**Project DBMS : Information About A sample Of a Population**

**Student: Jasser Saad**

**BA/IT**

**1/Introduction:**

This project describes how to treat a sample from a population in France which is consisting of two datasets; the first one is formed by personal information of the individuals of that sample like Job, AdresseId, Income … Whereas the second is formed by some Adresses and their Ids.

The total work is based on relating these two datasets to construct a coherent outpout 1, then using this latter to give a clear representation about the most important attributes in the dataset with graphs 2.

The main objective of this work is to interpret the results given to construct conclusions about the sample, then about the whole population.

**PS**: We will use TALEND OPEN STUDIO for 1 And POWER BI for 2.

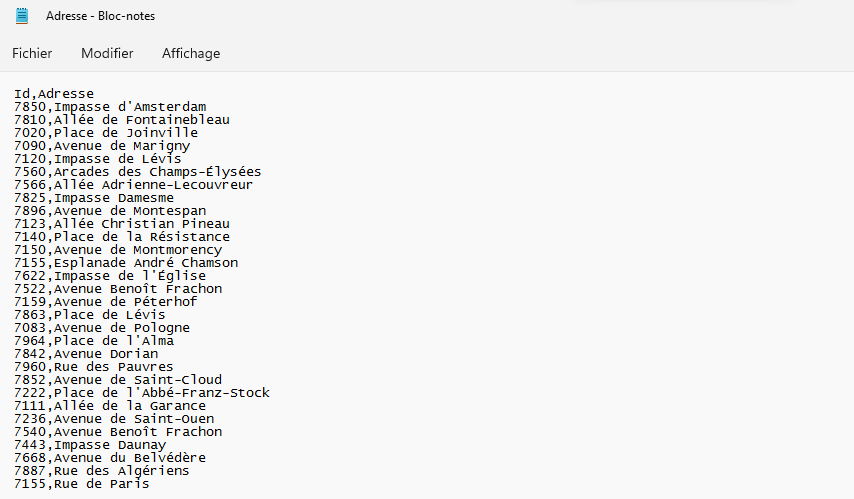
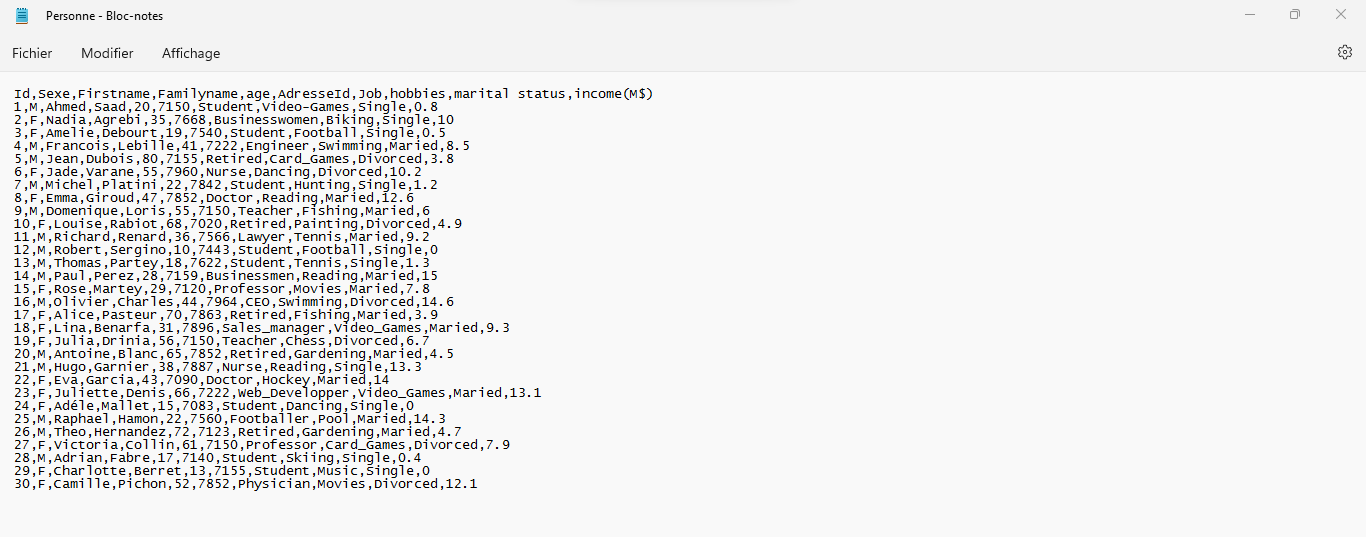
**2/Explanation of the main implemented phases in the database:**

**ETL PROCESS:**

*Phase 1: Data Extraction:*

The databases to work with looks like two samples from two population: one named **Personne** for information about people, and the other named **Adresse** for addresses of living in France and its Ids (two files of type csv).

