**Employee learning data and demographic information as an aid in the succession planning process - the role of data analytics.**

Sinéad M Duffy

A Thesis submitted in partial fulfilment

of the requirements for the

Degree of

Master of Science in Data Analytics

Diagram

Description automatically generated with medium confidence

September 2023

Supervisor: Marina Soledad Iantorno

**CCT College Dublin**

Assessment Cover Page

To be provided separately as a word doc for students to include with every submission.

|  |  |
| --- | --- |
| **Module Title:** | Capstone Project |
| **Assessment Title:** | Employee learning data and demographic information as an aid in the succession planning process - the role of data analytics. |
| **Supervisor Name:** | Marina Soledad Iantorno |
| **Student Full Name:** | Sinéad Duffy |
| **Student Number:** | SBA22229 |
| **Assessment Due Date:** | Friday, 22 September 2023 |
| **Date of Submission:** | Friday, 22 September 2023 |

Declaration

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**Acknowledgement**

I would like to thank the following people who have helped me undertake this research:

My supervisor Marina Soledad Iantorno., for her support, encouragement, and patience.

David, Debra and the Team at CCT

Colleagues in both Ireland and the UK that guided and directed me throughout the last 12 months, and most especially to those who kindly agreed to take part as Interviewees.

Kathleen, Helen, Eilis and Colm as well as Michelle and everyone else who tirelessly supported, encouraged and cajoled at various parts throughout this process.

And to my parents, who set me off on the road to this MSc a long time ago.

**Abstract**

Contents

[Table of Figures 7](#_Toc146039319)

[Table of Tables 7](#_Toc146039320)

[1. Chapter 1: Introduction 8](#_Toc146039321)

[2. Chapter 2: Research Design 10](#_Toc146039322)

[2.1. Problem Identification and Clarification 10](#_Toc146039323)

[2.2. Research Objectives 11](#_Toc146039324)

[2.3. Primary Research - Data Collection 12](#_Toc146039325)

[2.4. Secondary Data Collection 13](#_Toc146039326)

[2.5. Validity Type 13](#_Toc146039327)

[2.6. Ethical Considerations 15](#_Toc146039328)

[3. Chapter 3: Literature Review 16](#_Toc146039329)

[3.1. Human Resources and data analytics 16](#_Toc146039330)

[3.2. Learning Analytics 18](#_Toc146039331)

[3.3. Data Analysis of HR Data 21](#_Toc146039332)

[3.4. Conclusion 23](#_Toc146039333)

[4. Chapter 4: Methodology 24](#_Toc146039334)

[4.1. Dataset 24](#_Toc146039335)

[4.2. Python Libraries - Calculations, Graphs and Analysis 25](#_Toc146039336)

[4.3. Exploratory data analysis 26](#_Toc146039337)

[4.4. Encoding Data Types 27](#_Toc146039338)

[4.5. Machine Learning Models 27](#_Toc146039339)

[4.6. Algorithm 1 - Logistic Regression 28](#_Toc146039340)

[4.7. Algorithm 2 - Decision Trees 28](#_Toc146039341)

[4.8. Algorithm 3 - Support Vector Machines 29](#_Toc146039342)

[4.9. Algorithm 4 - Random Forest 29](#_Toc146039343)

[4.10. Algorithm 5 - Multi-Layer Perceptron (MLP) 30](#_Toc146039344)

[5. Chapter 5: Implementation and Results 31](#_Toc146039345)

[5.1. Dataset 31](#_Toc146039346)

[5.2. Independent Variables 31](#_Toc146039347)

[5.3. Algorithm 1 - Logistic Regression 32](#_Toc146039348)

[5.4. Algorithm 2 - Decision Tree 32](#_Toc146039349)

[5.5. Algorithm 3 - Support Vector Machines (SVM) 33](#_Toc146039350)

[5.6. Algorithm 4 - Random Forest 34](#_Toc146039351)

[5.7. Algorithm5 - Multi-Layer Perceptron (MLP) 36](#_Toc146039352)

[5.8. Analysis of Selected Variables 37](#_Toc146039353)

[6. Chapter 6: Discussion 40](#_Toc146039354)

[6.1. Objective 1 - Gender Variable 40](#_Toc146039355)

[6.2. Objective 2 - Studied\_Credits Variable 42](#_Toc146039356)

[6.3. Objective 3 - Tenure Variable 44](#_Toc146039357)

[7. Chapter 7: Conclusion 46](#_Toc146039358)

[8. Appendix A: Plan of Work 48](#_Toc146039359)

[9. Appendix B: Interview Transcripts 49](#_Toc146039360)

[9.1. Interview A - Transcript 49](#_Toc146039361)

[9.2. Interview B - Transcript 73](#_Toc146039362)

[9.3. Interview C - Transcript 89](#_Toc146039363)

[10. Appendix C: Sample Consent Form 98](#_Toc146039364)

[11. Reference List 100](#_Toc146039365)

# Table of Figures

[Figure 1- Charting the change in management requests for HR. Source: Diez et al (2019) page 5 17](#_Toc145970521)

[Figure 2- Python Libraries used for analysis. 25](#_Toc145970522)

[Figure 3- Exploratory Data Analysis steps (Peng et al., 2021) 26](#_Toc145970523)

[Figure 4 - Articles and Models reference table 28](#_Toc145970524)

[Figure 5- SVM analysis with hyperparameter tuning applied. 34](#_Toc145970525)

[Figure 6 - Random Forest with hyperparameter tuning applied. 35](#_Toc145970526)

[Figure 7 - Graph of Results by Gender. 41](#_Toc145970527)

[Figure 8 - Graph of Results by Studied\_Credits. 43](#_Toc145970528)

[Figure 9 - Graph of Results by Tenure 45](#_Toc145970529)

# Table of Tables

[Table 1 - Algorithm 1 Logistic Regression Results 32](#_Toc146048672)

[Table 2 - Algorithm 2 - Decision Tree Results (Tenure) 33](#_Toc146048673)

[Table 3- Algorithm 2 - Decision Tree Results (studied\_credits) 33](#_Toc146048674)

[Table 4 - Algorithm 3 SVM 34](#_Toc146048675)

[Table 5 - Algorithm 4 - Random Forest results. 35](#_Toc146048676)

[Table 6 - Algorithm 5 - MPL implementation 36](#_Toc146048677)

[Table 7 - MLP implementation results of loss and accuracy 36](#_Toc146048678)

[Table 8 - MLP Results with hyperparameter tuning. 37](#_Toc146048679)

[Table 9 - Accuracy Scores by Algorithm and by Target Variable 38](#_Toc146048680)

[Table 10- Initial results v's Test results 38](#_Toc146048681)

[Table 11 - Loss results for Algorithm 5 - MLP 39](#_Toc146048682)

# Chapter 1: Introduction

The author has been working in Human Resource Management for nearly 20 years, specifically within a multinational company for nearly half that time. They have observed how much data is collected throughout the department such as

1. **employee data** in databases such as SAP SuccessFactors, Workday, ADP Workforce etc.
2. **time and attendance data** in time management systems
3. **compensation and benefit data** in benefit platforms
4. **employee engagement data** using employee experience systems, including performance management data.
5. **employee expenses** in financial management systems
6. **talent management** software that allows management of the recruitment process, onboarding of employees as well as ongoing management of talent
7. **learning management systems** that structure learning experiences and ensuring compliance with training requirements or continuous professional development etc.

In the authors experience, all these systems operate independently of each other. For example, SAP and Workday may incorporate time and attendance tracking, talent management and some payroll processing or each may be a stand-alone system. A level of integration with APIs (Application Programming Interface) facilitating a connection to share basic data such as employee name and employee number as well as work email, manager, work area etc. Beyond this, there appears to be very little integration into the wider financial governance of expenses, benefit management platforms or indeed platforms that track and detail the employee experiences such as engagement.

The author wanted to utilise training data gathered from both local and corporate systems to see if there is any value in using this to support the succession planning process. As it currently stands, succession planning is largely a manual process, where Human Resource Business Partners speak with employees to identify areas that they would like to develop, what they feel their key skills are, identify if the employee is interested in moving within the company etc - see appendix 2. How this information is gathered is individualized and based on the reporters’ own experiences. At different times during the year, this information is collated, and meetings are held at functional area lead level who, in conjunction with other senior managers hold succession planning conversations. Succession Planning is a long-term planning strategy within HR, where every existing role within the company has a successor identified. It is an important planning process and helps the company develop internal talent as well as minimising business interruptions in the case where there is an unplanned departure (*Importance of Succession Planning (With Benefits and Tips) | Indeed.com Canada*, no date). Currently training data (such as courses completed) is not included as a metric in the process. The author would like to explore if there is a role for training data within the succession planning process. If it is possible to identify such a role, what it would look like.

Tambe et al in an article called ‘Artificial Intelligence in Human Resources Management: Challenges and a Path Forward’ (Tambe, Cappelli and Yakubovich, 2019) discuss the challenge that is faced when using HR data for machine learning. The article outlined that datasets from human resources can be small with not commonly repeating events (such as dismissals) or are influenced by external factors such as employment law (Equality Acts) or company policies (gender balance policies). With small datasets taking account of influences outlined previously, there is an opportunity to look at relationships in the data through the lenses of relationships rather than prediction from correlations of observed variables as in other areas of machine learning algorithms.

Initially, the author planned to use data from the learning management system (LMS) within the multinational company where they work. However, due to data privacy and ethical issues as well as the need for a non-disclosure agreement and limited access to proprietary information led the author to reevaluate this data source. An alternative data source was identified as OULAD - the Open University Learning Analytics Dataset which will be discussed in more detail in subsequent chapters. To aid their understanding of the succession planning process, the author was able to utilise the expertise of colleagues within the Human Resources community to discuss their experiences both at the current employing company and beyond.

# Chapter 2: Research Design

## Problem Identification and Clarification

The lack of any visible link between learning data and succession planning led the author to consider if some form of data analysis could be utilised to enhance the process. Experts in the area outline attempts to digitise the process by using standard templates to upload data to analytics tools used such as Power BI and Tableau, appendix 2. The author could not identify any attempts to create this link within the company, or the wider HR community, or even within literature. Indeed, the author could not find any literature where analysis was attempted on learning data within a work environment.

In the author’s own experience, employee interaction with learning systems can be mixed. Some employees complete mandatory learnings when assigned, others are assigned learnings as a development aid to their job. There is another category of learners that need to be considered though, and that is those who complete learnings based on their own interests, or to increase their knowledge of how the company works. It’s these outliers that interest the author as managers and human resources may not be familiar with either employees’ areas of interest, the learnings or are out the scope of their normal job. It is also possible that additional studies undertaken are done so outside the company, and additional learning is completed to support this.

There are other influences that can be analysed to determine their impact on an employee’s choice such age, gender, education level, access to education etc. The selected dataset from OULAD, has many of these features that can be used for such analysis such as age, education, result, gender, and the number of credits that are being studied for. There is a column for region which identifies where the student is from at the time that they registered for the course. Also included is the column imd\_band, which identifies the Index of Multiple Depravation that plots the areas in the UK based which are relatively deprived based on socio-economic factors (Alhakbani and Alnassar, 2022; ‘Multiple deprivation index’, 2023). The author had discounted the region and imd\_band columns, as in practice, employees generally live within commuting distance to the workplace, and as such this information is less relevant. The addition of a column on tenure will help mimic the relative experience the employee has with the company and will be included for analysis.

Taking all of this into account, the focus of this study will be on determining if learning data and demographic information for employees can be used as an aid in the succession planning process using data analytics to complete the analysis.

## Research Objectives

Now that the focus of the research project has been identified, it’s key to clarify what the research objectives are going to be support this analysis. The research project is focusing on the impact of learning data on succession planning using demographic features as the one common data point across the dataset, and in real data held within the company. Traditionally, demographic features are defined as age, gender, ethnicity, disability and family status (Tsui and Gutek, 1999; Clair *et al.*, 2019). It has to be acknowledged that current thinking within organisations show a more inclusive view of the definition of demographic traits such as gender fluid employees, whilst academia recognises that demographic traits are more fluid than previously defined (Clair *et al.*, 2019). The researcher acknowledges these definitions. However, the dataset selected for use as part of this analysis does not reflect such updated thinking, and standard demographic information will be used for analysis. This research paper will focus on identifying if there is a machine learning algorithm that is effective in accurately predicting, the success of an employee in the succession planning process using traditional demographic information.

Specifically, the initial research objectives identified are:

**Objective 1** - is an employee’s gender a reference point for succession planning?

**Objective 2** - using studied\_credits as a substitute for number of courses completed, what machine learning algorithm will provide an accurate measure for succession planning?

**Objective 3** - does employee tenure have an impact on a succession planning?

The author acknowledges that ethical issues and potential bias could become apparent when working with demographic variables, the author is not advocating that any resulting output model replaces the existing processes in use within the company. Rather any potential output be used as an aid to the process in terms of decision making.

## Primary Research - Data Collection

As part of the research proposal a quantitative approach to collecting primary research data was deemed the most appropriate method of gaining first person information from sources well versed in the succession planning process specifically within a multinational company. Such data collection was completed using in person interviews. Unstructured interviews were completed online using the Microsoft Teams platform which enabled transcription to be completed automatically. This method allows for in depth discussion on the area of succession planning, as well as allowing a level of observation to be used by the author to gauge reactions to questions posed and adjust the flow of the interview depending on the interviewee’s reply to the question (Kumar, 2011; Saunders, Lewis and Thornhill, 2012; Wilson, 2013). In essence, this method allows the author and interviewee to have a conversation that moves organically through topics giving an opportunity to probe for further understanding where necessary.

To ensure that the right mix of experts are chosen, Saunders et al (2012) outline that identifying the characteristics of the experts prior to selection will create a more rounded group of experts. To that end, the author has identified two characteristics that would be key in answering the research objectives posed as part of this research, namely;

• Have some involvement in implementing or improving processes with HR and the wider company.

• Be ‘outward looking’ in that they are knowledgeable of company strategy as well as best practices within the market.

To overcome any potential bias that occurs on the interviewee’s behalf, the author sought input from different individuals who have taken part in the succession planning process over several years, either in the multinational company itself, or as part of their previous roles with similar organisations. In terms of gaining a holistic view of the succession planning process, the interviewees have been identified from different parts of the organisation. One interviewee manages the succession planning process for the manufacturing division of the company, another manages the process for office-based employees located around the wider European Economic Area. Yet another interviewee has experience in both areas of the business outlined previously and is well placed to share similarities and differences in the two processes. Transcripts from the interviews can be seen in appendix 2.

## Secondary Data Collection

As outlined previously, initially, the author hoped to use data from the company’s own LMS system. As an alternative, the Open University Learning Analytics Dataset (OULAD) was selected as the dataset upon which analysis could be completed (Kuzilek, Hlosta and Zdrahal, 2017). The data file specifically selected for analysis was the ‘student\_info.csv’ file.

The dataset OULAD closely mimics data typically found in a commercial operation and extracted from an LMS within the company such as:

* **Student\_Id** = representation for employee number
* **Studied\_Credit** = as a simulation / substitute for level of interaction with the system i.e. the more credits gained, the more courses completed.
* **Tenure** = the addition of a column on tenure as a random variable to simulate the number of years the given employee / student has been with the company.
* **Gender** = the gender of the employee / student
* **Age** = grouped by age bands
* **Previous education** = highest level of education achieved by the student / employee.

The use of OULAD reduces the risks associated with data privacy and data ethics. It also allows for the data to be filtered to one module over one semester - thus allowing the data to represent one manufacturing site within a defined period. It also allows the author to closely mimic the number of employees within a manufacturing environment.

## Validity Type

Considering the research objectives outlined in the previous chapter, it is possible to say that the most relevant components of validity, relevant to this research are accuracy, currency, and bias (Kumar, 2011; Saunders, Lewis and Thornhill, 2012). It is however also possible to say all components of validity apply to the proposed research, some component’s more than others. The author has attempted to minimise overall validity by rooting concepts and models utilised in this research within a academic literature as well as practitioners experiences (Kumar, 2011). The concepts of accuracy, currency, and bias are explored below.

Accuracy in this instance relates to how comprehensive the data statistically is (Kumar, 2011; Saunders, Lewis and Thornhill, 2012). In terms of primary data, accuracy does not apply as the data is not statistically based. The data captured from interviews will need to be transcribed and included in the appendices of this report. Furthermore, the main points and sentiments expressed will be used to verify if the proposed model will be useful or not.

Currency in this instance is a potential barrier to the methodology of this research (Kumar, 2011; Saunders, Lewis and Thornhill, 2012). The author has chosen to use data extracted from an educational institute learning management system. The data was released in 2017 and contains data from 2013. It is true to say that the data is not current, however, it closely mimics the data is contained within the company’s own LMS. With the addition of simulated data in the form of the tenure column the author has chosen to accept the risk to the validity of the results of this research paper. As part of the future work of this research, the author advocates the need to compare results of this research with real world data.

Bias has already been identified as a possible threat to validity when conducting in-depth interviews for primary research. The author will attempt to limit bias by ensuring that there is a clear purpose of the interview which is communicated in advance. By working with known participants there is already a degree of trust established between the parties to facilitate a frank discussion. Finally, the author will create several prompts based on key research themes that will help guide the interview process and stay within the research area.

Although three components have been listed, it is not unreasonable to assert that other components may also become more apparent as this research progresses.

## Ethical Considerations

As with all research, there are ethical considerations that will need to be planned for, some of which have been outlined previously. In respect of the scheduled interviews, participants have been asked to take part, and have been given the option to withdraw their consent or have their data excluded at any stage of the process up to the final submission date. All interview participants are over 18 years of age and have not disclosed any medical condition or any other prohibition that will limit their ability to take part in the interview process. No incentives have been given to any participant to gain their support in the research process. As an added measure, all interviews have been transcribed for completeness and included in appendix 2 of this document. If any participants have questions at the end of the interview process, or in the time up to the submission date, the author has outlined a communication process that will allow for speedy resolution to these queries as quickly and sensitively as possible.

In respect of secondary data, due to data protection and sensitivity issues the author decided to use a dataset obtained from the Open University (OULAD) (Kuzilek, Hlosta and Zdrahal, 2017). This dataset was selected as it closely mimicked an extract of the LMS system used within the selected Company. The OULAD dataset contains more the 34000 data points which have already been anonymised, thus limiting any potential data breach. The General Data Protection Regulations (GDPR) outlines companies’ requirements to protect the private data of individuals (*General Data Protection Regulation (GDPR) – Official Legal Text*, no date). It also enshrines the concept of privacy by design -where anyone working with, or handling data needs to have sufficient security measures in place to secure the data from any potential risks. The decision to use a widely available dataset instead of actual employee data is a key reason that OULAD was selected for analysis. As such, the author has attempted to minimise potential risks to the company’s data, whilst also maintaining compliance with the companies own internal GDPR and ethics procedures.

