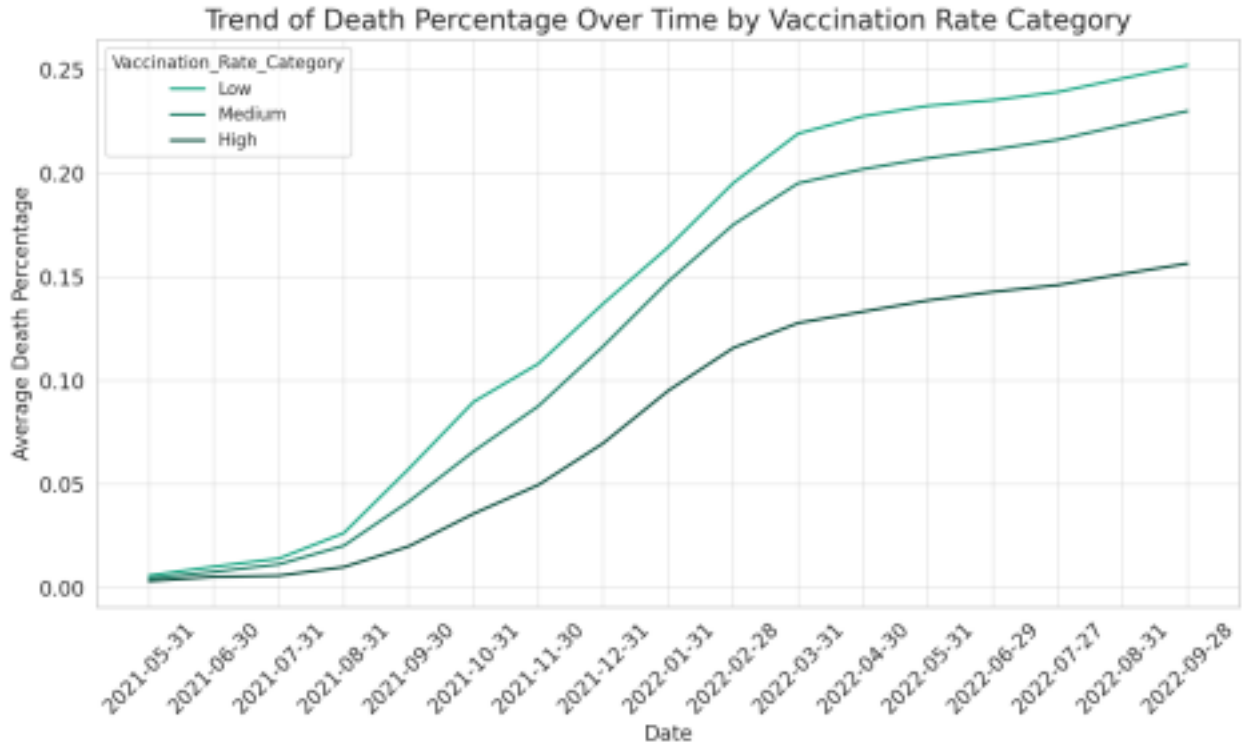
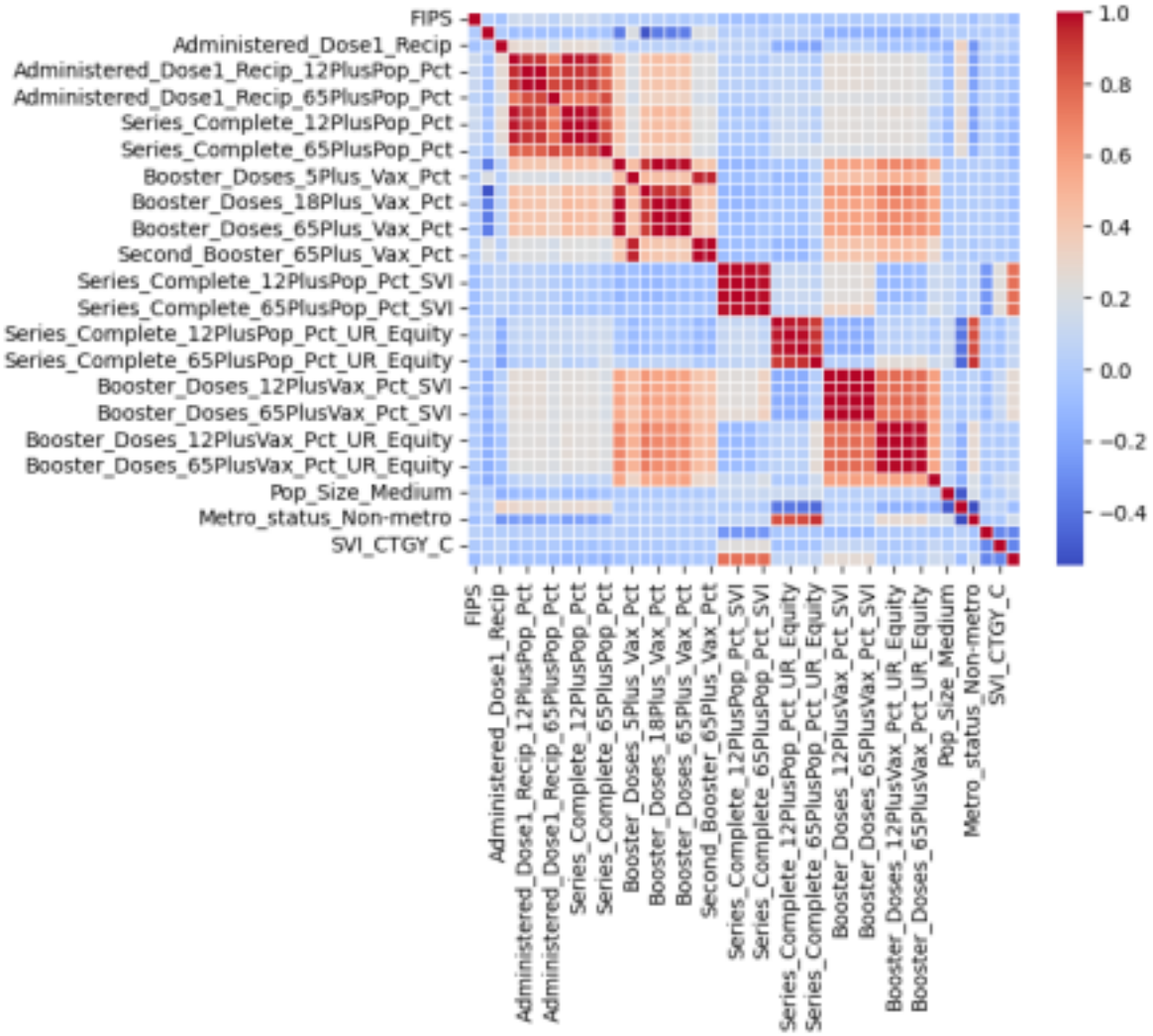
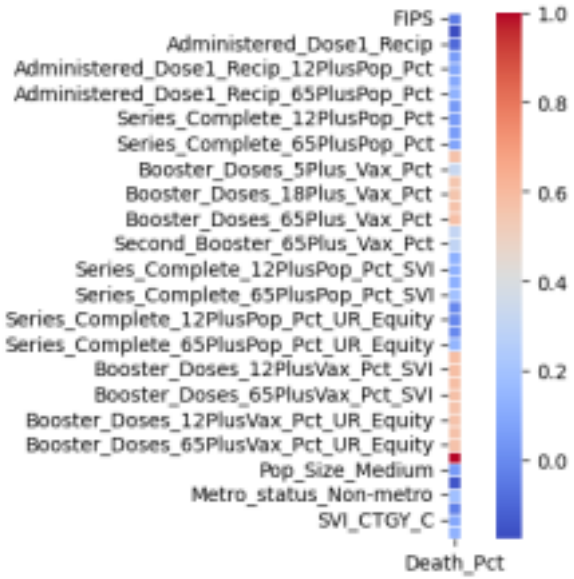
Visualizations



