IBM Employee Attrition Analysis

Work By Ameni Hsini

Major: BA, Minor: IT

Group: BA2

Introduction

Attrition is described as the gradual loss of employees over time, The HR manager of IBM Dallas, a branch of IBM company in Texas, is conducting a study to find the main reasons why a decent number of employees choose to resign over the past five years.
The goal of this report is to analyze **IBM Employee Attrition**, in order to uncover the factors that lead to employee attrition. We will consider if these factors are within the control of the organization and what actions can used to reduce attrition. Finally, based on our results, we will conclude with insights and recommendations.

1.Data Gathering:

This first source is a **csv** file which was exported from IBM HR database, which contains different information related to former and current employees such us:

- **-EmployeeNumber**: which a unique id assigned to each employee
- **-General information**: such as gender, education, address, contacts…
- **-Satisfaction**: which is related to job, environment and relationship, a scale from 1 to 4 was used to rate the satisfaction.
- **-Attrition**: which has two attributes; **yes,** means that the employee has resigned, and **No** that he is still hired.

The following screenshot shows the structure of the initial data source.

EmployeeNu	Age	Attrition	BusinessTrav	Department	phone	city	Education	EmployeeCount	Environment	Gender	JobInvolvem	JobLevel	JobRole	JobSatisfacti	MaritalStatu
1	41	Yes	Travel_Rarel	Sales	1254889779	cockrell_hill	2	1	2	Female	3	2	Sales Execut	4	Single
2	49	No	Travel_Freq	Research & D	1254889779	irving	1	1	3	Male	2	2	Research Sci	2	Married
4	37	Yes	Travel_Rarel	Research & D	1254889779	glenn_heigh	2	1	4	Male	2	1	Laboratory T	3	Single
5	33	No	Travel_Freq	Research & D	1254889779	combine	4	1	4	Female	3	1	Research Sci	3	Married
7	27	No	Travel_Rarel	Research & D	1254889779	glenn_heigh	1	1	1	Male	3	1	Laboratory T	2	Married
8	32	No	Travel_Freq	Research & D	1254889779	glenn_heigh	2	1	4	Male	3	1	Laboratory T	4	Single
10	59	No	Travel_Rarel	Research & D	1254889779	combine	3	1	3	Female	4	1	Laboratory T	1	Married
11	30	No	Travel_Rarel	Research & D	1254889779	farmers bran	1	1	4	Male	3	1	Laboratory T	3	Divorced
12	38	No	Travel_Freq	Research & D	1254889779	carrollton	3	1	4	Male	2	3	Manufacturi	3	Single
13	36	No	Travel_Rarel	Research & D	1254889779	plano	3	1	3	Male	3	2	Healthcare R	3	Married
14	35	No	Travel_Rarel	Research & D	1254889779	balch_spring	3	1	1	Male	4	1	Laboratory T	2	Married
15	29	No	Travel_Rarel	Research & D	1254889779	coppell	2	1	4	Female	2	2	Laboratory T	3	Single
16	31	No	Travel_Rarel	Research & D	1254889779	travis_ranch	1	1	1	Male	3	1	Research Sci	3	Divorced
18	34	No	Travel_Rarel	Research & D	1254889779	irving	2	1	2	Male	3	1	Laboratory T	4	Divorced
19	28	Yes	Travel_Rarel	Research & D	1254889779	farmers bran	3	1	3	Male	2	1	Laboratory T	3	Single
20	29	No	Travel_Rarel	Research & D	1254889779	rowlett	4	1	2	Female	4	3	Manufacturi	1	Divorced
21	32	No	Travel_Rarel	Research & D	1254889779	dallas	2	1	1	Male	4	1	Research Sci	2	Divorced
22	22	No	Non-Travel	Research & D	1254889779	balch_spring	2	1	4	Male	4	1	Laboratory T	4	Divorced
23	53	No	Travel_Rarel	Sales	1254889779	glenn_heigh	4	1	1	Female	2	4	Manager	4	Married
24	38	No	Travel_Rarel	Research & D	1254889779	glenn_heigh	3	1	4	Male	3	1	Research Sci	4	Single