# Chapter 3: Literature Review

## Human Resources and data analytics

HR data provides a lot of opportunity for analysis within companies (Rasmussen and Ulrich, 2015; Mattox II, Parskey and Hall, 2020; Ferrar and Green, 2021). Mattox et al (2020) in their book ‘*Learning Analytics*’ outline the pressure from business leaders to provide better and more insightful information in a timely manner. The demand for information is coming not just from Senior Managers, but also from stakeholders who want to know more about the people function and how effective it is (Topno, 2012; Mattox II, Parskey and Hall, 2020; Ferrar and Green, 2021). David Ulrich outlines how people analytics can add value to companies by allowing teams to make informed decisions led with data in support of the business (Ferrar and Green, 2021).

People analytics as defined by Ferrar et al (2021) is ‘*The analysis of employee and workforce data to reveal insights and provide recommendations to improve business outcomes*.’ Numerous authors outline the importance of using data analytics to empower business decisions within the Human Resources Function (Ferrar et al. 2021, Mattox et all 2020, Rasmussen and Ulrich 2015, Diez et al (Rasmussen and Ulrich, 2015; Diez, Bussin and Lee, 2019; Mattox II, Parskey and Hall, 2020; Ferrar and Green, 2021). Rasmussen and Ulrich (2015) however point out the need to ask the ‘*right question*’ when reviewing data generated by HR and propose that this question should be incorporated into the end-to-end analytics process to identify and confirm the impact of people on decisions. Diez et al (2019) in their book Fundamentals of HR Analytics outline how questions for HR are evolving and why there is now a need for HR speak the same language as senior management and other functions within the company. The specific fields within HR are called out in Figure 1 below.

A white rectangular box with text

Description automatically generated with medium confidence

Figure 1- Charting the change in management requests for HR. Source: Diez et al (2019) page 5

HR needs to prove its importance to the business, especially in terms of how impactful it’s action are on the overall financial health of the company (Dong, 2022; Losey, Meisinger and Ulrich, 2005), (pp 121). In monetary terms analysis has shown that small changes to processes can make cost savings for businesses such as implementing training reminders to cut down on the amount of time to complete induction, or to uncover a link between engagement data and business performance (Ferrar and Green, 2021) (pp. 4). There is an opportunity for data analytics within HR, whilst also recognising the need for help from HR experts to interpret the results of any analysis (Edwards and Edwards, 2019) - (pp. 5). In truth HR need to refocus their role to become a ‘strategic partner’ of the business helping it to achieve its strategic goals (Bhardwaj and Patnaik, 2019; Dahlbom *et al.*, 2020; Losey, Meisinger and Ulrich, 2005) (pp 150). Academics are aligned on the need for HR to upskill and become ‘ambassadors’ for data analytics as a means of driving data driven decision making (Martin, 2019).

To balance out this desire, HR data is uniquely different from other types of data used for analysis (Tambe, Cappelli and Yakubovich, 2019; Bankins, 2021). By its nature, data gathered by HR is formed of generally small datasets where events that companies want to model or predict are infrequent and nonstandard (for example dismissal of employees) or the data is subject to interpretation such as performance management where employees with different roles and responsibilities cannot easily be compared (Chadwick and Dabu, 2009; Tambe, Cappelli and Yakubovich, 2019; Bankins, 2021). Another issue with HR data relates to external requirements on the company which are not evident in other functions. For example, the recruitment process is influenced by internal factors such as the company’s own recruitment goals, as well as external ones such as the statutory landscape (Tambe, Cappelli and Yakubovich, 2019). This fact forces companies to limit the use of historical data such as recruitment data as it’s use could make incorrect predictions based on outdated information, or based on practices that are no longer the same (Rasmussen and Ulrich, 2015; Tambe, Cappelli and Yakubovich, 2019). As Bhardwaj et al (2019) stated ‘*Human resource analytics is an area of study that uses the mix of art and science on human capital in order to get measurable return on investment*’, (Bhardwaj and Patnaik, 2019).

One could argue that the future for HR data is to become integrated into the wider information stream of the company as a method to identifying how individual’s performance affects the wider company performance (Rasmussen and Ulrich, 2015; Tambe, Cappelli and Yakubovich, 2019). Rasmussen et al (2015) outline that impactful HR analytics are about linking to strategic business operations rather than trying to identify patterns in big data (Rasmussen and Ulrich, 2015). Some academic’s espouse the opinion that to be used successfully, HR data must be taken away from the HR department for analysis (Rasmussen and Ulrich, 2015; Ferrar and Green, 2021). Experience in one case-study outlined by Ferrar et al (pp 20 - 26), confirms that HR data is different to other types of data and to successfully analyse it HR must be included in system development (Ferrar and Green, 2021).

## Learning Analytics

Building on this, learning analytics on the other hand, focuses on the effectiveness of a learner’s experience (Mattox II, Parskey and Hall, 2020; ‘Handbook of Learning Analytics - Second edition’, 2022) and is routed in basic training evaluation models such as the Four Levels of Evaluation model developed by Don Kirkpatrick (Mattox et al 2020). Specifically in this research paper, the author will focus on training provided solely within a corporate structure, an area with limited research carried out (‘Handbook of Learning Analytics - Second edition’, 2022). Implementing a Learning Management Systems (LMS) has provided an effective way of gathering, analysing, and reporting on learning related data (Katrina Sin and Loganathan Muthu, 2015; Mattox II, Parskey and Hall, 2020; Karakolis *et al.*, 2022; ‘Handbook of Learning Analytics - Second edition’, 2022). LMS’s such as Moodle have long been used in academic circles and have provided rich data sources in understanding how students learn and interact with systems, and are becoming the ‘backbone’ of analysis within companies (Sin and Muthu, 2015, Arka et al 2022, (Katrina Sin and Loganathan Muthu, 2015; Mattox II, Parskey and Hall, 2020; Karakolis *et al.*, 2022; ‘Handbook of Learning Analytics - Second edition’, no date).

This paper is an attempt to identify if a link or relationship can be found between training undertaken by employees and area’s such as succession planning within a multinational environment. The literature review of academic and related papers has helped to uncover several opportunities for further analysis of data held within the HR Department focusing specifically on data relating to learners. The author could uncover very little research into learning data held outside of academia while Allison Littlejohn provided the closest match in her chapter on Professional Learning Analytics as part of the Handbook of Learning Analytics (Handbook of Learning Analytics - Second edition’, no date). Most articles found outlined analysis based on data from educational institutions. How analysis of education related data should be conducted was, as expected, discussed at length with different approaches being taken as outlined in Figure 5.

Focusing on data gathered as part of the learning process and how such analysis might be completed is discussed in the following sections. A range of methods are used to create a dataset, and this is particularly true for online or distance learning (Sin and Muthu, 2015, Arka et al 2022). Systems such as Moodle allow analysts to follow a student’s learning path through a module or full course of study (Sin and Muthu, 2015, Arka et al 2022). Shen and Chi (2016) analysed how different levels of learners reacted to different methods of learning using such online interactions. In practice companies use systems such as LMS’ to collate learning data from employee interactions. An LMS (a Learning Management System) is a system that allows companies to manage training within the company, which then allows companies to run reports, track training requirements, assign learnings etc (*The LMS Guidebook : Learning Management Systems Demystified*, 2018)(Chapter 1). The advantages of using such a system is advanced features such as dashboards and reports created displaying high-level overviews of the data contained within the LMS as well as the ability to interlink with existing systems within the HR department (*The LMS Guidebook : Learning Management Systems Demystified*, 2018). The amount of data incorporated into an LMS means that large datasets can potentially be extracted, and it may be necessary to use data mining techniques to focus such big data sources (Sin and Muthu, 2015, Arka et al 2022, Mattox et al 2021).

The advent of LMS systems has led to a culture of self-directed learning by employees within companies (*The LMS Guidebook : Learning Management Systems Demystified*, 2018). Self-directed learning is where the employee is in charge of their own learning journey, a method of learning that is gaining traction in recent times (Araka *et al.*, 2022; Mustafa Yağcı, 2022). The drive to this new method of learning is coming from both companies as they roll out new technologies and employees themselves as they become more data savvy (Mattox II, Parskey and Hall, 2020; Araka *et al.*, 2022). The drive towards digitisation has only increased since the onset of Covid-19 and the need for companies and employees to adapt to increasing digital offerings (Almeida, Duarte Santos and Augusto Monteiro, 2020). Kokoc et al (2021) present the theory that by giving learners (employees) access to a dashboard to support their individual learning journey they will have more motivation to develop based on consistent feedback on their progress (Kokoç and Altun, 2021).

The development of this new area of learning has given more scope to allow machine learning to analyse the resulting data to help predict different outcomes - especially within educational settings (Araka *et al.*, 2022; Mustafa Yağcı, 2022). Analysis completed by academics chart learner performance against system access, and compare the results to final exam results (Araka *et al.*, 2022; Mustafa Yağcı, 2022). In companies, a different but similar approach is needed to gauge employee progress. For clarity, learning analytics has many definitions, but the one used in this paper is that learning analytics is the method of collecting, analysing, interpreting and reporting data to inform and understand learning methods and environments with a view to making improvements (Mattox II, Parskey and Hall, 2020; Kokoç and Altun, 2021; Araka *et al.*, 2022; Mustafa Yağcı, 2022). Educational data mining has emerged as a new field in which to access learning data stored in data warehouses or data lakes and seeks to work to open learning data to new analysis methods (*Learning Analytics – A Growing Field and Community Engagement*, 2015; Araka *et al.*, 2022; Mustafa Yağcı, 2022).

Deloitte in their 2017 Global Human Capital Trends outline that HR leaders, and specifically Learning & Development (L&D) leaders should reassess how they think about an employee’s learning journey and ‘inspire’ employees to develop deeper skills with a view to enabling employees to change positions within their respective companies (‘2017 Deloitte Global Human Capital Trends’, 2017) (pp 36). The Deloitte report goes on to outline a case study about AT&T where they focus on career development for their employees and encourage them to change roles every four years as part of employees ongoing development (pp 36). Sources outline reasons that employee should ideally be seeking new experiences every three to five years such as keeping in touch with outside trends, that employees become comfortable with change as some of the key items (Ryan, 2016; Christian, 2022).

As outlined previously, the succession planning process is critical to the business’ ability to develop its employees (*Importance of Succession Planning (With Benefits and Tips) | Indeed.com Canada*, no date). Huselid et al (2005) agree and outline that it is better to identify roles that are critical for the business and then spend time investing in the development of employees going into those roles to ensure that the right people are in place to drive the business forward (Huselid, Beatty and Becker, 2005).

## Data Analysis of HR Data

Outlined in the previous section is the need for HR to become more integrated and aligned with Senior leaders within business (Ferrar et al. 2021, Mattox et all 2020, Rasmussen and Ulrich 2015, Diez et al (Rasmussen and Ulrich, 2015; Diez, Bussin and Lee, 2019; Mattox II, Parskey and Hall, 2020; Ferrar and Green, 2021). How can this be accomplished though?

Tambe et al (2019) put forward the argument that data within the HR Department generally contains small datasets which may not be suitable to use to clearly identify relationships or meaningful insights within the dataset (Tambe, Cappelli and Yakubovich, 2019). This fact is further complicated by any decision to use historical data for analysis, with the danger being that historical HR data may unwittingly contain a bias towards non-traditional employees within a workplace such as a bias towards men against women where historical data is largely collected on male employees as women were underrepresented at the time period (Loftus *et al.*, 2018; Tambe, Cappelli and Yakubovich, 2019; Vowels, Cihan Camgoz and Bowden, 2023). Therefore to limit potential bias when analysing HR data, whilst working with small datasets it is necessary to use other sources such as theory and prior research as a guide to identify models and potential relationships in the data (Tambe, Cappelli and Yakubovich, 2019).

In addition to bias, it is important to note that in some fields it is not ethical to seek relationships between variables, especially in scenarios where there may be moral considerations to be taken into account (Eberhardt, 2017; Malinsky and Danks, 2018; Vowels, Cihan Camgoz and Bowden, 2023). Bankins et al (2021) has proposed a framework to help with the ethical implementation of artificial intelligence within an organisation with scope to use for implementation of any type of machine learning or predictive analysis (Bankins, 2021).

Working forward, Muslim et al (2023) completed a review of literature on ‘Open Learning Analytics’ outlining key frameworks within the field of learning analytics (Muslim, Chatti and Guesmi, 2020). That paper showcased that in terms of implementation, when applying machine learning algorithms for learning analysis, there are a myriad of implementations possible, and indeed in use (Muslim, Chatti and Guesmi, 2020). Techniques used by authors across the research area are outlined in below:

* clustering / classification techniques (Araka *et al.*, 2022; Mustafa Yağcı, 2022)
* data mining based on time data (Araka *et al.*, 2022; Assaad, Devijver and Gaussier, 2022)
* Random forests / Support Vectors / Naïve Bayes techniques (Hussain *et al.*, 2018; Mustafa Yağcı, 2022; Zhang *et al.*, 2023)
* Neural networks such as MLP, ANN and CNN (Kulala and Rani, 2017; Chunqiao Mi, 2019; Poudyal, Mohammadi-Aragh and Ball, 2022; Wang, Guo and Shen, 2022; Al-azazi and Ghurab, 2023)

An example of where HR data has been implemented within an integrated approach to analysis including the use of neural networks, artificial intelligence, has been proposed by Dong (2022) for an integrated system called the Human Resource Intelligence System (IHRMS) (Dong, 2022).

As is expected, several alternative means of analysis is possible, and the author will need to experiment to identify the method and / or algorithm that will give the best results for the individual dataset chosen.

## Conclusion

Reviewing work by other authors is a key step in completing any analysis. In this instance, the review was able to highlight the need for a more analytical mindset when considering HR data or learning data and outlined the benefits of same in terms of closer strategic connections to the business. Opportunities exist for such data to be used to allow better decision making by business leaders, but the HR department must enable this by being ‘ambassadors’ for their data. Systems such as an LMS should be used to leverage learning analysis within companies, as well as directing more effective and strategic learning outcomes-based on sound data analysis.

# Chapter 4: Methodology

Chapter 3 provides the strong foundation on which the methodology for this research paper is based. The author reviewed the processes that several different research papers had previously taken when using the OULAD dataset, amongst others. This provided a launch point upon which this research could be built upon. Specifically, the author was able to identify models already used to explore the dataset, as well as the research focus of the article. Generally, research was dedicated to developing models predicting student performance or engagement with the systems within a learning environment. This research paper however is concerned with learning data accumulated within the work environment, where learning opportunities may be more limited or directed by the company rather than only at the interest of the learner. In addition, previous research focused on interaction with the system by students as a method of assessing engagement or performance. Neither of these areas are a focus for this research paper. However, the author felt an opportunity existed to use prior learning and apply it within this different area of study.

## Dataset

As outlined in previous chapters, the dataset used for analysis was sourced from the Open University Learning (Kuzilek, Hlosta and Zdrahal, 2017). This was selected as it provides similar data structure and content to that found in LMS within commercial companies. Using an existing dataset allowed the author to clearly see analysis conducted by other researchers around learning analytics. This research paper differs from those others in that the main aim of this research is not to identify students at risk of dropping out or of poor performance, but rather to identify if opportunities exist to utilise learning data to aid succession planning. It also minimises exposure to GDPR considerations, as well as releasing the company from sharing proprietary or otherwise sensitive data.

In terms of the dataset, it was necessary to apply filters to reflect the employee numbers within the target company. As outlined in the sampling strategy adopted for this research, the author is proposing to filter the dataset so that it reflects the employee makeup at the selected operating site. To implement this, the data will be filtered by module\_code, and again by semester. This will allow the author to select the best representative sample similar in number to the employment location.

As outlined previously, the use of OULAD reduces the risks associated with data privacy and data ethics. It also allows for the data to be filtered to one module over one semester - thus allowing the data to represent one manufacturing site within a defined period. It also allows the author to closely mimic the number of employees within a manufacturing environment.

## Python Libraries - Calculations, Graphs and Analysis

Python was the programming language that was used to complete the analysis for this research. Python was selected as it was relatively easy to learn and program for the researcher. The open source code libraries that are widely available within python and can be tailored depending on the needs of the programmer (M.Sc, 2023b, 2023a).

Figure 2- Python Libraries used for analysis.

Numerous articles outline the values of different python libraries that are available for use, with six core libraries being identified as being of use in this analysis. The selected libraries are outlined below .

* Pandas - used for ‘data wrangling’ or manipulation of data within a dataframe (M.Sc, 2023a).
* NumPy - used for calculations quickly and easily (M.Sc, 2023a).
* Matplotlib - helps display graphs and visualisations of the data using pandas and numpy (M.Sc, 2023a).
* seaborn - built to work alongside Matplotlib, the Seaborn library allows for statistical data to be graphed and displayed (*An introduction to seaborn — seaborn 0.12.2 documentation*, no date).
* Scikit-Learn - allows a programmer to quickly implement a range of machine learning algorithms in conjunction with other libraries such as pandas and matplotlib (*scikit-learn: machine learning in Python — scikit-learn 1.3.0 documentation*, no date).
* Scipy - enables a programmer to implement statistics and other mathematical computations with python (*SciPy*, no date).
* Keras - facilitates deep learning within Python by working in conjunction with TensorFlow to provide a simple but flexible platform for analysis (Team, no date).

## Exploratory data analysis

As with all data analysis projects, the exploratory data analysis (EDA) stage is crucial to ensure that the resulting analysis will be valid and not biased one way or the other. The steps undertaken as part of EDA in Notebook 00 - EDA on - Student\_Info.csv - as outlined by Peng are shown below (Peng *et al.*, 2021).

Figure 3- Exploratory Data Analysis steps (Peng et al., 2021)

## Encoding Data Types

The data used for analysis consists of both ordinal (age band / tenure band etc) and nominal (gender / credit studied etc) data types. Thus, it was necessary to apply a suitable methodology to handle such data. Label Encoding provided a simple solution to resolving data in the ‘gender’ column by converting the values from text to number (1 for Male, 0 for female) (*Categorical encoding using Label-Encoding and One-Hot-Encoder | by Dinesh Yadav | Towards Data Science*, no date) . One -Hot Encoding was selected as the basis to convert ordinal data into numerical values as it is relatively easy to implement, and allows for each categorical data value to give given its own column within the dataframe (Gefferth, 2023; *Categorical encoding using Label-Encoding and One-Hot-Encoder | by Dinesh Yadav | Towards Data Science*, no date; *sklearn.preprocessing.OneHotEncoder*, no date). Both encoding methods (Label Encoding and One-Hot Encoding) are part of the Scikit-Learn Python Library discussed previously.

