Version finale du projet

Idée de Projet : Analyse de la Rétention des Employés

Objectif : L'objectif de ce projet est d'analyser les facteurs qui influent sur la rétention des employés au sein d'une entreprise afin d'identifier les mesures à prendre pour améliorer la rétention du personnel.

Table de dimension

Employee :

- EmployeeID (Clé primaire)
- Age (Âge de l'employé)
- AgeGroup (Groupe d'âge de l'employé)
- Attrition (Si l'employé a quitté l'entreprise ou non)
- Department (Département dans lequel l'employé travaille)
- Education (Niveau d'éducation de l'employé)
- EducationField (Domaine d'éducation de l'employé)
- Gender (Genre de l'employé)
- MaritalStatus (Statut matrimonial de l'employé)
- Over18 (Si l'employé est majeur)
- OverTime (Si l'employé fait des heures supplémentaires ou non)

Date :

- DateID (Clé primaire)
- Date (Date)
- Year (Année)
- Month (Mois)
- DayOfMonth (Jour du mois)

Table de faits

ExperienceFact :

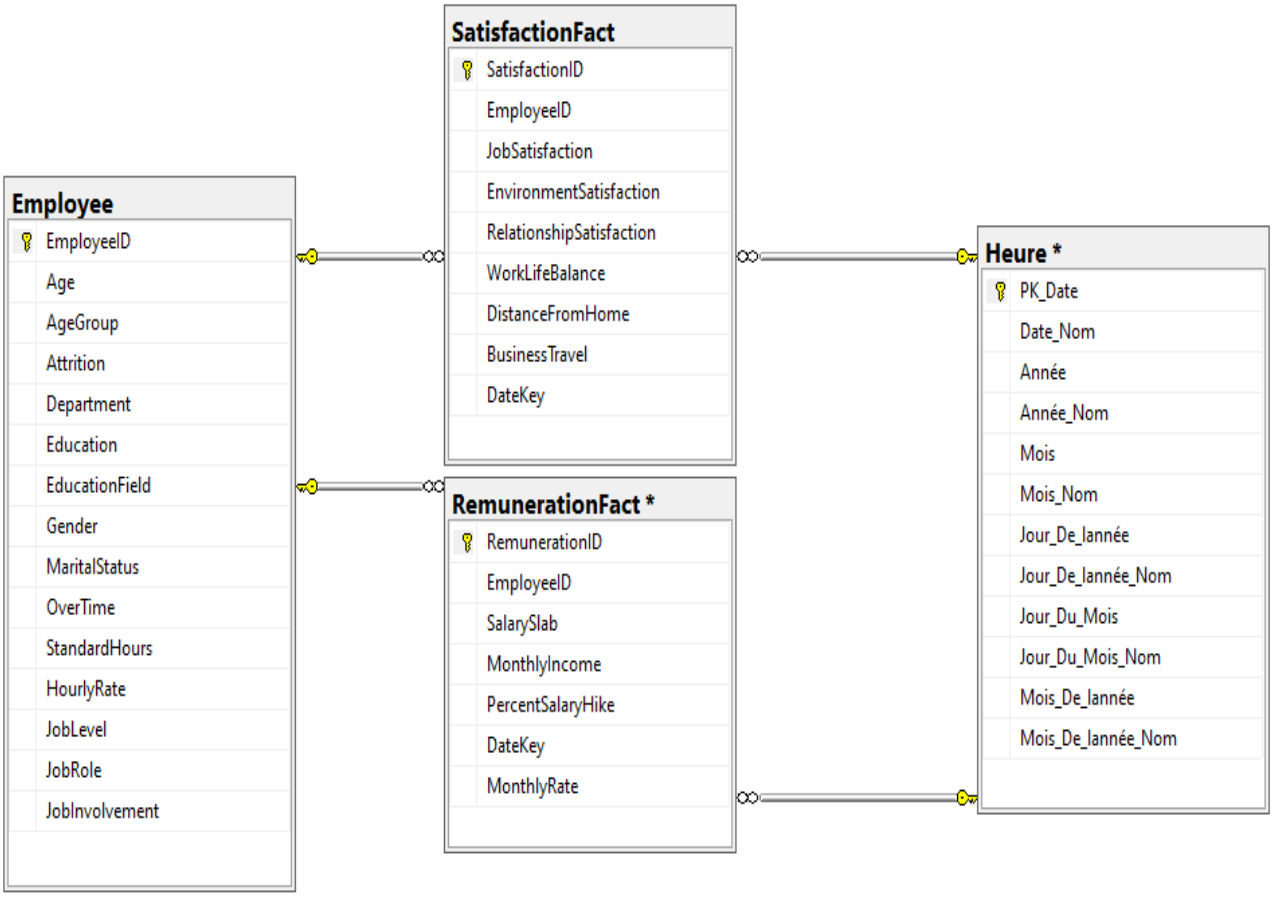
- RemunerationID : Identifiant de la rémunération (clé primaire)
- EmployeeID : Identifiant de l'employé (clé étrangère faisant référence à la table de dimension Employee)
- DateID (Clé étrangère vers la table Date)
- SalarySlab : Tranche de salaire
- MonthlyIncome : Revenu mensuel
- PercentSalaryHike : Pourcentage d'augmentation de salaire
- MonthlyRate : Taux mensuel

SatisfactionFact :

- SatisfactionID (Clé primaire)
- EmployeeID (Clé étrangère vers la table Employee)
- DateID (Clé étrangère vers la table Date)
- JobSatisfaction (Satisfaction du travail de l'employé)
- EnvironmentSatisfaction (Satisfaction de l'environnement de travail de l'employé)
- RelationshipSatisfaction (Satisfaction des relations au travail de l'employé)
- WorkLifeBalance (Équilibre entre vie professionnelle et vie personnelle de l'employé)
- DistanceFromHome (Distance entre le domicile de l'employé et son lieu de travail)
- BusinessTravel (Fréquence des voyages professionnels de l'employé)

Modélisation dimensionnelle :

Concevoir le schéma de la base de données dimensionnelle qui servira de base au cube OLAP.



Création des tables de faits et des dimensions :

```sql
-- Table de dimension Employee
CREATE TABLE Employee (
    EmployeeID VARCHAR(50) PRIMARY KEY,
    Age INT,
    AgeGroup VARCHAR(50),
    Attrition VARCHAR(3),
    Department VARCHAR(50),
    Education INT,
    EducationField VARCHAR(50),
    Gender VARCHAR(10),
    MaritalStatus VARCHAR(20),
    OverTime VARCHAR(3),
    StandardHours INT,
    HourlyRate INT,
    JobLevel INT,
    JobRole VARCHAR(50),
    JobInvolvement INT
);
-- Table de faits RemunerationFact
CREATE TABLE RemunerationFact (
    RemunerationID VARCHAR(50) PRIMARY KEY,
    EmployeeID VARCHAR(50),
    SalarySlab VARCHAR(50),
    MonthlyIncome INT,
    PercentSalaryHike INT,
    MonthlyRate INT,
    FOREIGN KEY (EmployeeID) REFERENCES Employee(EmployeeID)
);
```

```sql
-- Table de dimension SatisfactionDim
CREATE TABLE SatisfactionDim (
    SatisfactionID VARCHAR(50) PRIMARY KEY,
    EmployeeID VARCHAR(50),
    JobSatisfaction INT,
    EnvironmentSatisfaction INT,
    RelationshipSatisfaction INT,
    WorkLifeBalance INT,
    DistanceFromHome INT,
    BusinessTravel VARCHAR(50)
    FOREIGN KEY (EmployeeID) REFERENCES Employee(EmployeeID)
);
```

Messages

```
Commands completed successfully.

Completion time: 2024-05-09T15:19:56.1714089+01:00
```

Alimentation des données :

Charger les données dans les tables de faits et de dimensions à partir des sources de données.

The execution was successful

Success

11 Total 0 Error
11 Success 0 Warning

Details:

Action	Status	Message
Initializing Data Flow Task	Success	
Initializing Connections	Success	
Setting SQL Command	Success	
Setting Source Connection	Success	
Setting Destination Connection	Success	
Validating	Success	
Prepare for Execute	Success	
Pre-execute	Success	
Executing	Success	
Copying to [dbo].[RemunerationFact]	Success	1480 rows transferred
Post-execute	Success	

Filter Stop Report

Close

```
select * from Employee ;
```

100 %

Results Messages

	EmployeeID	Age	AgeGroup	Attrition	Department	Education	EducationField	Gender	MaritalStatus	OverTime	StandardHours	HourlyRate	JobLevel	JobRole
1	EMP1	18	18-25	Yes	Research & Development	3	Life Sciences	Male	Single	No	80	54	1	Laboratory Tec
2	EMP10	19	18-25	No	Research & Development	1	Medical	Female	Single	No	80	79	1	Laboratory Tec
3	EMP100	25	18-25	No	Sales	3	Life Sciences	Male	Married	No	80	41	2	Sales Executiv
4	EMP1000	40	36-45	No	Research & Development	3	Life Sciences	Female	Married	No	80	80	3	Manufacturing
5	EMP1001	40	36-45	No	Research & Development	3	Life Sciences	Female	Single	Yes	80	84	3	Manufacturing
6	EMP1002	40	36-45	Yes	Research & Development	3	Life Sciences	Male	Single	Yes	80	73	1	Laboratory Tec
7	EMP1003	40	36-45	No	Research & Development	2	Life Sciences	Female	Married	Yes	80	92	2	Manufacturing
8	EMP1004	40	36-45	No	Sales	4	Medical	Male	Single	No	80	88	5	Manager
9	EMP1005	40	36-45	No	Research & Development	2	Life Sciences	Female	Married	No	80	99	1	Laboratory Tec
10	EMP1006	40	36-45	No	Research & Development	3	Life Sciences	Female	Divorced	No	80	52	2	Healthcare Re
11	EMP1007	40	36-45	No	Research & Development	3	Life Sciences	Female	Divorced	No	80	47	2	Healthcare Re
12	EMP1008	40	36-45	No	Research & Development	4	Life Sciences	Male	Divorced	No	80	98	4	Manager
13	EMP1009	40	36-45	No	Research & Development	3	Technical D...	Female	Married	No	80	73	2	Laboratory Tec
14	EMP101	25	18-25	No	Research & Development	3	Life Sciences	Female	Married	No	80	62	2	Manufacturing
15	EMP1010	40	36-45	No	Sales	2	Marketing	Female	Married	No	80	92	2	Sales Executiv

