DePaul University

CAPSTONE PROJECT

Topic: IBM Attrition Data – HR Analytics

Using Microsoft Excel

By Divya Bhattiprolu

Introduction

Human Resource Analytics or HR Analytics refers to the analytical processes applied by the human resource department of an organization in the hope of improving employee performance and in turn getting a better return on investment. In the domain of human resources, attrition refers to a voluntary or involuntary reduction of a company’s workforce. This occurs when employees resign or retire and not replaced. It is a term used to describe the downsizing in an organization’s employees by the human resource department.

Predicting attrition, whether an employee will leave the job or not, has become an important concern for organizations in the recent days owing to many reasons. This project is about looking into those reasons by exploring the various factors. This would help the HR to intervene on time and handle the situation in a better way so as to prevent attrition. While this model can be routinely run to identify employees who are most likely to quit, the key driver of success would be the human element of reaching out the employee, understanding the current situation of the employee and taking action to remedy controllable factors that can prevent attrition of the employee.

About the Data

For this project, I am working on a dataset ‘IBM HR Analytics Employee Attrition & Performance’ from Kaggle. This dataset presents an employee survey from IBM, displaying if there is attrition or not. The dataset has 1471 rows and 33 columns with 51,450 observations. Given the limited size of the data set, the model should only be expected to provide modest improvement in identification of attrition vs a random allocation (any variable) of probability of attrition. It has numeric and categorical data types describing each employee’s characteristics pertaining to job and personal information. It has a column labelled attrition that tells if the employee is still working at the company or has already quit.

Data dictionary

Name	Description
Age	Age of Employee
Attrition	Employee leaving the company
Business Travel	Employee travel for business
DailyRate	Amount of money paid per day
Department	Employee works under which department
DistanceFromHome	Distance from work to home
Education	Education level of the employee
Education Field	Employee’s field of education
Employee Number	Employee ID
Environment Satisfaction	Satisfaction with the job environment
Gender	Gender
HourlyRate	Hourly Salary
JobInvolvement	How much is the employee involved in his job
JobLevel	Level of job
JobRole	Role of the employee in the company
JobSatisfaction	Satisfaction with the job

MaritalStatus	Employee is married or not
Monthly Income	Amount of money the employee earns monthly
Monthly Rate	Internal charge out rate which will be used to calculate the cost of each employee monthly, in general, the monthly rate will cover salary, social insurance, administration, logistics, over head etc.
NumCompaniesWorked	Number of companies employee worked at
Over18	Employee is above 18 years of age or no
OverTime	Employee works more than the designated hours
PercentSalaryHike	Percentage increase in salary
PerformanceRating	Employee’s rating based on his performance
RelationshipSatisfaction	Satisfaction level with the relationships at the job
StandardHours	Standard hours
TotalWorkingYears	Total years worked
TrainingTimesLastYear	Hours spent training
WorkLifeBalance	Time spent between work and outside
YearsAtCompany	Total number of years at the company
YearsInCurrentRole	Number of years in the current position

YearsSinceLastPromotion	Number of years since the employee is last promoted
YearsWithCurrManager	Number of years working with current manager

Explanation of the values in the dataset

Attrition

- No – 0
- Yes – 1

Business Travel

- Travel Rarely - 1
- Travel Frequently – 2
- Non-Travel – 3

Departments

- Human Resources
- Research and Development
- Sales

Education

- Below College – 1
- College – 2
- Bachelor – 3
- Master – 4
- Doctor – 5

Education Field

- HR
- Life Sciences
- Marketing
- Medical Sciences

- Others
- Technical

Environment Satisfaction

- Low – 1
- Medium – 2
- High – 3
- Very High – 4

Gender

- Male – 0
- Female – 1

Job Involvement

- Low – 1
- Medium – 2
- High – 3
- Very High – 4

Job Satisfaction

- Low – 1
- Medium – 2
- High – 3
- Very High – 4

Marital Status

- Divorced
- Married
- Single

Performance Rating

- Low – 1
- Good – 2
- Excellent – 3
- Outstanding – 4

Relationship Satisfaction

- Low – 1
- Medium – 2
- High – 3
- Very High – 4

Work Life Balance

- Bad – 1
- Good – 2
- Better – 3
- Best – 4

Hypotheses

Hypothesis 1: Higher the level of satisfaction, lower the chance of attrition.

Explanation: Attrition depends on the levels of satisfaction in various categories given by the employees in the survey

Hypothesis 2: Wages, Gender, Promotion and extra working hours affect the Attrition rate.

Explanation: The amount of money paid is a factor that would affect attrition along with the work done by the employees being recognized with either a promotion or salary hike. In addition to these factors, the amount of time spent by the employees working is another attribute that would make them consider leaving or staying in the company.

Hypothesis 3: Attrition rate does not depend on travelling locally or domestically.