## Machine Learning Models

All the models used as part of this research project were sourced from literature, as outlined in the figure below. Five different models were identified from the articles, and each are discussed in detail below. The models selected can work with classification data, contained within the OULAD dataset.



Figure 4 - Articles and Models reference table

## Algorithm 1 - Logistic Regression

The Logistic Regression method is widely recommended as a starting point for data analysis models, especially when working with classification data (Kohnke, Foung and Chen, 2022; Python, no date). The first pass of the algorithm was completed without any parameter tuning. Hyperparameter tuning, specifically GridSearchCV was then applied to the algorithm to improve results. Cross Fold Validation was applied in Kohnke et al (2022), the author applied GridSearchCV to the data as it automatically searches for all possible parameter combinations withing the data (*3.2. Tuning the hyper-parameters of an estimator*, no date).

## Algorithm 2 - Decision Trees

As logistic regression was not conclusive in terms of its results, the author decided to implement a similar model capable of working with categorical data. Decision Trees was selected as the second algorithm to test as it had been used by other researchers (Djoundourian, 2017; Hussain *et al.*, 2018). Different methods of hyperparameter tuning were applied in the form of cross validation and GridSearchCV from scikit-learn. Cross Validation was selected as it was used in previous papers (Kohnke, Foung and Chen, 2022; Zhang *et al.*, 2023), whilst scikit learn was selected as the library as both the model and hyperparameter tuning was selected from both. For Cross Validation, a range of features were tried to identify what number of features is optimal for the best results within the model. In addition, the author sought to look at tenure both as a grouped column, and an ungrouped column. This was completed to identify if the there was any difference in the results.

## Algorithm 3 - Support Vector Machines

Support Vector Machines (SVM) is also able to work with classification data, and is able to identify outliers in the data (*1.4. Support Vector Machines — scikit-learn 1.3.0 documentation*, no date). Zhang et al (2023) also used SVM to carry out their analysis on OULAD dataset. Scikit-learn documentation highly recommend that data used for SVM analysis be scaled using the StandardScaler(), and the author intends to follow this advice (*1.4. Support Vector Machines — scikit-learn 1.3.0 documentation*, no date). In terms of other variables, a range of kernels are to be applied to identify which is the most suitable. The kernel criteria utilise different mathematical bases to transform the data across different vectors (for example into 3D to better view the data). Hyperparameter tuning was applied following the first run in the form of GridSearchCV as recommended by Scikit-Learn, and further tuning was applied using a parameter grid (param\_grid) as a further addition to the parameter search to allow for a wider sequence of parameter settings to be included such as ‘C’ and ‘gamma’. The cross-validation parameter in this instance will be set to 5.

## Algorithm 4 - Random Forest

Based on the results achieved using Decision Trees, the author decided to use the Random Forest algorithm from scikit learn. Random Forest is an ‘ensemble’ algorithm which builds a ‘forest’ of decision trees to complete the desired analysis (*sklearn.ensemble.RandomForestClassifier*, no date; *1.11. Ensemble methods*, no date). Ensemble in this instance refers to the classifiers method of combing different estimators (Decision trees in this instance) to arrive at a single result (*1.11. Ensemble methods*, no date). It is possible to select one of three possible criteria to assess the quality of the splits in the data. Two of the methods were used to evaluate which was the most fitting - entropy and gini as the default. Only Entropy was selected as for the remaining criterion as both it and log\_loss use the same method of evaluation. As with previous models, no feature selection or hyperparameter turning was applied in the first run of the model. GridSearchCV and a parameter grid will be implemented to help identify the most appropriate tuning methods based on the results obtained when working with algorithm two. The model will also run with a different n\_estimators or ‘trees’ in the forest of the model to determine what is the optimal number for the algorithm.

## Algorithm 5 - Multi-Layer Perceptron (MLP)

The final algorithm selected is that of a Multi-Layer Perceptron (MLP) neural network. This algorithm was selected as it can work with classification data and is also relatively simple to implement as a baseline neural network. As with the previous algorithms, the author did not find an academic study which used this method for analysis in similar circumstances. However, the author felt it was important to see if applying a neural network algorithm to the dataset would provide any insights to the study. While no hyperparameter tuning was applied to the data, the algorithm was run several times with different number of neurons in the hidden layers to identify if an impact can be seen.

# Chapter 5: Implementation and Results

This chapter outlines how the various algorithms and python libraries were used to complete the analysis of the data. As with the previous chapter, implementation of each algorithm will be outlined on its own merits.

## Dataset

The OULAD data set was imported into Jupyter using the Panda’s library, where it was explored using numerous graphs and tables. The author included an additional column to the dataset as part of the EDA process to mimic employee tenure at the company. The author did not use any of the recommended tools for generation of synthetic data as there is no pattern to employee tenure. As such, the synthetic data created in this column was generated using random function randint, with a seed being set to keep the data consistent once created. The newly created tenure data was then grouped into bands to align with other categorized data in the dataset.

To follow the sampling strategy outlined as part of the research proposal, the author used the data from one semester and one course to complete the analysis. Semester 2013J was identified as the cohort that would more closely align with the number of employees at the manufacturing environment initially selected for evaluation. This resulted in a smaller dataset of approximately 382 rows of data with a new smaller dataset being created for ease of reference.

## Independent Variables

The first variable selected for analysis was that of gender to study if an employee’s gender has any impact on learning and may be useful in helping with the area of succession planning. Note that this data set only records gender as being male or female. The author recognises that more gender types are now common and not reflected in the selected dataset. When implementing the algorithm using real or company data, it will be necessary to reflect employee gender types as they appear within the company’s employee database.

The second independent variable selected for analysis was the data in ‘studied\_credits’. This column was selected as it more closely mimic’s employee interactions with training materials. Those with minimum interactions (such as only completing mandatory training) would have less credits earned than those with more credits would be seen as availing of the courses on offer more frequently. It is important to note here that employees are in control of their own learning journey, and trainings may be recommended by their manager as part of the employee’s own development, or they may source courses on their own initiative.

The final independent variable selected is that of tenure, which reflects the length of time the employee has been with the company. This is synthetic data and added using a random number generator as defined above.

## Algorithm 1 - Logistic Regression

As outlined previously, Logistic Regression was applied to the dataset, initially without hyperparameter tuning, then with GridSearchCV applied. The results of the analysis are outlined below. As the first model run, the test and train results are more than 50% accurate and will act as a base level for test following afterwards.



Table 1 - Algorithm 1 Logistic Regression Results

## Algorithm 2 - Decision Tree

Based on previous research, Decision Tree was the next algorithm implemented. Initially the author used tenure as the independent variable and ran the model twice - including tenure grouped by band in version 1, and using tenure ungrouped in version 2. Both versions of the algorithm were run multiple times, altering the number of features that were selected each time. The result of this analysis is shown below. As can be seen, for grouped tenure four features need to be selected to attain the best accuracy for the model, while only one feature needs to be selected in the model using ungrouped tenure.



Table 2 - Algorithm 2 - Decision Tree Results (Tenure)

As ‘studied\_credits’ has been identified as the independent variable, the Decision Tree model was rerun using this as the target variable, as well as the ‘gender’. As can be seen below when hyperparameter turning was applied in the form of GridSearchCV, the result was more than that achieved using other variables.



Table 3- Algorithm 2 - Decision Tree Results (studied\_credits)

## Algorithm 3 - Support Vector Machines (SVM)

Continuing with algorithms identified in previous literature, an SVM mode was the next one created. As with previous models, the initial run through was completed with no hyperparameter tuning applied. The algorithm was changed with each of the four possible kernels’ being employed. This was completed manually to confirm the best kernel and baseline results.



Table 4 - Algorithm 3 SVM

As with the other algorithms, GridSearchCV and yet another parameter grid was applied to the model, but an improvement in the accuracy of the results across any of the updated parameters used as part of the analysis was not seen. However, model output below show how each kernel performed once parameters were identified as part of the analysis.

A close-up of a graph

Description automatically generated

Figure 5- SVM analysis with hyperparameter tuning applied.

## Algorithm 4 - Random Forest

The next algorithm to be applied was the Random Forest Classifier. As with previous implementations, the model was run first without hyperparameter tuning, focusing on the number of ‘trees’ within the random forest. The optimal number of trees for the first model is 150 trees, going up to 200 trees when hyperparameter tuning in the form of GridSearchCV is applied. The result for entropy is most accurate with 200 trees selected.

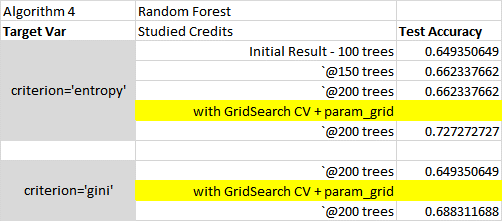


Table 5 - Algorithm 4 - Random Forest results.

The graph below ranks the important feature’s with GridSearchCv applied as hyperparameter tuning. Applying a parameter grid to aid in the search for the Random Forest with GridSearchCV identifies that the best parameters to be applied for the analysis. Figure 7 (below) identified the most important features in the model. Education level appears to be the least important feature withing the model.

A graph with a bar graph

Description automatically generated

Figure 6 - Random Forest with hyperparameter tuning applied.

## Algorithm5 - Multi-Layer Perceptron (MLP)

The final algorithm selected for analysis was the Multi-Layer Perceptron or MLP neural network. The model was run with differing neurons in the hidden layers. The more neurons added to the layers, the more varied the results - see table 6 below.



Table 6 - Algorithm 5 - MPL implementation

|  |  |  |
| --- | --- | --- |
| **Test 1 - Loss** | **Test 2 - Loss** | **Test 3 - Loss** |
| A graph with a blue line  Description automatically generated | A graph of a graph  Description automatically generated | A graph of a graph  Description automatically generated |
| **Test 1 - Accuracy** | **Test 2 - Accuracy** | **Test 3 - Accuracy** |
| A graph with a line  Description automatically generated | A graph with a line  Description automatically generated | A graph with a line  Description automatically generated |

Table 7 - MLP implementation results of loss and accuracy

The addition of hyperparameter turning on test two was complicated by the decision to use the keras library. It was necessary to create a function to allow for turning to take place between the keras and scikit-learn libraries. This involved creating the model within a function, then wrapping it in the KerasClassifier from keras.wrappers.scikit\_learn, which is a standalone module that allow keras and sci-kit learn to work in tandem (Brownlee, 2016, 2022; ‘Hyperparameter tuning using GridSearchCV and KerasClassifier’, 2020). As with other algorithms, a list of variables was created using a parameter grid with the intention of identifying the most appropriate variables that could be applied to the dataset. The output of the model is outlined in Table 8 in terms of accuracy and loss. It is clear to see that the application of hyperparameter tuning initially impacted on the accuracy of the model by the third epoch the model returned to a zero-accuracy result. In terms of the loss function, again the results are slightly better than previous models as the loss function gradually tuns to loss value rather than immediately doing so as its previous iterations.

|  |  |
| --- | --- |
| **Plot of Model Accuracy** | **Plot of Model Loss** |
| A graph with a line  Description automatically generated | A graph with a line going up  Description automatically generated |

Table 8 - MLP Results with hyperparameter tuning.

## Analysis of Selected Variables

Once all the algorithms had been tested, the models with the best results were run against the independent variables selected as part of the research paper. The results for each algorithm by the relevant variable are displayed in Table 9 below and will be discussed in detail in the following chapter.

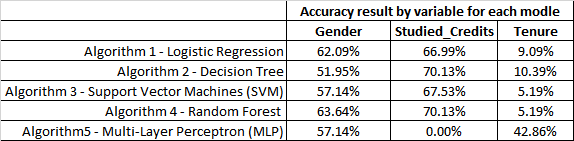


Table 9 - Accuracy Scores by Algorithm and by Target Variable

Reviewing the results of the analysis of all the algorithms it is possible to make some observations. Looking each variable on its own, for example, Gender, Algorithm 4 Random Forest provides the highest level of accuracy, whilst Algorithm 2 - Decision Trees is least accurate at 51.95%. For studied\_credits, 3 algorithms give the same level of accuracy - namely Decision Trees, Support Vector Machines and Random Forests. Algorithm5 on the other hand, the MLP network fails to give any degree of accuracy in respect of its results. In contrast, the variable for tenure has its strongest performance in MLP, whist results in all other variables are well below, with just one result (Decision Trees) achieving an accuracy result of just over 10%.

The initial results in Table 9 show that three Algorithms, 2 and 4 received the same accuracy result for studied\_credits. The author sought advice on how to test for statistical alignment or difference. The dataset was not suitable for an ANOVA analysis to be carried out - mainly in terms of the randomness of the sample (*One-way ANOVA - An introduction to when you should run this test and the test hypothesis | Laerd Statistics*, no date). A further subset of the OULAD dataset was created using the same sampling strategy as before - taking one module over one semester as the sample group. The idea being that for the test group of data to confirm if there is any clear over of under fitting of the data, or some other anomaly easily shown. The test group of data was larger than the initial module. The results of the initial run and the test results are outlined below.

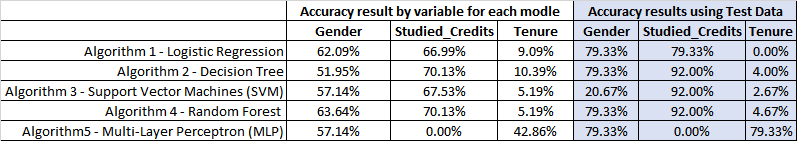


Table 10- Initial results v's Test results

As can be seen in the results for the variable gender, the test dataset shows considerable higher results, than the initial tests did, with the highest accuracy of 79.33% being reached across four of the five algorithms. The outlier in terms of results is the SVM algorithm, with a result of 20.67%. The test data result is significantly below that achieved in the initial tests of 57.14%

In respect of the variable studied\_credits, again the results using the test data are consistently the same for algorithms 2 and 4. Algorithm 3 also has the same result of 92% for studied\_credits, again a significant increase against the highest accuracy score from the initial tests completed. Again, the neural network MLP performed poorly in the tests even with using a larger dataset. An interesting point to note is that although the accuracy score was 0.00% in both tests, the loss score for the larger dataset did see an improvement - see Table 11.



Table 11 - Loss results for Algorithm 5 - MLP

The final variable analysed using the test dataset is the variable tenure. The results displayed are interesting. Algorithm 5 shows a marked improvement on the earlier test - an increase from 42.86% to 79.33% is displayed between the two different datasets. In contrast, algorithms 1, 2, 3 and 4 all see a marked decrease in the accuracy result for this variable.

The outcomes of the results will be discussed in more detail in the following chapter.

# Chapter 6: Discussion

The methodology that was outlined in Chapter 4 was applied to each of the five algorithms selected for analysis with the implementation and results being discussed Chapter 5. The results of the overall analysis cain be seen in Table 10, displaying that the most consistent algorithm was Algorithm 4 - Random Forest as it achieved the highest results for two variables - notably gender and studied credits. For the tenure variable, Algorithm 5 was the most accurate, although results were not above the 50% accuracy level on the initial test. However, it’s necessary to review the analysis based on the research questions posed in Chapter 2.

## Objective 1 - Gender Variable

As a brief reminder, Objective 1 asked if an employee’s gender Could provide a reference point for data analysis carried out as part of the succession planning process.

Using the initial dataset, and then the test dataset, Random Forest proved to be the most accurate predictor of success using gender, with a 63.64% result in the initial test, and 79.33 % when the test data was applied. What is interesting about these results is that Algorithms 4 and 5 gave the same percentage accuracy during initial analysis. The difference in results between the initial data and the test data could be because of scale, with the initial dataset being more realistic in style for the environment proposed for the completed model. As part of future work in respect of this research, the author would propose to confirm within the Jupyter notebook if the data is overfitting or underfitting and adjust accordingly. Strangely, the SVM algorithm performed adequately in the smaller dataset, but with the least amount of accuracy with the larger test dataset, despite using a parameter grid to identify the most appropriate variables for analysis.

Figure 7 - Graph of Results by Gender.

Several limitations have been identified in respect of this variable such as it not being possible to compare results gained with any previous work or research in this area, as the author was unable to locate a similar study using learning demographic data for analysis. In addition, using an openly available data set - OULAD is not fully reflective or representative of the employee population within the work environment. The timeframe is also not current in terms of validity, as the contents of the dataset was released in 2017 but relate to the time frame 2013 to 2014. Focusing specifically on demographic features such as gender / studied\_credits / tenure could unnecessarily import a bias into interpretation of results with, in this instance, a larger number of males being present over females. The dataset also does not consider the wider definition of gender that is accepted with HR circles in 2023. In addition, the author suspects that some degree of overfitting or underfitting is present in the larger test dataset as the results are the same across 3 distinct algorithm groups, and further analysis will need to be conducted to confirm if this is the case, or if there is another reason for the similar results. Finally, the author’s own inexperience in data analytics must also be considered in terms of fully understanding the results and opportunities for improvement identified as part of this research. Limitations are also present within the algorithms selected for analysis such as:

* Algorithm 1 - Linear Regression -the inclusion of too many features may affect the outcome of the model.
* Algorithm 2 - Decision Tree - does not work well with outliers, which are key to the analysis of this dataset.
* Algorithm 3 - Support Vector Machines (SVM) - works best for binary classification, so it may struggle with non-binary classification models such as that in use for this research.
* Algorithm 4 - Random Forest - this model was slow to run, due to the size of the dataset. Once the larger test dataset was applied, it took a noticeably longer time for the results to be displayed. This will need to be considered when considering future study or application of these models.
* Algorithm5 - Multi-Layer Perceptron (MLP) was noticeably the least accurate of all the models selected when completing an analysis of the gender variable.

Several opportunities for future work were identified within this research objective including the need to validate the findings against real data from the company as opposed to a freely available open dataset. As noted in the limitations above, the models will need to be tested in ‘real world’ conditions to see if they will be robust for future use with company data. Whilst this second phase of study is underway, the opportunity exists to confirm if the data is falling victim to overfitting or underfitting during analysis as outlined above, or if there is some other reason for the same results across two models in the initial dataset and four models in the larger test dataset. Further study will also be needed to understand why Algorithm 3 performed so badly on the larger test dataset.

Based on the points outlined above, the author can confirm that it is possible to accurately predict succession planning using gender as the target variable.

## Objective 2 - Studied\_Credits Variable

To recap, Objective 2 of this research paper queries if it is possible to use studied\_credits as a substitute for the number of courses completed, and to identify what (if any) machine learning algorithm could provide the most accurate measure for succession planning.