Query executed successfully. | DESKTOP-C5QF0NP (16.0 RTM) | DESKTOP-C5QF0NP\Amira ... | DataWehouseHR | 00:00:00 | 1,480 rows

```sql
selECT * FROM Heure ;
```

100 %

Results | Messages

	PK_Date	Date_Nom	Année	Année_Nom	Mois	Mois_Nom	Jour_De_Iannée	Jour_De_Iannée_Nom	Jour_Du
1	2023-01-01 00:00:00.000	Sunday, January 01 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	1	Day 1	1
2	2023-01-02 00:00:00.000	Monday, January 02 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	2	Day 2	2
3	2023-01-03 00:00:00.000	Tuesday, January 03 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	3	Day 3	3
4	2023-01-04 00:00:00.000	Wednesday, January 04 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	4	Day 4	4
5	2023-01-05 00:00:00.000	Thursday, January 05 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	5	Day 5	5
6	2023-01-06 00:00:00.000	Friday, January 06 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	6	Day 6	6
7	2023-01-07 00:00:00.000	Saturday, January 07 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	7	Day 7	7
8	2023-01-08 00:00:00.000	Sunday, January 08 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	8	Day 8	8
9	2023-01-09 00:00:00.000	Monday, January 09 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	9	Day 9	9
10	2023-01-10 00:00:00.000	Tuesday, January 10 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	10	Day 10	10
11	2023-01-11 00:00:00.000	Wednesday, January 11 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	11	Day 11	11
12	2023-01-12 00:00:00.000	Thursday, January 12 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	12	Day 12	12
13	2023-01-13 00:00:00.000	Friday, January 13 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	13	Day 13	13
14	2023-01-14 00:00:00.000	Saturday, January 14 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	14	Day 14	14
15	2023-01-15 00:00:00.000	Sunday, January 15 2023	2023-01-01 00:00:00.000	Calendar 2023	2023-01-01 00:00:00.000	January 2023	15	Day 15	15

Query executed successfully. | DESKTOP-C5QF0NP (16.0 RTM) | DESKTOP-C5QF0NP\Amira ... | DataWehouseHR | 00:00:00 | 365 rows

```sql
select * from SatisfactionFact;
selECT * FROM Heure ;
```

100 %

Results | Messages

	SatisfactionID	EmployeeID	JobSatisfaction	EnvironmentSatisfaction	RelationshipSatisfaction	WorkLifeBalance	DistanceFromHome	BusinessTravel
1	SAT1	EMP1	3	3	3	3	3	Travel_Rarely
2	SAT10	EMP10	2	2	4	3	3	Travel_Rarely
3	SAT100	EMP100	3	1	4	3	28	Travel_Rarely
4	SAT1000	EMP1000	3	1	4	3	15	Travel_Rarely
5	SAT1001	EMP1001	4	3	4	2	16	Travel_Rarely
6	SAT1002	EMP1002	1	1	2	1	7	Travel_Rarely
7	SAT1003	EMP1003	1	2	2	3	8	Travel_Rarely
8	SAT1004	EMP1004	2	2	2	2	7	Travel_Rarely
9	SAT1005	EMP1005	1	3	4	2	2	Travel_Rarely
10	SAT1006	EMP1006	4	3	4	3	1	Travel_Rarely
11	SAT1007	EMP1007	1	2	2	3	12	Travel_Rarely
12	SAT1008	EMP1008	1	1	1	3	1	Travel_Rarely
13	SAT1009	EMP1009	3	4	1	4	11	Travel_Frequently
14	SAT101	EMP101	3	1	4	3	1	Travel_Rarely
15	SAT1010	EMP1010	2	2	3	3	2	Travel_Rarely

Query executed successfully. | DESKTOP-C5QF0NP (16.0 RTM) | DESKTOP-C5QF0NP\Amira ... | Data

```sql
select * from RemunerationFact;
```

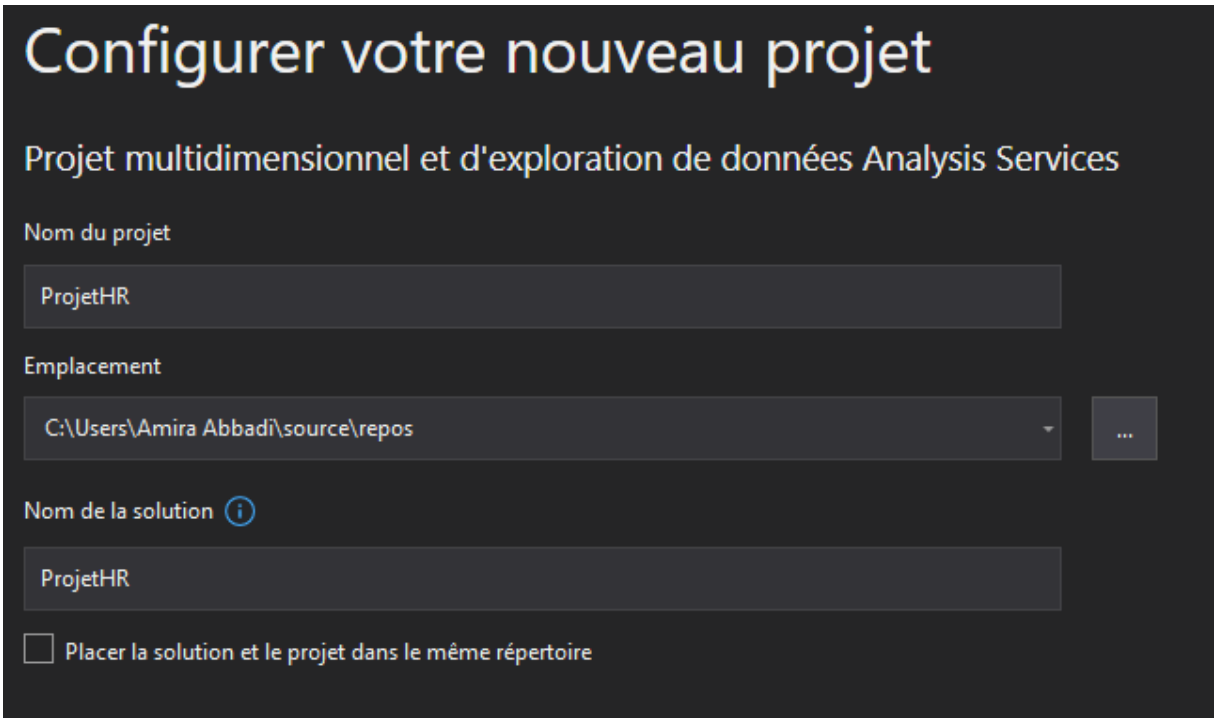
100 %

Results | Messages

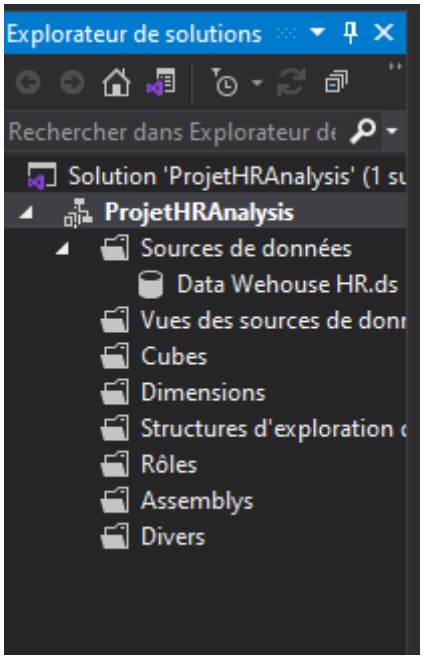
	RemunerationID	EmployeeID	SalarySlab	MonthlyIncome	PercentSalaryHike	MonthlyRate
1	REN1	EMP1	Upto 5k	1420	13	25233
2	REN10	EMP10	Upto 5k	1483	14	16102
3	REN100	EMP100	5k-10k	8639	18	24835
4	REN1000	EMP1000	10k-15k	10435	13	25800
5	REN1001	EMP1001	5k-10k	7945	15	19948
6	REN1002	EMP1002	Upto 5k	2166	14	3339
7	REN1003	EMP1003	5k-10k	6516	16	5041
8	REN1004	EMP1004	15k+	19833	14	4349
9	REN1005	EMP1005	Upto 5k	3377	17	25605
10	REN1006	EMP1006	5k-10k	6513	17	9060
11	REN1007	EMP1007	Upto 5k	4448	12	10748
12	REN1008	EMP1008	15k+	16823	11	18991
13	REN1009	EMP1009	5k-10k	6323	11	26849
14	REN101	EMP101	Upto 5k	4898	12	7505
15	REN1010	EMP1010	5k-10k	5677	14	4258

Query executed successfully. | DESKTOP

Création d'un projet SSAS :