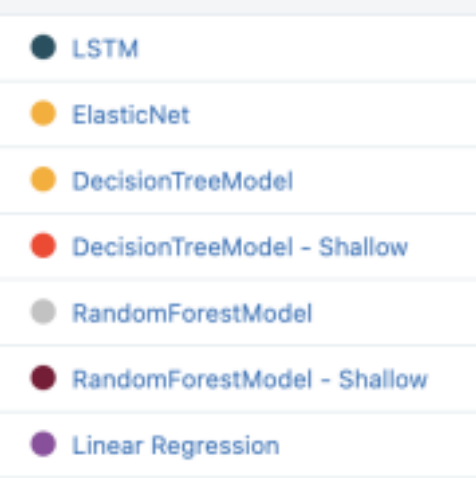
This line graph displays the trend of the death percentage over time, categorized by low, medium, and high vaccination rates. It helps in understanding how death rates evolved in different categories of vaccination rates.

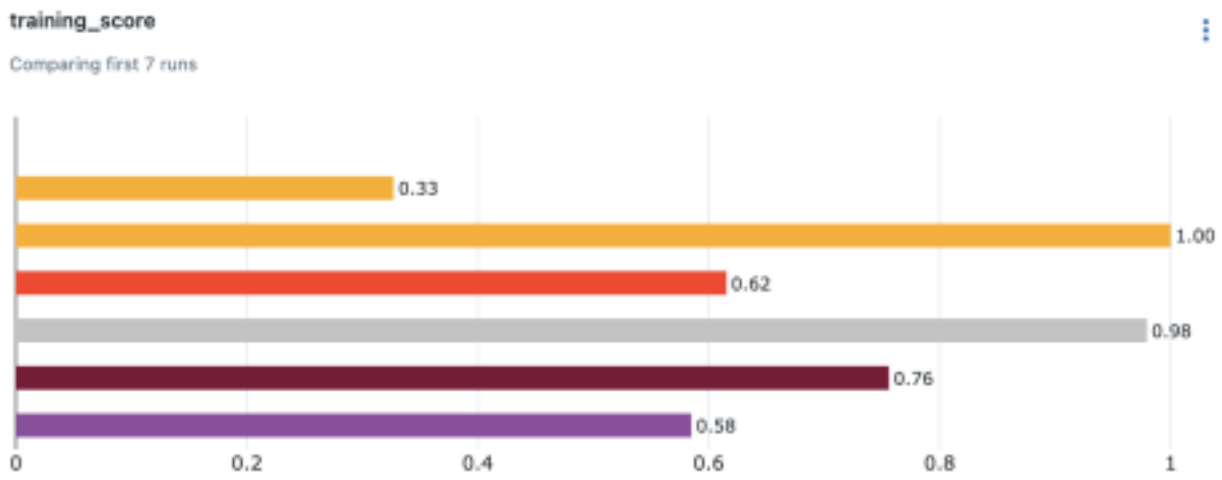


This diagram is a feature correlation map. It can reveal patterns or clusters of variables that are closely related, which might indicate shared underlying factors or causes. It helps in selecting features for machine learning models. If two features are highly correlated, one of them might be redundant, and the heatmap provides a visual tool to identify these relationships.

