**Q1) What were challenges for reproducibility in trying to reproduce the original results by Walt Hickey?**

There were multiple challenges faced for reproducing the original results of Walt Hickey.

The author has mentioned that contradictory claims were made many times in the article.

*“*

*Hickey argues that passing the test "didn't hurt its investors' returns", which is to say there was no significant negative relationship, but neither was there a significant positive relationship:*The model provides no evidence of a positive correlation between Bechdel scores and financial performance.*However, Hickey switches gears an in the conclusions, writes:*

*...our data demonstrates that films containing meaningful interactions between women****do better at the box office than movies that don’t****...*

*I don't know what analysis supports this interpretation.*

*The appropriate conclusion from Hickey's analysis is "there no relationship between the Bechdel test and financial performance," which he makes... then ignores.*

*“*

Moving forward the other challenges that were faced were as follows:

1. **Data extraction and interpretation was not clear:**

Data was used at many points without any clear explanation of the different features/columns that the data had. Hence making it difficult to reproduce the analysis conducted by Hickey. There was also no documentation of how data was used hence making it difficult to understand what exactly was going on.

1. **Unexplained variables:**

In the analysis different resulting variables were formed like

1. Return on investments
2. Gross Profits

These variables were made using different columns from the dataset present however, it was not clear which columns were used and what was the interpretations of the columns used. Like for example return on investments was with tax or without tax? Or, is the dollar value adjusted for inflation? These questions were not answered.

1. **Incorrect Assumptions:**

The article from Walt Hickey makes incorrect assumptions about the data while conducting tests. Like the variables created and mentioned above (Return on investments and Gross Profits) are considered to be independent. Whereas there is a high chance that they are related.

**Q2) How well do *you think* you would be able to reproduce the results of Dr. Keegan’s work? Why or why not?**

I think that I will be able to replicate the results of Dr. Keegan’s work pretty well. There are multiple reasons for that

1. The notebook is properly automated and documented. Everything is explained clearly and in depth. Moreover, each step in the notebook and its subsequent step is very clear in its content and relationship thus making it available easy to follow.
2. The programming language used is python which is widely understood and used. Many online resources are available to explain any function or method used that I might not understand at the first glance. Moreover, since everything is done being using python no extra software licensing is needed to reproduce the work thus making it easy for me.
3. The data is publicly available
4. Lastly, the results of the analysis are displayed very clearly and summarized in the end. Since I now know the conclusion of the work done by Dr. Keegan, I would be able to reproduce it.

There is one reason why if in future I want to reproduce the work I might face problem is that

1. The dataset is uploaded on a random link and URL is provided to us. However, in future there is a chance that the URL does not work or is removed. Thus, creating a problem for me to access the data. Therefore, if the data had been uploaded somewhere more reliable it would have been better.