

**Université Blida 1**  
**Faculté des sciences**

**TP\_ M1 TAL**

**2021/2022**

1. Lire le fichier 'DirectMail.txt' avec la fonction `pandas.read_table`, sépateur='\t' et afficher le nombre de lignes et de colonnes. Le fichier doit être stocké dans un dataframe nommé 'df'.
2. Supprimer les colonnes 'emp1', 'female' et 'emp2'.
3. Supprimer les lignes où la valeur de crédit dépasse 720 (`CRED>720`)
4. Créer une colonne 'TEST' qui reçoit comme valeur:
  - \* 'A' si `GENDER=M` et `RESTYPE=HOME` et `MILEAGE>30`
  - \* 'F' si `GENDER=M` et `RESTYPE=HOME` et `MILEAGE<30`
  - \* Autrement TEST reçoit la valeur 'R'
5. Supprimer les lignes ayant des personnes ayant un `AGE>62` et un `CRED>=750`
6. Lire le fichier 'emails.csv' avec la fonction `pands.read_csv`, sépateur='t'. Le fichier doit être stocké dans un dataframe nommé 'df2'.
7. Nommer la première colonne 'type' et la deuxième 'email'
8. Calculer le nombre de mots dans la colonne 'email' et mettre le résultat dans une nouvelle colonne nommée 'taille'.
9. Supprimer la ponctuation (.,;...etc.) et les symboles (@,/,#...) de la colonne 'email'
10. Concaténer horizontalement les dataframe df et df2 et supprimer les lignes contenant une valeur nulle.