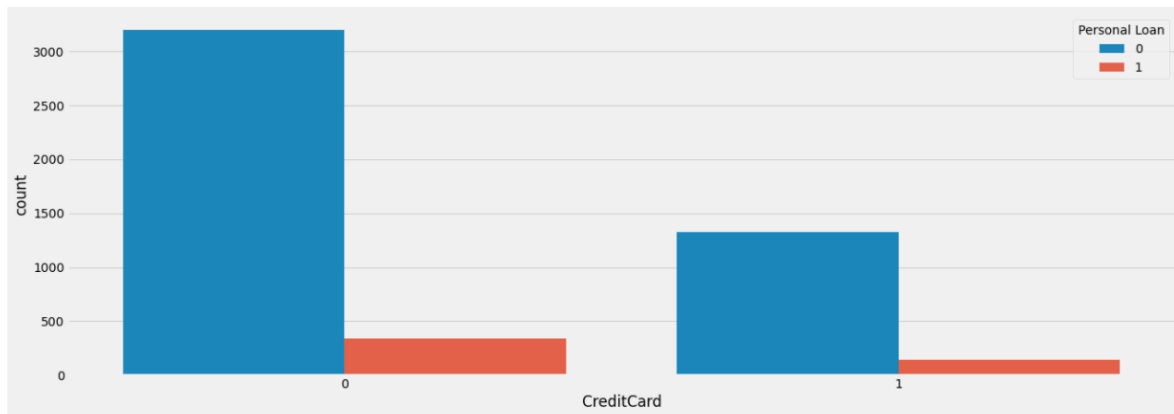


- Very few customers with Personal Loans have Security Account.

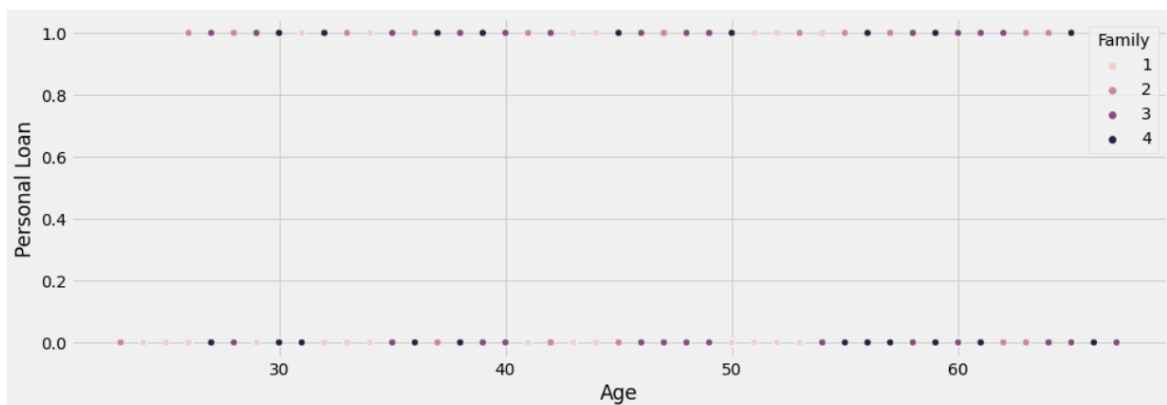


- No customer with Personal Loans has Deposits.





- Very few undergraduates take Personal Loan.
- Most of the Personal Loans are taken by professionals.



Results of Hypothesis Testing:

- Income of customers do have impact on Personal Loans as p value is less than 0.05, p value = 0.0
- Age of customers do not have any impact on Personal Loans as p value is greater than 0.05, p value = 0.584959263705325
- Family of customers do have impact on Personal Loans as p value is less than 0.05, p value = 1.4099040685673807e-05

Heat Map of Countries:

	Education	Income
Country		
AZ	7088.000000	1559.000000
Albania	1225.000000	171.000000
Argentina	1638.000000	266.000000
Armenia	1581.000000	154.000000
Australia	5919.400000	1400.800000
Austria	8124.000000	951.000000
Azerbaijan	2933.000000	151.000000
Belarus	2874.000000	224.000000
Belgium	9519.000000	1146.000000
Bosnia And Herzegovina	1268.500000	231.500000
Brazil	5243.000000	240.000000

