**Online Analytical Processing (OLAP) Model for Multi-Dimensional Analysis of HR Data**

*College of Computing and Software Engineering (CCSE), Kennesaw State University  
Marietta, Georgia*

Muhammad Sohel - msohel@students.kennesaw.edu

Tyler Sutherland - tsuther4@students.kennesaw.edu

Rushitha Buddharaju - rbuddhar@students.kennesaw.edu

***Abstract*— Human Resource departments all over the world have to deal with the business analytics of employee data, a lot of these departments have teams just for analyzing these metrics. Data warehousing techniques such as OLAP and ETL are essential for handling the data, hoping to get insights into employee turnover, performance trends, and compensation patterns. This paper aims to create a OLAP model for multi-dimensional analysis using data warehousing techniques to get useful insight on the data. Popular datasets will be used and the model will be evaluated to standards.**

***Keywords***— **OLAP, HR Analytics, Business Intelligence, BI, Cube**

1. ***Introduction***

In the world of human resource (HR) analytics, the goal almost always includes the ability to better understand employees, interpret data, and ultimately improve business performance. According to the Academy to Innovate HR (AIHR), HR analytics “involves gathering, analyzing, and reporting HR data to drive business results” (Vulpen, 2024). Deriving and interpreting data such as employee data of turnover rates, performance trends, and compensation patterns allow HR departments to make important business decisions that can help the business be successful.

Creating a system that can store and aid in analyzing data from different perspectives or dimensions can lead to a better understanding and decision making based on the data. Which is why we proposed an Online analytical processing (OLAP) system to provide practical insights to make informed decisions.

1. ***OLAP Systems***

Online analytical processing (OLAP) systems are used for “performing high-speed complex queries or multidimensional analysis on large volumes of data in a data warehouse, data lake or other data repository” (IBM, 2024). An OLAP system allows for processing and analyzing multiple data dimensions more efficiently and faster than a regular relational database. It does this by adding layers (dimensions) onto tables, ultimately creating a cube. This cube is the multi-dimensional data that can now be used for us to analyze and gain our insights from.

1. ***Related Works***

A related study by Hamoud and others focused on creating an HR data mart using a dataset of 484 retired employees of an Oil Company, with data on demographics, job details, and compensation. The data was imported into an SQL Server for cleaning and integration. A data mart is chosen over a data warehouse due to the small, departmental scope, with potential for future expansion and use. The implementation involves a bottom-up approach while focusing on efficient ETL processes and the method of Slowly Changing Dimensions (SCD). The final product is an OLAP cube that enables fast, multidimensional analysis to support HR decision-making for the company.

In the study of OLAP Analytical Solution for Human Resource Management by Debeljacki and others, they created a system for measuring and assessing HR management (HRM) performance to help in improving HR processes. They did this by enabling efficient tracking and analysis of HR data itself. They analyzed data such as employee motivation, productivity, compensation, training, and performance assessment. The model's architecture includes data sources, ETL processes, a data warehouse, and OLAP software for creating multidimensional reports and analyses. Data included employee attrition, compensation, recruitment, workforce development, and productivity, with KPIs like attrition rate, average salary, hiring rate, training costs, and overall productivity. The OLAP-based system provided ad-hoc reports, KPI monitoring, and a dashboard.

1. ***Methodology***
2. *Dataset*

The IBM HR Analytics Employee Attrition & Performance dataset is a “fictional dataset created by IBM data scientists, containing various demographic, qualitative, and quantitative variables about individual employees” (Ehlert-Mackie, 2023). Columns include information on employee demographics, job roles, performance ratings, satisfaction levels, and attrition status. Key attributes include education, environment satisfaction, job involvement, job satisfaction, performance rating, relationship satisfaction, and work-life balance, each with categorical values representing various levels of satisfaction and involvement.

1. *Data Pre-processing*

During our preprocessing phase of our project we had to take some steps in order to get our data to the correct position for us to start trying to answer our 3 questions. We first removed some irrelevant variables from our data that either had only unique values or only one value was presented throughout that variable in the dataset. We don’t want to generate patterns for values that only have one unique value to them. For instance in our data we have a variable called “Over18” this value is always presented in the dataset with a 1 which isn’t helpful for us when trying to determine patterns. In the case of attrition we could see a pattern between not leaving because they are 18 or over, but that pattern is created because of the amount of people in total that are 18 years old.

