

**An Exploration and Analysis of the Factors Relating to Employee Attrition**

***A Data-Driven Approach to Decision Making at PharmaCo***

MGT7182: HUMAN RESOURCE ANALYTICS

Word count: (3,153)

Module Coordinator: Dr. Byron Graham

Due date: 11/12/2020

**CONTENTS**

[**Introduction and Background**](#_7ilbx8ejek6s) **3**

[1.1 Employee Retention in Pharmaceutical Companies](#_f0a5ch6nd6ut) 3

[**Methodology**](#_19imlj8vb68e) **5**

[2.1 Data Quality Analysis](#_6uhjrjbo3qo3) 5

[2.2 Data Visualisations (Tableau)](#_ygypsqufl75h) 7

[2.3 Predictive Modelling (Knime)](#_inp0is7vy5ai) 7

[**Results and Discussion**](#_uve2zjgocqfh) **9**

[3.1 Tableau Visualisations and Interpretation](#_qrg4wx5nf2fm) 9

[3.2. Predictive Accuracy of modelling methods](#_f6tfumb4f5jn) 12

[3.3 Interpretation of Predictive Decision Tree Model](#_47so84nlmvwk) 13

[**Conclusion**](#_53aazbkd5qbb) **13**

[4.1. Practical implications and recommendations for PharmaCo](#_kcun4zhgivwt) 13

[4.2. Implications for future research](#_p4plaze5mh4l) 14

[4.3 Limitations of this study](#_khgl3b5s89x1) 14

[**REFERENCES**](#_fgodtxxanpfr) **15**

[**Appendices**](#_8ibmvq1646i5) **17**

# Introduction and Background

One of the most critical and challenging issues organisations face today is employee turnover. The CIPD (2020) refers to employee turnover as “the proportion of employees who leave an organisation over a set period, expressed as a percentage of total workforce numbers”. Turnover can be voluntary and initiated by the employee or involuntary and initiated by the organisation. Another distinction can be made between functional and dysfunctional turnover. Dysfunctional turnover includes the exit of high performers causing harm to the organisation. Functional turnover, on the other hand, might include the exit of poor performers or employees with easy-to-replace skills, and does not hurt the organisation.

Employee turnover is a common, persistent problem in organisations and one of the biggest human resource challenges. As such, it is a much studied phenomenon that has received considerable attention by HR professionals and senior management. Research has consistently identified certain factors linked with turnover including; demographic factors such as gender, age, marital status and nationality, work-specific factors such as pay, workload, job satisfaction and other elements such as poor morale and low levels of motivation, inadequate recruiting and selecting strategies and lack of career opportunities influence turnover and employee intentions to stay in the company (D’Addio et al., 2007; Higgins et al., 2000; Grissom et al., 2016; Pitts et al., 2011).

The reason research has focused so keenly on the causes of employee turnover is the impact high turnover rates have on performance and profitability. In today’s highly competitive talent market, companies strive to find top talent and invest resources in recruiting and retaining top performers with in-demand skills (Thibault et al., 2017). Employee turnover has a negative impact on company resources, costing both time and money. Research by the CIPD (2020) suggests costs associated with turnover can reach as high as 90% to 200% of an employee’s annual salary (Cascio, 2006; CIPD 2020). Many studies have linked high turnover rates to deficiencies in organisational performance, while competent and committed employees account for a significant contribution to the company’s overall productivity rates. Therefore, HR functions must strive to keep employee attrition at a minimum, in order to maintain high performance and competitive advantage.

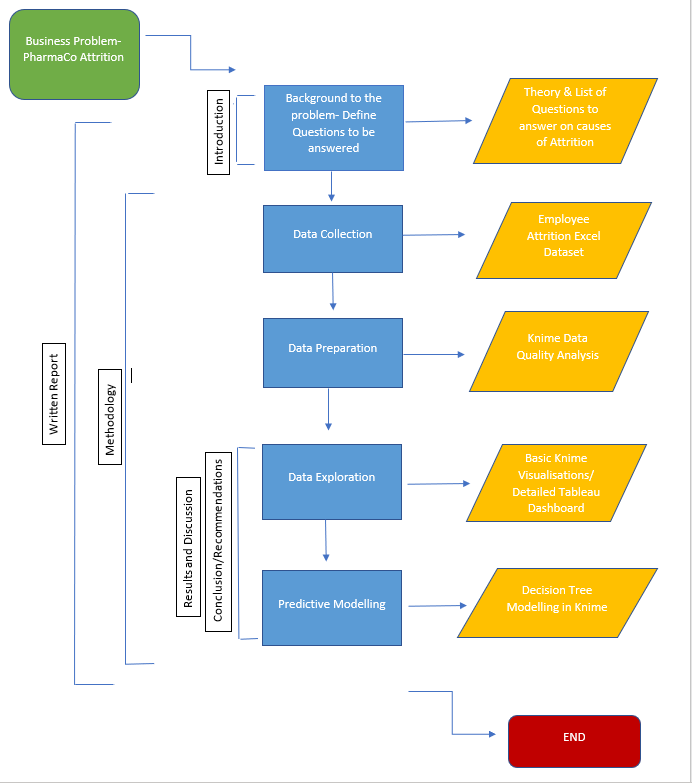
As technology and data-driven practices progress, HR Analytics has offered solutions to manage employee data and make evidence-based decisions to achieve organisational objectives. According to Levenson, “HR analytics includes the use of statistical techniques, research design, and algorithms to evaluate employee data and translating results into evocative reports’’ (Levenson 2005). By applying analytics, HR professionals can gain insights into employee data and reveal patterns that make it possible to *predict* employee behavioural patterns. Predictive analytics can be applied to forecast attrition, through the use of statistical data analysis, modelling, machine learning, data mining and artificial intelligence. Companies can leverage data including employees’ demographic information, work preferences, work hours, absence data, reasons for quitting etc. to assist HR to form evidence-based strategies. Analytics can provide firms with an opportunity to flag and prevent HR issues, such as attrition, before they seriously harm the company (Mohammed and Quddus, 2019).

#### *1.1 Employee Retention in Pharmaceutical Companies*

The pharmaceutical industry is considered one of the most influential industries worldwide with extensive investments in research and development which correspond with heavy investments in human resources. As it is a highly-regulated industry with a highly-skilled workforce, the effects of turnover can have a larger impact than in other fields and so, pharmaceutical companies continuously seek to retain their employees so as not to lose their best talent and hard-to-replace workers. More and more pharmaceutical companies have started to recognise the potential of the use of HR Analytics and utilise that tool to curb attrition. For example, in an innovative project, pharmaceutical company AstraZeneca mined thousands of pieces of historical data to predict the likelihood of someone resigning. Despite this the pharmaceuticals industry is grappling with the second highest level of attrition (after IT), with an annual attrition rate of 30-35%, according to a recent survey (Pawar & Chakravarthy, 2014).

This report examines a problem with high staff turnover at PharmaCo over the past 12 months. Some descriptive visualisations were produced in Tableau to show the relationships between key variables and turnover, while machine learning in Knime was used to build predictive models to help identify which staff are at risk of leaving over the next 12 months. Suggestions were then made based on the findings in order to aid management to reduce unwanted attrition. Figure 1 (below) is a graphical representation of the approach taken to examine the attrition problem, while the methodology provides a more detailed breakdown of each individual step. This provides an understanding of the task while also offering a guided framework for future analytics projects in the company.