Figure 8 below shows the accuracy results with studied\_credits as the independent variable used in the selected algorithms for analysis. In terms of the initial results, algorithms 2 and 4, Decision Trees and Random Forest respectively, provide the most consistent result at 70.13%. The most inaccurate result was received by Algorithm 5 - MLP which returned a 0.00% accuracy. Table 11 outlined that although the accuracy score was still 0.00%, there was an improvement on the loss score when re-run using a larger dataset. This can be interpreted as being either an issue with the data on which the model is run, or that the model is a poor choice for analysis of this variable. As other variables within the dataset achieve a much higher accuracy result, the author can suppose that the model is not a good fit for this variable - at a minimum, further analysis will need to be completed to understand what specifically is happening in Algorithm 5. Finally, Algorithm 1 on Logistic Regression performed quite well with an accuracy result of nearly 67% in the initial test, rising to 79.33% with the larger dataset.

Figure 8 - Graph of Results by Studied\_Credits.

When implementing the second larger dataframe, the results received were the same across algorithms 2, 3 and 4, which was not the case with the first dataset. This leads the author to suspect that the data may potentially be overfitting based on the high degree of accuracy received.

In terms of limitations, those outlined in the previous section would also apply here in that real data from the company will need to be imported to test the model for accuracy to confirm if there is any potential application outside of this research. With the lack of prior research or learning in this area, it is hard to compare the results received for model effectiveness. Also, as Algorithm 4 was implemented, the time taken to complete the analysis was more noticeable in the larger test dataset which had more than 800 rows of code. This should be considered if the intention is to test the model on larger datasets and should be planned for accordingly.

In terms of future work, more analysis will need to take place on the algorithms to understand how better accuracy was achieved with one model (70.13%), and a 0.00% accuracy was recorded with the final model tested (algorithm 5). If the opportunity arises for further work in this area using real data from the from the company, it is important to consider the ethical implications of doing so and try to limit any potential bias that could potentially arise. By implementing a parameter grid to allow for hyper parameter turning to take place, there is still clearly scope for the hyperparameters to be further tuned to potentially achieve even higher levels of accuracy. As with the previous variable, it is curious that the ensemble model created with the Random Forest classifier consistently performed the best across studied\_credits and gender. The author would suggest further work in this area as an additional opportunity for future analysis.

Overall based on the analysis completed as part of this research paper, it is possible to say that Studied\_Credits is a strong variable across all the models tested. As such, it may be considered when planning for succession planning.

## Objective 3 - Tenure Variable

The final objective of this research paper, Objective 3 queries if an employee’s tenure within the company may have an impact on a succession planning. Figure 9 displays the results of the variable tenure in all the algorithms selected for analysis. Across all the five models, the variable performed poorly, with algorithm 5 being the most accurate with a result of 42.86% in initial tests. What is even more interesting, is that the results for the larger test dataset were even less accurate than the smaller original dataset, apart from Algorithm 5 where a result of 79.33% was achieved. Further analysis should be performed to confirm if the data is being overfitted or underfitted, or if another method of parameter tuning may be more suitable. The author acknowledges that this variable (tenure) is the only one that was created using synthetic random data. It would be interesting to see if, when retesting with live data, the level of accuracy reacts differently.

Figure 9 - Graph of Results by Tenure

Another point to note is that algorithm 5, achieved a higher degree of accuracy using tenure than any other variable tested as part of this research. It will be necessary to confirm if this remains the case when the model is retested with real data, and indeed will be necessary to understand why a relatively high accuracy result was achieved for this variable that was not seen with the other variables selected. Although hyperparameter tuning was carried out across all algorithms, and a parameter grid was applied to the data, more focus is clearly needed to improve performance across all algorithms selected for analysis. Perhaps a different method of tuning will need to be considered also as part of opportunities for future work.

In terms of limitations, the implications of not using real world data, especially in terms of the value of the tenure column will need to be considered. In addition, accuracy of the algorithm is not the only measure of how a model performs, and potentially it is not the most appropriate measure in this instance. This point will need to be considered in any additional future work on this topic. Other general limitations and opportunities already identified as part of objectives one and two above should also be considered here.

Overall, based on the analysis carried out to date, tenure does not appear to be a good measure to aid succession planning.

## General Observations

Several unstructured interviews were carried out with senior colleagues within the Company as primary research for this work. As outlined previously, the interviews were carried out with colleagues having a variety of experience in different geographical areas as well as external companies to minimise any potential bias being included. It was interesting to note that the approach to succession planning outlined in interviews A and B were largely the same, despite different populations and seniority of roles being considered. For context, Interviewee A carries manages succession planning for the manufacturing section of the company across a wide geographical area in Europe, Asia, Africa, and Latin America. Skills for each role are governed by a framework, and employees are upskilled accordingly. Interviewee B completes the same process, but for very senior roles within Europe only.

The process of working closely with HR is mimicked in both interviews and inputs are captured based on impressions from discussions with managers and in development conversations. There is a degree of change taking place in both regions, but in different ways. Interviewee A reports using a standardised process with inputs collected and transferred into a Dashboard for review using automated tools. Interviewee B on the other hand reports an evolving process where employees are identified and measured against several key skills for each position, their current level of experience is graded based on the time needed to upskill, and a plan to provide the missing skills put in place as part of the employees ongoing development plan.

The process outlined by both participants are both very manual in nature but does outline the journey HR is on as regards the need to provide data drive decisions. In terms of learning data, once key skills are identified, suitable pathways to upskill can be identified, learning data may be used as a method to track progress. Interviewee C rightly points out though that learning data is not the only measure of readiness in terms of success, and other measures such as mobility, visibility and ability to partner are also keen indicators that need to be considered as part of succession planning. Identifying suitable data points for such indicators could be considered an area of future work for this research paper.

Additionally, identified in the literature review Tambe et al in their article Artificial Intelligence in Human Resources Management: Challenges and a Path Forward outlined the application causal discovery to investigate if machine learning algorithms using data derived from HR (Tambe, Cappelli and Yakubovich, 2019). Causal discovery and newly developed causal algorithms may have an application in the analysis of HR data supporting succession, as it can work with small datasets where expert knowledge is important in terms of design and interpretation. The author considers application of such algorithms a potential area for future research.

In terms of limitations, as outlined previously, this is a new area of research so comparison and clarity brought about by previous research is not easily available. The author has attempted to limit bias by holding interviews with a range of participants with a variety of experiences as well as years working with the company. It is also necessary to call out the ethical considerations that must be considered in terms of privacy and impartiality in the process, which is heavily managed and influenced by individual opinion. There is however, an ongoing drive to support decision making with data within the HR function, and the succession planning process is just one area of that.

# Chapter 7: Conclusion

This research paper was undertaken to determine if it would be possible to use learning data to support the succession planning process. Three research questions were proposed to direct the research to determine if demographic variables linked with learning data could be used as the basis for analysis. Several machine learning algorithms were selected based on prior research into the area of learning analytics, specifically those studies that utilised the OULAD dataset. Prior research gave a platform on which the study could be carried out, despite the research topics not correlating exactly. The analysis that was conducted as part of this research paper shows that two out of three variables identified could potentially be used to support the succession planning process, namely the variables of Gender and Studied\_Credits. Algorithm 4, that of Random Forrest reported the highest accuracy across these two variables. There is still the potential to improve the accuracy of the algorithm by completing additional fine tuning to the algorithm using the parameter grid that is already coded within the model.

As outlined earlier in Chapter 3, there is pressure coming to bear on the HR Department to become partners with the business and learn to harness the wealth of data that is available to help make better business decisions. This was borne out in interviews carried out as part of primary research. It is hoped that this analysis is such a first step in this process. When discussing this research with colleagues in the HR field, initially there was curiosity at how the proposed model would work, especially as it had never been proposed before within the company. Once the objectives were clarified, and the relevant variables selected, the conversation about the model became more meaningful. In addition, the selection of variable was confirmed as being relevant to the succession planning process by experts, which allowed for a greater degree of confidence on the part of the author.

Having completed the analysis of the data and taking on board the feedback from experts in the field, an additional application of this algorithm was suggested. As part of the talent plan within the company, it is possible to seek ‘experiences’ outside of the employee’s main roles and responsibilities. These experiences may be temporary - for a two-to-three-month period working as part of a project group or may be for a longer period of up 12 months. These opportunities can happen with short notice, as the business need develops. Managers sometimes struggle to identify employees with the relevant skills or interest to fill resultant gap in their team when an ‘experience’ position becomes available. It has been suggested by experts interviewed as part of this research paper, that this algorithm may be an additional mechanism that would aid manager and the HR department to identify employees who based on their prior learnings, may be interested in trying a new area of work for a short period of time. Thus, giving the employee an opportunity to try a new position, and the manager a potential new talent to help develop.

All that been said, the author acknowledges that there are flaws in the implementation of the models, and these limitations have been covered already in the previous Chapter. Should the opportunity for further research be given, further fine tuning of the model and application of current live data should permit for a robust model that will aid the HR Department in the succession planning process. As the content of primary interviews confirm, the process of succession planning is changing, and a more holistic view of employee experiences is considered. However, this holistic view is measured against several clear datapoints that successful candidates must have to progress. These development items are across a wide spectrum of skills, but where an employee is lacking this skill, then they must be supported to develop it within a supportive learning environment.

# Appendix A: Plan of Work

# Appendix B: Interview Transcripts

# Interview A - Transcript

0:0:0.0 --> 0:0:0.370  
Interviewee A  
But.

0:0:-1.-410 --> 0:0:3.530  
Sinead Duffy  
A couple of bits of it more because unfortunately, John, you are my, my.

0:0:5.340 --> 0:0:7.80  
Sinead Duffy  
Test subject for this.

0:0:6.720 --> 0:0:8.170  
Interviewee A  
But happy to be.

0:0:8.620 --> 0:0:26.970  
Sinead Duffy  
Thank you. And so look, as you know, I'm doing my thesis, it's it kind of moved a little bit I think from when we last talked. So what I'm kind of interested now are where the data is taking me is using learning data as a support for succession planning okay and is it possible, is it not and?

0:0:28.230 --> 0:0:39.370  
Sinead Duffy  
You know what kind of the data from a learning management system would be able to tell us in terms of, you know, both data analytics and as a support to the HR function, OK.

0:0:39.880 --> 0:0:40.300  
Interviewee A  
Hmm.

0:0:40.120 --> 0:0:53.750  
Sinead Duffy  
Kind of. It's it's interesting, but it might help. It might not, and I'm and that's what I'm trying to understand. So basically I got around the data standardisation part of the.

0:0:55.10 --> 0:0:57.730  
Sinead Duffy  
Requirement by I got a a learning analytics.

0:0:58.370 --> 0:1:12.140  
Sinead Duffy  
Data set from the Open University, so it gives you all of the standard clicks and that it takes people to click through a course and all of that. So I'm using that as a data set. But what I've done is I've added in a tenure.

0:1:13.280 --> 0:1:30.920  
Sinead Duffy  
A quantity or a 10 year band so that we can mimic almost like what's what we have here in like in COMPANY A or in The Companywhere you've got it's an employee database and they're with us for X amount of years and it's a random number between 0 and 20.

0:1:31.910 --> 0:1:32.320  
Interviewee A  
All right.

0:1:32.780 --> 0:1:34.770  
Sinead Duffy  
OK, so that's trying to mimic.

0:1:35.590 --> 0:1:37.760  
Sinead Duffy  
Like a bit more of a head county report.

0:1:38.330 --> 0:1:38.740  
Interviewee A  
Right.

0:1:39.250 --> 0:1:47.140  
Sinead Duffy  
Okay. So I suppose what I wanted, and I'm conscious that you have the like in COMPANY A, we have the.

0:1:47.900 --> 0:1:48.410  
Interviewee A  
Involved.

0:1:49.460 --> 0:2:13.820  
Sinead Duffy  
Neval. Yeah, which is and this is the same kind of a learning management system. So it tells you what courses you you've signed up for. You've been signed up for what's mandatory, what's not. And the only difference with this data set is that it works on studied credits. So, you know, with with her, with the universities they have, you have X amount of credits to study to get a degree.

0:2:14.360 --> 0:2:14.750  
Interviewee A  
Hmm.

0:2:15.160 --> 0:2:30.320  
Sinead Duffy  
That's the that's the main difference. So I've I've paired it down to like 1 module, 1 semester and it's about 300 people. Maybe 3:50 to try and mimic what would potentially be in a manufacturing environment. OK.

0:2:30.180 --> 0:2:30.620  
Interviewee A  
Right.

0:2:31.650 --> 0:2:47.580  
Sinead Duffy  
So I've done my best to mock up the data as best I can, and I think this is a good way for it. So we'll see. We'll see what the results come out of, but I suppose in terms of the succession, can you talk me through how the PDF process works in COMPANY A?

0:2:50.0 --> 0:2:51.150  
Interviewee A  
Yeah. So.

0:2:53.150 --> 0:2:55.690  
Interviewee A  
Just gone through this with Vincent yesterday and.

0:2:58.180 --> 0:2:58.460  
Interviewee A  
Ohh.

0:3:0.340 --> 0:3:0.940  
Interviewee A  
\* \*\*\*\*.

0:3:3.260 --> 0:3:4.60  
Interviewee A  
And.

0:3:6.220 --> 0:3:17.70  
Sinead Duffy  
While you're looking to ask, I'm just going to confirm that you know, if you want to take withdraw from this process, that's no bother at any stage, I'll pull your, your interview and whatever you tell me out of the.

0:3:18.270 --> 0:3:19.320  
Interviewee A  
Ohh you're safe enough.

0:3:32.790 --> 0:3:33.200  
Interviewee A  
Ohh.

0:3:20.560 --> 0:3:33.510  
Sinead Duffy  
But I'm going to totally like I'm using anonymous data and I'm calling it out at the front that you know I'm not sharing any proprietary information or anything like that. I'll make sure legal are okay with it before we we go any further.

0:3:34.130 --> 0:3:34.860  
Interviewee A  
Family.

0:3:39.950 --> 0:3:40.390  
Sinead Duffy  
Hmm.

0:3:36.780 --> 0:3:41.880  
Interviewee A  
So that's our PDF calendar and for COMPANY A, so basically.

0:3:42.650 --> 0:3:48.450  
Interviewee A  
It kind of starts in January with kind of the plans doing their work that matters most alignment. So looking at.

0:3:49.300 --> 0:3:56.350  
Interviewee A  
Client objectives and kind of allying with employees that the talent cards are career profiles. They ask people complete doors prior to their.

0:3:57.80 --> 0:4:3.630  
Interviewee A  
Client PDFs and then by the end of March each year. Basically all the plants are asked to complete their.

0:4:8.200 --> 0:4:8.600  
Sinead Duffy  
Hmm.

0:4:4.330 --> 0:4:15.390  
Interviewee A  
Their planned PDFs, right. I'll take you through the process in a second. We kind of then roll that up to kind of east and West regional kind of apps. And we also gonna do the supply chain services.

0:4:16.200 --> 0:4:28.350  
Interviewee A  
We then try and kind of get and we haven't been too successful on this, but kind of basically particularly supply chain where you got a lot of supply chain resource and plant and a lot of supply chain resources and supply chain services is kind of have a joint.

0:4:29.710 --> 0:4:41.500  
Interviewee A  
PDF to make sure kind of their sharing talent across the entire supply chain organisation and then in May we pick up the global functions and in June we kind of roll out up to kind of a COMPANY A level.

0:4:42.230 --> 0:4:42.710  
Sinead Duffy  
Okay.

0:4:42.830 --> 0:4:51.510  
Interviewee A  
Power to processes and it's linked to the job framework. We kind of have those formal peak career conversations in June and November and we do a.

0:4:52.230 --> 0:5:0.400  
Interviewee A  
Or supposed to do and our capability PDF in August. So our PDF process really looks at kind of our current rolls.

0:5:0.860 --> 0:5:1.310  
Sinead Duffy  
Hmm.

0:5:1.160 --> 0:5:18.950  
Interviewee A  
That kind of says have you got kind of a a succession pipeline in place to kind of plug those gaps, the idea the our capability is to kind of using the arc capability model is to identify what are the rules, the skins and the structures that we need for the future but more future looking?

0:5:19.590 --> 0:5:20.40  
Sinead Duffy  
Okay.

0:5:20.440 --> 0:5:31.150  
Interviewee A  
And then in terms of rest of the PDF process is really kind of the back end of your own September through to October, November, it's a talent segmentation PDF where we basically turn where your.

0:5:37.550 --> 0:5:37.920  
Sinead Duffy  
M.

0:5:32.400 --> 0:5:40.10  
Interviewee A  
But used to be hypertension casts and Cortana, and it's now obviously just high potential. And then at the back end of the year, you'll find it at the P/E.

0:5:41.580 --> 0:5:43.850  
Interviewee A  
PDFs in terms of that differentiation process.

0:5:44.450 --> 0:5:44.940  
Sinead Duffy  
Atcha.

0:5:44.520 --> 0:5:47.550  
Interviewee A  
What actually happens in the?

0:5:50.700 --> 0:5:51.270  
Interviewee A  
Compare.

0:5:56.680 --> 0:6:9.770  
Interviewee A  
So the objective is that basically you go to the process and you can meet your kind of current future tenant needs in terms of specific outcomes that you identify 2 to 3 candidates for each position. So you got problem with candidates.

0:6:10.430 --> 0:6:10.850  
Sinead Duffy  
Umm.

0:6:10.640 --> 0:6:18.650  
Interviewee A  
They need you identify want development action. So we're trying to differentiate between development for your current role and development for succession plan role.

0:6:19.930 --> 0:6:20.340  
Sinead Duffy  
Okay.

0:6:22.660 --> 0:6:23.50  
Sinead Duffy  
Hmm.

0:6:49.20 --> 0:6:49.560  
Sinead Duffy  
OK.

0:6:19.680 --> 0:6:49.570  
Interviewee A  
And because they're kind of two different animals and and it's really trying to get the focus on what's the one development action that were most accelerate this person's readiness. So if there are two to five, alright, what's the gap? What do they missing in terms of the scaler and experience? And that stopped them being zero to 2 and have an identified what that is. But the one thing you can do of next 12 months that one most accelerate their readiness must have one big ticket item you can focus on. We asked it then.

0:6:49.640 --> 0:6:50.710  
Interviewee A  
Because I am.

0:6:51.950 --> 0:6:57.330  
Interviewee A  
Retention risks for the pipe, attention of cast and then we also look at there's anyone need a move.

0:6:59.350 --> 0:7:2.980  
Interviewee A  
And and then obviously all that data gets rolled up into our PDF.

0:7:4.760 --> 0:7:6.180  
Interviewee A  
Dashboard reporting you seem to.

0:7:5.190 --> 0:7:6.190  
Sinead Duffy  
This dashboard.

