1. **Could you quickly tell me what you changed?**

Nothing changed I just added more explanations on the statistical significance of the models and the relationship between the relevant models.

1. **The distribution plotting of revenue size is new - why did you do it**

All the models specified that revenue size was log transformed, I therefore transformed the variable to log , instead of doing it everytime in the model initialization. This reduced redundancy, and saved time. I plotted the density plot of the var before and after log transformation.

1. **0.2.1 ⁃ Stupid question: What is on the x and y axes of the plots?**

On which plots: For density plots, x is tha range of values a variable occupies, y is the frequency of those values.

1. **0.3.0.1 ⁃ Correlation value of 1 would mean perfectly correlated - correct? What number would be still acceptable about the correlation?**

One of the things that instructions told us to look at was multicollinearity: Multicollinearity exists whenever an independent variable is highly correlated with one or more of the other independent variables in a multiple regression equation. Multicollinearity is a problem because it undermines the statistical significance of an independent variable.

A value for multicollineariy is not set, but 0.7 and -0.7 is the value that was used to determine multicolinearity.

1. **0.3.1 Assessing Significance of Relationship between numeric variables ⁃ Doesn’t this already has a degree of interpretation? That this and that variable has a significant relationship with CAR1 or 2 - when focusing on the first 2 lines?**

That was the case, I tested all the numerical variables with the each other. Pay attention to the correlation with CAR variables. All the variables have a statistically significant relationship at a 95% level of confidence, except, the CAR variables.

1. **0.3.1.3 Distribution of Car2 generally and During and After Covid19 Which method of winterizing was used here? Lower weight or similar value for the outliers?**

Extreme values that were considered for removal were 5% percentile.

1. **You wrote „Outliers negatively affect the accuracy of the model, hence the winsorizing“. How can I detect that now, after the winsorize, the distribution looks better?**

By taking a look at the change in density plots, the tails are no longer thin and extended.

1. **0.4 HYPOTHESIS 1 Could you add a description of the box plot?** Done
2. **0.2 HYPOTHESIS 2 What happened? => The coefficients and also the significances for the first regression are now different to the output yesterday. Why? Size revenue is statistically significant now.**

That’s because I had not set the random/state or seed therefore, everyitme a model samples data and models the results are different.

1. **0.6.1 COVID 19 VS EBITDA Conclusion p.28: Mistake or? The revenue change and not PE dummy is meant in the last sentence?**

That was a mistake was supposed to be revenue\_pos

1. **Conclusion p. 29: Where do you see the significant p-value? „Taking the p-value of the model at a 90% level of significance, the relationship between car1, p\_e\_dummy and COVID 19 is statistically significant.“**

The model’s p-value is last value on the model’s summary.

1. **Conclusion p. 31: „Note, however, that when relatively small sample sizes it is not recommended to estimate a relatively high number of parameters, so it would be better tbo test for simple effects with the given datsaet.“ => What would be better to test? Which simple regression do you have in mind?**

The cash models that you suggested.

1. **Are the regressions just done with the subsample? Because it seems, see first regression model „151 degrees of freedom“ would suggest that we have a sample of 151, which would be the full sample size Is here therefore COVID-19 set =1?**

I have set covid19 to 1 as requested on theextra models.