They are extracted from an open source in Internet for people living in France .



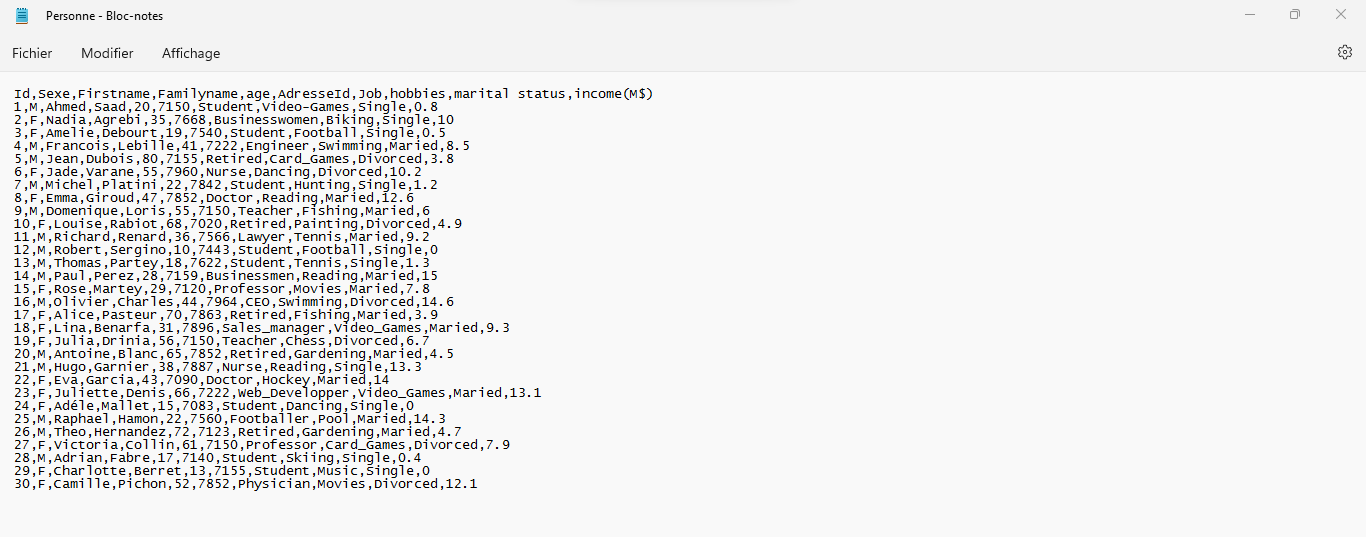
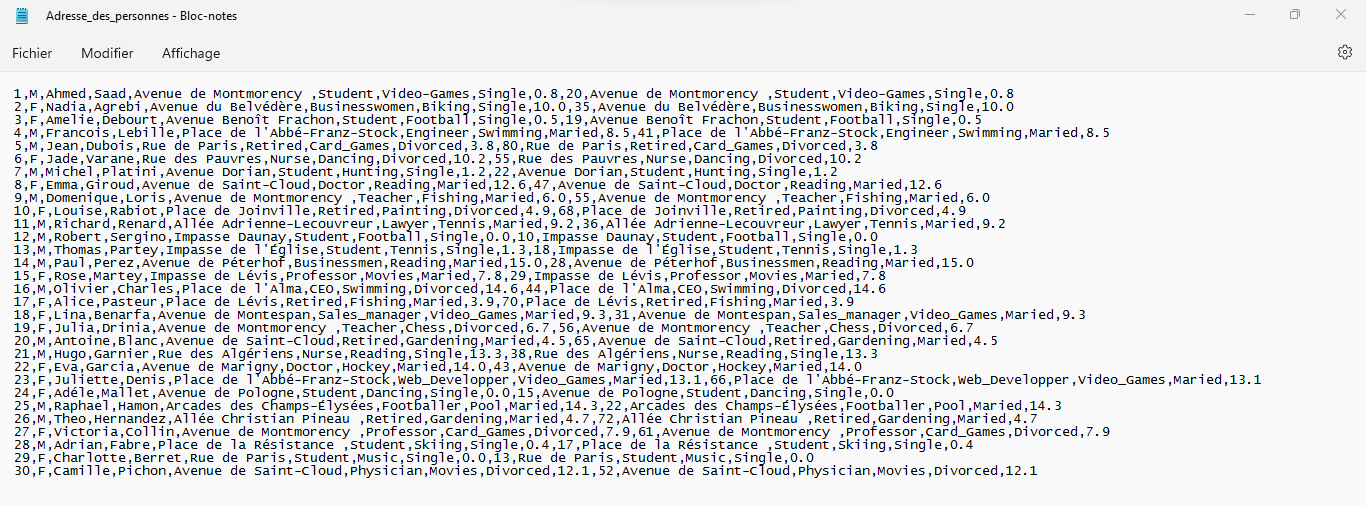
Database NUM :2

