

Porjet2 : A réaliser Sous DataBricks (5 points)

La Direction des ventes de LatDior Data cherche un Data Analyste afin d'exploiter leurs data(s)

Description de la donnée (https://github.com/idiattara/sda_2024/blob/main/TP_Databricks.json)

```
{
  "Transaction": [
    {"IdTransaction": 1, "Prix": 20, "TypeProduit": "Laitier", "Devis": "EUR", "Pays": "France"},
    {"IdTransaction": 2, "Prix": 5000, "TypeProduit": "Boisson", "Devis": "TND", "Pays": "Tunisie"},
    {"IdTransaction": 3, "Prix": 200, "TypeProduit": "Electronique", "Devis": "USD", "Pays": "Senegal"},
    {"IdTransaction": 4, "Prix": 200, "TypeProduit": "Laitier", "Devis": "EUR", "Pays": "France"},
    {"IdTransaction": 5, "Prix": 205, "TypeProduit": "Boisson", "Devis": "EUR", "Pays": "France"},
    {"IdTransaction": 6, "Prix": 500, "TypeProduit": "Laitier", "Devis": "TND", "Pays": "Tunisie"},
    {"IdTransaction": 7, "Prix": 145000, "TypeProduit": "Electronique", "Devis": "TND", "Pays": "Tunisie"},
    {"IdTransaction": 8, "Prix": 200, "TypeProduit": "Electronique", "Devis": "CFA", "Pays": "Senegal"}
  ],
  "Devis": [
    {"Taux": 1.0, "Devis": "EUR"},
    {"Taux": 0.88, "Devis": "USD"},
    {"Taux": 0.0015, "Devis": "CFA"},
    {"Taux": 0.31, "Devis": "TND"}
  ]
}
```

Lib à importer

```
import org.apache.spark.sql.functions._
import org.apache.spark.sql.types._
import org.apache.spark.storage.StorageLevel
import spark.sqlContext.implicits
import org.apache.spark.sql.DataFrame
import org.apache.spark.sql.expression
```

Partie1 : Load Data

Question1 :

```
def readerjson(path: String):Seq[DataFrame]
```

Cette fonction lit un fichier **JSON** et retourne deux DataFrames, un qui contient les informations des transactions et un autre celles des devis

```
val dftransaction=readerjson(path)(0)
dftransaction.show
```

▶ (2) Spark Jobs

▶  dftransaction: org.apache.spark.sql.DataFrame = [Devis: string, IdTransaction: long ... 3 more fields]

Devis	IdTransaction	Pays	Prix	TypeProduit
EUR	1	France	20	Laitier
TND	2	Tunisie	5000	Boison
USD	3	Senegal	200	Electronique
EUR	4	France	200	Laitier
EUR	5	France	205	Boison
TND	6	Tunisie	500	Laitier
TND	7	Tunisie	145000	Electronique
CFA	8	Senegal	200	Electronique

```
val dftauxdevis=readerjson(path)(1)
dftauxdevis.show
```

▶ (2) Spark Jobs

▶  dftauxdevis: org.apache.spark.sql.DataFrame = [Devis: string, Taux: double]

Devis	Taux
EUR	1.0
USD	0.88
CFA	0.0015
TND	0.31

Partie 2 Agrégation simple

Question1 :

def NbCommandeParProduitDansPay():DataFrame

Cette fonctionne calcule le **nombre** de commande de chaque produit pour chaque **pays**

```
NbCommandeParProduitDansPay().show
```

► (6) Spark Jobs

```
+-----+-----+-----+-----+
| TypeProduit|France|Senegal|Tunisie|
+-----+-----+-----+-----+
|      Boison|      1|      0|      1|
|      Laitier|      2|      0|      1|
|Electronique|      0|      2|      1|
+-----+-----+-----+-----+
```

Partie3 Cross data

Question 1 :

def `PrixEur()`:DataFrame

La fonction fait la **jointure** du Dataframe **dftransaction** et celui de **dfauxdevis** afin de convertir les prix en euro

```
PrixEur().show
```

