Case Study 3: Investigating Crime Patterns in Los Angeles (2020-2024) Rubric

---

DS 4002 - Fall 2024 - Instructor: Adam Chow
Due Date: Monday, December 9th
Submission Format: Upload link to Github Repository

General Description

Submit a link to your GitHub repository, which should include all necessary components to understand and replicate the analysis. This case study will guide you through exploring seasonal crime trends, evaluating time series models, and drawing meaningful insights from predictive analytics.

Why am I doing this?

This case study will be using the Github Repository, https://github.com/a-chow3/Case-Study-3/tree/main, as the backbone of the analysis and evaluation methods. After you download the dataset from the link in the DATA folder, you will first wrangle the data in order to put our dataframe in a position to conduct meaningful analysis for our goals. After you load in the dataset and wrangle the data, you can proceed to the EDA section. Now, you should explore and analyze temporal first patterns in the crime data. Next, you will run the Part A code in order to perform ARIMA and SARIMA analysis on the crime data. After this step, you will run the Part B code in order to perform Random Forests on the crime data. Following the two Juypter Notebook scripts, you should then draw your own insights and conclusions to guide your potential solutions. You'll leverage your knowledge of time series modeling and develop insights into crime patterns, learning how data science impacts urban safety and policy.

Course Learning Objectives:

1. Explore and analyze temporal patterns in crime data.
2. Apply ARIMA and SARIMA models to understand seasonality and trends.
3. Draw insights and conclusions to guide decision-making.

Tips for Success

- Ensure all file paths are correctly adjusted to your local directory when running the notebooks.
- Focus on interpreting outputs—your ability to derive meaningful insights is a key component of the deliverable.
- Approach the analysis with an open mindset since handling time series data requires multiple levels of nuance that need to be given adequate thought.

- Pay attention to seasonal trends and feature importance as these will guide your conclusions.

Evaluation Criteria

Spec Category	Details
Formatting	- One GitHub repository submitted via Canvas. - Repository follows the specified structure: `README.md`, `LICENSE.md`, folders (`DATA`, `REFERENCES`, `SCRIPTS`, `OUTPUT`). - Proper organization for reproducibility.
README.md	- Summarizes the analysis process and results clearly and concisely. - Outlines the hierarchy of folders and subfolders. - Includes instructions for setting up and running the notebooks.
LICENSE.md	- Uses MIT License for project sharing.
DATA Folder	- Contains a link to the dataset and a data dictionary.
REFERENCES	- All references need to be listed at the end of the document in IEEE Documentation Style.
SCRIPTS Folder	- Includes two well-documented Jupyter notebooks: - Part A: ARIMA and SARIMA Models.ipynb: Implement and run time series models without modifying code. - Part B: Random Forest Model.ipynb: Run and analyze Random Forest predictions without modifying code. - Students must draw their own conclusions and interpret the results.
OUTPUT Folder	- Contains visualizations (e.g., actual vs. predicted values, feature importance) and final conclusions.

How Will I Know I've Succeeded?

You will meet specifications when you:

1. Follow the GitHub repository structure and include all required components.
2. Run the provided scripts correctly and produce accurate visualizations.
3. Provide thoughtful and well-supported conclusions in your final deliverable.
4. Present your findings clearly in the README.md file.