**Figure 1:** Project Flowchart

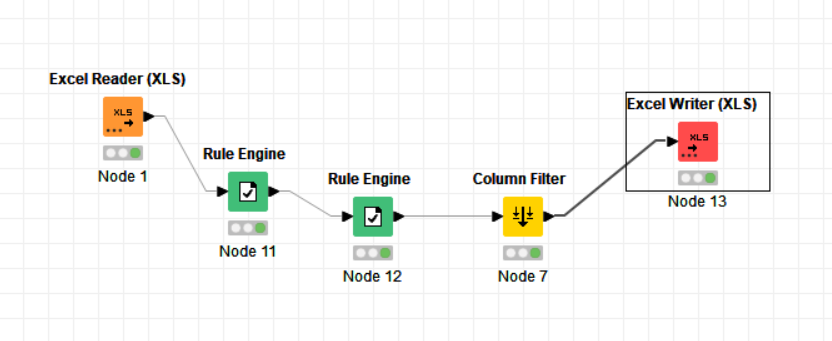


# Methodology

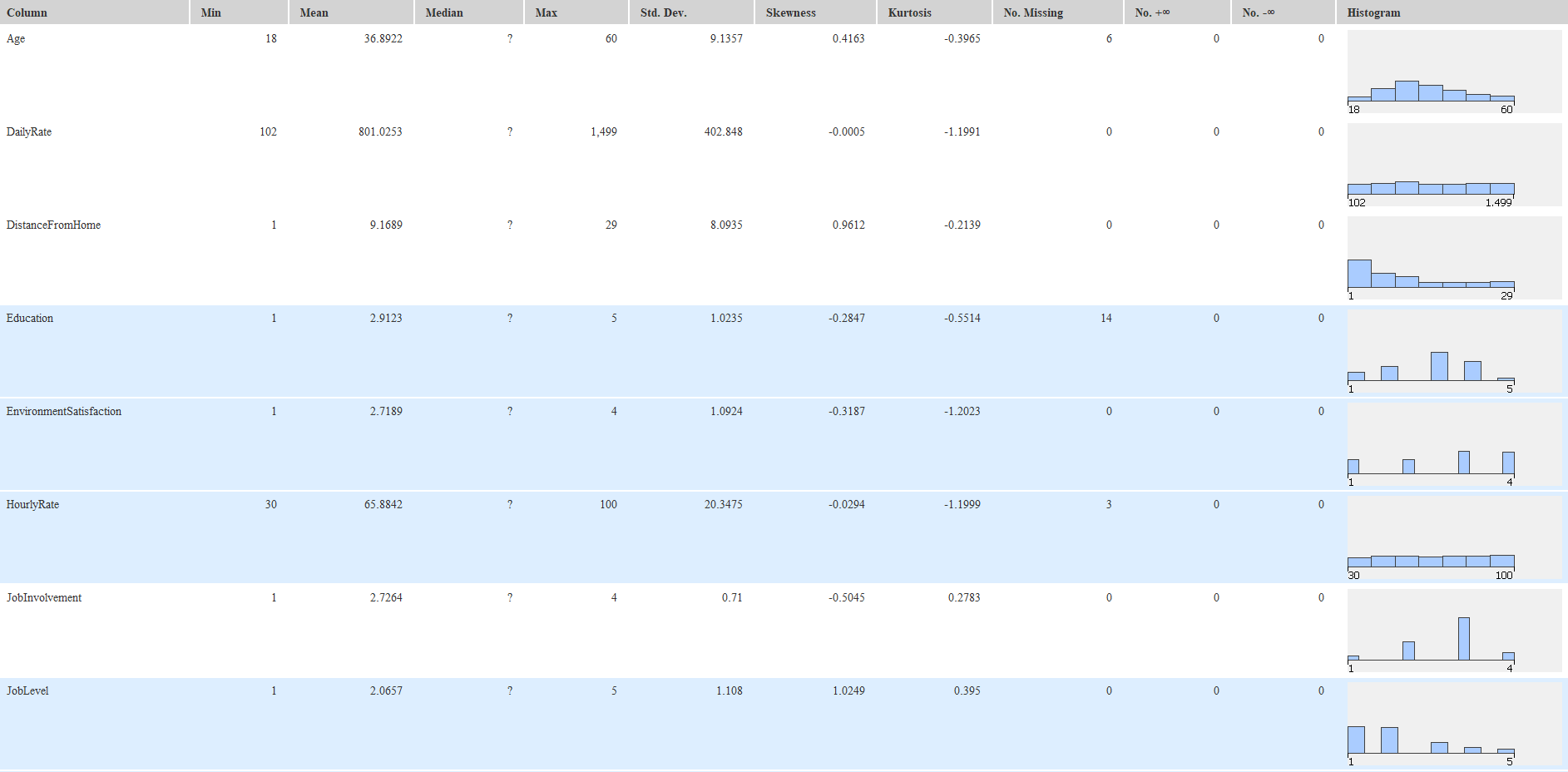
#### *2.1 Data Quality Analysis*

KNIME was used to identify and fix a number of quality issues in the dataset. The dataset was held as an excel file, and so the ‘Excel Reader’ node was used to read-in the data. The ‘Statistics’ node was used to provide an initial descriptive summary of the data and highlighted data quality issues in a number of columns. Figure 2 shows the full workflow used to address these issues.

**Figure 2**: Data Quality Workflow (Knime)



The ‘Age’ column was adjusted to remove an incorrect value. Initially, the ‘Row Filter’ node was used to do this, however the ‘Rule Engine’ was determined to be the best solution so as not to remove multiple rows. The new age range was set between 16-80. There was also an extreme outlier in the hourly rate variable, of $700 per hour, which was removed using the “Rule Engine” node. Three columns (“ID”, “Employee Count” and “Over 18”) were determined to be of little value and were removed using the ‘Column Filter node’. There were a total of 36 missing values across 3 columns in the dataset. These were not deemed to be errors, nor significant enough to be removed, while keeping them was of use in Predictive Modelling carried out in Knime, and Tableau automatically filtered out missing data.

**Figure 3**: Statistical Summary Sample of Clean Dataset

A new descriptive summary using the clean data was produced, shown in Figure 3. Some basic visualisations showing ‘Attrition’ and ‘Monthly Income’ were also produced (appendices 1, 2 and 3), before the clean dataset was exported for use in Tableau.