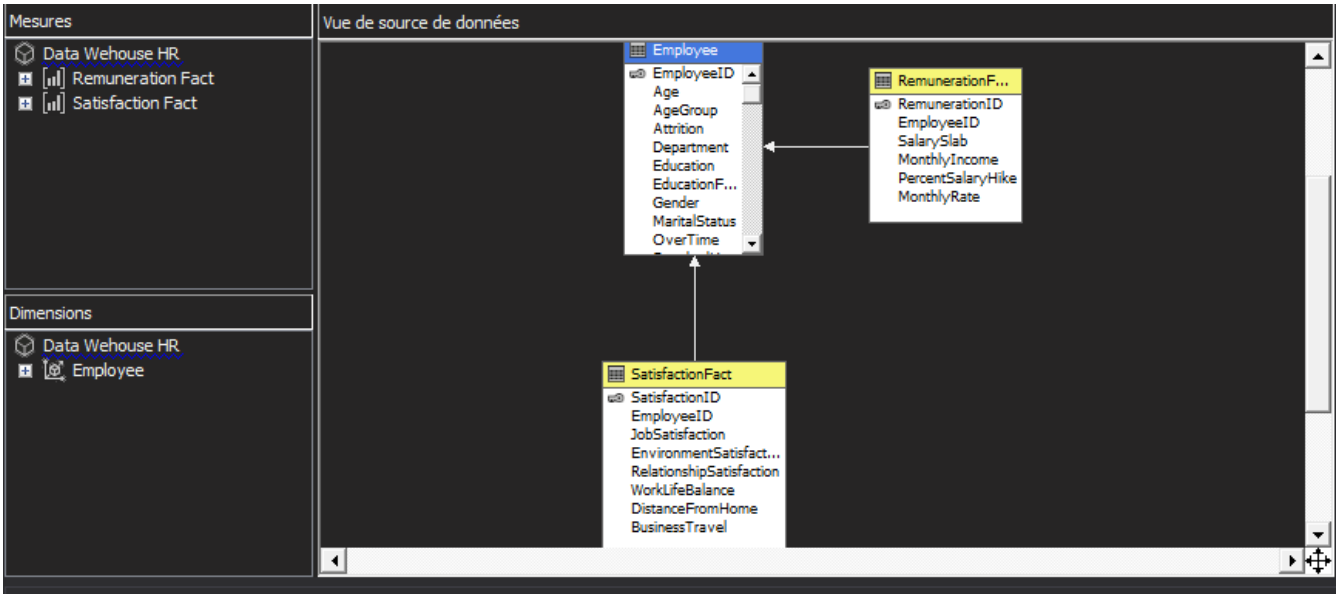


Définition des sources de données :

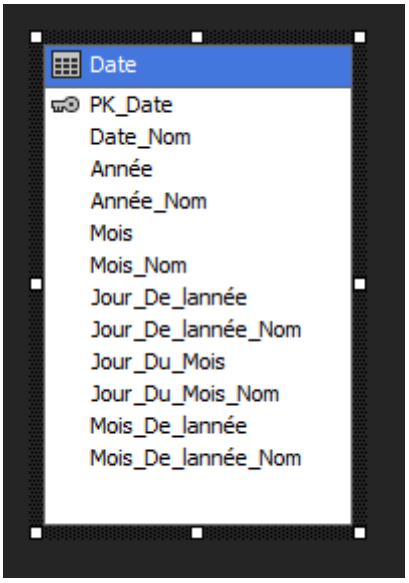


Création de la vue de la source de données :

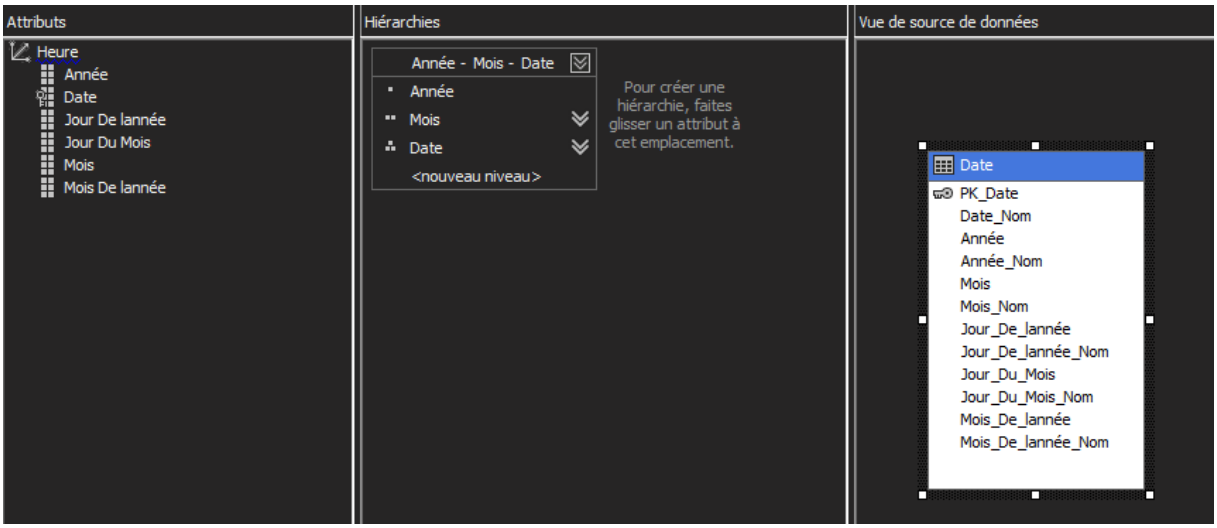
Création de cube :

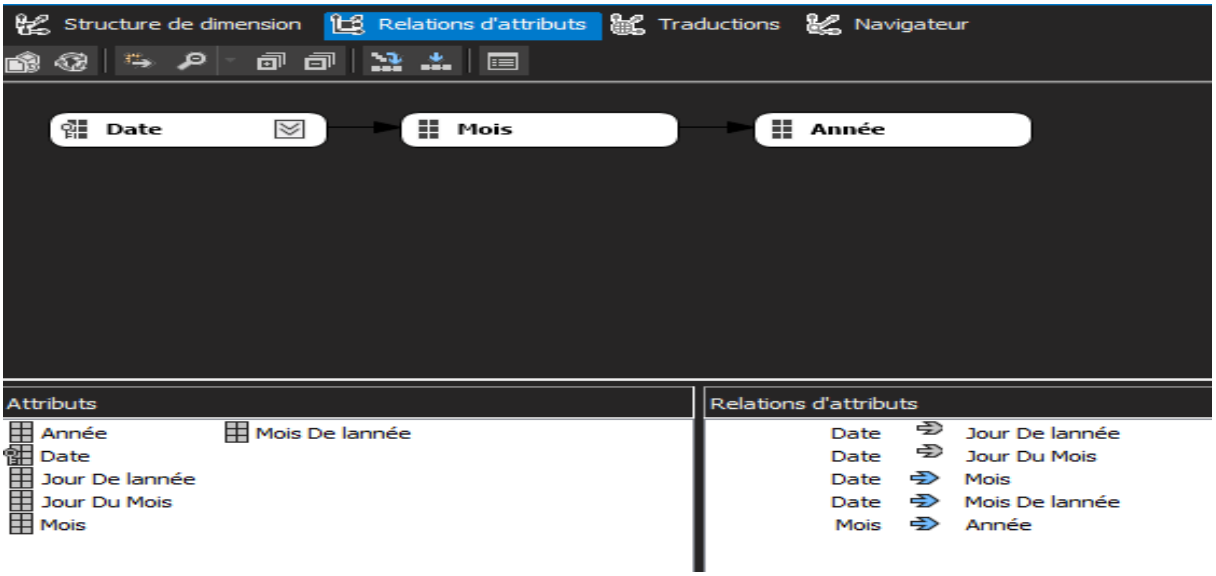


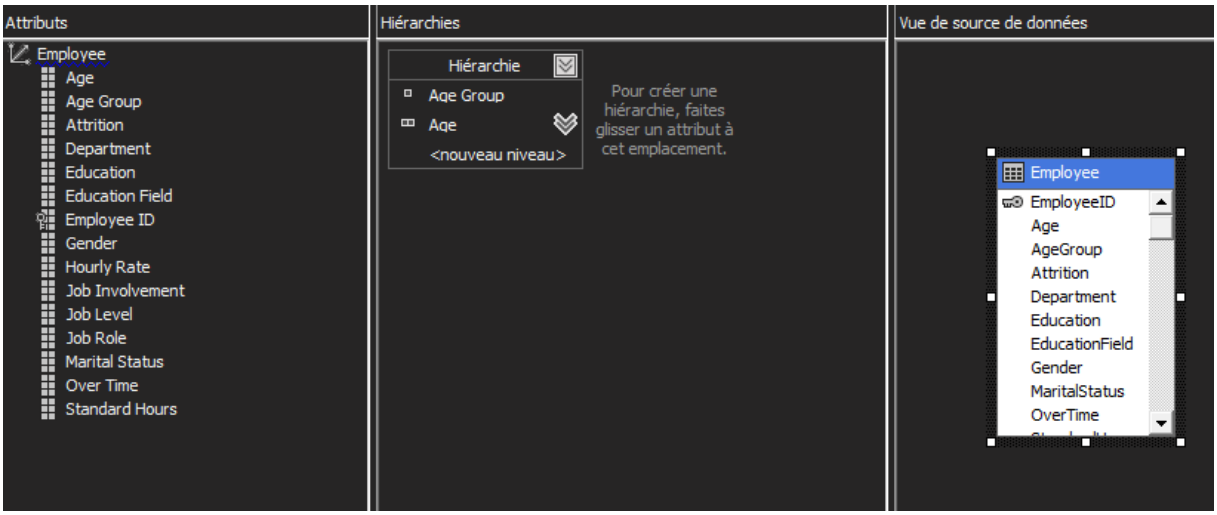
Définition de la dimension de temps :

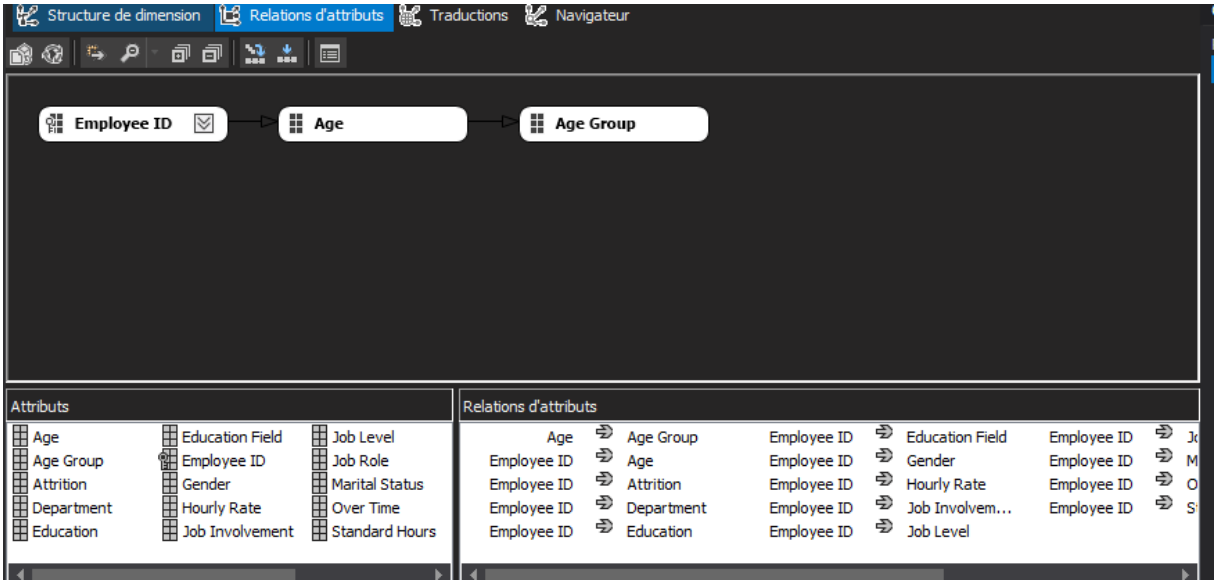


Création des hiérarchies :

