

Projet Spark : Apple

Le rendu du projet sera un script Python nommé : **NOM_PRENOM_CLASSE.py**

Vous allez devoir travailler sur un Dataset de données sur les prix de ventes des produits Apple dans différents pays.

Les données ont été récupérées à l'aide d'outil de scrapping sur les sites d'Apple des différents pays.

Nous partirons du principe que les prix des produits vendus aux United States est le vrai prix des produits.

Le but du projet est de réaliser un script Python à rendre qui devra répondre aux différentes demandes.

Le projet est composé de 2 fichiers :

- ApplePrice.csv
- CurrencyConversion.csv

Schéma de ApplePrice.csv :

```
root
|-- Model_name: string (nullable = true)
|-- Price: double (nullable = true)
|-- Country: string (nullable = true)
|-- Currency: string (nullable = true)
```

Le prix des produits est dans la monnaie du pays.

Chaque pays ne possède pas forcément tous les produits Apple.

Schéma de CurrencyConversion.csv :

```
root
|-- ISO_4217: string (nullable = true)
|-- Dollar_To_Curr_Ratio: double (nullable = true)
```

Le ratio est établi de la sorte de 1 \$ = xx Currency

Un des objectifs de ce projet est de **travailler sur des prix en dollar \$**.

Travail demandé :

- Réaliser un script Python qui prend en argument les chemins vers les deux fichiers csv.
Exemple :
`spark-submit --master local nomDeMonScript.py pathVersCsv1.csv pathVersCsv2.csv`
- Convertir les prix des différents produits en dollar \$ (USD)
(pas besoin d'enregistrer, toutes les questions suivantes devront se baser sur des prix en dollar)
- Etablir la moyenne des prix des produits de chaque pays et les comparer à l'équivalent du prix classique US. Ajouter une colonne du pourcentage d'écart entre la moyenne des produits aux US et la moyenne des produits vendus dans le pays.
Trier le DataFrame par ordre décroissant.
Enregistrer ce DataFrame au format .csv
Le nommer : `moyennePrix.csv`
- Réaliser un DataFrame comprenant le pays et le coût total afin d'acheter tout les produits Apple disponible au sein du pays.
Enregistrer ce DataFrame au format .csv
Le nommer : `coutTotal.csv`
- Réaliser un DataFrame avec le nom des produits disponible.
Enregistrer ce DataFrame au format .csv
Le nommer : `listeProduit.csv`
- Réaliser un DataFrame afin de trouver le pays proposant les Airpods Pro au prix le plus abordable.
Enregistrer ce DataFrame au format .csv
Le nommer : `airpodsPro.csv`
- **Bonus** : Chaque fichier enregistré doit n'enregistrer **qu'un seul** fichier .csv

Il ne doit y avoir qu'un seul script pour toutes les tâches demandées.

Les commentaires sont appréciés !

Vous serez notés aussi sur la qualité du code proposé.