|  |  |
| --- | --- |
| A picture of a winding road and trees  eda REPORT  ON CREDIT CARD DATA SET | Abstract  Analyzing the credit card customer data  Rayon Susan Koshy  Data Science with Python 2020 |

**OBJECTIVE**

Deep dive into the Credit Card Customer data set using Exploratory Data Analytics & identify the hidden pattern in the customer features. Using the pattern to answer Who, Why, How

* Who are more being in defaulters & why?
* Why is the default customers number increasing?
* How to ensure a smaller number of defaulters?

Contents

[**I.** **DATA SET & FEATURES** 3](#_Toc39325036)

[**II.** **AGE DISTRIBUTION OF THE CUSTOMERS** 4](#_Toc39325037)

[UNDERSTANDING AGE DISTRUBUTION OF MEN & WOMEN 4](#_Toc39325038)

[**III.** **HOW MUCH THEY SPEND? BILL AMOUNT IN EACH CATEGORY** 7](#_Toc39325039)

[**IV.** **PAYMENT STATUS FOR EACH CATEGORY OF CUSTOMERS** 8](#_Toc39325040)

[PAYMENT STATUS WITH RESPECT TO MARITAL STATUS 9](#_Toc39325041)

[PAYMENT STATUS WITH RESPECT TO MARITAL STATUS 10](#_Toc39325042)

[**V.** **WHO ARE THE DEFAULTERS AMOUNG THE FEMALE CUSTOMERS?** 12](#_Toc39325043)

[**VI.** **WHO ARE THE DEFAULTERS AMOUNG THE MALE CUSTOMERS?** 14](#_Toc39325044)

1. **DATA SET & FEATURES**

The below table gives attribute information:

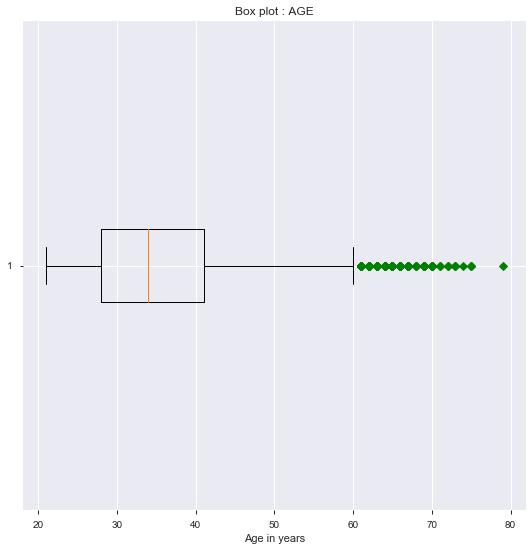
|  |  |  |
| --- | --- | --- |
| Variable | Description | Category |
| ID | The unique number for each row |  |
| LIMIT\_BAL | Amount of the given credit (NT dollar) including both the individual consumer credit and his/her family (supplementary) credit |  |
| SEX | Gender | 1 = male  2 = female |
| EDUCATION | Educational qualification of the customers | 1 = graduate school  2 = university  3 = high school  0, 4, 5, 6 = others |
| MARRIAGE | Marital status | 1 = married  2 = single  3 = divorce  0=others |
| AGE | Age in year |  |
| PAY\_0 - PAY\_6 | History of past payment (April to September) | -2 = No consumption  -1 = Paid in full  0 = The use of revolving credit  1 = payment delay for one month  2 = payment delay for two months  8 = payment delay for eight months  9 = payment delay for nine months and above. |
| BILL\_AMT1 - BILL\_AMT6 | Bill amount from April to September (NT dollar) |  |
| PAY\_AMT1 - PAY\_AMT6 | Amount of previous payment (NT dollar) |  |
| default payment next month | Client's behavior | Default  Non-Default |

**TABLE1**: Attribute information

1. **AGE DISTRIBUTION OF THE CUSTOMERS**

Age of the customers varies in wide range, 21 being the youngest customer & 79 being the eldest.

The highlights on the age of the customer’s:



**Fig1**: Age Boxplot

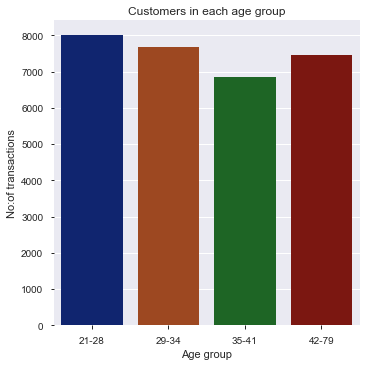
|  |  |
| --- | --- |
| Minimum | **25** |
| First quartile(Q1) | **28** |
| Median (IQR) | **34** |
| Third quartile(Q1) | **41** |
| Maximum | **79** |

## UNDERSTANDING AGE DISTRUBUTION OF MEN & WOMEN

**KDE Plot** described as Kernel Density Estimate is utilized for visualizing the Probability Density of AGE. It depicts the probability density at various age of the customers, thereby help us to understand age distribution of customers. Let us consider two sets:

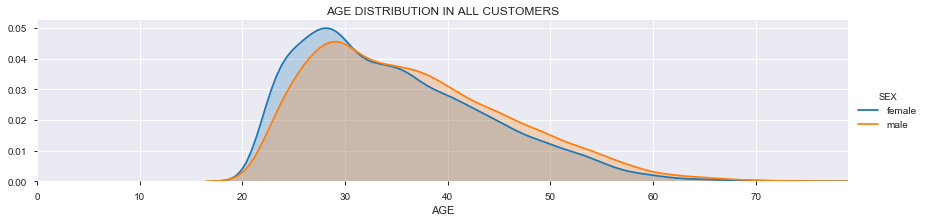
* Data consisting of all customers
* Data consisting of just default customers.

