Project Report
On

HR-Employee-Attrition

Akhilesh Godishala

AIT-580 Analytics: Big Data to Information

1. Abstract

Employee attrition is a circumstance which an organization or employer is faced with when employees quit. People tend to change occupations to other organizations where they are dissatisfied with their current job or have issues with the office facilities and culture. In general, when there is high demand for people in an industry due to mass retirements or organizational expansion, employee attrition will be very high. When there are more job opportunities on the market, there will be a higher attrition rate. Due to the demand for software goods across all industries, the software industry once suffered a significant attrition rate from employers due to large openings globally in the software business. The Human Resources Department maintains the records of employees in-order to identify and mitigate the factors causing attrition.

2.Introduction

The dataset I've chosen is that of HR-Employee-Attrition. I've always been wanting to understand the working of an organization (i.e. What all are the attributes of an employee that are tracked by an organization). By using this dataset for my project, I will understand the factors considered in a particular decision-making process.

The Questions which I want to answer through this dataset are:

i) Is there higher attrition in a particular department?
ii) How is the attrition being influenced by the Percentage of Salary Hike?
iii) What are major factors causing employees to quit?
iv) To which field do the most educated people belong to?

By Answering these questions, organizations can take measures to improve the retention rate of their employees.

3.Literature Review

Employee attrition is a circumstance in which an organization or employer is faced with when employees quit. People tend to change occupations to other organizations where they are dissatisfied with their current job or have issues with the office facilities and cultureWhen there are more job opportunities on the market, there will be a higher attrition rate. Due to the demand for software goods across all industries, the software industry once suffered a significant attrition rate from employers due to large openings globally in the software business.

The above thread explains the topic and briefly states a few factors causing Employee attrition.

The concept of employee attrition & retention is the representation of employee movement in an organization, which is thought of by HR researchers. They are two sides of same coin. Employee attrition & retention may be result of the negative or positive influence of the various factors. The needs and wants of employees are ever-changing. Their expectation can be expressed as their wish to be employed at a place to work to that of being employed at a Great place to work. The organizations that are able to cope up with these varying factors of employee profile will be able to improve employee commitment, reduce attrition and hence will be able to retain their employees.

In the 1980s, a great deal of literature evolved calling for a strategic role for human resources. The increased interest in HR stems from the assumption that employees and the way they are managed are critically important to the achievement of organizational goals, and provide an edge which can sustain a competitive advantage over the competition. Research and professionals agree that HR management will play an increasingly important role in the future. HR will need to identify the relevant factors and take steps to minimize attrition in order to maintain high performer talent pools.

The above thread explains the topic and also implies the importance of HR management in the analysis and control of Employee attrition.

The majority of voluntary turnovers are caused by people who hunt for more money, better benefits, an improved work/life balance, more opportunities for career advancement, time for health problems or relocations, greater flexibility, or to escape toxic or ineffective managers or workplaces.

Exit interviews should be facilitated by Human Resources for all departing employees. Developing a deeper understanding of the causes of voluntary turnover -and finding solutions to those problems -is an essential part of talent management. Assuring employees that their answers will remain confidential and that they won't affect the company's response to references or employment confirmation requests can encourage them to be honest in exit interviews.

The above thread briefly sheds light on some of the factors causing Employee attrition. Our Analysis of the dataset for this project will help us attain a clearer idea on the topic.

4. Data Pre-Processing

Data preprocessing refers to the modification of data before it is used in order to enhance performance during its analysis and model building. It is an important step in the data mining process. The Data which we have used for the project is in the form of a csv file. It contains 1470 rows and 33 columns.

Our data contains few Categorical variables like Attrition, Gender, Department and EducationField.

It also contains some numeric variables like MonthlyIncome, DistanceFromHome and YearsAtCompany

Data Cleaning: Data cleaning is the process of modifying the data in order to analyze or build models using the data.