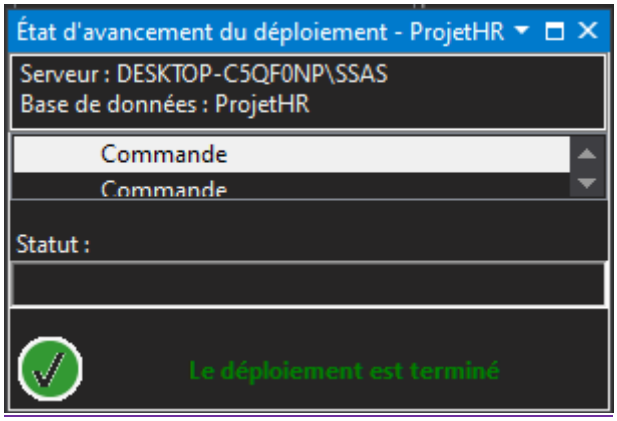


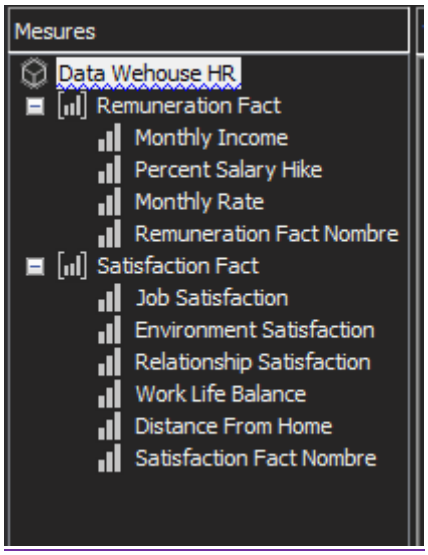


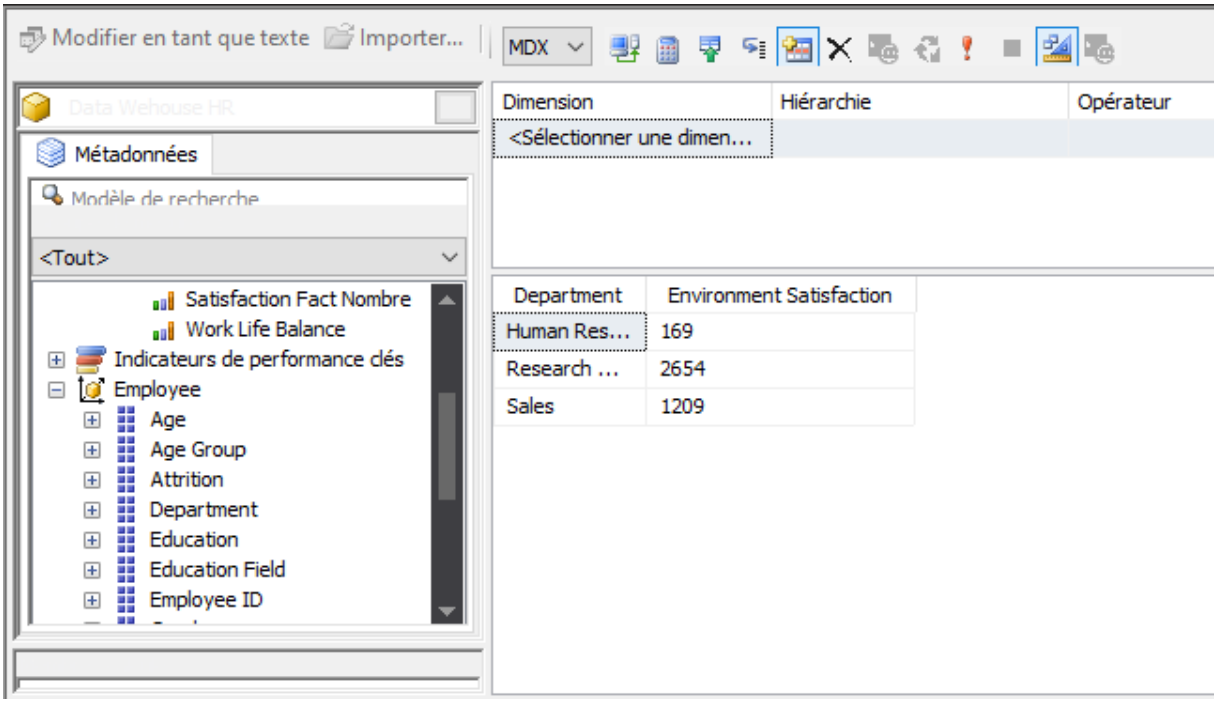


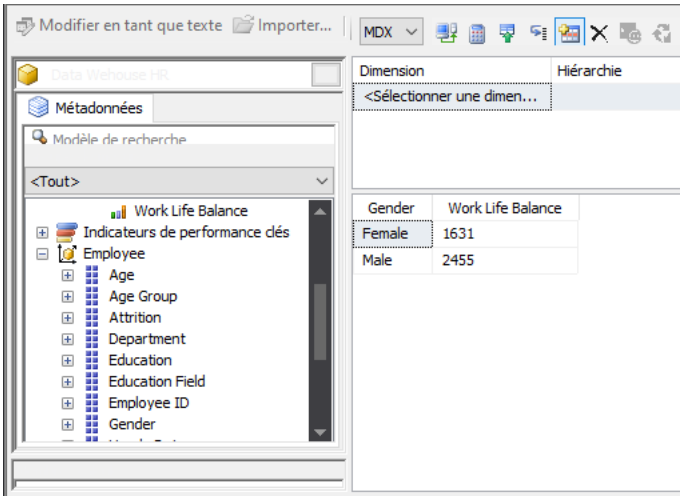


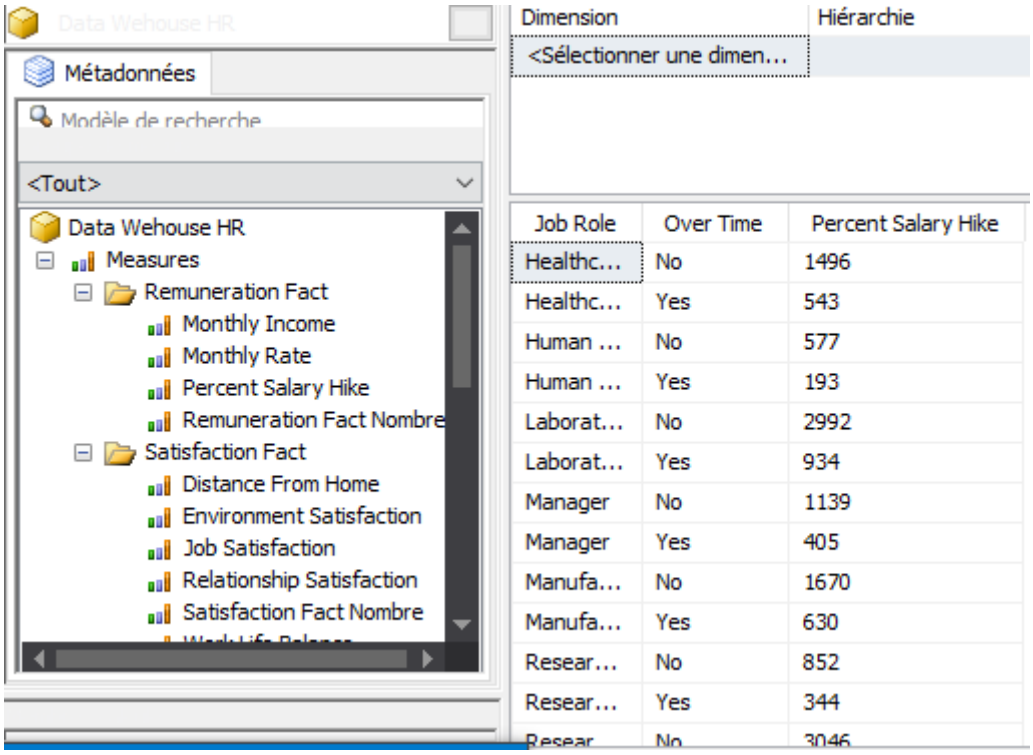
Traitement du cube :