Moldova	1445.000000	148.000000
Montenegro	4267.000000	246.000000
Morocco	2087.000000	205.000000
NV	6647.000000	1568.000000
NY	18274.000000	2335.000000
Nepal	605.000000	98.000000
Netherlands	4924.500000	1223.000000
New Zealand	6454.000000	1175.000000
Norway	5650.000000	1494.000000
OR	8732.000000	1704.000000
Panama	3162.000000	356.000000
Peru	2636.000000	224.000000
Philippines	3067.000000	183.000000
Poland	2769.833333	421.333333

Switzerland	15304.000000	2858.000000
TX	9281.666667	1980.000000
Taiwan	5538.000000	671.000000
Thailand	6993.000000	337.000000
Tunisia	859.000000	148.000000
Turkey	2766.250000	224.000000
Ukraine	1930.000000	204.000000
United Arab Emirates	4961.500000	1239.000000
United Kingdom	11423.000000	1380.000000
Uruguay	3012.000000	325.000000
Uzbekistan	2061.000000	112.000000
Vietnam	3884.000000	180.000000
WA	10650.000000	2373.000000

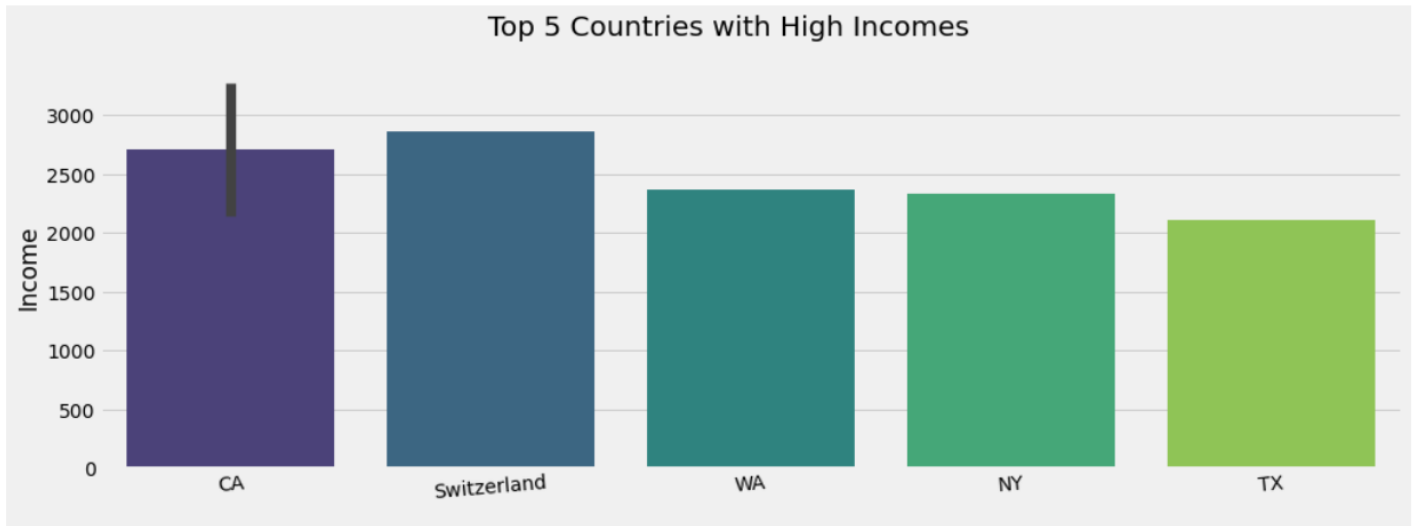


Figure 6 CA has the highest income but with huge deviations.