Before Data Cleaning:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Age	Attrition	BusinessTrav	DailyRate	Department	DistanceFror	Education	EducationFie	EmployeeNu	Environment	Gender	HourlyRate	JobInvolvem	JobLevel	JobRole	JobSatisfacti	MaritalStatu	MonthlyInco	MonthlyRate	NumCompar	Over18	OverTime
2	41	Yes	Travel_Rarel	1102	Sales	1	2	Life Sciences	1	2	Female	94	3	2	Sales Execut	4	Single	5993	19479	8	Y	Yes
3	49	No	Travel_Frequ	279	Research & I	8	1	Life Sciences	2	3	Male	61	2	2	Research Sci	2	Married	5130	24907	1	Y	No
4	37	Yes	Travel_Rarel	1373	Research & I	2	2	Other	4	4	Male	92	2	1	Laboratory T	3	Single	2090	2396	6	Y	Yes
5	33	No	Travel_Frequ	1392	Research & I	3	4	Life Sciences	5	4	Female	56	3	1	Research Sci	3	Married	2909	23159	1	Y	Yes
6	27	No	Travel_Rarel	591	Research & I	2	1	Medical	7	1	Male	40	3	1	Laboratory T	2	Married	3468	16632	9	Y	No
7	32	No	Travel_Frequ	1005	Research & I	2	2	Life Sciences	8	4	Male	79	3	1	Laboratory T	4	Single	3068	11864	0	Y	No
8	59	No	Travel_Rarel	1324	Research & I	3	3	Medical	10	3	Female	81	4	1	Laboratory T	1	Married	2670	9964	4	Y	Yes
9	30	No	Travel_Rarel	1358	Research & I	24	1	Life Sciences	11	4	Male	67	3	1	Laboratory T	3	Divorced	2693	13335	1	Y	No
10	38	No	Travel_Frequ	216	Research & I	23	3	Life Sciences	12	4	Male	44	2	3	Manufacturi	3	Single	9526	8787	0	Y	No
11	36	No	Travel_Rarel	1299	Research & I	27	3	Medical	13	3	Male	94	3	2	Healthcare R	3	Married	5237	16577	6	Y	No
12	35	No	Travel_Rarel	809	Research & I	16	3	Medical	14	1	Male	84	4	1	Laboratory T	2	Married	2426	16479	0	Y	No
13	29	No	Travel_Rarel	153	Research & I	15	2	Life Sciences	15	4	Female	49	2	2	Laboratory T	3	Single	4193	12682	0	Y	Yes
14	31	No	Travel_Rarel	670	Research & I	26	1	Life Sciences	16	1	Male	31	3	1	Research Sci	3	Divorced	2911	15170	1	Y	No
15	34	No	Travel_Rarel	1346	Research & I	19	2	Medical	18	2	Male	93	3	1	Laboratory T	4	Divorced	2661	8758	0	Y	No
16	28	Yes	Travel_Rarel	103	Research & I	24	3	Life Sciences	19	3	Male	50	2	1	Laboratory T	3	Single	2028	12947	5	Y	Yes
17	29	No	Travel_Rarel	1389	Research & I	21	4	Life Sciences	20	2	Female	51	4	3	Manufacturi	1	Divorced	9980	10195	1	Y	No
18	32	No	Travel_Rarel	334	Research & I	5	2	Life Sciences	21	1	Male	80	4	1	Research Sci	2	Divorced	3298	15053	0	Y	Yes
19	22	No	Non-Travel	1123	Research & I	16	2	Medical	22	4	Male	96	4	1	Laboratory T	4	Divorced	2935	7324	1	Y	Yes
20	53	No	Travel_Rarel	1219	Sales	2	4	Life Sciences	23	1	Female	78	2	4	Manager	4	Married	15427	22021	2	Y	No
21	38	No	Travel_Rarel	371	Research & I	2	3	Life Sciences	24	4	Male	45	3	1	Research Sci	4	Single	3944	4306	5	Y	Yes
22	24	No	Non-Travel	673	Research & I	11	2	Other	26	1	Female	96	4	2	Manufacturi	3	Divorced	4011	8232	0	Y	No
23	36	Yes	Travel_Rarel	1218	Sales	9	4	Life Sciences	27	3	Male	82	2	1	Sales Repres	1	Single	3407	6986	7	Y	No
24	34	No	Travel_Rarel	419	Research & I	7	4	Life Sciences	28	1	Female	53	3	3	Research Dir	2	Single	11994	21293	0	Y	No
25	21	No	Travel_Rarel	391	Research & I	15	2	Life Sciences	30	3	Male	96	3	1	Research Sci	4	Single	1232	19281	1	Y	No
26	34	Yes	Travel_Rarel	699	Research & I	6	1	Medical	31	2	Male	83	3	1	Research Sci	1	Single	2960	17102	2	Y	No
27	53	No	Travel_Rarel	1282	Research & I	5	3	Other	32	3	Female	58	3	5	Manager	3	Divorced	19094	10735	4	Y	No
28	32	Yes	Travel_Frequ	1125	Research & I	16	1	Life Sciences	33	2	Female	72	1	1	Research Sci	1	Single	3919	4681	1	Y	Yes
29	42	No	Travel_Rarel	691	Sales	8	4	Marketing	35	3	Male	48	3	2	Sales Execut	2	Married	6825	21173	0	Y	No
30	44	No	Travel_Rarel	477	Research & I	7	4	Medical	36	1	Female	42	2	3	Healthcare R	4	Married	10248	2094	3	Y	No
31	46	No	Travel_Rarel	705	Sales	2	4	Marketing	38	2	Female	83	3	5	Manager	1	Single	18947	22822	3	Y	No
32	33	No	Travel_Rarel	924	Research & I	2	3	Medical	39	3	Male	78	3	1	Laboratory T	4	Single	2496	6670	4	Y	No
33	44	No	Travel_Rarel	1459	Research & I	10	4	Other	40	4	Male	41	3	2	Healthcare R	4	Married	6465	19121	2	Y	Yes
34	30	No	Travel_Rarel	125	Research & I	9	2	Medical	41	4	Male	83	2	1	Laboratory T	3	Single	2206	16117	1	Y	No
35	39	Yes	Travel_Rarel	895	Sales	5	3	Technical De	42	4	Male	56	3	2	Sales Repres	4	Married	2086	3335	3	Y	No
36	24	Yes	Travel_Rarel	813	Research & I	1	3	Medical	45	2	Male	61	3	1	Research Sci	4	Married	2293	3020	2	Y	Yes
37	43	No	Travel_Rarel	1273	Research & I	2	2	Medical	46	4	Female	72	4	1	Research Sci	3	Divorced	2645	21923	1	Y	No
38	50	Yes	Travel_Rarel	869	Sales	3	2	Marketing	47	1	Male	86	2	1	Sales Repres	3	Married	2683	3810	1	Y	Yes
39	35	No	Travel_Rarel	890	Sales	2	3	Marketing	49	4	Female	97	3	1	Sales Repres	4	Married	2014	9687	1	Y	No

