How I Completed this entire project?

To complete this project, I began by gathering the time series data from the provided CSV file, which consisted of over 18,000 rows. My first step was to dive into Exploratory Data Analysis (EDA) to understand the data better. This involved checking the quality of the data, including identifying and addressing any missing values or outliers that could affect the analysis. Ensuring the data was clean and consistent was crucial for reliable results.

Next, I focused on the date column to make sure it was in the correct datatype, as this would be important for any time-based analysis. Afterward, I moved on to Feature Engineering, where I created new features or transformed existing ones to enhance the model's performance. This stage was about refining the data to make it more predictive.

With the data prepared, I split it into training and testing sets, carefully applying a sampling distribution to find the optimal split. This was essential to ensure that the model would perform well on unseen data. I then selected the most appropriate machine learning model for the task, trained it on the training data, and evaluated its performance using the test data.

For evaluation, I chose metrics that would accurately reflect the model's ability to detect anomalies, with the goal of achieving an accuracy of over 75% on the test set. To further improve the model, I performed hyperparameter tuning, adjusting the model's parameters to enhance its performance.

Finally, I developed a plan for deploying the model in a production environment. I also packaged the entire solution in a zip file, including a README that provides clear instructions on how to install and run the end-to-end pipeline. Additionally, I documented the process thoroughly, highlighting how these steps align with industry needs and how they could benefit the company in a real-world scenario.