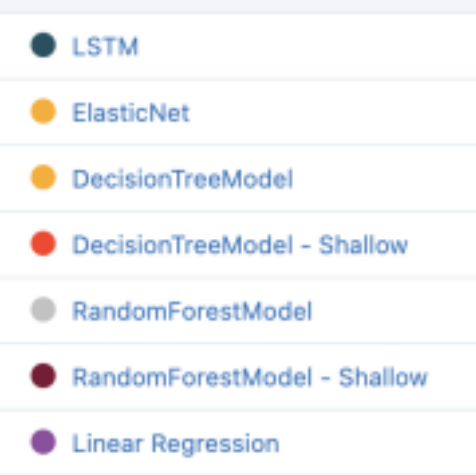


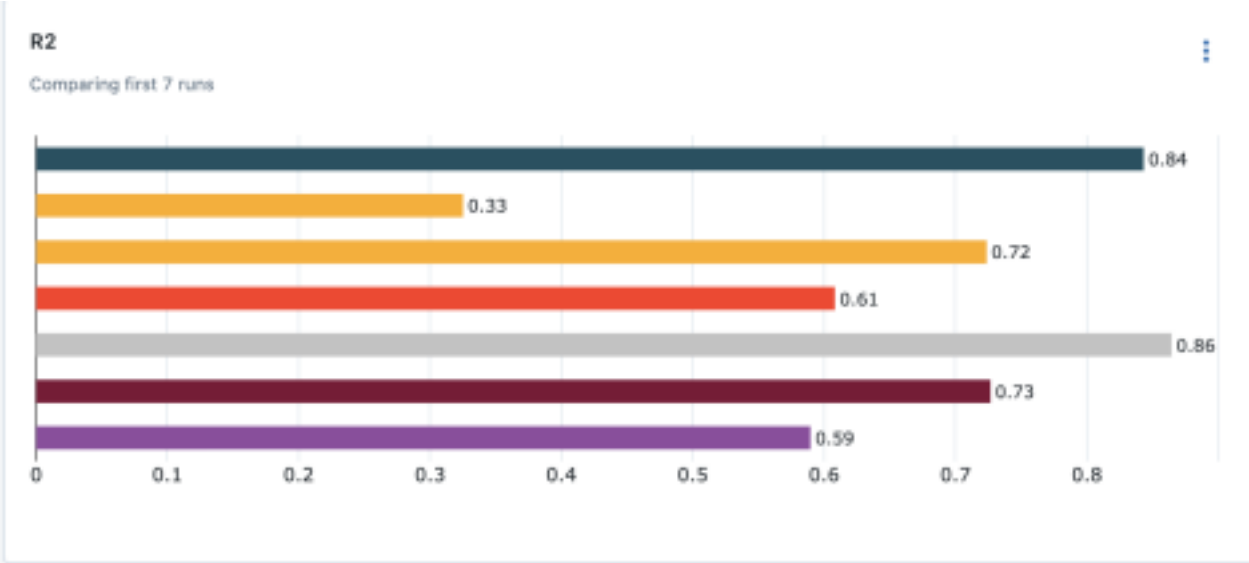
This visual demonstrates the correlation between each feature and our dependent variable Death_Pct. It can be seen that booster doses have relatively higher correlation. This diagram help us assess which of the parameters are good to use to fit models



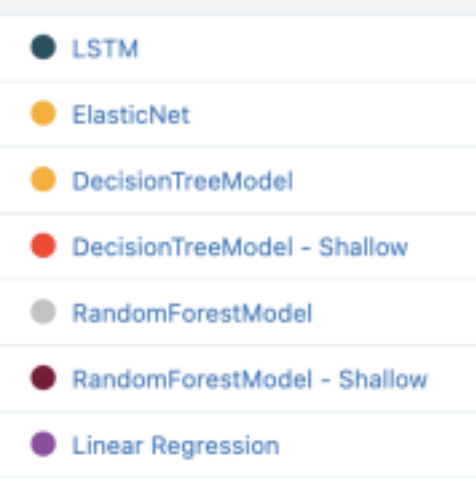


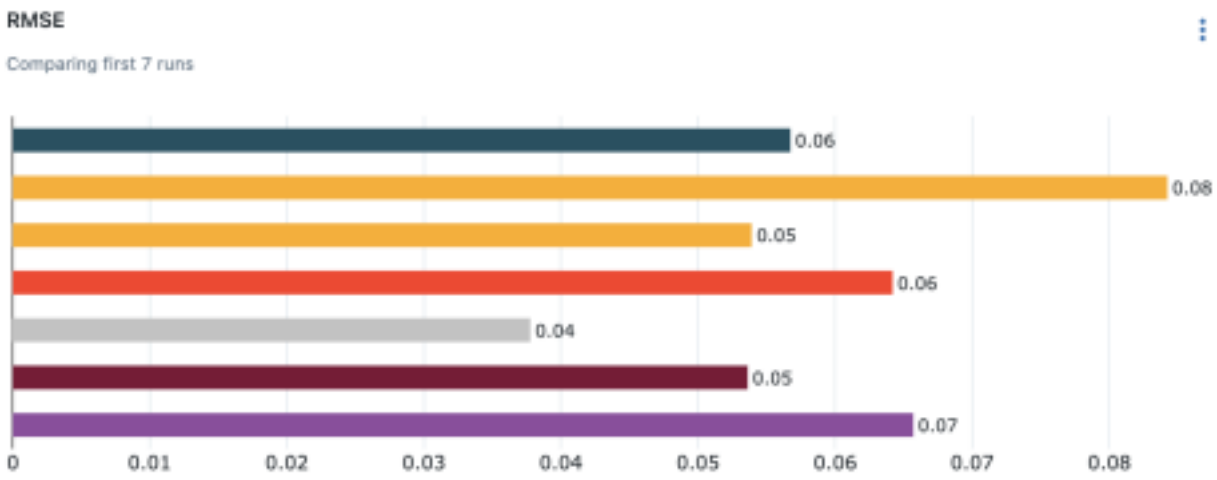
The training score diagram helps us determine if a model has been overfitting: It seems like the Decision Tree model is overfitting since it has a training score of 1.