Database NUM :1

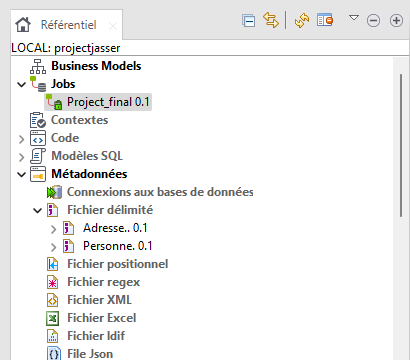
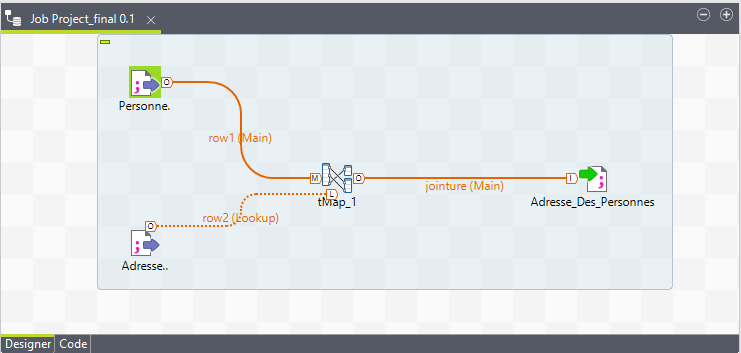
*Phase 2: Data Transformation:*

This step is consisting of make some transformations or changes on the latter databases using an ETL tool: TALEND OPEN SOURCE:

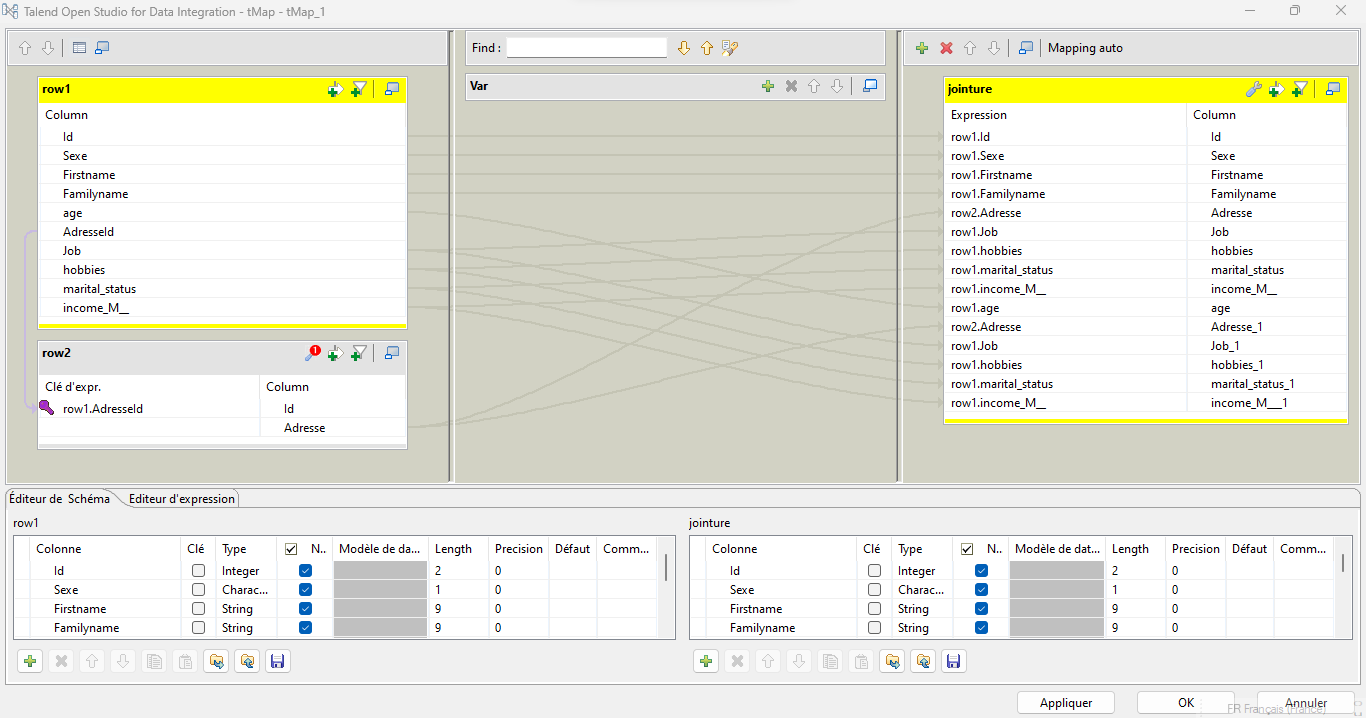
We will use this tool in order to transform the AdresseId attribute in the first database by the corresponding AdresseName from the second database resulting in a new Output database named **Adresse\_des\_personnes**.