0:7:7.480 --> 0:7:11.630  
Interviewee A  
In terms of what the guys covering the agenda is?

0:7:12.400 --> 0:7:28.560  
Interviewee A  
Just gonna fly betterments to it. First one is kind of just looking at the data you have so you know what's your percent ready nose, percent ready now female percent. The rules. They got two or more candidates. And so just really kind of doing assessment in terms of how we doing.

0:7:44.470 --> 0:7:44.990  
Sinead Duffy  
I get you.

0:7:29.640 --> 0:7:48.650  
Interviewee A  
In terms of those metrics versus previous years and based on where do we need to focus in terms of succession plan detail itself was literally in the plans going through all those roles and being sure we got kind of robust succession plans in place. And then once you've done a succession plan, there's kind of that development action planning piece.

0:7:49.970 --> 0:7:50.320  
Sinead Duffy  
Hmm.

0:7:49.370 --> 0:7:53.400  
Interviewee A  
And where you identify, right, so where we've got gaps.

0:7:54.60 --> 0:8:2.150  
Interviewee A  
What a development action plan put in place and then you kind of track doors and obviously then the retention assessments and the people in place so often.

0:8:3.370 --> 0:8:5.100  
Interviewee A  
Development perspective.

0:8:7.50 --> 0:8:20.840  
Interviewee A  
Basically, they're complete this template and so I kind of listen your name project, you know current. So I will location, etcetera, etcetera, what the successor role is. And then in the scription of.

0:8:21.850 --> 0:8:26.140  
Interviewee A  
The development action to accelerate their readiness and then we kind of have.

0:8:26.790 --> 0:8:40.420  
Interviewee A  
So later knowing different kind of development categories like you know is an STASTE, is it on the job? Is it kind of classroom training? Is the leadership development and specify who's responsible and then?

0:8:41.330 --> 0:8:44.740  
Interviewee A  
Specify whether it's kind of not started in progress or complete.

0:8:47.190 --> 0:8:50.0  
Interviewee A  
But that's like it 32nd overview of.

0:8:51.250 --> 0:8:51.780  
Sinead Duffy  
Elvis.

0:8:51.320 --> 0:8:52.490  
Interviewee A  
The PDF process.

0:8:52.970 --> 0:9:12.620  
Sinead Duffy  
Okay and I know we've got a 7020 ten kind of approach to development. Do does the does any of the learnings that are stored and evolve or the learning management system? Are they kind of taken out and looked at it any part of this project, this processor is it?

0:9:17.900 --> 0:9:18.270  
Sinead Duffy  
Hmm.

0:9:14.280 --> 0:9:23.840  
Interviewee A  
The Stephen involved probably wouldn't be hugely relevant to this, because certainly at this point, because more so stopped and bomb relates to job framework roles.

0:9:24.440 --> 0:9:26.460  
Sinead Duffy  
Yeah. OK. So grades nine and below.

0:9:25.970 --> 0:9:34.300  
Interviewee A  
And so it's. Yeah and no, that's beginning to change. So for example, we're doing some work for the SO owns in terms of identifying.

0:9:35.350 --> 0:9:38.0  
Interviewee A  
Like 7 core technical skills and three.

0:9:38.710 --> 0:9:45.680  
Interviewee A  
Leadership skills that that we're gonna build into their BSM. So in the future that's gonna draw health drive to PDF process.

0:9:46.530 --> 0:9:46.810  
Sinead Duffy  
OK.

0:9:46.520 --> 0:9:53.190  
Interviewee A  
And you'll be looking at specific skills, but generally just stopping involved the minute is is planned this kind of.

0:9:54.110 --> 0:9:57.500  
Interviewee A  
Plant based, so I don't think you're going to use a lot of succession planning.

0:9:58.440 --> 0:9:59.190  
Interviewee A  
The.

0:9:59.940 --> 0:10:6.770  
Interviewee A  
But what you do see in the the stuff we specifically put in here is more around succession plan, so.

0:10:8.260 --> 0:10:14.180  
Interviewee A  
You've got the information involved, you've got information in work day, in terms of the development items.

0:10:14.820 --> 0:10:15.210  
Sinead Duffy  
Hmm.

0:10:15.170 --> 0:10:19.20  
Interviewee A  
Which you could link him, but generally a lot of those development items.

0:10:20.390 --> 0:10:21.980  
Interviewee A  
Most were most people for the win.

0:10:22.740 --> 0:10:22.990  
Sinead Duffy  
Yeah.

0:10:22.630 --> 0:10:25.810  
Interviewee A  
Feed the ones that do is primarily around their current role.

0:10:28.100 --> 0:10:28.430  
Sinead Duffy  
Yeah.

0:10:26.740 --> 0:10:29.70  
Interviewee A  
Current development and.

0:10:29.930 --> 0:10:32.110  
Interviewee A  
But where we pick up the?

0:10:33.170 --> 0:10:36.270  
Interviewee A  
Succession planning development items is actually in.

0:10:36.980 --> 0:10:37.990  
Interviewee A  
The dashboard here.

0:10:38.740 --> 0:10:39.90  
Sinead Duffy  
Okay.

0:10:39.90 --> 0:10:41.760  
Interviewee A  
Or sorry. So when you go into the dashboard.

0:10:46.850 --> 0:10:47.370  
Interviewee A  
This one.

0:11:0.10 --> 0:11:1.0  
Interviewee A  
That's the cards.

0:11:2.630 --> 0:11:5.250  
Interviewee A  
That's just for general manager on their website.

0:11:6.220 --> 0:11:6.620  
Sinead Duffy  
Hmm.

0:11:12.850 --> 0:11:18.860  
Interviewee A  
Okay that's wear because it was XXXX rolls there yesterday. So I need to figure out was going on there.

0:11:19.380 --> 0:11:19.850  
Sinead Duffy  
OK.

0:11:21.50 --> 0:11:22.780  
Sinead Duffy  
OK so.

0:11:22.400 --> 0:11:25.230  
Interviewee A  
Behind us. Sorry. Some reason it's 2022.

0:11:29.850 --> 0:11:31.340  
Interviewee A  
I know. I'm sorry I'm in the wrong.

0:11:33.630 --> 0:11:34.720  
Interviewee A  
Family were shot.

0:11:39.870 --> 0:11:47.180  
Sinead Duffy  
Like I love the dashboard. I don't know if he's anything like that here in EU. I'll find out next week when I talk to Helen.

0:11:49.10 --> 0:11:53.980  
Interviewee A  
I don't think so cause it kind of for time to couple of people about I haven't had anyone near you but.

0:11:54.730 --> 0:11:57.30  
Sinead Duffy  
Hmm, it was fantastic. Dashboard.

0:11:57.710 --> 0:12:1.800  
Interviewee A  
It is. And to be fair, kind of it's, you know, it doesn't take any grade.

0:12:2.440 --> 0:12:5.190  
Interviewee A  
Managing because all the plans basically go in.

0:12:5.890 --> 0:12:11.730  
Interviewee A  
And update the SharePoint so each plant has its own folder and it goes in and puts the data into.

0:12:12.450 --> 0:12:16.50  
Interviewee A  
3 Excel files and then basically everyday 12:00.

0:12:17.530 --> 0:12:18.70  
Sinead Duffy  
Think so?

0:12:16.890 --> 0:12:18.360  
Interviewee A  
It pulls the data across.

0:12:19.600 --> 0:12:25.870  
Sinead Duffy  
Perfect. And I can see you've got a link there for the development plan which is coming from work day, right, so.

0:12:25.690 --> 0:12:27.220  
Interviewee A  
No, it's not.

0:12:26.910 --> 0:12:27.380  
Sinead Duffy  
Okay.

0:12:28.200 --> 0:12:28.590  
Sinead Duffy  
Okay.

0:12:27.960 --> 0:12:34.480  
Interviewee A  
It's that's actually coming from the Excel files that showed you earlier because to be honest, it.

0:12:33.150 --> 0:12:35.40  
Sinead Duffy  
OK, here's the development calling.

0:12:35.990 --> 0:12:36.450  
Interviewee A  
It's.

0:12:37.100 --> 0:12:45.770  
Interviewee A  
What's in work? Work is probably a lot of crap, like from a they just, they'll just be reams and reams of stuff and it's not probably.

0:12:47.560 --> 0:12:47.990  
Sinead Duffy  
Helpful.

0:12:47.470 --> 0:12:52.890  
Interviewee A  
Hugely relevant and so where it's coming from is here.

0:12:56.600 --> 0:12:57.70  
Sinead Duffy  
OK.

0:12:58.100 --> 0:13:17.170  
Sinead Duffy  
And it's this. What is there is a specific level of rules that you look at for this. Is it just like I know for manufacturing, it's the DSM is is is key but for when you're doing the PDF or you're doing it on like the specialist roles of manufacturing, are you doing it on the support roles?

0:13:17.270 --> 0:13:19.990  
Sinead Duffy  
Um, so like the lab.

0:13:22.20 --> 0:13:22.300  
Interviewee A  
Would.

0:13:21.960 --> 0:13:24.620  
Sinead Duffy  
Finance, engineering, that kind of stuff or.

0:13:25.570 --> 0:13:27.920  
Interviewee A  
No, we're typically do it on.

0:13:28.650 --> 0:13:31.990  
Interviewee A  
Is the minimum that have to do the planned leadership team orders.

0:13:32.530 --> 0:13:33.80  
Sinead Duffy  
OK.

0:13:33.310 --> 0:13:34.240  
Interviewee A  
And.

0:13:35.220 --> 0:13:39.700  
Interviewee A  
After that, we're kind of saying, well, we want to see kind of direct reports of.

0:13:40.890 --> 0:13:41.870  
Interviewee A  
The leadership team.

0:13:42.320 --> 0:13:42.760  
Sinead Duffy  
Hmm.

0:13:43.230 --> 0:13:44.360  
Interviewee A  
And people managers.

0:13:45.100 --> 0:13:45.820  
Sinead Duffy  
I get you. Yeah.

0:13:51.790 --> 0:13:52.190  
Sinead Duffy  
OK.

0:13:46.70 --> 0:13:54.340  
Interviewee A  
And so that's why we asked, we don't know, we don't always get it, feel you'll get as a minimum you get the leadership team and but there's a good few roles now in this.

0:13:56.60 --> 0:13:57.960  
Interviewee A  
Well, in our version and.

0:13:58.230 --> 0:14:1.60  
Sinead Duffy  
Yeah, that's the biggest, isn't it? So.

0:14:1.280 --> 0:14:5.350  
Interviewee A  
Yeah. In fairness, they kind of do a pretty good job in terms of populating it and.

0:14:6.30 --> 0:14:7.910  
Interviewee A  
So basically this is the Excel file.

0:14:9.270 --> 0:14:11.200  
Interviewee A  
And the guy's updated.

0:14:12.300 --> 0:14:17.680  
Interviewee A  
The template here in terms of each of the rules to the successes there and then it kind of populates this.

0:14:18.850 --> 0:14:20.50  
Interviewee A  
Big spiders that are front.

0:14:21.380 --> 0:14:26.50  
Interviewee A  
Basically then the the dashboard pulls this data 12:00 everyday.

0:14:26.750 --> 0:14:28.480  
Sinead Duffy  
Bring it and uploaded it into the.

0:14:36.260 --> 0:14:36.770  
Sinead Duffy  
Yeah.

0:14:29.760 --> 0:14:39.80  
Interviewee A  
Into this thing and and look, my this thing does for us is just kind of makes it easy to to get the metrics that's like otherwise you're gonna spend an hour or so.

0:14:39.810 --> 0:14:42.60  
Interviewee A  
This stuff here is coming from.

0:14:43.400 --> 0:14:49.60  
Interviewee A  
Family succession development. If you're gonna change actually work day would be much better if you can actually.

0:14:50.420 --> 0:14:53.90  
Interviewee A  
As development items, which has Lords are.

0:14:57.950 --> 0:14:58.340  
Sinead Duffy  
Umm.

0:14:53.800 --> 0:14:59.290  
Interviewee A  
But you really want development items that relate to succession planning rather than your your current law.

0:15:0.410 --> 0:15:4.700  
Sinead Duffy  
I got you. I got you. So do you think cause I know you've been.

0:15:5.800 --> 0:15:28.420  
Sinead Duffy  
Leading the charge on this, I mean and the the dashboard is fantastic. It's great like it's simple and easy to fill in the spreadsheets and and it's quick to do so. But my questions on is do you think that and you kind of had said earlier that you're starting to evolve and evolve to bring in a kind of specialist stuff. Do you think that would be helpful in the future?

0:15:29.270 --> 0:15:32.680  
Sinead Duffy  
Like to have to link in learning.

0:15:34.190 --> 0:15:34.800  
Sinead Duffy  
Um.

0:15:35.750 --> 0:15:39.650  
Sinead Duffy  
Data and maybe start small, but specialist roles and I suppose.

0:15:43.160 --> 0:15:43.610  
Interviewee A  
Hmm.

0:15:40.500 --> 0:15:44.570  
Sinead Duffy  
If I look at my own particular case where I work in HR.

0:15:45.590 --> 0:15:47.140  
Sinead Duffy  
Doesn't really have a strong.

0:15:50.470 --> 0:15:50.890  
Interviewee A  
No.

0:15:47.930 --> 0:15:54.460  
Sinead Duffy  
Desa needs phone to a better word than now, but in the future I think it will.

0:15:55.90 --> 0:15:57.710  
Sinead Duffy  
But a lot of the learnings I would have done would have been.

0:15:58.470 --> 0:16:5.70  
Sinead Duffy  
Well, they've been HR related, but they would have also been data relators, so I would be, I suppose, turned as an outlier.

0:16:6.120 --> 0:16:8.430  
Sinead Duffy  
In the stuff that I'm looking at on.

0:16:10.330 --> 0:16:17.690  
Sinead Duffy  
Like The Companyconnecting all about? You think like looking at outliers would have a would have a benefit to kind of say.

0:16:18.550 --> 0:16:28.230  
Sinead Duffy  
To manage as well, have you when when they're doing this to kind of say look well, what about say she needs she has this interest obviously cause she's doing all of these courses.

0:16:32.300 --> 0:16:33.670  
Interviewee A  
Yeah, and.

0:16:34.980 --> 0:16:39.350  
Sinead Duffy  
And it's specifically outliers cause I mean if you look at I I know how.

0:16:40.90 --> 0:16:51.800  
Sinead Duffy  
You know the LMS works, everyone has mandatory training to do so you've got your code of business conduct, you've got your health and safety stuff. Everyone has to do that and then the mandatory training changes per role, so.

0:16:52.540 --> 0:16:58.300  
Sinead Duffy  
It's what else's people look at us, and I suppose it's the outliers in systems like that that.

0:16:59.160 --> 0:17:7.780  
Sinead Duffy  
I think where the value is, it's not there like what everyone is doing. Everyone's gonna have to do the mandatory stuff. It's the additional stuff that is kind of.

0:17:8.640 --> 0:17:20.800  
Sinead Duffy  
Up to you to decide, because it's, you know, with the learning journeys, you're empowered to do whatever you want in terms of the online available courses. It doesn't have to necessarily be job related.

0:17:21.900 --> 0:17:22.330  
Sinead Duffy  
Now.

0:17:21.870 --> 0:17:26.420  
Interviewee A  
So what you're saying is one this if we identify another's certain.

0:17:27.570 --> 0:17:28.560  
Interviewee A  
Courses.

0:17:29.650 --> 0:17:31.940  
Interviewee A  
That relate to certain competencies.

0:17:32.590 --> 0:17:36.220  
Interviewee A  
For specific roles, if people aren't in those roles.

0:17:36.940 --> 0:17:38.440  
Interviewee A  
And looking at that stop.

0:17:40.170 --> 0:17:45.760  
Interviewee A  
While there is trying to pull them into the P, the PDF, the talent pool for that particular role at the.

0:17:46.120 --> 0:18:9.610  
Sinead Duffy  
Not holding in, but at least have them as a have them as a an outlier in terms of like. I'm looking for someone with data analytics skills that has maybe done this course and at least should be able to pull up a list of of people who've who have completed the course but who may not necessarily already be working in that area. If you get me.

0:18:9.970 --> 0:18:10.400  
Interviewee A  
Yeah.

0:18:13.830 --> 0:18:16.120  
Interviewee A  
You could and it's.

0:18:17.730 --> 0:18:21.790  
Interviewee A  
I think it might be more relevant if you're kind of looking for an SDA or you're kind of.

0:18:23.220 --> 0:18:26.310  
Interviewee A  
Because if someone's not working in the area and.

0:18:27.120 --> 0:18:28.30  
Interviewee A  
You are gonna jump.

0:18:28.990 --> 0:18:30.150  
Interviewee A  
Yeah, I.

0:18:26.800 --> 0:18:30.610  
Sinead Duffy  
Yeah. You're not gonna put them in a yeah. You're not gonna take the risk on a permanent role.

0:18:31.190 --> 0:18:39.850  
Interviewee A  
And I think the other thing is because they're interested in the, the court or something, they're necessarily interested in the roles that they, you know, when you actually look at the.

0:18:40.650 --> 0:18:42.380  
Interviewee A  
Data. Maybe you need to look at.

0:18:43.820 --> 0:18:45.550  
Interviewee A  
Want their career aspirations there?

0:18:46.270 --> 0:18:46.610  
Sinead Duffy  
Yeah.

0:18:47.290 --> 0:18:48.980  
Interviewee A  
And look at the.

0:18:49.840 --> 0:18:50.790  
Interviewee A  
The learning data.

0:18:51.510 --> 0:18:56.580  
Interviewee A  
And see what it termed here. And I think it would be really useful. I think they kind of do it for.

0:18:58.20 --> 0:19:9.260  
Interviewee A  
You know, I think you can run a report somewhere in work day that says, but show me everyone. That's gas. SAP are done, of course. And a PRSP listed as a skill.

0:19:9.750 --> 0:19:10.90  
Sinead Duffy  
Yeah.

0:19:11.690 --> 0:19:15.220  
Sinead Duffy  
Yeah. And I know you can do that, but I suppose like I.

0:19:23.270 --> 0:19:23.600  
Interviewee A  
Yeah.

0:19:16.480 --> 0:19:36.610  
Sinead Duffy  
And I and I can only talk about me cause I know I meant I'm an odd ball or an outlier in that kind of a setting where I may not have. I may not have, like I might have power BI on my my list of skills, but I may not have the machine learning and that kind of stuff that I'm doing now.

0:19:37.150 --> 0:19:37.520  
Interviewee A  
Ohh.

0:19:47.60 --> 0:19:47.480  
Interviewee A  
Yeah.

0:19:37.380 --> 0:19:55.20  
Sinead Duffy  
Listed in my development items, but it might and it probably will pop up in terms of the courses that I've done cause I know I've done courses on statistics that I haven't put in my development items, so I suppose that's I can only talk about me because I think I am an interesting case in terms of developments.

0:19:54.370 --> 0:19:55.250  
Interviewee A  
Yeah, yeah.

0:19:57.10 --> 0:19:57.600  
Interviewee A  
I support.

0:19:56.870 --> 0:20:0.550  
Sinead Duffy  
And I suppose I'm not alone. That's my my thought processes.

0:20:2.600 --> 0:20:10.120  
Interviewee A  
So let's say kind of you were right now tall Shaheen or something. You're looking for someone in Daleks and he went through and pulled.

0:20:11.680 --> 0:20:16.870  
Interviewee A  
Something kinda says. Show me everyone who's not in the analytics who's actually doing courses and data analytics.

0:20:17.580 --> 0:20:17.960  
Sinead Duffy  
Yeah.

0:20:19.100 --> 0:20:20.10  
Interviewee A  
And then.

0:20:22.160 --> 0:20:28.760  
Interviewee A  
Yeah, I suppose if you're kind of looking for someone and you're looking for someone, got an interest in us, what you do then kind of proactively reach out to them and say love.

0:20:30.290 --> 0:20:34.600  
Interviewee A  
I know you've got an interest in it. Gonna kind of explore it further. Type thing is it?

0:20:40.90 --> 0:20:40.420  
Interviewee A  
Yeah.

0:20:35.170 --> 0:20:50.690  
Sinead Duffy  
Yeah, a little bit because I know sometimes we struggle for UMST&STE people. That's it's a good point like succession is is a lot more. And I suppose I went with succession. So I think it's a lot more structures and but my question, yeah.

0:20:51.750 --> 0:21:0.630  
Sinead Duffy  
You know that that's the point that I hadn't thought about is, you know, it might be more helpful in terms of STA experience where we struggled to get people.

0:21:1.870 --> 0:21:9.660  
Interviewee A  
Yeah. No, I'm dealing, say, generally the plan's can put in whatever role they want for the succession.

0:21:10.410 --> 0:21:10.750  
Sinead Duffy  
Yeah.

0:21:11.30 --> 0:21:15.900  
Interviewee A  
Database promise kind of management roles that we we tend to.

0:21:15.310 --> 0:21:17.520  
Sinead Duffy  
Yeah. Well, well, that's.

0:21:16.760 --> 0:21:19.300  
Interviewee A  
To look at more solid and kind of you know.

0:21:21.90 --> 0:21:27.380  
Interviewee A  
Job framework rules. We typically don't look at doors as perhaps succession planning process. I don't anyway to the the plans might.

0:21:28.760 --> 0:21:30.350  
Sinead Duffy  
Yeah. And I think that's.

0:21:31.800 --> 0:21:35.160  
Sinead Duffy  
That's fine, because you know those types of roles are.

0:21:44.220 --> 0:21:44.530  
Interviewee A  
Move.

0:21:36.300 --> 0:21:45.30  
Sinead Duffy  
Um, you know, there's very clear training path and career path and people might swap between different like areas, but they don't generally.

0:21:46.510 --> 0:21:54.410  
Sinead Duffy  
Go off and if you do go up, they those kind of those levels you get on the management stage cause the it's a very flat organisation.

0:21:55.490 --> 0:21:56.200  
Interviewee A  
Yeah.

0:21:56.630 --> 0:22:4.810  
Sinead Duffy  
Like if we had supervisors or line supervisors, I think they probably might be a little bit more included, but we don't. We just have operations a BCD.

0:22:5.950 --> 0:22:6.220  
Sinead Duffy  
Right.

0:22:8.520 --> 0:22:11.390  
Interviewee A  
We have operated a BCD aim involved. We're just.

0:22:12.40 --> 0:22:13.970  
Sinead Duffy  
No, no, no. Just in general in the crew.

0:22:14.640 --> 0:22:15.620  
Sinead Duffy  
Framework structure.

0:22:16.20 --> 0:22:16.890  
Interviewee A  
Yeah, yeah.

0:22:18.510 --> 0:22:21.580  
Interviewee A  
We do, and that's probably gonna go. I'm just gonna change a bit with.

0:22:22.810 --> 0:22:26.140  
Interviewee A  
The new things coming in January would have bothered either in there as well.

0:22:26.610 --> 0:22:31.130  
Sinead Duffy  
Yeah. Anyway, look, I'm conscious that it's half 12 and you have a hard stop.

0:22:31.80 --> 0:22:31.470  
Interviewee A  
Yep.

0:22:34.610 --> 0:22:34.920  
Interviewee A  
Ohh.

0:22:31.760 --> 0:22:37.290  
Sinead Duffy  
So I just wanted to say thank you very much and anything that you say to me would be confidential, I'll be.

0:22:38.10 --> 0:22:38.970  
Sinead Duffy  
Reviewing all of this.

0:22:38.70 --> 0:22:40.430  
Interviewee A  
I think you're safe enough and nothing I said anything opening.

0:22:41.280 --> 0:22:43.300  
Interviewee A  
Insightful our sensitive so.

0:22:43.690 --> 0:23:0.410  
Sinead Duffy  
**No, but you never know. Um, but look at. I might reach back out in a couple of weeks just to kind of maybe, maybe follow up on a couple of questions that I might that might come up as I I start delving a little bit deeper into this, but I really like your STAST kind of thing. I think that might be.**

0:23:1.230 --> 0:23:3.780  
Sinead Duffy  
A different Ave that I hadn't thought of, so thank you.

0:23:7.110 --> 0:23:7.440  
Sinead Duffy  
Come on.

0:23:4.220 --> 0:23:7.800  
Interviewee A  
Okay no problem, I'm sure. If you need anything, just let me know and I'll see you on Monday.

0:23:9.700 --> 0:23:10.20  
Interviewee A  
Ohh.

0:23:13.570 --> 0:23:14.20  
Interviewee A  
I.

0:23:8.540 --> 0:23:14.150  
Sinead Duffy

# Interview B - Transcript

0:0:0.0 --> 0:0:2.90  
Sinead Duffy  
Ohh it's going to take any just cause I'm not at work.

0:0:2.590 --> 0:0:2.970  
Interviewee B  
That.

0:0:3.340 --> 0:0:24.170  
Sinead Duffy  
Okay. Perfect, right. So thank you. First off, to Evie for for agreeing to meet with me. So basically, let me let me kind of roll back and let me give you a bit of a background. So I'm doing my my thesis in data analytics for Masters and science and basically what I'm kind of I'm doing now is reaching out to Heitor professionals and colleagues.

0:0:35.950 --> 0:0:36.320  
Interviewee B  
Hmm.

0:0:24.950 --> 0:0:55.110  
Sinead Duffy  
To understand their succession planning process. So my thesis is based on succession planning and can you use learning data to add value to succession planning process. Because I'm looking at me as an outlier where as you know as the in learning learning is either a signed by the manager assigned by the company, it can be developmental, but then there's another cohort of people myself who are doing stuff outside of their job.

0:0:55.800 --> 0:1:11.470  
Sinead Duffy  
And it's those outliers that I think might be of interest in terms of succession planning because they don't fall into any neat category and it's kind of how can we use the that data and can we use that data to enhance the succession planning process. Does that make sense?

0:1:11.700 --> 0:1:13.880  
Interviewee B  
It does make sense, yeah.

0:1:13.740 --> 0:1:30.350  
Sinead Duffy  
OK, so that's the premise of it. Now I have not used company data. I went out into the Open University and used an anonymized public data set, and I've added in a column called tenure. But I'm looking at it to see.

0:1:30.700 --> 0:1:42.350  
Sinead Duffy  
Um can if will you get a degree of accuracy if you have like demographic information on it. So like male and female data, um, you know the highest point of education.

0:1:43.610 --> 0:2:11.590  
Sinead Duffy  
What else? There's tenure in it as well, and that's an that's actually my van today in and it's just a random number. And then the number of credits that people are studying for. So the idea behind that column is, is that, you know the the lowest amount is somebody who's just doing the mandatory training. They're not doing anything else. And then the higher level would be the ones who are developing themselves outside them, like they're taking control of their own learning journey. And that's that. That's exposed the scale. Does that make sense?

0:2:11.970 --> 0:2:12.370  
Interviewee B  
Yeah.

0:2:13.200 --> 0:2:13.600  
Sinead Duffy  
Okay.

0:2:13.630 --> 0:2:13.950  
Interviewee B  
Ohh.

0:2:15.90 --> 0:2:15.460  
Sinead Duffy  
And.

0:2:23.180 --> 0:2:23.670  
Sinead Duffy  
Perfect.

0:2:35.270 --> 0:2:35.850  
Sinead Duffy  
Hmm.

0:2:14.910 --> 0:2:44.840  
Interviewee B  
Shall I tell you how it's done here or because it's probably a little bit different from what you're looking at in terms of data because and Helen will be able to talk to you more broadly cause she's got a lot more experience than I have. But in The Company succession planning is only done for like the global function and leads the occult and the Ault minus one roles, right? So that's the and the rationale behind that of my understanding of it is.

0:2:44.910 --> 0:3:15.70  
Interviewee B  
Because you need to be able to do it in a way the you know is feasible for people's career paths and everything, because otherwise you know and basically you have the succession plans and then jobs at job grade 14 plus are done by slating. So the succession plans, because they're done in three groupings. So it's like ready zero to two, which is generally they're ready to move into that role as the next role ready to to five which is where they would have like maybe one to two experiences and ready 5 plus which would.

0:3:15.200 --> 0:3:23.120  
Interviewee B  
Eight to three experiences, so the data it's a bit different because it's you're talking about relatively senior roles.

0:3:23.830 --> 0:3:24.230  
Sinead Duffy  
Yeah.

0:3:24.820 --> 0:3:48.490  
Interviewee B  
And the way, and there's allsorts of reports that are kept in work days, we're basically manage it through global pools where there's several rows like so take it from for example AVP of franchise one for the big LATAM Europe, North America, that would have a pool of people or a plan which is an individual role that has succession for it. So.

0:3:52.700 --> 0:3:53.330  
Sinead Duffy  
Okay.

0:4:3.610 --> 0:4:4.350  
Sinead Duffy  
Perfect shift.

0:3:49.810 --> 0:4:5.550  
Interviewee B  
Learning credits and things don't really come into it because of the level that you're talking about, but I can show you if you want how it's now being done because it's a completely new process and you'll probably see that there are data points. It's just the one without names, right?

0:4:8.60 --> 0:4:9.350  
Interviewee B  
So basically.

0:4:12.220 --> 0:4:43.80  
Interviewee B  
What's happened is there's, and this is very small, but this is a global piece where actually we're creating succession profiles. So currently they're being created by the global network community. So the global functions and also in Europe were sort of going ahead and they are basically being created for sort of the Lt roles that are global functional level. And then in Europe, we're creating them right now for the orult roles. And then eventually we want to do the Oulton.

0:4:43.190 --> 0:5:1.20  
Interviewee B  
Nine is 1 rolls, so that will be created for a either a pool or a plan basically. So for example, for a VP of Strategy you would have two different ones. You'd have VP of strategy one, which is basically for the big O US and then VP St Too which is the smaller OUS.

0:5:13.90 --> 0:5:13.640  
Sinead Duffy  
Okay.

0:5:1.950 --> 0:5:33.610  
Interviewee B  
And then the idea is that these having these um succession profiles, you're able to be much more scientific about succession. So what can happen is that people can say ohh I think so and so is ready or I think this person should be here. But this is actually saying and let's forget succession health at the moment. But this is actually saying these are the functional skills and there's too many here by the way that everybody in a function needs to have. These are the targeted skills that that specific role.

0:5:33.700 --> 0:5:49.80  
Interviewee B  
Needs to have. These are the experiences that people need to have had, and these are sort of the leadership and growth behaviours that they need to have. So that basically means that if somebody's in in a ready zero to two, as in, they could move into that in the next role.

0:6:4.770 --> 0:6:5.150  
Sinead Duffy  
Hmm.

0:5:50.100 --> 0:6:5.530  
Interviewee B  
They need to have multi year experience leading strategy team. They need to have multi year year experience. They need to have led and developed a large diverse team. So these will all depend on the specific role and this is sort of also work in progress because this is one that I pulled together. But.

0:6:6.970 --> 0:6:40.440  
Interviewee B  
And it also means that they, for the targeted skills, they need to have negotiations and scared but share value creation. So this will depend on the specific role and these are a bit working progress, but I suppose where I'm saying to you this becomes more data-driven is that you can actually see that this is a much more informed discussion. So for example if somebody's in ready to to five which means they'll have another one to two experiences those experiences then can be targeted against the experiences in this list they're missing and also help them build the skills that they need to build and also the leadership piece.

0:6:40.970 --> 0:6:41.720  
Sinead Duffy  
OK.

0:6:40.890 --> 0:7:11.230  
Interviewee B  
So what we're saying is, because it already zero to two, there's only a certain amount that you can do because they're going to be theoretically moving in at the next role, but are ready to to five. And we're asking them to pull together what we call acceleration development plans. Now this is only part of it because global level, they're also asking them to list the skills that they do or don't have. And then the development options to to promote these skills. But we're saying okay. So what experiences they missing? And quite often, it will be an issue with visibility as well.

0:7:11.450 --> 0:7:36.360  
Interviewee B  
So it could be and I know at the global functional discussions they'll sometimes bring up names and they don't know who they are, so they'll be visibility actions. It could be with the bottler, it could be with global, it could be within the European U itself. And then they'll also have against these, they'll have a view that actually they're missing these skills, they need to build their negotiation skill, and that's what, therefore, what's the development plan?

0:7:37.360 --> 0:7:40.0  
Interviewee B  
And then the succession health basically.

0:7:40.810 --> 0:7:41.930  
Interviewee B  
Within Europe.

0:7:42.190 --> 0:7:42.670  
Interviewee B  
Um.

0:7:45.410 --> 0:7:47.720  
Interviewee B  
Let me just share something else.

0:7:50.90 --> 0:7:51.290  
Sinead Duffy  
Okay so.

0:7:51.990 --> 0:7:53.600  
Sinead Duffy  
That's really interesting, isn't it?

0:7:55.90 --> 0:7:59.30  
Interviewee B  
Yeah. So if I show you then so this is all sort of new.

0:7:59.600 --> 0:8:0.0  
Sinead Duffy  
Hmm.

0:8:0.90 --> 0:8:4.60  
Interviewee B  
I'm just gonna share my stuff. I share my screen quickly.

0:8:5.50 --> 0:8:6.310  
Interviewee B  
Just going to go back.

0:8:8.10 --> 0:8:9.530  
Interviewee B  
To beginning here.

0:8:12.90 --> 0:8:25.280  
Interviewee B  
So this is basically saying I can send, you know this succession health piece. So basically this is saying for a particular pool or or particular pool or particular sort of individual role.

0:8:45.80 --> 0:8:45.640  
Sinead Duffy  
Yeah.

0:8:26.650 --> 0:8:49.540  
Interviewee B  
What are the skills, experiences and expected leadership behaviours? And then it's also saying how what does our succession help look health look like. So for example on the one that I showed you, if you don't have successors in each of the categories and the succession health isn't looking health but also that then comes into our gender balance. So obviously we're looking at 5050 geography, 14 plus.

0:8:57.210 --> 0:8:57.740  
Sinead Duffy  
Yeah.

0:8:50.190 --> 0:9:7.660  
Interviewee B  
And so that means that at least 50% of the the candidates need to be female, ideally, right? So the principles that we have in Europe and this is this is just the principles that we have in Europe for European successes we're saying for I know you Lt role.

0:9:8.470 --> 0:9:17.60  
Interviewee B  
We would like at least one ready zero to two, two, ready, two to five and three ready 5 plus unique European successes, right?

0:9:26.620 --> 0:9:27.60  
Sinead Duffy  
Hmm.

0:9:17.680 --> 0:9:38.950  
Interviewee B  
And then we want a 5050 gender balance, because otherwise we're not going to get. We've already reached 5050 GEOGRID 14 plus, but we need to maintain it and then we don't want to spread people too thin. So an individual can only be included into session pools or plans at the same readiness level at any one time, because otherwise they're being spread really thin and we're overcooking our succession, right.

0:9:39.260 --> 0:9:39.640  
Sinead Duffy  
Yeah.

0:9:40.70 --> 0:9:47.800  
Interviewee B  
We view European talent more holistically, so we also look at European talents that are based in other EU or they might be part of the system.

0:9:49.230 --> 0:10:20.320  
Interviewee B  
And then what we've said actually is you know, it's not good enough just to say this person's in succession, then we've got to actually really identify where the gaps are, because that's where you know you're you're analytical piece comes in. So we we're asking for him ready to to five those people to actually have clear development plans in place and what are the skills gaps, what are the visibility actions and what are the experiences they need. And then the five plus level which is quite often why we forget about people, they also need to be mentored.

0:10:20.390 --> 0:10:50.610  
Interviewee B  
So it's less hands on, but it's more about being clear what those development plans are and we're moving them towards the goal. And then there's also a much bigger push to have consistency between succession and segmentation. So basically we expect at least it job grades or 12 plus if you're a hypo that you should be on a succession plan for something, right, it doesn't necessarily be the Ault roles, but the Ault minus one roles because other.

0:10:52.560 --> 0:10:52.930  
Sinead Duffy  
Yeah.

0:10:50.700 --> 0:11:2.570  
Interviewee B  
Otherwise, what are they hypo for? And likewise where you filled out the high potential templates that you've just done, we've asked what succession plans are in because if our typos aren't in succession it like job grade 12 plus.

0:11:3.610 --> 0:11:23.200  
Interviewee B  
Less so. It's sort of like the 1011, right, because they might, but it job grade 12 plus they should be in succession for these URTOU t -, 1 rolls or the global functional Lt roles, because otherwise why they hypo, you know, they've got the opportunity to accelerate. So we would expect to have that strong linkage.

0:11:37.390 --> 0:11:37.830  
Interviewee B  
Ohh.

0:11:24.410 --> 0:11:54.380  
Sinead Duffy  
Okay. So you're looking at it very much at more holistic view of succession rather than what I'm looking at, OK. And I suppose I'm looking at it from the manufacturing environment, which is slightly different. Again, you're looking at it for the top grades. I'm looking at it for the bottom and kind of looking at it, see right, where can the development, but it's common themes, it's all about development and how do we get the development items and figure out who's where on that development profile for what we're better word?