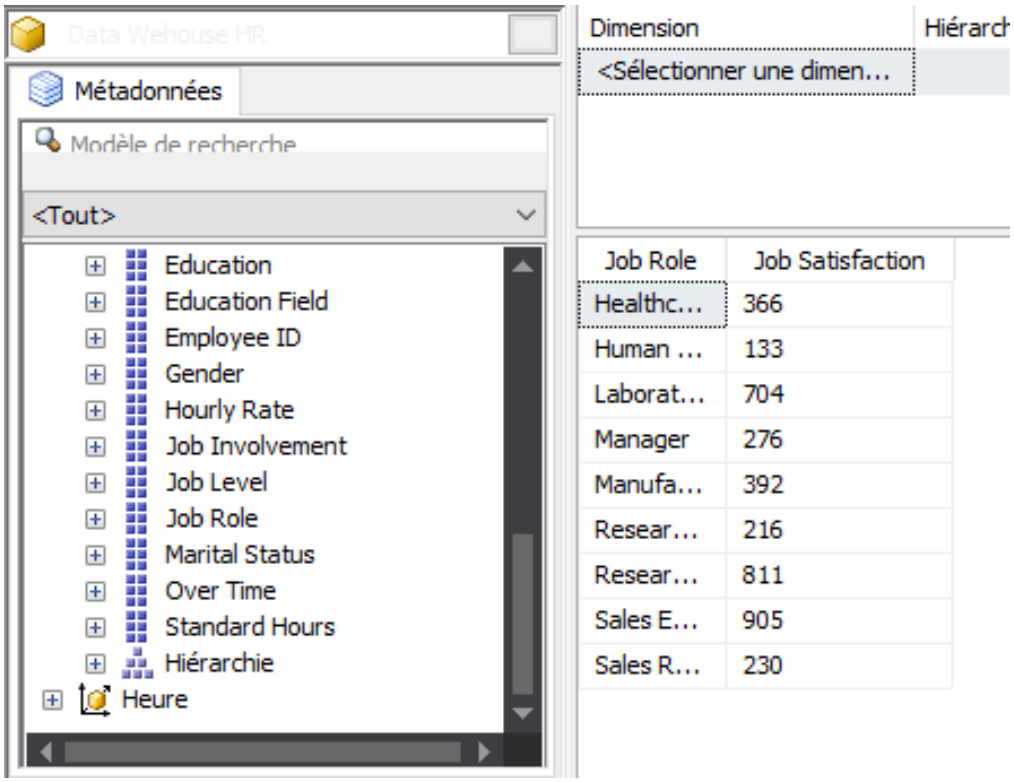


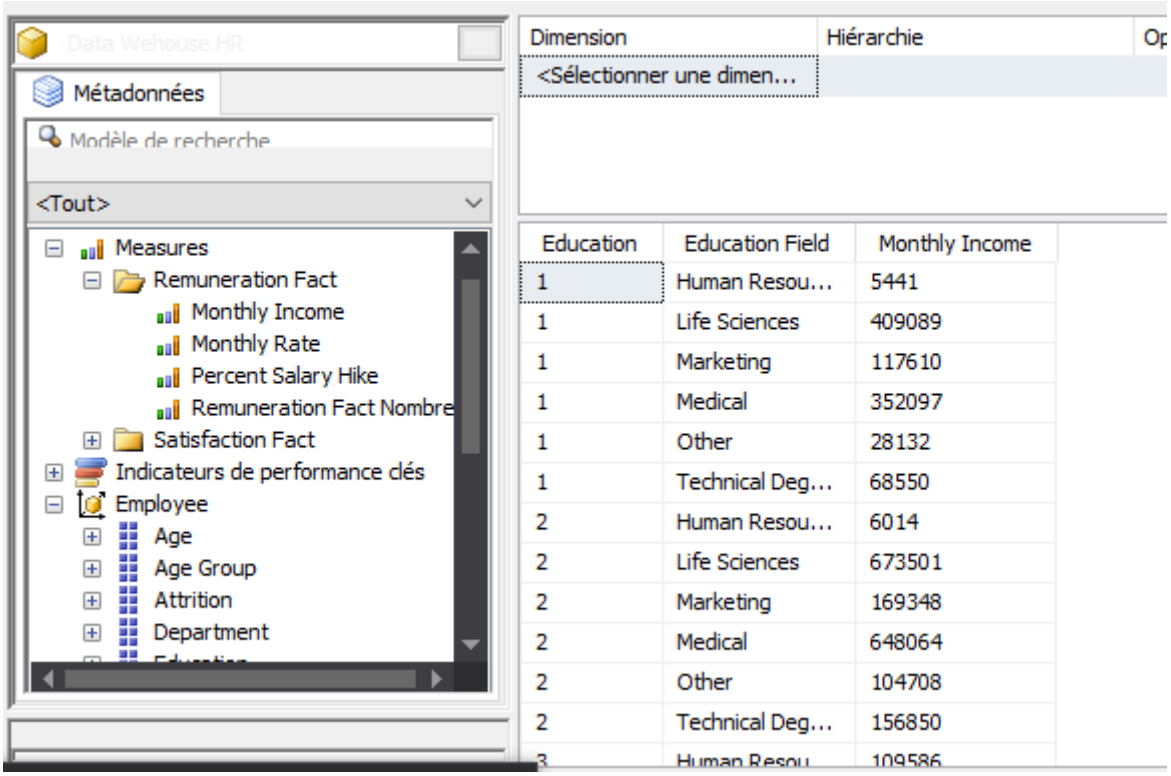

Mesures
Data Wehouse HR
Remuneration Fact
Monthly Income
Percent Salary Hike
Monthly Rate
Remuneration Fact Nombre
Satisfaction Fact
Job Satisfaction
Environment Satisfaction
Relationship Satisfaction
Work Life Balance
Distance From Home
Satisfaction Fact Nombre

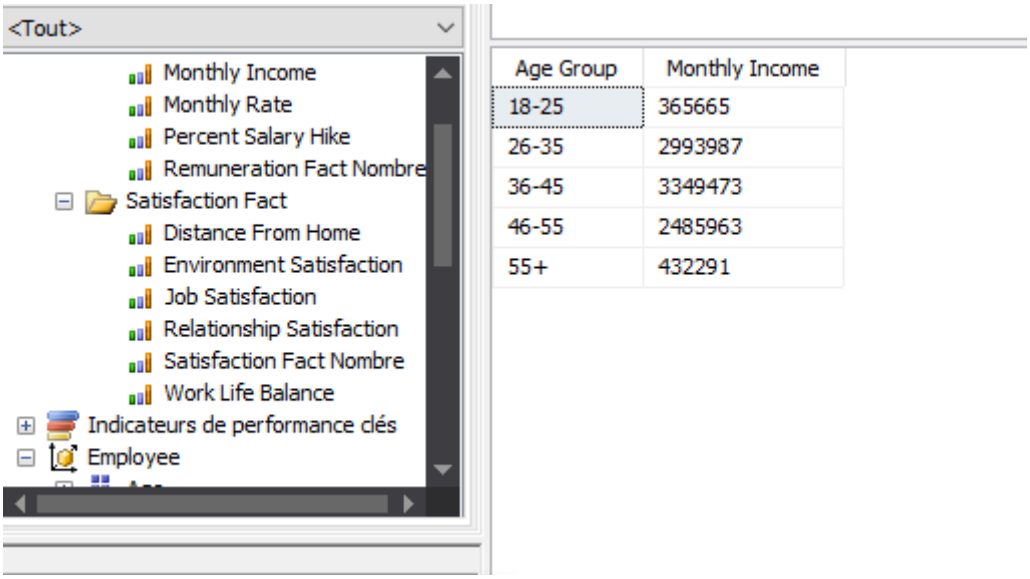

Modifier en tant que texte
Importer...
MDX
Dimension
Hiérarchie
Opérateur
<Sélectionner une dimen...
Métadonnées
Modèle de recherche
<Tout>
Satisfaction Fact Nombre
Work Life Balance
Indicateurs de performance clés
Employee
Age
Age Group
Attrition
Department
Education
Education Field
Employee ID
Department
Environment Satisfaction
Human Res...
169
Research ...
2654
Sales
1209

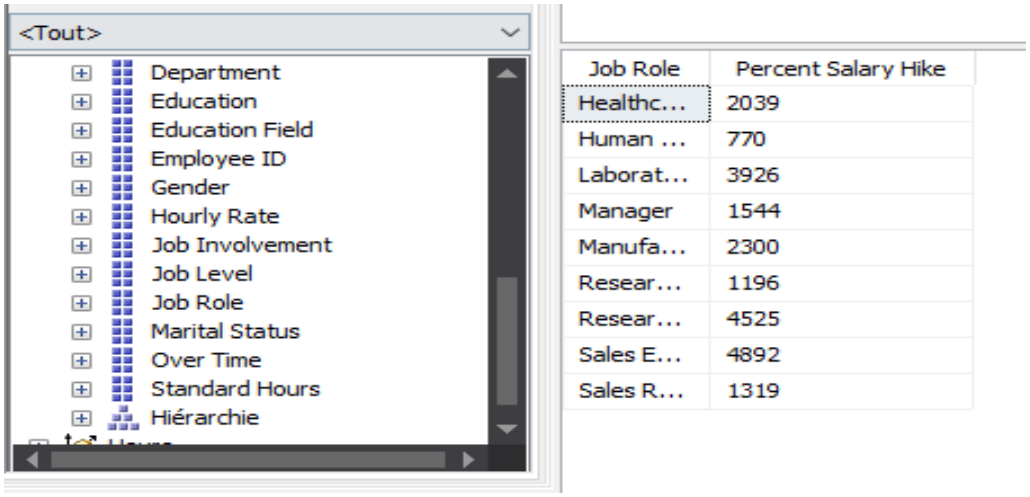

Modifier en tant que texte
Importer...
MDX
Dimension
Hiérarchie
<Sélectionner une dimen...
Métadonnées
Modèle de recherche
<Tout>
Work Life Balance
Indicateurs de performance clés
Employee
Age
Age Group
Attrition
Department
Education
Education Field
Employee ID
Gender
Gender
Work Life Balance
Female
1631
Male
2455


Data Wehouse HR
Métadonnées
Modèle de recherche
<Tout>
Data Wehouse HR
Measures
Remuneration Fact
Monthly Income
Monthly Rate
Percent Salary Hike
Remuneration Fact Nombre
Satisfaction Fact
Distance From Home
Environment Satisfaction
Job Satisfaction
Relationship Satisfaction
Satisfaction Fact Nombre
Dimension
Hiérarchie
<Sélectionner une dimen...
Job Role
Over Time
Percent Salary Hike
Healthc... No 1496
Healthc... Yes 543
Human ... No 577
Human ... Yes 193
Laborat... No 2992
Laborat... Yes 934
Manager No 1139
Manager Yes 405
Manufa... No 1670
Manufa... Yes 630
Resear... No 852
Resear... Yes 344


Data Wehouse HR
Métadonnées
Modèle de recherche
<Tout>
Education
Education Field
Employee ID
Gender
Hourly Rate
Job Involvement
Job Level
Job Role
Marital Status
Over Time
Standard Hours
Hiérarchie
Heure
Dimension
<Sélectionner une dimen...
Job Role
Job Satisfaction
Healthc... 366
Human ... 133
Laborat... 704
Manager 276
Manufa... 392
Resear... 216
Resear... 811
Sales E... 905
Sales R... 230


