TECHNOCOLABS SOFTWARES

Internship Mini-Project

We completed this mini-project at Technocolabs Softwares which honed our skills and bolstered our experience for real-time problems. The technologies majorly used are GitBash/Git CLI/Linux, MS Excel, Python and Tableau.

OVERVIEW

A comprehensive analysis of the attrition at ACME Corporation.

This project aims to provide insights into the factors influencing employee attrition and predict which employees will likely leave the company. Acme Corporation, a leading tech company, faces a significant challenge with employee turnover. The HR department is concerned about the increasing attrition, which negatively impacts team dynamics, project continuity, and overall company morale. Acme Corporation wants to leverage data analytics and machine learning to understand the factors influencing employee turnover and predict which employees might leave soon.

The Objective

The purpose is to find a relation between different domains of the dataset and attain reasonable conclusions that can be further worked upon.

- Need #1: to identify the reasons for attrition
- Need #2: reason out the development of staff to reduce attritions
- Need #3: predict who is more likely to quit

The Opportunity

Opportunity to perform an in-depth analysis and execute tasks accordingly

- Goal #1: Data Exploration and Visualization
- Goal #2: Descriptive Analytics
- Goal #3: Predictive Analytics
- Opportunity: Learn to Use Power BI

The Solution

Refer to visualisations made in Power BI to comply with the following inferences and recommendations. It is recommended that you follow Key Value Influencers to target the issues at various departments.

- Recommendation #1: Target Sales Department
- Recommendation #2: Follow a Particular Pattern Depending on Particular Variables for Monthly Income
- Recommendation #3: Administer the Departments at the Administration and Interpersonal Levels to Restore the Faith of Youth

OUR PROPOSAL

During this case study Sales department was particularly highlighted.

Acme shares high reviews on GlassDoor and AmbitionBox. But the current rate of attrition has understandably reduced the trust in youth and the current average rate of attrition 16.12% is affected by some key factors along with many others.

The key factors though might not seem to be related but have a lasting impact on any company and the sentiments of employees, and attrition acts as fuel to fire. During my study we found these factors to impact mostly the Sales department. The analysis suggests distress among employees in the Sales department. Furthermore, unexpectedly the trend doesn't seem to follow any gender or marital status bias. It is also to be mentioned the relation between hourly rates, daily rates, monthly rates, total working experience, years in current role and monthly income is also a little ambiguous irrespective of departments. Therefore, it is highly recommended that Acme officials work on administrative as well as well as interpersonal levels to regain the confidence of youth.

Kindly note that although out of 237, 122 resignations supposedly represent they will still impact the organisation and have been considered, considering 51.44% of 237 is not a small amount and the data might have been falsified under pressure from superiors. Therefore, it is highly recommended that Key Value Influencers are studied in detail and actions are taken accordingly.

Rationale

Description of following this pattern.

- Research: Provided me with ample opportunities to study and delve deeper into the data along with honing my analysis skills.
- Each step of analysis led me to conclusive details of the dataset.
- Alignment with the Mission: The reason and departments affected by attrition became clearer by effectively representing data using Power BI.
- Current resources/technology: Excel, Python, MS SQL Server, MS Power BI, MS Word (in order of timeline)

Execution Strategy

Our execution strategy incorporates proven methodologies, extremely qualified personnel, and a highly responsive approach to managing deliverables. Following is a description of our project methods, including how the project will be developed, a proposed timeline of events, and reasons for why we suggest developing the project as described.

Technical/Project Approach

> *To complete the project, I have primarily focused on using MS SQL Server, Python, MS Excel and Tableau in order of extent of use.*

The first step was to download the project from the GitHub link by forking it and pulling it using Git CLI commands through WSL2 which was preinstalled in my system. After the successful download, I opened the CSV file in MS Excel to view the dataset. Since the initial dataset consists of numerous rows and columns (which would be, but not necessarily constrained to, referred to as measures and dimensions later), identification of the Primary Key and total number of columns were necessary along with their names. I designed a module named Functions.py which has (as supposed) all the Python functions using Pandas in Python. The functions defined in the file identify the Primary Key and store the names of all the columns in rows of an MS Excel file.

After that, the second step called for the grouping of related data, therefore I used the SmartArt tool of MS Excel to identify related columns (as shown in the figures below) and formed different tables using MS SQL Server, hence enabling me to define a quantitative analysis before leaping directly into visualisation. Such an approach is more apt to deal with and allows us to configure visualisation, as we already know what to find. In addition, the purpose of SmartArts is visualisation and the actual tables in MS SQL Server can be different.