The below bar graph shows the number of customers per age group:

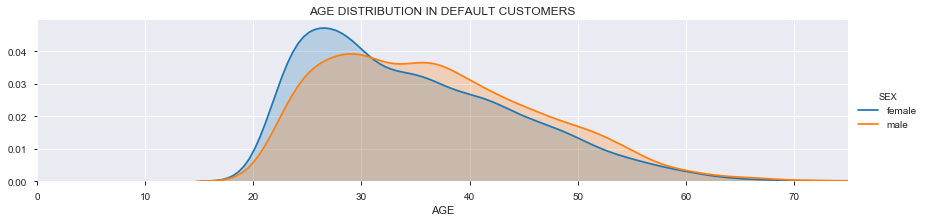


**Fig2**: Age Bar graph

To understand how the age is distributed among the customers let us investigate the below age distribution KDE plot of both men & women in a set of all customers & set of only defaulted customers, respectively:

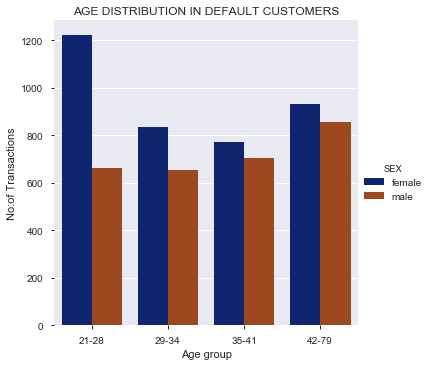
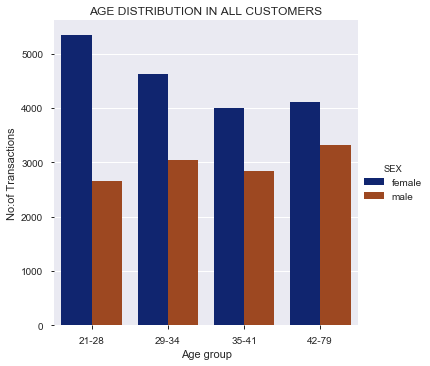


**Fig3**: Age Distribution in all customers



**Fig4**: Age Distribution in default customers

To have a closer look into the count per age group, let us refer the below bar graph:



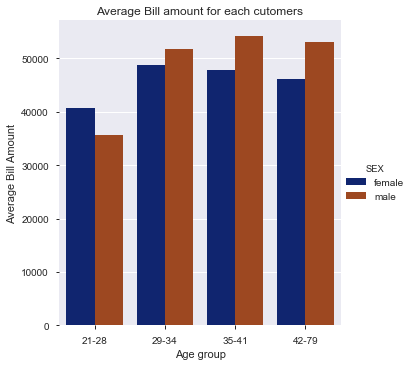
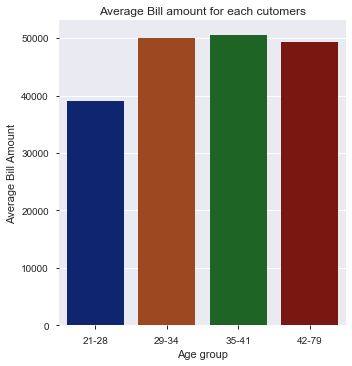
**Fig5**: Bar plot showing age distribution in all customers(left) & default customers(right)

INFERENCE:

* Most of the women defaulters are in the age group 21 to 28 & with the increase in age the chances of they are in default is less.
* It is also important to note that the women around the age 21 to 28 are the main customers
* There is a high density of default male customers in the age group from 24 to 40

1. **HOW MUCH THEY SPEND? BILL AMOUNT IN EACH CATEGORY**

Shopping pattern is different in men & women. They indeed are different in different age group. The below fig shows average bill amount in each age group & how they differ based on the gender.



**Fig6**: Average bill amount bar graph for each age group(left) & gender(right)

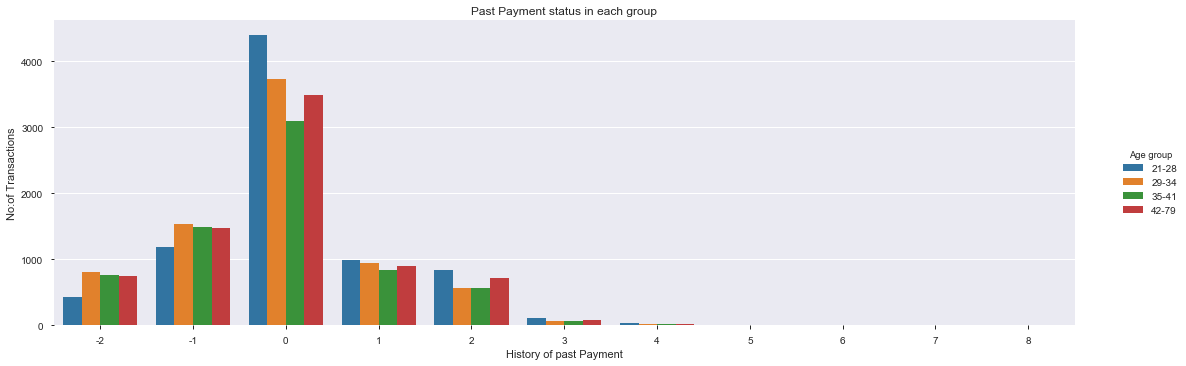
INFERENCE

* From Fig2, it was noted that the age group 21 to 28, tends to shop more. But the above graph shows mean amount spend in the age group from 21 to 28, is less comparatively.
* Other than the age group 21 to 28, male customers spend more on an average. But the male customers in the age group 42 to 79 tends to more on default list (Fig5)
* Women customers in the age group 21 to 28, spend less on an average and shop more. They are also high in number among the defaulters list.

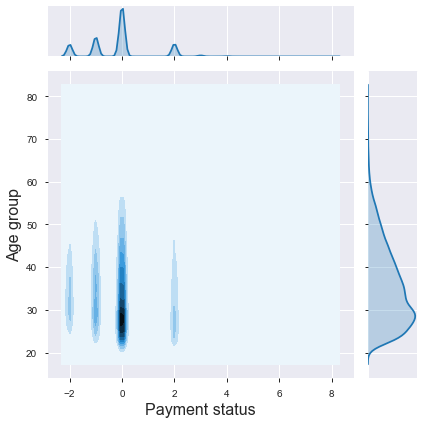
1. **PAYMENT STATUS FOR EACH CATEGORY OF CUSTOMERS**

Payment status for each customer is different in 6 months. For visualizing the payment status for various category of customers, we consider payment status just for one month

The below shows the payment status for each group:



**Fig7**: Bar plot showing payment status in all customers with respect to the age group



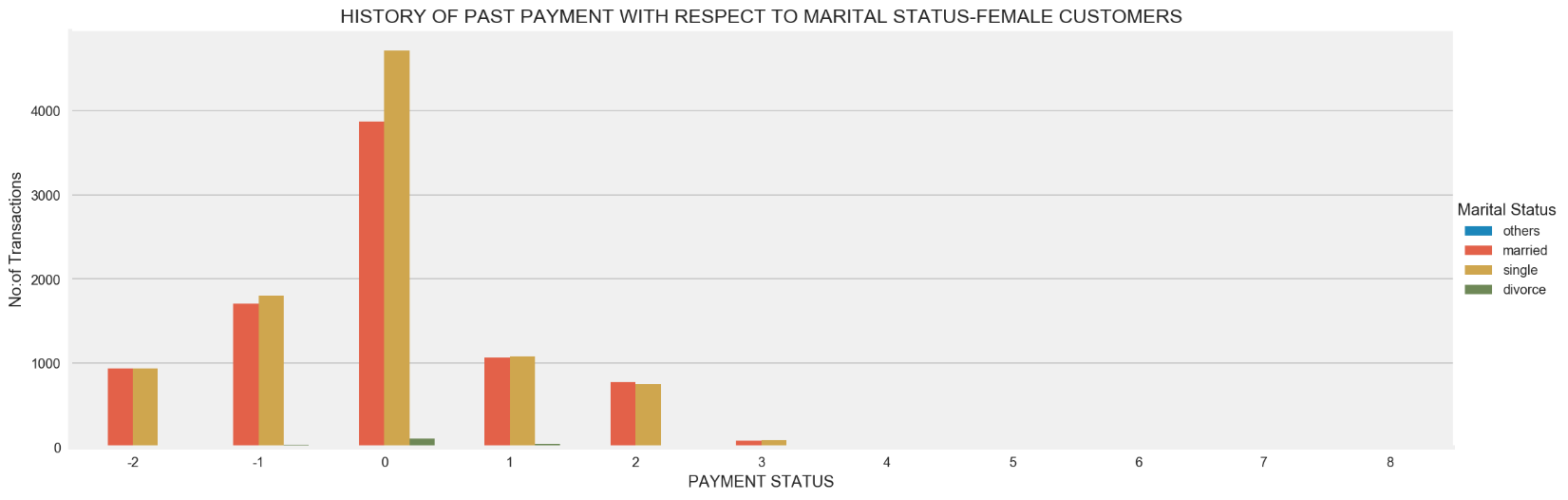
**Fig8**: Distribution of age in each payment status

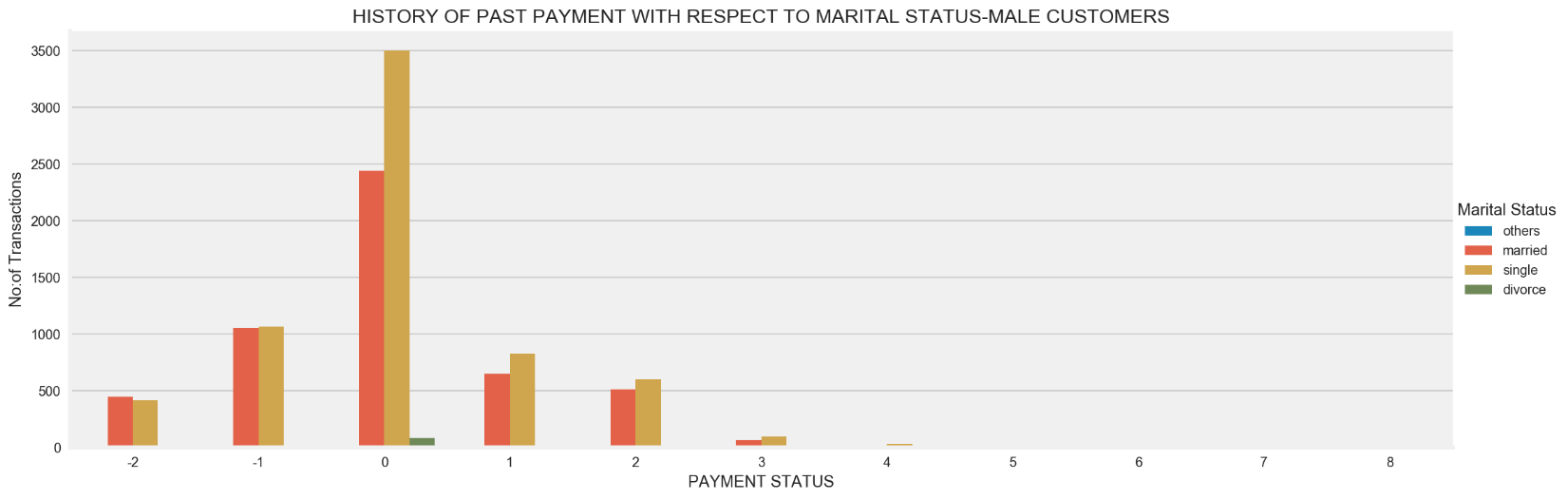
INFERENCE

* Most customers like to use revolving credit
* Age group 20 to 28, is less among the customers that pay in full & customers with no consumptions. That concludes they usually do not pay in full. Rather use the revolving credit to do more purchase.