Explanation: People do not tend to consider leaving a job because of the distance they have to travel to work nor because of travelling to other places for business purpose. This might be a factor to consider but there will not be much change in the decision of attrition with regard to travel.

Hypothesis 4: Average salary varies from department to department.

Explanation: The salaries of people do vary from one department to another department based on the skill set and the work they do. The productivity levels of the employee also depends on the department that he/she is working in and hence will exist a difference.

Descriptive Data

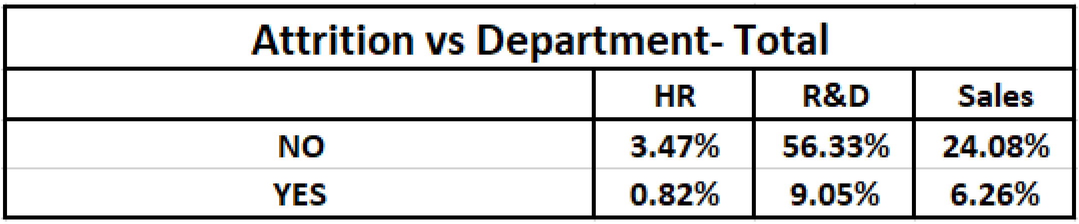
The first step I did was to clean the data of the attributes that I knew would not be needed for the analysis. The attribute ‘employee count’ is just assigned 1 to each employee and hence is of no use for the analysis of data. The attribute ‘stock option level’ is an attribute whose data description was not understood or found in the data dictionary and hence I removed this attribute as well.

Next, I wanted to check the attrition based on the departments – Human Resources, Research & Development and Sales. So, I segregated the data according to the employees tending to leave or stay in each department.

Part 1: Calculated the total number of entries and the number of ‘yes and no’ (attrition) to obtain the percentage of the attrition in each department.

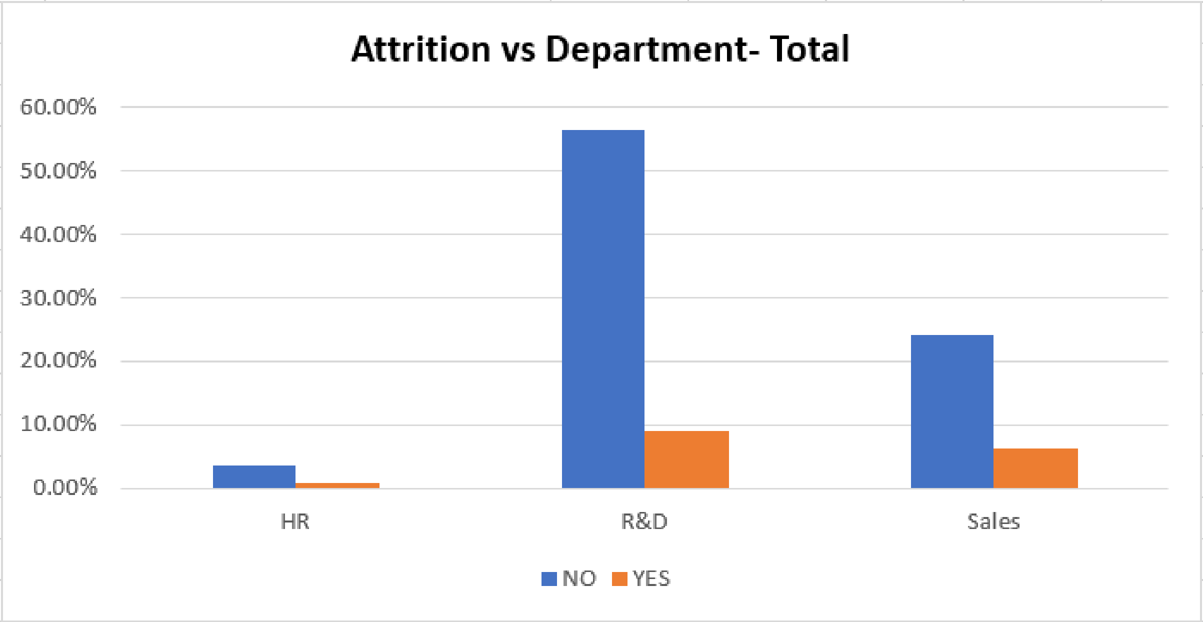
Part 2: Also, I calculated the total number of people in each department and the attrition. The results are as follows:

Part 1:



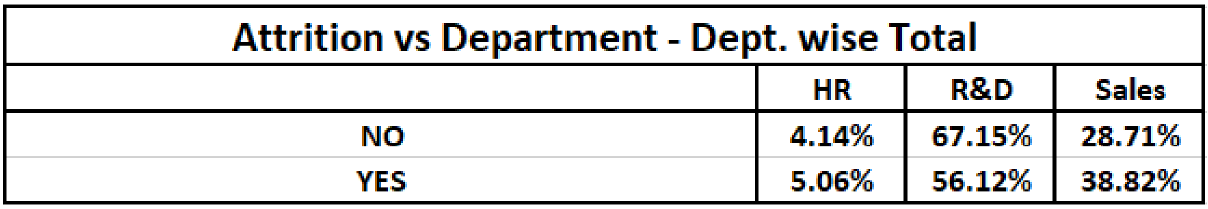
Attrition vs Department- Total

	HR	R&D	Sales
NO	3.47%	56.33%	24.08%
YES	0.82%	9.05%	6.26%

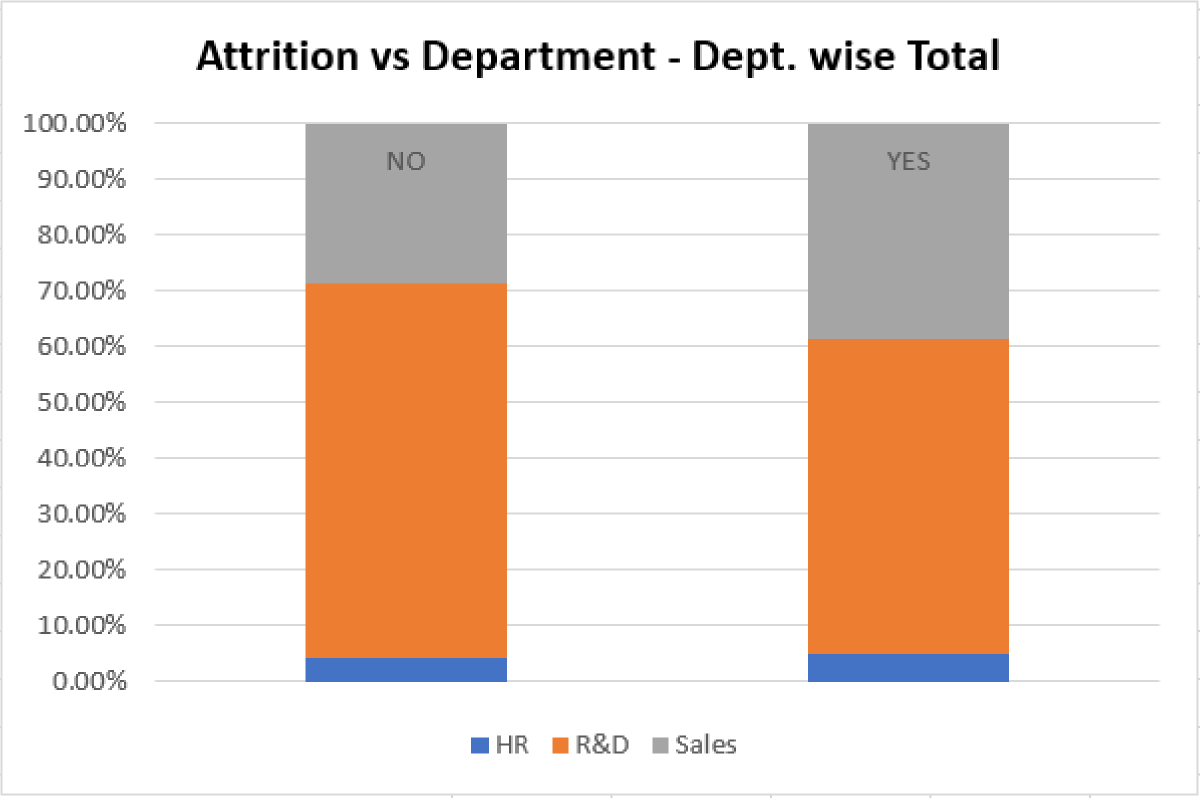


From the above result, it is seen that employees from R&D are more likely to leave the company, followed by Sales department. The HR department is of the least concern as the percentage is dismissible.

Part 2:

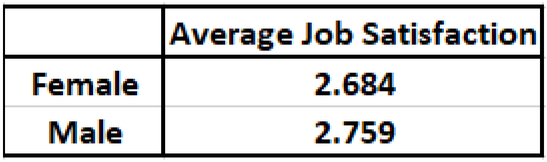


Attrition vs Department - Dept. wise Total			
	HR	R&D	Sales
NO	4.14%	67.15%	28.71%
YES	5.06%	56.12%	38.82%



But when we look at the attrition based on the total number of employees data calculated department wise, HR and Sales department seem to be of concern. As the percentage of employees that are likely to stay back is lower than the ones that are likely to leave. Hence, HR should concentrate on these departments in order to mend these values.