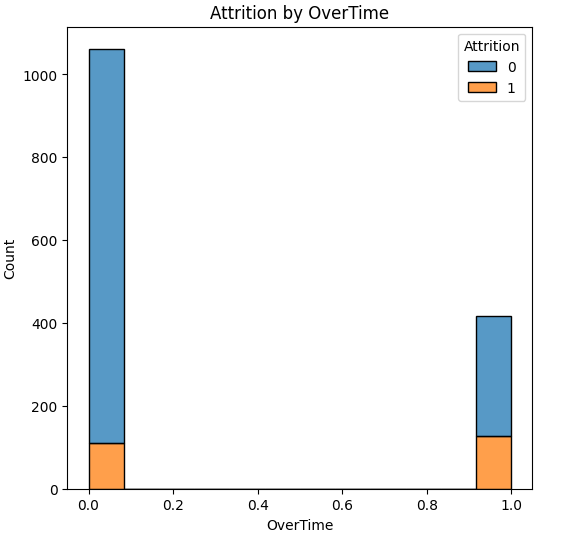
The next step in our preprocessing stage we did was to convert from string variables such as (“Single”, “Married”, “Divorced”) to int variables like (0,1,2). This conversion acts as if we swap the string out for a foreign key value.We did these conversions so that when forming OLAP cubes, correlation matrix, charts, and heatmaps it would be easier than trying to do later and on the fly. This also allows us to incorporate more correlations than just removing the string values from our dataset.

1. *Implementation*

We decided to split up our three main problems into their own independent implementation. Our problem categories are Attrition (leaving company), Performance Rating, and Compensation (monthly income). All of these problems follow a similar implementation structure with only minor differences between each of the problems.

First, we find correlations between our primary variables we are testing for and our other variables. We do this by creating a correlation matrix and looking at each of our variables to see which values are close to one. Being close to one means that it is highly correlated with our primary variable while a zero is the opposite. Once we have that correlation matrix we can then create our heatmap to more visually see how variables are being correlated to each other.

Second, we now want to do more direct comparisons with our primary variables and our other variables. We did this by creating histogram charts for each of the comparisons to look for correlation between our primary variable and each of our correlatable variables. For instance when looking at attrition as it relates to overtime we found that employees that work over time are more likely to leave their job versus those who do not work overtime.



***Histogram 1: Attrition by Overtime***

Third, we want to evaluate our hypothesis results that we came to from our charting phase. Here we group our comparisons and look at the number counts to confirm our hypothesis. We do this by grouping our primary value and a secondary value together and getting their sizes of each grouping. For example for overtime and attrition, if they worked overtime and left their job they made up 30% of those who worked overtime, versus if they didn’t work overtime and still left, that grouping only made up 10% of the total of those who didn’t work overtime. This difference is substantial because there is a 20% gap between them.