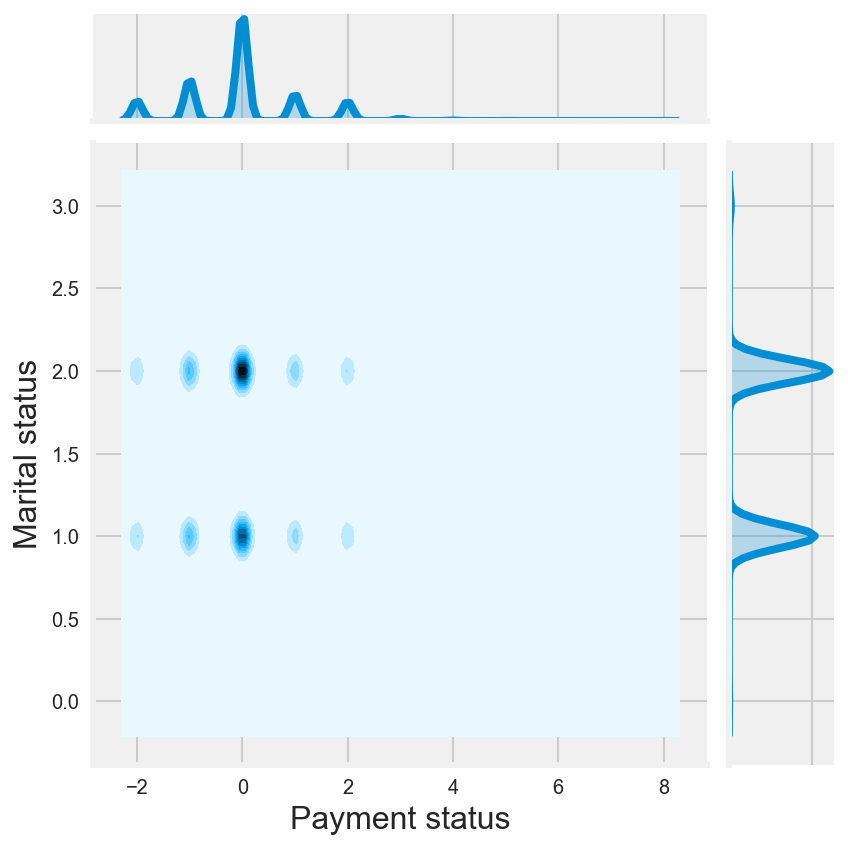
## PAYMENT STATUS WITH RESPECT TO MARITAL STATUS

Let us investigate the payment status for women & men with respect to their marital status.





**Fig9**: Bar plot showing payment status among female (upper) & male customers(lower) with respect to their marital status



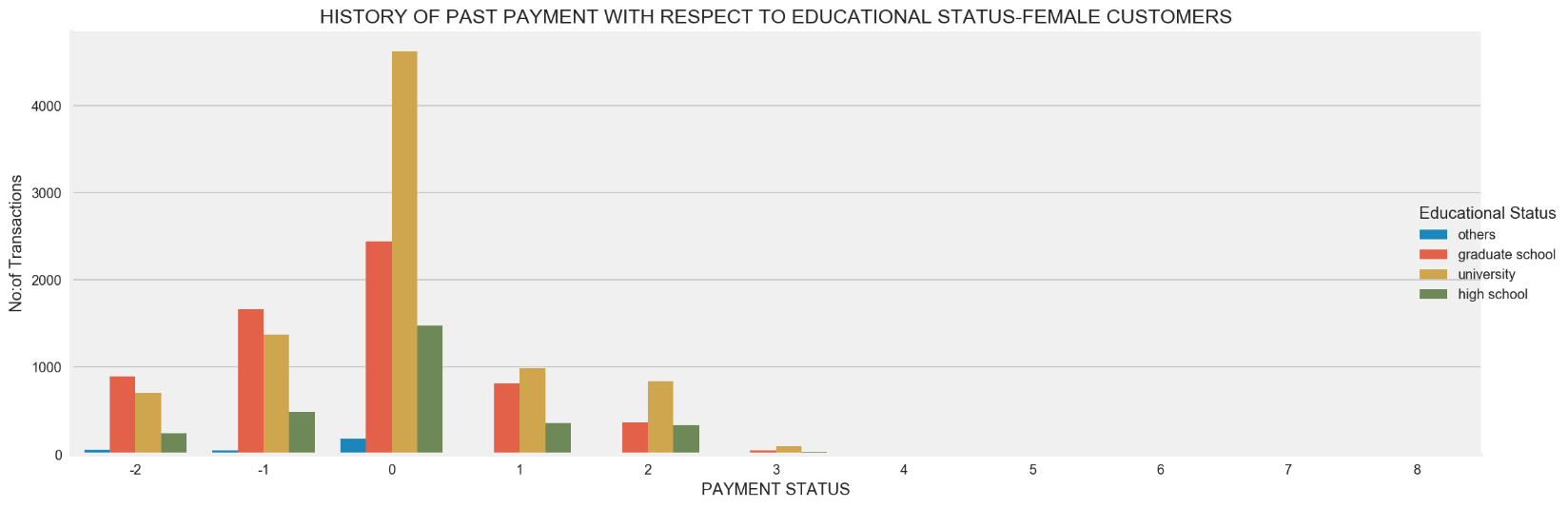
**Fig10:** Distribution of customers in each payment status with respect to their marital status

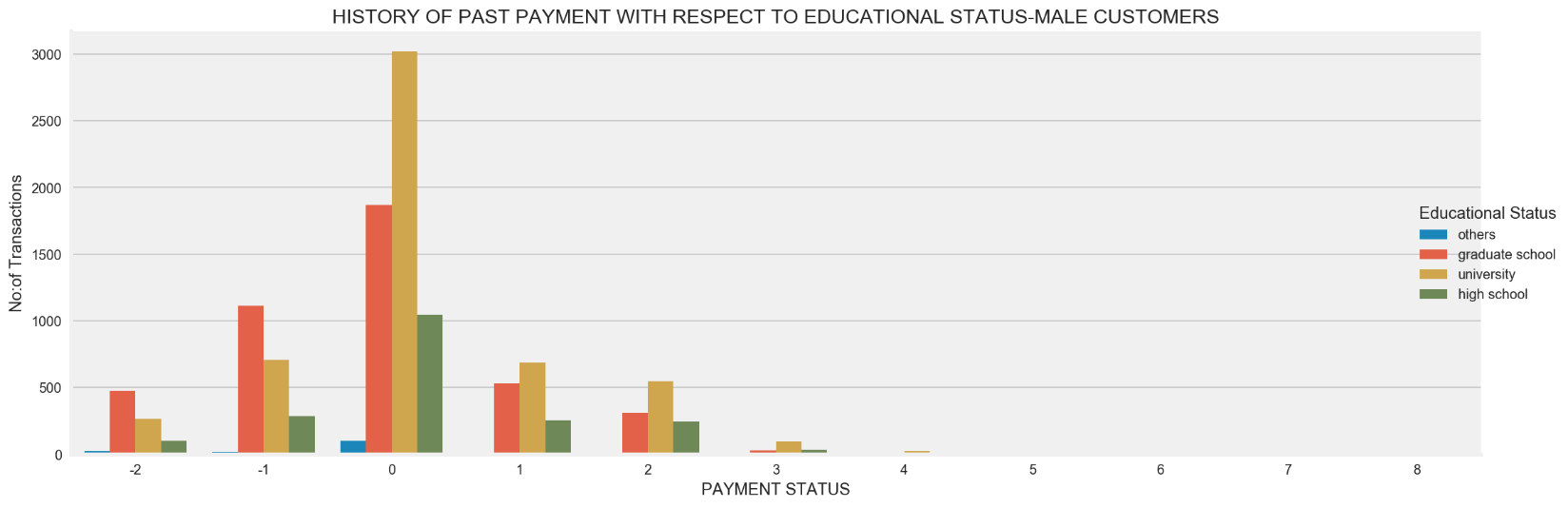
INFERENCE

* Among both female & male customers, most of them are single and like to use the revolving credit.
* Considering payment delay in the female customers, most of them are single & married
* The pattern of payment delay among the women is same to that of male. But there is slight difference that lies in the type male customers who have more payment delay status. Single male customers are likely to be more in payment delay status.