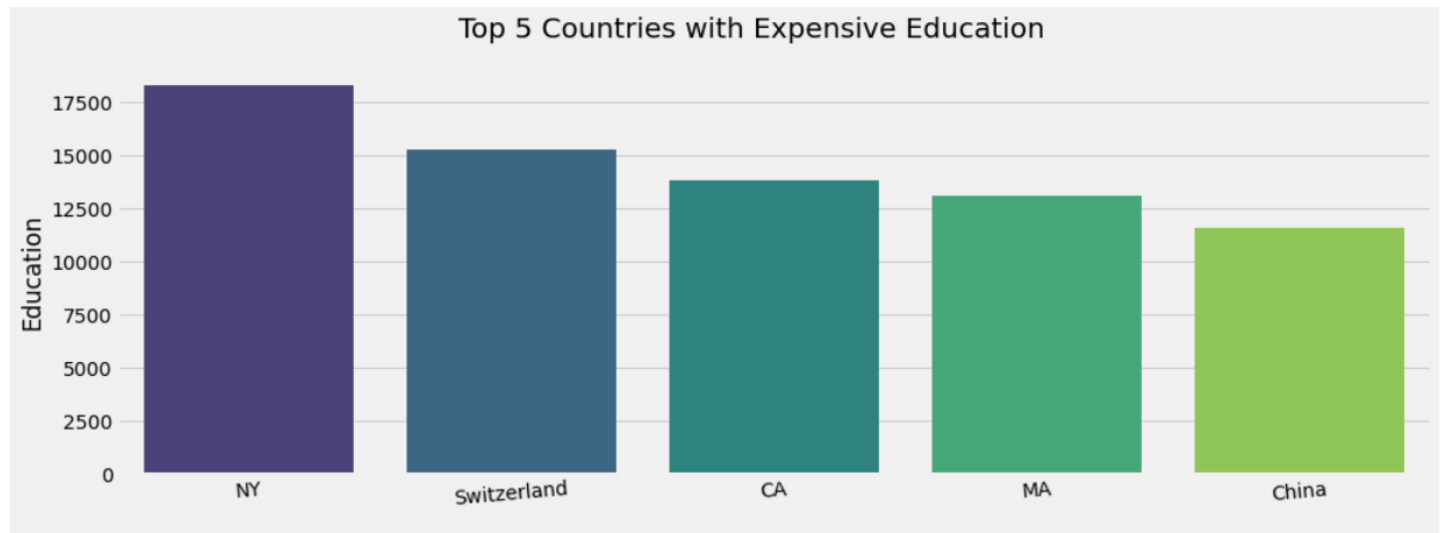


Figure 7 NY provides the most expensive education.

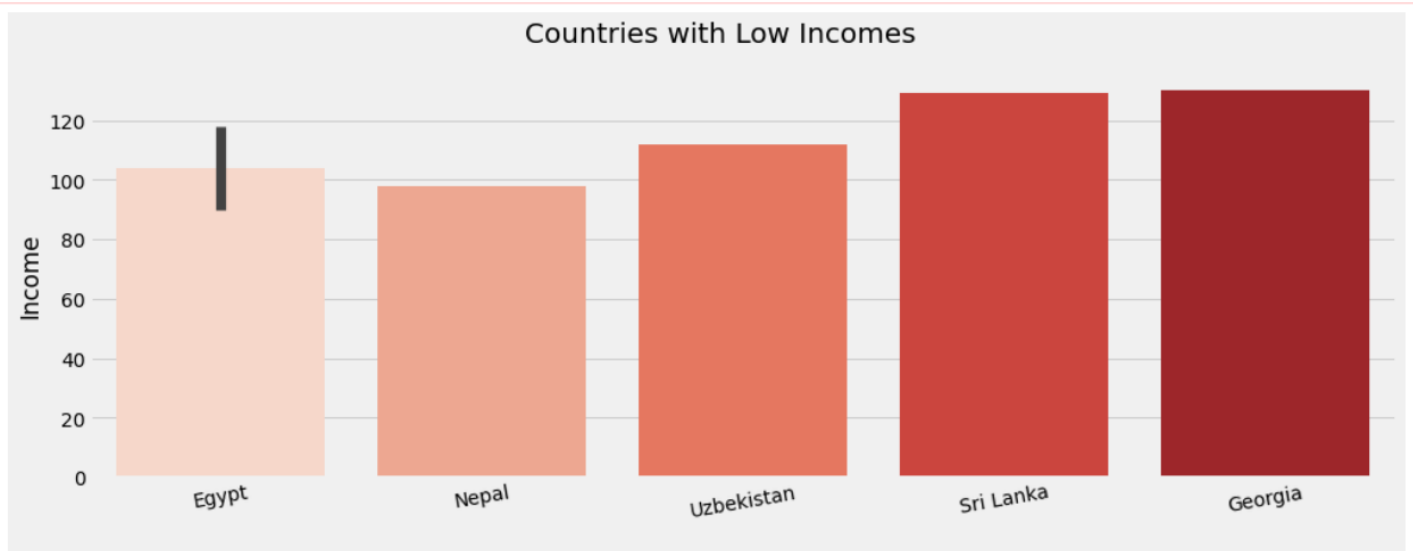


Figure 8 Egypt has the lowest income with huge deviations.

