Palantir Stock Prediction Case Study Create

DS 4002 – Emma Wunderly
Due: Monday after Presentation week of project cycle 3

Submission format: Upload document to canvas

Individual Assignment

Why am I doing this? This case study gives you a high-stakes, real-world mission: to apply fundamental data science skills to financial forecasting. The stock market is a non-linear, unpredictable environment, making it an excellent challenge to test your skills in time-series analysis, data wrangling, and model building. You will learn to tackle data integrity issues by merging two fundamentally different datasets—structured stock prices and unstructured media sentiment (GDELT)—to see if you can find an edge. Ultimately, you are doing this to gain practical experience that goes beyond textbook examples, giving you the confidence to approach a complex, messy, and fascinating data science problem.

What am I going to do? Your mission, as laid out in the Hook Document, is to build a predictive model for **Palantir Technologies (PLTR)**. You will use the provided **GitHub repository and data set** (historical stock prices and GDELT GKG media sentiment) to construct a full data science pipeline. Your core deliverable is a functional model that forecasts the **5-day stock return** for PLTR. You will be guided by this Rubric, which details the criteria for your work in feature engineering, model selection, and performance evaluation. By the end, you will produce a complete solution (Jupyter Notebook) demonstrating whether media coverage related to AI and war truly influences Palantir's stock volatility, and how well your model can predict it.

Github link: https://github.com/ewunder32/DS4002_CS3

Your final deliverables should include:

- Merged, cleaned, time-series dataset
- The predictive model used to generate the 5-day returns for PLTR
- Risk-adjusted return metrics
- Model vs. Benchmark comparisons
- Trading strategy investment summary

Tips for success:

- **Be Clear:** A clear presentation of fundamentals is more valuable than an unclear presentation of cutting-edge techniques.
- **Don't Overthink It:** You are solving a data problem, not predicting the future. Focus on building a sound, reproducible process.
- **The *Why* Matters:** Your interpretation of the GDELT data—linking news sentiment to stock volatility—is a key insight in this case study.

- **Resources:** Use the provided scripts and supplemental materials (blog post, technical article) to guide your understanding.

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

Spec Category	Spec Details
Formatting	• One GitHub repository (submitted via link on Canvas) • Create a new GitHub repository for this assignment titled '**DS4002_CS3**' that contains" • README.md • LICENSE.md • Source Code File (with results) • References
README.md	• Write a summary discussing what you've produced for the case study. This should provide enough information to orient people to your repository. • You should summarize the problem presented in the case study and discuss how you met the demands of the deliverable.
Source Code File	• Well documented Jupyter Notebook file and R Script that contains the code used to execute stock prediction analyses. The source code should include: • Merged, cleaned, time-series dataset • The predictive model used to generate the 5-day returns for PLTR • Risk-adjusted return metrics • Model vs. Benchmark comparisons • Trading strategy investment summary
References	• Cite any resources (journal articles, websites, etc.) referenced in helping you create your model in IEEE Documentation style.

Acknowledgements: Thank Professor Alonzi for providing the rubric!