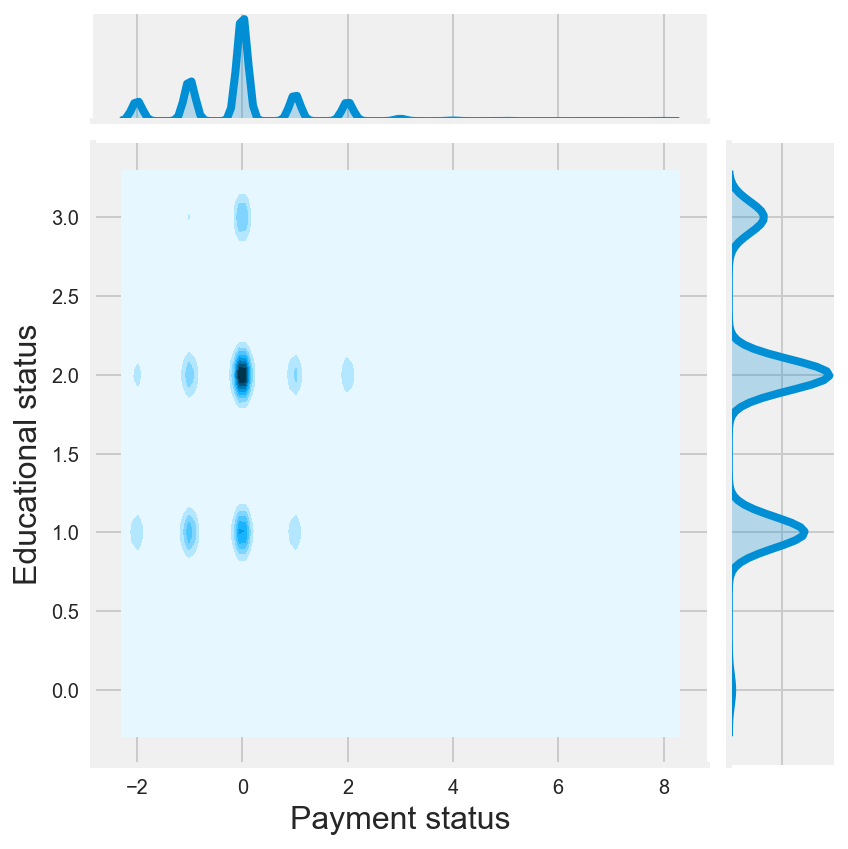
## PAYMENT STATUS WITH RESPECT TO EDUCATIONAL STATUS

After having a look into customer’s payment status based on the marital status, the next factor to consider is education. The below shows the influence of education as a factor in payment status.





