

TP Big Data : Pig

Partie 1 : Analyse des employés

Salaire moyen par département

```
(101,53200.0)
(102,57400.0)
(103,63000.0)
(104,62600.0)
```

Nombre d'employés / département

```
(101,5)
(102,5)
(103,5)
(104,5)
```

Liste des employés avec leurs départements

```
(1,Dupont Jean, Male, 50000, Informatique, Paris, France)
(5,Lemoine Lucie, Female, 48000, Informatique, Bordeaux, France)
(13,Pichon Aline, Female, 66000, Informatique, Nice, Suisse)
(17,Bourgeois Sandrine, Female, 53000, Informatique, Lille, France)
(9,Garnier Jules, Male, 49000, Informatique, Rennes, Suisse)
(10,Dubois Sara, Female, 72000, Marketing, Lille, Belgique)
(18,Delaunay R?my, Male, 49000, Marketing, Toulouse, France)
(14,Rousseau Louis, Male, 54000, Marketing, Rennes, Suisse)
(6,Muller Hugo, Male, 52000, Marketing, Nice, France)
(2,Martin Sophie, Female, 60000, Marketing, Lyon, France)
(15,Boucher C?line, Female, 70000, Ventes, Lyon, Suisse)
(7,Roux Camille, Female, 62000, Ventes, Lille, Belgique)
(19,Caron Sophie, Female, 75000, Ventes, Bordeaux, France)
(3,Leblanc Pierre, Male, 55000, Ventes, Marseille, France)
(11,Fournier Amandine, Female, 53000, Ventes, Toulouse, Suisse)
(16,Marchand Olivier, Male, 68000, Ressources Humaines, Nantes, France)
(8,Bernard L?a, Female, 58000, Ressources Humaines, Nantes, Belgique)
(20,Leroy Gabriel, Male, 56000, Ressources Humaines, Nice, France)
(4,Durand Alice, Female, 70000, Ressources Humaines, Toulouse, France)
(12,Giraud Nicolas, Male, 61000, Ressources Humaines, Bordeaux, Suisse)
```

Employés gagnant plus de 60 000

```
(4,Durand Alice, Female, 70000, 104, Toulouse, France)
(7,Roux Camille, Female, 62000, 103, Lille, Belgique)
(10,Dubois Sara, Female, 72000, 102, Lille, Belgique)
(12,Giraud Nicolas, Male, 61000, 104, Bordeaux, Suisse)
(13,Pichon Aline, Female, 66000, 101, Nice, Suisse)
(15,Boucher C?line, Female, 70000, 103, Lyon, Suisse)
(16,Marchand Olivier, Male, 68000, 104, Nantes, France)
(19,Caron Sophie, Female, 75000, 103, Bordeaux, France)
```

Maximum du salaire par département

```
(101,66000)
(102,72000)
(103,75000)
(104,70000)
```

Département avec le salaire le plus élevé (global)

```
(103,75000)
```

Nombre total d'employés

(28)

Employés de Paris

(1,Dupont,Jean,Male,50000,101,Paris,France)

Salaire total par ville

```
(Lyon,130000)
(Nice,118000)
( Nice,56000)
(Lille,134000)
(Paris,50000)
( Lille,53000)
(Nantes,58000)
(Rennes,103000)
( Nantes,68000)
(Bordeaux,109000)
(Toulouse,123000)
( Bordeaux,75000)
( Toulouse,49000)
(Marseille,55000)
```

Départements ayant des femmes

```
(101,Informatique)
(102,Marketing)
(103,Ventes)
(104,Ressources Humaines)
```

Partie 2 : Analyse des films (MovieLens)

Nombre de films par genre

```
(War,12)
(Crime,131)
(Drama,1176)
(Action,503)
(Comedy,1024)
(Horror,262)
(Sci-Fi,46)
(Fantasy,2)
(Musical,25)
(Mystery,36)
(Romance,58)
(Western,33)
(Thriller,181)
(Adventure,155)
(Animation,89)
(Film-Noir,25)
(Children's,90)
(Documentary,123)
```

Top 10 genres

```
(Drama,1176)
(Comedy,1024)
(Action,503)
(Horror,262)
(Adventure,155)
(Crime,131)
(Documentary,123)
(Thriller,181)
(Children's,90)
(Animation,89)
```