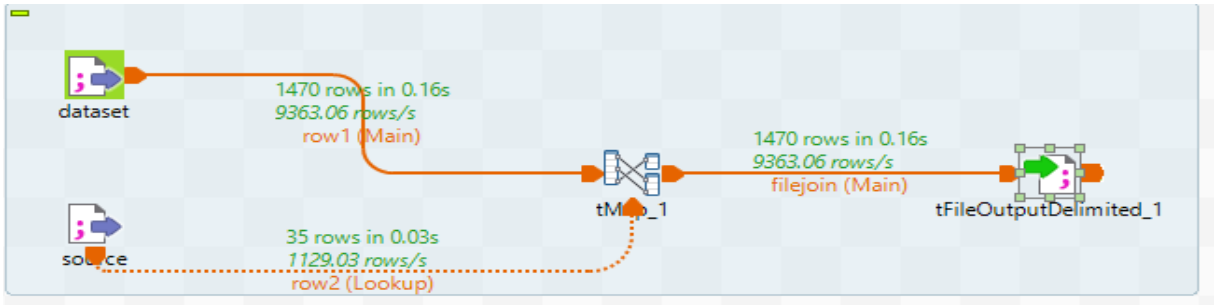
-Since the aim of this analysis is to identify important factors leading to attrition, another dataset of text format was collected; it contains different cities near Texas and the approximate distance between each city and the company location in cockrell_hill, Dallas. The aim of this dataset is to include the effect of the distance from work to home on employee attrition in our analysis.

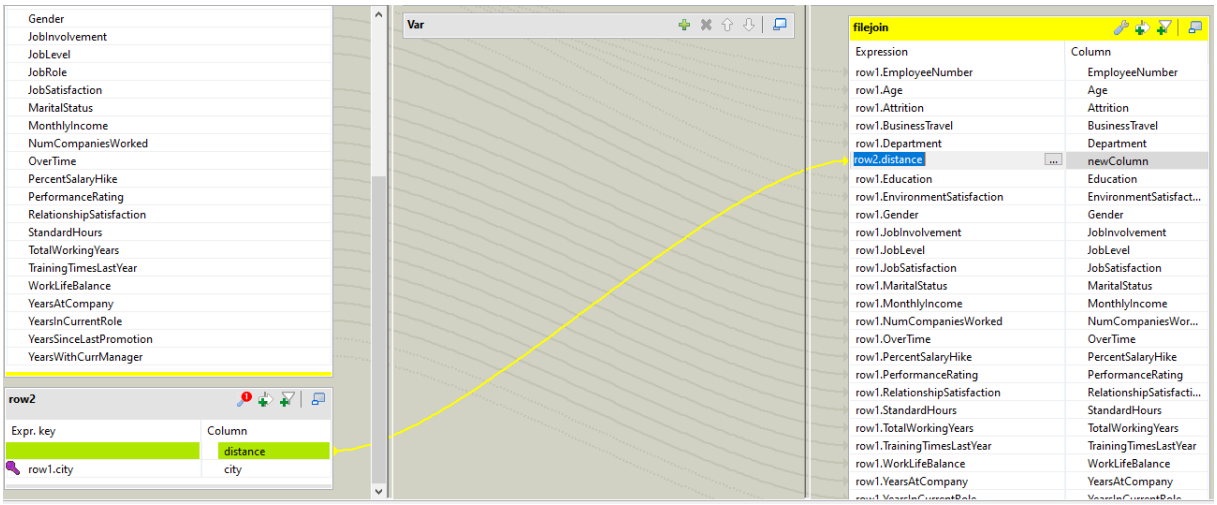
```
distance;city
5;dallas
21;rockwall
29;collin
9;kaufman
14;ellis
24;tarrant
35;denton
45;garland
16;balch_springs
18;grand_prairie
20;hutchins
```

2.ETL process:

This phase was realized via **Talend Open Studio** in order to make some transformation on our dataset so it becomes more adequate to the analysis goals. The used tools were:

- **-data cleaning**: the initial dataset contains a column **irrelevant** to the analysis which is Phone number, **tMap** tool was used to remove it.
- **-file join**: as mentioned earlier, a dataset containing distances was made to highlight distance factor in our analysis, **tMap** tool was used to join the **csv** and **text** type files and replace the column '**city**'(indicating the city in which the employee lives) in the final output with '**distance**' separating that city and the company location, the following screenshots illustrate the process:





The output is as follows:

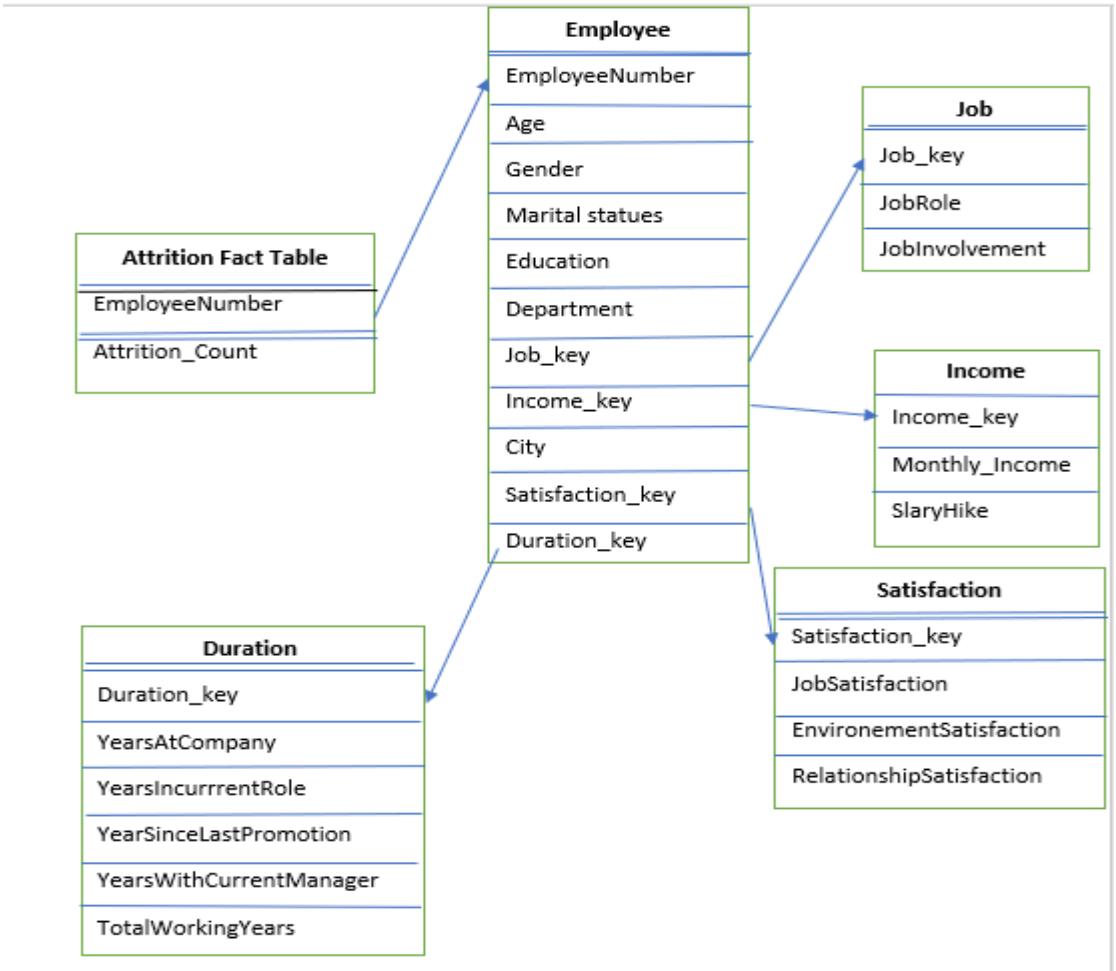
EmployeeNumber	Age	Attrition	BusinessTrav	Department	distance	Education	EmployeeCount	Environment	Gender	JobInvolvem	JobLevel	JobRole	JobSatisfacti	MaritalStatu	Monthl
1	41	Yes	Travel_Rarel	Sales	1	2	1	2	Female	3	2	Sales Execut	4	Single	
2	49	No	Travel_Frequ	Research & D	8	1	1	3	Male	2	2	Research Sci	2	Married	
4	37	Yes	Travel_Rarel	Research & D	2	2	1	4	Male	2	1	Laboratory T	3	Single	
5	33	No	Travel_Frequ	Research & D	3	4	1	4	Female	3	1	Research Sci	3	Married	
7	27	No	Travel_Rarel	Research & D	2	1	1	1	Male	3	1	Laboratory T	2	Married	