**Fig11**: Bar plot showing payment status among female (upper) & male customers(lower) with respect to your education



**Fig12:** Distribution of customers in each payment status with respect to their educational status

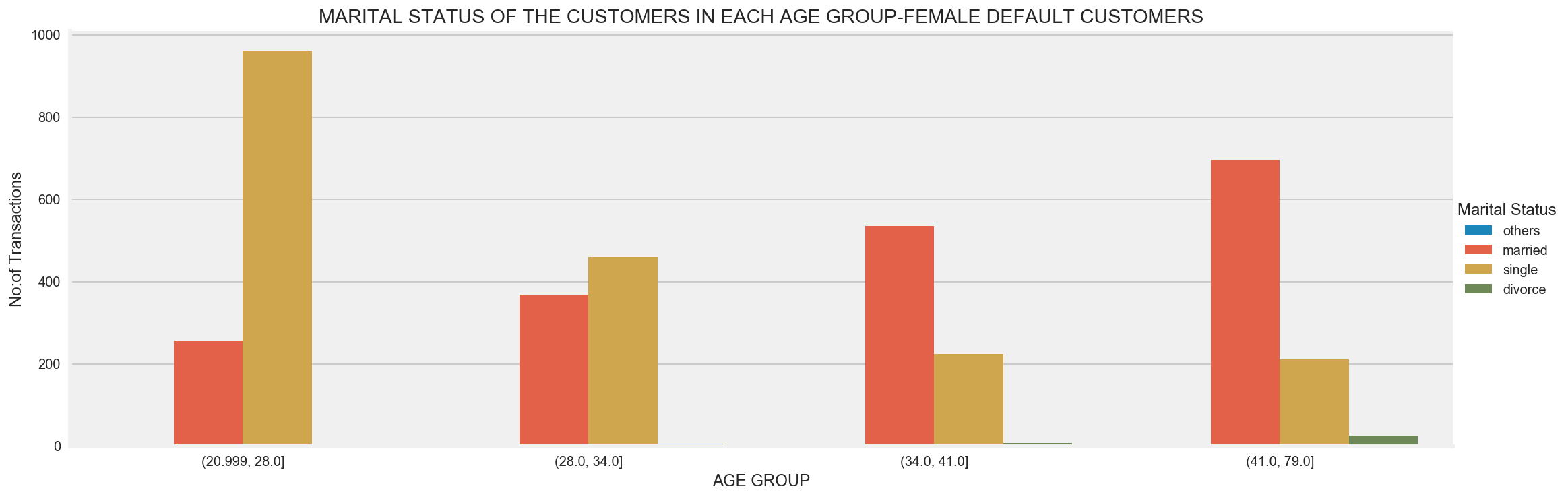
INFERENCE

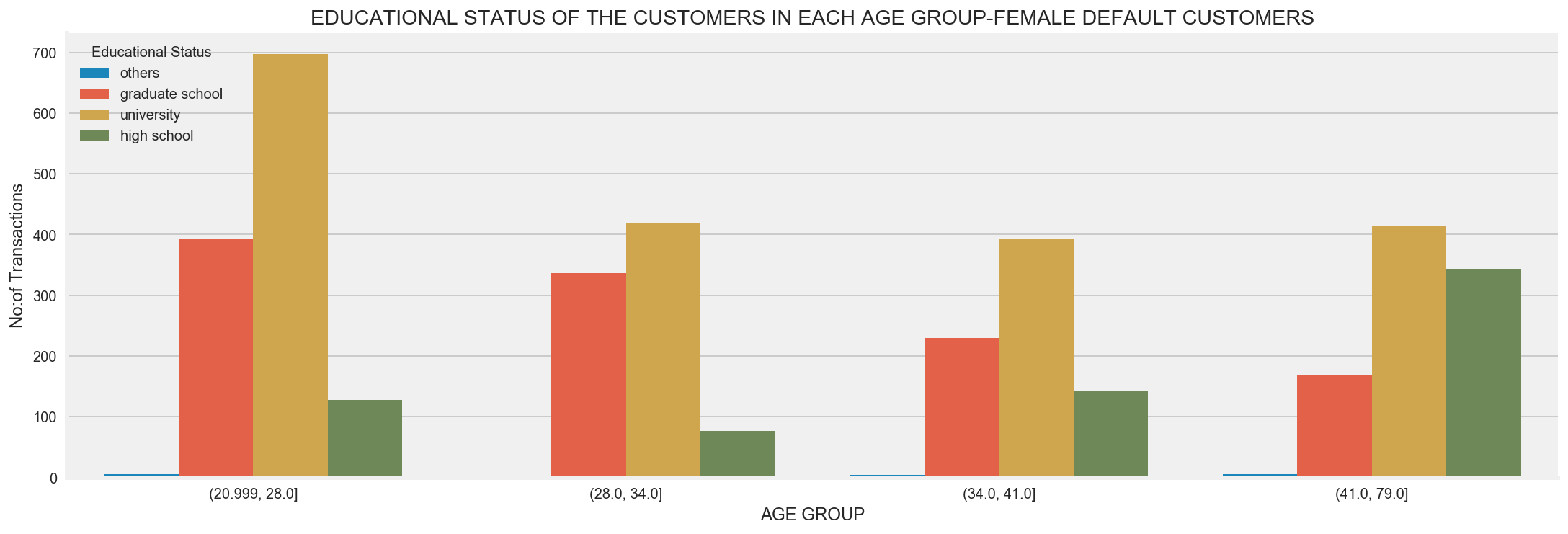
* By looking into the above graphs of both female & male, they show same pattern in payment status when considering the educational status. So, we can say sex alone is not factor in predicting the payment status. Seeing the above graph, even though sex does change the pattern, it does change the count. Female customers are more compared to the male customers.
* Customers with university education mostly use the revolving credit. They are also high when coming to the customers with payment delay.
* Customers with graduate education mostly make full payment

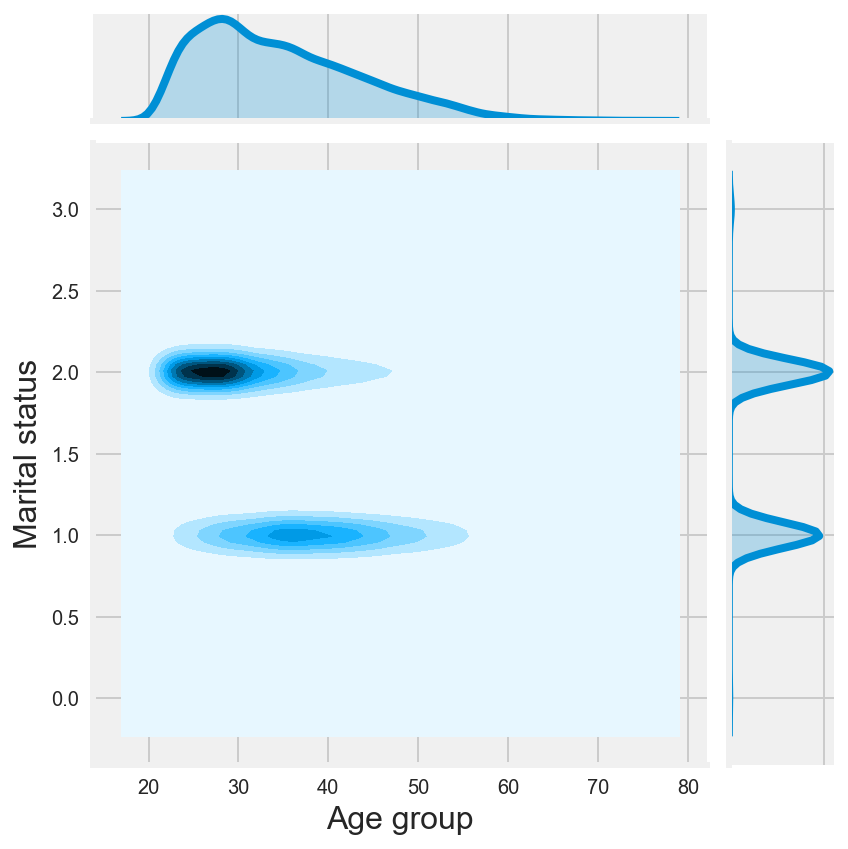
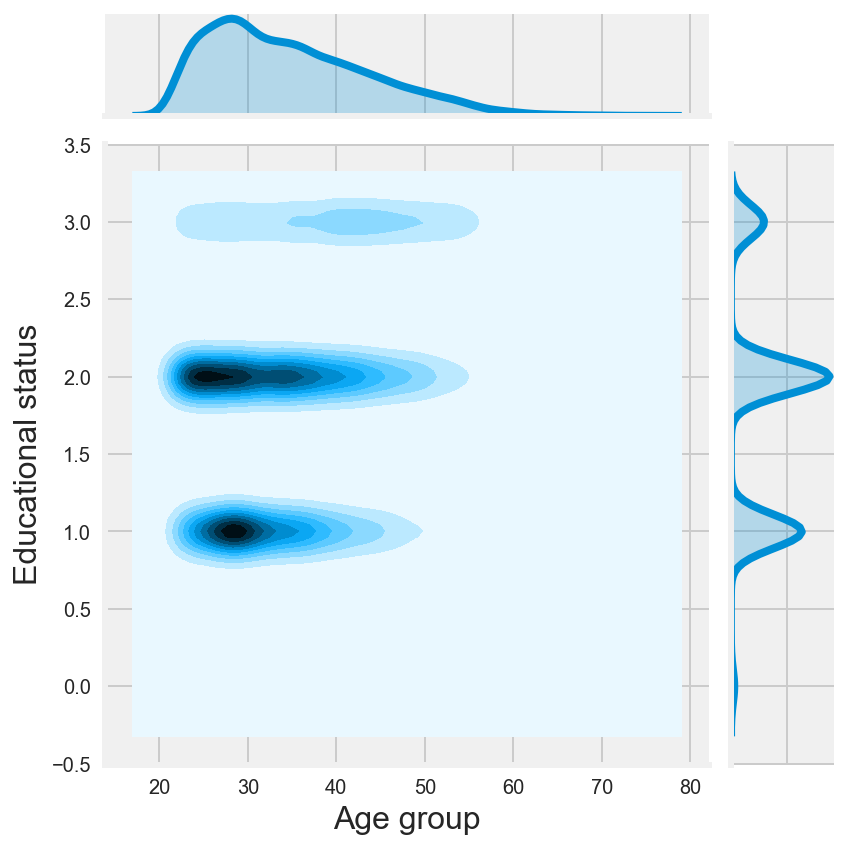
1. **WHO ARE THE DEFAULTERS AMOUNG THE FEMALE CUSTOMERS?**