Nombre de films par année

```
(1900,0)
(1901,0)
(1902,0)
(1903,0)
(1904,0)
(1905,0)
(1906,0)
(1907,0)
(1908,0)
(1909,0)
(1910,0)
(1911,0)
(1912,0)
(1913,0)
(1914,0)
(1915,0)
(1916,0)
(1917,0)
(1918,0)
(1919,0)
(1920,0)
(1921,0)
(1922,0)
(1923,0)
(1924,0)
(1925,0)
(1926,0)
(1927,0)
(1928,0)
(1929,0)
(1930,0)
(1931,0)
(1932,0)
(1933,0)
(1934,0)
(1935,0)
(1936,0)
(1937,0)
(1938,0)
(1939,0)
(1940,0)
(1941,0)
(1942,0)
(1943,0)
(1944,0)
(1945,0)
(1946,0)
(1947,0)
(1948,0)
(1949,0)
(1950,0)
(1951,0)
(1952,0)
(1953,0)
(1954,0)
(1955,0)
(1956,0)
(1957,0)
(1958,0)
(1959,0)
(1960,0)
(1961,0)
(1962,0)
(1963,0)
(1964,0)
(1965,0)
(1966,0)
(1967,0)
(1968,0)
(1969,0)
(1970,0)
(1971,0)
(1972,0)
(1973,0)
(1974,0)
(1975,0)
(1976,0)
(1977,0)
(1978,0)
(1979,0)
(1980,0)
(1981,0)
(1982,0)
(1983,0)
(1984,0)
(1985,0)
(1986,0)
(1987,0)
(1988,0)
(1989,0)
(1990,0)
(1991,0)
```

Films par année ET par genre

```
(1998,Crime,15)
(1998,Drama,116)
(1998,Action,44)
(1998,Comedy,94)
(1998,Horror,12)
(1998,Sci-Fi,1)
(1998,Mystery,2)
(1998,Romance,6)
(1998,Thriller,5)
(1998,Adventure,8)
(1998,Animation,7)
(1998,Film-Noir,2)
(1998,Children's,7)
(1998,Documentary,18)
(1999,Crime,8)
(1999,Drama,183)
(1999,Action,27)
(1999,Comedy,93)
(1999,Horror,9)
(1999,Sci-Fi,1)
(1999,Mystery,2)
(1999,Romance,3)
(1999,Thriller,10)
(1999,Adventure,1)
(1999,Animation,7)
(1999,Children's,4)
(1999,Documentary,15)
(2000,Crime,4)
(2000,Drama,36)
(2000,Action,19)
(2000,Comedy,65)
(2000,Horror,4)
(2000,Sci-Fi,2)
(2000,Romance,1)
(2000,Thriller,6)
(2000,Adventure,4)
(2000,Animation,6)
(2000,Children's,2)
(2000,Documentary,7)
```

Partie 3 : Analyse de vols (Airline Delay Analysis)

Nombre total de vols

```
2025-11-15 17:02:28,846 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(100)
```

Retard moyen par compagnie aérienne

```
(WN,15.09)
grunt>
```

Retard moyen par aéroport d'origine

```
(IAD,2.0)
(TND,11.35319011776070589)
(CSP,3.357142857142857)
(CAN,48.111111111111114)
(CJYX,9.2686956521739)
(LAS,26.0909690909090909)
```

Retard moyen par aéroport de destination

```
(AHO, 23.0)
(ALB, 78.0)
(AMA, 38.0)
(AUS, 28.666666666666668)
(BDL, 49.0)
(BHM, 7.5)
(BNA, 18.875)
(BOM, 13.0)
(BUF, 52.0)
(BUR, 5.0)
(BWI, 9.0)
(FLL, 1.5555555555555556)
(HOU, 71.4)
(IND, 6.0)
(LAB, 1.0)
(LAS, 5.666666666666667)
(MCI, -7.0)
(MCO, 19.0)
(MDW, 11.5)
(ORF, -0.5)
(PBI, -1.3333333333333333)
(PHL, 23.5)
(PHX, 0.0)
(RSM, -26.0)
(TPA, 5.875)
```

Nombre de vols par compagnie

```
(WN, 190)
```

Retard moyen par compagnie et par aéroport d'origine

```
(CMI, IAD, 2.0)
(CMI, MIA, 11.352941176070889)
(CMI, TSP, 3.35712885712857)
(CMI, JDN, 48.111111111111114)
(CMI, JAX, 9.26086956521739)
(CMI, LAS, 26.090909090909099)
```

Retard moyen par distance (regroupée par tranches)

```
(0_17, 384615384615383)
(500_14, 8)
(1000_0, 7333333333333333)
(1500_24, 025)
(2000_35, 025)
```

Retard total par compagnie

```
(WN, 1509)
```