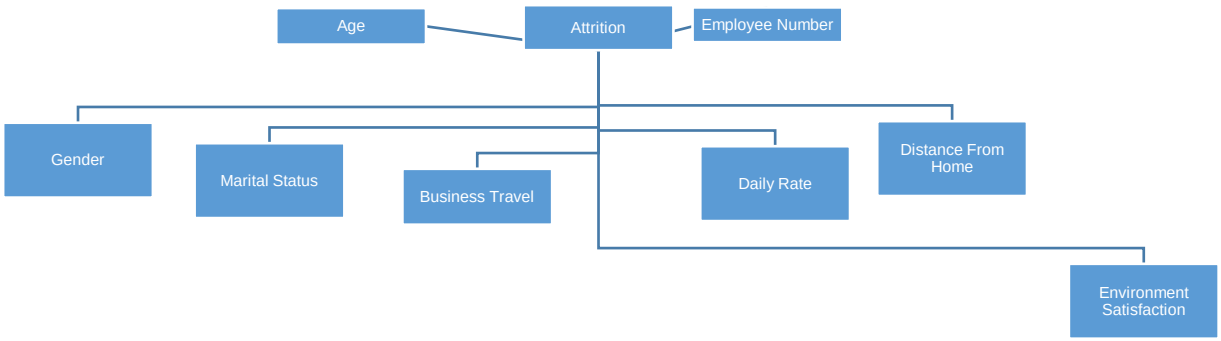

Table 1:EmployeePersonalDetails

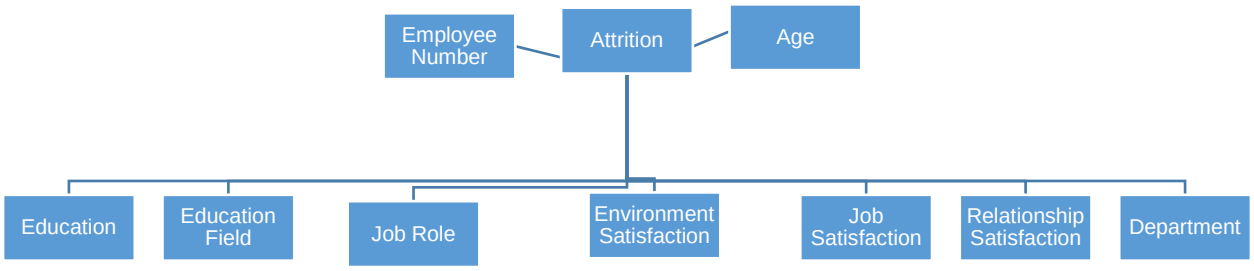

Table 2:EmpEduJobDetails

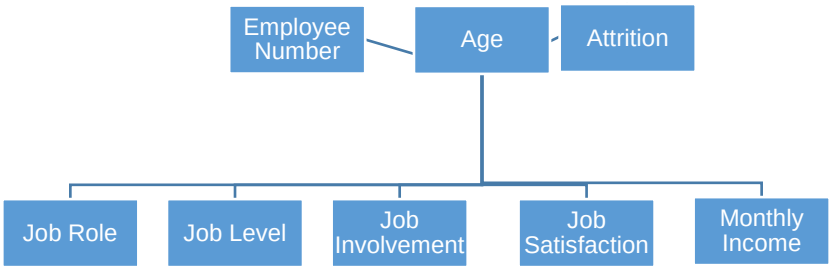

Table 3:EmpJobDetails

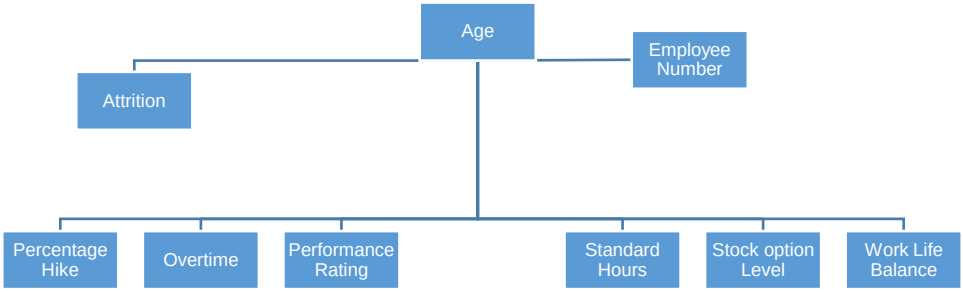

Table 4:EmpPerformance

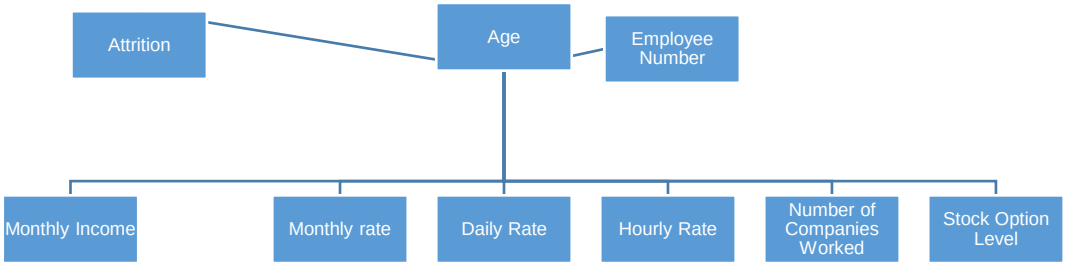

Table 5:EmpExpRate

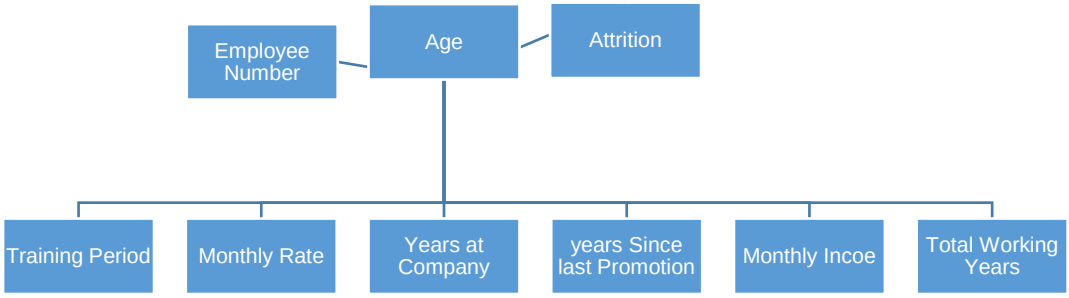

Table 6:EmpYearsSalAcme

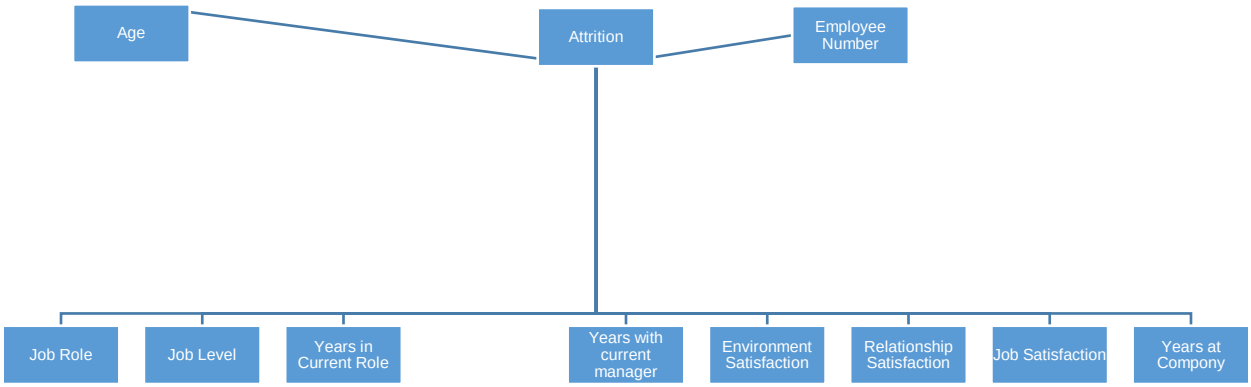

Table 7:EmpWithMgr

After designing and modifying these tables in and through MS SQL Server 2019, I designed the necessary queries for Tableau Public through the SQL database. Furthermore, I also performed combined analysis on different tables so created through SQL, to get a general idea of the data trend. That way, it was easier to filter and understand trends in Tableau. I have kept comments to show the output of each query and how each query can be related to different tables. In addition, I have used comments to depict my thought process and maintain coherence among different tasks. First, I generated general queries from EmployeePersonalDetails and studied why people with high environment satisfaction left. In the second, part within the same query I tried to analyse the relation between gender and attrition, by counting the number of total women in the company and women who left. This case study had to be more detailed as the jobs of women can be influenced by their marital status. Further studies showed possible outliers in the data.

The preliminary study of possible outliers shows the possible relations among different variables of the dataset which are studied in detail through key value influencers and visualisations. The following figures tend to show the possible relations, the results of which will be studied through key value influencers. The solutions will be developed based on this case study.

In addition, the software used for this purpose is Power BI and not Tableau Public Desktop as Power BI supports a direct import of SQL Queries, unlike Tableau Public Desktop. I would like to thank Technocolabs Softwares as it provided me with the opportunity to learn Power BI. Also, each visualisation has a key value influencer associated with it for the detailed case study.

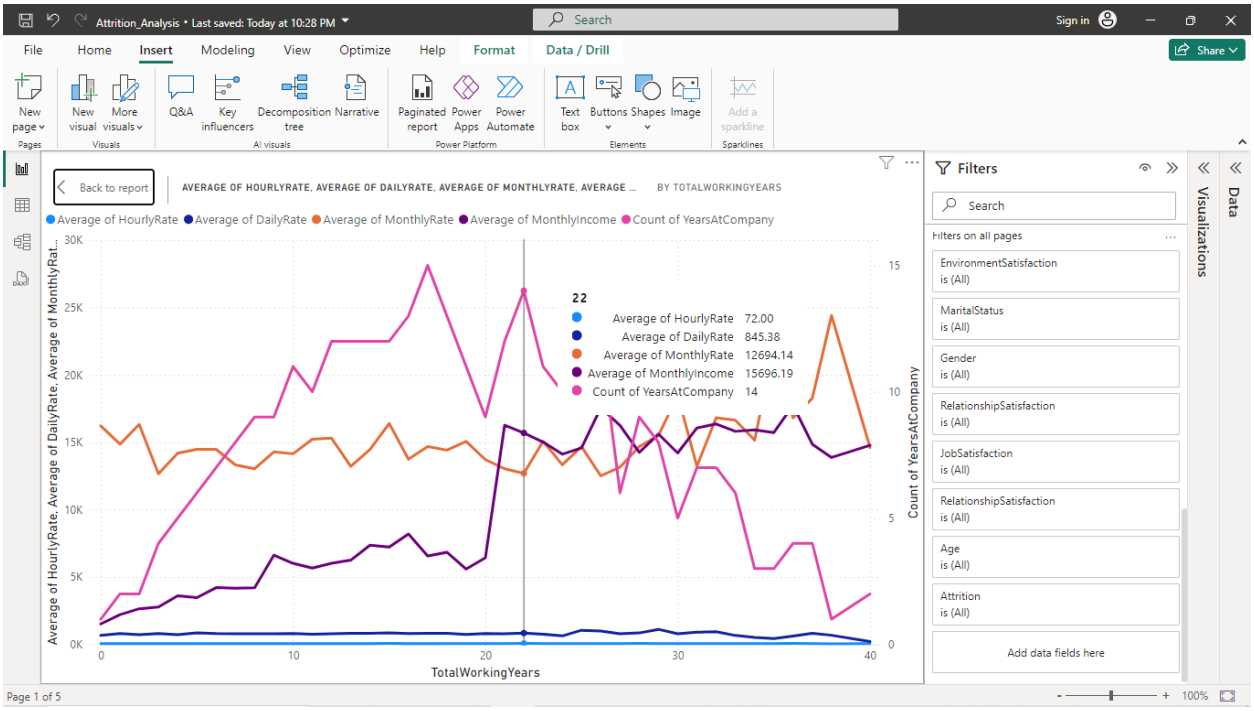

Figure 1:Reation between Monthly Income, Daily Rate, Hourly Rate Count of Years at the Company and Total Working Years

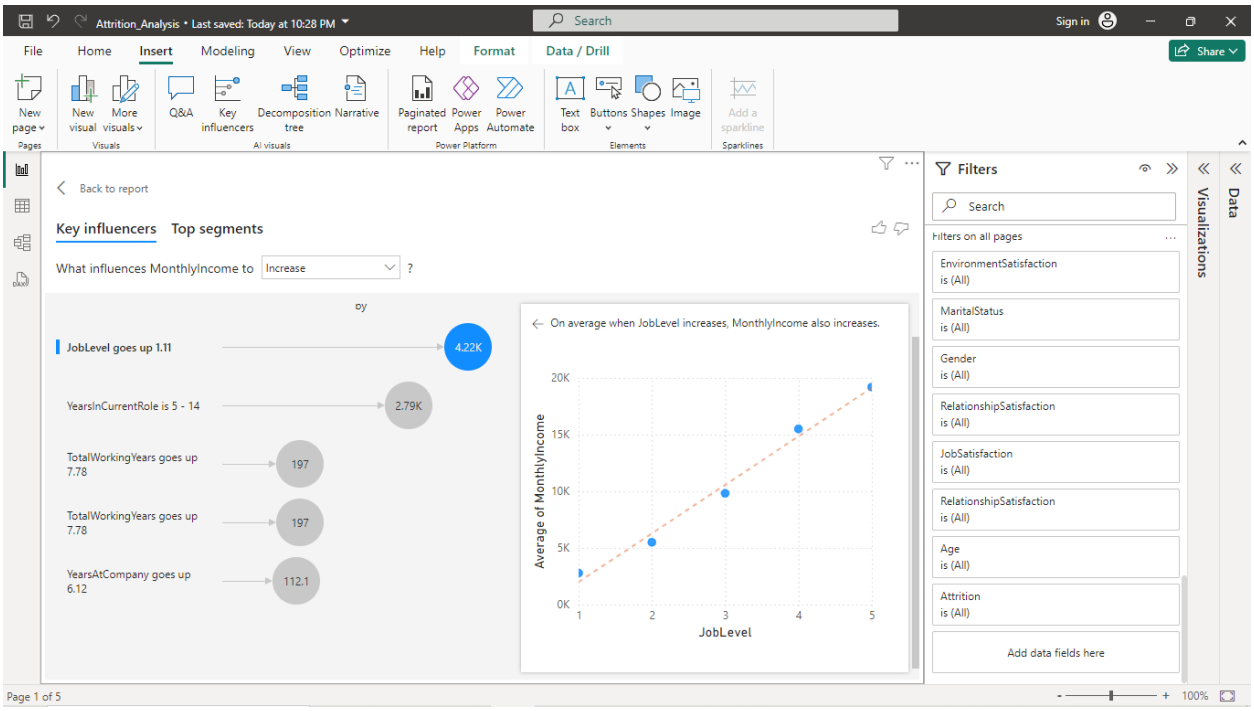

Figure 2Reation between Monthly Income, Daily Rate, Hourly Rate Count of Years at Company and total Working Years_ Key Value Influencers (a)

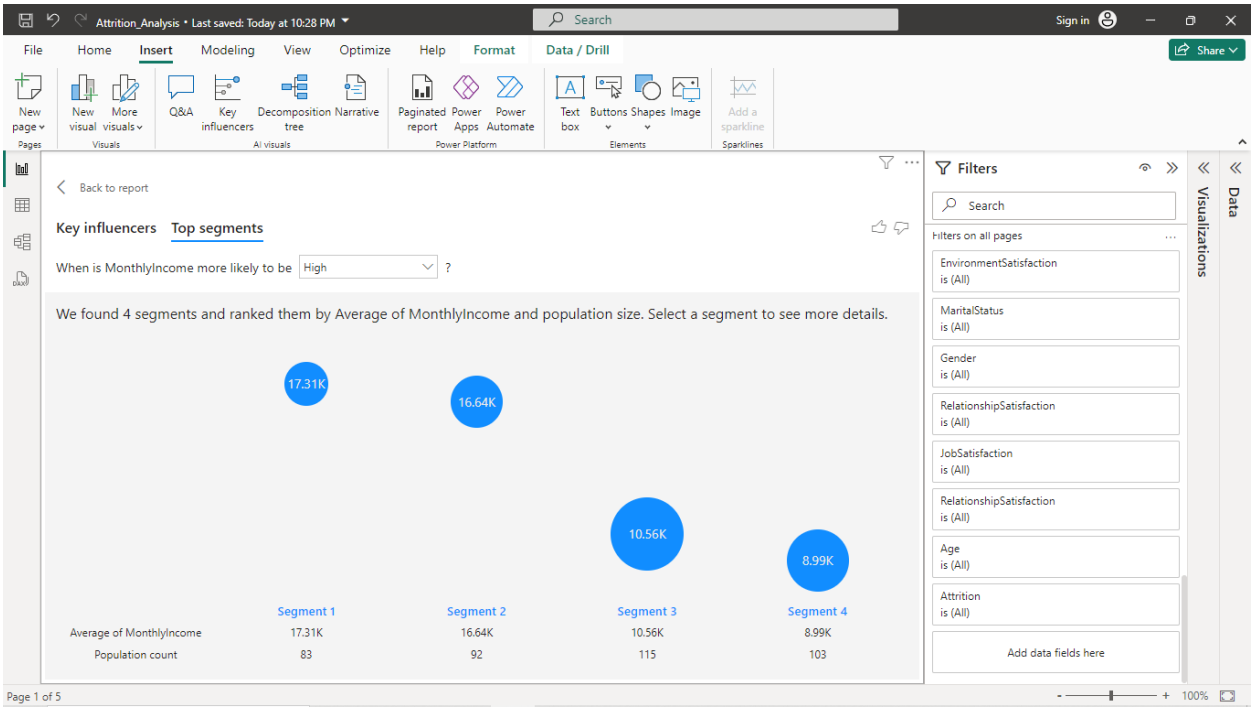

Figure 3:Reation between Monthly Income, Daily Rate, Hourly Rate Count of Years at Company and total Working Years_ Key Value Influencers (b)

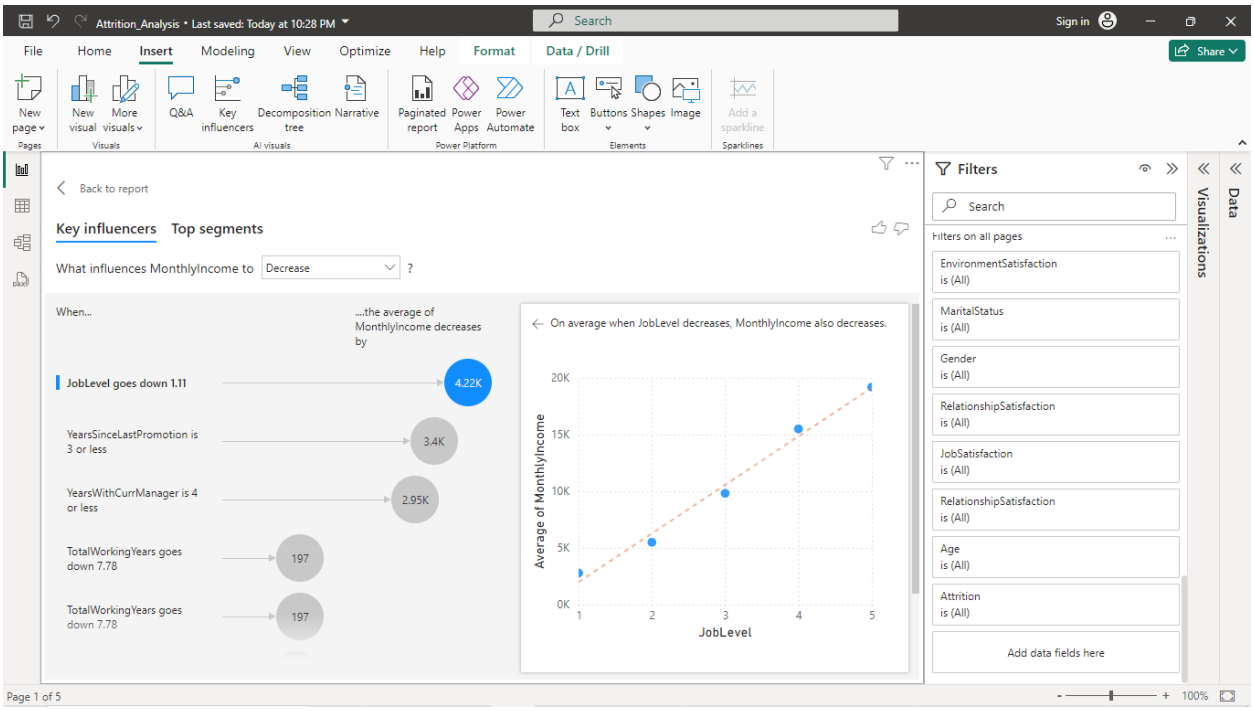

Figure 4:Reation between Monthly Income, Daily Rate, Hourly Rate Count of Years at Company and total Working Years_ Key Value Influencers (c)

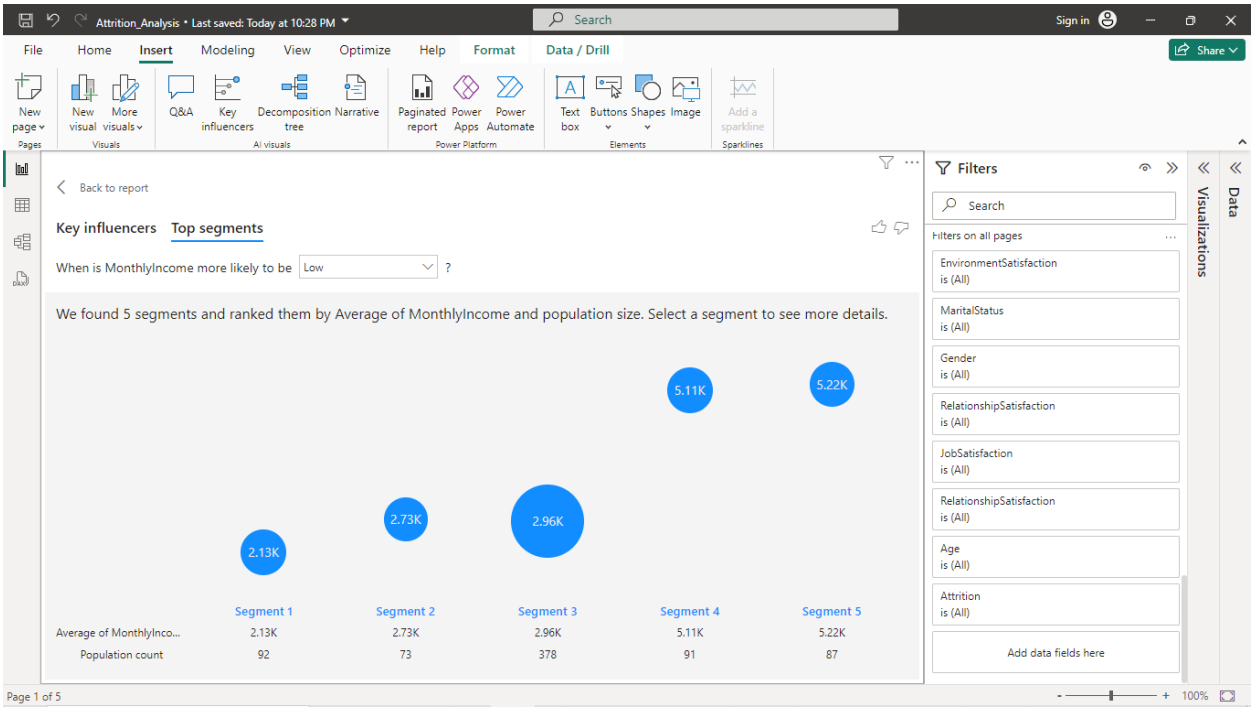

Figure 5:Reation between Monthly Income, Daily Rate, Hourly Rate Count of Years at Company and total Working Years_ Key Value Influencers (d)

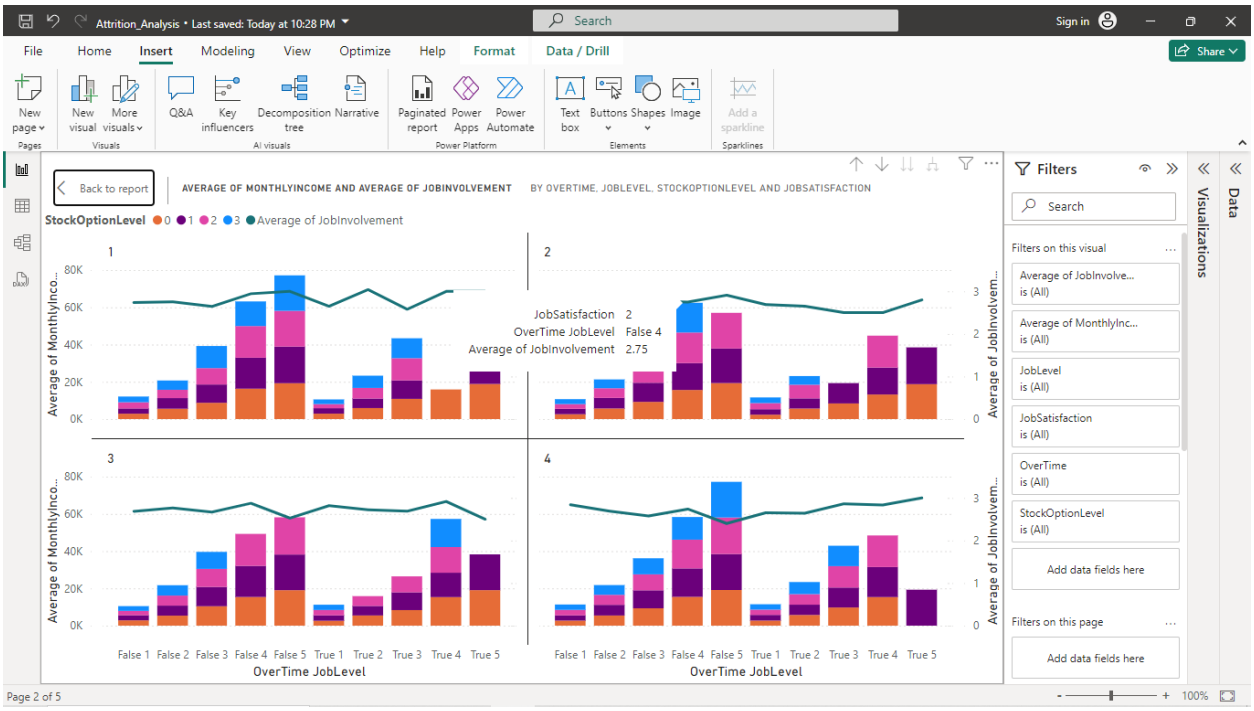

Figure 6:Monthly Income and Average of Job Involvement concerning Overtime, Job Level, Stock Option Level and Job Satisfaction

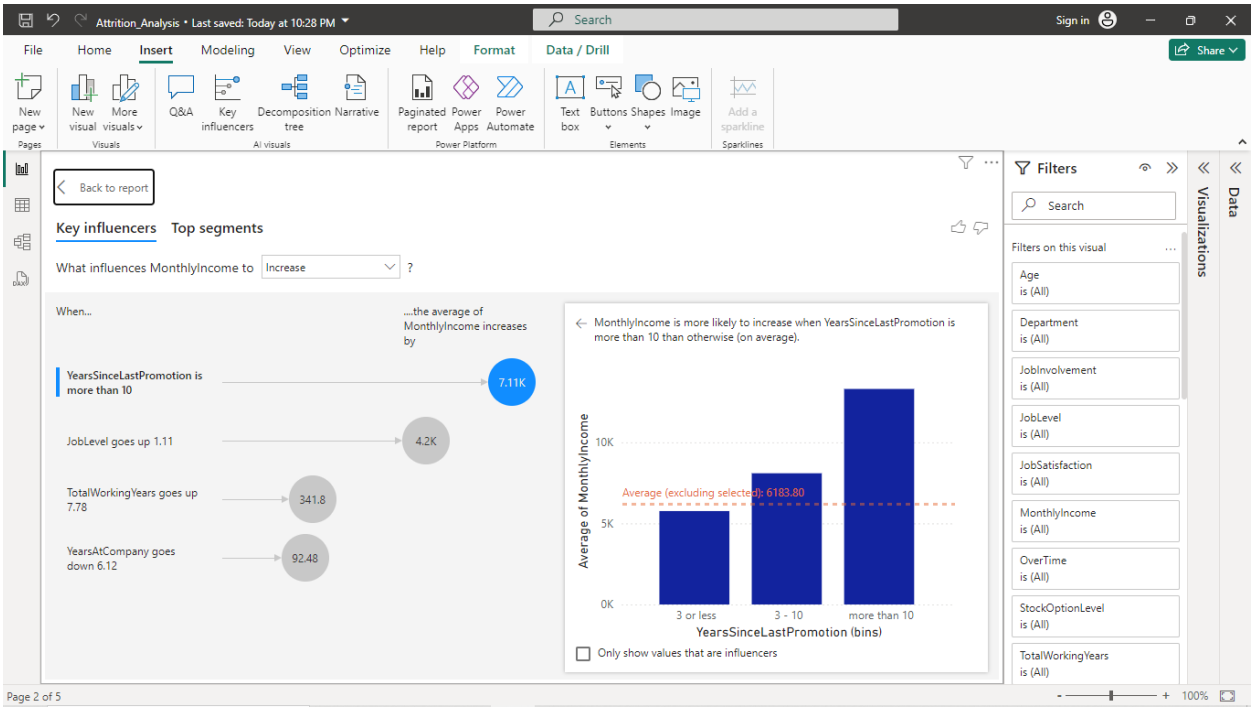

Figure 7:Monthly Income and Average of Job Involvement concerning Overtime, Job Level, Stock Option Level and Job Satisfaction_Key Value Influencers (a)

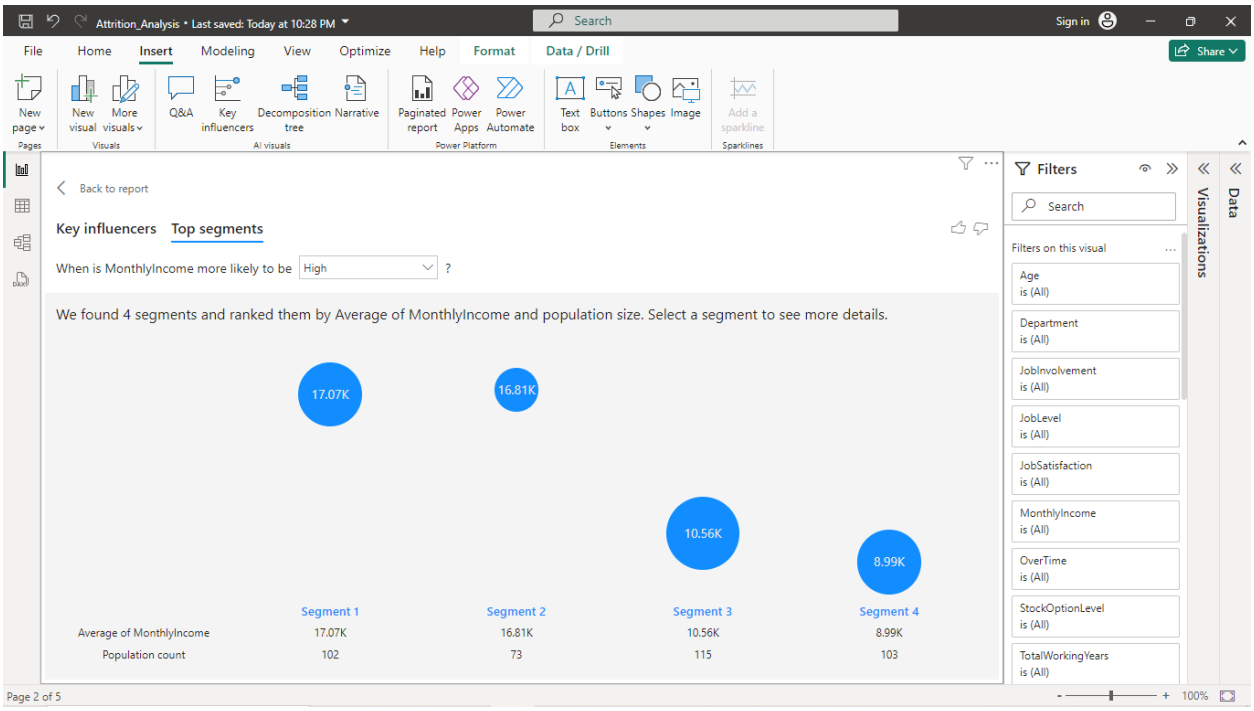

Figure 8:Monthly Income and Average of Job Involvement concerning Overtime, Job Level, Stock Option Level and Job Satisfaction_Key Value Influencers (b)

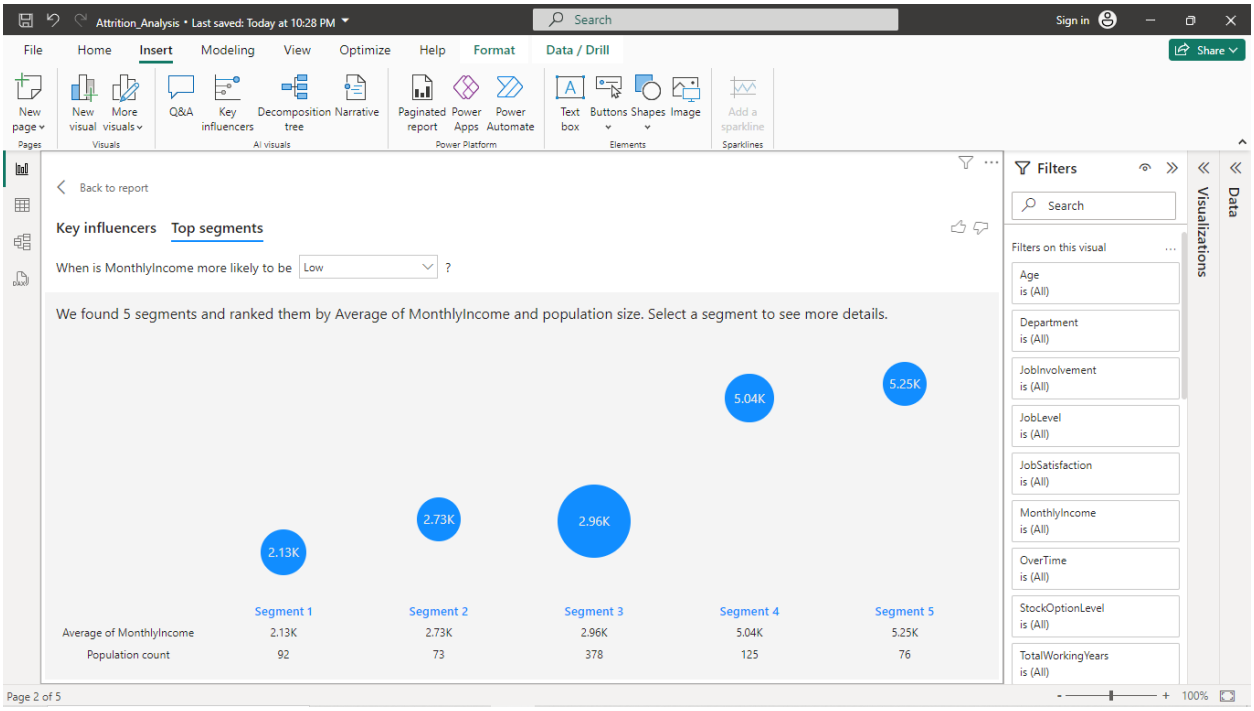

Figure 9:Monthly Income and Average of Job Involvement concerning Overtime, Job Level, Stock Option Level and Job Satisfaction_Key Value Influencers (c)

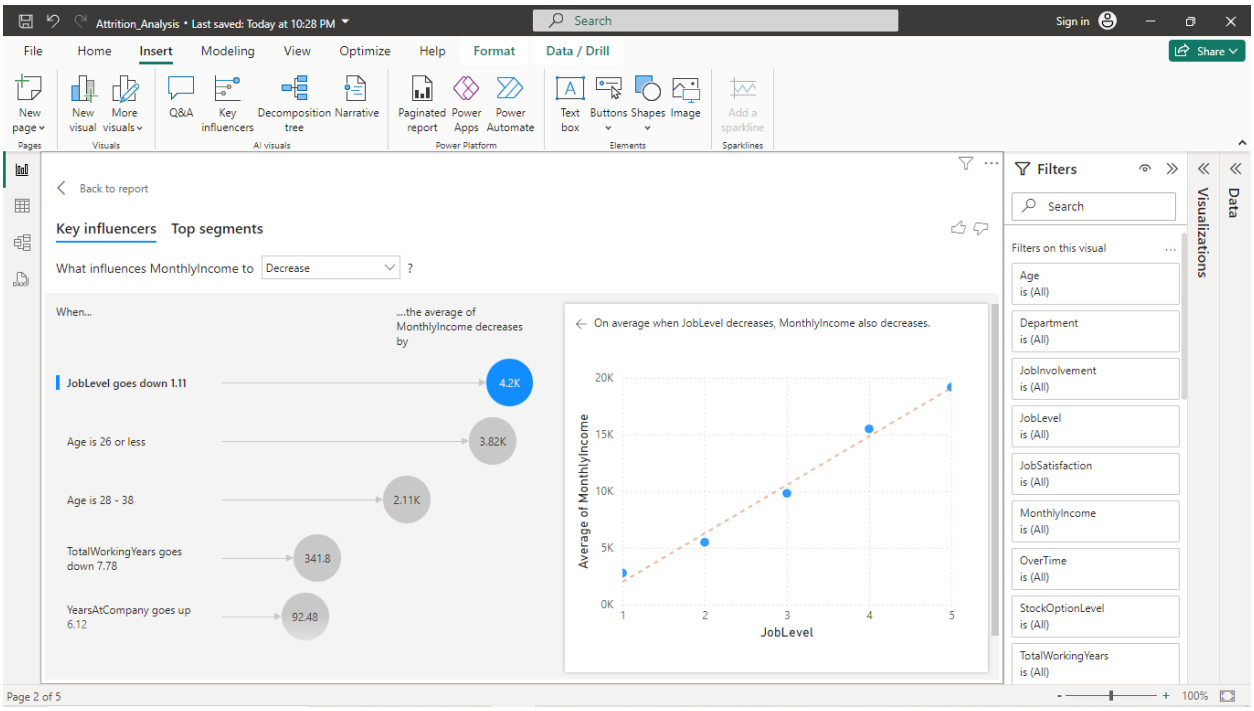

Figure 10:Monthly Income and Average of Job Involvement concerning Overtime, Job Level, Stock Option Level and Job Satisfaction_Key Value Influencers (d)

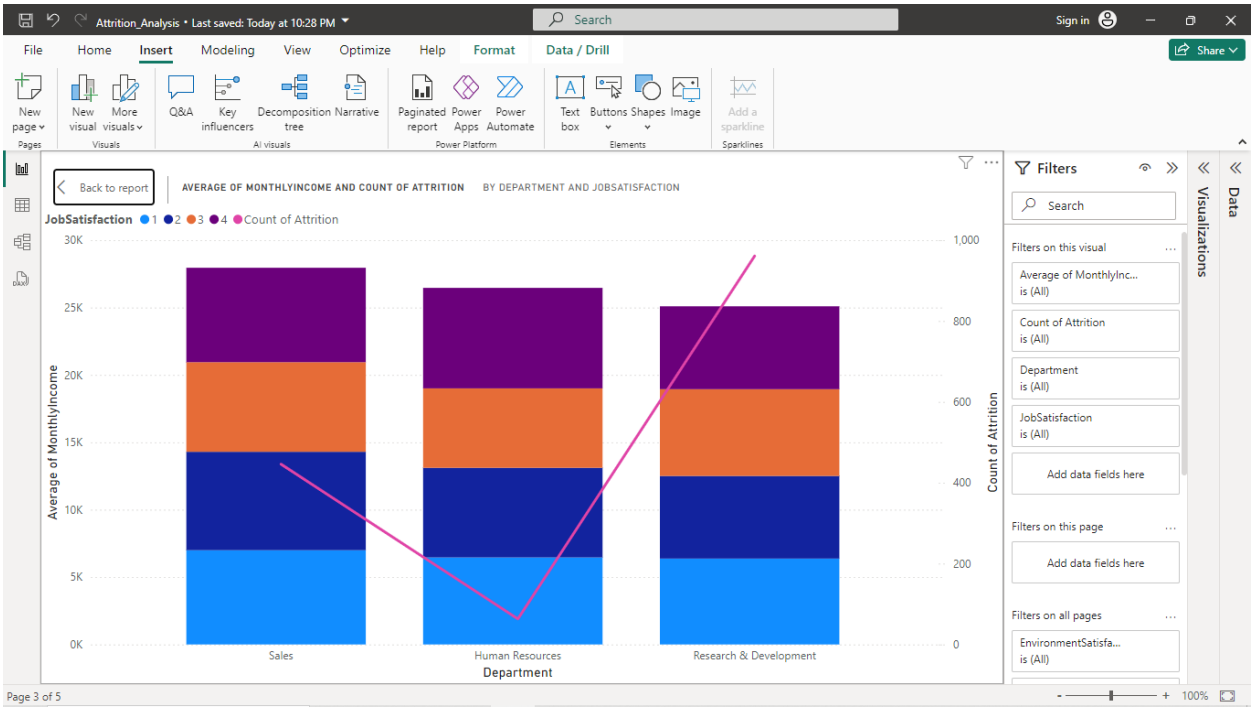

Figure 11:Monthly income Attrition by Department and Job Satisfaction

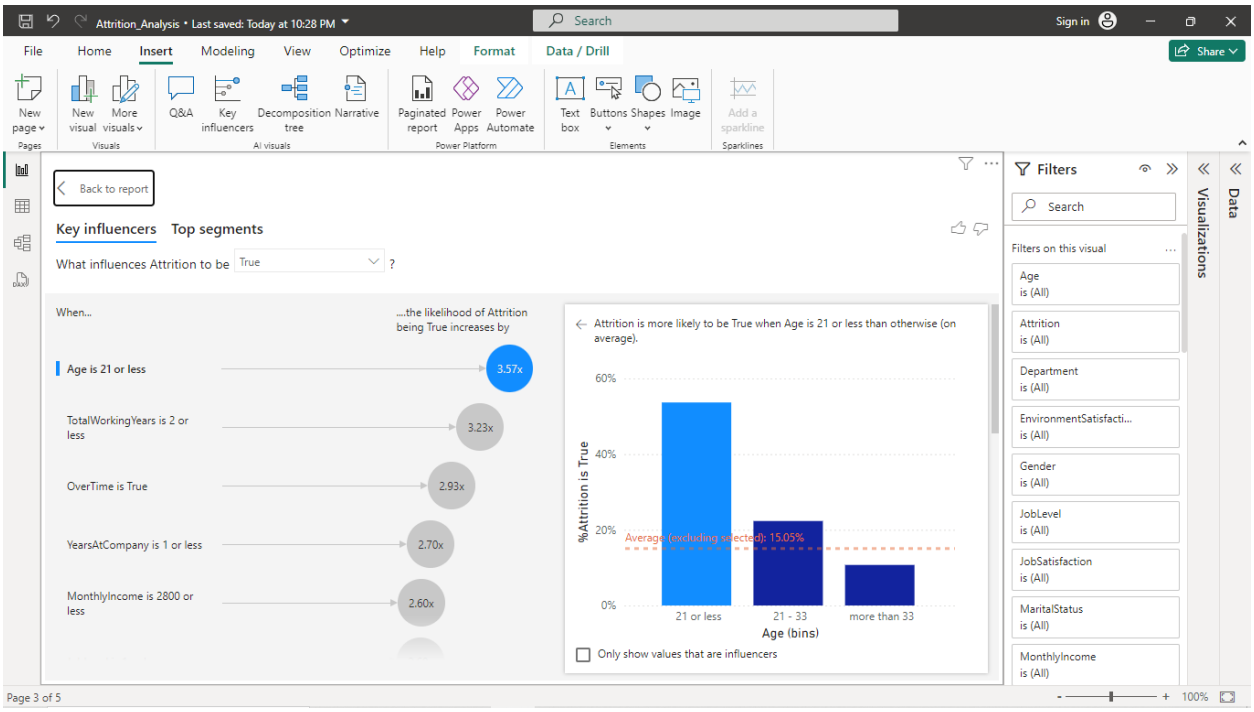

Figure 12:Monthly income Attrition by Department and Job Satisfaction_Key Value Influencers (a)

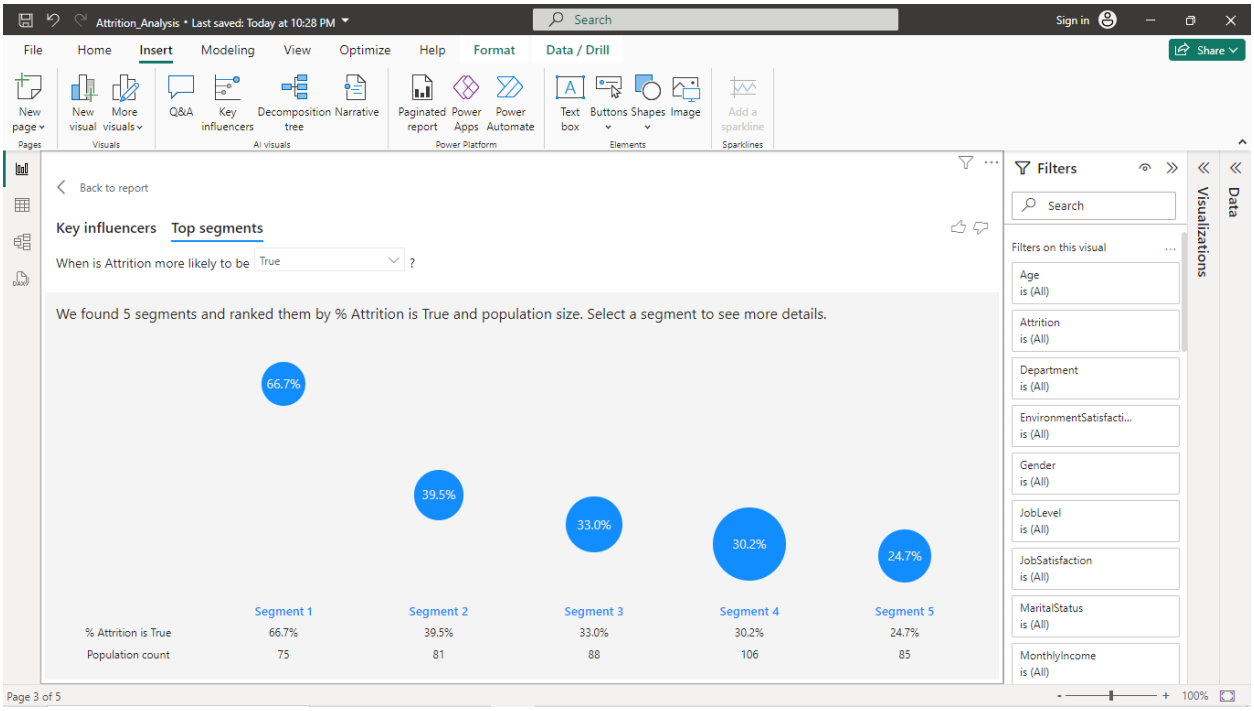

Figure 13:Monthly income Attrition by Department and Job Satisfaction_Key Value Influencers (b)

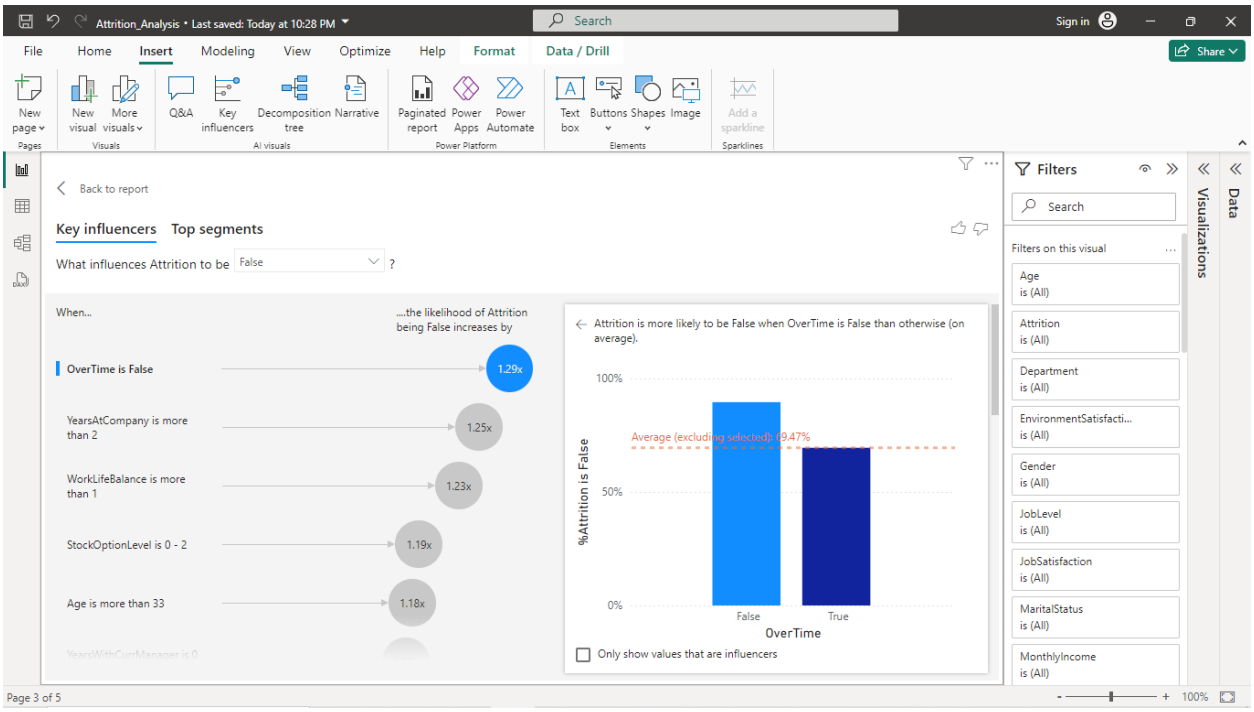

Figure 14:Monthly income Attrition by Department and Job Satisfaction_Key Value Influencers (c)

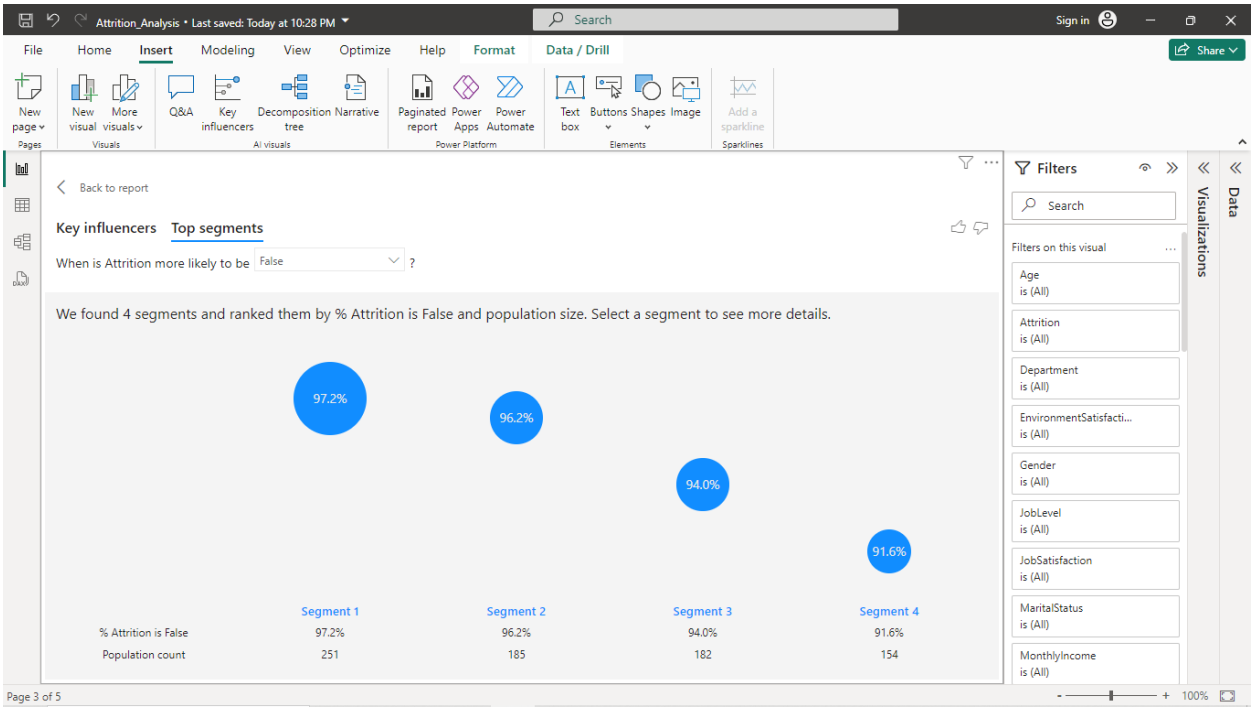

Figure 15:Monthly income Attrition by Department and Job Satisfaction_Key Value Influencers (d)

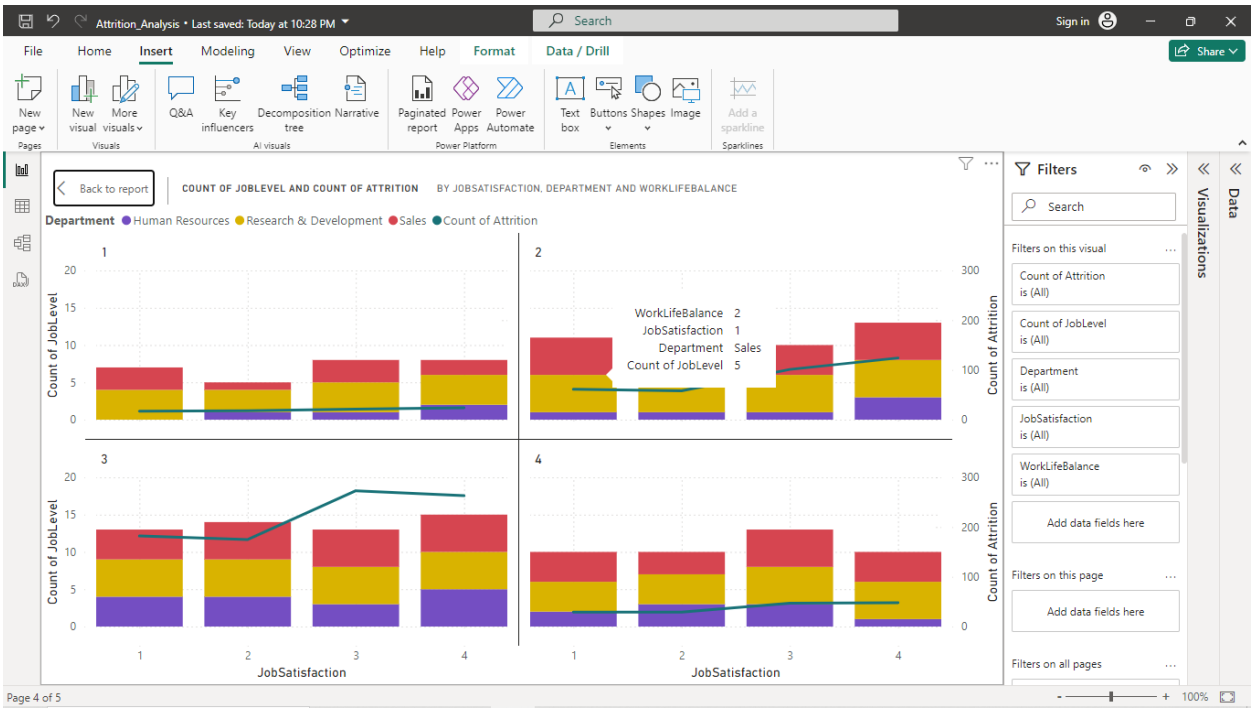

Figure 16:Count of Job Level and Count of Attrition by Job Satisfaction, Department and Work-Life Balance

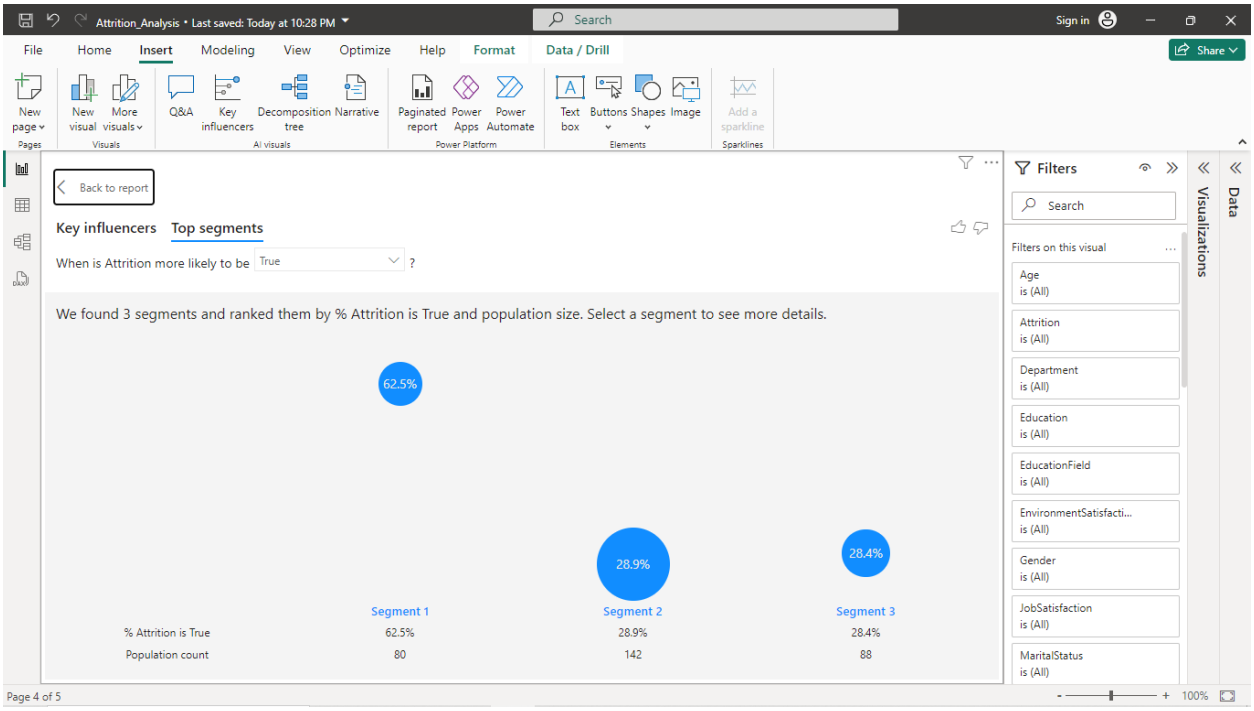

Figure 17:Count of Job Level and Count of Attrition by Job Satisfaction, Department and Work-Life Balance Key Value Influencers (a)

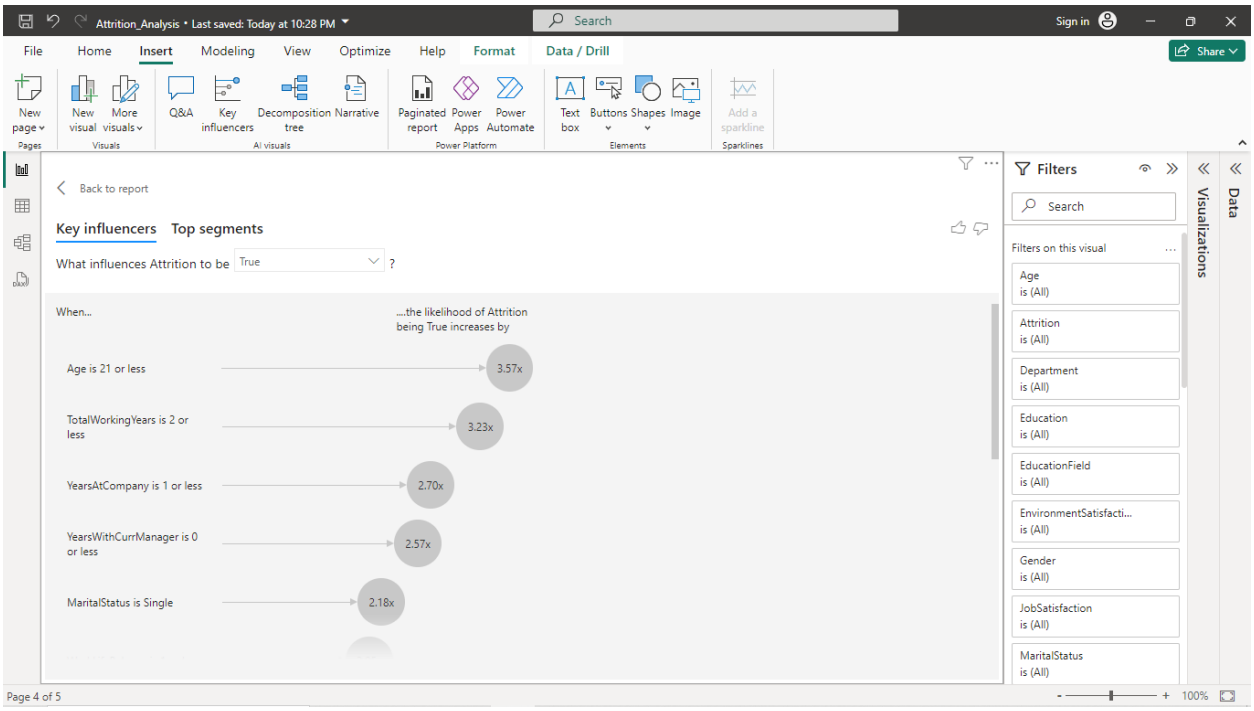

Figure 18:Count of Job Level and Count of Attrition by Job Satisfaction, Department and Work-Life Balance Key Value Influencers (b)

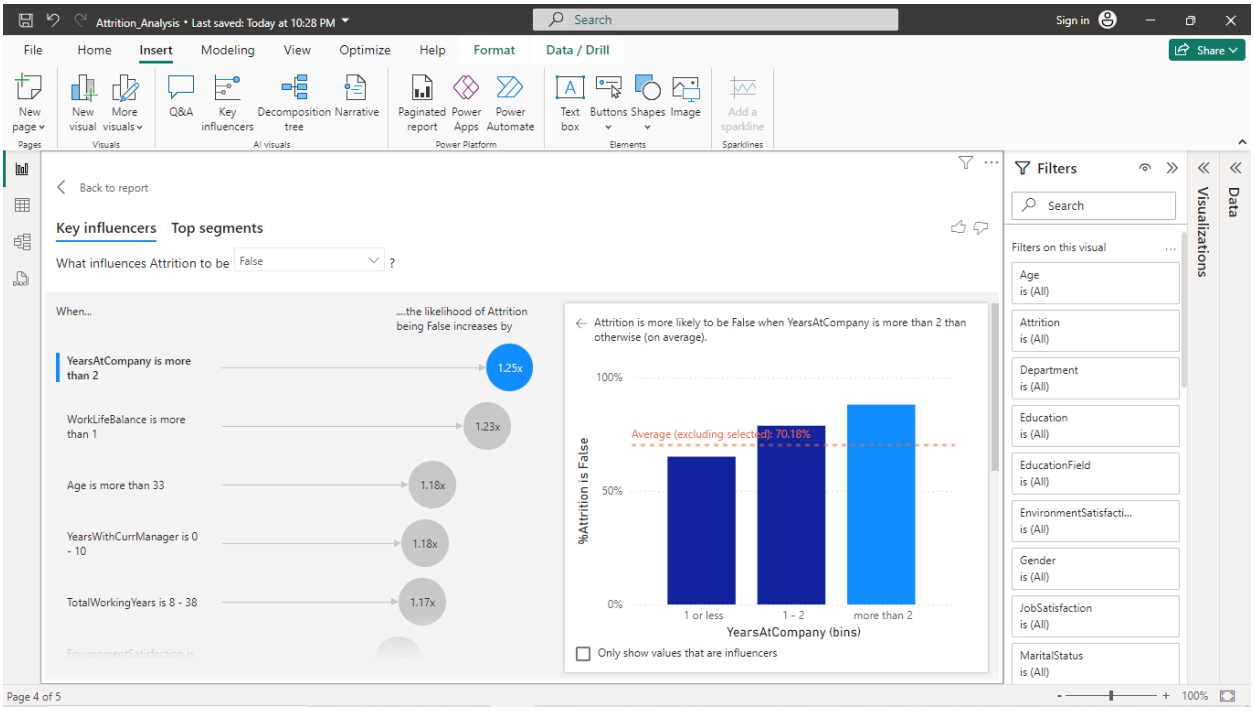

Figure 19:Count of Job Level and Count of Attrition by Job Satisfaction, Department and Work-Life Balance Key Value Influencers (c)

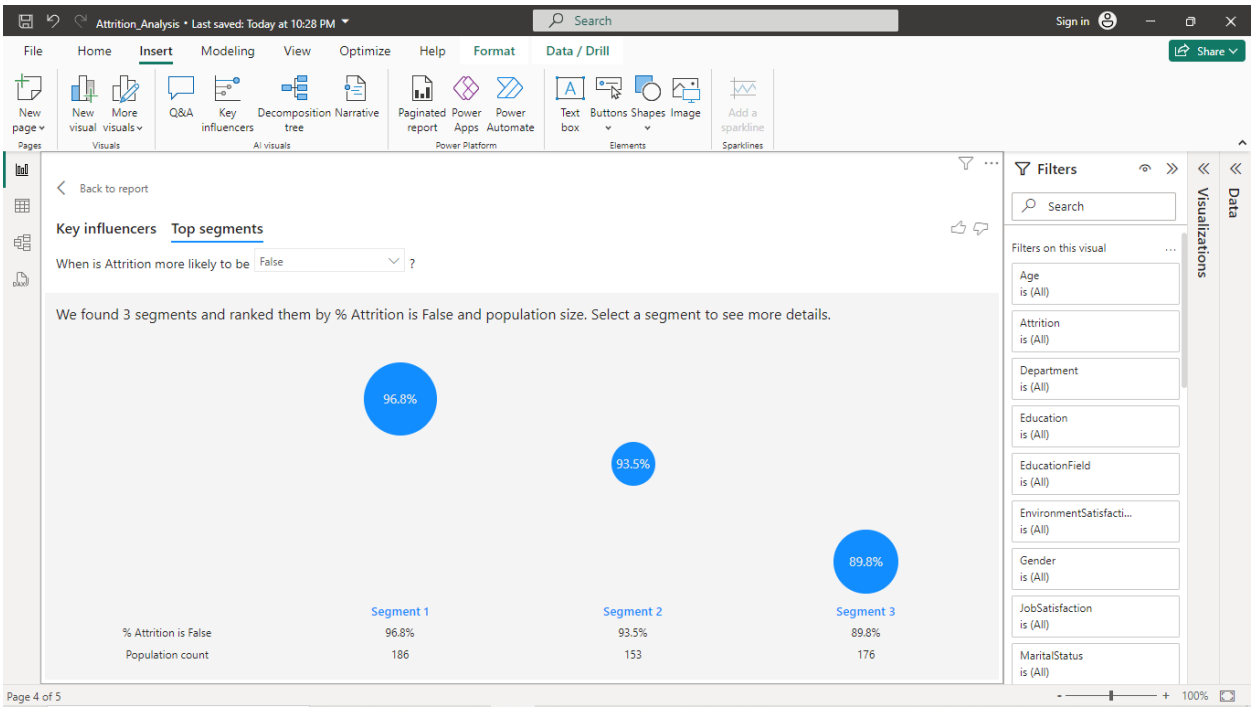

Figure 20:Count of Job Level and Count of Attrition by Job Satisfaction, Department and Work-Life Balance Key Value Influencers (d)

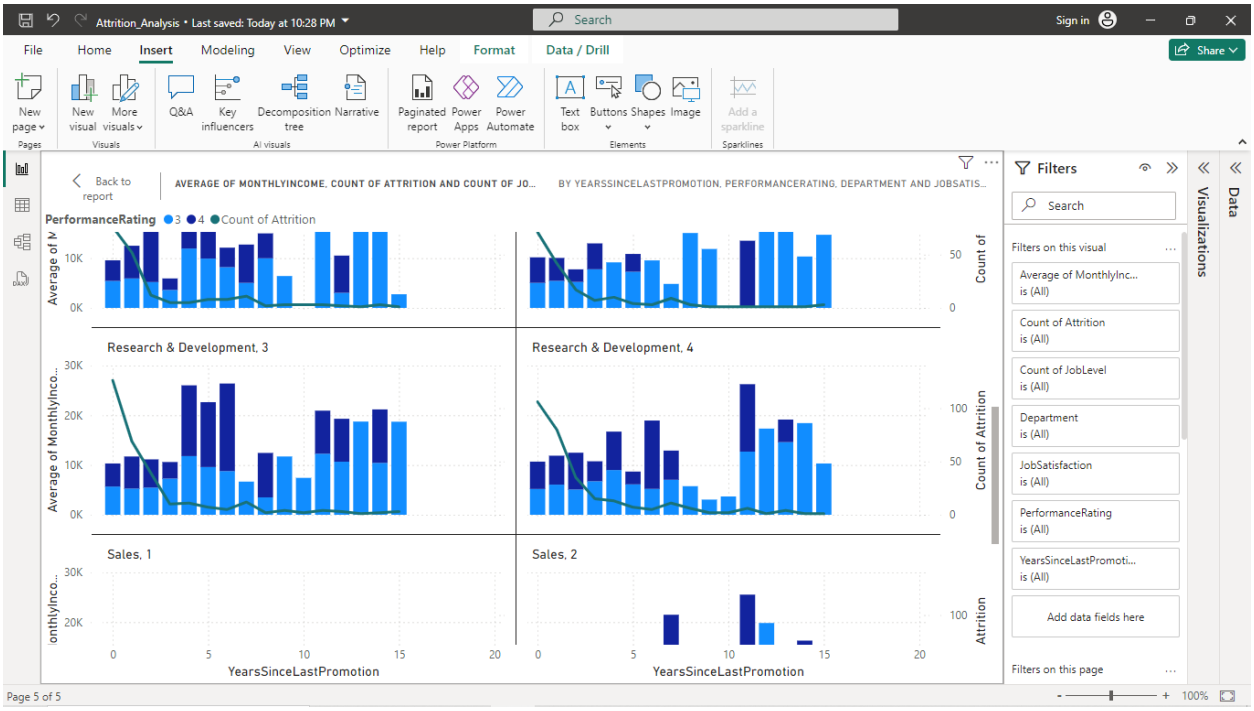

Figure 21:Average pf Monthly Income, Count of Attrition and Count of Job Level by Years Since Last Promotion, Performance Rating, Department and Job Satisfaction

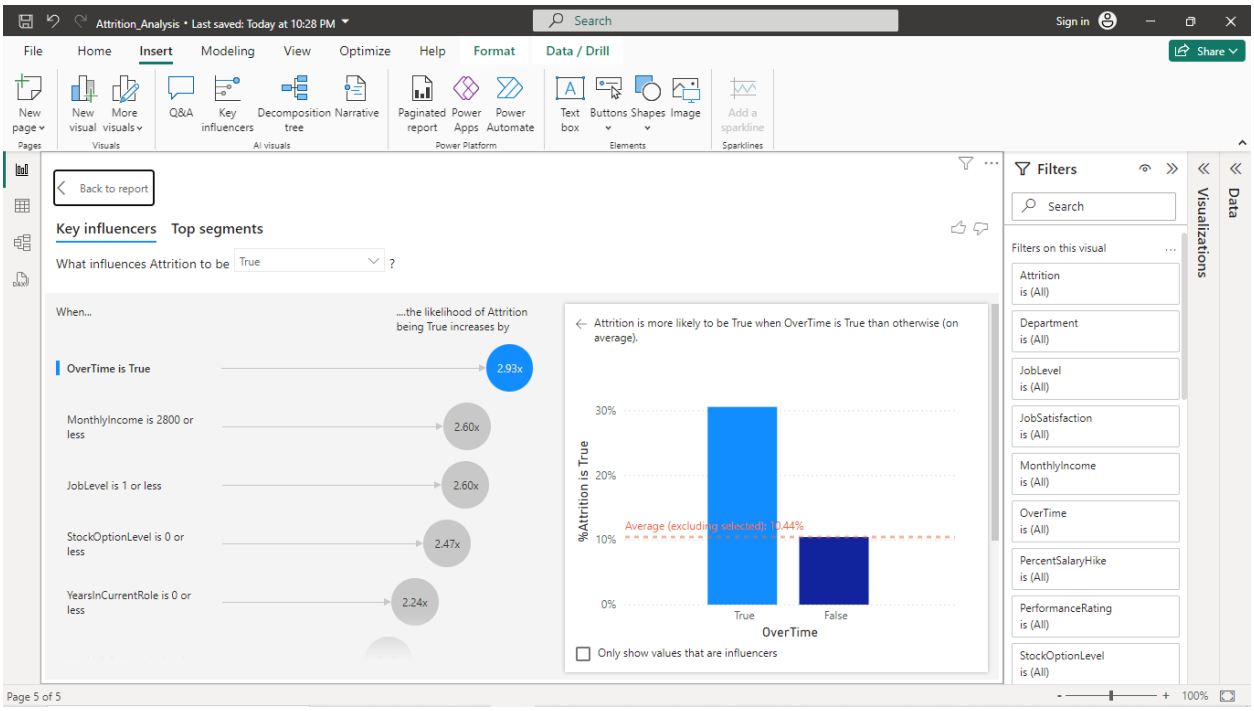

Figure 22:Average pf Monthly Income, Count of Attrition and Count of Job Level by Years Since Last Promotion, Performance Rating, Department and Job Satisfaction_Key Value Influencers (a)

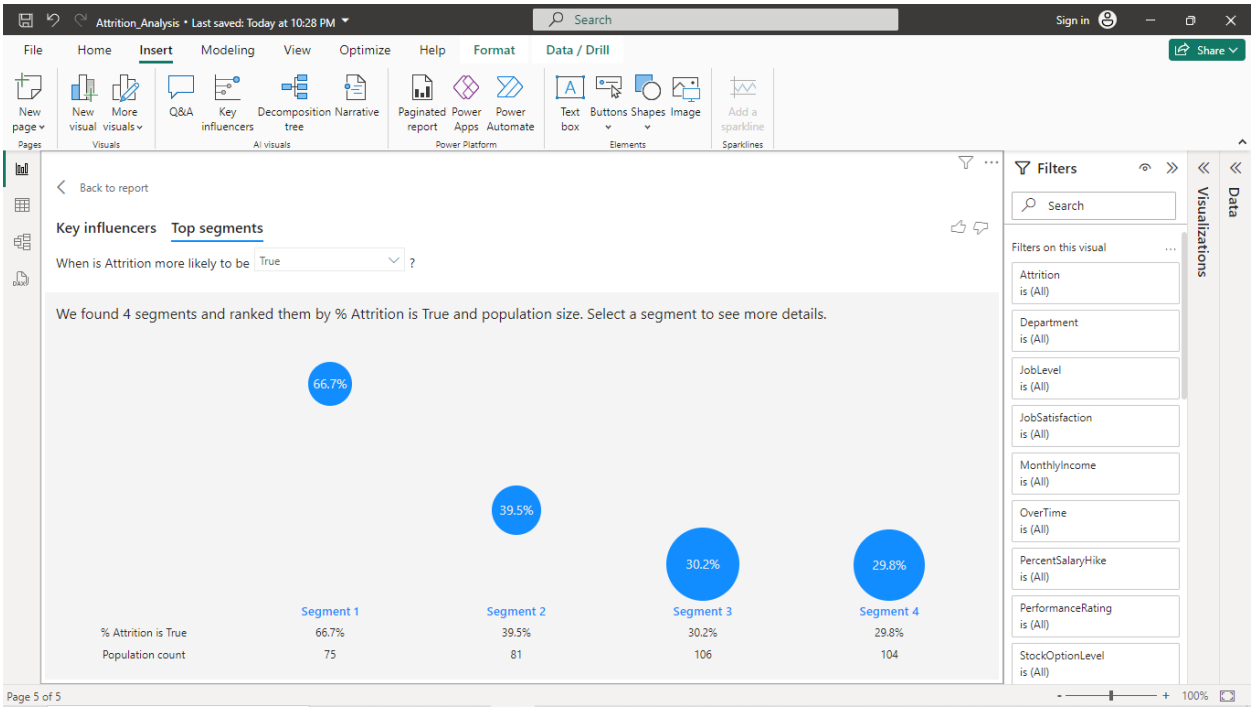

Figure 23:Average pf Monthly Income, Count of Attrition and Count of Job Level by Years Since Last Promotion, Performance Rating, Department and Job Satisfaction_Key Value Influencers (b)

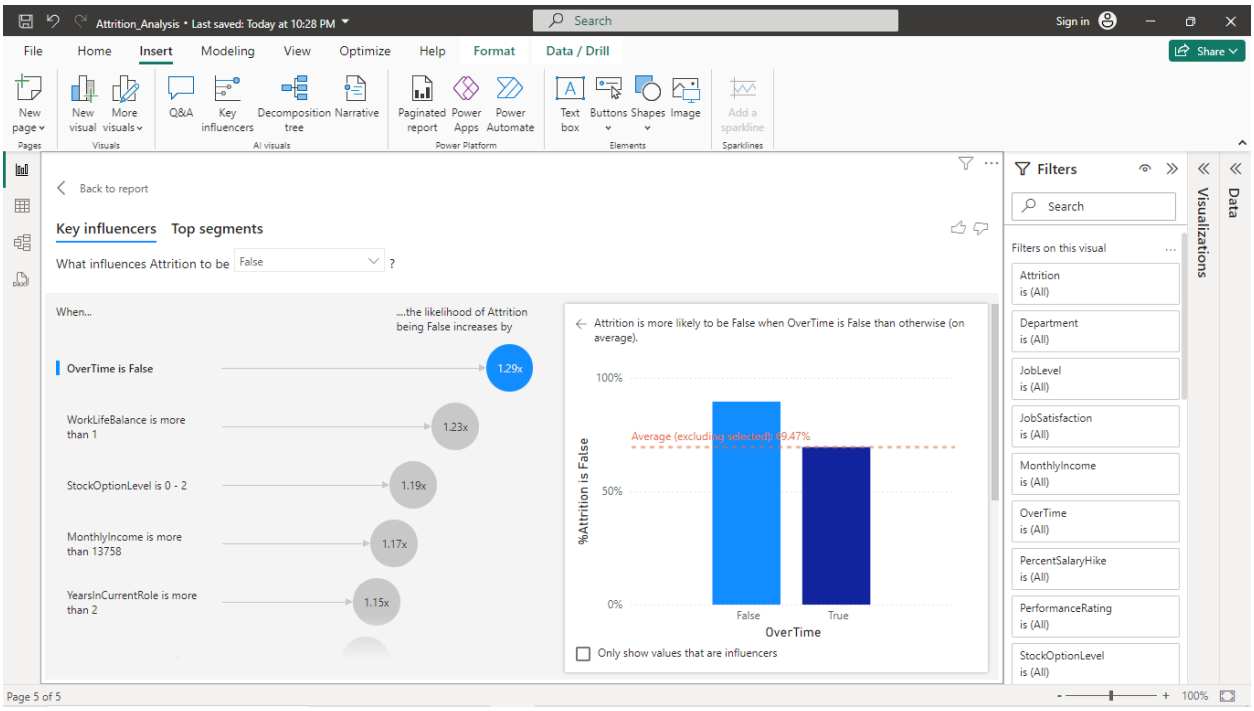

Figure 24:Average pf Monthly Income, Count of Attrition and Count of Job Level by Years Since Last Promotion, Performance Rating, Department and Job Satisfaction_Key Value Influencers (c)

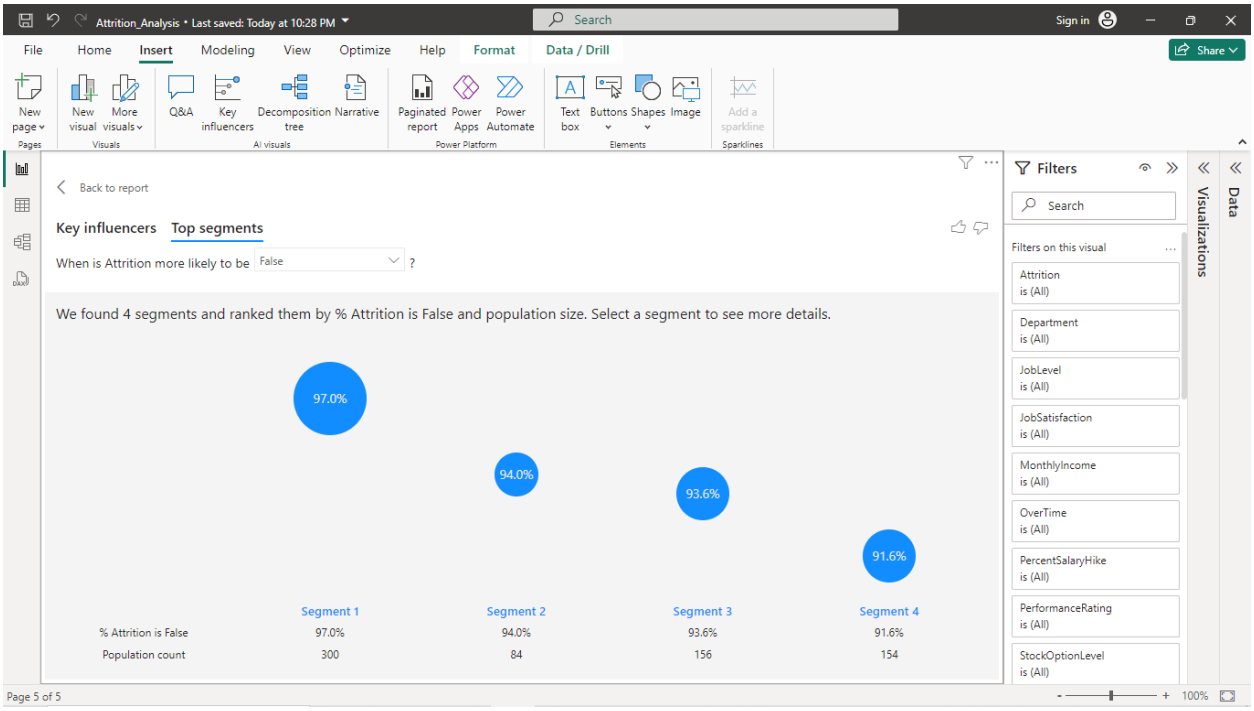

Figure 25:Average pf Monthly Income, Count of Attrition and Count of Job Level by Years Since Last Promotion, Performance Rating, Department and Job Satisfaction_Key Value Influencers (d)

Kindly note that the above figures are actual representations of the visuals, and they depict the relations among different factors such as job satisfaction, environment satisfaction, marital status and average monthly income, to list a few. A detailed study can be made by studying and discussing the details in visualisations.

Resources

List of Resources Accessible During the Project

- GitHub Repository
- Technocolabs Softwares GitHub repository to Download Dataset From
- Unhindered Support from Technocoloabs Softwares
- Python
- MS Excel
- MS SQL SERVER
- MS Power BI
- MS Word
- WSL2

Timeline for Execution

Details of Timeline of Execution

Key project dates are outlined below. Dates are best-guess estimates and are subject to change until a contract is executed.

The following table shows how the tasks were divided and how they were executed.

Description	Start Date	End Date	Duration
Project Start	June 1, 2024	June 1, 2024	
Milestone 1: Understanding GitHub	June 1, 2024	June 1, 2024	3 Hours
Milestone 2: Downloading Data Identifying Primary Key and Storing Column Names in an Excel file.	June 2, 2024	June 2, 2024	1 Hour
Phase 1 Complete	June 2, 2024	June 4, 2024	3 Days
Milestone 3: Making Diagrams Possible Tables in Excel	June 2, 2024	June 3, 2024	1.5 Days
Milestone 4: Executing SQL Queries to Make Tables	June 3, 2024	June 4, 2024	1.5 Days
Phase 2 Complete			4 Days
Milestone 5: Identifying Possible Outliers and Inferring Their Possible Reason	June 4, 2024	June 5, 2024	2 Days
Milestone 6: Making Visual Representation of Data with Power BI	June 5, 2024	June 6, 2024	2 Days
Milestone 7: Writing Detailed Project Report and Pushing it to GitHub	June 6, 2024	June 7, 2024	1.5days
Project End	June 7, 2024	June 7, 2024	

Supplied Material

The following materials were supplied by Technocolabs Softwares beforehand along with the problem statement through GitHub.

Materials Supplied by Technocolabs Softwares	Due Date*
Dataset of Acme’s Employee Detail	June 1, 2024

Anurag Sharma
Intern