These are some of the questions:

1. What is your research question?
   * Can we predict how many deaths each state will have in the next 30 days
2. What are your hypotheses?
   * NA
3. What are your citations from statistical research?
   * Gecili, E., Ziady, A., & Szczesniak, R. D. (2021). Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy. PloS one, 16(1), e0244173.
   * Lee, J. H., Shi, Z., & Gao, Z. (2021). On LASSO for predictive regression. Journal of Econometrics.
   * Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2020). Machine Learning Advances for Time Series Forecasting. arXiv preprint arXiv:2012.12802.
4. What statistical methods did those researchers use?
   * Holt Modeling (linear exponential smoothing)
   * ARIMA Modeling
   * Trigonometric Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal component
   * cubic smoothing spline model
   * LASSO regression
5. What attributes did they find important?
   * Date
   * Number of Deaths due to COVID-19
   * Time
6. Which statistical methods will you use? Normality test, parametric or nonparametric test, post-hoc test. (Remember you must use one or more of the eight principal statistical methods: t-test, z-test, ANOVA, regression analysis, Kruskal-Wallis, Wilcoxon, Chi-Square, Fisher’s Exact Test.)
   * Direct Time Series Forecasting
     + Creating a series of distinct horizon-specific models
     + I will be using a LASSO and Random Forest
7. Does the data fit what are you intending to do?
   * Yes
8. Which tool(s) will you use? R, Python, SAS. Provide intext citations supporting your choice.
   * The programming language that I will use is R and I will present my finding using RMarkdown. R along with its many libraries allows for extensive analysis of data. RMarkdown provides a framework to combine your code, results, and narration for quick, reproducible reporting.
9. Has the data been studied statistically by other researchers? If so, what statistical methods did they use? What were their findings?
   * Other researchers have studied death due to COVID-19 using ARIMA modeling. Their findings were that death cases of COVID-19 may have an exponential growth, they were able to give semi-accurate death predictions.
10. What are the limitations and delimitations of your study?
    * The limitation of my study is that COVID-19 is still fairy new and a solid understanding of the transmissibility and death rates is not there.
    * The number of deaths is underreported.
    * The model will be constrained by the data.
11. What are the number of rows of data in your dataset? Min. 7,000 required.
    * 800,437
12. What is the data sparsity (%)?
    * 3.03%