Data Wehouse HR
Métadonnées
Modèle de recherche
<Tout>
Measures
Remuneration Fact
Monthly Income
Monthly Rate
Percent Salary Hike
Remuneration Fact Nombre
Satisfaction Fact
Indicateurs de performance clés
Employee
Age
Age Group
Attrition
Department
Dimension
Hiérarchie
<Sélectionner une dimen...
Education
Education Field
Monthly Income
1 Human Resou... 5441
1 Life Sciences 409089
1 Marketing 117610
1 Medical 352097
1 Other 28132
1 Technical Deg... 68550
2 Human Resou... 6014
2 Life Sciences 673501
2 Marketing 169348
2 Medical 648064
2 Other 104708
2 Technical Deg... 156850


<Tout>
Monthly Income
Monthly Rate
Percent Salary Hike
Remuneration Fact Nombre
Satisfaction Fact
Distance From Home
Environment Satisfaction
Job Satisfaction
Relationship Satisfaction
Satisfaction Fact Nombre
Work Life Balance
Indicateurs de performance clés
Employee
Age Group
Monthly Income
18-25 365665
26-35 2993987
36-45 3349473
46-55 2485963
55+ 432291


<Tout>
Department
Education
Education Field
Employee ID
Gender
Hourly Rate
Job Involvement
Job Level
Job Role
Marital Status
Over Time
Standard Hours
Hiérarchie
Job Role
Percent Salary Hike
Healthc... 2039
Human ... 770
Laborat... 3926
Manager 1544
Manufa... 2300
Resear... 1196
Resear... 4525
Sales E... 4892
Sales R... 1319

Requêtes MDX :

```
--calculer le salaire mensuel moyen par département :
SELECT
[Measures].[Monthly Income] ON COLUMNS,
[Department].[Department].MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

	Monthly Income
Human Resources	419234
Research & Developm...	6073113
Sales	3135032
Unknown	(null)

```
-- Satisfactiond'environnement par département :
SELECT
[Measures].[Environment Satisfaction] ON COLUMNS,
[Department].[Department].MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

	Environment Satisfaction
Human Resources	169
Research & Developm...	2654
Sales	1209
Unknown	(null)

```
-- Augmentation du salaire par département :
SELECT
[Measures].[Percent Salary Hike] ON COLUMNS,
[Department].[Department].MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

	Percent Salary Hike
Human Resources	930
Research & Developm...	14792
Sales	6789
Unknown	(null)

```sql
-- calculer le salaire mensuel moyen des employés de 55 ans
SELECT
[Measures].[Monthly Income] ON COLUMNS
FROM [Data Wehouse HR]
WHERE [Age].[Age].&[55]
```

100 %

Messages | Results

Monthly Income
290492

```sql
--calculer le salaire mensuel moyen des employés de 30 ans dans le département "Sales"
SELECT
[Measures].[Monthly Income] ON COLUMNS
FROM [Data Wehouse HR]
WHERE ([Age].[Age].&[30], [Department].[Department].&[Sales])
```

100 %

Messages | Results

Monthly Income
115185

```sql
--Pour ajouter une condition sur le niveau d'éducation des employés ( niveau d'éducation 3) :
SELECT
[Measures].[Monthly Rate] ON COLUMNS
FROM [Data Wehouse HR]
WHERE [Education].[Education].&[3]
```

100 %

Messages | Results

Monthly Rate
8125763

```sql
--calculer le salaire mensuel moyen des employés de niveau d'éducation 3 dans le département "Human
SELECT
[Measures].[Monthly Income] ON COLUMNS
FROM [Data Wehouse HR]
WHERE ([Education].[Education].&[3], [Department].[Department].&[Human Resources])
```

100 %

Messages | Results

Monthly Income
182535

```sql
-- calculer le salaire mensuel moyen par genre et par département :
SELECT
[Measures].[Monthly Income] ON COLUMNS,
[Department].[Department].MEMBERS * [Gender].[Gender].MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

		Monthly Income
Human Resources	Female	145280
Human Resources	Male	273954
Research & Developm...	Female	2472474
Research & Developm...	Male	3600639
Sales	Female	1332729
Sales	Male	1802303
Unknown	Unknown	(null)

```sql
-- calculer job satisfaction par genre et par département :
SELECT
[Measures].[Job Satisfaction] ON COLUMNS,
[Department].[Department].MEMBERS * [Gender].[Gender].MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

		Job Satisfaction
Human Resources	Female	55
Human Resources	Male	109
Research & Developm...	Female	999
Research & Developm...	Male	1633
Sales	Female	532
Sales	Male	705
Unknown	Unknown	(null)

```sql
--calculer le salaire mensuel moyen des employés age= 18 dans le département "Sales" :
SELECT
[Measures].[Monthly Income] ON COLUMNS
FROM [Data Wehouse HR]
WHERE ([Employee].[Age].&[18], [Department].[Department].&[sales])
```

100 %

Messages | Results

Monthly Income
4647

```sql
-- calculer work life balance  par genre et par département :
SELECT
[Measures].[Work Life Balance] ON COLUMNS,
[Department].[Department].MEMBERS * [Gender].[Gender].MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

		Work Life Balance
Human Resources	Female	59
Human Resources	Male	125
Research & Developm...	Female	1031
Research & Developm...	Male	1605
Sales	Female	541
Sales	Male	725
Unknown	Unknown	(null)

```sql
-- calculer revenue mesuel  par employer  et par attrition :
SELECT
[Measures].[Monthly Income] ON COLUMNS,
[Employee].[Employee ID]*[Employee].[Attrition] .MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

		Monthly Income
All	All	9627379
All	No	8481984
All	Yes	1145395
All	Unknown	(null)

```sql
-- calculer environment satisfaction par attrition :
SELECT
[Measures].[Environment Satisfaction] ON COLUMNS,
[Employee].[Attrition] .MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

	Environment Satisfaction
All	4032
No	3444
Yes	588
Unknown	(null)

```sql
-- calculer environment satisfaction par attrition et par departement :
SELECT
[Measures].[Environment Satisfaction] ON COLUMNS,
[Employee].[Attrition] .MEMBERS*[Department].[Department].MEMBERS  ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

		Environment Satisfaction
All	Human Resources	169
All	Research & Developm...	2654
All	Sales	1209
All	Unknown	(null)
No	Human Resources	141
No	Research & Developm...	2325
No	Sales	978
Yes	Human Resources	28
Yes	Research & Developm...	329
Yes	Sales	231
Unknown	Unknown	(null)

```sql
-- calculer Relationship Satisfaction par age groupe  et par departement :
SELECT
[Measures].[Relationship Satisfaction] ON COLUMNS,
[Employee].[Age Group].MEMBERS * [Employee].[Department].MEMBERS ON ROWS
FROM [Data Wehouse HR]
```

100 %

Messages | Results

		Relationship Satisfaction
All	All	4009
All	Human Resources	182
All	Research & Developm...	2614
All	Sales	1213
All	Unknown	(null)
18-25	All	332
18-25	Human Resources	13
18-25	Research & Developm...	213
18-25	Sales	106
26-35	All	1649
26-35	Human Resources	70
26-35	Research & Developm...	1062
26-35	Sales	517
36-45	All	1249
36-45	Human Resources	68
36-45	Research & Developm...	846
36-45	Sales	335
46-55	All	641
46-55	Human Resources	22
46-55	Research & Developm...	409

Partie exploration des données avant de faire le projet :

Importation des bibliothèques

+ Code + Texte

```
[8] import numpy as np # est une bibliothèque pour effectuer des calculs numériques
    import pandas as pd # est une bibliothèque pour la manipulation et l'analyse des données.
    import matplotlib.pyplot as plt # utilisée pour la visualisation des données.
    import seaborn as sns # utilisée pour la visualisation des données.
```

Chargement des données

```
HRData = pd.read_csv('/content/HR_Analytics.csv')
# charger un jeu de données de hr à partir d'un fichier CSV en utilisant
#la fonction read_csv de la bibliothèque Pandas.
```

Exploration des données

```
HRData.head() # Afficher par défaut les cinq premières lignes du DataFrame
```

	EmpID	Age	AgeGroup	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField
0	RM297	18	18-25	Yes	Travel_Rarely	230	Research & Development	3	3	Life Sciences
1	RM302	18	18-25	No	Travel_Rarely	812	Sales	10	3	Medical
2	RM458	18	18-25	Yes	Travel_Frequently	1306	Sales	5	3	Marketing
3	RM728	18	18-25	No	Non-Travel	287	Research & Development	5	2	Life Sciences
4	RM829	18	18-25	Yes	Non-Travel	247	Research & Development	8	1	Medical

5 rows × 42 columns

```
HRData.shape # Afficher tuple qui représente les dimensions du DataFrame
```

```
(1480, 42)
```

```python
HRData.info() # Afficher des informations sur le DataFrame
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1480 entries, 0 to 1479
Data columns (total 42 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   EmpID                     1480 non-null   object
 1   Age                       1480 non-null   int64
 2   AgeGroup                  1480 non-null   object
 3   Attrition                 1480 non-null   object
 4   BusinessTravel            1480 non-null   object
 5   DailyRate                 1480 non-null   int64
 6   Department                1480 non-null   object
 7   DistanceFromHome          1480 non-null   int64
 8   Education                 1480 non-null   int64
 9   EducationField            1480 non-null   object
 10  EmployeeCount             1480 non-null   int64
 11  EmployeeNumber            1480 non-null   int64
 12  EnvironmentSatisfaction   1480 non-null   int64
 13  Gender                    1480 non-null   object
 14  HourlyRate                1480 non-null   int64
 15  JobInvolvement            1480 non-null   int64
 16  JobLevel                  1480 non-null   int64
 17  JobRole                   1480 non-null   object
 18  JobSatisfaction           1480 non-null   int64
 19  MaritalStatus             1480 non-null   object
```


```
 19  MaritalStatus             1480 non-null   object
```


```
 20  MonthlyIncome             1480 non-null   int64
 21  SalarySlab                1480 non-null   object
 22  MonthlyRate               1480 non-null   int64
 23  NumCompaniesWorked        1480 non-null   int64
 24  Over18                    1480 non-null   object
 25  OverTime                  1480 non-null   object
 26  PercentSalaryHike         1480 non-null   int64
 27  PerformanceRating         1480 non-null   int64
 28  RelationshipSatisfaction  1480 non-null   int64
 29  StandardHours             1480 non-null   int64
 30  StockOptionLevel          1480 non-null   int64
 31  TotalWorkingYears         1480 non-null   int64
 32  TrainingTimesLastYear     1480 non-null   int64
 33  WorkLifeBalance           1480 non-null   int64
 34  YearsAtCompany            1480 non-null   int64
 35  YearsInCurrentRole        1480 non-null   int64
 36  YearsSinceLastPromotion   1480 non-null   int64
 37  YearsWithCurrManager      1475 non-null   float64
 38  IDEMP                     1480 non-null   object
 39  IDEXP                     1480 non-null   object
 40  IDREN                     1480 non-null   object
 41  IDSAT                     1480 non-null   object
dtypes: float64(1), int64(25), object(16)
memory usage: 485.8+ KB
```

```
HRData.isnull().sum() # Afficher le nombre de valeurs nulles pour chaque colonne de votre DataFrame
```

```
EmpID                     0
Age                       0
AgeGroup                  0
Attrition                 0
BusinessTravel            0
DailyRate                 0
Department                0
DistanceFromHome          0
Education                 0
EducationField            0
EmployeeCount             0
EmployeeNumber            0
EnvironmentSatisfaction   0
Gender                    0
HourlyRate                0
JobInvolvement            0
JobLevel                  0
JobRole                   0
JobSatisfaction           0
MaritalStatus             0
MonthlyIncome             0
SalarySlab                0
MonthlyRate               0
NumCompaniesWorked        0
```

```
Over18                    0
OverTime                  0
PercentSalaryHike         0
PerformanceRating         0
RelationshipSatisfaction  0
StandardHours             0
StockOptionLevel          0
TotalWorkingYears         0
TrainingTimesLastYear     0
WorkLifeBalance           0
YearsAtCompany            0
YearsInCurrentRole        0
YearsSinceLastPromotion   0
YearsWithCurrManager      5
IDEMP                     0
IDEXP                     0
IDREN                     0
IDSAT                     0
dtype: int64
```

Les attributs :

- EmpID (clé primaire) : Identifiant de l'employé. Valeur unique pour chaque employé.

- Age : Âge de l'employé. Valeurs possibles sont des nombres entiers.
- AgeGroup : Groupe d'âge de l'employé. Valeurs possibles sont des plages d'âge comme "26-35", "36-45", etc.
- Attrition : Attrition de l'employé. Valeurs possibles sont "Yes" (oui) ou "No" (non).
- BusinessTravel : Fréquence des voyages d'affaires de l'employé. Valeurs possibles sont "Travel_Rarely", "Travel_Frequently", "Non-Travel".
- DailyRate : Taux journalier de l'employé. Valeurs possibles sont des nombres entiers.
- Department : Département de l'employé. Valeurs possibles sont "Sales", "Research & Development", "Human Resources".
- DistanceFromHome : Distance entre le domicile et le lieu de travail de l'employé. Valeurs possibles sont des nombres entiers.
- Education : Niveau d'éducation de l'employé. Valeurs possibles sont des nombres entiers.
- EducationField : Domaine d'éducation de l'employé. Valeurs possibles sont "Life Sciences", "Medical", "Marketing", "Technical Degree", etc.
- EmployeeCount : Nombre d'employés. Valeurs possibles sont des nombres entiers.
- EmployeeNumber : Numéro de l'employé. Valeur unique pour chaque employé.
- EnvironmentSatisfaction : Satisfaction de l'environnement de l'employé. Valeurs possibles sont des nombres entiers.
- Gender : Sexe de l'employé. Valeurs possibles sont "Male", "Female".
- HourlyRate : Taux horaire de l'employé. Valeurs possibles sont des nombres entiers.
- JobInvolvement : Implication dans le travail de l'employé. Valeurs possibles sont des nombres entiers.
- JobLevel : Niveau de poste de l'employé. Valeurs possibles sont des nombres entiers.
- JobRole : Rôle de l'employé. Valeurs possibles sont "Sales Executive", "Research Scientist", "Laboratory Technician", "Manufacturing Director", "Healthcare Representative", etc.
- JobSatisfaction : Satisfaction professionnelle de l'employé. Valeurs possibles sont des nombres entiers.
- MaritalStatus : Statut marital de l'employé. Valeurs possibles sont "Single", "Married", "Divorced".
- MonthlyIncome : Revenu mensuel de l'employé. Valeurs possibles sont des nombres entiers.
- SalarySlab : Tranche de salaire de l'employé. Valeurs possibles sont "Upto 5k", "5k-10k", "10k-15k", etc.
- MonthlyRate : Taux mensuel de l'employé. Valeurs possibles sont des nombres entiers.
- NumCompaniesWorked : Nombre d'entreprises dans lesquelles l'employé a travaillé. Valeurs possibles sont des nombres entiers.
- Over18 : Si l'employé a plus de 18 ans. Valeurs possibles sont "Y" (oui) ou "N" (non).
- OverTime : Si l'employé fait des heures supplémentaires. Valeurs possibles sont "Yes" (oui) ou "No" (non).
- PercentSalaryHike : Pourcentage d'augmentation de salaire de l'employé. Valeurs possibles sont des nombres entiers.
- PerformanceRating : Évaluation de la performance de l'employé. Valeurs possibles sont des nombres entiers.

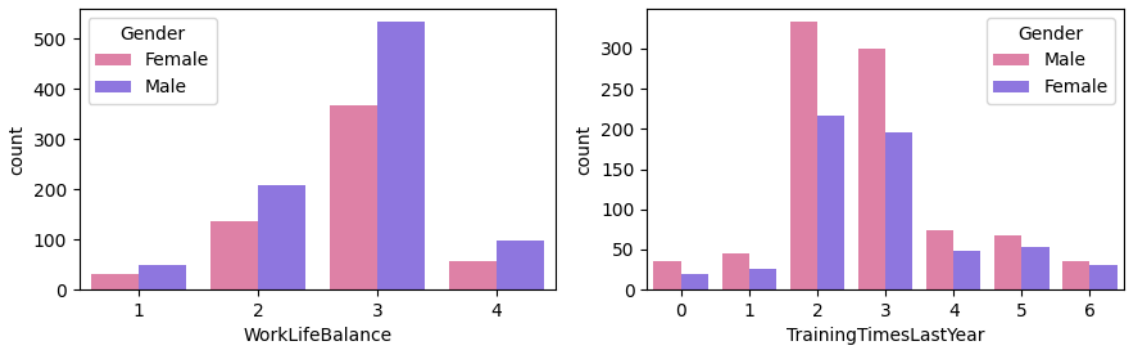
- RelationshipSatisfaction : Satisfaction relationnelle de l'employé. Valeurs possibles sont des nombres entiers.
- StandardHours : Heures standard de l'employé. Valeurs possibles sont des nombres entiers.
- StockOptionLevel : Niveau d'option d'achat d'actions de l'employé. Valeurs possibles sont des nombres entiers.
- TotalWorkingYears : Nombre total d'années de travail de l'employé. Valeurs possibles sont des nombres entiers.
- TrainingTimesLastYear : Nombre de fois que l'employé a été formé l'année dernière. Valeurs possibles sont des nombres entiers.
- WorkLifeBalance : Équilibre entre vie professionnelle et vie privée de l'employé. Valeurs possibles sont des nombres entiers.
- YearsAtCompany : Nombre d'années passées par l'employé dans l'entreprise. Valeurs possibles sont des nombres entiers.
- YearsInCurrentRole : Nombre d'années passées par l'employé dans son rôle actuel. Valeurs possibles sont des nombres entiers.
- YearsSinceLastPromotion : Nombre d'années depuis la dernière promotion de l'employé. Valeurs possibles sont des nombres entiers.
- YearsWithCurrManager : Nombre d'années passées par l'employé avec son manager actuel. Valeurs possibles sont des nombres entiers.

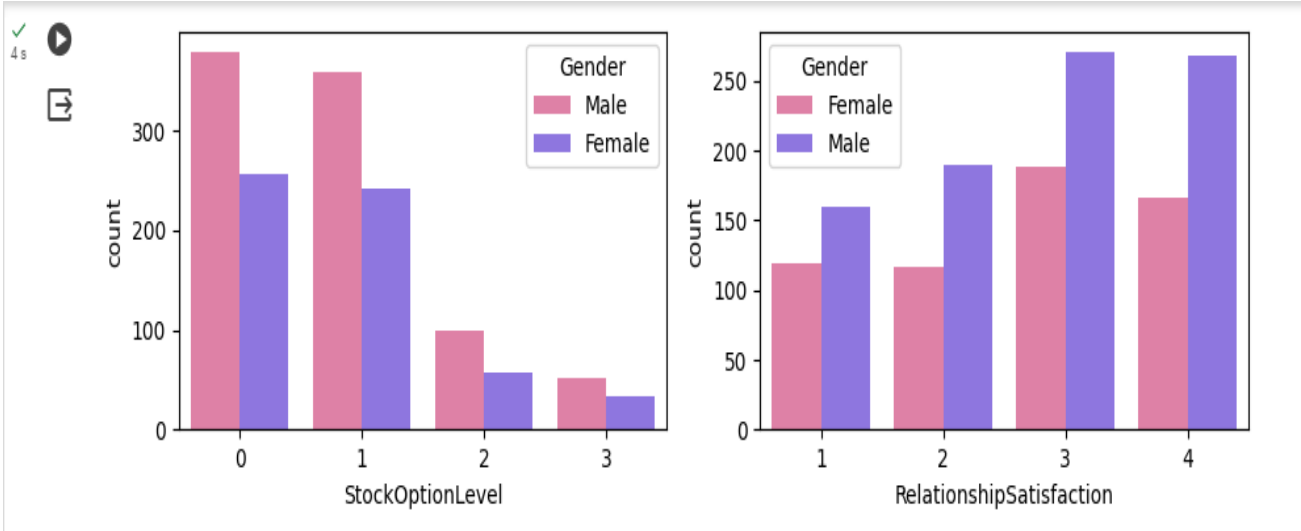
Partie visualisation des données :

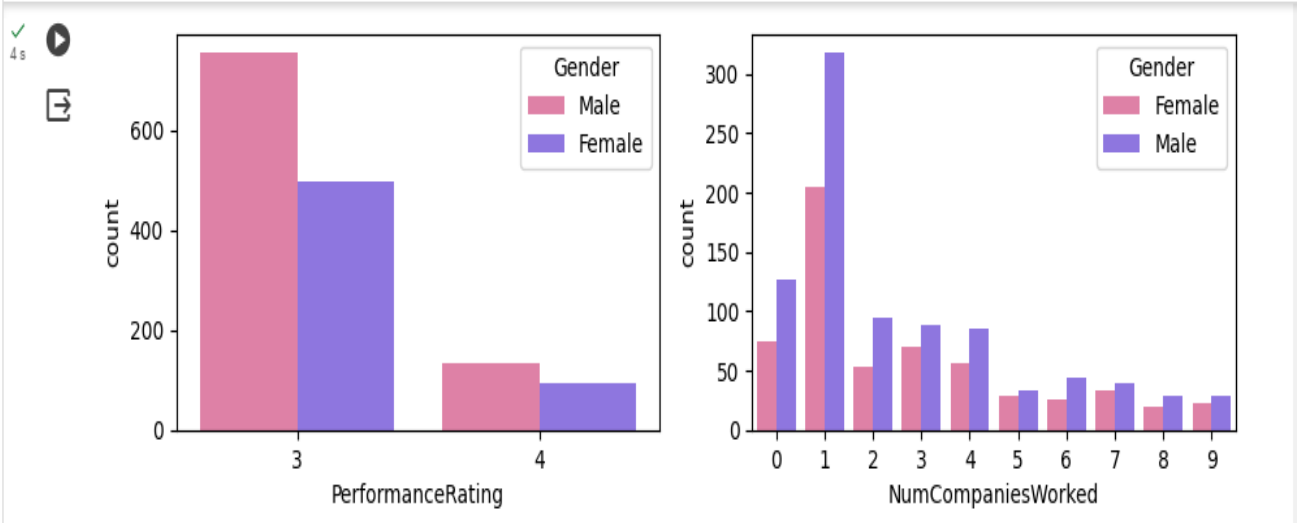
visualisation des données

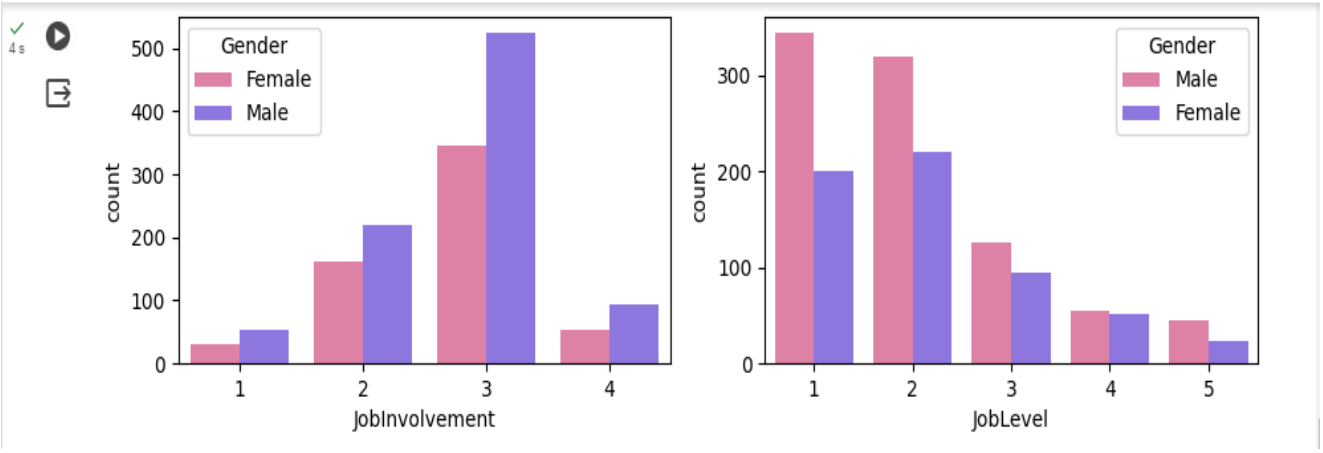
```
colunas = ['WorkLifeBalance','TrainingTimesLastYear','StockOptionLevel',
    'RelationshipSatisfaction','PerformanceRating','NumCompaniesWorked',
   'JobInvolvement', 'JobLevel', 'JobSatisfaction',
   'EnvironmentSatisfaction','Education']