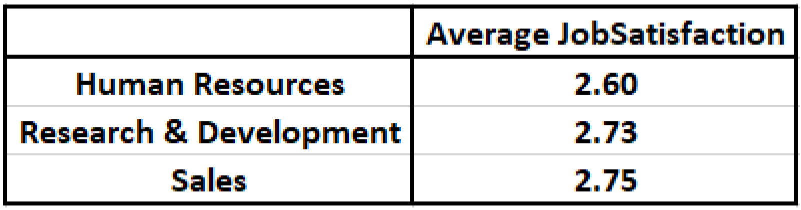
Then, comparing the average job satisfaction based on the gender I received the following results.



	Average Job Satisfaction
Female	2.684
Male	2.759

From the above table, we can see that the job satisfaction among the male employees is slightly higher than the job satisfaction among the female employees. Hence, the HR should maybe reach out to the female employees to understand the areas that are lacking due to which the job satisfaction is low.

Later, I compared the job satisfaction level based on the departments. The results were as follows:



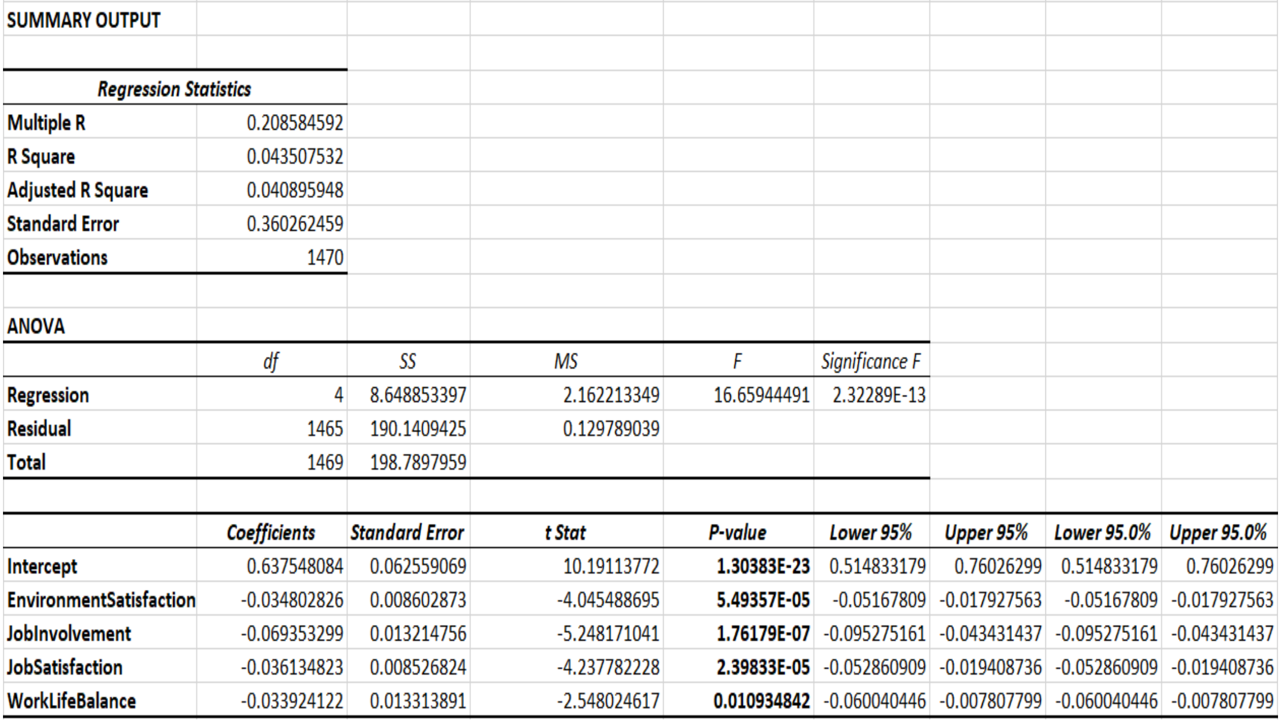
	Average JobSatisfaction
Human Resources	2.60
Research & Development	2.73
Sales	2.75

From the above table, it is seen that the average job satisfaction overall is almost the same. Sales department is doing very well in terms of employees enjoying their time with the work they are doing followed very closely by the Research & Development department. HR department employees though do not seem to be satisfied with the work they are doing when compared to the other two departments.

Analysis – Regression

Hypothesis 1: Higher the level of satisfaction, lower the chance of attrition.