The steps I have followed for Data cleaning are:

i) I have checked for Null values and Duplicate rows. There were no such values.

ii) I have removed the following columns (Over18, EmployeeNumber, HourlyRate, MonthlyRate, DailyRate) as they had a single value or they were of no use in predicting the attrition.

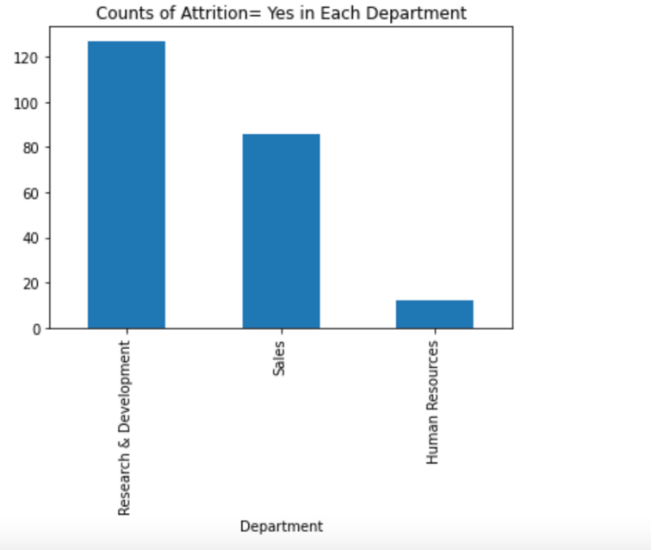
iii) I have checked and removed the outliers using Boxplot visualization from the library matplotlib