8	32	No	Travel_Frequ	Research & D	2	2	1	4	Male	3	1	Laboratory T	4	Single	
10	59	No	Travel_Rarel	Research & D	3	3	1	3	Female	4	1	Laboratory T	1	Married	
11	30	No	Travel_Rarel	Research & D	24	1	1	4	Male	3	1	Laboratory T	3	Divorced	
12	38	No	Travel_Frequ	Research & D	23	3	1	4	Male	2	3	Manufacturi	3	Single	
13	36	No	Travel_Rarel	Research & D	27	3	1	3	Male	3	2	Healthcare R	3	Married	
14	35	No	Travel_Rarel	Research & D	16	3	1	1	Male	4	1	Laboratory T	2	Married	
15	29	No	Travel_Rarel	Research & D	15	2	1	4	Female	2	2	Laboratory T	3	Single	
16	31	No	Travel_Rarel	Research & D	26	1	1	1	Male	3	1	Research Sci	3	Divorced	
18	34	No	Travel_Rarel	Research & D	8	2	1	2	Male	3	1	Laboratory T	4	Divorced	
19	28	Yes	Travel_Rarel	Research & D	24	3	1	3	Male	2	1	Laboratory T	3	Single	
20	29	No	Travel_Rarel	Research & D	21	4	1	2	Female	4	3	Manufacturi	1	Divorced	
21	32	No	Travel_Rarel	Research & D	5	2	1	1	Male	4	1	Research Sci	2	Divorced	
22	22	No	Non-Travel	Research & D	16	2	1	4	Male	4	1	Laboratory T	4	Divorced	
23	53	No	Travel_Rarel	Sales	2	4	1	1	Female	2	4	Manager	4	Married	1
24	38	No	Travel_Rarel	Research & D	2	3	1	4	Male	3	1	Research Sci	4	Single	

