Greetings from ResoluteAI.in!

Thank you for your interest in our internship opening. As a next step in the screening process, you are required to complete the below mentioned assignment.

Role: Data Science - Intern

Dataset: [Datasets' Link](#)

Choose any one from the options below.

Option 1: Clustering/Pattern Mining

Complexity: Beginner

Task: Use any unsupervised technique to extract patterns or segregate data into groups.

User Story: User should be able to provide a data point (a row) and program should be able to identify to which group given data point belongs to and why?

Hint: ARM/ Clustering

Option 2: Prediction

Complexity: Beginner

Task:

User Story: User should be able to provide a Customer ID and Date, and program should be able to predict quantity.

Hint: Data pre-processing, pivoting

Option 3: Recommendation

Complexity: Intermediate

Task: Recommend item to the given customer id for a given date.

User Story: User should be able to provide a Customer ID and Date, and program should be able to recommend item to be purchased.

Hint: Approach would be given importance over result

Option 4: Big Data Analysis using Pyspark

Complexity: Intermediate

Task:

Explore the shared dataset using pyspark and answer the following questions:

1. Who are the winners of the D1 division in the Germany Football Association (Bundesliga) in the

last decade?

2. Which teams have been relegated in the past 10 years?

3. Does Octoberfest affect the performance of Bundesliga?

4. Which season of Bundesliga was the most competitive in the last decade?

5. What's the best month to watch Bundesliga?

Enter the Spark Environment Of google collab. For your help can refer to the following Link

https://medium.com/analytics-vidhya/getting-started-spark3-0-0-with-google-colab-9796d350d78

Take a screenshot of the Answer and also create a small video explaining your code.

Data is provided check the data folder.

Option 5: Text Analytics using NLP

Complexity: Beginner

Task: Reading PDF content and summarization

User Story: User should be able to upload pdf. Application/solution should be able to read the content of pdf and summarize the content into 500 words. Please use "Operations Management.pdf" as input pdf.

Option 6: Q&A using LLMs

Complexity: Intermediate

Task:

User Story: User should be able to upload a pdf (having textual data) and ask questions, application/solution should provide with answers from the context of uploaded pdf. You are free to use enterprise LLM APIs.

Submission: Send a screen recorded video(zipped) of the user story or upload into your google drive and share link with view and download access (please ensure to rename the video to your full name before sending it).

Once approved we will ask for the code.

Important points to consider before submission:

1. You are free to choose open-source models and codes. Please ensure that there is no complete copy paste done.

2. For **ease of processing**, can **filter and use data for any one country**.

3. You are free to add external features relevant to the dataset.

4. Please share **only screen recorded videos** and not the code files.

5. Videos **without applicant's name in filename would not be considered for evaluation**.

Good Luck !!