The previous work is done using the **tmap component in TALEND** under a job named Project\_final 0.1; Relating the two “fichier délimité” Personne and Adresse to construct the “combination” Adresse\_des\_personnes shown in this file using **tfileouptutdelimiter component**.

The configuration of the **tmap component** is the following:



*Phase 3: Data analysis:*

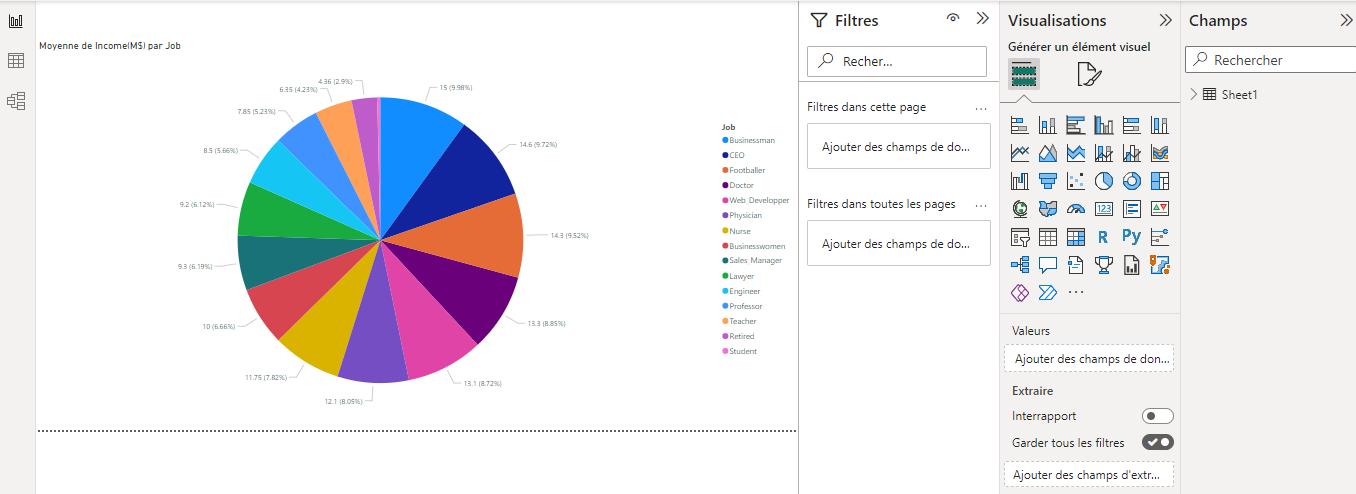
This step will be based on the tool **POWER BI**;

It will used to form a graphical representation of some attributes of the output file in function of some other attributes from the same file (We will use file source .txt this time: Excel file): First **Load** the database in the **power bi** tool :



