Medicaid Case Study Rubric

DS4002 – Fall 2024 – Anastasia Nicholson
Due: Dec 9th
Submission Format: Upload link of GitHub repository on UVA Canvas

Individual Assignment

Why am I doing this?
This case study allows you to utilize and expand your data science knowledge by using moving average time analysis to determine if the adoption of Medicaid in certain states in the US will decrease the average cost of healthcare over time compared to states in the US that did not accept an expansion policy.

What am I going to do?
The GitHub repository for this case study can be found at https://github.com/anajonicholson/DS4002_CS3/. You will first obtain the datasets consisting of quantitative yearly trends in health expenditures and the annual medicaid enrollment across the various states from the Centers for Medicare and Medicaid Services. The repository contains these datasets that have been cleaned to classify all of the quantitative data as integers, and remove rows and columns that do not need to be used. Then, you will normalize this data according to the population and average annual income in each state. This can also be found in the data folder of the repository. From here, you will perform moving-average time analysis through ARIMA methods, beginning with checks for stationarity, then, if the series is non-stationary, you will apply differencing. Next, parameters for ARIMA will be set through various autocorrelation plots to further fit the model to the data. From there, forecasts for future expenditures can be calculated and reviewed using time series plots.

Your final deliverables should include:
- A visualization of the moving averages of medicaid costs overtime
- A data dictionary
- Well documented commented source code
- A GitHub repository containing all materials used

Tips for success:

1. Do not become overwhelmed by the various states and data
    a. Search for general trends and alternative correlations rather than looking at each state individually.
2. Make your visuals of moving time average analysis very simplistic and not over complicated to look at.
    a. This could include adding a legend or organizing sections rather than showing everything possible.
3. Familiarize yourself with python and ARIMA techniques to perform the analysis with lower confusion.

How will I know if I have succeeded?

Spec Category	Spec Details
Formatting	One GitHub repository (submitted via link on Canvas) Create a new GitHub repository for this assignment titled 'CS3_Medicaid' that contains ▪ README.md ▪ LICENSE.md ▪ Source Code File ▪ Your data (i.e., the enrollment numbers and state populations/annual income) ▪ REFERENCES.md
README.md	Brief summary of what you've produced for the case study, this does not have to be very detailed but should provide enough information to orient people to your repository.
Source Code File	Well documented Jupyter Notebook file that contains the code used to execute your time average and statistical analyses. In the source code, you must include: ▪ All datasets used, including medicaid enrollment, medicaid costs, annual income, and state populations. ▪ Moving time average analysis using ARIMA. ▪ Visual outputs expressing the correlation between the medicaid enrollment and the

	medicaid costs annually.
References	Markdown File titled “REFERENCES.md” with citing any resources (journal articles, websites, etc.) referenced in helping you create your model in IEEE Documentation style. Also include brief annotations under each citation on how each reference informed/helped you for this case study.

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