DATA ANALTICS &
VISUALIZATION TEAM

RECRUITMENT ASSIGNMENT

Assignment Deadline

20th June 2024

Team Heads

Aman – 8075610955

Amish – 8120349405

You are the head of the sales and marketing division of “The Renewables”, a company that sells renewable energy solutions to people. Recognizing India's vast potential for renewable energy growth, you decide to extend your operations into this dynamic market. To make informed decisions, you need to understand the country's energy landscape, including its actual and installed energy capacities.

For this you turned to the NDAP for comprehensive data on India's energy sector,and obtained this dataset from there:

Q1. Ask 10 reasonably involved questions and try to answer them by analyzing the Dataset.

Q2. Seeing high levels of pollution in the country, you get curious about the trends in pollution and try to correlate it with the dataset for power plants you are given. How would you go about doing that? Is the current dataset sufficient to identify pollution trends? If not, what additional data would you need, and where would you obtain it? List the potential sources for acquiring this necessary data.

After extensive research, suppose you decided to enter the Indian Market with your existing products (P, Q, R, S and T).
Your team has determined that the behavior of the new market in India closely mirrors that of your existing market. In your current market, the sales team categorizes all customers into four segments (A, B, C, D). They then tailor their outreach and communication strategies to each specific segment. This approach has proven to be highly effective for them.

They plan to use the same strategy for the Indian markets and have identified about 2500 new potential trial customers. For this they have an existing database of about 8000 customers which they have already classified.

Q3. Implement methods to assist your team in identifying and predicting the appropriate group (A, B, C, or D) for each customer in the test data.

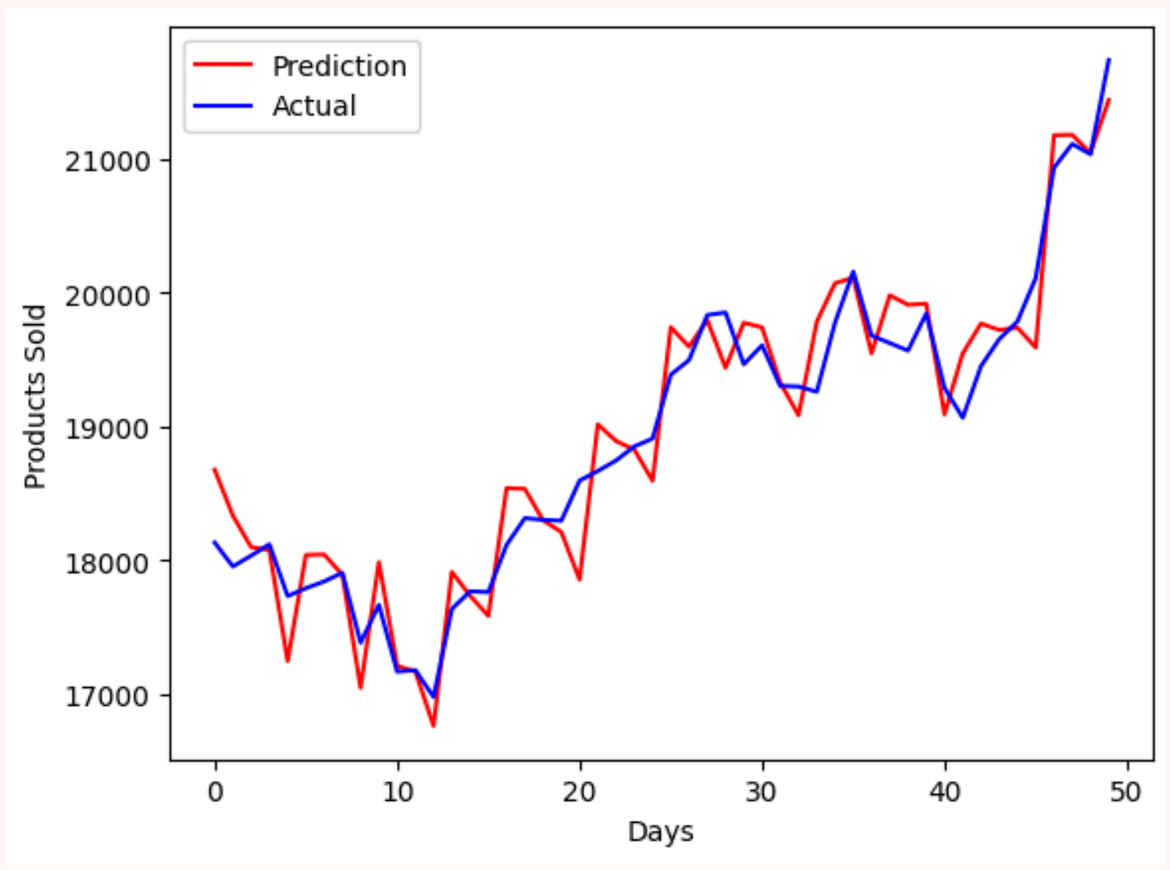
Q4. In the above case your team already had data from which customers were classified in groups. How will you predict the classes if the groups and it's data weren't available? State the method you would have used.