Result:



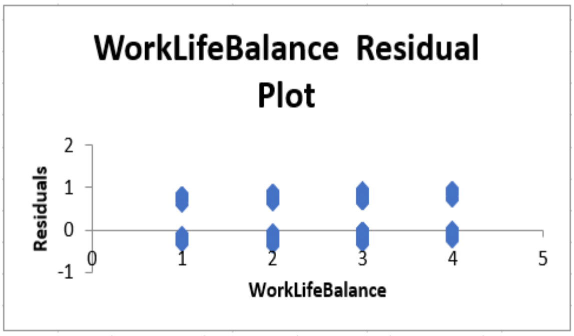
SUMMARY OUTPUT

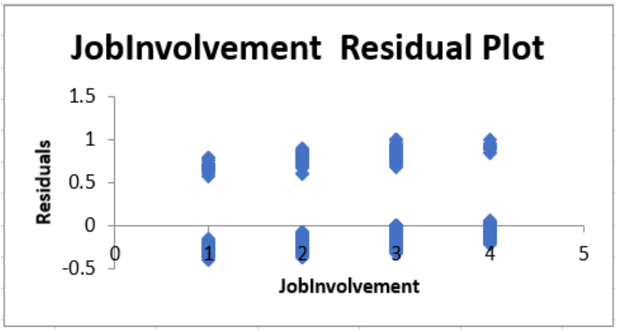
Regression Statistics	
Multiple R	0.208584592
R Square	0.043507532
Adjusted R Square	0.040895948
Standard Error	0.360262459
Observations	1470

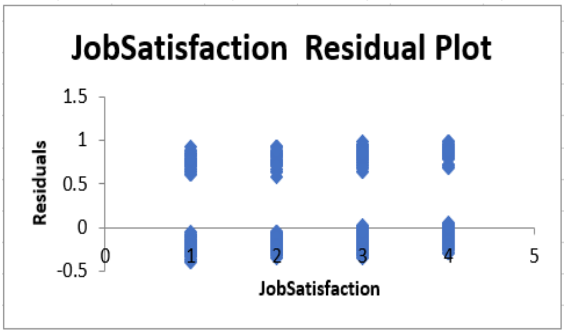
ANOVA

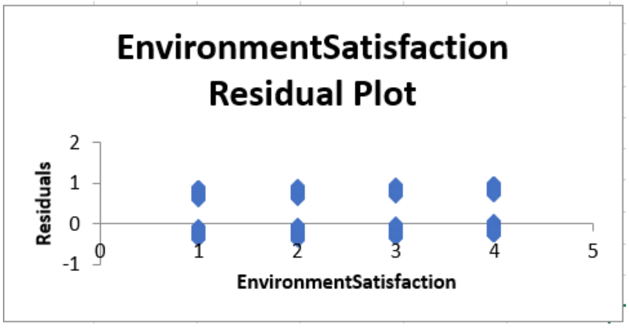
	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	8.648853397	2.162213349	16.65944491	2.32289E-13
Residual	1465	190.1409425	0.129789039		
Total	1469	198.7897959			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.637548084	0.062559069	10.19113772	**1.30383E-23**	0.514833179	0.76026299	0.514833179	0.76026299
EnvironmentSatisfaction	-0.034802826	0.008602873	-4.045488695	**5.49357E-05**	-0.05167809	-0.017927563	-0.05167809	-0.017927563
JobInvolvement	-0.069353299	0.013214756	-5.248171041	**1.76179E-07**	-0.095275161	-0.043431437	-0.095275161	-0.043431437
JobSatisfaction	-0.036134823	0.008526824	-4.237782228	**2.39833E-05**	-0.052860909	-0.019408736	-0.052860909	-0.019408736
WorkLifeBalance	-0.033924122	0.013313891	-2.548024617	**0.010934842**	-0.060040446	-0.007807799	-0.060040446	-0.007807799









