Data Engineer position Assignment:

- **Problem**
  In this test, you'll be tasked with designing a machine learning model to predict the similarity between pairs of manufacturing materials based on their descriptions. The dataset contains various material descriptions, including specifications such as dimensions, material type, and standards.

  Your objective is to accurately rank or classify the similarity between material pairs. This challenge is highly relevant to industries where comparing and identifying similar materials for production, procurement, or substitution is crucial.

- **Evaluation**
  The evaluation metric for this competition will be **Mean Average Precision at K (MAP@K)**, a common metric used for ranking problems.

- **Dataset**

  ### Training Data:

  You are provided with a dataset of 1,000 unique material descriptions, each associated with an auto-incrementing ID.

  materials.csv

  - **ID**: Unique identifier for the material
  - **Material_Description**: A detailed description of the material

  ### Sample Training Data:

ID	Material_Description
1	INSULATION GASKET KIT - 8" - 300# - G-10 RETAINER W/PTFE…
2	BARREL NIPPLE - 1.5" - SCH.80 - SMLS - ASTM A106 GR.B, HOT…
…	…

  ### Test Data:

  You will be provided with a test file containing pairs of material IDs. Your task is to predict the similarity between the pairs.

  test_pairs.csv

  - **ID_1**: ID of the first material in the pair
  - **ID_2**: ID of the second material in the pair

Sample Test Data:

ID_1	ID_2
12	45
3	89

- ## Submission Format:

  Your submission will be a file that contains a similarity score for each material pair. The score should be a floating-point number between 0 and 1, where:

  - **1** indicates the materials are very similar.
  - **0** indicates the materials are not similar.

  submission.csv

  - **ID_1**: ID of the first material in the pair
  - **ID_2**: ID of the second material in the pair
  - **Similarity_Score**: Predicted similarity score between 0 and 1

  ### Sample Submission:

ID_1	ID_2	Similarity_Score
12	45	0.78
3	89	0.25

Note: Please find below csv files attached with the email.

materials.csv test_pairs.csv submission.csv