#### *2.2 Data Visualisations (Tableau)*

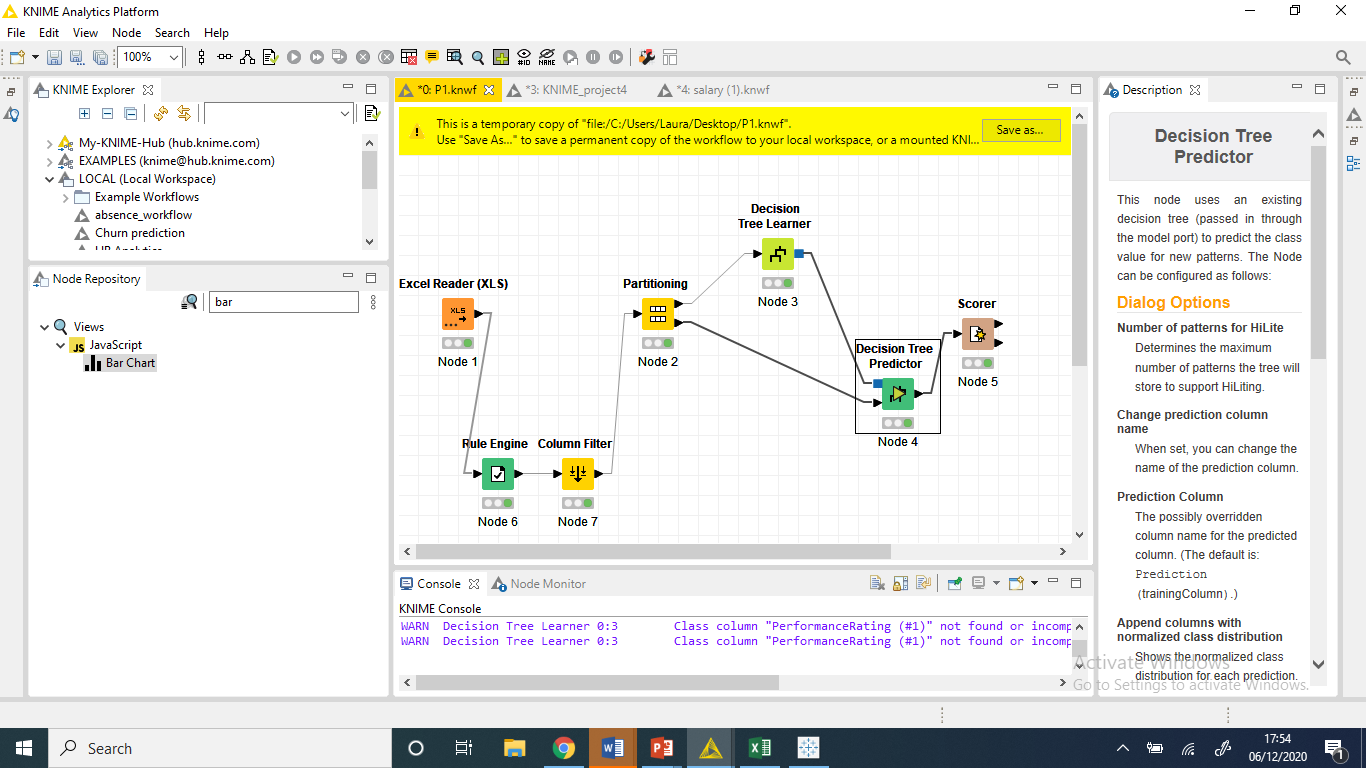
Tableau Desktop was used to visualise the clean data and analyse factors relating to attrition. Tableau Desktop uses a drag & drop approach to data analysis, allowing for quick and clean visualisations, that can assist decision-making for HR managers at PharmaCo. The data was imported and several factors were explored which often relate to turnover rates, as described in the literature (D’Addio et al., 2007; Higgins et al., 2000; Grissom et al., 2016; Pitts et al., 2011). These factors include job satisfaction, job role, monthly rate, monthly income, education field, overtime. Some visualisations were explored before 4 final visualisations were created based on the key factors, while a Tableau Dashboard showing all visualisations was produced for practical use (appendix 5).

#### *2.3 Predictive Modelling (Knime)*

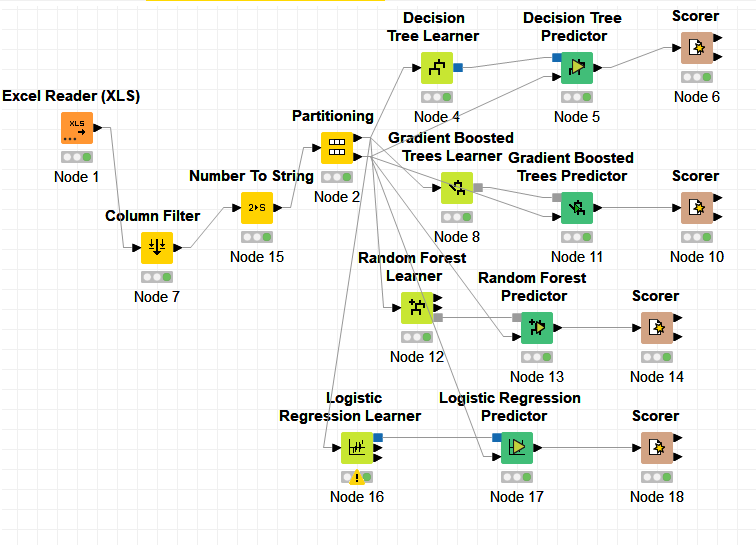
A basic model to predict attrition was created using Knime. Using the clean dataset, a decision tree was produced. A decision tree is a commonly used data-mining method for developing prediction algorithms for a target variable, in this case attrition (Song & Ying, 2015). This model classifies a population into branch-like segments, to create an inverted tree. The decision tree was chosen as a first exploration of prediction, to highlight the most salient characteristics associated with attrition at PharmaCo. To retain generalisability and avoid overfit, the data was split using the ‘partitioning’ node at an 80% training and 20% test ratio, before implementing the ‘decision tree learner’ node, a ‘decision tree predictor’ node and lastly a ‘scorer’ node (Figure 4).

To improve accuracy and further reduce the risk of overfit, MDL pruning was implemented to the ‘decision tree learner’ node. A further model removed the inclusion of protected characteristics (age, gender and marital status). The removal of these characteristics was found to insignificantly affect model accuracy, suggesting discriminatory practices are not critical to attrition at Pharmaco. Multiple pay variables were present, with monthly income having a slight predictive effect. The other variables were removed to avoid potential issues of multicollinearity. For example, monthly pay/daily rate/wage variables are likely to be related, so only one of these variables was included.

**Figure 4**: Decision Tree Workflow in Knime



Some additional algorithms were applied to the workflow (Figure 5). The ‘Random Forest Predictor’ consists of multiple decision trees with each class producing their own result. The overall accuracy is often better than a single tree.

**Figure 5**: Knime workflow with additional prediction algorithm

Similarly, a ‘Gradient Boosted Trees Predictor’ was used. Again, as multiple decision trees are used in this model, the prediction accuracy should be higher. The main difference between this model and the Random Forest Learner is that it builds one tree at a time, while Random Forest builds each tree independently. Configuring the number of models produced to 250 yielded the most accurate results.

Lastly the ‘Logistic Regression’ model was used. This model is highly effective for predicting the outcome of categorical variables. In this case, the category is a yes/no binary (attrition). A mix of 14 categorical and numerical variables such as monthly rate, job satisfaction, job role and overtime produced the most accurate results.

# Results and Discussion

#### *3.1 Tableau Visualisations and Interpretation*

The following section analyses the results from the four key visualisations.

