The impact of COVID-19 on happiness

Mantas Rumskas, Zeno Kujawa and Andrej Jovanović

During the coronavirus pandemic, many people were subjected to stay-at-home orders in order to curb the spread of the disease. With this, individuals had to sacrifice their livelihoods and endure great hardship. In order to quantify this sentiment, a questionnaire was conducted by NORC at The University of Chicago, then later shared with the Associated Press, asking US citizens to share the experiences based on social, economic and physical factors relating to the coronavirus pandemic. The question that we investigated was which of the afore-mentioned factors affected the happiness of individuals the most.

The questionnaire was conducted on two occasions: once in April and once in May. This was to give a fair reflection of how the personal implications of the coronavirus changed over a sustained period of time rather than a particular instance. The dataset that was used can be accessed publicly from the following link: https://data.world/associatedpress/covid-impact-survey-public-data

Upon embarking on our question, the first task was to clean the data so that both the continuous and categorical data can be used in the polynomial regression models we would ultimately create. We created an unhappiness score which would allow us to quantify the emotions that participants were experiencing. With this, we had to assign values to certain categorical variables so that they could be eventually totaled in the unhappiness score. The explanation of this is that a minimum score of 0 indicates that a person experienced no emotional difficulty, per se, during the pandemic, whereas a maximum score of 15 indicated significant emotional difficulty and trauma.

In order to create the polynomial models to predict happiness, we split the April and May datasets into train-test. Secondly, we split each of the polynomial models based upon the category of question that was asked, namely economic, social and physical. We then used forward and backward selection to see if the regressions could be simplified. This was the case only with the economic factor model.

Our predictions yielded strong results, where our adjusted R-squared values were well over 15%, which can be considered high given we were trying to predict something as complex as happiness. Our results for each model can be viewed more specifically in our slide deck. A limitation of our findings is that we did not have a comparative dataset pre-COVID-19. With this, we would be able to discern whether these levels of unhappiness were caused directly by the pandemic or that it falls as a part of a larger mental health issue in the United States. Another limitation that would be possible is overfitting as many variables were used for the models. Because of that we used the adjusted R squared hoping that it would point to any bad models that only seemed to work due to overfitting.

Ultimately, we see that the coronavirus had a great effect on people's levels of happiness across the board. This was inline with our original hypothesis.