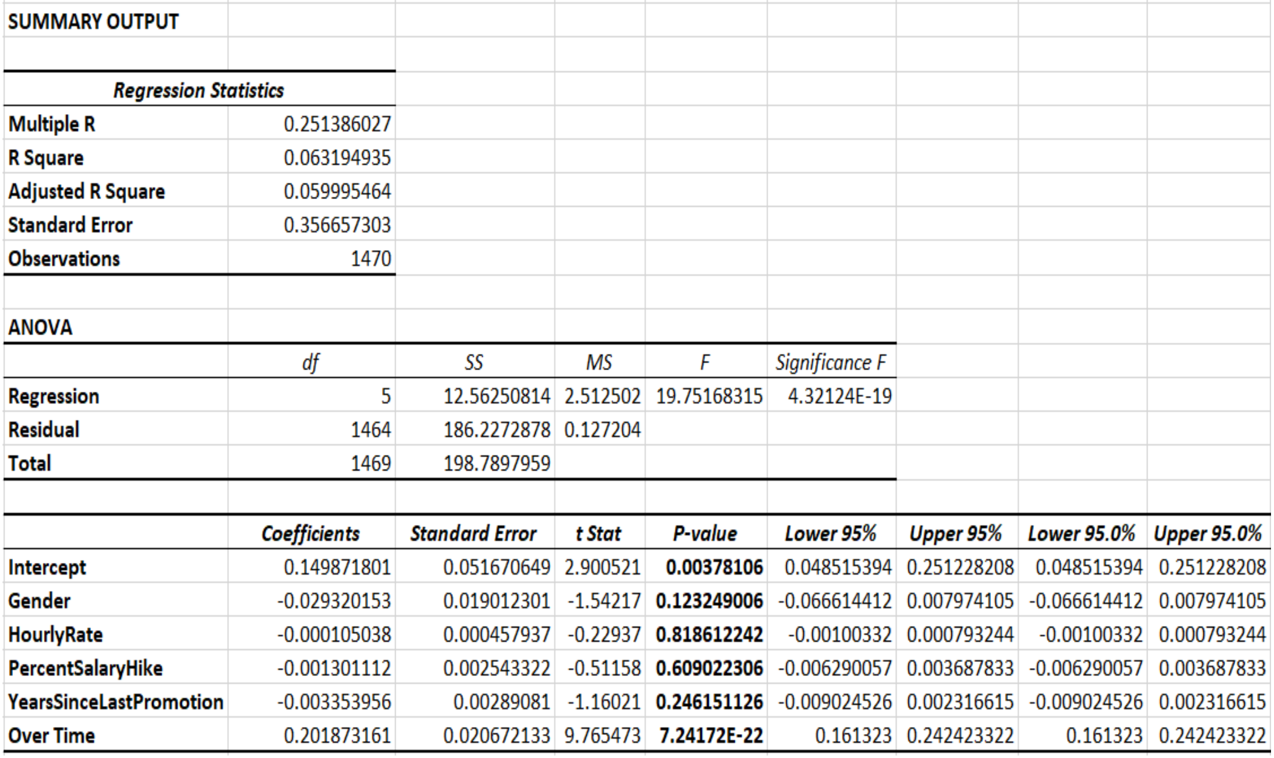
Attrition = 1.30E-23 + 5.49E-05 * EnvironmentSatisfaction + 1.76E-07 * JobInvolvement + 2.39E-05 * JobSatisfaction + 0.01 * WorkLifeBalance + E

For every 5.49E-05 units, 1.76E-07 units, 2.39E-05 units, 0.01 units change in Environment Satisfaction, JobInvolvement, Job Satisfaction and WorkLife Balance respectively, there will be one unit change in the attrition.

The above results show us that the satisfaction levels of each factor – Work life balance, Job involvement, Job satisfaction and environment satisfaction are the attributes that are significant enough to affect the decision of an employee to either stay with or leave an organization. Or simply put, these factors play a major role in an employee deciding whether to be with a company or no. Hence, we accept the hypothesis.

Hypothesis 2: Wages, Gender, Promotion and extra working hours affect the Attrition rate.

Result:



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.251386027
R Square	0.063194935
Adjusted R Square	0.059995464
Standard Error	0.356657303
Observations	1470

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	12.56250814	2.512502	19.75168315	4.32124E-19
Residual	1464	186.2272878	0.127204		
Total	1469	198.7897959			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.149871801	0.051670649	2.900521	**0.00378106**	0.048515394	0.251228208	0.048515394	0.251228208
Gender	-0.029320153	0.019012301	-1.54217	**0.123249006**	-0.066614412	0.007974105	-0.066614412	0.007974105
HourlyRate	-0.000105038	0.000457937	-0.22937	**0.818612242**	-0.00100332	0.000793244	-0.00100332	0.000793244
PercentSalaryHike	-0.001301112	0.002543322	-0.51158	**0.609022306**	-0.006290057	0.003687833	-0.006290057	0.003687833
YearsSinceLastPromotion	-0.003353956	0.00289081	-1.16021	**0.246151126**	-0.009024526	0.002316615	-0.009024526	0.002316615
Over Time	0.201873161	0.020672133	9.765473	**7.24172E-22**	0.161323	0.242423322	0.161323	0.242423322

