Question 1.3: Does the result of the graph above align with your intuition? Why or why not?

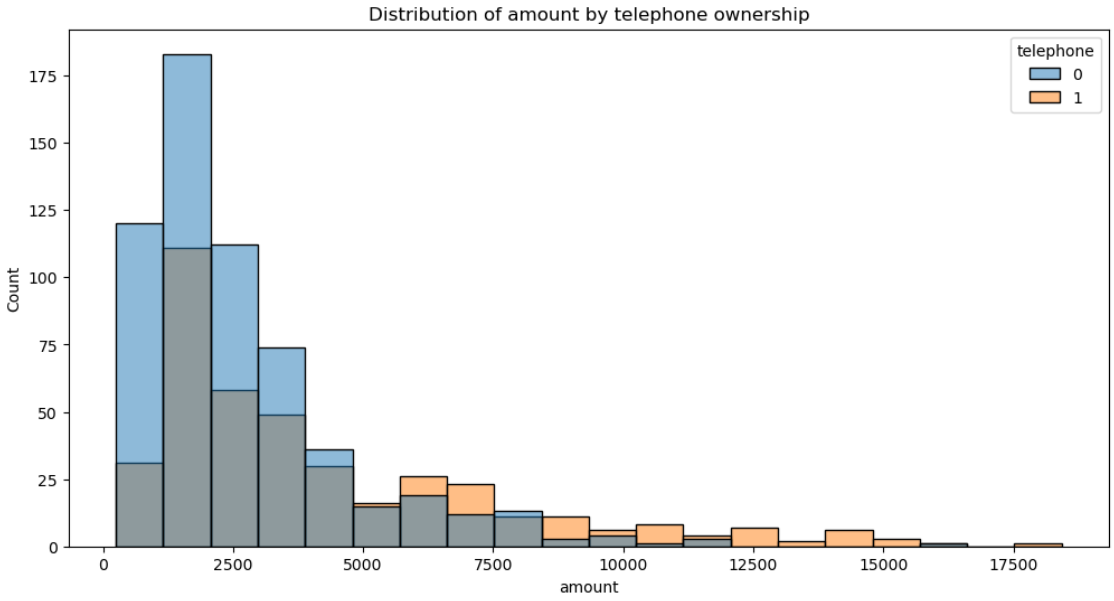
In the graph, I find that borrowers with past delays in paying off credit have a lower survival rate than others, indicating that they are more likely not to pay off their loans as quickly. This intuitively makes sense: If you previously had issues paying off your debt, you might still have some financial struggles and thus would be slower to pay off your loan. This could reasonably be used to classify a loaner as one of higher risk.

The critical account borrowers have the lowest survival rate, i.e., they are the quickest to pay off their loan. I have a hard time figuring out exactly what is meant by *Critical Account/Other Credits Existing* as this is not clear from the limited description provided in the UCI documentation or the GitHub documentation other than they either have other credits existing at other banks or have the been classified as a *critical account* – whatever that means. I do not have any specific intuition for why people with credits existing in other banks would be quicker to pay off their loan other than that they might have a better credit score and thus easily can get a loan elsewhere. However, the name *critical account* would indicate that they are not in a good financial situation and thus would be slower to pay off their loan.

Question 1.5: Does this align with your intuition? What are some possible reasons for this to be the case?

I find that people with telephones are slower to pay off their loans, which does not make much sense. There must be something else going on here. One possible explanation could be that people with telephones have a larger loan than those without one. We are here only considering the number of months a given loaner takes to pay off their loan—not the size of the loan itself. It seems intuitive that a larger loan implies a longer time to pay it off. Plotting a histogram split on whether or not the loaner has a telephone shows that many people without a telephone tend to have a smaller loan than those with a telephone. The smaller loan would make it easier to pay back, and additionally, it is plausible that this could very well be the reason why people with telephones take longer to pay back their debt.

```
In [21]: sb.histplot(data=credit, x="amount", hue="telephone", bins=20)
         plt.title('Distribution of amount by telephone ownership')
         plt.show()
```

