

Peer Assessment

In assessment of the provided project, each section will be assessed individually with individual criticisms, and then the project as a whole will be assessed.

1. Introduction:

The introduction is strong, but succinct, almost too much so. I would argue that a bit more might strengthen the reader's understanding of what the project was about and some more background information about data itself and the vaccinations. It does not say that the data comes from the Google Trend data for Denmark which I think is an important distinction (as opposed to worldwide or another country). The use of personal pronoun I is another mechanical criticism albeit less important. Otherwise, a good, concise introduction where the vaccines are clearly identified and the project is successfully overviewed to introduce the reader to the subject.

2. Methodology:

a. Trends:

This section is well written with a clear description for "stopwords" as well as some valid self critique. Methods for stripping punctuation are left out (Not positive or negative just something I noticed). The writer clearly understands the academic level of the project and does not waste words or the reader's time rambling unnecessarily.

B. Mining of Trends

This section is again, very compact, which is good but misses a couple key points about the nature of the data. There is no mention of the fact that the data gathered is not absolute data but relative data, Google Trends compares the queries against each other not against a global measurement index. And again the personal I usage here is probably not applicable to an academic paper. Otherwise, the section is neatly organized, with consecutive tasks described quickly and effectively and saves space for possible room for error (a more advanced script, and poor trend data criticisms are covered).

C. Prediction:

Clearly describes and annotates the libraries used for prediction. Perhaps the summary of lasso and k fold cross validation were a bit excessive, but nonetheless relevant to the project (I would assume the reader would understand these concepts beforehand).

3. Findings:

The graphs provided were applicable and relevant to the project. The writing is very compact, again, I would say more explanation might be helpful to the reader. For example the actual frequency data was left out of the findings, which might help the reader understand the project more clearly. I found it easy to follow, but those that have not taken the subject in a long time might find it more difficult, especially with regards to the root mean square error. Overall, the section clearly and effectively presents the data in an understandable way.

4. Conclusions.

Here the writer puts all the conclusive evidence towards an assessment of the predictive model, yet there is little expansion upon the connection between the predictive dataset and the observed dataset. It simply says "But it's hard to be sure of the exact usefulness of the models as is." Which is true but does not explain why this is the case other than stating that the error is larger in the predictive model. I would say that more here would strengthen the project, especially when elaborating on the connection between the frequency output from google and the relevancy towards a predictive model. The writer shows a clear understanding of the methodology and concepts around the project, but I believe a critical thinking (i.e. what do the predictions mean? How do outliers affect the predictive model? What other factors might be involved?) section might be helpful in expanding upon the findings of the data.

Overall I thought it was a very good project, the writer clearly understands what he/she is doing and how the data is taken and analyzed. However, I think more expansion upon the findings and perhaps connections between data could strengthen the project enormously. Again, elaborating upon the connection between the mined and observed data is one area for improvement. Otherwise, it is cleanly and effectively presented, with well labeled graphs and tables. The english is very easily understood and the ideas are relevant and helpful. I thought it was a very well done project.