After Data Cleaning:

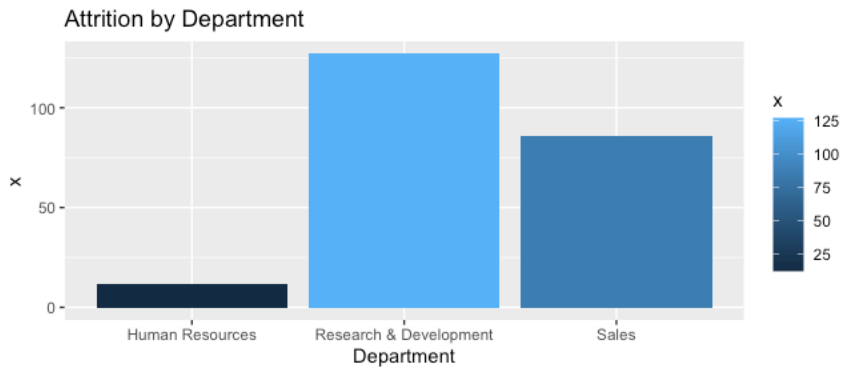
Age	Attrition	BusinessTrav	Department	DistanceFror	Education	EducationFie	Environment	Gender	JobInvolvem	JobLevel	JobRole	JobSatisfacti	MaritalStatu	MonthlyInco	NumCompar	OverTime	PercentSalar	Performance	RelationshipS	StockOption
41	Yes	Travel_Rarel	Sales	1	2	Life Sciences	2	Female	3	2	Sales Execut	4	Single	5993	8	Yes	11	3	1	0
49	No	Travel_Frequ	Research & D	8	1	Life Sciences	3	Male	2	2	Research Sci	2	Married	5130	1	No	23	4	4	1
37	Yes	Travel_Rarel	Research & D	2	2	Other	4	Male	2	1	Laboratory T	3	Single	2090	6	Yes	15	3	2	0
33	No	Travel_Frequ	Research & D	3	4	Life Sciences	4	Female	3	1	Research Sci	3	Married	2909	1	Yes	11	3	3	0
27	No	Travel_Rarel	Research & D	2	1	Medical	1	Male	3	1	Laboratory T	2	Married	3468	9	No	12	3	4	1
32	No	Travel_Frequ	Research & D	2	2	Life Sciences	4	Male	3	1	Laboratory T	4	Single	3068	0	No	13	3	3	0
59	No	Travel_Rarel	Research & D	3	3	Medical	3	Female	4	1	Laboratory T	1	Married	2670	4	Yes	20	4	1	3
30	No	Travel_Rarel	Research & D	24	1	Life Sciences	4	Male	3	1	Laboratory T	3	Divorced	2693	1	No	22	4	2	1
38	No	Travel_Frequ	Research & D	23	3	Life Sciences	4	Male	2	3	Manufacturi	3	Single	9526	0	No	21	4	2	0
36	No	Travel_Rarel	Research & D	27	3	Medical	3	Male	3	2	Healthcare R	3	Married	5237	6	No	13	3	2	2
35	No	Travel_Rarel	Research & D	16	3	Medical	1	Male	4	1	Laboratory T	2	Married	2426	0	No	13	3	3	1
29	No	Travel_Rarel	Research & D	15	2	Life Sciences	4	Female	2	2	Laboratory T	3	Single	4193	0	Yes	12	3	4	0
31	No	Travel_Rarel	Research & D	26	1	Life Sciences	1	Male	3	1	Research Sci	3	Divorced	2911	1	No	17	3	4	1
34	No	Travel_Rarel	Research & D	19	2	Medical	2	Male	3	1	Laboratory T	4	Divorced	2661	0	No	11	3	3	1
28	Yes	Travel_Rarel	Research & D	24	3	Life Sciences	3	Male	2	1	Laboratory T	3	Single	2028	5	Yes	14	3	2	0
29	No	Travel_Rarel	Research & D	21	4	Life Sciences	2	Female	4	3	Manufacturi	1	Divorced	9980	1	No	11	3	3	1
32	No	Travel_Rarel	Research & D	5	2	Life Sciences	1	Male	4	1	Research Sci	2	Divorced	3298	0	Yes	12	3	4	2
22	No	Non-Travel	Research & D	16	2	Medical	4	Male	4	1	Laboratory T	4	Divorced	2935	1	Yes	13	3	2	2
38	No	Travel_Rarel	Research & D	2	3	Life Sciences	4	Male	3	1	Research Sci	4	Single	3944	5	Yes	11	3	3	0
24	No	Non-Travel	Research & D	11	2	Other	1	Female	4	2	Manufacturi	3	Divorced	4011	0	No	18	3	4	1
36	Yes	Travel_Rarel	Sales	9	4	Life Sciences	3	Male	2	1	Sales Repres	1	Single	3407	7	No	23	4	2	0
21	No	Travel_Rarel	Research & D	15	2	Life Sciences	3	Male	3	1	Research Sci	4	Single	1232	1	No	14	3	4	0
34	Yes	Travel_Rarel	Research & D	6	1	Medical	2	Male	3	1	Research Sci	1	Single	2960	2	No	11	3	3	0
32	Yes	Travel_Frequ	Research & D	16	1	Life Sciences	2	Female	1	1	Research Sci	1	Single	3919	1	Yes	22	4	2	0
42	No	Travel_Rarel	Sales	8	4	Marketing	3	Male	3	2	Sales Execut	2	Married	6825	0	No	11	3	4	1
44	No	Travel_Rarel	Research & D	7	4	Medical	1	Female	2	3	Healthcare R	4	Married	10248	3	No	14	3	4	1
33	No	Travel_Rarel	Research & D	2	3	Medical	3	Male	3	1	Laboratory T	4	Single	2496	4	No	11	3	4	0
44	No	Travel_Rarel	Research & D	10	4	Other	4	Male	3	2	Healthcare R	4	Married	6465	2	Yes	13	3	4	0
30	No	Travel_Rarel	Research & D	9	2	Medical	4	Male	2	1	Laboratory T	3	Single	2206	1	No	13	3	1	0
39	Yes	Travel_Rarel	Sales	5	3	Technical De	4	Male	3	2	Sales Repres	4	Married	2086	3	No	14	3	3	1
24	Yes	Travel_Rarel	Research & D	1	3	Medical	2	Male	3	1	Research Sci	4	Married	2293	2	Yes	16	3	1	1
43	No	Travel_Rarel	Research & D	2	2	Medical	4	Female	4	1	Research Sci	3	Divorced	2645	1	No	12	3	4	2
50	Yes	Travel_Rarel	Sales	3	2	Marketing	1	Male	2	1	Sales Repres	3	Married	2683	1	Yes	14	3	3	0