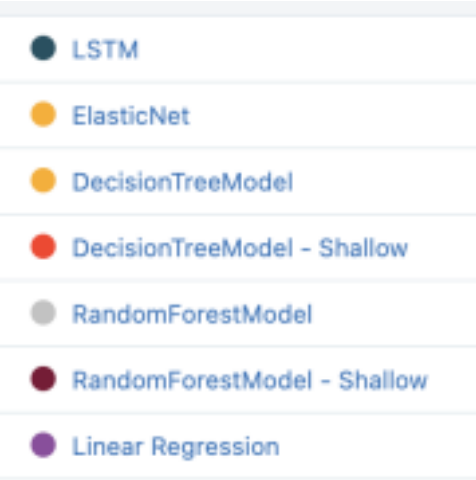


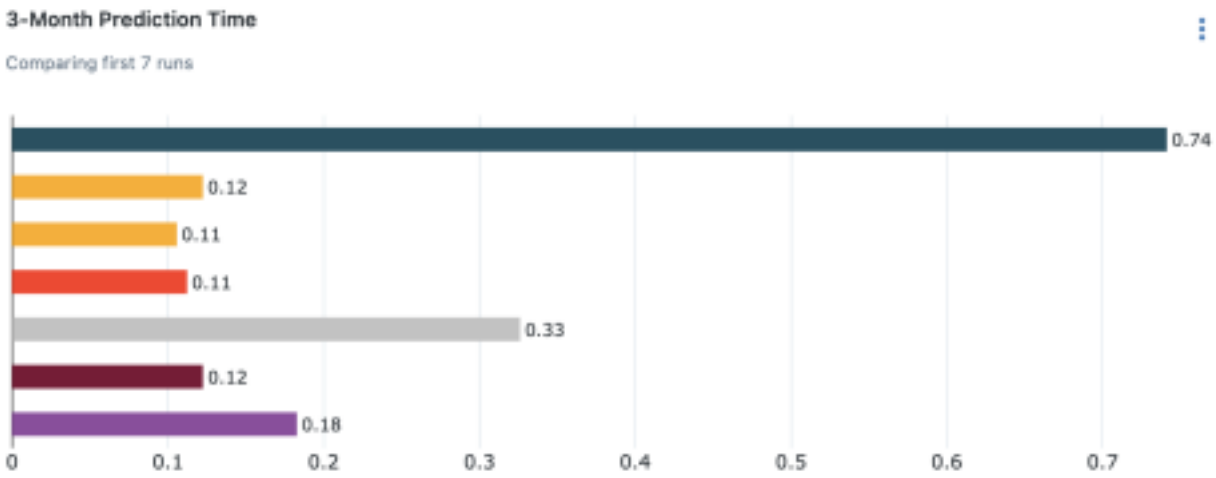
The R2 bar chart shows how much the variations were explained by our model. It can be seen that Random Forest has explained the most with a R-squared of 0.86, followed by the LSTM model with a R-squared value of 0.84