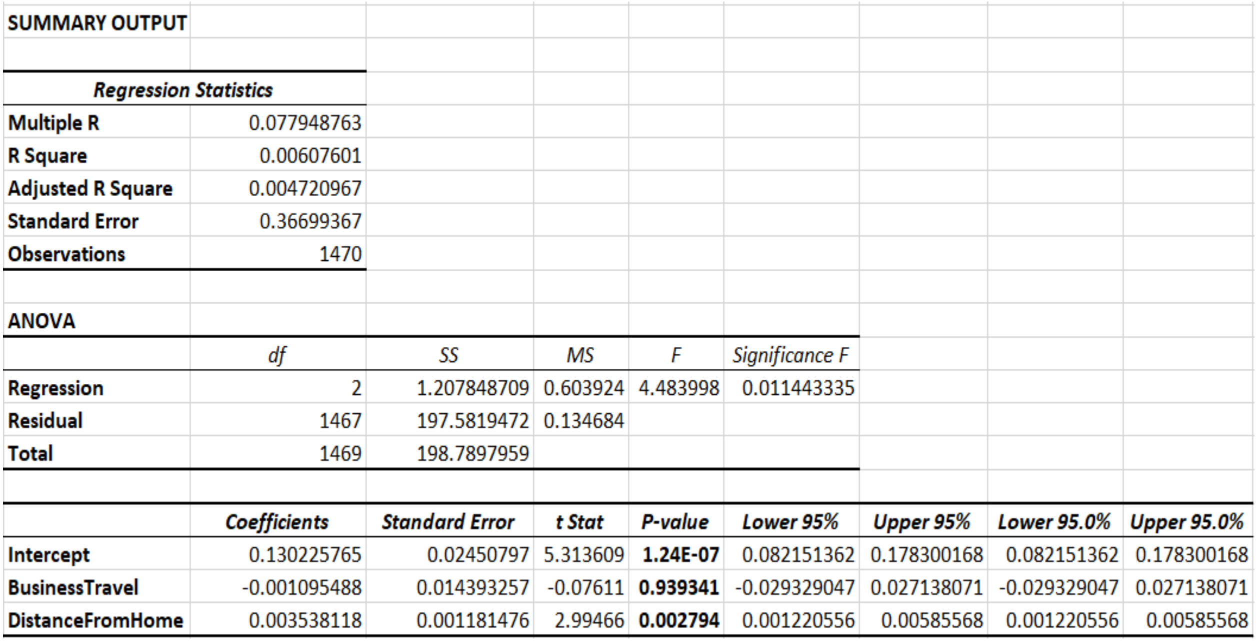
Attrition = 0.003 + 0.12 * Gender + 0.81 * HourlyRate + 0.60 * PercentSalaryHike + 0.24 * YearsSinceLastPromotion + 7.24E-22 * OverTime + E

For every 0.12 units in Gender, 0.81 units in HourlyRate, 0.60 units in PercentSalaryHike, 0.24 units YearsSinceLastPromotion and 7.24E-22 units change in OverTime, there will be one unit change in the Attrition.

Irrespective of the contrary belief that money and recognition are the only main factors for a person to continue in a job, we can see that these factors do not affect the attrition as much as we expect it to. There are many other factors that come into the lime light when an employee considers being in a job or leaving it. But, the factor of working extra hours than that designated to an employee will matter as a factor in case of attrition. So, we partially accept the hypothesis.

Hypothesis 3: Attrition rate does not depend on travelling locally or domestically.

Result:



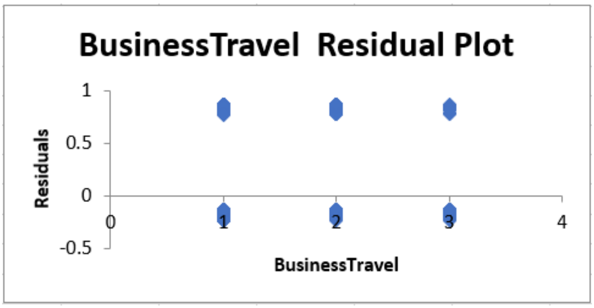
SUMMARY OUTPUT

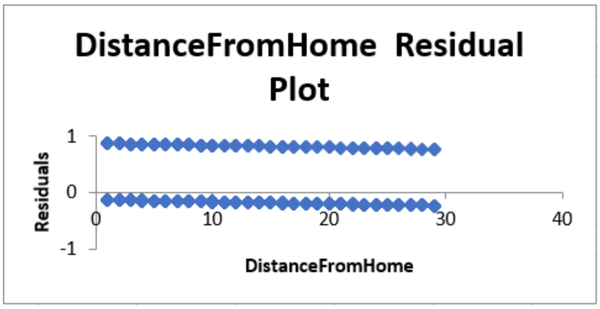
Regression Statistics	
Multiple R	0.077948763
R Square	0.00607601
Adjusted R Square	0.004720967
Standard Error	0.36699367
Observations	1470

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1.207848709	0.603924	4.483998	0.011443335
Residual	1467	197.5819472	0.134684		
Total	1469	198.7897959			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.130225765	0.02450797	5.313609	**1.24E-07**	0.082151362	0.178300168	0.082151362	0.178300168
BusinessTravel	-0.001095488	0.014393257	-0.07611	**0.939341**	-0.029329047	0.027138071	-0.029329047	0.027138071
DistanceFromHome	0.003538118	0.001181476	2.99466	**0.002794**	0.001220556	0.00585568	0.001220556	0.00585568