Answering the Research Questions:

1) Is there higher attrition in a particular department?



Visualization in Python

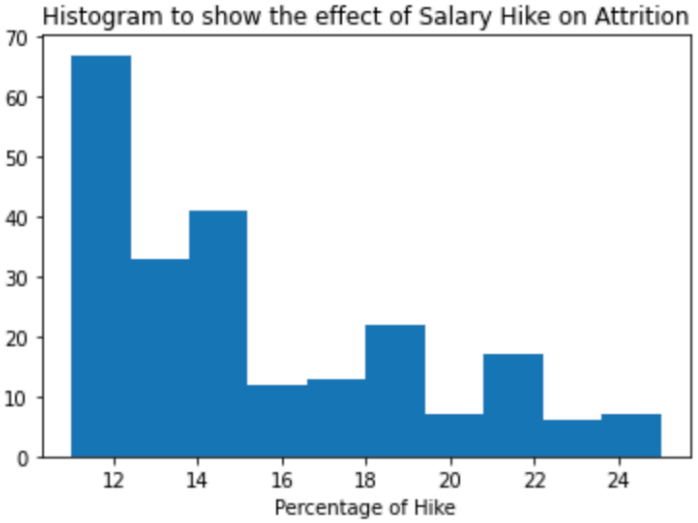


Visualization in R

From the above visualizations we can say that there is higher attrition in Research and Development Department.

