

# Exercise: Shop Revenue Analysis

- The revenues of a souvenir shop that is present in different cities in France is given in text files. Each file contains the monthly income of a shop's branch over a one year period (each line contains a month and the corresponding revenue).
- Each store (shop's branch) is identified by a name as follows:
  - city\_i; i=1, 2, .... Depending on the number of stores in the city
  - If only one store is present in a city X, it is identified by the name of the city
- You need to write a small spark script that allows to display different statistics on the shop's performance:
  - Average monthly income of the shop in France
  - Average monthly income of the shop in each city
  - Total revenue per city per year
  - Total revenue per store per year
  - The store that achieves the best performance in each month

# Exercise: Shop Revenue Analysis

- Deliverable:
  - One PDF file including the pyspark script (using RDD transformations and actions)
  - You shall include necessary screen shots showing the results of the execution.
  - The scripts shall be readable and well indented (and possibly commented), Ex:

```
df2 = df1.transformation1()
df3 = df2.transformation2()
df3.action1()
```

Or

```
df1.transformation1() /
    .transformation2() /
    .action1()
```
- The deadline for submitting the deliverable is **Thursday 11 October 2018 midnight**.