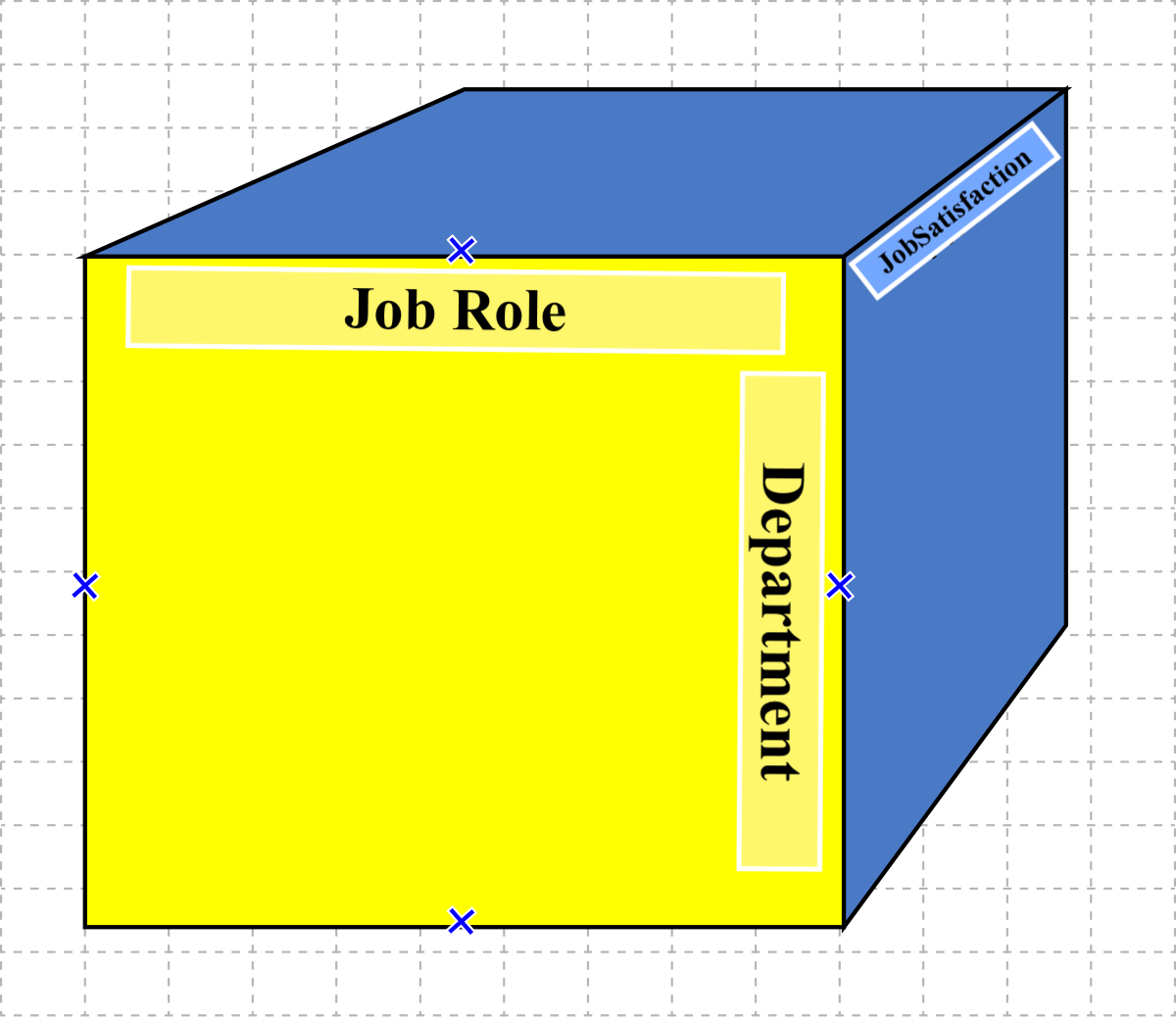
| Overtime | Attrition | Count | % |
| --- | --- | --- | --- |
| NO | NO | 952 | 89.64 |
| NO | YES | 110 | 10.36 |
| Total of NO Overtime | | 1062 |  |
| YES | NO | 290 | 69.38 |
| YES | YES | 128 | 30.62 |
| Total of Overtime | | 418 |  |