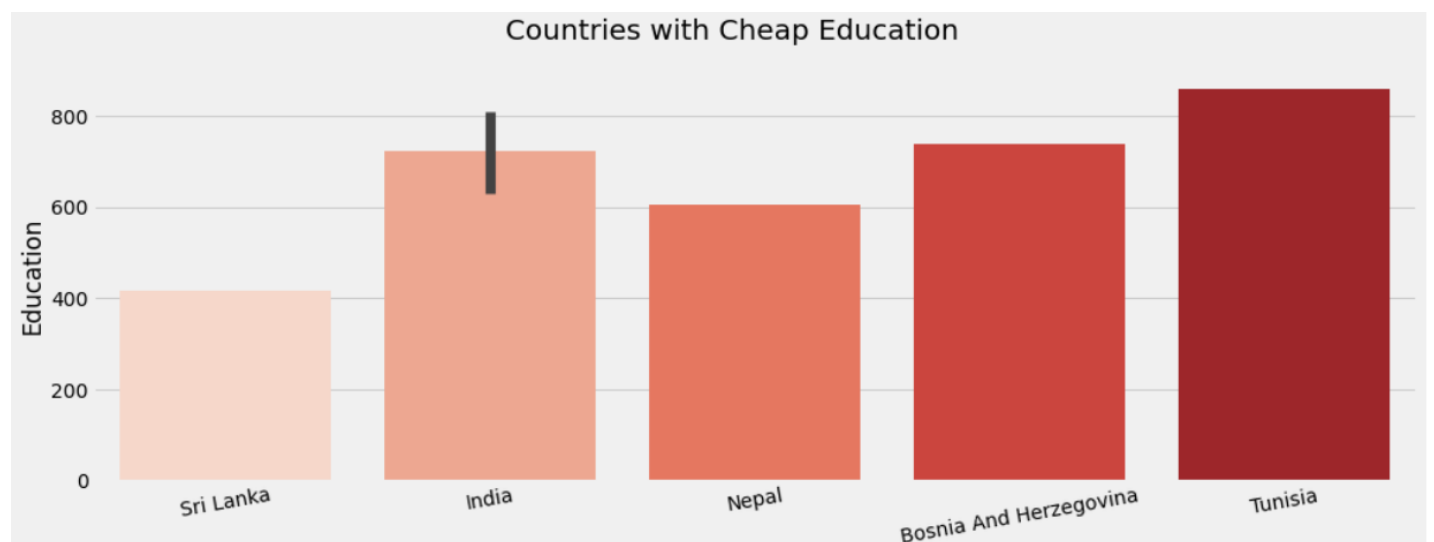
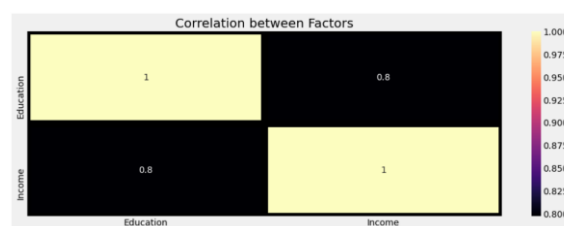


Figure 9 Sri Lanka provides the cheapest education.



- Income and Education are highly correlated.