Female customers are high when compared to the male. Each customer can vary depending upon their age, education & marital status. Each feature can play a major role in financial status of customers, which in turn determines whether they are defaulters or not.

The below shows the number of default customers in each age group, considering marital & educational status:





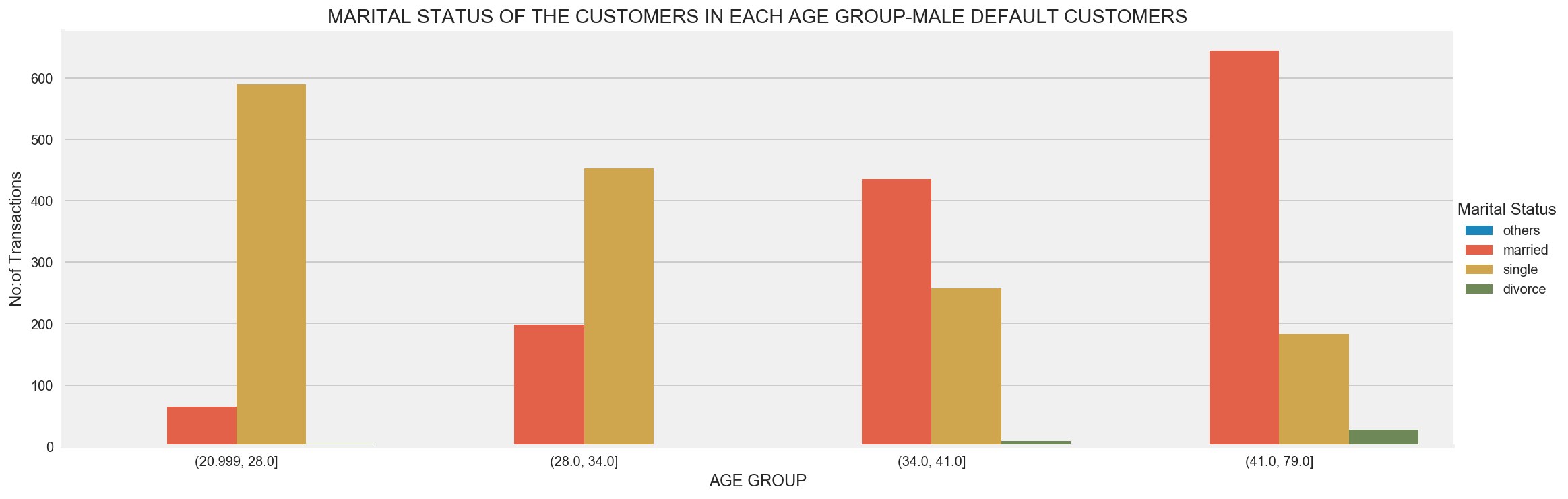
INFERENCE

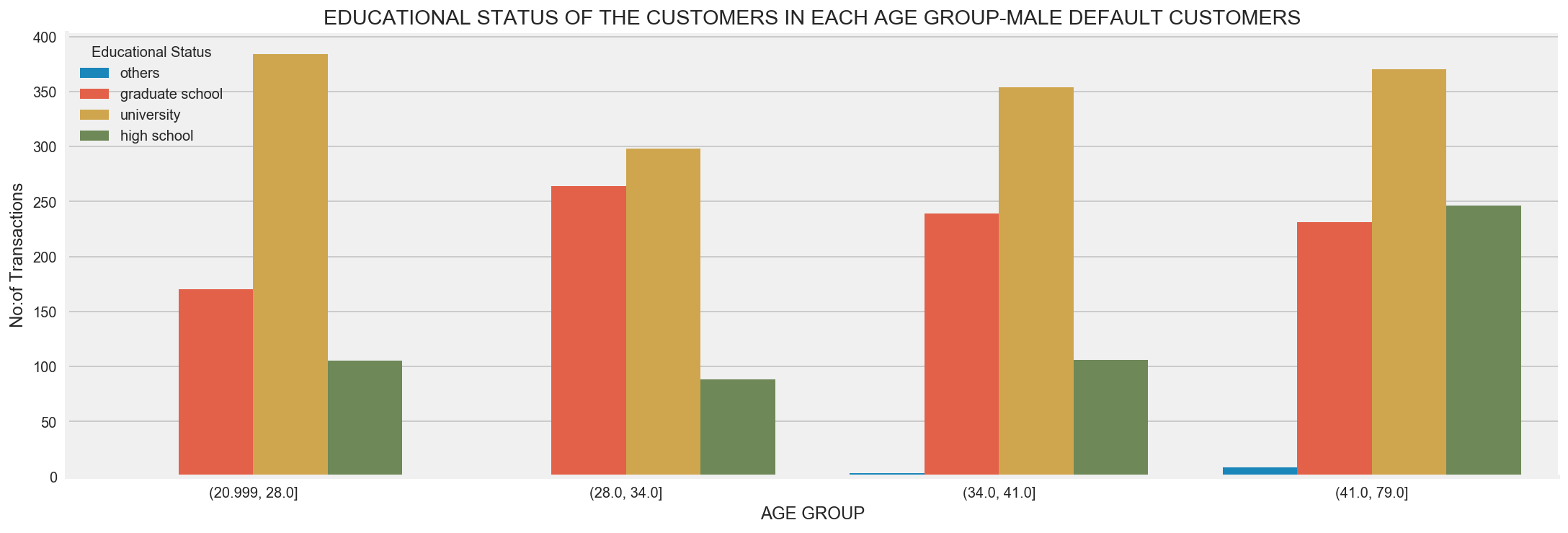
* In age group 21 to 34, most of the women customers are single with the educational qualification of university
* In age group 35 to 79, most of the women customers are married with the educational qualification of university
* Women customers with high school educational qualification, tends to be more in default list as the age increases. They are quite high in the age 42 to 79
* On contrary women customers with graduate school educational qualification, tends to be less in default list as the age increases. They are quite high in the age 21 to 28.

