

Roberto Reynoso

Analytics Deployment App

With in my GitHub Repo, All my work is in app_folder (Includes: Md. Doc, DataBricks HTML, and files to create my app.)

- My GitHub Repo: https://github.com/RvRproduct/app_challenge_sp22
- My Analytics Deployment App: <https://stormy-ravine-99578.herokuapp.com/>
- DataBricks and Contrast Thoughts: Comparing using Pyspark to Pandas, I see the Pros of Using Databricks than just using pandas in say visual studio. Everything is set-up so nicely and it becomes easier and faster to manipulate data, since pyspark and sparksql is really the back bone of looking at data.
I found it super useful switching from the two. I feel pretty confident now, with my ability to go through big data and pretty confident in using databricks as one of my main resources to work in data science.
- Docker: Is very similar to a virtual environment. You are able to make an image of your data/application(folder), while having certain dependencies that docker will for said program. Then it can be used as a container, which anyone (with the correct permissions) can use, without having to download dependencies on their own or having to download all the files for the program, Docker takes care of that for you.
- Statistical Regression and Machine Learning Regression Difference: Machine learning models are made to make the most accurate predictions as possible, while statistical models are made for inference about the relationships between variables.