0:11:54.130 --> 0:12:24.700  
Interviewee B  
Yeah. So there's the two parts of it. There's ensuring that you nominate the right successes and that's what we're trying to do by making this a much more facts based conversation. And the other thing is that once you've nominated them, then having that really clear gap analysis and experiences and visibility in place so that you make sure that they actually accelerate and making sure that the succession conversation isn't separate to other HR conversations as well. And then data wise, we have, it's.

0:12:35.260 --> 0:12:36.90  
Sinead Duffy  
So.

0:12:25.180 --> 0:12:37.320  
Interviewee B  
In work day and you can call succession Pauls and succession profile and blah blah. But the succession profile things felt relatively new. So that kicked off this year. So we're in the process of developing them.

0:12:37.780 --> 0:12:45.210  
Sinead Duffy  
But it's kind of you've still wait, you're still basing it on people filling in their talent profiles and filling in their development items, right?

0:12:49.40 --> 0:12:49.550  
Sinead Duffy  
Okay.

0:12:46.200 --> 0:13:16.210  
Interviewee B  
So this is being done a bit outside of that, but the expectation is that they would have they, you know, everyone needs to have career profiles up to date because at the level we're talking about their, the conversations are happening between the ult and all of that sort of stuff. But the idea is for hypos, for example, the expectation is that they all have 702010 plans in place and they're updated in work day. So whilst these plans are sitting outside of that, that shouldn't mean that these people also don't have plans in work, but we haven't gone.

0:13:16.610 --> 0:13:38.40  
Interviewee B  
Because of work day and development items and everything is not very helpful and also this is a very targeted conversation, but the expectation would be that for those acceleration development plans and everything that when all the conversations are happening after this because obviously people know that they're in succession for X and blah blah, then the they are reflecting those in their broader development plans as well.

0:13:39.40 --> 0:13:43.610  
Sinead Duffy  
OK. Very interesting. Very good. Thank you. Thank you for letting me through.

0:13:42.420 --> 0:13:54.50  
Interviewee B  
So I know this is an exactly what you are looking for, but I think the the idea is that that succession needs to become increasingly fact and data-driven.

0:13:54.690 --> 0:13:55.210  
Sinead Duffy  
Hmm.

0:13:56.520 --> 0:14:9.830  
Interviewee B  
Depending on what level you're looking at, what facts and data they are or not. But yes, gender needs to. If you have gender targets, gender needs to be in there. She needs to be A to identify them and we do have gender balance targets, so we need to have gender in there.

0:14:10.520 --> 0:14:15.990  
Sinead Duffy  
And his tenure is tenure in the role more important than tenure in the company.

0:14:18.610 --> 0:14:19.480  
Sinead Duffy  
10 years doesn't.

0:14:21.810 --> 0:14:21.970  
Sinead Duffy  
Yeah.

0:14:17.20 --> 0:14:26.990  
Interviewee B  
I not really. I mean it's more about the experience as you can get. So when you actually look at these or critical experiences and things, it will be more of an issue with visibility.

0:14:27.660 --> 0:14:33.800  
Interviewee B  
So because if someone's run actively new, then obviously you're gonna do more actions around visibility, so people know who they are.

0:14:34.510 --> 0:14:46.470  
Interviewee B  
Um, but again, 10. You're enrol, we would say generally it's about three years, but for high potentials it ends up with a quicker acceleration cycle. So they end up doing a couple of years and before they move on.

0:14:47.180 --> 0:14:48.380  
Sinead Duffy  
Okay okay, September.

0:14:47.450 --> 0:14:57.110  
Interviewee B  
But there's also because our hyper potential expectation is that they're in, like stretch assignments and stuff. So that doesn't necessarily have to be a new role, but it has to have elements of the role that continued to stretch them.

0:14:57.800 --> 0:15:15.330  
Sinead Duffy  
OK. OK, very good. That's very, very helpful. Thank you. One of the things I had meant to say was that you can withdraw your consent to to take part in all of this at any stage. And I will well any stage up till Friday when I publish as in when I submit not publish submit.

0:15:15.950 --> 0:15:16.550  
Sinead Duffy  
Ohh.

0:15:20.260 --> 0:15:20.710  
Sinead Duffy  
No.

0:15:30.350 --> 0:15:30.510  
Sinead Duffy  
And.

0:15:30.780 --> 0:15:31.400  
Sinead Duffy  
Ohh no no no.

0:15:32.900 --> 0:15:33.330  
Sinead Duffy  
Here.

0:15:15.770 --> 0:15:37.140  
Interviewee B  
But you're not sharing the detailed content or anything, are you? So yeah, I just mentioned to Helen tomorrow that I've taken you through how we're doing it, but I haven't shown you any confidential data, and it's not like you're publishing the actual slides or anything. So it's the principles behind it as not the principles are all fine. That's why I said to you, let me just quickly show you one without the names on it.

0:15:38.150 --> 0:15:39.950  
Interviewee B  
And then, yeah.

0:15:37.720 --> 0:15:49.120  
Sinead Duffy  
Yeah. Which is perfect. Yeah. And I'm. I'm not sharing the video. I'm just sharing the transcript. And I will take um, I will take some of the. Well, if there's anything that I think is confidential, I'll just block it out.

0:15:50.260 --> 0:15:56.290  
Sinead Duffy  
Rather than keep it in but um and if they want to ask for it, they can come back and ask me for it, so that's fine.

0:16:4.760 --> 0:16:5.510  
Sinead Duffy  
No, no, no.

0:15:55.690 --> 0:16:9.520  
Interviewee B  
I can't see that anything that I've just told you would be considered as particularly common cause I've taught we've talked about whilst I've shown you an example, I haven't gone through the details and I I can't see that this is this. There's anything in there that would be.

0:16:12.660 --> 0:16:14.280  
Interviewee B  
Yeah, exactly.

0:16:18.280 --> 0:16:18.540  
Interviewee B  
Yeah.

0:16:10.160 --> 0:16:29.490  
Sinead Duffy  
It it's more about process than than anything else which is and that's that's 100% reason why I went and I got an external data set because it's totally anonymized and it's a level. It's a step back from the company and the date is freely available. So I just added that one column of tenure because I think.

0:16:30.260 --> 0:16:32.440  
Sinead Duffy  
It gives you good balance on on what?

0:16:33.510 --> 0:16:34.820  
Sinead Duffy  
Of represents it nothing.

0:16:33.730 --> 0:16:38.820  
Interviewee B  
I I I think it's useful to have and time enrol is useful to have as well.

0:16:40.440 --> 0:16:53.250  
Interviewee B  
Because also you don't want people who are cycling through roles consistently because they need to get the depth of experience. And it's also very dependent on, but whether they're in succession for something that's like it's more the breadth of experience that you have rather than the depth of experience.

0:17:0.230 --> 0:17:0.510  
Sinead Duffy  
Yeah.

0:16:54.910 --> 0:17:19.960  
Interviewee B  
But personally I'm in a place where the more data fields you can get the better right? Because you never know and I'm finding with the talent role, you never know what you're needing to analyse against for. So even things like short term assignments, they've been on and you know a lot of because we've done a bit of work with our talent data here and with all the other talent consultants. So we have eaten employee headcount report which you've seen.

0:17:21.120 --> 0:17:35.870  
Interviewee B  
Satellites are stuff, right? So we did some work to shuffle it around a bit, but what I find is having everything in one place is super helpful. So you need to know whether the hypo you need to know that gender balance you do it is useful to know their time in the company.

0:17:59.100 --> 0:17:59.590  
Sinead Duffy  
Yeah.

0:17:37.290 --> 0:18:5.930  
Interviewee B  
Things like that, because all of and past roles what they've been in succession for in the past, what they've been slate, we haven't got all of this in one place by the way. But it would be useful to have it like what they've been slated for where they've been successful on the slating where they haven't been successful on slating. But obviously The Company has as I'm finding out quite a unique way of doing it with the whole slating and it's more but I understand what Helen says that succession is for senior roles, but purely because you've got to then be able to.

0:18:6.10 --> 0:18:14.20  
Interviewee B  
Well, through all that and you don't want to give people a you if you've got people mix succession, you've gotta have, you know, the ability to sort of get them there or move them towards it.

0:18:14.600 --> 0:18:15.390  
Sinead Duffy  
Exactly.

0:18:14.810 --> 0:18:23.300  
Interviewee B  
But Helen will be able to give you a much broader view, because she's obviously been in the space for significantly longer than me, who's been in it since June.

0:18:26.890 --> 0:18:30.620  
Sinead Duffy  
Yeah. Well, and you're both longer in it than I am, which is April.

0:18:31.100 --> 0:18:31.500  
Interviewee B  
On it.

0:18:31.20 --> 0:18:33.270  
Sinead Duffy  
Um, what was it gonna say?

0:18:37.670 --> 0:18:38.300  
Sinead Duffy  
It is.

0:18:32.190 --> 0:18:39.950  
Interviewee B  
Yeah. June is when I started the talent consultant stuff. So I'm I'm hope that made sense because this is sort of what I've learned about it soon.

0:18:41.200 --> 0:18:55.160  
Sinead Duffy  
It did well. You were very coherent. Somebody's only learning about it. Can I just ask one other question? And what I'm here, what I'm reading and what I'm saying is, is that the demand for HR has been more data-driven is I can see it in my job. Can you see it reflected in your job?

0:18:56.760 --> 0:18:57.430  
Sinead Duffy  
And it's.

0:19:25.840 --> 0:19:26.340  
Sinead Duffy  
Yeah.

0:18:56.390 --> 0:19:27.90  
Interviewee B  
Yeah, and. And we have to be, um. And even when I think about when we've prepped for the PDFs, the amount of data that we we put in there, the problem we're having is access to data and that sort of an ongoing conversation. That's what we're having because what you find and what we're finding is that data in work day, for example, is often is reported for a specific reason. But that's not the reason that HR wants to use it. And a good example is timing.

0:19:28.60 --> 0:19:28.540  
Sinead Duffy  
Hmm.

0:19:27.170 --> 0:19:57.80  
Interviewee B  
No, but it can be really difficult to work out how long someone's been enrolled because it can change because their position changed or something like that. And that's all over the place. But I think if you look at sort of top skills for the company, like data-driven, storytelling, data and analytics, they're increasingly important and we have to be the same in HR. So we can't things like succession need to be made off facts, not feelings. And unless you've got the data to it will never become 100% data, right, cause you've got to interpret it and all of that.

0:20:1.100 --> 0:20:1.710  
Sinead Duffy  
Yeah.

0:19:57.220 --> 0:20:12.250  
Interviewee B  
But I think, um, yes, absolutely. Then the the issue becomes access to that data interpretation of that data that where the data sits and it's difficult and the purpose that the data has been created for versus the one that you want to use it for.

0:20:13.500 --> 0:20:14.10  
Sinead Duffy  
Who like?

0:20:13.170 --> 0:20:32.560  
Interviewee B  
Which were finding is very painful, but you know things like we try and calculate what I'm trying to at the moment in import export data. So how many people from outside of the EU have we imported into EU? Cause we can't fill ourselves or it's been part of their development versus how many people were exporting. So they're getting those experiences.

0:20:33.540 --> 0:20:34.250  
Interviewee B  
Nightmare.

0:20:34.910 --> 0:20:40.460  
Sinead Duffy  
Yeah. Yeah, I can imagine. Month. Yeah. Yeah, yeah.

0:20:35.310 --> 0:20:46.350  
Interviewee B  
So nightmare to get hold of that it's very manual so. But I think increasingly yes, that it has to be there has to be robust data behind it.

0:20:53.730 --> 0:20:54.230  
Interviewee B  
Yeah.

0:20:47.900 --> 0:21:0.190  
Sinead Duffy  
Very good. Thank you so much. I'll shoot on the the consent forms. Debbie, in the next day or two and I'll, I'll shoot it through and DocuSign. So if you want to have a look and read it, then it's there any questions or anything like that, let me know.

0:20:59.670 --> 0:21:2.620  
Interviewee B  
Was that okay? Cause I know it wasn't exactly what you were looking for, but.

0:21:2.440 --> 0:21:24.530  
Sinead Duffy  
No, it's perfect. It's exactly what I need. It's it's the real life. And I mean basically what you've what you've confirmed is what I'm reading in the literature over the last while and the demands. And I can see it in my own job. But it's your dealing with, I supposed the senior executives a lot more than I am and it's good to see that that's being filtered down. But also you know.

0:21:24.990 --> 0:21:35.360  
Sinead Duffy  
Ohh it which is interesting is that you guys are recruiting for the higher rules which is fine and we're in CPS there. They're recruiting for a wider breadth of roles.

0:21:36.290 --> 0:21:36.760  
Interviewee B  
Yeah.

0:21:36.240 --> 0:21:52.490  
Sinead Duffy  
So they have to go for a bit there generally. Well the, you know, the top levels that's that's done at one level, but they have to recruit for within functions. So if you're looking at like manufacturing the guys down on the floor, they've got they have the.

0:21:56.210 --> 0:21:56.530  
Interviewee B  
Hmm.

0:21:53.630 --> 0:21:59.690  
Sinead Duffy  
It's the career, it's the job framework. It's a, it's a version of the carer job architecture that's coming out for us.

0:22:5.630 --> 0:22:5.910  
Interviewee B  
Hmm.

0:22:16.970 --> 0:22:17.450  
Interviewee B  
Yeah.

0:22:26.0 --> 0:22:26.420  
Interviewee B  
Yeah.

0:22:0.670 --> 0:22:30.520  
Sinead Duffy  
But they get to see and then they stop. See is where they're fully developed, and then D is when they go into like, a specialist role, but they have to be advertised. But it's the recruitment they do at the succession planning they do is kind of more to the deed level. And then how did goes into like management into like other functions and stuff like that. So it's different and you have the same. It's just A at at at a broader level than what he is.

0:22:32.620 --> 0:22:32.950  
Interviewee B  
Hi.

0:22:30.640 --> 0:22:35.270  
Sinead Duffy  
This update here, so it's it's just it's different but it's the process is the same.

0:22:36.80 --> 0:22:40.90  
Interviewee B  
Perfect, right. I'm going off to talk about talent segmentation with the yard now, so.

0:22:40.560 --> 0:22:43.770  
Sinead Duffy  
Have fun. Thank you so much. I really appreciate this.

0:22:41.810 --> 0:22:46.760  
Interviewee B  
Thank you and I hope, sorry, I'm I know that wasn't exactly, but hopefully that was somewhat helpful.

0:22:49.530 --> 0:22:51.580  
Interviewee B  
You're welcome. Bye.

0:22:46.950 --> 0:22:52.300  
Sinead Duffy  
It's perfect. Thank you so much for your time. I'm gonna talk to you soon. Take care. Bye. Bye. Bye.

# Interview C - Transcript

0:0:5.560 --> 0:0:21.330  
Sinead Duffy  
Right. So um, thanks a million. First off for agreeing to do this, NAME. So just I supposed to set the scene and let you know what I'm doing. So I'm in the kind of in the stage of completing my masters in data analytics through my MCC. And my topic is.

0:0:30.850 --> 0:0:31.320  
Interviewee C  
From.

0:0:22.650 --> 0:0:51.680  
Sinead Duffy  
It's learning data and how it can, might or maybe used to support succession planning. OK, and I suppose I'm kind of taking myself as an example. OK, so I'm in HR. Data analytics isn't normally a part of like a sexual hate or that is widely popular or people would consider going well together. So that's, I suppose I'm an outlier in that sense. And that's where I think the value in the learning data is because.

0:0:52.140 --> 0:0:55.720  
Sinead Duffy  
When you look at data of the learning data you've got, you know.

0:0:56.40 --> 0:1:26.450  
Sinead Duffy  
I'm training that is mandatory and I suppose I'm. I'm thinking mostly of a manufacturing environment, so or even ours, but it's a it's mandatory. That's the basic minimum that everyone does. Then you've got maybe developmental or manager assigned training. So that kind of upset a little bit and then you've got people who are actively using the system engaged, driving their own learning during eight point for a better world. And I think I'm kind of one of those kind of people because I'm looking at stuff that likes.

0:1:26.540 --> 0:1:34.430  
Sinead Duffy  
Statistics and a I and all of that, which doesn't really link into HR, was not really Hitler related, but it is.

0:1:35.260 --> 0:1:35.720  
Interviewee C  
Hmm.

0:1:35.620 --> 0:1:40.450  
Sinead Duffy  
OK. So that's the kind of they're the kind of people that I'm interested in in seeing.

0:1:41.210 --> 0:1:47.40  
Sinead Duffy  
Look, with this work for so. Um, I've had conversations with um Interviewee B on.

0:1:53.270 --> 0:1:53.800  
Interviewee C  
Hmm.

0:1:48.80 --> 0:2:3.490  
Sinead Duffy  
On how the new process is and I don't know if you know Interviewee A, he's the capabilities director for THE COMPANY. So Interviewee A is taking me through the THE COMPANY version of it, which is very interesting because they're it's coming out of from two different levels, but the process is still the same.

0:2:4.220 --> 0:2:16.850  
Sinead Duffy  
Which is good to know and but I kind of wanted to ask you because I know you've had a lot of experience outside of the Company and how you find the succession planning process and your feelings on it.

0:2:17.960 --> 0:2:20.490  
Interviewee C  
Ohh, I find the succession planning process here.

0:2:21.80 --> 0:2:28.430  
Sinead Duffy  
Or either either I'm I'm. I'm not proud. I'll. I'll take your wealth of experience. I know you're next person.

0:2:27.910 --> 0:2:33.260  
Interviewee C  
Okay. So is it in general the practise of succession planning and how I feel about it?

0:2:33.570 --> 0:2:42.220  
Sinead Duffy  
Yeah, exactly. And then do you think that there is a there is space for data analysis in it and if so, what would that look like?

0:2:42.340 --> 0:3:7.300  
Interviewee C  
Yeah, absolutely. There is an absolute huge space for data analysis in it, because I think the key thing around succession planning is that it has to work for the business. So I have worked in businesses where at employer A, for instance, it was such an integral part of the work that they did and they truly believed in investing in kind of key successors.

0:3:7.460 --> 0:3:37.670  
Interviewee C  
Planning it and then utilising that data in order to look at moving people around, giving them the right experiences, understanding what experiences are needed and when people were appointed internally, you could always see that there were part of the succession. So it has really worked and they leave her succession and it supported the business and they the business was doing really well. If I looked at all of the CEO, they were all Internet internally greened.

0:3:37.770 --> 0:3:40.780  
Interviewee C  
At employer A, which was kind of really, really great.

0:3:42.550 --> 0:3:49.660  
Interviewee C  
An organisation Rd didn't work was employer B where from a financial services perspective there was.

0:3:49.810 --> 0:4:19.850  
Interviewee C  
Requirement from a regulatory body to do succession planning in order to mitigate risks and in terms of business continuity. And so people didn't really believe in the practise, they just felt like they had to do it because it's part of a regulatory requirement and there it didn't work. You would kind of see a lot of people on the succession plan because they just put their fingers in the air. There wasn't really a criteria to kind of measure it on. And so the plants.