**Figure 6: Turnover and Salary Across Job Roles**

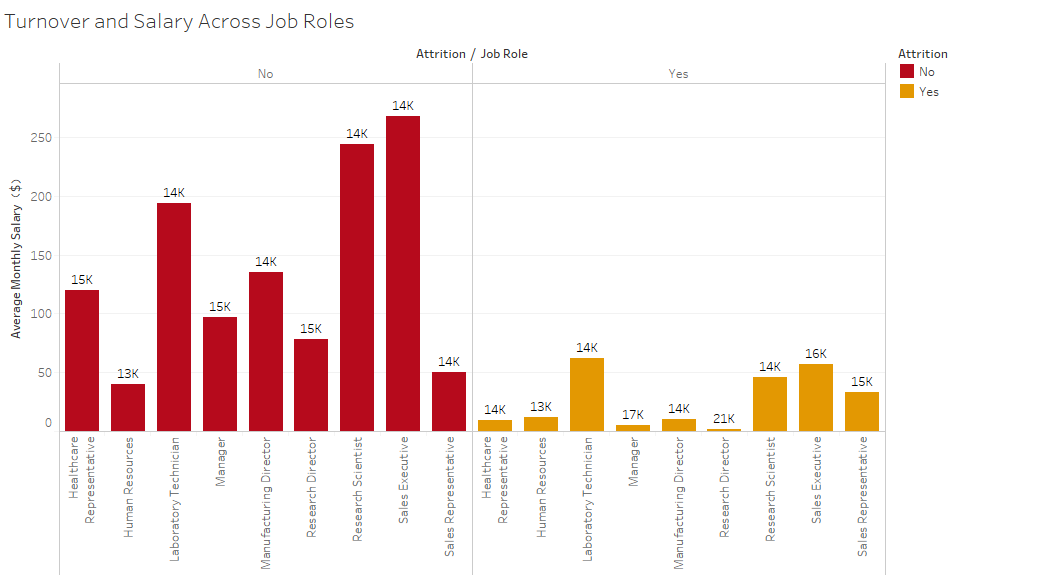


Figure 6 shows how ‘Job Role’ and ‘Monthly Rate’ affect attrition. The visualisation shows a correlation between salary and attrition rates, with those on a lower average salary more likely to leave the company. Furthermore, certain job roles have a more equal proportion of employees leaving/staying than others. Sales representatives, for example, show an almost even distribution between those that leave and stay, although more information is needed to determine if this is more related to salary or position. Research suggests that one of the most important elements of job satisfaction is indeed related to the characteristics of the position, such as job content, autonomy, compensation, promotion and career prospects (Bonache, 2005, Vidal et al., 2007).

Different positions have different job content and different salaries. These elements can easily become important catalysts for resignation. Therefore, the finding that employees occupying higher positions and receiving higher salaries, such as Research Directors, are less likely to leave the organisation was not surprising. However, it is not clear whether remuneration was the only factor that played a role in turnover at PharmaCo as the attrition rates were inconsistent in employees receiving average salaries.

**Figure 7:** The relationship between job satisfaction and turnover rates of different genders

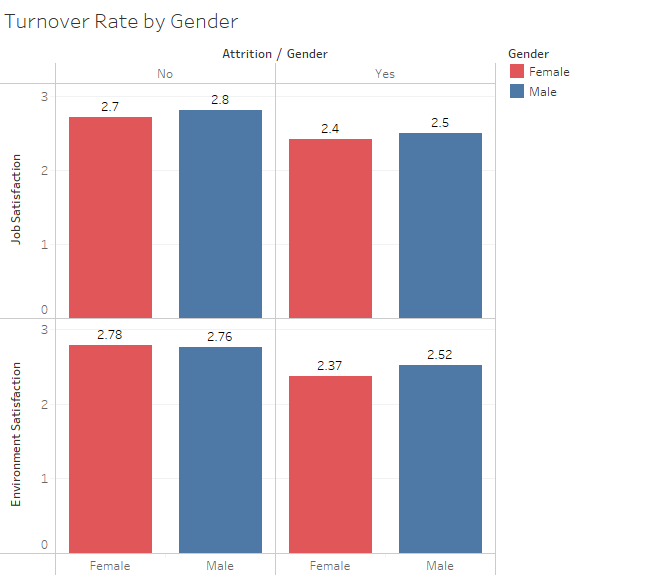
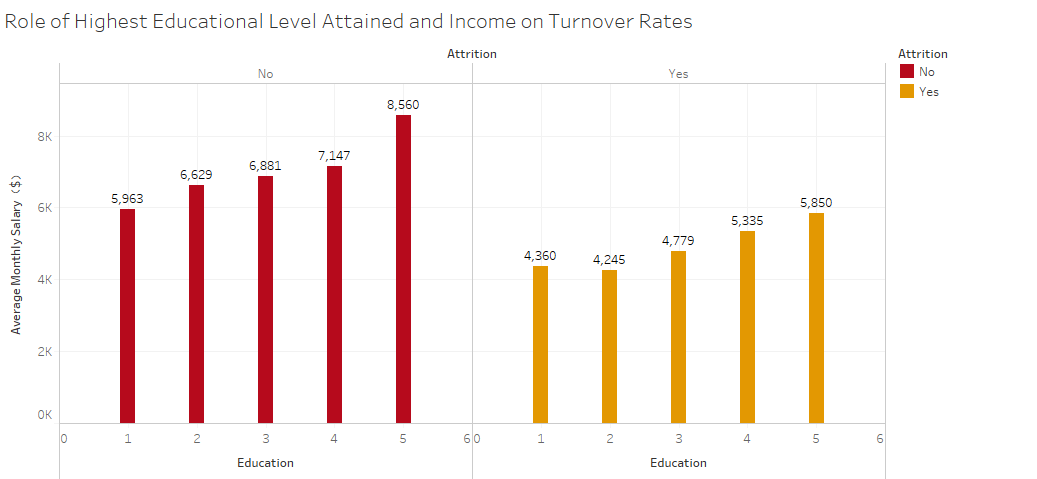


Figure 7 shows the relationships between job/environment satisfaction, gender and attrition at PharmaCo. There is no clear correlation between gender and job/environment satisfaction. The visualisation does however show a correlation between lower average and job/environment satisfaction and Attrition. This is in line with findings from research conducted by Griffeth et al. (2000) that indicated that job satisfaction was the best predictor of turnover as it displayed the highest relationship to employee attrition. The company should continue to strive for gender equality in the workplace. Surveys could be used to determine how job and environment satisfaction can be improved in order to reduce attrition.

**Figure 8: Role of Highest Educational Level Attained and Income on Turnover Rates**

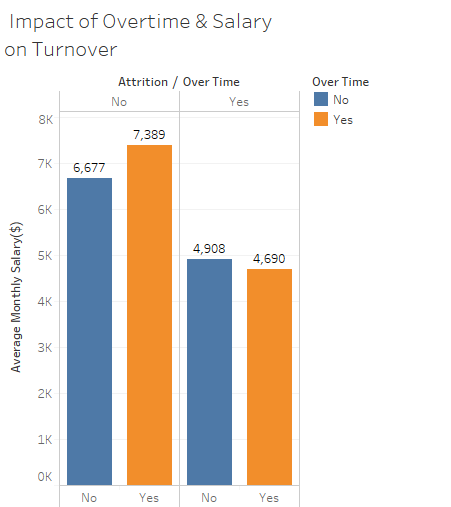
****

**(1'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'**

**Education=Education levels)**

As shown in Figure 8, the turnover rate rises as the level of education increases. Higher-educated employees are more likely to leave because they believe they should be paid more and also have more potential career opportunities. The finding that education levels impact turnover rates was interesting as it is inconsistent with other studies that have found no significant relationship between level of education and turnover intent (Lambert et al., 2001). One explanation for this finding could be that different occupations require different educational levels and therefore it is normal that there would be inconsistencies between educational levels and needed work skills.

**Figure 9: Impact of Overtime & Salary on Turnover**



This graph shows that for an employee working overtime, earning an average salary of $7,000 is more likely to stay with the company than someone earning $4,000. This means that there may be loopholes and inequities in a company's remuneration system. Since overtime exists, companies should design a matching remuneration system and compensation mechanism to ensure a balance between the interests of the employee and the company.

Studies have shown a correlation between excessive overtime hours and occupational Injuries, turnover rate. Turnover can be also exacerbated by wages that do not match the labour effort (Anderson et al., 1988).

#### *3.2. Predictive Accuracy of Modelling Methods*

|  |  |
| --- | --- |
| Decision Tree Predictor | *82.82%* |
| Random Forest Predictor | *87.03%* |
| Gradient Boosted Trees Predictor | *85.324%* |
| Logistic Regression Model | *87.37%* |

The accuracy results for the table above show the most accurate model as the Logistic regression model. This is likely due to the inclusion of specific variables such as “Job Role” and “Overtime” which were proven to be highly correlated in previous models, as well as its usefulness in predicting binary outcomes such as attrition. Both the Random Forest and Gradient Boosted Trees predictor, as expected, are more accurate than the single decision tree, as they include multiple decision trees. These models however are limited in interpretative value in Knime, which is why the decision tree output, despite having a lower accuracy than others, remains highly practical in this analysis.

#### *3.3 Interpretation of Predictive Decision Tree Model*

The decision tree model revealed that the variable which is most strongly correlated with employee attrition within PharmaCo was ‘Job Role’. Attrition typically varies widely between industries however in the US, companies have an average turnover rate of 22%, with 15% attributed to voluntary turnover (Mercer, 2019). At PharmaCo, attrition for 21/22 is predicted to be 16%. The job roles identified which would suffer from high attrition within the next 12 months were; Sales Representatives (38.7%), Sales Executives (19.4%) Laboratory Technicians (24.8%) and Human Resources professionals (21%). Attrition at PharmaCo is characterized by high attrition in the aforementioned roles and low attrition in other roles e.g. Research Director which has a 3% attrition rate.

Overtime is a salient factor for Sales Executives (37.7%), Representatives (68.8%) and Laboratory Technicians (46.2%). For HR professionals a shorter working career, *specifically working less than 2.5 years* is a risk factor for attrition. For laboratory technicians, a lack of work/life balance, environmental factors and a monthly income less than $3727 are all risk factors for their leaving.

# Conclusion

#### *4.1. Practical Implications and Recommendations for PharmaCo*

On a positive note, PharmaCo’s attrition rates are not unusually high, nor do they suggest gender discrimination. This investigation has provided a series of useful insights which may help PharmaCo to further reduce their attrition. Essentially, the above findings indicate PharmaCo may need to adopt a differentiated approach to their HRM policies, as their current strategy appears to work well for certain job roles but not for others, particularly those in sales. Further research is required to explore the issues within specific job roles. Job content and tasks for these positions may be adapted in order to keep employees more engaged and satisfied.

In addition, overtime was found to be particularly problematic for three of the four roles which experience high turnover. This may be self-fulfilling; as employees leave, their peers are required to work more and then also leave as a result of burnout (Junaidi et al., 2020). Reducing the amount of overtime permitted to be performed by employees within the organisation is recommended, while also adopting a proactive approach to hiring new employees for those job roles which see a correlation between overtime and attrition. Using this predictive attrition model, a ‘Goldilocks model’ of overtime may be achieved; where those employees who want to work overtime are able to do so, but the organisation is not under-staffed and reliant on its employees working overtime (Goldenhar et al., 2003).

Salary was also found to be a key factor in determining attrition levels, with literature also considering remuneration as a key resource for reducing turnover rates (Ducharme et al. 2005). Workers who make more, but whose salaries fall short of the going market rate, may feel undervalued at their current companies and look for a company that will pay them what they're worth (Firth et al., 2004). PharmaCo should ensure it’s salaries are at least in line with the market rate in the pharmaceutical sector. Appropriate market research must be conducted in this area for each job role, and should be reviewed on an annual or quarterly basis if possible. Salary reviews could also be utilised to not only enable the organisation to remain fair with employee compensation, but stay competitive in the job market, and reduce dysfunctional turnover.

PharmaCo should also look to further improve their data-collection process, with a greater emphasis on data that provides insight into *why* past employees have left. For example, an exit survey or past feedback could be quantified and processed. Data on more external factors, such as economic climate, could also be collected for use in analysis to provide deeper insight into factors affecting attrition.

#### *4.2. Implications for Future Research*

Studies have suggested that demographic factors such as gender, age, marital status and nationality can influence turnover and employee intentions to stay in the company (Grissom et al., 2016; Pitts et al., 2011; Griffeth et al., 2000), however removing demographic characteristics had little effect on decision-tree accuracy. Therefore, we suggest that future research is less focused on demographic factors when examining employee turnover and more considerate of other elements that HR can tangibly effect, such as work environment satisfaction, diversity or innovation. The study indicated that job satisfaction plays a major role in employee turnover rates. However, job satisfaction is a combination of environmental circumstances including benefits, promotion, supervision, working condition, relationship with co-workers and communication (Westlund & Hannon, 2008). Therefore, in the future there is a need for deeper research to identify which of these factors are the most important drivers of employee turnover, so HR professionals can make more evidence-based decisions in order to reduce attrition levels. Research in this area should also influence deeper data collection practices into these factors, facilitating better analysis and decision making as a result.

#### *4.3 Limitations of this study*

In terms of functional and dysfunctional turnover, it is hard to gauge at PharmaCo. Based on performance ratings alone, all PharmaCo turnover unfortunately *appears* to be dysfunctional*,* as every employee achieved a rating of either ‘good’ or ‘excellent’. The dataset also does not reveal the cost of replacing employees of each ‘at-risk’ job role. The consistently high ratings of employees might reflect an excellent hiring system or a dysfunctional performance management system; however it is clear PharmaCo should keep a record of their functional and dysfunctional turnover to better target interventions in the future.

There are also limitations in terms of the data itself. There is no quantifiable feedback from those that have left PharmaCo on exactly *why* they did so. This data is also a snapshot, and may change overtime, so it is important that users are aware of this and that data is continuously collected and analysed. It would also be useful to have information and data on what action has already been taken, if any, to reduce attrition levels, which can be taken into account for future studies and recommendations. Lastly, some contextual economic and industry data could have improved this study, to determine whether attrition levels were affected by economic climate or consistent with the pharmaceutical industry, rather than a characteristic of PharmaCo itself.

# REFERENCES

Anderson, E. E. and Buchholz, R. A. (1988) ‘Economic Instability and Occupational Injuries: The Impact of Overtime Hours and Turnover Rates’, *Labor Studies Journal*, 13(4), p. 33. Available at: http://search.ebscohost.com.queens.ezp1.qub.ac.uk/login.aspx?direct=true&db=buh&AN=6121630&site=eds-live&scope=site (Accessed: 7 December 2020)

Bonache J. (2005) Job satisfaction among expatriates, repatriates and domestic employees, *Personnel Review*, 34(1), pp. 110–124

Cascio, W.F. (2006) Managing Human Resources: Productivity, Quality of Work Life, Profits (7th ed.). Burr Ridge, IL: Irwin/McGraw-Hill.

CIPD (2020) *Retaining Talent: A Guide to Analyzing and Managing Employee Turnover*, London: SHRM Foundation.

D’Addio, A.C., Eriksson, T. and Frijters, P. (2007) “An analysis of the determinants of job satisfaction when individuals baseline satisfaction levels may differ”, *Applied Economics,* Vol. 39 No. 19, pp. 2413-2423.

Ducharme, M., Singh, P., and Podolsky, M. (2005). Exploring the links between performance appraisals and pay satisfaction. *Compensation Benefits Review*, 37(5), pp. 46-52

Griffeth, R.W., Hom, P.W. and Gaertner, S. (2000), “A meta-analysis of antecedents and correlates of employee turnover: update, moderator tests, and research implications for the next millennium”, *Journal of Management*, Vol. 26 No. 3, pp. 463-488.

Grissom, J.A., Viano, S.L. and Selin, J.L. (2016), “Understanding employee turnover in the public sector: insights from research on teacher mobility”, *Public Administration Review*, Vol. 76 No. 2, pp. 241-251.

Higgins, C., Duxbury, L. and Johnson, K.L. (2000), “Part-time work for women: does it really help balance work and family?”, *Human Resource Management*, Vol. 39 No. 1, pp. 17-32.

Junaidi, A., Sasono, E., Wanuri, W. and Emiyati, D. (2020) ‘The Effect of Overtime, Job Stress, and Workload on Turnover Intention’, *Management Science Letters*, *10*(16), pp.3873-3878. [Online] Available at: <http://growingscience.com/msl/Vol10/msl_2020_248.pdf> (Accessed 5th December 2020)

Lambert E.G., Hogan N.L. & Barton S.M. (2001) The impact of job satisfaction on turnover intent: a test of a structural measurement model using a national sample of workers, *The Social Science Journal*, 38(2), pp. 233-250

Levenson, A., Lawler III, E. E., & Boudreau, J. W. (2005). *Survey on HR Analytics and HR transformation*: Feedback report. Los Angeles, CA: Center for Effective Organizations, University of Southern California.

Mercer (2019) ‘*North American Employee Turnover Trends and Effects* ’ [online] Available at: <https://www.imercer.com/articleinsights/North-American-Employee-Turnover-Trends-and-Effects> (Accessed 12 November 2020)

Mohammed, D. and Quddus, A. (2019) HR Analytics: A Modern Tool in HR for Predictive Decision Making. *Journal of Management*, *6*(3).

Mitchell, T.R., Holtom, B.C., & Lee, T.W. (2001) How to keep your best employees: Developing an effective retention policy. *Academy of Management Executive*, 15, pp. 96-108.

Nyberg, A. (2010), “Retaining your high performers: moderators of the performance–job satisfaction–voluntary turnover relationship”, *The Journal of Applied Psychology,* Vol. 95 No. 3, pp. 440-453.

Pawar, I.A. and Chakravarthy, V. (2014) ‘Factors Influencing Employee Turnover in Fusion Healthcare Organisation’ *International Journal of Management Research and Reviews*, *4*(9), p.834.

Pitts, D., Marvel, J. and Fernandez, S. (2011), “So hard to say goodbye? Turnover intention among us federal employees”, *Public Administration Review*, Vol. 71 No. 5, pp. 751-760

Price, J. L. (1989) ‘The Impact of Turnover on the Organization’, *Work and Occupations*, 16(4), pp. 461–473. doi: 10.1177/0730888489016004005.

Schmidt, J.A., Pohler, D. and Willness, C.R. (2018) ‘Strategic HR System Differentiation Between Jobs: The Effects on Firm Performance and Employee Outcomes’, *Human Resource Management*, *57*(1), pp.65-81. [Online] Available at: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/hrm.21836?saml_referrer> (Accessed: 29th November 2020)

Song, Y.Y. and Ying, L.U. (2015) ‘Decision Tree Methods: Applications for Classification and Prediction’, *Shanghai Archives of Psychiatry*, *27*(2), p.130. [Online]. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/> (Accessed 6th December 2020)

Thibault Landry, A., Schweyer, A. and Whillans, A. (2017) ‘ Winning the war for talent: Modern motivational methods for attracting and retaining employees’ *Compensation & Benefits Review*, *49*(4), pp.230-246. [Online] Available at: <https://www.awhillans.com/uploads/1/2/3/5/123580974/0886368718808152__1_.pdf> (Accessed 14 November 2020)

Vidal M.S., Valle R.S., and Aragón M.B. (2007) Antecedents of repatriates' job satisfaction and its influence on turnover intentions: Evidence from Spanish repatriated managers, *Journal of Business Research*, 60(12), pp. 1272-1281.

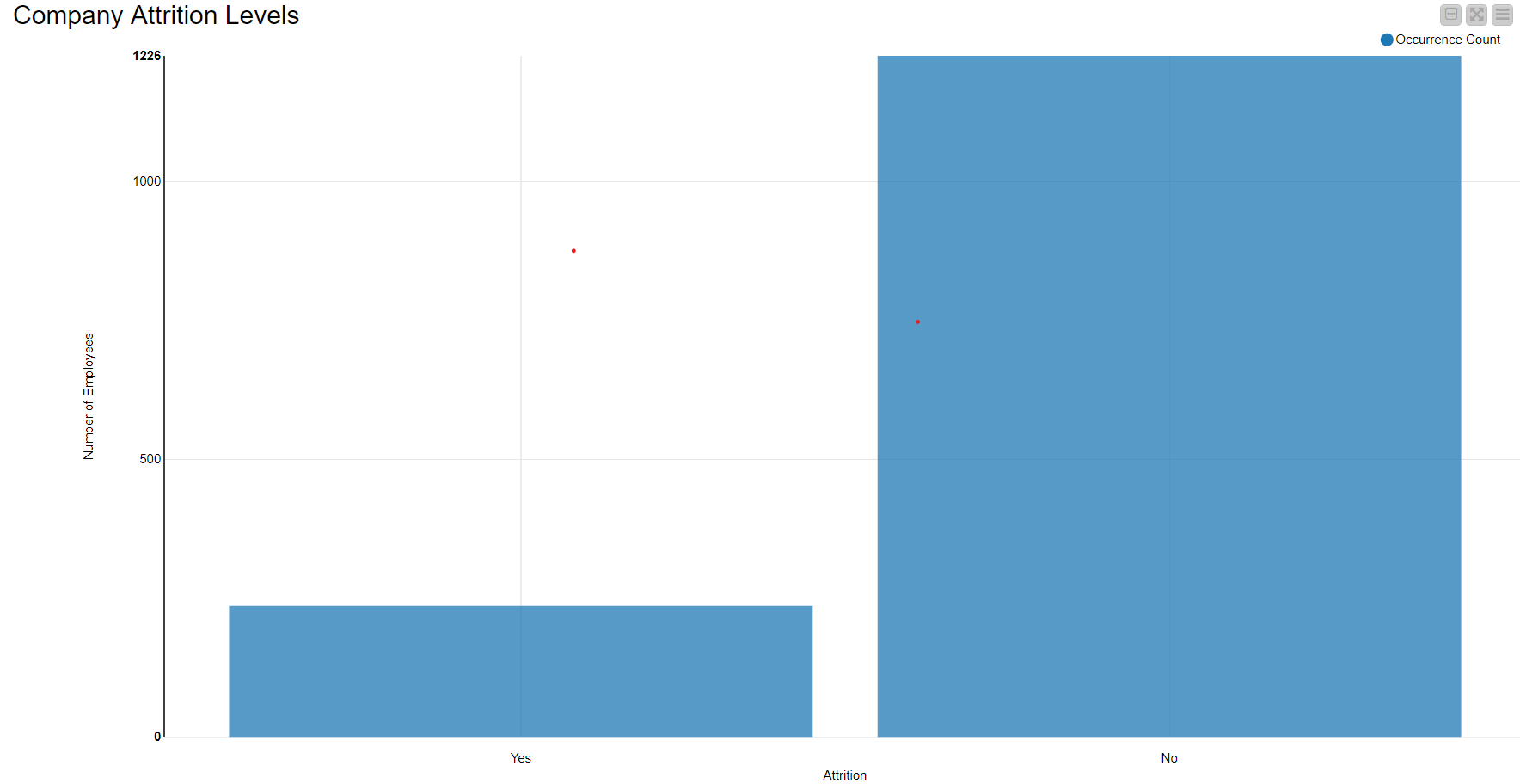
Westlund S. and Hannon J. (2008) Retaining talent: Assessing job satisfaction facets most significantly related to software developer turnover intentions, *Journal of Information Technology Management*, 19(4), pp. 1-15

