

BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

Faculty of Business Administration in Foreign Languages

**Data-Driven Talent Management: The Blueprint for Building a High-Performing Organization**

**Authors:**

Cazan Ioana-Anca

Adam Estera-Lorena

Malecu Laura-Georgiana

Bucharest, 2024

**Introduction**

The competitive and dynamic business environment necessitates a data-driven approach to talent management.

This study aims to dive deeper into the connections between how to bring employees on board most efficiently and which attributes are essential for a successful employee retention strategy. Our objective for this analysis is to analyze our company’s situation from a Human Resources standpoint and assist the Board of Directors (BOD) into making the best decisions going forward. By digging into the intricate web of HR variables, this research seeks to provide valuable insights that can inform decision-makers and contribute to the development of effective retention and performance enhancement strategies in the contemporary organizational landscape.

**Research Question**

RQ: How can we leverage data-driven insights such as recruitment sources, departmental distribution, gender balance, employee tenure, satisfaction levels, engagement metrics, productivity measures, and compensation structures, to unlock strategies to elevate our overall performance and establish as a high-performing talent magnet?

**Method**

**Data Collection:**

* Collected HR data, from a consulting service company we work for, in the past year: this data includes employee demographics, recruitment sources, departmental distribution, tenure, productivity, satisfaction surveys, engagement surveys & salary data. The data has been of course anonymized.
* Ensured data quality and consistency by validating and cleaning the data.
* Dataset source:



**Data Analysis:**

* Conducted quantitative analysis using statistical methods in Spyder to identify correlations and trends between the variables of interest.
* Used visualization tools to present data findings in a clear and understandable manner.

**Interpretation & Recommendations:**

* Interpreted data findings and developed actionable recommendations based on the research outcomes.
* Provided insights into how the HR department can improve recruiting, retention, engagement, and employee performance based on the research findings.

**Dissemination:**

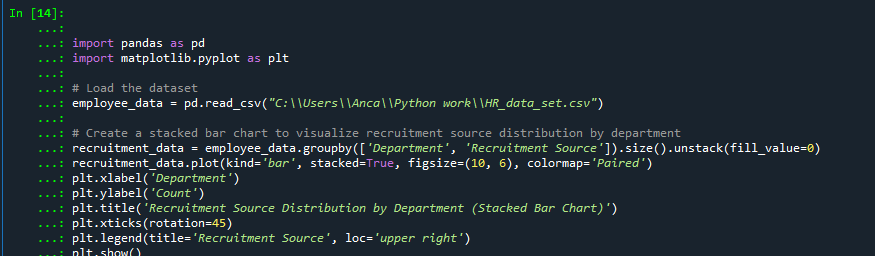
* Presented the research findings to Board of Directors to recommend talent management strategies.

**Description, Analysis and Methodology**

Going from abstract to specific into our analysis, the first variables in question are the recruitment source and department.

More specifically, we want to answer the following question:

***“Which recruitment sources are most effective for each department?”***



* We have imported the libraries (panda to read data from csv, Matplotlib for graphs & charts) and our dataset.
* Defined which data to be included (Department, Recruitment Source, then counted the number of employees in each group. The unstack() function is used to convert the resulting DataFrame into a more readable format.
* Created a stacked bar chart to visualize the recruitment source distribution by department. The kind='bar' parameter specifies that a bar chart should be created. The stacked=True parameter specifies that the bars should be stacked on top of each other. The figsize=(10, 6) parameter sets the figure size to 10 inches by 6 inches. The colormap='Paired' parameter specifies that the paired colormap should be used to color the bars.
* Set the x-axis label to "Department “& y-axis label to "Count".
* Set the title of the chart to "Recruitment Source Distribution by Department (Stacked Bar Chart)".
* Rotated the x-axis labels by 45 degrees to make them more readable.
* Added a legend to the chart, with the title "Recruitment Source". The loc='upper right' parameter specifies that the legend should be placed in the upper right corner of the chart.
* Displayed the chart.



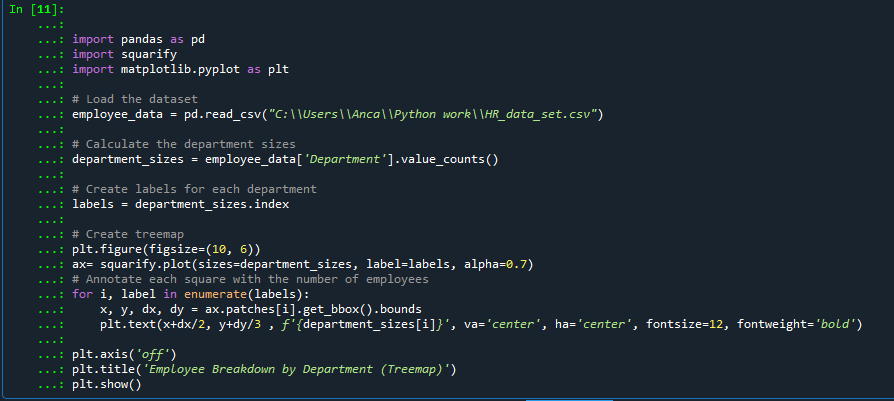
Our analysis & findings:

* IT department has the highest percentage of employees who were recruited through referrals. This suggests that the company's employees are particularly well-connected in the IT industry.
* Sales & Marketing departments have the highest percentage of employees who were recruited through online job platforms. This suggests that the company should focus more on using these channels for future recruitments for the two departments, but also should consider increasing the number of referrals.
* Overall, the graph suggests that the company is using a variety of recruitment sources to attract a diverse group of talent, but there is room for improvement.

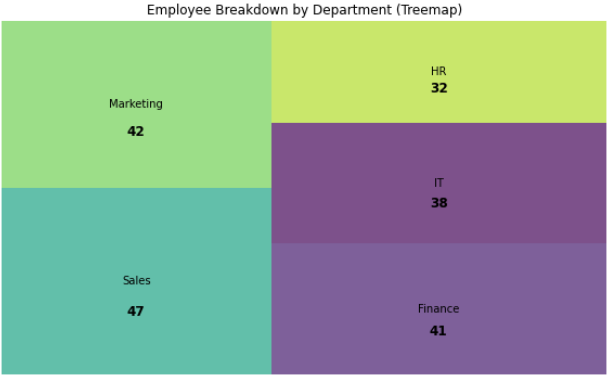
Going forward, it is important to understand whether our consulting company is balanced & well-structured across departments or whether our organization has in focus only certain departments where it tends hires more.

Therefore, the next question that arises is:

***“What is the number of employees in each department? Is it evenly structured or not?”***



* We have imported the libraries (panda to read data from csv, squarify to create treemeap, Matplotlib for graphs & charts) and our dataset.
* Calculated the department sizes by counting the number of employees in each department.
* Created labels for each department using the index attribute of the department\_sizes Series.
* Created a treemap using the squarify.plot() function. The sizes parameter specifies the sizes of the squares in the treemap. The label parameter specifies the labels for the squares in the treemap. The alpha parameter specifies the transparency of the squares in the treemap. Alpha 0.7 value of 0.7 is a common choice for treemaps, as it provides a good balance of visual appeal and clarity.
* Annotated each square in the treemap with the number of employees in the corresponding department. The enumerate() function is used to iterate over the labels and the corresponding indices. The ax.patches[i].get\_bbox().bounds property returns the bounding box of the i-th square in the treemap. The plt.text() function is used to add text to the treemap
* Turns off the axes of the treemap.
* Sets the title of the treemap to "Employee Breakdown by Department (Treemap)".
* Displays the treemap.



Our analysis & findings:

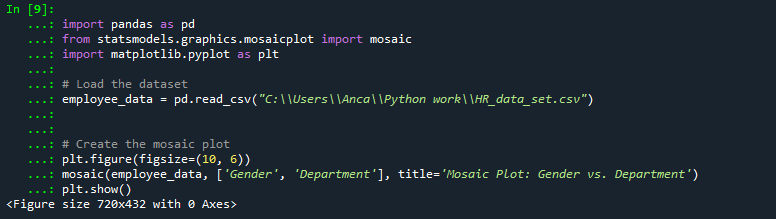
* Our company's departmental distribution aligns with its identity as a consulting service firm. The relatively even distribution of employees across IT, Finance, HR, Marketing & Sales reflects the company's focus on providing comprehensive professional services to its clients. This also suggests that our company is well-diversified in terms of its employee skillsets and expertise.

As a company, it is important to be balanced and well-structured so that the resources are allocated optimally, the decision making, adaptability and resilience improves while the culture enables enhanced innovation and creativity.

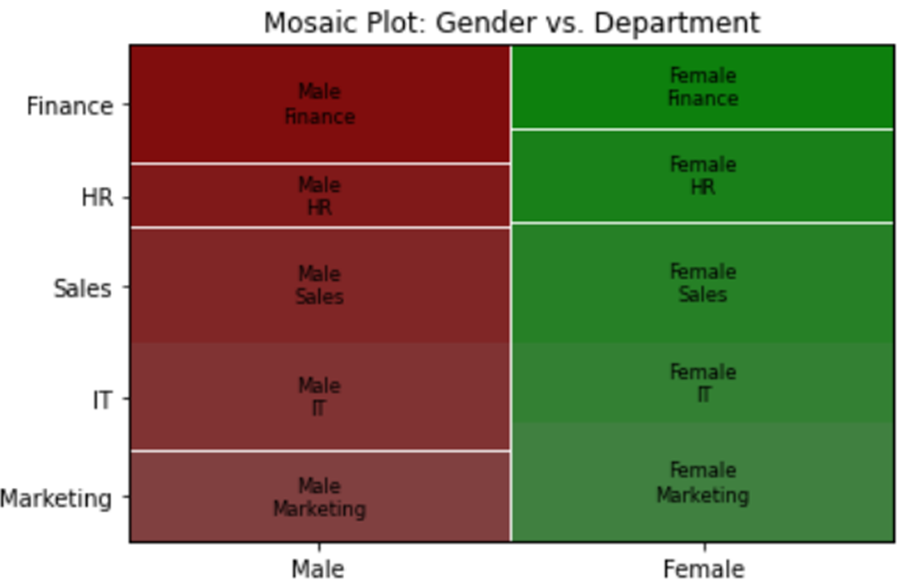
We’ve seen that from a departmental perspective the structure is balanced. What about the dynamics between departments? We should also pay careful attention to the relationships within its individual departments, achieving harmony on both fronts.

Therefore, the next question that arises is:

***“Is there a gender imbalance in the distribution of employees across different departments?”***



* We have imported the libraries (panda to read data from csv, Matplotlib for graphs & charts, Statsmodels for statistical functions for data analysis and modeling) and our dataset.
* Created a new figure window with a size of 10 inches by 6 inches.
* Called the mosaic() function from the statsmodels.graphics.mosaicplot module to create a mosaic plot of the employee data. The mosaic() function takes the following arguments: from employee\_data, ['Gender', 'Department']: to create the mosaic plot, adding as title for mosaic plot='Mosaic Plot: Gender vs. Department.
* Displayed the mosaic plot.



Our analysis & findings:

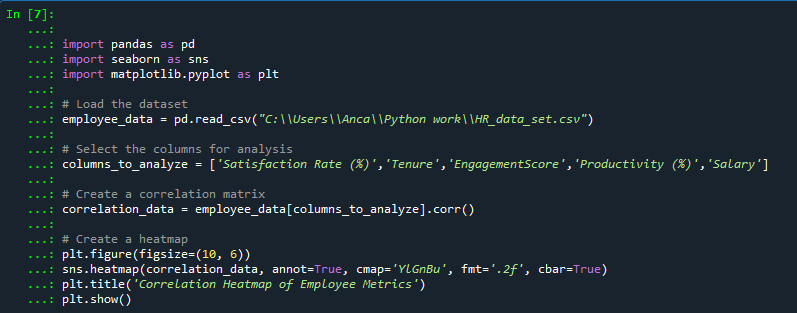
* Alignment with industry norms for consulting firms in following departments:
  + Finance (with slight male majority)
  + HR (with slight female majority)
* Significant gender disparity compared to the industry standard in the following departments:
  + Marketing (higher proportion of women vs norm), while industry standard is fairly balanced in terms of gender representation.
  + IT (less significant gender disparity which is not the case vs norm). IT is another traditionally male-dominated industry.
* Overall, our company demonstrates a well-established & balanced gender representation across its departments.

Going more in depth towards employee retention & metrics, we aim to uncover the relationship between tenure, satisfaction, engagement, and salary. Tenure is the length of time an employee has worked for a company. Satisfaction is an employee's feeling of contentment with his or her job. Engagement is an employee's level of involvement and commitment to his or her work. Salary is the amount of money an employee is paid.

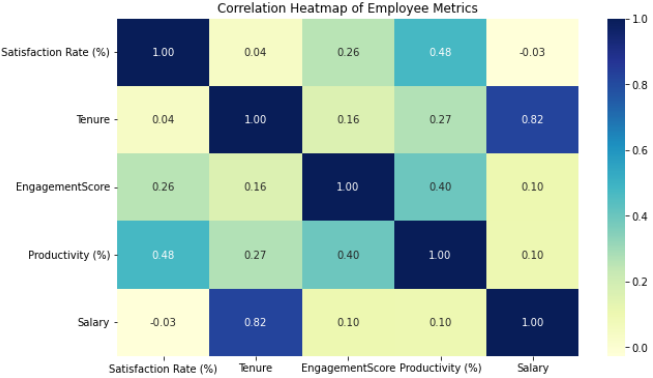
Research has shown that tenure is positively correlated with satisfaction and engagement, Satisfaction and engagement are also positively correlated with salary and that satisfaction and engagement can also impact salary. We’re here to understand whether this is the case for our company as well.

Therefore, our next question we aim to answer is:

***“Are there any interactions between tenure, satisfaction, engagement, and salary?”***



* We have imported the libraries (panda to read data from csv, Matplotlib for graphs & charts, Seaborn for data visualization) and our dataset.
* Selected the 'Satisfaction Rate (%)', 'Tenure', 'Engagement Score', 'Productivity' and 'Salary' columns to investigate their relationship.
* Calculated the correlation matrix between 'Satisfaction Rate (%)', 'Tenure', 'Engagement Score', 'Productivity' and 'Salary'. The correlation matrix is a numerical representation of how strongly two variables are related.
* Created & displayed a heatmap based on the correlation matrix. The heatmap uses a color scale to represent the strength of correlation, from blue (strong) to yellow (weak). The 'annot=True' option displays the correlation values on the cells, the 'cmap='YlGnBu' option specifies the colormap, the 'fmt='.2f' option formats the correlation values to two decimal places, and the 'cbar=True' option shows the colorbar for interpreting the color scale. The title 'Correlation Heatmap of Employee Metrics' provides context.
* Set the title and labels of the chart using the plt.title(), plt.xlabel(), and plt.ylabel() functions.
* Displayed the chart using the plt.show() function.



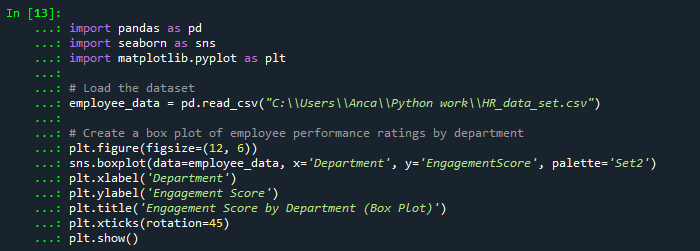
Our analysis & findings:

* Tenure & Salary are positively correlated, with a correlation coefficient of 0.82;
* Satisfaction & Productivity are positively correlated, with a correlation coefficient of 0.48;
* Employee Engagement & Productivity are positively correlated, with a correlation coefficient of 0.40.
* The graph indicates that there is a strong positive correlation between Employee Satisfaction, Engagement, & Salary. This suggests that these factors are interdependent and that improving one factor can lead to improvements in others.

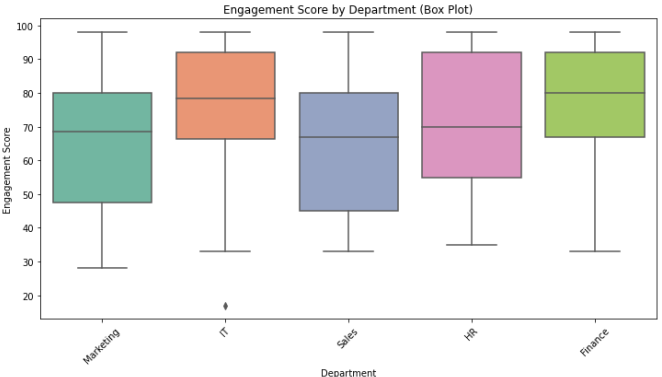
As our previous exercise suggests, engagement is one of the most important factors within an organization that can further affect employee satisfaction.

It is now the time to dive into a question purely related to the engagement variable, namely:

***“Which are the Departments with the highest engagement score within our organization?”***



* We have imported the libraries (panda to read data from csv, Matplotlib for graphs & charts, Seaborn for data visualization) and our dataset.
* Created a box plot to visualize the distribution of employee engagement scores across different departments. The 'plt.figure()' function sets the figure size to 12 inches wide and 6 inches tall. The 'sns.boxplot()' function generates a box plot using the employee data, grouped by the 'Department' column and measuring the 'EngagementScore'. The 'palette='Set2'' option specifies the color scheme for the boxes.
* Customize plot:
  + xticks(rotation=45)': Rotates the labels on the x-axis'plt.xlabel('Department')': Adds a label 'Department' to the x-axis.
  + 'plt.ylabel('Engagement Score')': Adds a label 'Engagement Score' to the y-axis.
  + 'plt.title('Engagement Score by Department (Box Plot)')': Adds a title 'Engagement Score by Department (Box Plot)' to the plot.
  + 'plt. by 45 degrees for better readability.
* Displayed the chart using the plt.show() function.



Our analysis & findings:

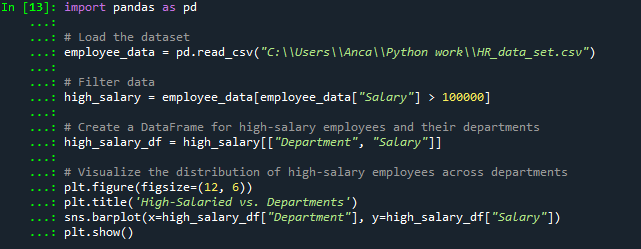
* The Finance department has the highest median engagement score, followed by the IT & Marketing departments.
* The Sales and HR departments have the lowest median engagement scores.
* Overall, the graph indicates that there is some variation in engagement scores across different departments in the company. Sales & Marketing have the lowest engagement score overall. This can be caused by leadership, workload, cultural factors, and this is where we should consider improvement measures.

To further capture the best insights for the Board of Directors, we wanted to analyse the salary variable as well, and to answer the following questions.

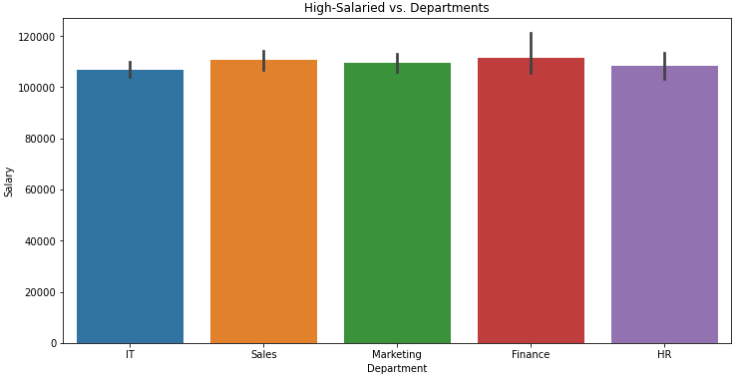
***“Which departments are known for having the highest salaries?”***

***“Is there a discrepancy in high salaries across different department?”***

.



* We have imported the library (panda to read data from csv, and our dataset.
* Filtered employee data set to select only employees with a salary greater than $100,000. The resulting DataFrame is stored in a variable named "high\_salary".
* Created a new DataFrame that only includes the "Department" and "Salary" columns from the "high\_salary" DataFrame. The resulting DataFrame is stored in a variable named "high\_salary\_df".
* Created a bar chart of the salary distribution for high-salary employees across departments. The chart is titled "High-Salaried vs. Departments" and has a figure size of 12x6 inches. The sns.barplot() function is used to create the bar chart, and the high\_salary\_departments\_df DataFrame is used to provide the data for the chart. The chart is then displayed using the plt.show() function.
* The chart title is "High-Salaried vs. Departments", and the x-axis shows the department, while the y-axis shows the salary.
* Displayed the box plot.



Our analysis & findings:

* Overall, the data suggests that Sales and Finance are the two departments with the highest average salaries in this company. Marketing and HR have lower average salaries, but there is still a wide range of salaries within each department.
* There are no high discrepancies across different departments.

And finally, a very hot topic nowadays, let’s assess the gender pay gap.

That being said, the final questions we aim to answer are:

***“Do men or women have higher salaries on average?”***

***“Is there a gender pay gap in the organization?***

A computer screen shot of a program code

Description automatically generated

* We have imported the libraries (panda to read data from csv, Matplotlib for graphs & charts) and our dataset.
* Defined the function called "calculate\_average\_salary“, that calculates the average salary for a specific gender. It takes two arguments: the DataFrame and the gender. It filters the DataFrame to only include employees of the specified gender and then calculates the average salary for that group. If there are no employees of the specified gender, it returns 0.
* Created a pie chart that compares the average salaries of male and female employees:
  + average\_salary\_male: The average salary for male employees.
  + average\_salary\_female: The average salary for female employees.
* Created a list of labels for the pie chart. The labels are "Male" and "Female".
* Created a list of the average salaries for male and female employees. The list is created by assigning the average\_salary\_male and average\_salary\_female variables to the list.
* Created a pie chart that shows the percentage of employees in each gender and their corresponding average salaries. It uses the Matplotlib library to create the chart. The chart title is "Average Salary Comparison (Male vs Female)", and the x-axis shows the gender, while the y-axis shows the average salary.
* Created the pie chart.

A blue and orange pie chart

Description automatically generated

Our analysis & findings:

* The average salary distribution shows a surprising trend: women earn slightly more than men in our company. This could be due to the fact that women are more likely to have a college degree and work in professional occupations.

Further Analysis & Discussions

Following the analysis performed on our organization, it is imperative we end it with a one-pager summary and future indications for our Board of Directors:

**Recruitment and Hiring**

* *Focus on referrals & employee networks*: implement referral bonus program to incentivize employees to refer friends & colleagues.
* *Increase use of online job platforms*: job postings - well-written, up-to-date & posted on relevant job platforms.

**Departmental Distribution**

* *Review departmental distribution:*keeping it flexible and aligned with company strategic goals.
* *Continue to have a balanced departmental distribution:*to reduce risk of silos & promote collaboration across company but considering company strategic evolving needs.

**Gender Diversity**

* *Review company's recruiting & hiring practices:*ensure job descriptions are not gender-biased.
* *Create a more inclusive work culture:*foster a culture that is welcoming & supportive for all employees, regardless of gender. Introducing Leadership programs that can focus on developing inclusive decision-making strategies to involve diverse perspectives and ensure that all voices are heard and valued.

**Employee Retention Metrics**

* *Focus on improving employee satisfaction:*providing opportunities for professional development, offering flexible work arrangements, recognizing, and rewarding employees for their contributions*.*
* *Increase employee engagement:*investing in training and development, creating positive & supportive culture, providing opportunities to participate in decision-making.

**Salary**

* *Review the company's salary structure:*ensure that salary structure is fair & equitable & does not discriminate based on gender or any other protected characteristic.

Our comprehensive analysis has revealed areas of strength and opportunity within our organization. By implementing the strategic initiatives outlined above, we can further enhance our talent acquisition, departmental structure, gender diversity, employee retention, and salary practices, laying the foundation for sustainable growth and success.

By implementing these strategic initiatives, we can create a workplace that attracts, retains, and empowers top talent, leading to organizational success and growth.