***Table1 : Evaluating Numbers of Overtime and Attrition***

1. *OLAP Cubes*

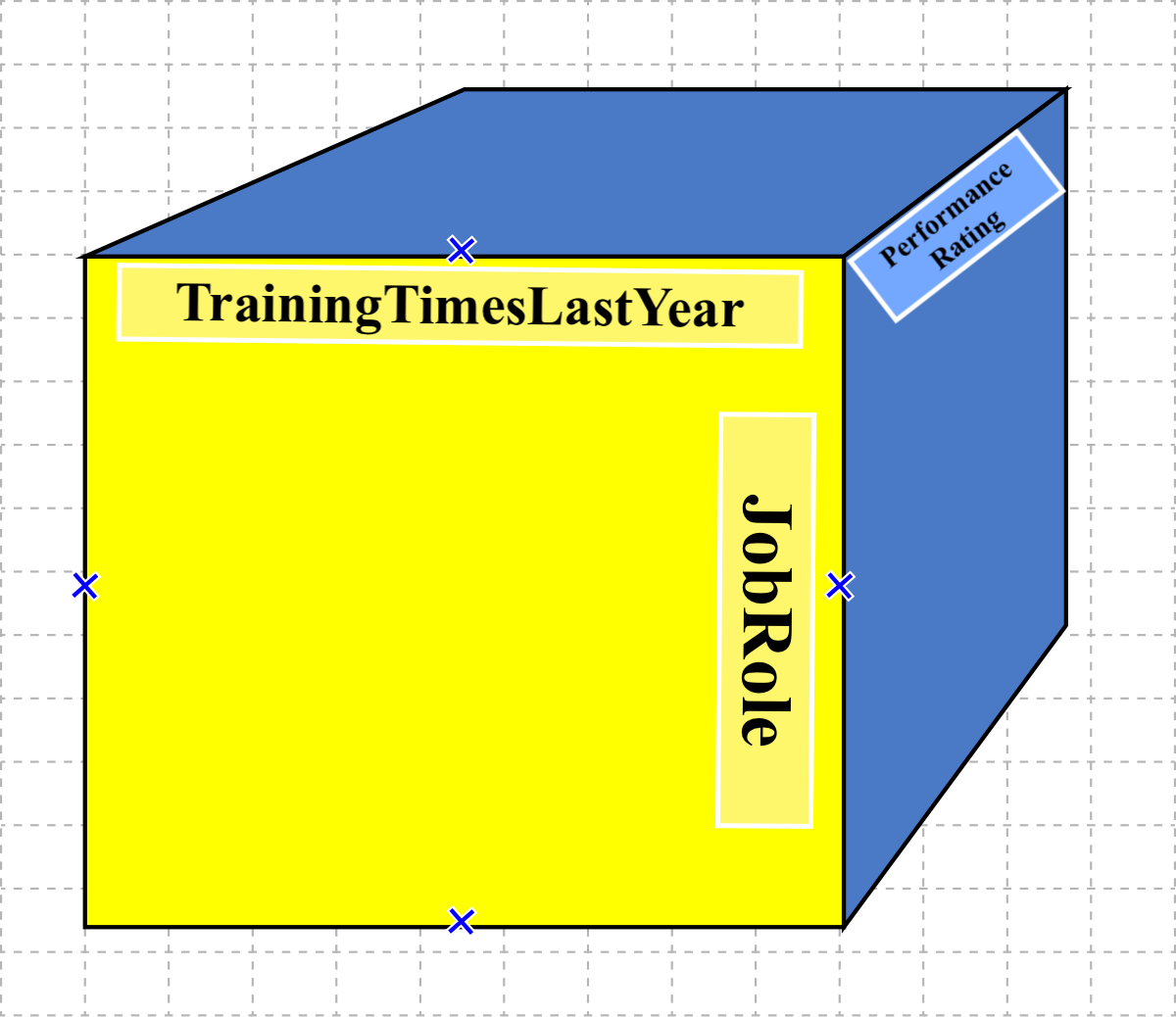
Visualizing some of the OLAP cubes along with their dimensions and queries can provide meaningful insight to what we aimed to accomplish and extract with this model. Understanding the different dimensions along the (X,Y,Z) coordinates and performing OLAP operations such as roll up, drill down, slice, dice, and pivot can give us results that won’t usually be seen in a regular 2-dimensional table or model.

In the first cube, the objective is to analyze job satisfaction across different job roles and departments, with the x-axis being the job role, y-axis being the department, and the z-axis representing job satisfaction. The main query this cube would help us solve is which job roles in which departments have the highest and lowest job satisfaction. Hoping to provide insight to the HR department on job satisfaction across the company.

**

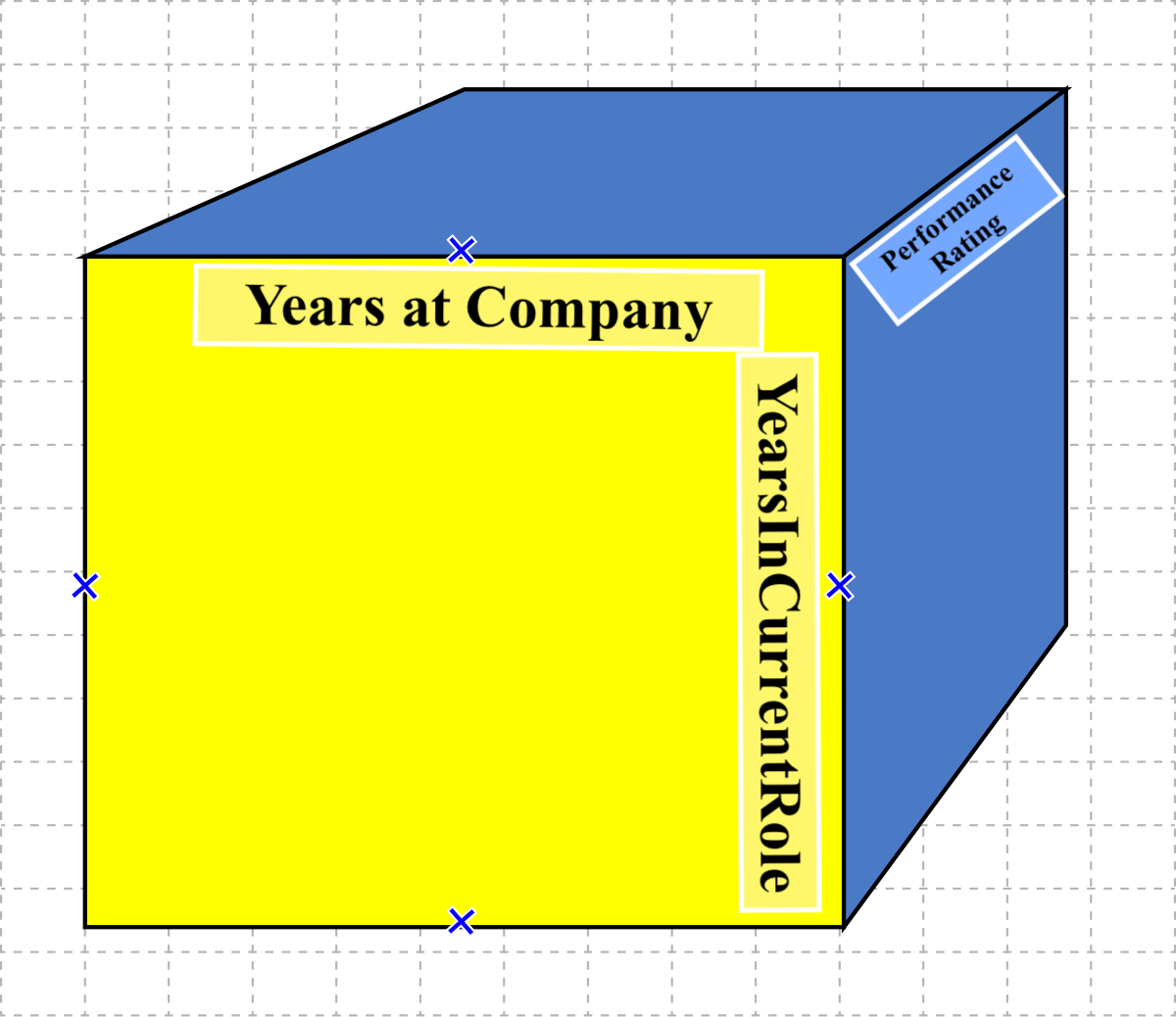
***Cube 1: Job Satisfaction Analysis***

In the cube below, we hope to ​​assess the impact of training on employee performance ratings. The x-axis represents the amount of training times in the past year, the y-axis shows job role, and z-axis represents the performance rating. The query will fall along the lines of evaluating if employees who receive more training have higher performance ratings across different job roles in the company. Giving HR insight on how training affects job performance across different jobs.

**

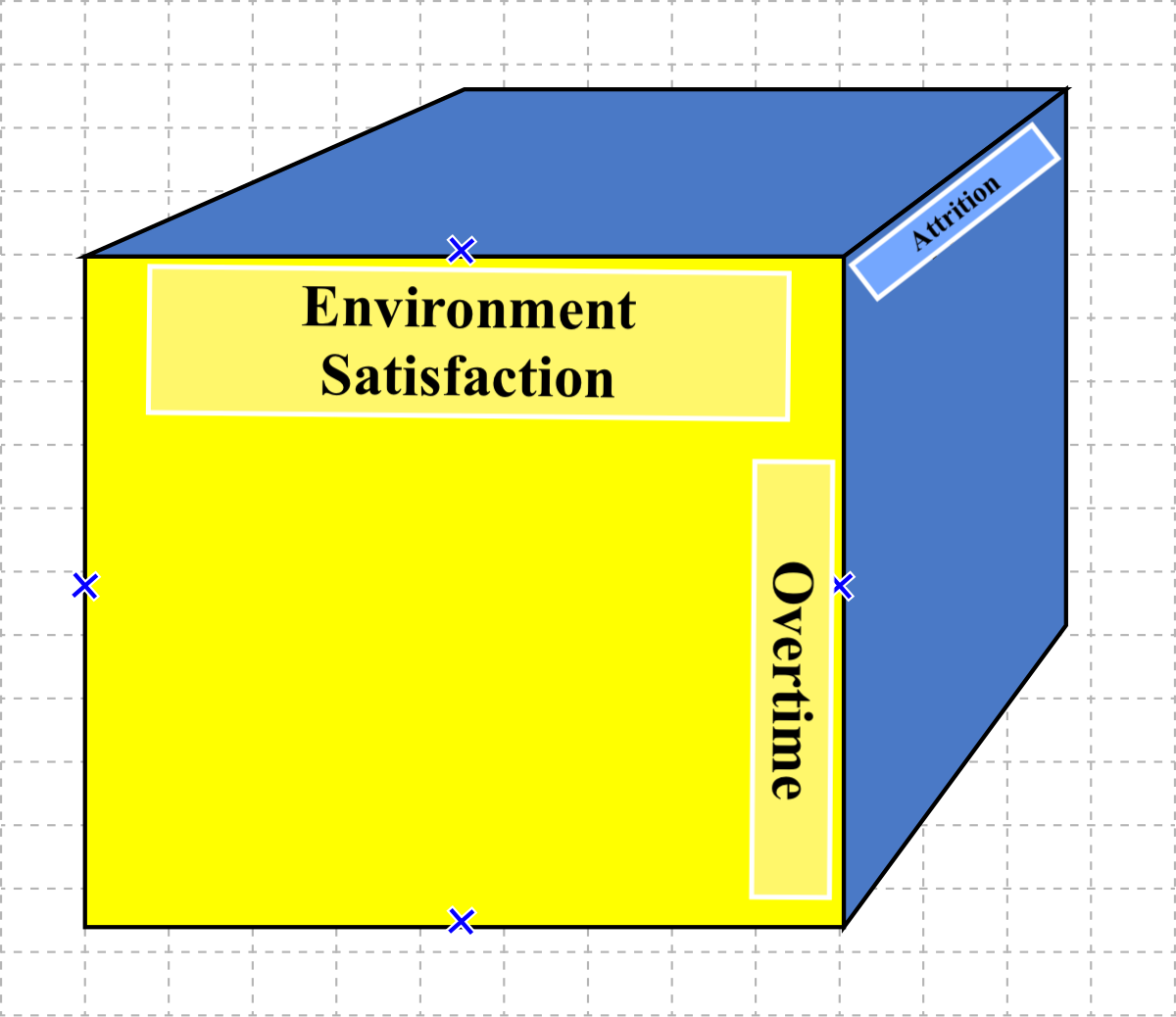
***Cube 2: Training and Performance***

The following cube evaluates performance ratings and its relationship with the number of years employees have been at the company and in their current job roles. The x-axis represents the years an employee was at the company, the y-axis shows their years in their current role, and the z-axis shows their performance rating. The query here would be to analyze if employees with longer tenure or time at the company have higher performance ratings. Giving insight to whether tenure influences performance.

**

***Cube 3: Performance Rating by Tenure***

In the final cube below, we aim to analyze the attrition rate in relation to environment satisfaction and overtime worked. Hoping to get insight on what influences attrition rates within the company. Our query here would be to understand if lower environment satisfaction at the job and working overtime contribute to higher attrition rates among employees.

**

***Cube 4: Attrition by Satisfaction and OverTime***

1. **Results**

Our analysis of employee turnover, derived from different heatmaps, correlation tables, correlation charts, and OLAP operations, reveals significant insights into the factors influencing attrition rates. We saw that employees who work overtime show a considerably higher attrition rate of 30.6%, compared to 10.4% for those who do not work overtime. We also saw that single employees are more likely to quit than when compared with married or divorced counterparts. Employees in lower job levels tend to leave more frequently, with Job Level 1 showing an attrition rate of 26.2%, while higher job levels show lower rates. Employees with lower total working years and those with 30 to 40 years at the company are both more prone to leave. Low monthly income significantly contributes to higher attrition, as does the absence of stock options. We also observed that male employees are more likely than females employees to leave.

Our data shows distinct patterns concerning job tenure and promotion history. Employees in their first three years in a current role with low monthly income have an attrition rate of 3.1%, which drastically increases to 36.9% for those with higher incomes. For those with 4-7 years in the current role, the attrition rate for low monthly income is 3.97%, and 12.9% for higher incomes. Employees with 8-11 years in the current role show an attrition rate of 7.2% for low income and 5.8% for higher income. Interestingly, those with 16-18 years in the current role and low income have an attrition rate of 38.5%.

When examining the time since an employee's last promotion, employees in their first three years with low income show an attrition rate of 3.1%, rising to 28.9% for higher incomes. The attrition rate for those with 4-7 years since the last promotion is 6.1% for low income and 12.6% for higher income. Employees with 8-11 years since the last promotion have an attrition rate of 10.6% for low income and 4.5% for higher income. Those with 12-15 years since the last promotion show a significant attrition rate of 26.2% for low income, but only 2.4% for higher income.

1. **Evaluation**

The evaluation of our results show the critical factors influencing employee turnover within the organization. The difference in attrition rates between employees who work overtime and those who do not underscores the impact of work-life balance on retention. Single employees showing higher attrition rates suggest the importance of personal life stability in job satisfaction. The correlation between lower job levels and higher turnover rates indicates that career progression opportunities might be limited, prompting employees to seek better prospects elsewhere. The analysis also reveals that employees with lower total working years and those in the late stages of their careers (30-40 years) are more likely to leave, emphasizing the need for tailored retention strategies for these groups. We also saw that monthly income influences attrition rates, this can correlate for the necessity of competitive compensation packages. The lack of stock options leading to higher turnover suggests that equity incentives could play a crucial role in employee retention. The higher attrition rates among male employees could indicate underlying issues that require further investigation, such as workplace culture or job satisfaction disparities.

The patterns observed in relation to job tenure and promotion history further convey the importance of understanding employee turnover. The sharp increase in attrition rates for employees in their early years in a current role with higher incomes indicates possible dissatisfaction despite higher pay, possibly due to lack of career growth or job satisfaction. The data also highlights the importance of regular promotions and career advancement opportunities for employees, as we saw higher attrition rates among those with more extended periods since their last promotion.

1. **Conclusions**

In conclusion, our comprehensive analysis of employee attrition rates show several critical insights that can inform the organization's retention strategies. Our analysis and the resulting model provide useful insights into the factors driving employee turnover. By identifying key determinants such as overtime work, marital status, job level, total working years, and compensation, our model highlights critical areas where intervention can significantly impact employee retention. The stark differences in attrition rates between employees who work overtime and those who do not underline the importance of promoting a healthy work-life balance. Companies can benefit greatly from implementing policies that minimize overtime and support a balanced lifestyle for their employees.

Overall, our model provides a robust baseline foundation for companies to understand and address the factors influencing attrition rates. By implementing the insights gained from this analysis, organizations can look to create a more supportive and engaging work environment, reduce attrition rates, and foster long-term success and stability. Companies that adopt our model or a similar approach can benefit from a deeper understanding of their workforce dynamics, ultimately leading to better retention and improved organizational performance.

***References***

1. Vulpen, E. van. (2024, May 31). *What is HR analytics? all you need to know to get started*. AIHR. https://www.aihr.com/blog/what-is-hr-analytics/
2. Hamoud, A. K., Abd Ulkareem, M., Hussain, H. N., Mohammed, Z. A., & Salih, G. M. (2020, May). Improve HR decision-making based on data mart and OLAP. In Journal of Physics: Conference Series (Vol. 1530, No. 1, p. 012058). IOP Publishing.
3. Debeljački, R., & Grljević, O. (2014). Olap Analytical Solution for Human Resource Management Performance Measurement And Evaluation: From Theoretical Concepts To Application.
4. What is OLAP?. IBM. (2021a, October 15). https://www.ibm.com/topics/olap
5. Ehlert-Mackie, C. (2023a, May 27). *IBM HR Analytics: Understanding employee data*. LinkedIn. https://www.linkedin.com/pulse/ibm-hr-analytics-understanding-employee-data-christy/
6. Moussa, M., Bright, M., & El-Gayar, O. (2017). "Customer analytics: Current trends and applications." Journal of Information Technology Management, 28(3), 1-1
7. Davenport TH, Harris J, Shapiro J. Competing on talent analytics. Harv Bus Rev. 2010 Oct;88(10):52-8, 150. PMID: 20929194.
8. Lazzari, M., Alvarez, J.M. & Ruggieri, S. Predicting and explaining employee turnover intention. Int J Data Sci Anal 14, 279–292 (2022). <https://doi.org/10.1007/s41060-022-00329-w>