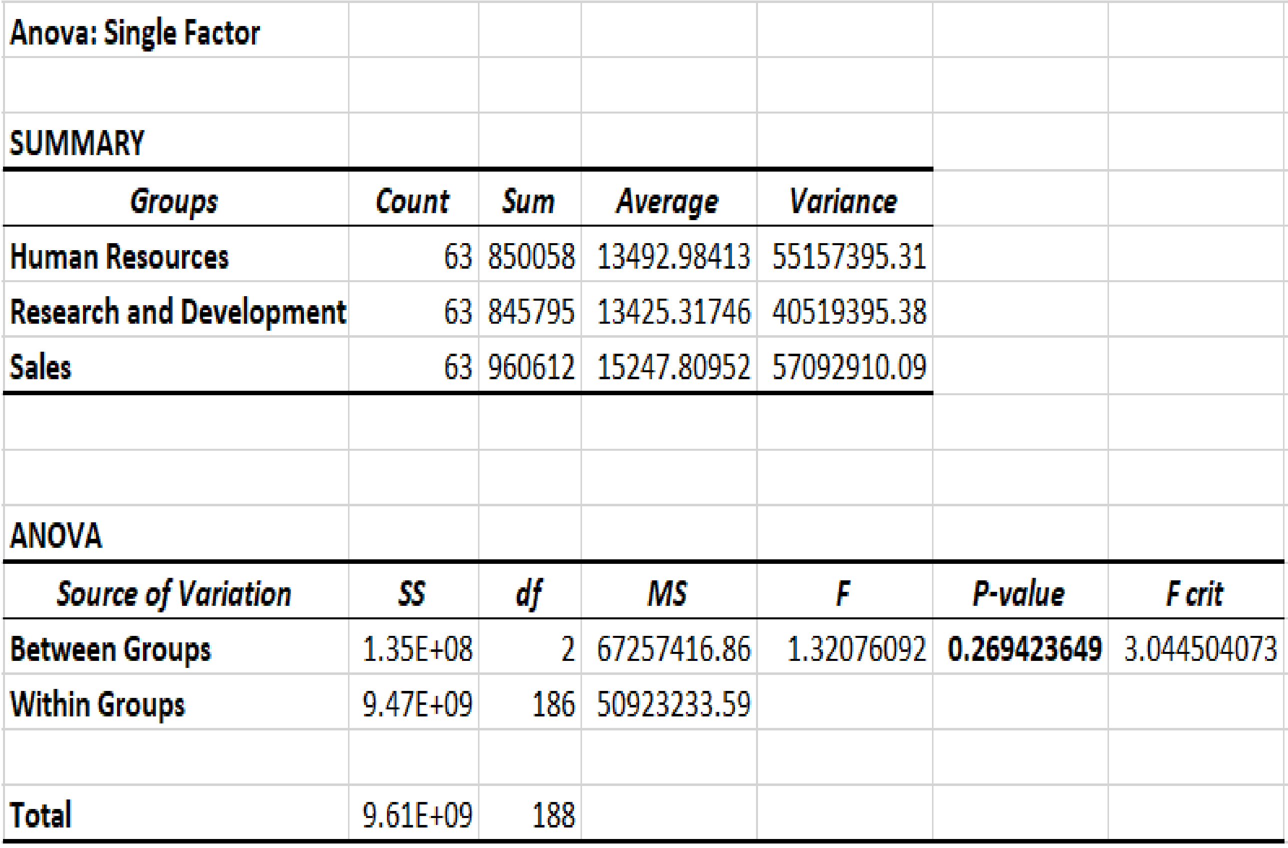
Attrition = 1.24E-07 + 0.93 * BusinessTravel + 0.001 * DistanceFromHome

For every 0.93 units and 0.001 units increase in the Business Travel and DistanceFromHome variables there will be one unit change in the attrition.

From the above results, we can see that travelling for business is something that employees do not mind doing but the distance from the work place to home will be of concern while considering leaving a company. Hence, we reject the hypothesis.

Hypothesis 4: Average salary varies from department to department.

Result:



Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Human Resources	63	850058	13492.98413	55157395.31		
Research and Development	63	845795	13425.31746	40519395.38		
Sales	63	960612	15247.80952	57092910.09		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.35E+08	2	67257416.86	1.32076092	**0.269423649**	3.044504073
Within Groups	9.47E+09	186	50923233.59			
Total	9.61E+09	188				

We can see from above that the average price does not vary much between the HR and R&D departments but there is a significant increase in the average price of the Sales department. Since the p-value is more than 0.05, we reject the hypothesis.

Conclusion

Minimizing the level of attrition and being prepared for instances that cannot be helped will significantly improve the operations of most businesses. As a future development, with a sufficiently large data set, it would be used to run a segmentation on employees, to develop certain “at risk” categories of employees. This could generate new insights for the business on what drives attrition, insights that cannot be generated by merely informational interviews with employees.

Works Cited

Dataset: https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset