BTA 350 ANALYTICS PROJECT

IBM HR ANALYTICS EMPLOYEE ATTRITION

RICKY TRUONG

974061616

6/2/2022

PROJECT SECTION 1

Dataset: IBM HR Analytics Employee Attrition & Performance

Link: IBM HR Analytics Employee Attrition & Performance

Schedule:

- **Week 1 (3/28 – 4-3)**
  - Look for datasets on Kaggle.com
  - Find a dataset that interests you and email Marta about it for approval
  - Explore and familiarize the selected dataset in excel file
  - Complete Project Section 1 Writeup
- **Week 2 (4/4 – 4/10)**
  - Create a data dictionary to define what each column means
  - Create pivot tables and charts to explore patterns/trend of dataset
  - Complete Project Section 2 Writeup
- **Week 3 (4/11 – 4/17)**
  - Complete Project Section 3 Writeup
- **Week 4 (4/18 – 4/24)**
  - Complete Project Section 4 Writeup
- **Week 5 (4/25 – 5-1)**
  - Complete Project Section 5 Writeup
  - Continue analyzing dataset
- **Week 6 (5/2 – 5/8)**
  - Complete Project Section 6
  - Revise Project based on Marta's Feedback
- **Week 7 (5/9 – 5/15)**
  - Complete Project Section 7
  - Revise Project based on Marta's Feedback
- **Week 8 (5/16 – 5/22)**
  - Complete Project Section 8
- **Week 9 (5/23 – 5-29)**
  - Complete Project Section 9
- **Week 10 (5/30 – 6/5)**
  - Work on Presentation

PROJECT SECTION 2: CONTEXT

The dataset I selected for my project is about employee attrition at IBM. This dataset was obtained through Kaggle.com and is a fictional dataset created by IBM data scientists. Although this dataset is fictional, its great practice for solving real business problems, such as attrition, which is another word for turnover rate, and a real problem in some companies.

Since this dataset revolves around attrition, my potential business problem will most likely revolve around employee attrition. Employee attrition can be impactful in any company. In this scenario, attrition mainly impacts the HR department at IBM. Since employees are leaving, the HR department is responsible for finding replacements for the unfilled positions. While searching for replacements, HR is held accountable for analyzing why former employees leave? This is the part I will focus on, the why part of employees leaving IBM.

To tackle this dataset, I will be using the DATA analytics framework methodology from our textbook:

1. **Decisions:**
   The first part of the DATA analytics framework is to decide what I intend to make based on the results I may find throughout this project. In this case, I would like to know how to reduce employee attrition based on the leading causes I may find.

2. **Acquisition:**
   After decisions, the next part would be determining what data I need to collect to help inform the decisions I plan to make. Once I know what decisions I'm making, I'll have a better idea of the type of data required to create my decisions. In this case, I would collect data that correlate to attrition, such as job satisfaction, job role, hourly rate, monthly income, work-life balance, age, gender, etc.

3. **Time:**
   The third step of the DATA analytics framework is to determine how long I need to collect and analyze my data to support my decision. Since this dataset has everything I needed, I would be analyzing it from week 2 to week 6.

4. **Analysis**:
   The last step of this methodology is to determine how I will analyze this dataset for the next four weeks. Based on the lecture on the type of analytics, I would be mostly leaning towards conducting visual analytics due to the size of the dataset. I will be creating visuals/graphs from pivot tables to spot any patterns or trends that may correlate to attrition.

PROJECT SECTION 3: DEFINING THE PROBLEM

1. **Problem Definition:** What are the main factors that cause IBM employees to attrite?

   Attrition is best described when employees leave an organization for multiple reasons such as termination, resignation, retirement, or death. The impact of attrition can be a severe problem for IBM if there's a high rate of employees leaving, resulting in losing customers/sales. They're numerous reasons employees may leave a company, such as strict company policies, the environment, the relationship with management, low-performance level, low pay, no benefits, etc.

2. **Question Definition:** Which department are most heavily affected by attrition?

   Since attrition can impact IBM's performance and create challenges for the HR department and management, I will use exploratory and inferential methods to find the leading causes of attrition. My hypothesis is why employees leave may relate to the relationship with management, job satisfaction, and workload. Figuring out why attrition happens and how to satisfy employees based on their needs should make them stay in IBM longer.

3. **Business Relevance:** IBM is known for producing and selling computer software/hardware products to their clients and providing consulting, cloud computing, data analytics, and many more. Attrition acts like a chain reaction to companies like IBM. If they're fewer employees in IBM due to them leaving, the efficiency of producing computer parts will decrease, and services will be delayed resulting in unhappy clients. If clients are dissatisfied with IBM's services, they're most likely to leave. Unhappy clients lead to decreased sales and customer loyalty. Worst case scenario, if this continues long enough, IBM will go bankrupt.

PROJECT SECTION 4: DATA

The data type for my dataset is structured data because it was already organized into rows and columns. Since this dataset revolves around current and former employees, the data source is internal data because the data was collected within the company and owned by the company. If this dataset weren't owned by IBM and were collected and received outside the organization, it would've been external data. Overall, this dataset should help me answer the question definition I stated previously. Some of the variables/fields look promising to analyze as the whole dataset revolves around HR analytics.

During the exploratory data analysis stage, the first thing I did was define the field names (columns) in a data dictionary (table below) to understand the dataset better. I can always refer to this table if I don't know or have forgotten what some of the field names mean. In the table I created, the field name column lists out all the column names in the dataset. The data type column lists whether the field name is a number or text data. The description column defines the field names. To summarize my data dictionary table, there are 35 field names; 24 fields consist of number data, and 11 fields consist of text data.

After creating my data dictionary, the next step was conducting summary statistics (the next image after the table) on some of the variables I'm interested in that correlate to attrition. Out of curiosity, I picked the following variables: years at the company, years with the current manager, years since the last promotion, the satisfaction levels (1 = low, 4= very high), the employees' age, and hourly rate (wage). I picked these specific variables to help me understand the employee data I'm working with. Understanding how long employees spent working at IBM plus their satisfaction level is a step closer to answering my question. For example, the three satisfaction summaries on my summary statistic, job environment, and relationship have an average rating of 2.70: on a scale of 1-4, 1 being low satisfaction and 4 being very high satisfaction. This indicates that employees have an average experience working at IBM.

Field Name	Data Type	Description
Age	Number	How old the employees are
Attrition	Text	The departure of employees from the organization for any reason (voluntary or involuntary): resignation, termination, retirement or death
BusinessTravel	Text	How often do employees travel for business trips
DailyRate	Number	the amount of money a company has to pay their employees to work for them for a day **($)**
Department	Text	What department are the employees from
DistanceFromHome	Number	The distance from the employees' home to the workplace (commuting)
Education	Text	The education level the employee received
EducationField	Text	The education the employee specialized in
EmployeeCount	Number	The bodycount of the employee **(unnecessary for the dataset)**
EmployeeNumber	Number	Employee ID **(unnecessary for the dataset)**
EnvironmentSatisfaction	Number	How satisfied are the employees based on work environment?
Gender	Text	Is the employee male or female
HourlyRate	Number	The amount paid to an employee for each hour worked **($)**
JobInvolvement	Number	The psychological and emotional extent to which someone participates in his/her work, profession, and company
JobLevel	Number	Categories of authority in an organization
JobRole	Text	The employee's job title
JobSatisfaction	Text	How satisfied are the employees based on their position?
MartialStatus	Text	The employee's state of being single, married, or divorced
MonthlyIncome	Number	The amount of income the employee earns in one month **($)**
MonthlyRate	Number	The internal charge out rate which will be used to calculate the cost of each employee monthly, in general, the monthly rate will cover salary, social insurance, administration, logistics, over head etc **($)**
NumCompaniesWorked	Number	The number of companies the employee worked in previous years
Over18	Text	Is the employee over the age of 18? **(unnecessary for the dataset)**
OverTime	Text	Any hours worked by an employee that exceed their normally scheduled working hours
PercentSalaryHike	Number	the amount a salary is increased (ex: raise)
PerformanceRating	Number	The step in the work measurement in which an analyst observes the employee's performance and records a value representing that performance
RelationshipSatisfaction	Number	How well do the employees get along with others
StandardHours	Number	The number of hours defined for a normal workweek (Ex: 40 hrs a week is typical)
StockOptionLevel	Number	stock options in the company's stock granted by an employer to certain employees
TotalWorkingYears	Number	The number of work experience before working at IBM (Ex: person A has 5 years of work experience)
TrainingTimesLastYear	Number	The number of training sessions the employee received last year
WorkLifeBalance	Number	An equilibrium state, where one effectively balances work or career demands and those of their personal life
YearsAtCompany	Number	How long has the employee been working at the company?
YearsInCurrentRole	Number	How long has the employee been working in their position?
YearsSinceLastPromotion	Number	When was the last time the employee had a promotion?
YearsWithCurrManager	Number	How long has the employee been working under their current manager?
Total Columns: 35		

YearsAtCompany	
Mean	7
Standard Error	0
Median	5
Mode	5
Standard Deviation	6.13
Sample Variance	37.53
Kurtosis	3.94
Skewness	1.76
Range	40
Minimum	0
Maximum	40
Sum	10,302
Count	1,470
Largest(1)	40
Smallest(1)	0

YearsWithCurrManager	
Mean	4
Standard Error	0
Median	3
Mode	2
Standard Deviation	3.57
Sample Variance	12.73
Kurtosis	0.17
Skewness	0.83
Range	17
Minimum	0
Maximum	17
Sum	6,061
Count	1,470
Largest(1)	17
Smallest(1)	0

YearsSinceLastPromotion	
Mean	2
Standard Error	0
Median	1
Mode	0
Standard Deviation	3.22
Sample Variance	10.38
Kurtosis	3.61
Skewness	1.98
Range	15
Minimum	0
Maximum	15
Sum	3,216
Count	1,470
Largest(1)	15
Smallest(1)	0

Age	
Mean	37
Standard Error	0
Median	36
Mode	35
Standard Deviation	9.14
Sample Variance	83.46
Kurtosis	-0.40
Skewness	0.41
Range	42
Minimum	18
Maximum	60
Sum	54,278
Count	1,470
Largest(1)	60
Smallest(1)	18

JobSatisfaction	
Mean	2.73
Standard Error	0.029
Median	3
Mode	4
Standard Deviation	1.10
Sample Variance	1.22
Kurtosis	-1.22
Skewness	-0.33
Range	3
Minimum	1
Maximum	4
Sum	4,011
Count	1,470
Largest(1)	4
Smallest(1)	1

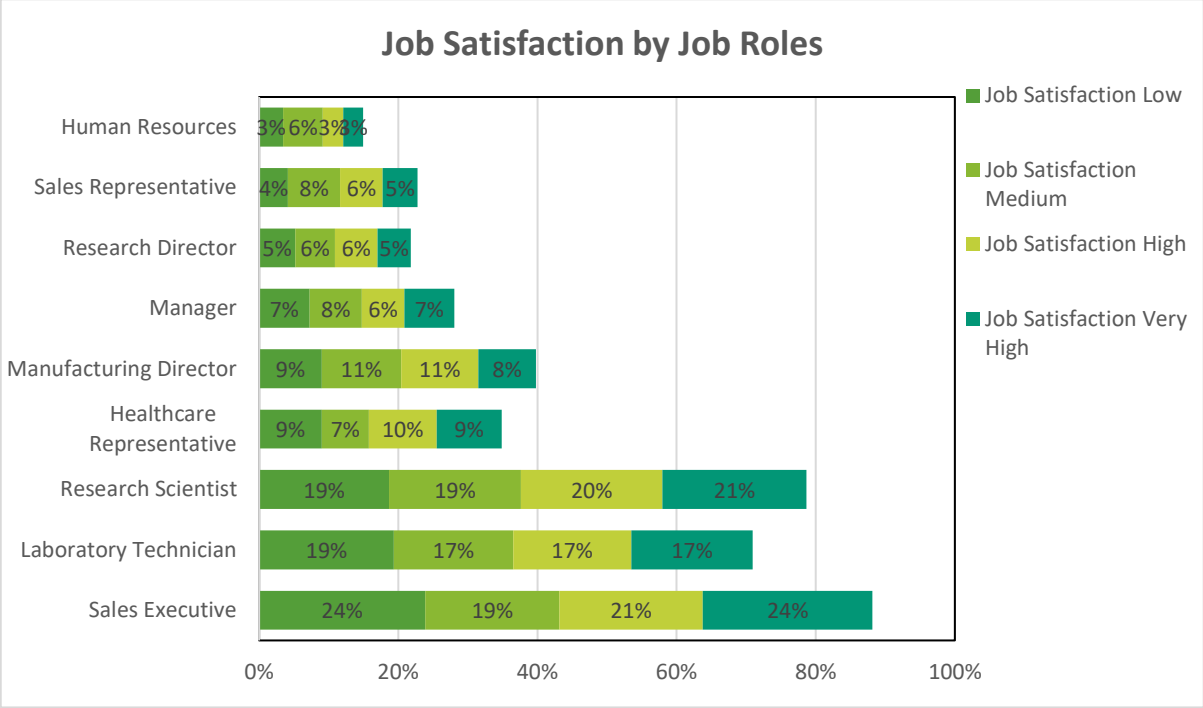
EnvironmentSatisfaction	
Mean	2.72
Standard Error	0.029
Median	3
Mode	3
Standard Deviation	1.09
Sample Variance	1.19
Kurtosis	-1.20
Skewness	-0.32
Range	3
Minimum	1
Maximum	4
Sum	4,001
Count	1,470
Largest(1)	4
Smallest(1)	1

RelationshipSatisfaction	
Mean	2.71
Standard Error	0.028
Median	3
Mode	3
Standard Deviation	1.08
Sample Variance	1.17
Kurtosis	-1.18
Skewness	-0.30
Range	3
Minimum	1
Maximum	4
Sum	3,987
Count	1,470
Largest(1)	4
Smallest(1)	1

HourlyRate	
Mean	65.89
Standard Error	0.53
Median	66
Mode	66
Standard Deviation	20.33
Sample Variance	413.29
Kurtosis	-1.20
Skewness	-0.03
Range	70
Minimum	30
Maximum	100
Sum	96,860
Count	1,470
Largest(1)	100
Smallest(1)	30

PROJECT SECTION 5: ANALYSIS

The outcome variable for my dataset would be attrition, whether an employee leaves or not. My outcome variable correlates to my problem and question because I'm looking for trends or patterns that cause attrition. In short, I'm focusing on what could potentially be the leading cause of attrition.



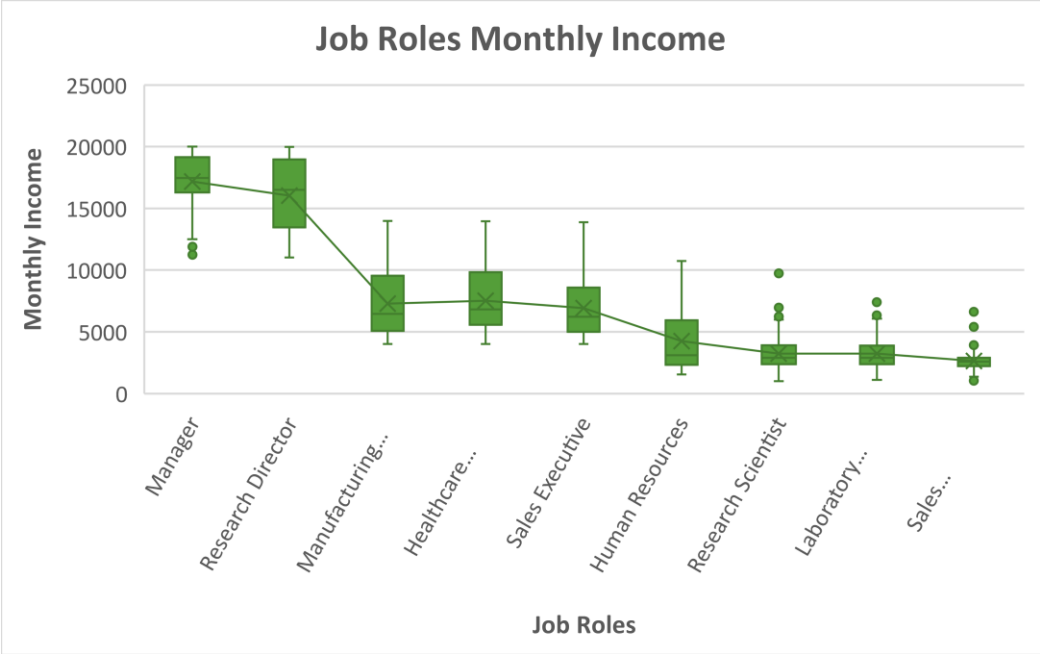
	Attrition rate	
Job Roles	**No**	**Yes**
Laboratory Technician	16%	26%
Sales Executive	22%	24%
Research Scientist	20%	20%
Sales Representative	4%	14%
Human Resources	3%	5%
Manufacturing Director	11%	4%
Healthcare Representative	10%	4%
Manager	8%	2%
Research Director	6%	1%

	Job Satisfaction			
Job Role	**Low**	**Medium**	**High**	**Very High**
Sales Executive	24%	19%	21%	24%
Laboratory Technician	19%	17%	17%	17%
Research Scientist	19%	19%	20%	21%
Healthcare Representative	9%	7%	10%	9%
Manufacturing Director	9%	11%	11%	8%
Manager	7%	8%	6%	7%
Research Director	5%	6%	6%	5%
Sales Representative	4%	8%	6%	5%
Human Resources	3%	6%	3%	3%

Count of Attrition	**Attrition**	
Job Roles	**No**	**Yes**
Laboratory Technician	197	62
Sales Executive	269	57
Research Scientist	245	47
Sales Representative	50	33
Human Resources	40	12
Manufacturing Director	135	10
Healthcare Representative	122	9
Manager	97	5
Research Director	78	2
Grand Total	**1233**	**237**

Job Satisfaction by Job Roles	**Job Satisfaction**			
Job Role	**Low**	**Medium**	**High**	**Very High**
Sales Executive	69	54	91	112
Laboratory Technician	56	48	75	80
Research Scientist	54	53	90	95
Healthcare Representative	26	19	43	43
Manufacturing Director	26	32	49	38
Manager	21	21	27	33
Research Director	15	16	27	22
Sales Representative	12	21	27	23
Human Resources	10	16	13	13
Grand Total	**289**	**280**	**442**	**459**

The first model I conducted was comparing the job satisfaction levels amongst the job roles in IBM (stacked bar chart). The stacked bar graph shows that sales executives have the lowest job satisfaction (24%), second by laboratory technicians (19%) and research scientists (19%). The two variables I selected to find the correlation in the bar chart were job roles and job satisfaction; created from a pivot table. To further detail, I wanted to confirm if low job satisfaction correlates with attrition, so I created another pivot table with job roles and attrition, and the results were nearly identical. The laboratory technicians had the highest attrition count, sales executives were the second-highest, and research scientists were third. Now that low job satisfaction is one of the leading causes of attrition, I wanted to know the why part of being unsatisfactory?



Count of Attrition	OverTime	
Job Roles	No	Yes
Research Scientist	195	97
Sales Executive	232	94
Laboratory Technician	197	62
Manufacturing Director	106	39
Healthcare Representative	94	37
Manager	75	27
Sales Representative	59	24
Research Director	57	23
Human Resources	39	13
Grand Total	**1054**	**416**

Count of WorkLifeBalance	Work Life Balan			
Job Roles	Bad	Good	Better	Best
Laboratory Technician	20	58	156	25
Research Scientist	16	86	166	24
Sales Executive	12	76	202	36
Healthcare Representative	10	30	80	11
Manufacturing Director	7	34	90	14
Manager	6	23	61	12
Research Director	4	15	49	12
Human Resources	4	6	32	10
Sales Representative	1	16	57	9
Grand Total	**80**	**344**	**893**	**153**

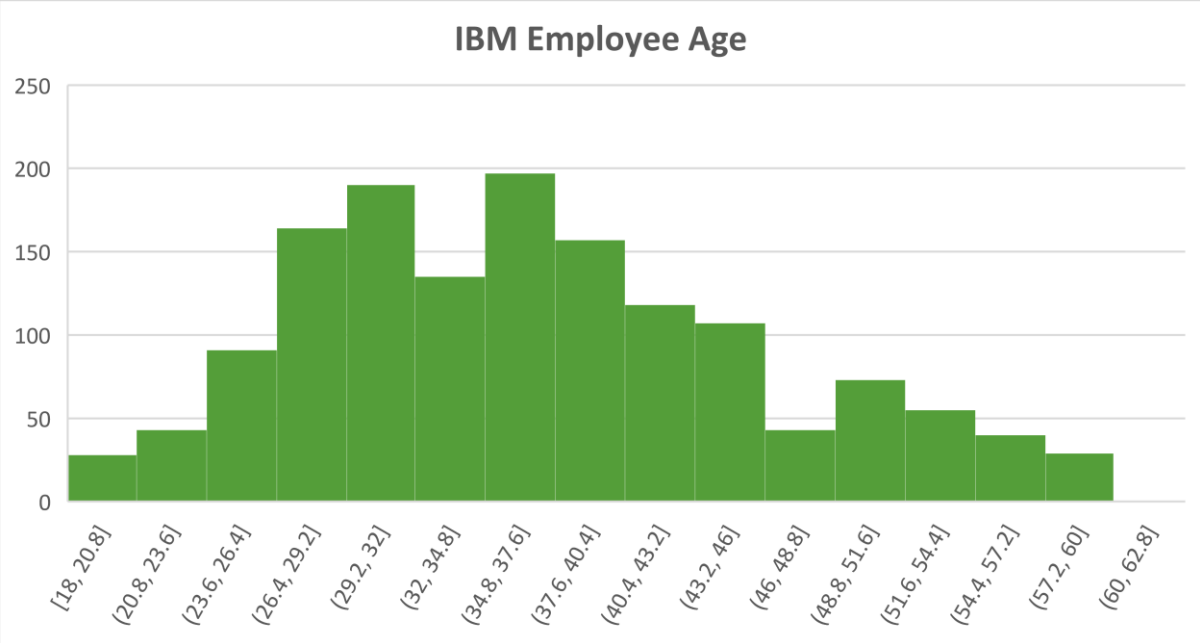
	OverTime	
Job Roles	**No**	**Yes**
Research Scientist	19%	23%
Sales Executive	22%	23%
Laboratory Technician	19%	15%
Manufacturing Director	10%	9%
Healthcare Representative	9%	9%
Manager	7%	6%
Sales Representative	6%	6%
Research Director	5%	6%
Human Resources	4%	3%

	Work Life Balance			
Job Role	**Bad**	**Good**	**Better**	**Best**
Laboratory Technician	25%	17%	17%	16%
Research Scientist	20%	25%	19%	16%
Sales Executive	15%	22%	23%	24%
Healthcare Representative	13%	9%	9%	7%
Manufacturing Director	9%	10%	10%	9%
Manager	8%	7%	7%	8%
Research Director	5%	4%	5%	8%
Human Resources	5%	2%	4%	7%
Sales Representative	1%	5%	6%	6%

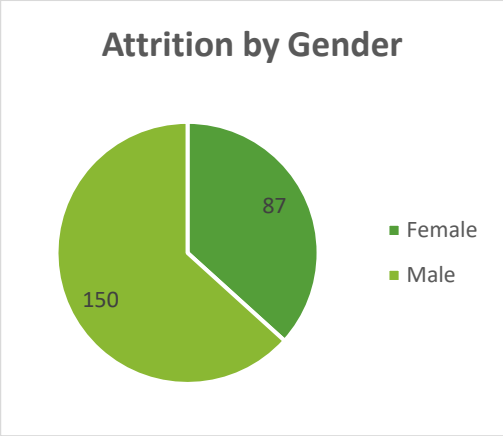
One of the interesting findings that may correlate to unsatisfied positions was the monthly income. The two variables I used for my box plot were monthly income and job roles. Based on the box plot I created, Research Scientists and Laboratory Technicians were one of the bottom three least paid positions; Sales Representatives were the least. Sales Executives was surprisingly in the middle of the chart, meaning its an average paid position. After creating the boxplot to get an overview of the monthly income per position, I can see the correlation between being unsatisfied due to being underpaid. But at the same time, being unhappy with your pay can't be the only reason leading to attrition, so I dug deeper. According to the two pivot tables below the box plot, the top three highlighted job roles have the most overtime correlating to having the worst work-life balance. Working overtime plus having difficulty balancing work-life makes sense in leaving IBM.

	Attrition			
	No		Yes	
Job Level	Count of Attrition	Average of YearsAtCompany	Count of Attrition	Average of YearsAtCompany
1	400	4	143	3
2	482	7	52	5
3	186	9	32	10
4	101	14	5	16
5	64	15	5	23

Besides being unsatisfactory with your position, there's a correlation between attrition with job levels. Based on the pivot table, job level 1 (entry-level positions) seems to have the highest count of attrition, averaging staying at the company for three years. There are multiple reasons why this happens; one of the main reasons is that the employee could be at a younger age who recently graduated from college and is job-hopping. They're job-hopping due to either advancing through their careers or may leave the company due to the causes I found in previous charts. The main takeaway from this pivot table is that higher job levels are less likely to leave IBM compared to lower job levels. The higher job levels are occupied by mostly older employees with more experience than lower job levels.



Job Roles	Average of Age
Manager	47
Research Director	44
Healthcare Representative	40
Manufacturing Director	38
Sales Executive	37
Human Resources	36
Research Scientist	34
Laboratory Technician	34
Sales Representative	30

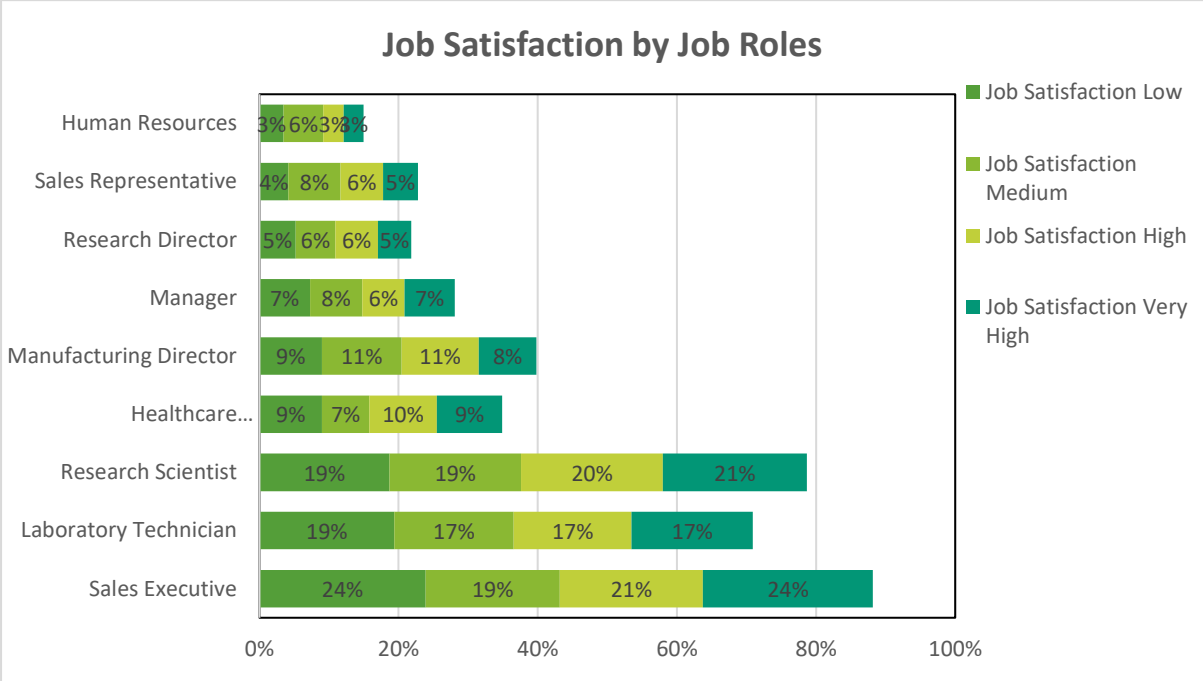


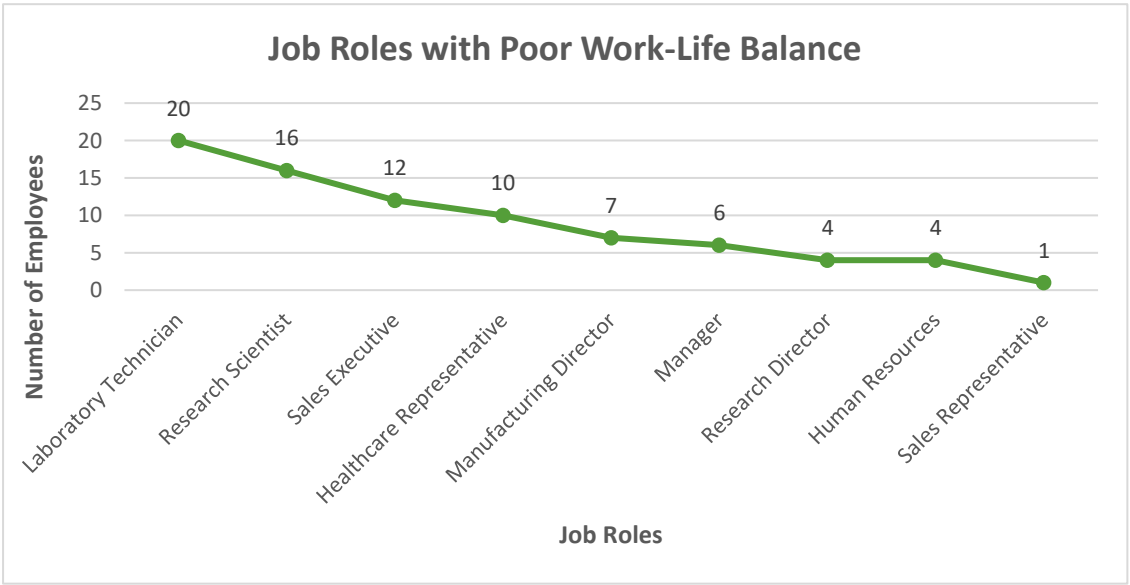
	Attrition rate	
Gender	No	Yes
Female	41%	37%
Male	59%	63%

In addition to age, I decided to create a histogram to get an overview of IBM's demographics. The youngest employee to work is 18, and the oldest is 60. Based on the histogram, employees between mid-twenties and thirties make up the age average. This concludes that they're more IBM employees in their 20s-30s than employees in their 40s-60s. The pivot table below the histogram showcases the average age for each job role to look further into age. Comparing the relationships in previous charts I mentioned, attrition occurs more in younger employees than older employees. In addition to attrition, the pie chart showcases that more male employees are affected by attrition than female employees; 63% for males vs. 37% for females.

PROJECT SECTION 6: EVALUATION

1. **Cofounding:**
   The stacked bar chart and line chart have an apparent relationship between job roles and the satisfaction level and work-life balance. As I mentioned before, the top 3 job roles with the lowest satisfaction levels plus poor work-life balance are sales executives, laboratory technicians, and research scientists. The sales executives have the lowest out of them all due to a heavy workload from responsibilities of being at the top of the pecking order. The higher your position, the more responsibilities you need to take, resulting in increased stress and a poor work-life balance. But while looking at the line chart (created from the table in project section 5), laboratory technicians have the lowest work-life balance of all job roles. This may be because, like being in the highest position, an entry-level position requires more responsibilities from the higher-ups like management. An entry-level position like a laboratory technician requires heavy assigned tasks from management, resulting in long hours and poor work-life balance and stress.





Job Satisfaction by Job Roles	Job Satisfaction			
Job Role	Low	Medium	High	Very High
Sales Executive	69	54	91	112
Laboratory Technician	56	48	75	80
Research Scientist	54	53	90	95
Healthcare Representative	26	19	43	43
Manufacturing Director	26	32	49	38
Manager	21	21	27	33
Research Director	15	16	27	22
Sales Representative	12	21	27	23
Human Resources	10	16	13	13
Grand Total	**289**	**280**	**442**	**459**

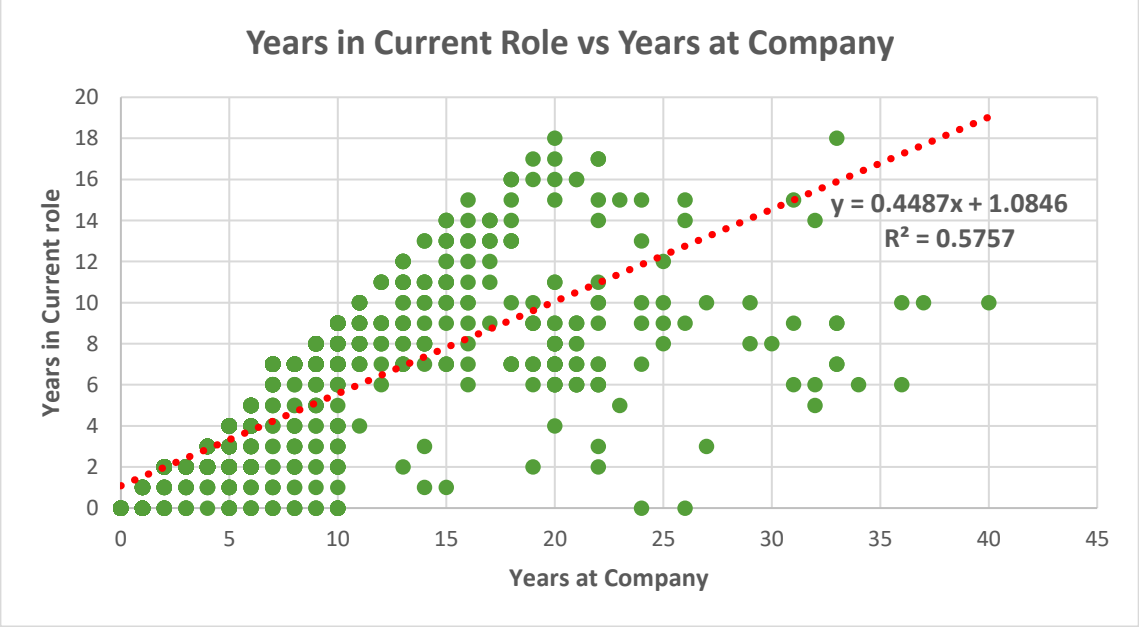
WorkLifeBalance by Job Roles	Work Life Balance			
Job Roles	Bad	Good	Better	Best
Laboratory Technician	20	58	156	25
Research Scientist	16	86	166	24
Sales Executive	12	76	202	36
Healthcare Representative	10	30	80	11
Manufacturing Director	7	34	90	14
Manager	6	23	61	12
Research Director	4	15	49	12
Human Resources	4	6	32	10
Sales Representative	1	16	57	9
Grand Total	**80**	**344**	**893**	**153**

	Job Satisfaction			
Job Role	Low	Medium	High	Very High
Sales Executive	24%	19%	21%	24%
Laboratory Technician	19%	17%	17%	17%
Research Scientist	19%	19%	20%	21%
Healthcare Representative	9%	7%	10%	9%
Manufacturing Director	9%	11%	11%	8%
Manager	7%	8%	6%	7%
Research Director	5%	6%	6%	5%
Sales Representative	4%	8%	6%	5%
Human Resources	3%	6%	3%	3%

	Work Life Balance			
Job Role	Bad	Good	Better	Best
Laboratory Technician	25%	17%	17%	16%
Research Scientist	20%	25%	19%	16%
Sales Executive	15%	22%	23%	24%
Healthcare Representative	13%	9%	9%	7%
Manufacturing Director	9%	10%	10%	9%
Manager	8%	7%	7%	8%
Research Director	5%	4%	5%	8%
Human Resources	5%	2%	4%	7%
Sales Representative	1%	5%	6%	6%

2. **Overfitting:**

Based on the two models I had chosen plus my findings in project section 5, none of the charts display overfitting due to not using the correct data model. I was more focused on the visual end of my analysis instead of the statistics. But after conducting additional analysis, I created a scatter plot to find any correlation between years at the company and years in current role. I hypothesized that there might be a relationship between these two variables to correlate with attrition. Still, based on my results, the scatter plot did not showcase overfitting but rather under-fitting due to the model being too simple by having less variance for predictions.



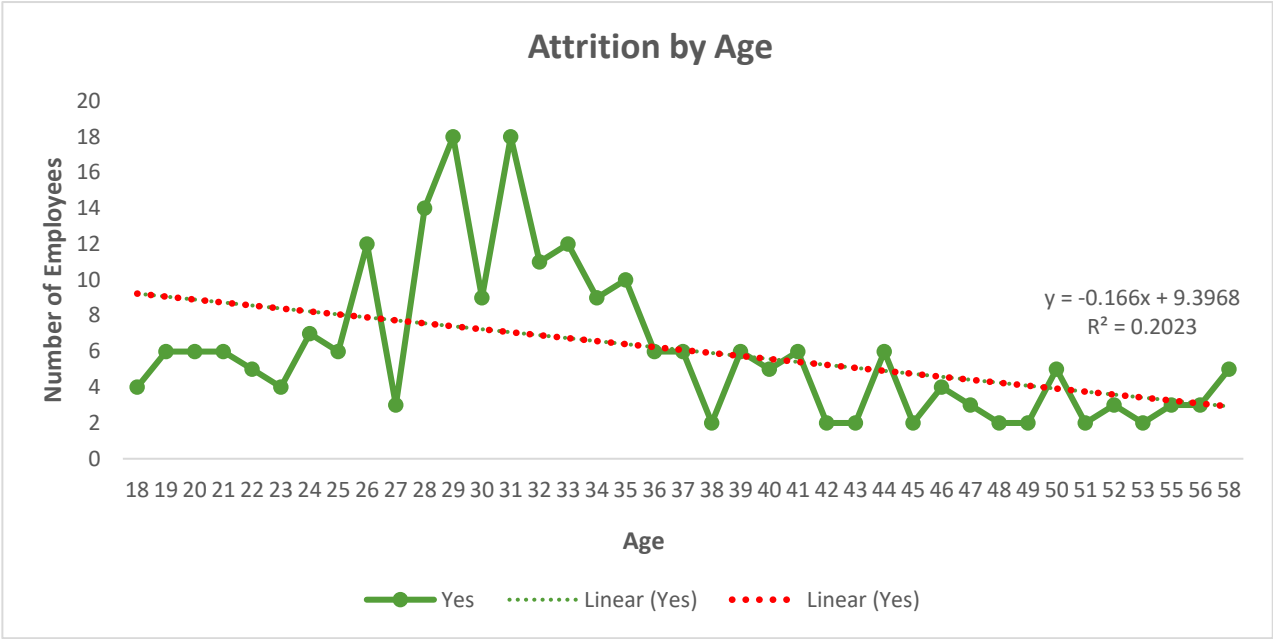
3. **Causality:**
   Going back to my two models about job roles and the relationship between work-life balance and job satisfaction, I can conclude that one of the leading causes of employees leaving is low job satisfaction. Low job satisfaction can mean numerous things, such as having a poor work-life balance. Employees who experience poor work-life balance are caused by factors such as excessive workload, time pressure, lack of job security, hostile work environment, strict rules, etc. These factors then connect to low job satisfaction, soon resulting in attrition. The analysis I conducted in the previous project section supports my causation, but I would like to conduct more analysis to find any other leading causes.

4. **Significance:**
   Based on my findings, they showcased significance correlated to my business problem. Most of my findings correlate to the attrition factors, and the results aren't based on chance. When you look back at my findings in the two models and my analysis in project section five, the first three results are the same: Sales Executives, Laboratory Technicians, and Research Scientists. Every analysis I conducted that revolves around job roles will most likely output the three job roles I listed. This proves that these job roles suffer the most from attrition.

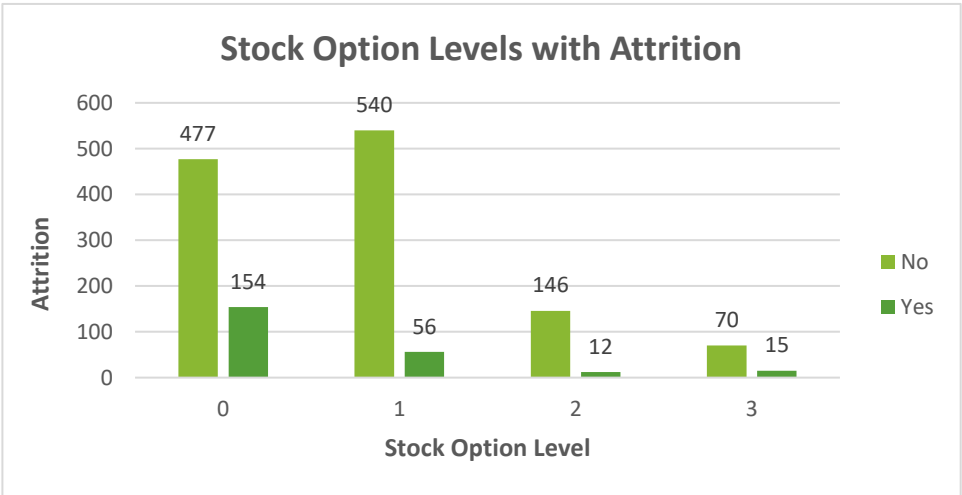
5. **Effect Size:**
   Relating to significance, there are some effect sizes in the job satisfaction bar chart. For example, there's a vast difference between the sales executive and human resources value in terms of low job satisfaction with the value of 69 vs. 10. The results conclude that human resources have the smallest low job satisfaction value compared to sales executives. In addition, the top three job roles with the smallest low job satisfaction are human resources, sales representatives, and research directors. Overall, the purpose of the bar chart was to interpret which job roles have the highest low job satisfaction level, which demonstrates it well.

Overall, the models plus the analysis in project section five has interesting results correlated to the leading cause of attrition. Based on my evaluation, I would like to conduct a few more analysis to further support and reinforce my findings for the leading cause of my business problem.

One of the interesting results while conducting more analysis was the correlation between age with attrition. Based on my results, employees ages 26 to 35 exceed the attrition trendline, meaning that their mid-20s to mid-30s have higher attrition than the rest of the age group. The line chart correlates with the average age in job roles back in project section five, where the top three job roles are most likely to leave, Sales Executives, Laboratory Technicians, and Research Scientists. The three job roles have an average age of 34 to 37.

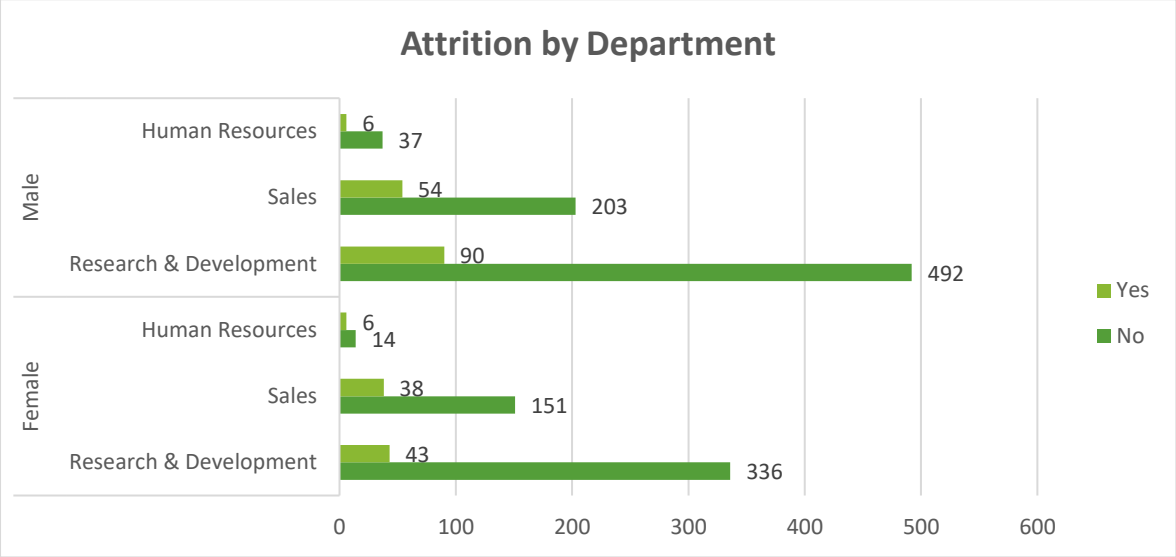


The bar chart showcases the relationship between stock options levels with attrition. Based on the bar chart, employees with 0 or 1 stock options level have the highest attritions compared to stock options levels 2 and 3. Stock options are simply defined as a form of compensation employees can receive from the company. The compensation is contracts that the employees receive, and it explains they have the right to buy and sell a certain number of company shares. Level 0 stock options levels mean the employee didn't receive any stock options. A level 1 stock option means employees can buy/sell limited company shares; the higher the level, the more shares employees can buy/sell. The graph clearly shows that employees with level 0 or 1 stock options are more likely to leave the company.

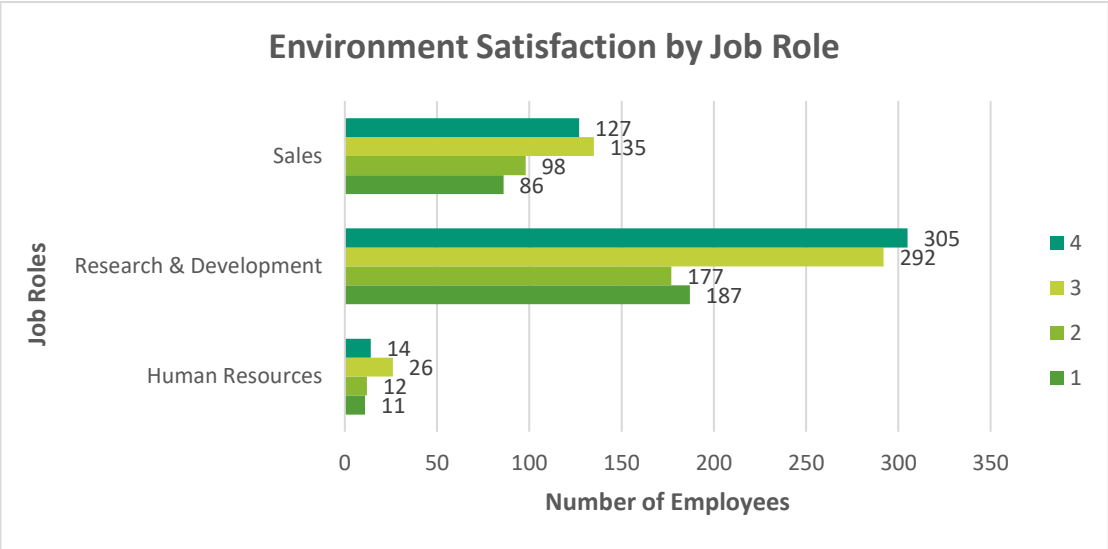


	Attrition rate	
Stock Option Level	No	Yes
0	39%	65%
1	44%	24%
2	12%	5%
3	6%	6%

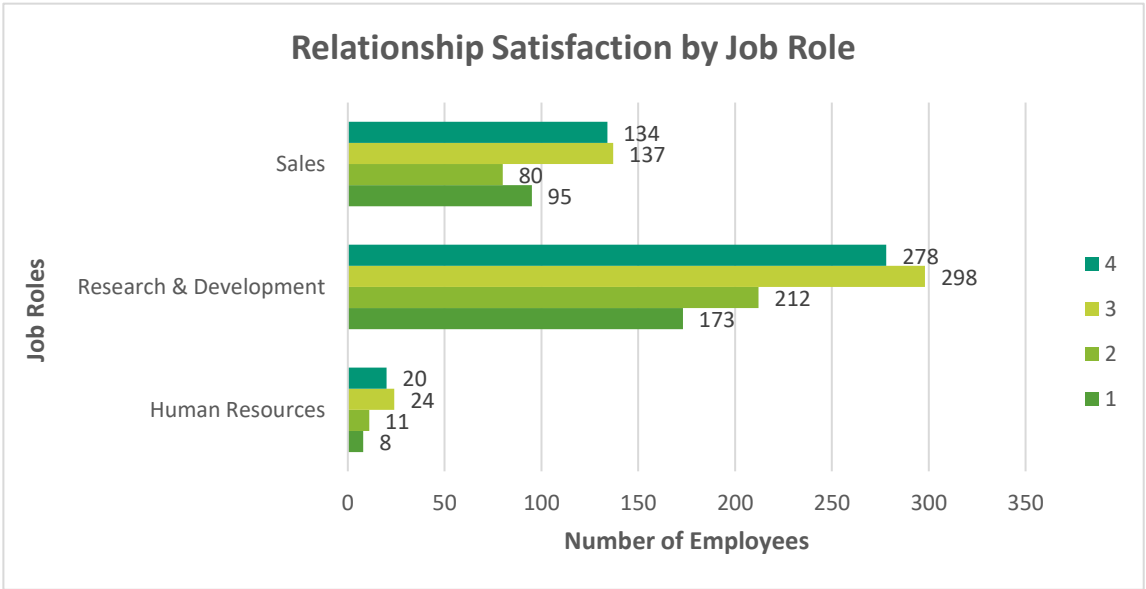
Besides the stock options level bar graph, this bar chart displays the number of attritions by the department. According to the bar graph, the research/development and sales department are the top two with the most attrition, reinforcing my findings from my previous analysis of the job roles. To reinforce my findings, the two other bar graphs showcase the department's environment and relationship satisfaction levels (1 = low satisfaction, 4 = very high satisfaction). Environment satisfaction is defined as the employee's satisfaction with their work environment, and relationship satisfaction is defined as the employee's relationship with their co-workers. Based on the bar graphs, The top two departments with the lowest satisfaction levels are research/development and sales, correlating again to the three job roles.



	Attrition rate	
Job Department	**No**	**Yes**
Female		
Research & Development	27%	18%
Sales	12%	16%
Human Resources	1%	3%
Males		
Research & Development	40%	38%
Sales	16%	23%
Human Resources	3%	3%

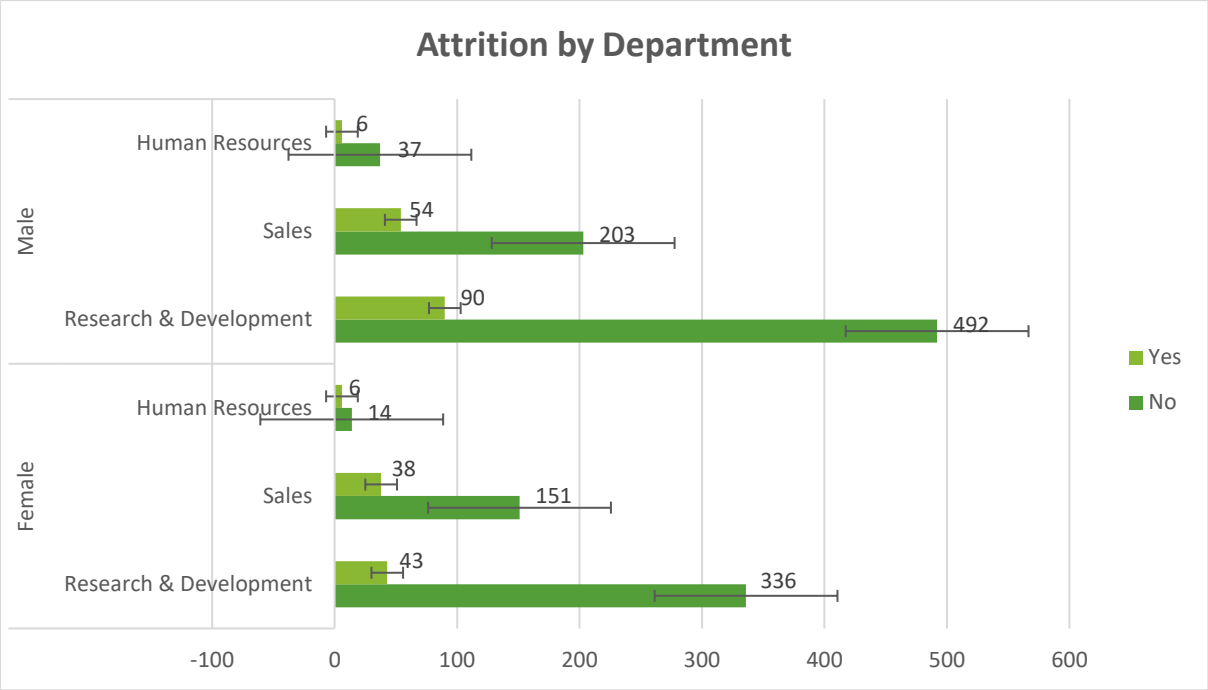


	Environment Satisfaction			
Job Department	**Low**	**Medium**	**High**	**Very High**
Human Resources	4%	4%	6%	3%
Research & Development	66%	62%	64%	68%
Sales	30%	34%	30%	28%



	Relationship Satisfaction			
Job Department	**Low**	**Medium**	**High**	**Very High**
Human Resources	3%	4%	5%	5%
Research & Development	63%	70%	65%	64%
Sales	34%	26%	30%	31%

After conducting additional analysis to support my findings in previous project sections, I will use this model during my additional analysis as my final model. I selected this chart as my final model because it summarizes everything I found. Most of the analyses I conducted fall under low job satisfaction levels. Low job satisfaction is an umbrella for low environment/relationship satisfaction, bad stock options levels, low monthly income, and a poor work-life balance. All the findings resulted in the same outcome. The research/development and sales department have the lowest satisfaction level and highest attrition correlating to the top three job roles with the lowest job satisfaction: Sales Executives, Laboratory Technicians, and Research Scientists. The average age for the three job roles is between ages 34 to 37, and the age that is most likely to leave are the ages between 26 to 35.



Attrition by Department	Attrition	
Department	No	Yes
Female	**501**	**87**
Research & Development	336	43
Sales	151	38
Human Resources	14	6
Male	**732**	**150**
Research & Development	492	90
Sales	203	54
Human Resources	37	6

	Attrition rate	
Job Department	No	Yes
Female		
Research & Development	27%	18%
Sales	12%	16%
Human Resources	1%	3%
Males		
Research & Development	40%	38%
Sales	16%	23%
Human Resources	3%	3%

PROJECT SECTION 7: INTERPRETATION

Job Satisfaction by Job Roles	Job Satisfaction			
Job Role	Low	Medium	High	Very High
Sales Executive	69	54	91	112
Laboratory Technician	56	48	75	80
Research Scientist	54	53	90	95
Healthcare Representative	26	19	43	43
Manufacturing Director	26	32	49	38
Manager	21	21	27	33
Research Director	15	16	27	22
Sales Representative	12	21	27	23
Human Resources	10	16	13	13
Grand Total	**289**	**280**	**442**	**459**

	Job Satisfaction			
Job Role	Low	Medium	High	Very High
Sales Executive	24%	19%	21%	24%
Laboratory Technician	19%	17%	17%	17%
Research Scientist	19%	19%	20%	21%
Healthcare Representative	9%	7%	10%	9%
Manufacturing Director	9%	11%	11%	8%
Manager	7%	8%	6%	7%
Research Director	5%	6%	6%	5%
Sales Representative	4%	8%	6%	5%
Human Resources	3%	6%	3%	3%

WorkLifeBalance by Job Roles	Work Life Balance			
Job Roles	Bad	Good	Better	Best
Laboratory Technician	20	58	156	25
Research Scientist	16	86	166	24
Sales Executive	12	76	202	36
Healthcare Representative	10	30	80	11
Manufacturing Director	7	34	90	14
Manager	6	23	61	12
Research Director	4	15	49	12
Human Resources	4	6	32	10
Sales Representative	1	16	57	9
Grand Total	**80**	**344**	**893**	**153**

	Work Life Balance			
Job Role	Bad	Good	Better	Best
Laboratory Technician	25%	17%	17%	16%
Research Scientist	20%	25%	19%	16%
Sales Executive	15%	22%	23%	24%
Healthcare Representative	13%	9%	9%	7%
Manufacturing Director	9%	10%	10%	9%
Manager	8%	7%	7%	8%
Research Director	5%	4%	5%	8%
Human Resources	5%	2%	4%	7%
Sales Representative	1%	5%	6%	6%

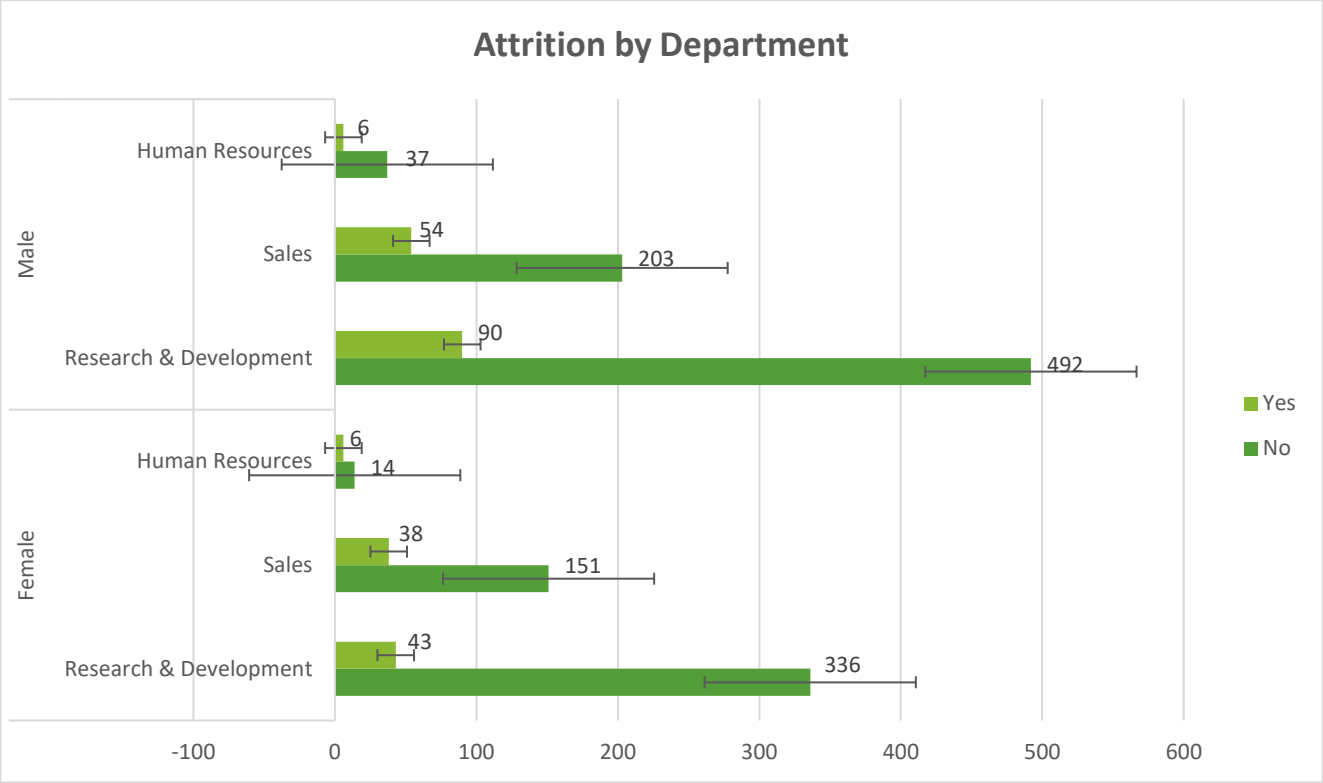
Problem Definition: What are the main factors that cause IBM employees to attrite?

Question Definition: Which departments are most heavily affected by attrition?

Directionality: According to the two models in project section 6, it's difficult to determine the directionality as their different values for each group. The order by grand total from least to greatest in the job satisfaction table starts medium, low, high, and very high. The order by grand total from least to greatest in the work-life balance table starts bad, best, good, and better.

Magnitude: The two models are collected by the total employees in the dataset. They're 1,470 employees in the IBM dataset. Since we're focusing on the characteristics of attrition, the low job satisfaction and bad work-life balance columns are the focus. 289/1470 or 19.67% of IBM employees experience low job satisfaction in the job satisfaction table, and 80/1,470 or 5.44% of IBM employees have a bad work-life balance.

Uncertainty: Based on the two models, they're no uncertainties as the results differ by each job role.



Attrition by Department	Attrition	
Department	No	Yes
Female	**501**	**87**
Research & Development	336	43
Sales	151	38
Human Resources	14	6
Male	**732**	**150**
Research & Development	492	90
Sales	203	54
Human Resources	37	6

	Attrition rate	
Job Department	No	Yes
Female		
Research & Development	27%	18%
Sales	12%	16%
Human Resources	1%	3%
Males		
Research & Development	40%	38%
Sales	16%	23%
Human Resources	3%	3%

Traits/Patterns: The final model depicts attrition across IBM's three core job departments: Human resources, Sales, and Research & Development. Not only are the results split into departments but is also categorized by gender. The model showcases a few interesting characteristics. For one, there's more attrition among male employees than female employees. In terms of which department, Research & Development has the most attrition, next is sales, and last is human resources. Since the human resource department has the least amount of attrition compared to the other departments, they are the least likely to leave IBM. The attrition pattern is the same for female employees but not as much for male employees. The overall model answers my question as I found which department is most affected by attrition. For my original problem, I wanted to see the leading causes of attrition which I did during the analysis stage. The result of the final model strongly correlates to the results in project sections five and six. For example, since the final model proves that most of the attrition consists of male employees in the research & development department, two of the top three job roles most likely to leave are Laboratory Technicians and Research Scientists, Sales executives being the first.

For the gender statistic, 63% that makes up attrition were male employees vs. female employees with 37%. Based on the final model, there were no outliers to be found.

PROJECT SECTION 8: RECOMMENDATIONS

After collecting results from IBM's dataset, I can make recommendations to upper management. Based on their request to find the leading cause of attrition and which department is heavily impacted, I can conclude that male employees were impacted the most, research & development department with 38% and the sales department with 23%.. Based on the two departments, Sales Executives (24%), Laboratory Technicians (19%), and Research Scientists (19%) are the top three job roles most likely to leave due to low job satisfaction, which correlates to the top two departments. Many factors connect to low job satisfaction, such as poor work-life balance, low environment satisfaction, low relationship satisfaction with other employees/management, poor stock option level benefits, and low monthly income. To reduce the attrition amongst the top two departments, I would recommend the following suggestions: implementing flexibility in work schedules to minimize poor work-life balance, promoting a positive work environment, better communication and transparency between employees to management, acknowledging and appreciating the employee's hard work through rewards, and hiring the right candidate/improve training. To tackle pay-related issues, I would suggest implementing better compensation and benefits for all roles, no matter the employees' job level. Incorporating the following suggestions will reduce attrition and increase the chance of employees staying longer.

	Job Satisfaction			
Job Role	**Low**	**Medium**	**High**	**Very High**
Sales Executive	24%	19%	21%	24%
Laboratory Technician	19%	17%	17%	17%
Research Scientist	19%	19%	20%	21%
Healthcare Representative	9%	7%	10%	9%
Manufacturing Director	9%	11%	11%	8%
Manager	7%	8%	6%	7%
Research Director	5%	6%	6%	5%
Sales Representative	4%	8%	6%	5%
Human Resources	3%	6%	3%	3%

	Environment Satisfaction			
Job Department	**Low**	**Medium**	**High**	**Very High**
Human Resources	4%	4%	6%	3%
Research & Development	66%	62%	64%	68%
Sales	30%	34%	30%	28%

	Relationship Satisfaction			
Job Department	**Low**	**Medium**	**High**	**Very High**
Human Resources	3%	4%	5%	5%
Research & Development	63%	70%	65%	64%
Sales	34%	26%	30%	31%

	Work Life Balance			
Job Role	**Bad**	**Good**	**Better**	**Best**
Laboratory Technician	25%	17%	17%	16%
Research Scientist	20%	25%	19%	16%
Sales Executive	15%	22%	23%	24%
Healthcare Representative	13%	9%	9%	7%
Manufacturing Director	9%	10%	10%	9%
Manager	8%	7%	7%	8%
Research Director	5%	4%	5%	8%
Human Resources	5%	2%	4%	7%
Sales Representative	1%	5%	6%	6%

	Attrition rate	
Job Department	**No**	**Yes**
Female		
Research & Development	27%	18%
Sales	12%	16%
Human Resources	1%	3%
Males		
Research & Development	40%	38%
Sales	16%	23%
Human Resources	3%	3%

	Attrition rate	
Gender	**No**	**Yes**
Female	41%	37%
Male	59%	63%

	Attrition rate	
Stock Option Level	**No**	**Yes**
0	39%	65%
1	44%	24%
2	12%	5%
3	6%	6%

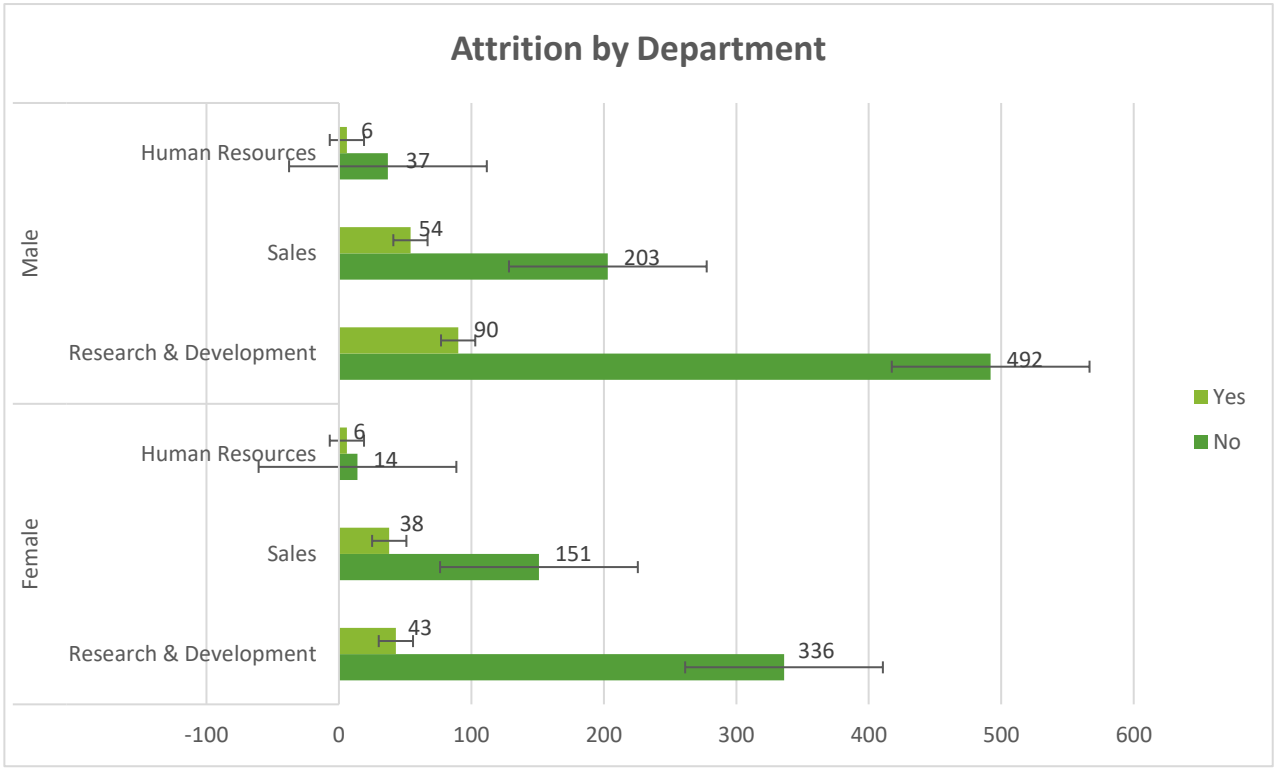
PROJECT SECTION 9: EXECUTIVE SUMMARY

The purpose of the project analysis is to find the correlation between attrition rate and the job departments by exploring the root cause. Attrition is defined as employees leaving an organization for various reasons, either voluntarily or involuntarily: termination, resignation, retirement, or death. Attrition can impact the company's health by reducing performance and productivity due to the lack of employment. If performance and productivity decrease over time, eventually, IBM will go bankrupt. In addition to attrition, they're various causes to why employees leave.

The following below is a summary of the findings I found throughout my analysis of the IBM dataset:

- Male employees have a higher attrition rate than female employees
    - Males: 63%
    - Females: 37%
- Research & development and sales department are the top two departments that make up the majority of attrition
    - Research & development: 56%
    - sales department: 39%
- Sales executive, laboratory technician, and research scientist are the top three job roles with the highest low job satisfaction levels corresponding to the top two departments that make up most of IBM's attrition rate
    - Sales executive: 24%
    - Laboratory technician: 19%
    - Research scientist: 19%
- Research & development and sales department are the top two departments with the highest low environment/relationship satisfaction levels
    - Low environment satisfaction
        - Research & development: 66%
        - Sales: 30%
    - Low relationship satisfaction
        - Research & development: 63%
        - Sales: 34%
- Sales executive, laboratory technician, and research scientist are the top three job roles with the worst work-life balance
    - Sales executive: 15%
    - Laboratory technician: 25%
    - Research scientist: 20%
- 65% of Employees with level 0 stock options levels are most likely to leave
- Research scientists and laboratory technicians are the top two job roles with the lowest monthly income

Based on my findings, research & development and the sales department make up most of IBM's attrition rate. Under these departments, the top three job roles most likely to leave are sales executives, laboratory technicians, and research scientists. The leading causes of these job roles leaving are low job satisfaction, poor work-life balance, low monthly income, bad stock options, and low environment/relationship satisfaction. The following chart below summarizes all my findings in one chart.


Attrition by Department
Male
Female
Human Resources
Sales
Research & Development
Human Resources
Sales
Research & Development
6
37
54
203
90
492
6
14
38
151
43
336
Yes
No
-100
0
100
200
300
400
500
600

PROJECT SECTION 10: PRESENTATION

Link to my presentation: https://youtu.be/PfTXbA0LJSc