3.DW Modeling:

The fact and dimensions were identified as follows:

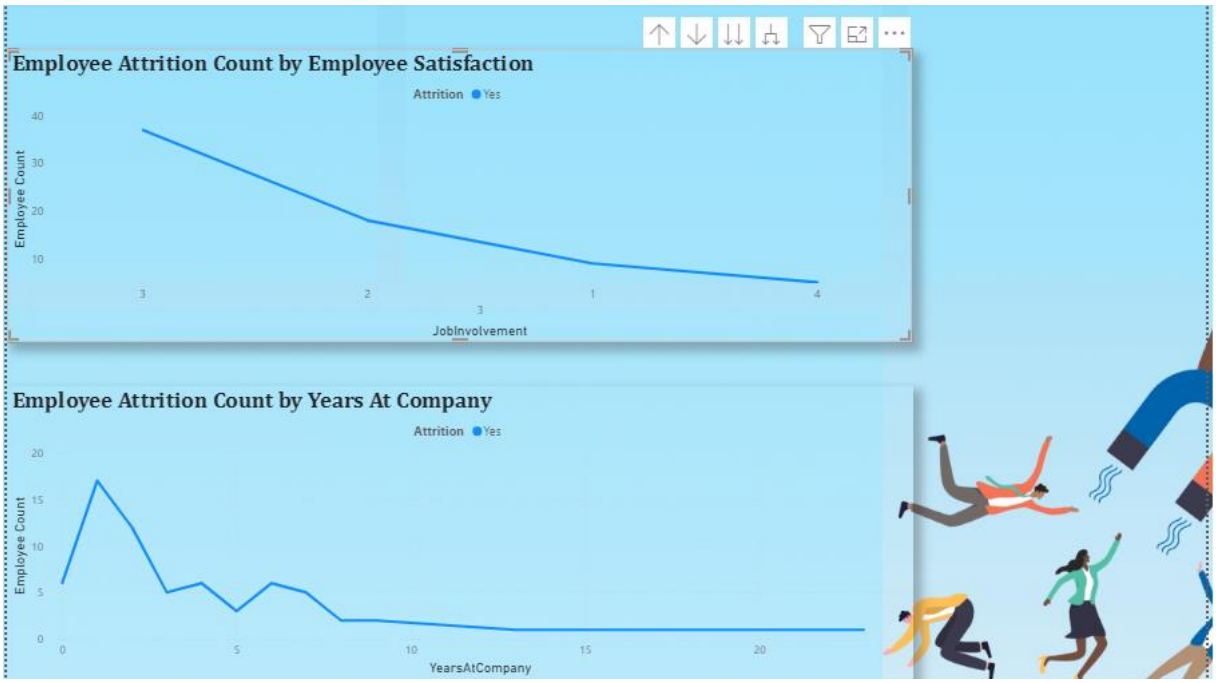
- **-fact**: attrition
- **-measures**: attrition rate
- **-dimension**: employee (Age, Gender, Marital status, Education, Department, Job, Income, City, Satisfaction, Duration)

-Snowflake schema:



4-Data analysis:

Power BI generated the following visualizations:



Insights from the analysis:

Below we will discuss insights from the above analysis and recommendations that can result in action.

- Attrition rate is about 16% over the past 5 years.
- The maximum number of employees (133 employees) are from the Research & Development department which is likely to attrite the company.

We found that among the most important attributes in determining attrition rates are monthly income, job level, satisfaction, training frequency and years at the company.

- **Monthly income**: employees with non-satisfactory monthly income, are much more likely to leave the company.
  - ⇨ a straightforward recommendation for the company is to increase their income level, whether it be promotions or a bonus.
- **Satisfaction and trainings**: measuring job, environment and relationship satisfaction and job involvement indicated that many leaving employees were not satisfied with their working atmosphere. Furthermore, analyzing trainings effects has demonstrated that a high percentage of attired employees lacked trainings.
  - ⇨ low level of satisfaction can be explained by poor company culture, limited professional development opportunities, or a poor work-life balance, lack of trainings, or even the distances that they need to travel to work.
  - ⇨ It is recommended to train the management to lead with empathy and have a human approach at work that can also help employees feel supported and satisfied
  - ⇨ Invest In onboarding, training, and mentoring programs would be efficient.
- **Job level:** employees with entry to intermediate job levels are expected to resign, the manager must look at options to help these workers become more engaged such as more stretching goals, additional training or incentives.
- **Years at company**: When looking at the number of years worked, we also find this to be of interest to human resources in evaluating attrition. We found that the first three years of an employee's time with the company is the most critical.

- ⇨ This time period should be carefully monitored by HR. If these employees begin clocking in overtime or put in requests for promotions and pay increases, then they are at a very high risk of leaving the company