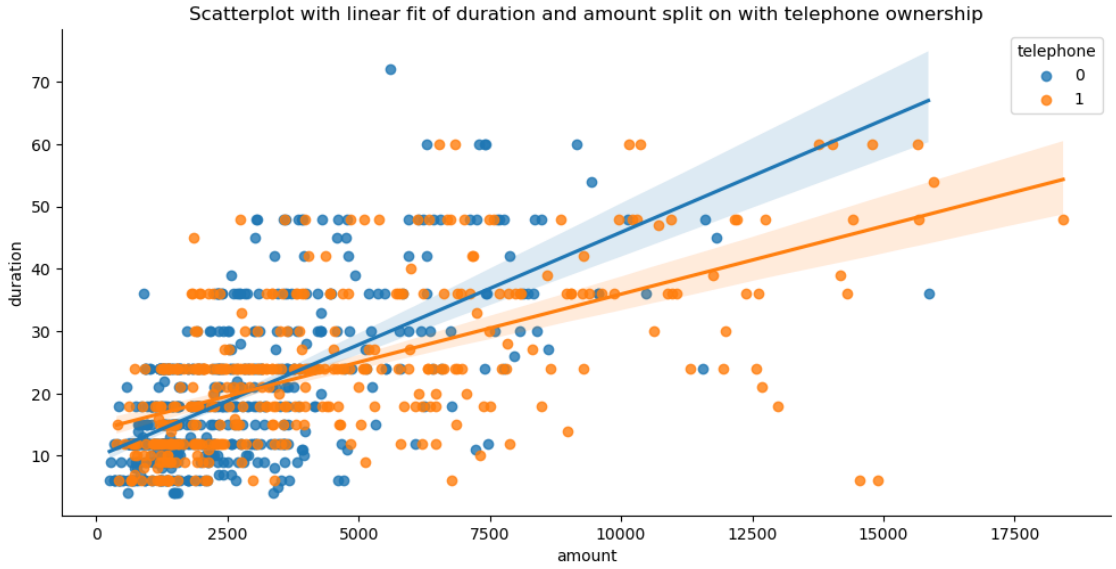


Looking further into this correlation with a scatterplot of the loan size and the duration of the loan shows that there is a positive correlation between the two and that people without a phone, though slightly faster at paying back small loans, are, in fact, slower when paying back larger loans.

```
In [22]: sb.lmplot(data=credit, x='amount', y='duration', hue='telephone', facet_kws={'legend_out':Fals
         plt.title('Scatterplot with linear fit of duration and amount split on with telephone ownershi
         plt.show()
```

```
/Users/johanoelgaard/anaconda3/lib/python3.11/site-packages/seaborn/axisgrid.py:118: UserWarning: The f
```

```
self._figure.tight_layout(*args, **kwargs)
```



Thus, though it might not seem intuitive at first glance – it does make sense that people with telephones are slower to pay off their loans as they are more likely to have smaller loans that they will pay off quickly.

Question 2.2: In the cell below, explain the regression summary. How do `age` and `telephone` affect loan duration? Are the findings significant? How do you interpret these findings in relation to your answer from 1.5?

I find

$$duration = \underbrace{0.009708}_{(0.003325)} \cdot age - \underbrace{0.245381}_{(0.077616)} \cdot telephone$$

Thus, the coefficient for age is 0.009708, and the coefficient for telephone is -0.245381. The p-values for both coefficients are less than 0.05, indicating that both coefficients are significant. The coefficient for age is again positive, indicating that as age increases, the duration of the loan increases – this could make sense from the intuition that older people have a harder time paying off their loans, or it could be a relation similar to what I mentioned in Q1.5 older people might have larger loans. The coefficient for telephone is negative, indicating that those with a telephone are quicker to pay off their loan. This is interesting as it contradicts the survival curves from question 1.3. However, it does align with the intuition I mentioned in Q1.5. People with telephones have larger loans and, thus, take longer to pay them off, but especially for larger loans, they are quicker to pay them off than those without a telephone.