Tasks:

1. **ML:** R-Programing: Fuzzy Text Match
2. Read excel file. (DocumentData.xlsx)
3. Create test/training sample set
4. Predict Best Match of Test data docs from Training data docs
5. Output Matched docs, Count of Matched docs and Confidence score.

ML: P-Prog: Model Fitting

1. Read csv file(TestParam.csv)
2. Create test/training sample set
3. Train model on training data
4. Predict label for test data ( Label is outcome)
5. Show accuracy of model.
6. **R-Programming:** Read EmployeeData xlsx file(3 sheets)
7. Add all files
8. File has fields: Name | Dept | Salary | …
9. Remove Duplicate Records based on Key (Name, Dept, Salary)
10. Club data by Dept and show output as: Dept | Salary\_Range | Num\_Of\_Employees.
11. **Text based clustering (NLP) :** Perform unsupervised topic modelling of unlabeled text descriptions
    1. Perform preprocessing and cleaning text data
    2. Execute feature engineering method to extract text features (use TF-IDF or BOW or any other appropriate method)
    3. Run topic modelling clustering algorithm to label text descriptions with proper category
12. **Deep learning :** Build a neural net model for binary classification (Data :- deep\_learning.csv)
    1. Create train/test set and construct a model (using Keras or any other appropriate library)
    2. Predict labels for new data (test data)
    3. Evaluate your model and show relevant performance metrics (like precision, recall etc.)
    4. Fine tune your model using appropriate optimization parameters and attempt improvement in model performance
    5. And save the model object