Parking Ticket Predictor Case Study - Rubric

DS 4002 – Julian Krese
Due: TBD

Submission format:

- **A GitHub repository uploaded as a link.**

Individual Assignment

Preparatory Assignments – Everything in the course (Projects 1 and 2).

Why am I doing this? This case study allows you to leverage your data science knowledge by using time series data to train Isolation Forest models and use them as classification models. As you work through this assignment, you will be exposed to the ways that data analysis can be used in a real-world context and around your daily life.

- Course Learning Objective: accurately replicate experiments.

What am I going to do? You will replicate the provided case study, including cleaning data and then using it to train multiple Isolation Forest models. Afterwards, you will score their performances to determine the best model. There will be provided code and other materials necessary in a GitHub repository found at https://github.com/JulianKrese/DS4002-CS3-Ticket-Predictor/. You will begin by forking the existing repository. Next, you will read the hook and supplementary documents to better understand the context of this project. Now you will be ready to begin creating the deliverables, which will result from the existing scripts. The deliverables are as follows…

- A cleaned dataset from the provided original
- 5 trained models across time ranges…
    - All time (2000-2023)
    - Recent (2023)
    - 5 year gaps (2023-2019, 2018-2014, 2013-2009)
- Model performances displayed visually
- A GitHub repository containing all of the above in an “OUTPUT” folder.

Tips for success:

- Follow the steps outlined in the “README” to streamline the overall process
- Read the provided code and comments in “SCRIPTS”. They are well documented and can help speed the process along.

- You will be working with Python and related packages (sklearn, pandas, matplotlib). Familiarize yourself with the language and packages to streamline the process.

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

Spec Category	Spec Details
Formatting	● A GitHub repository containing the following **new** materials. Note that the existing materials from the original repository should still be there from the original fork, but add the following in the root. ○ DATA folder (new part) ○ OUTPUT folder (new part) ○ Reflection.txt
DATA	● Goal: Clean the original provided data ● Store this in a new directory, “DATA/Final/” ● There should be two resulting files, named “cleaned_parking_tickets.csv” and “encoded_parking_tickets.csv”
OUTPUT	● Goal: Reproduce the outputs of the case study ● Store this in a new directory, “OUTPUT”, which contains… ○ Models and their scalers in a subdirectory called “models” (10 files total) ○ A file containing each models accuracy called “model_performance.csv” ○ A bar chart showing comparing performances visually, called “model_performance.png”
Reflection.txt	● Goal: Reflect on the work done in the case study ● Split into three sections… ○ Data Cleaning ■ Summarize how the cleaning was done. ■ Write about one choice from the cleaning script and why the original author may have chose to do it this way. If it was a bad choice and it can be improved upon, explain why and what should be changed. ○ Models Training and Testing ■ Summarize how the training and testing for the models was conducted.

	■ Write about one choice from these scripts and why the original author may have chose to do it this way. If it was a bad choice and it can be improved upon, explain why and what should be changed. ○ Learnings ■ Write about one new concept that you learned while completing this project. This could be how the underlying model works, a particular import, etc. ● Each section should be concise but thorough enough to explain your perspective, roughly 1 paragraph. The goal is not to create a writing assignment but think about your results.

Acknowledgements: Thank you to Professor Guadagni for providing the overall case study structure and rubric.