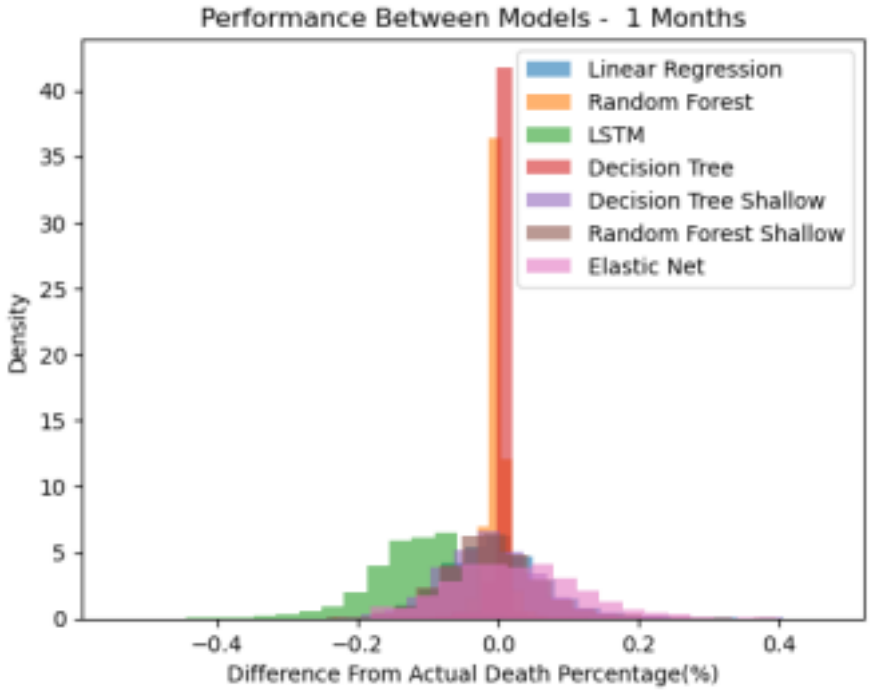


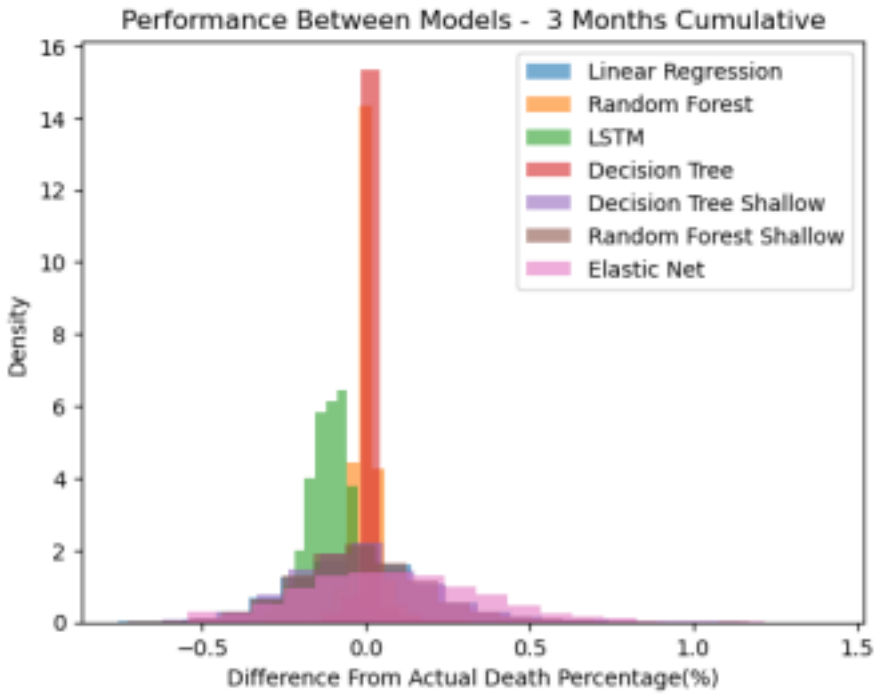
The RMSE value is a measurement of the magnitude of the errors. A smaller RMSE value indicates a better fit between the predicted and the actual values. In our case, random Forest Model has the lowest RMSE.