2) How is the attrition being influenced by the Percentage of Salary Hike?

Visualization in R



Visualization in Python

From the above visualizations we can say that as the Salary Hike increases Attrition Decreases.

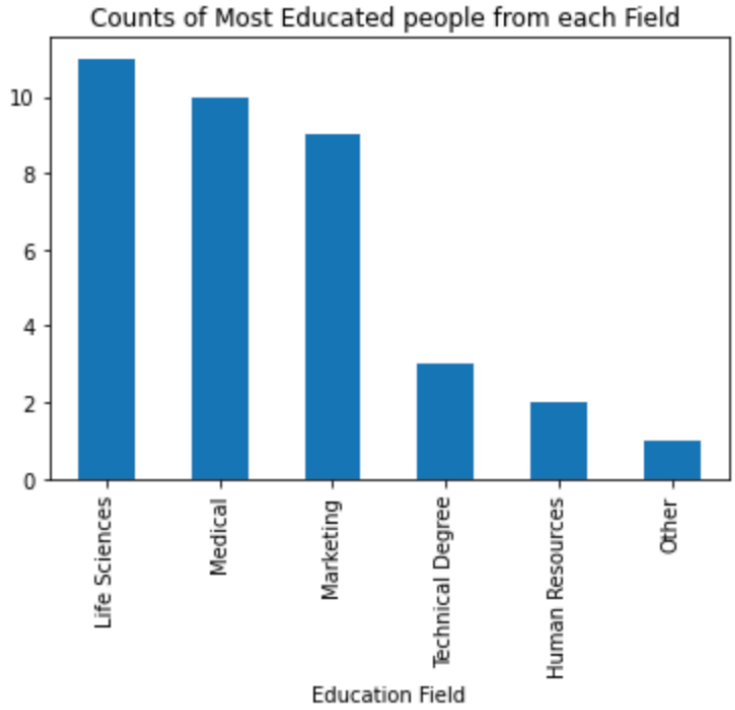
3) What are major factors causing employees to quit?

This can be known by using the function dataframe.corr() . After looking at the output we can conclude that the major factors influencing attrition are:

i) Did the employee work Overtime
ii) Years in Current Role
iii) Years with Current Manager
iv) Years At company
v) Job Level

-ve Correlation indicates Inversely proportional and +ve correlation indicates directly proportional

4) To which field do the most educated people belong to?



Most Educated people belong to Life Sciences.

I have also built a Decision Tree Classifier to implement the learnings of my course. The code and its output has been attached in python file attached in the submission(.ipynb)