plt.figure(figsize=(9,36))
for i,col in enumerate(colunas):
    axes = plt.subplot(13,2, i + 1)
    sns.countplot(x=HRData [col], hue=HRData ['Gender'], palette=['#ED72A3','#8565F0'])
plt.tight_layout()
plt.show()
```

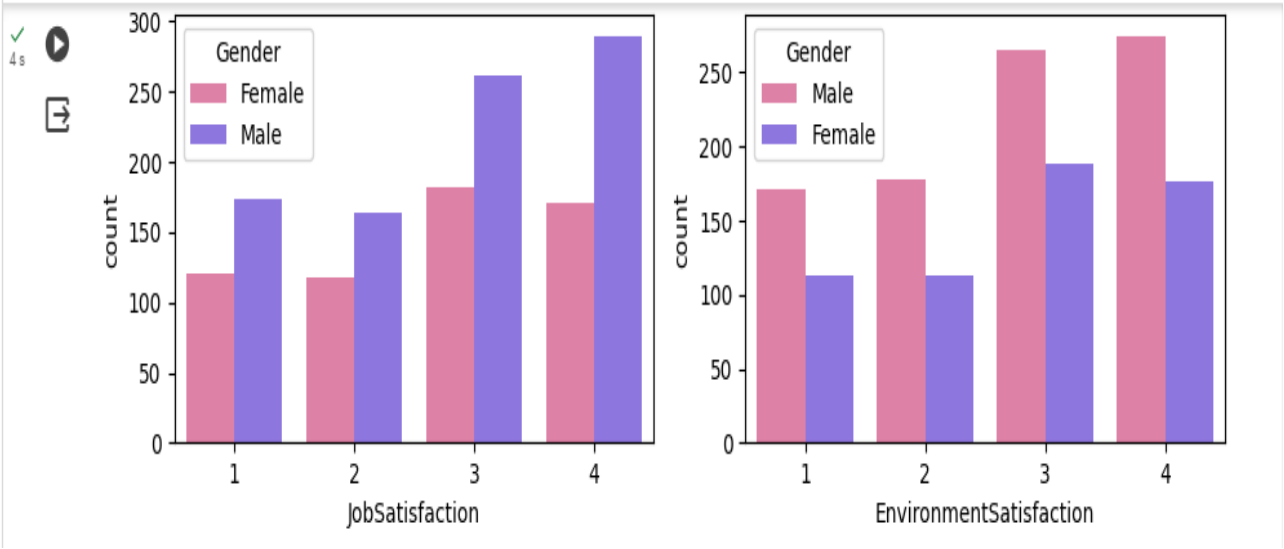
[13]

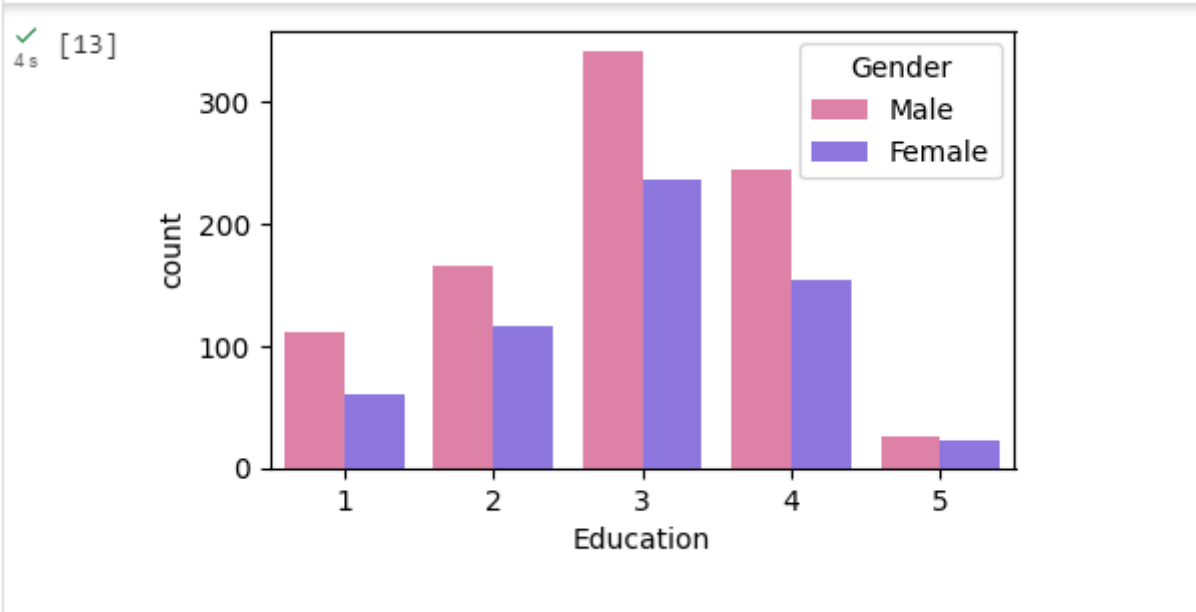



Gender
Male
Female
count
StockOptionLevel
Gender
Female
Male
count
RelationshipSatisfaction


Gender
Male
Female
count
PerformanceRating
Gender
Female
Male
count
NumCompaniesWorked


Gender
Female
Male
count
JobInvolvement
Gender
Male
Female
count
JobLevel


Gender
Female
Male
count
JobSatisfaction
Gender
Male
Female
count
EnvironmentSatisfaction


[13]
Gender
Male
Female
count
Education