# 

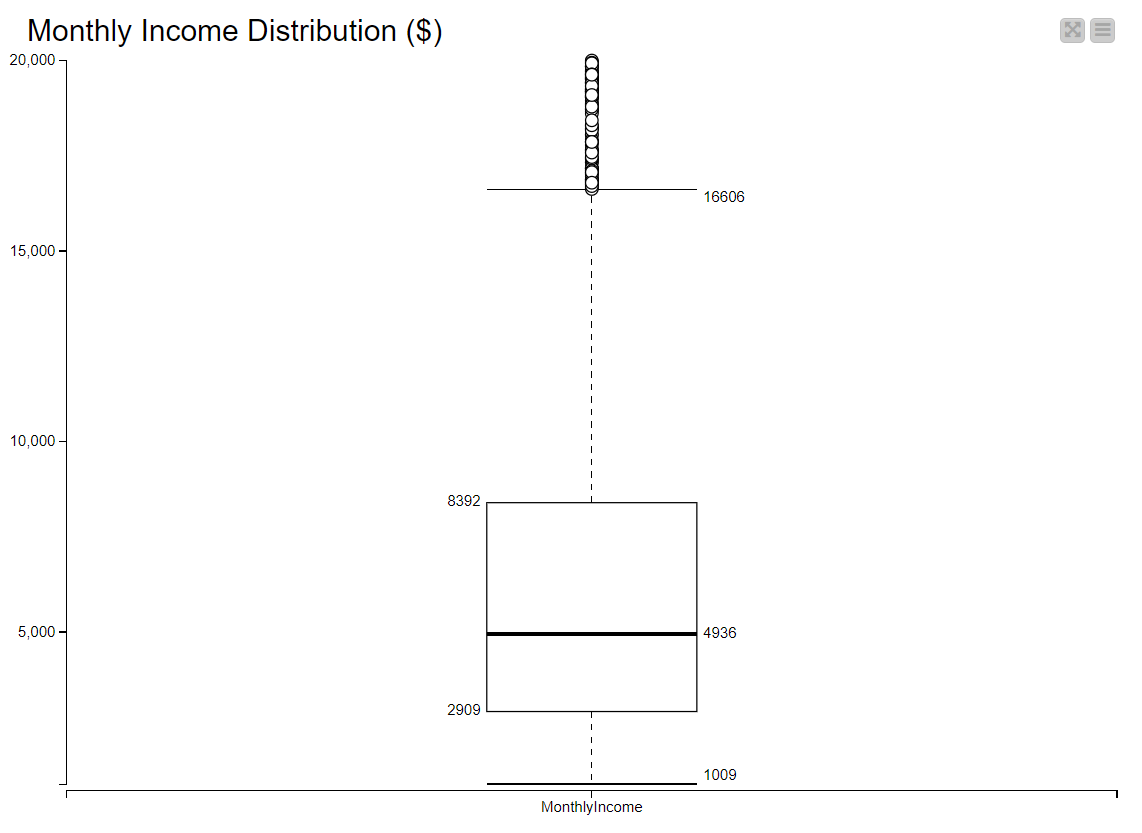
# Appendices

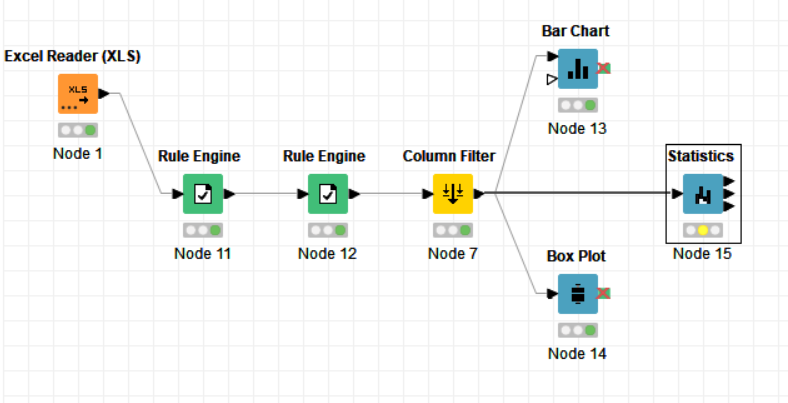
**Appendix A: Screenshots**

**Appendix 1**: Basic KNIME Barchart showing employee Attrition levels

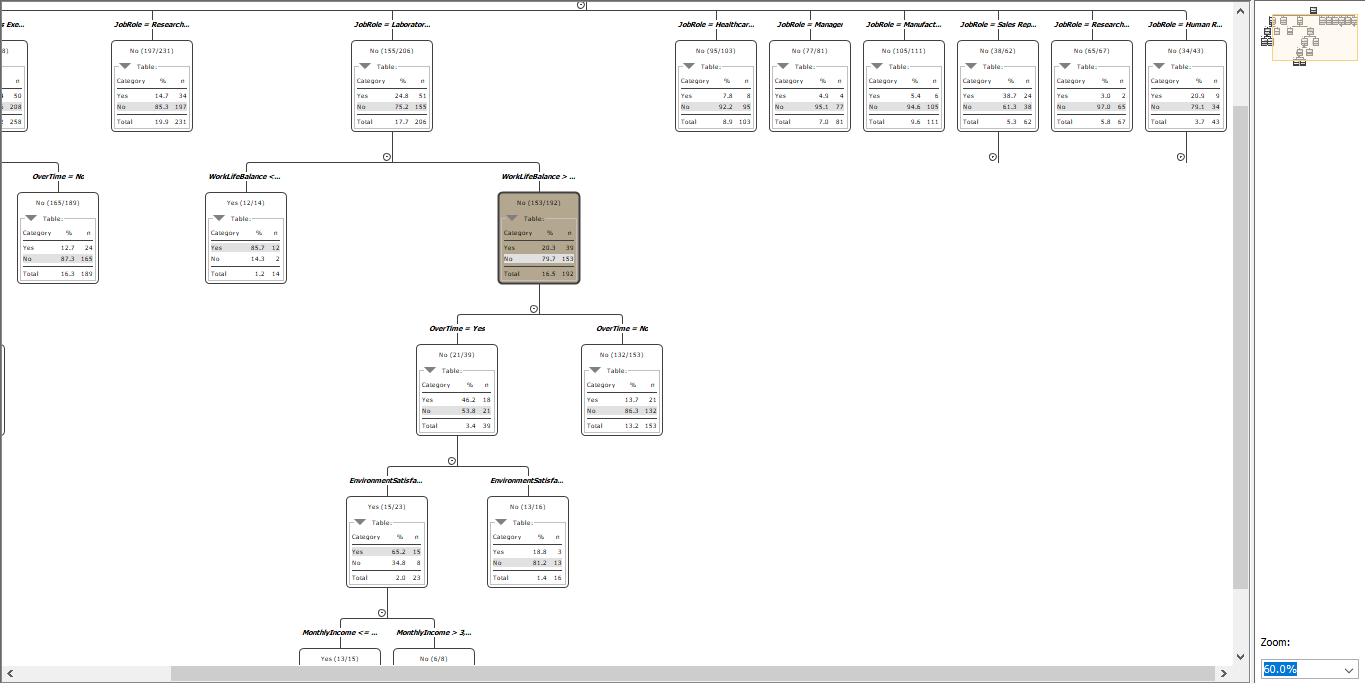


**Appendix 2**: KNIME Boxplot showing Employee Income distribution

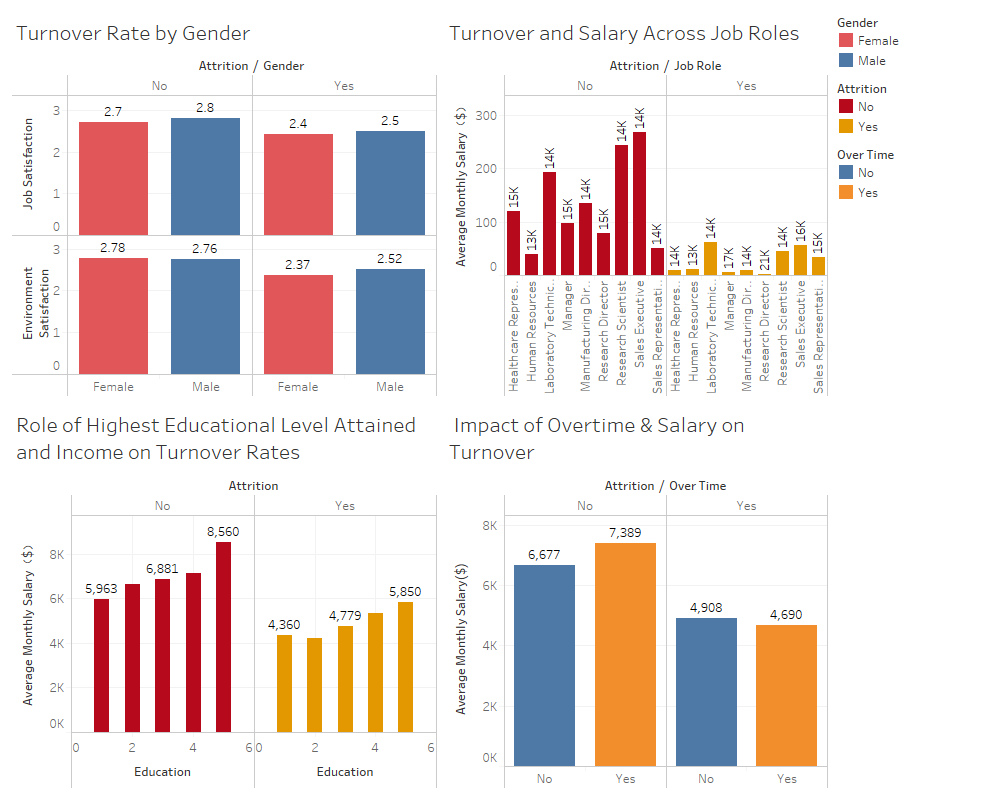


**Appendix 3:** Visualisation and Summary Workflow 

**Appendix 4:** Decision Tree Output (Knime)

****

**Appendix 5：**HRdashboard

****

**Appendix B: Activity Report**

5/11/2020:

First Group meeting, defining early group roles, minutes recorded.

11/11/2020:

Joe: Looked at Data Quality Issues in Knime-produced provisional clean dataset for use in Tableau. Set up Google Drive.

Laura: Getting familiar with Knime, also looking at data quality issues in Dataframe

Maria: Looking at literature to begin writing Background section of Written Report

12/11/2020:

Maria: Research + Writing introduction and background of the problem section

13/11/2020:

2nd meeting- defined group tasks split for next week (Joe and Laura prediction in Knime, Maria, Shiyan, Tao visualisations in Tableau) and reviewed what was done so far, minutes recorded

Joe: Completed final quality fixes in Knime, also some basic Knime visualisations , sent out for use in Tableau, started writing methodology for this part

Laura: Working on excluding certain variables to increase accuracy of basic predictive knime model built.

Shiyan and Tao: made some provisional visualisations in Tableau using clean dataset.

20/11/2020

Third meeting, further discussion of visualisations and modelling, and decision to do extra models.

05/12/2020

Laura: Writing methodology and results based on decision tree

Shiyan and Tao: Writing methodology (Tableau) and results, based on visualisations

9/12/2020

Maria: Linked results with previous literature and research

10/12/2020

Everyone: Final run-through and corrections

**DECLARATION**

All students contributed equally to the group and we understand that each member will receive the same mark for the assignment.

Signed:

*Maria Giannopoulou (40307373)*

*Laura Boden (40199156)*

*Joseph Deery (40078793)*

*Tao Luo (40287847)*

*Shiyan Xu (40285819)*