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,predict)
```

```
0.8063241106719368
```

5. **SQL Commands**

```
7
8  select * from attrition limit 10;
9
```

100% 1:8

Result Grid | Filter Rows: Search | Export: | Fetch rows:

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHo...	Education	EducationField	EmployeeNumber	EnvironmentSatisfacti...	Gender	H
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	2	Female	9
49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	2	3	Male	6
37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	4	4	Male	9
33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	5	4	Female	5
27	No	Travel_Rarely	591	Research & Development	2	1	Medical	7	1	Male	4
32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	8	4	Male	7
59	No	Travel_Rarely	1324	Research & Development	3	3	Medical	10	3	Female	8
30	No	Travel_Rarely	1358	Research & Development	24	1	Life Sciences	11	4	Male	6
38	No	Travel_Frequently	216	Research & Development	23	3	Life Sciences	12	4	Male	4
36	No	Travel_Rarely	1299	Research & Development	27	3	Medical	13	3	Male	9

Brief Overview of data

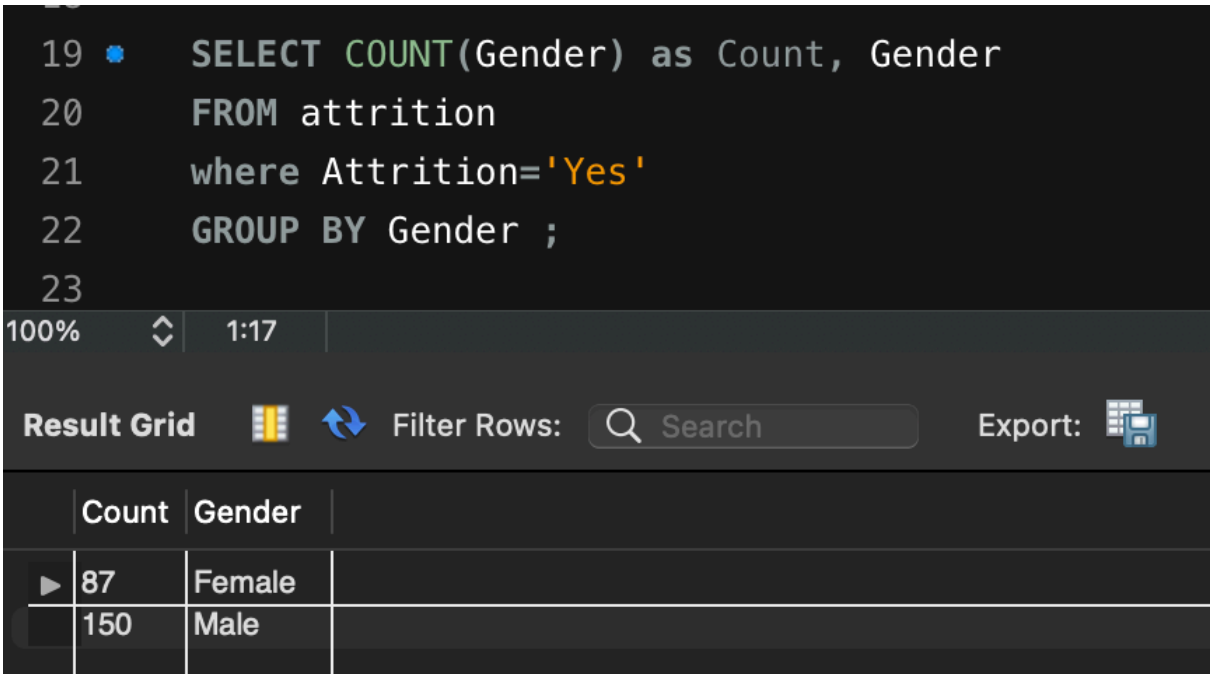
```
15  Select Department,count(*) from attrition  Group By Department;
16
```

100% 1:13

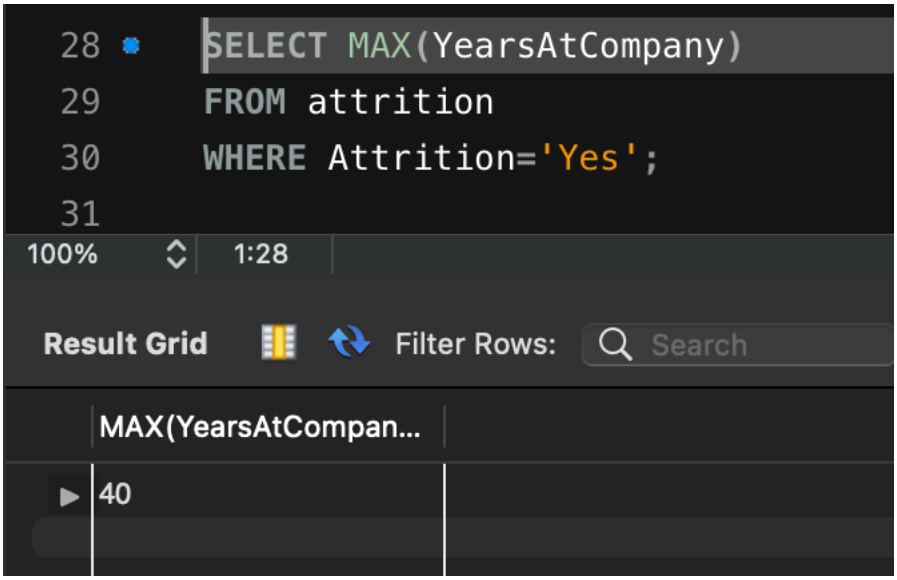
Result Grid | Filter Rows: Search | Export:

Department	count(*)
Sales	446
Research & Development	961
Human Resources	63

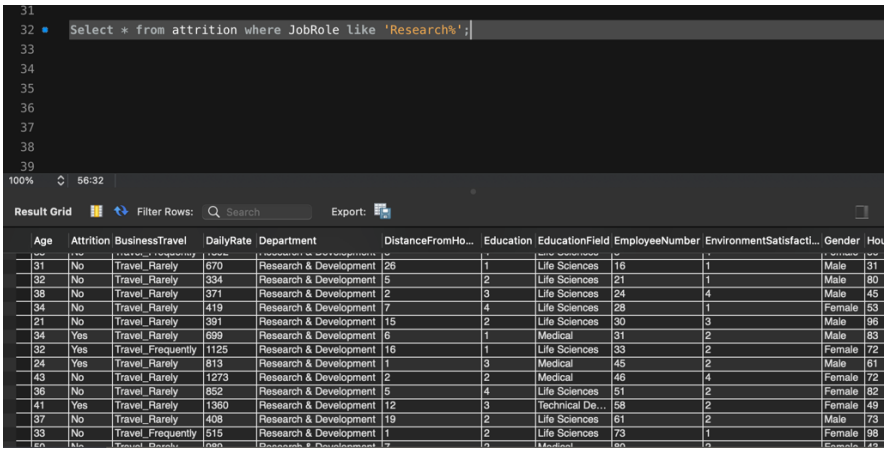
Use of group By command



Use of Count and Group by Function



Use of max function



Use of 'Like' command

6. Conclusion

The outcomes of the Research Questions can be analyzed and incorporated into the values of an organization by the human resources department in order to improve the retention rate of employees .

Citations:

i) Holliday, M. (n.d.). *Why employees leave and how to prevent it*. Oracle NetSuite. Retrieved April 8, 2022, from https://www.netsuite.com/portal/resource/articles/human-resources/employee-turnover-causes.shtml

ii) Shivani Mishra. REVIEW OF LITERATURE ON FACTORS INFLUENCING ATTRITION AND RETENTION. International Journal of Organizational Behaviour & Management Perspectives © Pezzottaite Journals

iii)Literature review on employee attrition. (n.d.). Retrieved April 8, 2022, from https://www.ipl.org/essay/Literature-Review-On-Employee-Attrition-P3TLQF7ESJP6

iv) Hr-employee-attrition - dataset by Aaizemberg. data.world. (2020, August 20). Retrieved March 27, 2022, from https://data.world/aaizemberg/hr-employee-attrition/workspace/file?filename=WA_Fn-UseC_-HR-Employee-Attrition.tsv.txt

v) *Data Cleaning*. Sisense. (n.d.). Retrieved May 8, 2022, from https://www.sisense.com/glossary/data-cleaning/

vi) Turp, M. (n.d.). YouTube. Retrieved May 8, 2022, from https://www.youtube.com/?watch

vii) Aruchamy, V. (2021, June 1). *Data Analytics with pandas – how to drop a list of rows from a pandas dataframe*. freeCodeCamp.org. Retrieved May 8, 2022, from

https://www.freecodecamp.org/news/drop-list-of-rows-from-pandas-dataframe/#:~:text=To%20drop%20a%20row%20or%20column%20in%20a%20dataframe%2C%20you,method%20in%20the%20docs%20here.&text=Rows%20are%20labelled%20using%20the,Columns%20are%20labelled%20using%20names.

viii) *GGPLOT2 histogram plot : Quick Start Guide - R Software and Data Visualization*. STHDA. (n.d.). Retrieved May 8, 2022, from http://www.sthda.com/english/wiki/ggplot2-histogram-plot-quick-start-guide-r-software-and-data-visualization

ix) SQL syntax. (n.d.). Retrieved May 8, 2022, from https://www.w3schools.com/sql/sql_syntax.asp