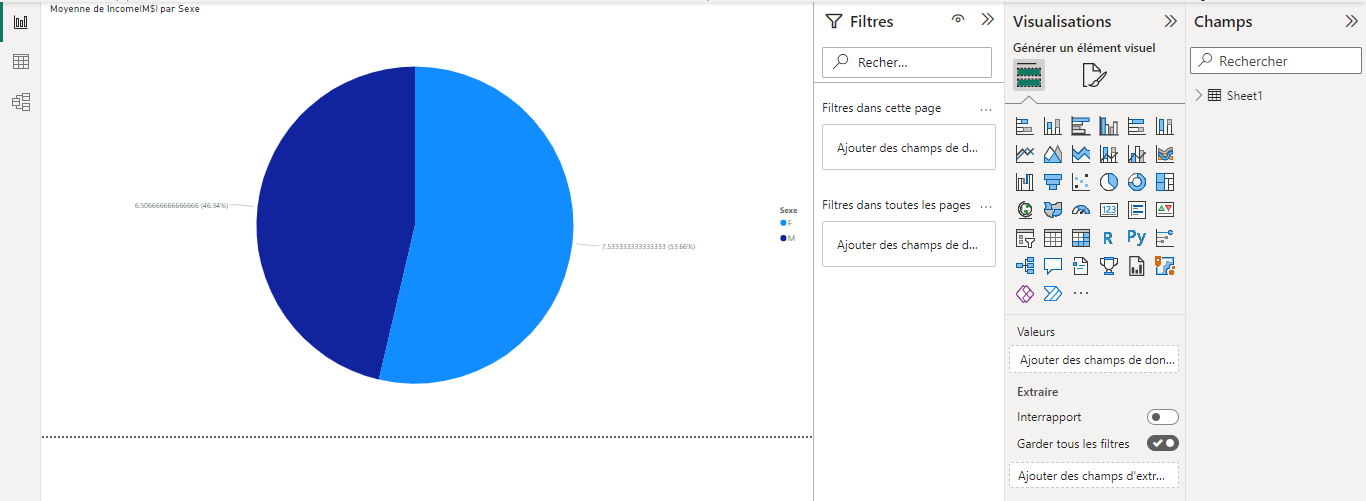
We will design two **Pie charts:**

**\*/Average Income in function of Job:**



**Interpretation:** the job generating the best income in this simple is businessman with 15 million dollars, followed by CEO with 14.6 million Dollars, Whereas Teachers have the smallest income by 6.53 million dollars.

**\*\*/Average Income in function of sexe:**

****

**Interpretation:** In this sample of individuals, the females are payed more than the males by 1 million dollars.

PS: The sample doesn’t reflect the population overall.

**3/Conclusion:**

*Project problem:* We all know that the gap between the income of males and the income of females everywhere. Also, we cannot believe that businessmen earn more than engineers, doctors or nurses …

*Project enhancement:* So that, we can conclude that the sample doesn’t reflect the real facts in the population. Then, Statisticians of this project should revise their work, by changing the sampling method.

**4/Project code:** (for data transformation in Talend):

**/\*\* \* [tFileOutputDelimited\_1 main ] start \*/**

**currentComponent="tFileOutputDelimited\_1";**

**StringBuilder**

**sb\_tFileOutputDelimited\_1 = new StringBuilder();**

**if(jointure.Id != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Id ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1);**

**if(jointure.Sexe != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Sexe );**

**}**

**sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1);**

**if(jointure.Firstname != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Firstname ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.Familyname != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Familyname ); }**

**sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.Adresse != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Adresse ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1);**

**if(jointure.Job != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Job ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.hobbies != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.hobbies ); }**

**sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.marital\_status != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.marital\_status ); }**

**sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.income\_M\_\_ != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.income\_M\_\_ ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.age != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.age ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.Adresse\_1 != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Adresse\_1 ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.Job\_1 != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.Job\_1 ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.hobbies\_1 != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.hobbies\_1 ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.marital\_status\_1 != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.marital\_status\_1 ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_tFileOutputDelimited\_1); if(jointure.income\_M\_\_\_1 != null)**

**{ sb\_tFileOutputDelimited\_1.append( jointure.income\_M\_\_\_1 ); } sb\_tFileOutputDelimited\_1.append(OUT\_DELIM\_ROWSEP\_tFileOutputDelimited\_1);**

**nb\_line\_tFileOutputDelimited\_1++;**

**resourceMap.put("nb\_line\_tFileOutputDelimited\_1", nb\_line\_tFileOutputDelimited\_1); outtFileOutputDelimited\_1.write(sb\_tFileOutputDelimited\_1.toString()); tos\_count\_tFileOutputDelimited\_1++;**

**/\*\* \* [tFileOutputDelimited\_1 main ] stop \*/**