

Sprint – 1

Understanding the Dataset

Date	04 november 2022
Team ID	PNT2022TMID14579
Project Title	Corporate Employee Attrition Analytics

This project is based on an understanding of the factors to keep employees at the Company and which prompt others to leave. The data can be downloaded from the [Dataset](#) : We need to use only (3 files - General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv) for the current project.

Let us try to understand each field of the data (general_data.csv)

Below are the values each column has. The column names are pretty self-explanatory.

1. AGE Numerical Value
2. ATTRITION Employee leaving the company (0=no, 1=yes)
3. BUSINESS TRAVEL (1=No Travel, 2=Travel Frequently, 3=Travel Rarely)
4. DEPARTMENT (1=HR, 2=R&D, 3=Sales)
5. DISTANCE FROM HOME Numerical Value - THE DISTANCE FROM WORK TO HOME
6. EDUCATION Numerical Value. (1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor')
7. EDUCATION FIELD (1=HR, 2=LIFE SCIENCES, 3=MARKETING, 4=MEDICAL SCIENCES, 5=OTHERS, 6= TECHNICAL)
8. EMPLOYEE COUNT Numerical Value
9. EMPLOYEE ID Numerical Value
10. GENDER (1=FEMALE, 2=MALE)
11. JOB LEVEL Numerical Value
12. JOB ROLE (1=HR REP, 2=HR, 3=LAB TECHNICIAN, 4=MANAGER, 5=MANAGING DIRECTOR, 6= RESEARCH DIRECTOR, 7= RESEARCH SCIENTIST, 8=SALES EXECUTIVE, 9= SALES REPRESENTATIVE)
13. MARITAL STATUS (1=DIVORCED, 2=MARRIED, 3=SINGLE)
14. MONTHLY INCOME Numerical Value - MONTHLY SALARY
15. NUMCOMPANIES WORKED Numerical Value - NO. OF COMPANIES WORKED AT
16. OVER 18 (1=YES, 2=NO)
17. PERCENT SALARY HIKE Numerical Value - PERCENTAGE INCREASE IN SALARY
18. STANDARD HOURS Numerical Value - STANDARD HOURS
19. STOCK OPTIONS LEVEL Numerical Value - STOCK OPTIONS (Higher the number, the more stock option an employee has)
20. TOTAL WORKING YEARS Numerical Value - TOTAL YEARS WORKED
21. TRAINING TIMES LAST YEAR Numerical Value - HOURS SPENT TRAINING

22. YEARS AT COMPANY Numerical Value - TOTAL NUMBER OF YEARS AT THE COMPANY
 23. YEARS SINCE LAST PROMOTION Numerical Value - LAST PROMOTION
 24. YEARS WITH CURRENT MANAGER Numerical Value - YEARS SPENT WITH CURRENT MANAGER
- b. Let us try to understand about each field of the data (employee_survey_data.csv)
1. Employee ID
 2. Environment Satisfaction (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
 3. Job Satisfaction (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
 4. Work Life Balance (1 'Bad', 2 'Good', 3 'Better', 4 'Best')
- c. Let us try to understand about each field of the data (manager_survey_data.csv)
1. Employee ID
 2. Job Involvement (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
 3. Performance Rating (1 'Low', 2 'Good', 3 'Excellent', 4 'Outstanding')

The screenshot shows the IBM Cognos Analytics interface. On the left, the 'Data module' pane lists 'manager_survey_data.csv' as the selected module. The main area displays a grid view of the data with the following columns: Row Id, EmployeeID, JobInvolvement, and PerformanceRating. The data consists of 10 rows.

Row Id	EmployeeID	JobInvolvement	PerformanceRating
1	1	3	3
2	2	2	4
3	3	3	3
4	4	2	3
5	5	3	3
6	6	3	3
7	7	3	4
8	8	3	4
9	9	3	4
10	10	3	3

us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=3343241405_de15e4d147de4c139c71948fb7eb7600_sessionTemp&objRef=&tid=334324...

IBM Cognos Analytics with Watson

New data module

Search content

Properties

Data module

Grid Relationships Custom tables

ge	Attrition	BusinessTravel	Department	DistanceFromHome	EducationalAttainment
1	No	Travel_Rarely	Sales	6	2
1	Yes	Travel_Frequently	Research & Development	10	1
2	No	Travel_Frequently	Research & Development	17	4
3	No	Non-Travel	Research & Development	2	5
2	No	Travel_Rarely	Research & Development	10	1
5	No	Travel_Rarely	Research & Development	8	3
3	Yes	Travel_Rarely	Research & Development	11	2
2	No	Travel_Rarely	Research & Development	18	3
1	No	Travel_Rarely	Research & Development	1	3
5	No	Non-Travel	Research & Development	7	4

27°C Haze

Search the web

ENG IN

10:12 08-11-2022

us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=3343241405_de15e4d147de4c139c71948fb7eb7600_sessionTemp&objRef=&tid=334324...

IBM Cognos Analytics with Watson

New data module

Search content

Properties

Data module

Grid Relationships Custom tables

Row Id	EmployeeID	Environment...	JobSatisfaction	WorkLifeBalance
1	1	3	4	2
2	2	3	2	4
3	3	2	2	1
4	4	4	4	3
5	5	4	1	3
6	6	3	2	2
7	7	1	3	1
8	8	1	2	3
9	9	2	4	3
10	10	2	1	3

27°C Haze

Search the web

ENG IN

10:12 08-11-2022