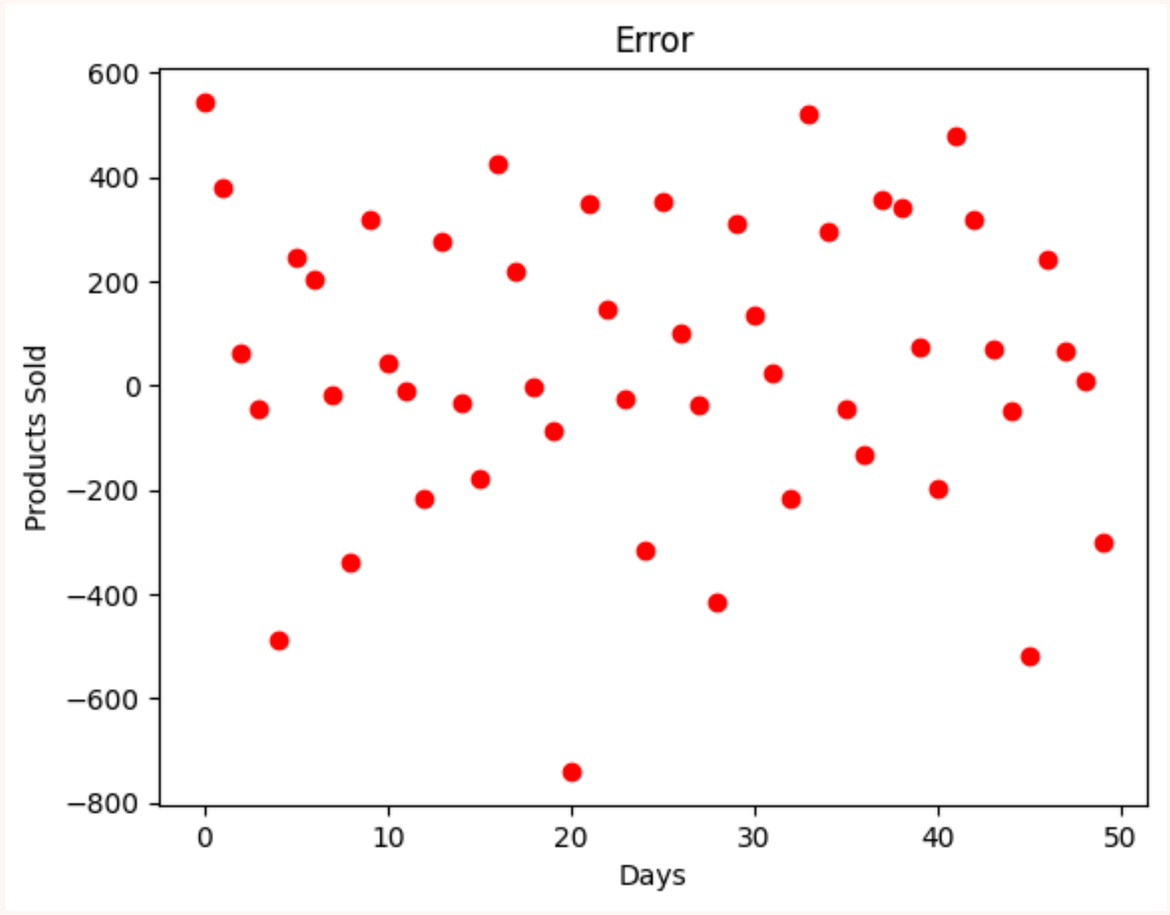
BONUS

Q5. Implement the method described in Q4 and compare the groups formed with A,B,C,D (Q3).

After launching your renewable energy products in the Indian market, you notice a positive response from customers. Sales are increasing steadily, and the demand for your products is reaching new heights. To ensure your company can meet this growing demand, you need to predict the number of goods that can be produced over time. This forecast will help you make informed decisions about scaling operations and supply chain management.

Given that your company relies on renewable energy sources, daily production can vary based on factors like weather, season, and other variables. Your team has developed a model that accounts for these factors, but its performance on test data has been disappointing, as shown in the accompanying figure.





Q6. What do you think might be causing the poor performance of the model? To improve the model's accuracy, what steps would you take? Provide a detailed justification for each of your proposed methods.

BONUS QUESTION

You have been provided with a dataset containing emails categorized as either spam or non-spam and another dataset with emails not yet categorized .

Q7. Develop a model to predict whether each email is spam or not, and use it to classify the uncategorized emails.

SUBMISSION INSTRUCTIONS

Please submit your findings in a report pdf format and include your workings in a Jupyter Notebook.
Also include excel file/tableau file/ whatever software used

Submit your predictions for Q2 and Q7 as a csv file

Q1, Q2, Q3, Q4, Q6 are compulsory questions

Remove debugging outputs from the Jupyter Notebook.