► (1) Spark Jobs

Devis	IdTransaction	Pays	Prix	TypeProduit	Taux	PrixEur
EUR	1	France	20	Laitier	1.0	20.0
TND	2	Tunisie	5000	Boison	0.31	1550.0
USD	3	Senegal	200	Electronique	0.88	176.0
EUR	4	France	200	Laitier	1.0	200.0
EUR	5	France	205	Boison	1.0	205.0
TND	6	Tunisie	500	Laitier	0.31	155.0
TND	7	Tunisie	145000	Electronique	0.31	44950.0
CFA	8	Senegal	200	Electronique	0.0015	0.3

Partie4 agrégation avec windows partition

Question1 :

def Top2TransactionPerPays():DataFrame

Donne les informations des deux premières transactions qui ont génère plus d'argent pour chaque Pays

```
Top2TransactionPerPays().show
```

► (2) Spark Jobs

Devis	IdTransaction	Pays	Prix	TypeProduit	Taux	PrixEur	rank
EUR	5	France	205	Boison	1.0	205.0	1
EUR	4	France	200	Laitier	1.0	200.0	2
USD	3	Senegal	200	Electronique	0.88	176.0	1
CFA	8	Senegal	200	Electronique	0.0015	0.3	2
TND	7	Tunisie	145000	Electronique	0.31	44950.0	1
TND	2	Tunisie	5000	Boison	0.31	1550.0	2

Partie 5 Agrégation Combinée

Question 1 :

def `Cube()`:DataFrame

Fait une agrégation combinée sur TypeProduit et Pays en utilisant le fonction `cube`

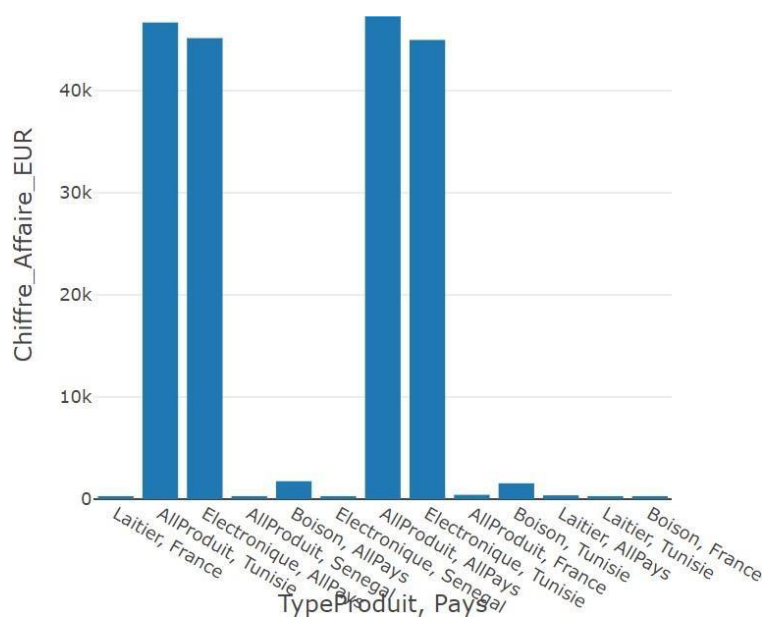
```
Cube().show
```

► (2) Spark Jobs

```
+-----+-----+-----+
| Pays| TypeProduit|Chiffre_Affaire_EUR|
+-----+-----+-----+
| France| Laitier| 220.0|
| Tunisie| AllProduit| 46655.0|
| AllPays| Electronique| 45126.3|
| Senegal| AllProduit| 176.3|
| AllPays| Boison| 1755.0|
| Senegal| Electronique| 176.3|
| AllPays| AllProduit| 47256.3|
| Tunisie| Electronique| 44950.0|
| France| AllProduit| 425.0|
| Tunisie| Boison| 1550.0|
| AllPays| Laitier| 375.0|
| Tunisie| Laitier| 155.0|
| France| Boison| 205.0|
+-----+-----+-----+
```

Question 2 :

a) Réaliser le graphe de la question Question1 en utilisant l'Api SQL



b) Tirez une conclusi