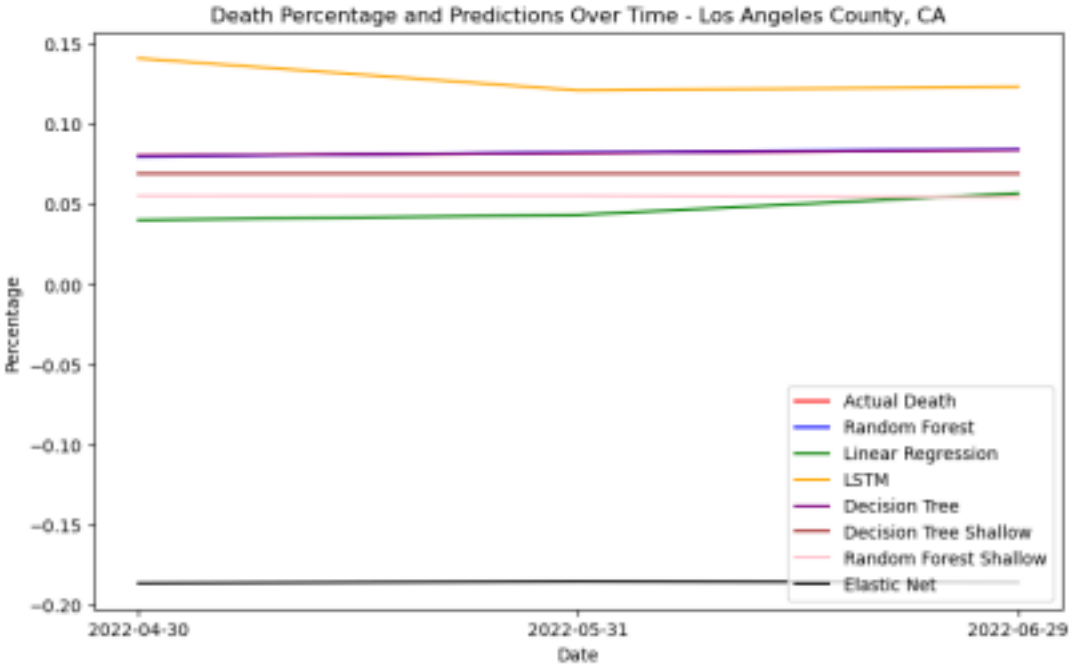


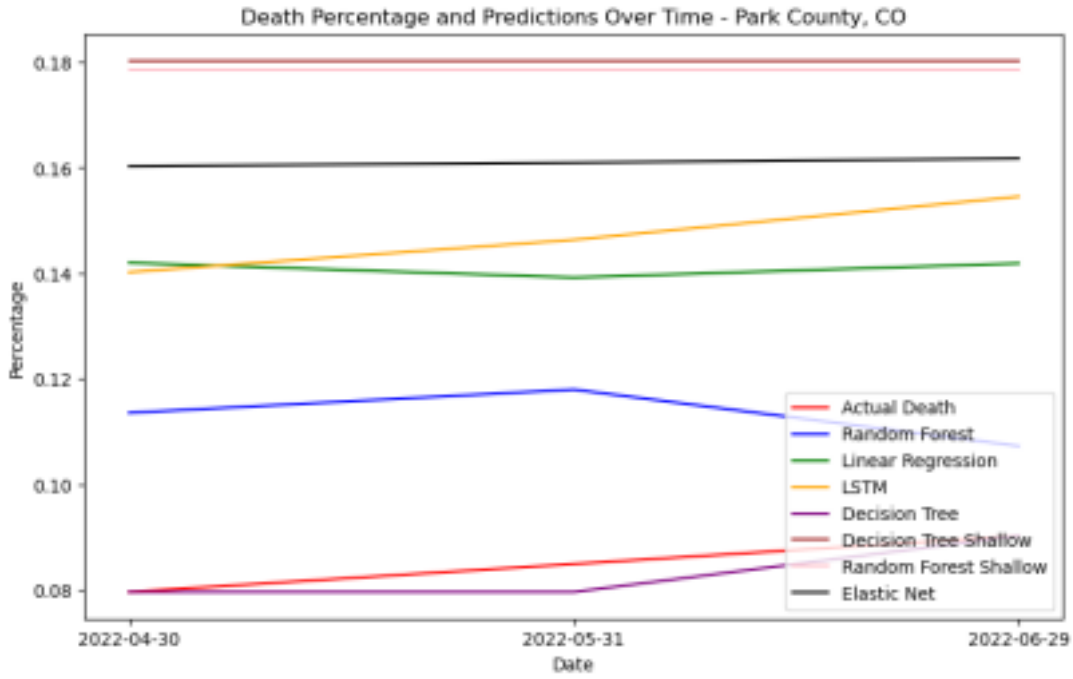
This diagram is the comparison of run time to make three-month predictions, LSTM model shows a significantly slower running time compared to other models.





These two graphs show the distribution of the differences between the actual and predicted death percentage, where the y-axis indicates their frequency of occurrences. Based on these visualizations, we can conclude that the random forest model performs the best among all models. It's also quite interesting to see that the LSTM model was underestimating for all time periods.





These two line charts show the difference between Actual Death percentage predicted for the two counties compared to predictions from each model. It's once again obvious that the random forest model makes the best predictions. In addition, the random forest model seems to be doing especially good when the population size is large. One possible explanation could be that individual effect is less significant in large counties and the pattern learned by random forest model is more applicable.