Conclusion Write Up
Suhas, Bini, Sheryl, Morgan

Marriage and Home Loans:

Does being married at any instance of an individual's life (currently married, divorced, or widowed) affect whether or not the individual has a home loan?

Using the Confidence Intervals for population proportions of currently/been married with a house loan and single with a house loan. I found that the two CI's overlap meaning there is a possibility of the two population proportions being the same.

Using a 2 sample z-test for proportions to determine if the two proportions are the same or if the proportion for the currently/been married population is greater than our single population since that would tell us its effect of owning a house loan. From that z test the p-value was 0.021. This tells us the currently/been married with a house loan population proportion is greater than the single with a house loan population proportion telling us that if you currently/been married you are more likely to have a house loan.

Education and Personal Loans:

Does one's level of education (4y, 6y, 9y,High school, illiterate, Prof.Course,Univ.) affect whether or not the individual has an individual loan?

Plotting a bar chart showing how many people from each educational level is the first test I used. Through this it was found there is a positive correlation between the amount of personal loans had and education. However, there was not much of a relationship between the ratio of those with a personal loans and education

I also created a heat map by turning the loan and education values into integers. The intersection value between education and personal loans is only 0.009. This value was too low to prove correlation between the 2 factors.

Conclusion Write Up
Suhas, Bini, Sheryl, Morgan

How does the job type and the age affect the kind of loan an individual takes?

Firstly, I created individual countplots to show the relationship between job and the loans(housing & personal). In the countplot for the job type and the housing loan, we can see that admins form a majority of the job type(approx 28%). Out of which, 15.7% of them opt for housing loans. Blue-collar professionals and technicians are next in line in taking the maximum number of house loans. We can see that students and the housemaids take relatively fewer house loans.

Moving onto the countplot for job and personal loans, we can see that a considerable percentage of admins, blue-collar professionals and technicians don't take personal loans.

Comparing both these countplots we can say that, irrespective of the job type, people take more housing loans when compared to personal loans.

Secondly, I created the individual countplots for age groups and loans. Initially the age was discrete data which I converted into categories using pandas cut function and bins. Now looking at the relation between age groups and housing loans we see that, people who fall in the age range of (29-49) take a majority of the house loans. People in the age range of (59-99), are considered senior citizens and hence don't opt for housing loans. Students from the age of (9-19) also don't take housing loans for obvious reasons.

The countplot displaying the relationship between age and personal loans shows that a large percentage of the people falling into the age group(19-59) don't take personal loans.

Finally, I created heatmaps to visualize the job type and the age with each of the two loans. Firstly, speaking about the heatmap for the housing loans, I took an example to explain this, considering a person whose jobtype is “housemaid” whose age ranges from (49-59) and checked the heatmap. From the heatmap it was evident that the housemaids satisfying this criteria dont take house loans. I have also verified the same from the dataset.

Speaking of the heatmap for personal loans, for this case I took the example of a blue-collar professional whose age falls in the range from (69-79). The heatmap shows that blue-collar professionals who satisfy this criteria apply for personal loans. This was also verified from the dataset.

We know from the previous countplots, that more people opt for housing loans when compared to personal loans, which explains why the heatmap for job, age and housing loans are more colored in green and the heatmap for job, age and personal loans are more colored in red.

**Note: These heatmap calculations have been made by performing aggregations on the data.(mode())

Conclusion Write Up
Suhas, Bini, Sheryl, Morgan

Finally, I made use of both the above heatmaps to identify what kind of loan a person takes on the basis job and the age category. For this I considered the example of a person who is an entrepreneur and falls in the age group of (29-39). From the heatmaps we see that the person applies/has a housing loan but doesn't have a personal loan. The same is verified from the dataset.

Again, all heatmap calculations are considered on aggregated data.

4) a) How do housing loans impact the subscription of term deposit(y)?

Analysis is made based on the impact of housing loan on term deposit after cleaning the data frame and dropping out the unknown terms.
Observations:- It is observed that the number of customers who have housing loans is 13174 and those who don’t have housing loans is found to be 11131. Inorder to describe the relationship between two categorical variable ie housing loan & term deposit,special type of table called crosstab is used. The relation between housing loan and term deposit is analysed and it is observed that total number of customers with housing loan and term deposit subscription is 1692 & the percentage dependency is 12.84% and those without housing loan and without term deposit is 9755 & the percentage dependency is 87.63%, total number of customers who have housing loan but no subscription is 11482 & the percentage dependency is 87.15% and total number of customers who don't have housing loan but have subscription is 1376 & the percentage dependency is 12.63%.
Analysis:- More number of customers have housing loans . Housing loans don't seem to be a priority to check for the subscription to term deposit as we can see that almost an equal number of customers who have and don't have housing loans did subscribe to the term deposit.

b) How do personal loans impact the subscription of term deposit(y)?

Analysis is made based on the impact of personal loan on term deposit after cleaning the data frame and dropping out the unknown terms.
Observations:- It is observed that the number of customers who have personal loans is 20491 and those who don’t have personal loans is found to be 3814. Inorder to describe the relationship between two categorical variable ie personal loan & term deposit, special type of table called crosstab is used. The relation between personal loan and term deposit is analysed and it is observed that total number of customers with personal loan and term deposit subscription is 481 & the percentage dependency is 12.84% and those without housing loan and without term deposit is 17904 & the percentage dependency is 87.37%, total number of customers who have personal loan but no subscription is 3333 & the percentage dependency is 87.38% and total number

of customers who don't have personal loan but have subscription is 2587 & the percentage dependency is 12.62%.

Analysis:- More number of customers don't have personal loan. More number of customers who don't have personal loans also haven't made a subscription to the term deposit. 3) Only very few people having a personal loan have subscribed to the term deposits(The logic may be that they already have some existing liability that they need to repay which makes it difficult to arrange money for the subscription to the term deposit)

c) How does credit in default impact the subscription of term deposit(y)?

Analysis is made based on the impact of personal loan on term deposit after cleaning the data frame and dropping out the unknown terms.

Observations:- It is observed that the number of customers who don't have credit in default is 24302 and those who don't have credit in default is found to be 3. Inorder to describe the relationship between two categorical variable ie credit in default & term deposit, special type of table called crosstab is used. The relation between credit in default and term deposit is analysed and it is observed that total number of customers with credit in default and term deposit subscription is 0 & the percentage dependency is 0% and those without credit in default and without term deposit is 21234 & the percentage dependency is 87.37%, total number of customers who have credit in default but no subscription is 3 & the percentage dependency is 100% and total number of customers who don't have credit in default but have subscription is 3063 & the percentage dependency is 12.62%.