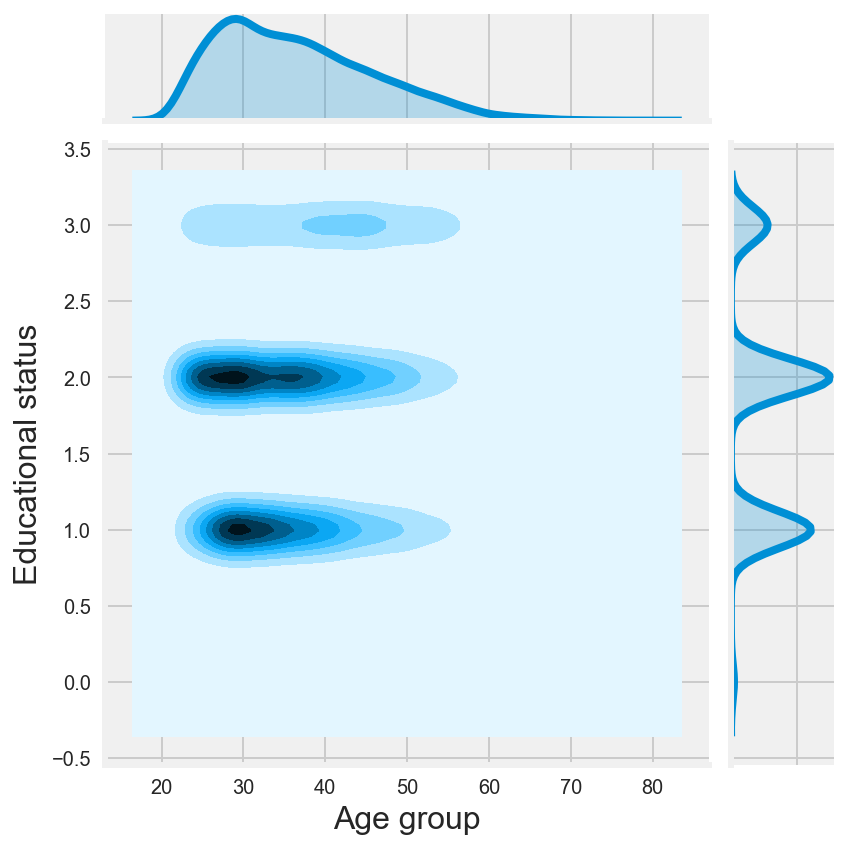
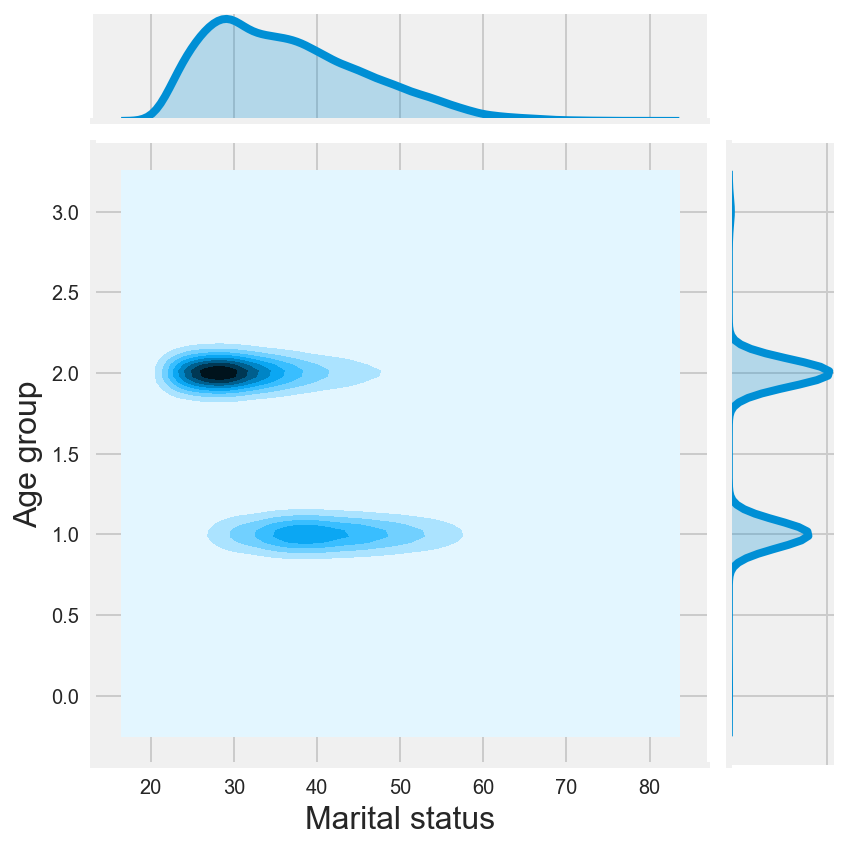
1. **WHO ARE THE DEFAULTERS AMOUNG THE MALE CUSTOMERS?**

As said earlier female customers are high when compared to the male. Influence of customer features like age, educational status & marital status on the default customers is not exactly same on male & female customers. We already saw the who are most likely to be defaulters in female customer list. Let us now consider the male customers.

The below shows the number of default customers in each age group, considering marital & educational status:







INFERENCE

* In age group 21 to 34, most of the men customers are single with the educational qualification of university. This is same to pattern found in female customers.
* In age group 35 to 79, most of the male customers are married with the educational qualification of university
* Male customers with high school educational qualification, tends to be more in default list as the age increases. They are quite high in the age 42 to 79. This also same to women’s trend
* Male customers with graduate school educational qualification, tends to be more in default list as the age increases. They are quite high in the age 29 to 79. This is different from the women customer’s trend.