Faculté : NTIC, **Département :** IFA
Année Universitaire : 2023/2024
Module: TDSAI (Master 1 SDIA)

TP 02: Spark

-sparkSQL-

Exercice 1

Nous considérons le dataset de l'exercice 4 du TP 1.

Utilisez SparkSQL pour résoudre l'exercice selon deux manières :

1. En appliquons les opérations sur les dataframes.
2. En appliquant les requêtes SQL.

Exercice 2

Soit le jeu de données relatif aux commandes des clients extrait à partir du site Web d'Amazon (commandes_client.csv). Le fichier contient trois colonnes, la première colonne représente l'ID du client, la seconde représente l'ID de l'article commandé et la dernière représente le montant total dépensé pour cet article.

Écrire un code SparkSQL qui répond aux questions suivantes :

1. Donnez le nombre de clients.
2. Donner le nombre total des articles commandés.
3. Donner le nombre de chaque article commandé.
4. Donnez le montant dépensé par chaque client.
5. Donnez le montant dépensé le plus élevé pour chaque article.
6. Donnez la moyenne puis le maximum puis le minimum du montant dépensé par client.

NB. Utilisez les opérations sur les dataframes puis les requêtes SQL.