The following is the Talend transformation code:

```
/**
         * [tMap_1 main ] start
         */

        currentComponent="tMap_1";

                boolean hasCasePrimitiveKeyWithNull_tMap_1 = false;


        // ###############################
        // # Input tables (lookups)
                  boolean rejectedInnerJoin_tMap_1 = false;
                  boolean mainRowRejected_tMap_1 = false;

                                ////////////////////////////////////////////
                                // Starting Lookup Table "row2"
                                ////////////////////////////////////////////


                                 boolean forceLooprow2 = false;


                                   row2Struct row2ObjectFromLookup = null;

                                   if(!rejectedInnerJoin_tMap_1) { // G_TM_M_020


                                    hasCasePrimitiveKeyWithNull_tMap_1 = false;

                                                    row2HashKey.city = row1.city ;


                                        row2HashKey.hashCodeDirty = true

                                       tHash_Lookup_row2.lookup( row2HashKey );


                                                        } // G_TM_M_020


                                                        if(tHash_Lookup_row2 != null &&
tHash_Lookup_row2.getCount(row2HashKey) > 1) { // G 071

//System.out.println("WARNING: UNIQUE MATCH is configured for the lookup 'row2' and it contains more
one result from keys :  row2.city = '" + row2HashKey.city + "'");
                                                                } // G 071


                                                        row2Struct row2 = null;

                                                       row2Struct fromLookup_row2 = null;
                                                        row2 = row2Default;


                                                                if (tHash_Lookup_row2 !=null &&
tHash_Lookup_row2.hasNext()) { // G 099



                                                                fromLookup_row2 =
tHash_Lookup_row2.next();



                                                                } // G 099
```

```
                                                    if(fromLookup_row2 != null) {
                                                            row2 = fromLookup_row2;
                                                    }


                        // ###############################
                { // start of Var scope

                        // ###############################
                // # Vars tables

Var__tMap_1__Struct Var = Var__tMap_1;// ###############################
                // ###############################
                // # Output tables

joining = null;


// # Output table : 'joining'
joining_tmp.EmployeeNumber = row1.EmployeeNumber ;
joining_tmp.Age = row1.Age ;
joining_tmp.Attrition = row1.Attrition ;
joining_tmp.BusinessTravel = row1.BusinessTravel ;
joining_tmp.Department = row1.Department ;
joining_tmp.distance = row2.distance ;
joining_tmp.Education = row1.Education ;
joining_tmp.EmployeeCount = row1.EmployeeCount ;
joining_tmp.EnvironmentSatisfaction = row1.EnvironmentSatisfaction ;
joining_tmp.Gender = row1.Gender ;
joining_tmp.JobInvolvement = row1.JobInvolvement ;
joining_tmp.JobLevel = row1.JobLevel ;
joining_tmp.JobRole = row1.JobRole ;
joining_tmp.JobSatisfaction = row1.JobSatisfaction ;
joining_tmp.MaritalStatus = row1.MaritalStatus ;
joining_tmp.MonthlyIncome = row1.MonthlyIncome ;
joining_tmp.NumCompaniesWorked = row1.NumCompaniesWorked ;
joining_tmp.OverTime = row1.OverTime ;
joining_tmp.PercentSalaryHike = row1.PercentSalaryHike ;
joining_tmp.PerformanceRating = row1.PerformanceRating ;
joining_tmp.RelationshipSatisfaction = row1.RelationshipSatisfaction ;
joining_tmp.StandardHours = row1.StandardHours ;
joining_tmp.TotalWorkingYears = row1.TotalWorkingYears ;
joining_tmp.TrainingTimesLastYear = row1.TrainingTimesLastYear ;
joining_tmp.WorkLifeBalance = row1.WorkLifeBalance ;
joining_tmp.YearsAtCompany = row1.YearsAtCompany ;
joining_tmp.YearsInCurrentRole = row1.YearsInCurrentRole ;
joining_tmp.YearsSinceLastPromotion = row1.YearsSinceLastPromotion ;
joining_tmp.YearsWithCurrManager = row1.YearsWithCurrManager ;
joining = joining_tmp;
// ###############################

} // end of Var scope

rejectedInnerJoin_tMap_1 = false;


        tos_count_tMap_1++;

/**
 * [tMap_1 main ] stop
 */
```

Project sources:

[IBM HR Analytics Employee Attrition & Performance | Kaggle]

[List of Towns and Cities Near Cockrell Hill (Texas) and suburbs - Within 45 Miles Distance of Cockrell Hill Texas United States Between 0 and 72.41 Kilometers Radius List with Population Data (distantias.com)]