COVID Sports Viewership Impacts

Case Study Rubric

DS 4002 – Fall 2024 – Adam Snyder
Due: Dec 10
Submission format:

- Document containing all deliverables assigned below

Individual Assignment

Why am I doing this?

In this case study, you will use your data science knowledge by performing time series analyses on historical data to determine the impacts of COVID on American sports viewership and assess how well these sports have recovered since the introduction of COVID. As you complete this assignment, you will practice time series analysis techniques and learn ways to predict trends in time series datasets. You will assess differences between expected and measured viewership in all years since the spread of COVID impacted each sport, such that you can assess immediate impacts as well as recovery.

What am I going to do?

The GitHub repository for this case study can be found at https://github.com/arsnyder23/DS4002_CS3. You will access datasets showing recorded viewership of major league sports championships, including NFL, NBA, MLB, and NHL. These datasets are available in the GitHub repository. Next, you will do some background research on double exponential smoothing, learning why this technique can be used for this application and learning about key parameters you will adjust. Then you will split data for each sport into pre- and post-COVID data, separating data from start-2019 from data 2020-end (except for NFL, which splits between 2020 and 2021). You will then create a double exponential smoothing model from each sport's pre-COVID data to forecast viewership as if the COVID pandemic had never occurred. Lastly, you will compare this forecast to the measured post-COVID data to assess how much viewership differed as a response to the pandemic.

Your final deliverables include:

- A summary, describing the process you completed and why you took the steps you did.
- A short discussion, stating key findings from your case study and answering your research question
- A plot of viewership over time for each of the 4 leagues, showing 3 curves:
    - The recorded data over all years (pre- and post-COVID)
    - Data from the fitted model over pre-COVID years
    - Forecasted data from the model for post-COVID years
- A plot showing percentage difference residuals between forecasted data and recorded data for each year, starting in 2020.
- A python notebook containing all analysis code and appropriate documentation

Tips for success:

- Use available resources, such as references and documentation to understand the techniques
- If you are unfamiliar with this type of coding, use AI tools for assistance to get you started
- Question each step, asking *Why am I doing this? Is this step the best way to reach my goal?*

How will I know I have Succeeded? You will meet expectations this case study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	● A 1-2 page PDF document with the following information: ○ Name, course, and date ○ Title ○ Summary Paragraph ○ Discussion Paragraph ○ 4 Viewership vs. Time Plots ○ 1 Percentage Difference Residual vs. Time Plot ● A python notebook of source code and outputs, appropriately documented
Document	● The document will contain: ○ Appropriate headers, such as name, course, date, and title ○ A summary paragraph describing the process you completed and why you took the steps you did ○ Each of the 5 plots assigned
Code	● Your python notebook will meet the following criteria: ○ Organized and documented, allowing for easy comprehension or reproduction of results ○ Read in data in .csv format ○ Perform expected analysis ○ Print assigned plots

Acknowledgements: Thank you to Professor Alonzi for providing the rubric structure.