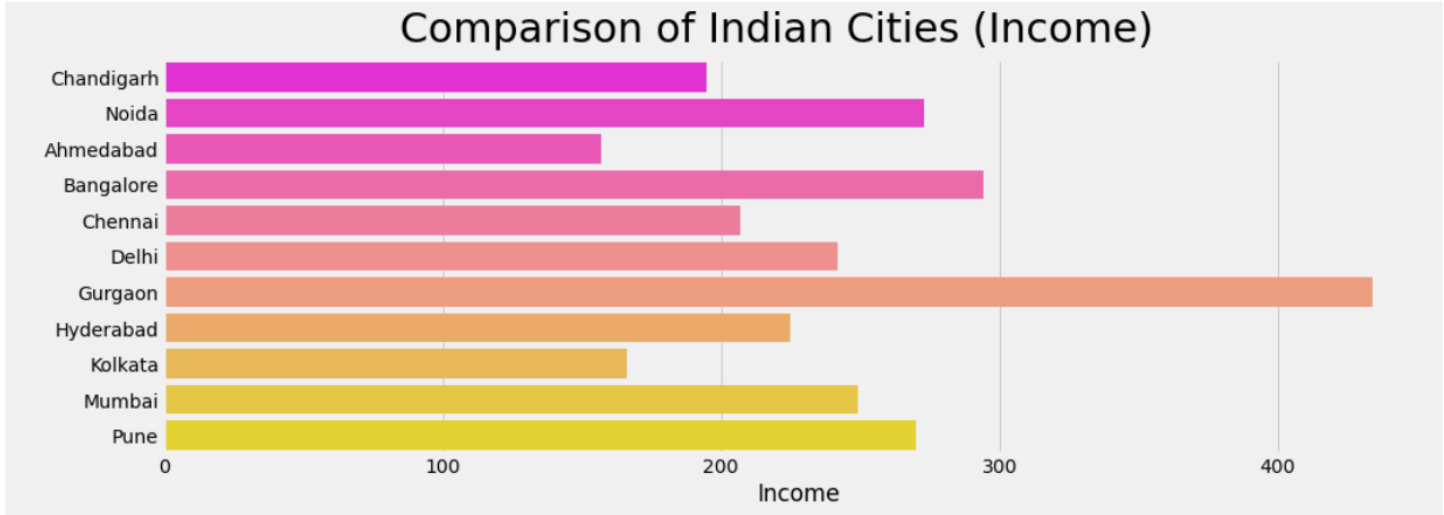


Figure 10 Gurgaon is an IT city with the highest income.

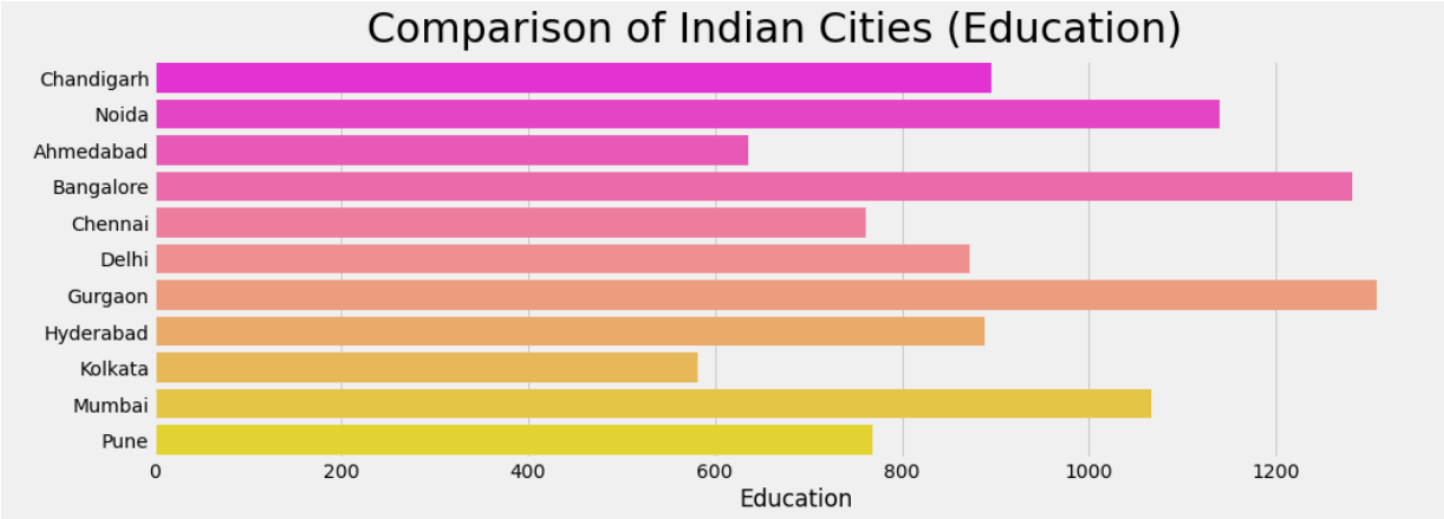


Figure 11 Education in Gurgaon is most expensive.