0:4:19.950 --> 0:4:21.640  
Interviewee C  
South were not activated.

0:4:22.630 --> 0:4:23.40  
Sinead Duffy  
OK.

0:4:22.570 --> 0:4:52.300  
Interviewee C  
And so here it the Company. I can truly say that we believe in the practise and that we utilise the practise to our advantage to make sure that we have the best people in the in the right, the right talent and the right roles at the right time and to also kind of understand how do we need to move people in order to get them the experiences for them to move into bigger, greater roles that deliver more impact travel to us as a business and.

0:4:53.100 --> 0:5:6.970  
Interviewee C  
That's why I feel like succession planning. You has to work for the business and has to be an integral part, and people need to believe in it in order to make it work in terms of data.

0:5:7.890 --> 0:5:10.100  
Interviewee C  
It's so important because.

0:5:11.110 --> 0:5:41.300  
Interviewee C  
Just simple data such as has our succession plan actually been activated, so when we appoint someone internally into a senior role, so drop rate 14 and above all UL t -, 1 where we should have succession plans. Has this person actually been on a succession plan or pool or not? So you can kind of with that data kind of say actually our practises working or it's not or maybe it's the mindset or the operating model or whatever it is. So having that data is super important.

0:5:41.480 --> 0:5:52.190  
Interviewee C  
Another data that we really need and succession planning is when you look at the readiness categories. So you have ready zero to two. You have ready to to five, you have ready 5 plus.

0:5:52.640 --> 0:5:53.330  
Sinead Duffy  
OK.

0:5:52.960 --> 0:6:23.250  
Interviewee C  
And so we need to see people moving from those readiness categories into others. And so kind of just having data on the movement of people are moving would be amazing or kind of looking at data on high potentials on succession plans, ideally 100% of hypertension should be unsuccessful plan. But it doesn't mean that all success, 100% of successes need to be high potential, but just kind of like that data to make sure that our talent.

0:6:23.320 --> 0:6:36.900  
Interviewee C  
Practises are also integrated. It's a talent segmentation as feeding succession planning basically and so there's a lot of data that we need in order to make this really helpful. You know, I could go on and on.

0:6:39.320 --> 0:7:7.450  
Sinead Duffy  
No, but it it is an interest for me. It's interesting because I wouldn't have been so gung ho about it. And for me, I think we we invest a lot in training, we we spend loads of money like I know you guys have just finished the the employees distance programme for across Europe and I know in, in THE COMPANY we do a lot, they do a lot of training as well. So the company is has there's such a fantastic resources out there and that's why I think it's really important to kind of maybe.

0:7:7.530 --> 0:7:13.160  
Sinead Duffy  
See if there's a link between the two and can one be used to enhance the other and again.

0:7:21.890 --> 0:7:22.320  
Sinead Duffy  
Yeah.

0:7:31.280 --> 0:7:31.730  
Sinead Duffy  
Yeah.

0:7:13.460 --> 0:7:43.420  
Interviewee C  
And I also would see it much broader than that, Sinead. I would not only kind of look at the training because that's only 10% education is just 10%. I would also look at things like mobility packages, how much as an organisation actually do we spend on mobility in order for our people to get the right experience. And I can say it's extortionate. It's much more than any other company that I worked with. And so you need to kind of like almost combine different measures and kind of stop thinking more holistically around what.

0:7:54.910 --> 0:7:55.440  
Sinead Duffy  
Okay.

0:7:43.500 --> 0:8:2.590  
Interviewee C  
Actually contributes towards somebody being ready for a succession role. It's primarily experiences. It's also visibility and it is the skills and knowledge. And so if you kind of combine the three, visibility is a little bit more harder to quantify in terms of investment. But the other two definitely.

0:8:20.130 --> 0:8:20.480  
Interviewee C  
Yeah.

0:8:3.120 --> 0:8:22.150  
Sinead Duffy  
Yeah, okay, that's really interesting cause I hadn't thought of that. And I suppose I've had to the nature of this is that I've had only look at at learning, but as a future opportunities, I think it's it kind of does widen it open like you could because you could if you have access to the data you could link all of them and see because it's something.

0:8:25.690 --> 0:8:26.10  
Sinead Duffy  
We will.

0:8:29.320 --> 0:8:29.610  
Sinead Duffy  
Yeah.

0:8:21.300 --> 0:8:52.310  
Interviewee C  
So, like, look at Candidate A. She's on a mobility contract. She's on a succession plan. Candidate B and mobility. He's on a succession plan, you know. Then you kind of look at Candidate C. I don't know if he's on on a mobility or not on an international assignment or not, but you look at Candidate D, you look at Candidate E, definitely on success. You know, you kind of look at all of the people that are moving around who are successors and you can easily pull that data together for Europe and kind of say this is how much we spend at job grade.

0:8:52.380 --> 0:9:0.820  
Interviewee C  
14 and above, this is how much our own, like an international assignment to get the experiences that they need, even STAS, you know, like all of this.

0:9:1.960 --> 0:9:3.630  
Interviewee C  
Sorry I couldn't come on and talk.

0:9:3.910 --> 0:9:7.950  
Sinead Duffy  
No, but it's it is, and I think that there it's an area that.

0:9:8.640 --> 0:9:10.790  
Sinead Duffy  
I think hasn't been given.

0:9:11.600 --> 0:9:26.120  
Sinead Duffy  
Well, I don't know. I find that in HR and definitely in the literature, there's no research out there on using heat or data and learning data in this way. And like it's very difficult to get hit or lead.

0:9:29.320 --> 0:9:29.660  
Interviewee C  
Yeah.

0:9:26.930 --> 0:9:42.410  
Sinead Duffy  
Research being honest, it it's generally on on, I don't know, like on the processes in Haiti or something like that, or or how we can use AI to do basic things like, you know, type letters or, you know, do standard things.

0:9:41.430 --> 0:9:53.980  
Interviewee C  
You know what would be really interesting is if you looked at the investment that we put into grooming our successors. Even if you just take one or two versus recruiting somebody externally.

0:9:55.400 --> 0:9:55.730  
Sinead Duffy  
Yeah.

0:9:55.160 --> 0:10:24.150  
Interviewee C  
It would be really interesting to kind of find out how that measures up and whether we as a company want to make a decision based on that. I think we'll still continue to groom and develop people internally simply because they then also have our knowledge or institutions, the relationship, the network which you can't pay for. But it's also kind of like thinking about it because either you recruit somebody externally or you develop internally for these kind of senior roles.

0:10:24.600 --> 0:10:51.490  
Sinead Duffy  
Yeah, exactly. Can I ask, do you find? And? And this is, I suppose, away from kind of away from succession planning in a way, but more open to like they hate our part of the general part of it. Do you find that there's from the business to know you work really closely with the business? Is there more requests for data like for HR to take an active part in like having dated driven results. So rather than you know?

0:11:2.50 --> 0:11:2.310  
Sinead Duffy  
Yeah.

0:11:18.730 --> 0:11:19.170  
Sinead Duffy  
Yeah.

0:10:50.420 --> 0:11:20.570  
Interviewee C  
Ohh absolutely. So I'm constantly asked about import export data so I drop grades fourteen and above or UL t -, 1 rows. How many do we import and how many people do we export? And you know because obviously some of the experiences we just not able to give them an hour OU and therefore they need to kind of be able to move out. And then similarly if we're importing too many people that means we don't give the opportunities to our people in Europe in order to grieve that pipeline. So that's kind of really important data.

0:11:20.640 --> 0:11:32.190  
Interviewee C  
That they're always asking about, um. And what was the other kind of data, the hit rate amounts kind of like success in on us being on slates, you know, so it job rights 14 and above.

0:11:32.370 --> 0:11:43.10  
Interviewee C  
And there is no open marketplace at the moment. People cannot apply the talent and development directors put slates forward in kind of say, you know, these are the kind of people that we would like to.

0:11:43.550 --> 0:12:0.110  
Interviewee C  
Um to put on the slide enduring the dude, and throughout the interview process, and we're currently doing an analysis of how many slates were we long listed for our Europe talent, shortlisted for our Europe talent, and how many rolls did they get, and how does it compare to other olus?

0:12:1.240 --> 0:12:3.870  
Sinead Duffy  
Hey, so be really interesting to kind of have those.

0:12:5.10 --> 0:12:22.280  
Sinead Duffy  
Have that kind of data before you put someone forward because like obviously it's not. The data is the decision is going to be made on the data at the end of it. It's HR, there is the people part of it, but it might give you a strong indication of whether or not that person is gonna be a fit for the role.

0:12:21.770 --> 0:12:25.950  
Interviewee C  
Ohh yeah, and Pte does as well. PT is very um.

0:12:27.290 --> 0:12:31.570  
Interviewee C  
Yeah, it is. Um, I don't know if you know about, you know, PT, right.

0:12:32.110 --> 0:12:32.320  
Sinead Duffy  
No.

0:12:32.390 --> 0:12:34.190  
Sinead Duffy  
Ohh, what's PC positions?

0:12:37.140 --> 0:12:39.990  
Sinead Duffy  
Ohh yes, yes, yes, yes, I do, yeah.

0:12:33.170 --> 0:12:46.580  
Interviewee C  
Ohh Okay Pte is an assessment for job grades fourteen and above. Yeah, yeah, yeah. And so, I mean, seriously, the way that the data is kind of coincidence coinciding and combining with.

0:12:47.770 --> 0:12:52.520  
Interviewee C  
Um with reality? It's just mind boggling. It's really good.

0:13:9.870 --> 0:13:10.560  
Interviewee C  
Nice.

0:12:53.150 --> 0:13:11.150  
Sinead Duffy  
Yeah. Yeah, cause we did a lying actually with our with the GB and I Lt team when we went to Athens last week and that was really interesting because we did the align. No one got the results and then we had the discussion and what was raised in the discussion is exactly what was raised on the line. So it was great to have the.

0:13:11.730 --> 0:13:30.20  
Interviewee C  
And you know what's also really great is that you have the right team dynamics and culture for people to be be able to open up so that it kind of actually reflects because in some teams what happens is that everything is positive and great and blah, blah. But then actually you look at the aligning perspective and it isn't so.

0:13:30.20 --> 0:13:42.870  
Sinead Duffy  
Yeah. Yeah. No, that's really, really good. And, but I suppose it just showed the power of it just reinforced for me the power of of data and like a 5 minute exercise and you get exactly where where you are on the line. Sounds really interesting.

0:13:43.950 --> 0:13:47.110  
Sinead Duffy  
Very good. I don't have any other questions. Sorry. That was short and sweet.

0:13:48.560 --> 0:13:49.70  
Sinead Duffy  
Ohh.

0:13:46.450 --> 0:13:49.820  
Interviewee C  
Okay perfect. You can stop the recording.

0:13:50.410 --> 0:13:52.770  
Sinead Duffy  
I can let me stop this now.

0:13:55.30 --> 0:13:57.260  
Sinead Duffy  
Do I stop, stop recording?

# Appendix C: Sample Consent Form

Informed Consent for Research Interviews

**Working Title:** Can demographic data support the succession planning process: exploring the role of data analytics.

**Introduction**:

Hello, my name is Sinead Duffy, and I am a graduate student at CCT College Dublin. Thank you for showing interest in participating in our research study titled "Can demographic data support the succession planning process: exploring the role of data analytics." Before we begin the interview process, I would like to provide you with some important information regarding your participation and obtain your informed consent.

**Purpose of the Study**:

The purpose of this study is to explore the succession planning process within a multinational company and examines if there is a role for data analytics to support the process. Data from the learning management system is used to complete the analysis including demographic information on employees. We hope to gather insights that will help us to understand if development of a model to explore demographic information will have any application in a real-world setting.

**Participant Requirements**:

To participate in this study, you must have at least 10 years’ experience in the succession planning process within a multinational company.

**Procedure**:

During the interview, I will ask you a series of open-ended questions related to the study topic. The interview will take approximately 30 minutes to complete. We may also ask follow-up questions to clarify your responses or gather additional information. This conversation will have a transcript recorded for analysis purposes.

**Confidentiality**:

We assure you that your identity and the information you provide will be kept confidential. In any publications resulting from this study, we will use a pseudonym, or generic designations to maintain your anonymity. The transcripts will be stored securely and will only be accessible to members of the research team.

**Voluntary Participation**:

Your participation in this study is completely voluntary. You have the right to withdraw from the study at any time or refuse to answer any questions without any consequences.

If you decide to withdraw your agreement to participate in the study, you must email the main researcher (Sinead Duffy) to confirm the withdrawal of your consent. The main researcher will then remove the transcript of your interview and any data / insights that may have been used in the report.

**Risks and Benefits**:

There are minimal risks associated with participating in this study. You may experience feelings of discomfort or distress when discussing sensitive topics. Should this occur, you can choose to skip a question or withdraw from the study.

**Contact Information**:

If you have any concerns or questions about the research project, you may contact Sinead Duffy, at [sba2229@student.cct.ie](mailto:sba2229@student.cct.ie)

**Consent**:

By participating in this interview, you acknowledge that you are at least 18 years old, have been informed about the purpose of the study, the risks and benefits involved, and your rights as a participant. You voluntarily agree to participate in the study and provide your informed consent.

*(If issuing consent in writing the following will be included in the document)*

|  |  |
| --- | --- |
| Participant Name: |  |
| Participant Signature |  |
| Date: |  |

**Actual signed consent forms available on request.**

# Reference List

*1.4. Support Vector Machines — scikit-learn 1.3.0 documentation* (no date). Available at: https://scikit-learn.org/stable/modules/svm.html (Accessed: 30 August 2023).

*1.11. Ensemble methods* (no date) *scikit-learn*. Available at: https://scikit-learn/stable/modules/ensemble.html (Accessed: 31 August 2023).

*3.2. Tuning the hyper-parameters of an estimator* (no date) *scikit-learn*. Available at: https://scikit-learn/stable/modules/grid\_search.html (Accessed: 30 August 2023).

‘2017 Deloitte Global Human Capital Trends’ (2017).

Al-azazi, F.A. and Ghurab, M. (2023) ‘ANN-LSTM: A deep learning model for early student performance prediction in MOOC’, *Heliyon*, 9(4), p. e15382. Available at: https://doi.org/10.1016/j.heliyon.2023.e15382.

Alhakbani, H.A. and Alnassar, F.M. (2022) ‘Open Learning Analytics: A Systematic Review of Benchmark Studies using Open University Learning Analytics Dataset (OULAD)’, in *Proceedings of the 2022 7th International Conference on Machine Learning Technologies*. New York, NY, USA: Association for Computing Machinery (ICMLT ’22), pp. 81–86. Available at: https://doi.org/10.1145/3529399.3529413.

Almeida, F., Duarte Santos, J. and Augusto Monteiro, J. (2020) ‘The Challenges and Opportunities in the Digitalization of Companies in a Post-COVID-19 World’, *IEEE Engineering Management Review*, 48(3), pp. 97–103. Available at: https://doi.org/10.1109/EMR.2020.3013206.

*An introduction to seaborn — seaborn 0.12.2 documentation* (no date). Available at: https://seaborn.pydata.org/tutorial/introduction (Accessed: 29 August 2023).

Araka, E. *et al.* (2022) ‘Using Educational Data Mining Techniques to Identify Profiles in Self-Regulated Learning: An Empirical Evaluation’, *The International Review of Research in Open and Distributed Learning*, 23(1), pp. 131–162. Available at: https://doi.org/10.19173/irrodl.v22i4.5401.

Assaad, C.K., Devijver, E. and Gaussier, E. (2022) ‘Survey and Evaluation of Causal Discovery Methods for Time Series’, *Journal of Artificial Intelligence Research*, 73, pp. 767–819. Available at: https://doi.org/10.1613/jair.1.13428.

Bankins, S. (2021) ‘The ethical use of artificial intelligence in human resource management: a decision-making framework’, *Ethics & Information Technology*, 23(4), pp. 841–854. Available at: https://doi.org/10.1007/s10676-021-09619-6.

Bhardwaj, S. and Patnaik, S. (2019) ‘People Analytics: Challenges and Opportunities - A Study Using Delphi Method’, *IUP Journal of Management Research*, 18(1), pp. 7–23. Available at: https://search.ebscohost.com/login.aspx?direct=true&db=bsh&AN=134816499&site=eds-live (Accessed: 14 May 2023).

Brownlee, J. (2016) ‘Use Keras Deep Learning Models with Scikit-Learn in Python’, *MachineLearningMastery.com*, 30 May. Available at: https://machinelearningmastery.com/use-keras-deep-learning-models-scikit-learn-python/ (Accessed: 3 September 2023).

Brownlee, J. (2022) ‘How to Grid Search Hyperparameters for Deep Learning Models in Python with Keras’, *MachineLearningMastery.com*, 1 July. Available at: https://machinelearningmastery.com/grid-search-hyperparameters-deep-learning-models-python-keras/ (Accessed: 3 September 2023).

*Categorical encoding using Label-Encoding and One-Hot-Encoder | by Dinesh Yadav | Towards Data Science* (no date). Available at: https://towardsdatascience.com/categorical-encoding-using-label-encoding-and-one-hot-encoder-911ef77fb5bd (Accessed: 29 August 2023).

Chadwick, C. and Dabu, A. (2009) ‘Human Resources, Human Resource Management, and the Competitive Advantage of Firms: Toward a More Comprehensive Model of Causal Linkages’, *Organization Science*, 20(1), pp. 253–272. Available at: https://doi.org/10.1287/orsc.1080.0375.

Christian, A. (2022) *The case for job hopping*, *bbc.com*. Available at: https://www.bbc.com/worklife/article/20220720-the-case-for-job-hopping (Accessed: 16 May 2023).

Chunqiao Mi (2019) ‘Student Performance Early Warning based on Data Mining’, *International Journal of Performability Engineering*, 15(3), pp. 822–833. Available at: https://doi.org/10.23940/ijpe.19.03.p11.822833.

Clair, J.A. *et al.* (2019) ‘Loosening Categorical Thinking: Extending the Terrain of Theory and Research on Demographic Identities in Organizations’, *Academy of Management Review*, 44(3), pp. 592–617. Available at: https://doi.org/10.5465/amr.2017.0054.

Dahlbom, P. *et al.* (2020) ‘Big data and HR analytics in the digital era’, *Baltic Journal of Management*, 15(1), pp. 120–138. Available at: https://doi.org/10.1108/BJM-11-2018-0393.

Diez, F., Bussin and Lee, V. (2019) *Fundamentals of HR Analytics : A Manual on Becoming HR Analytical*. Available at: https://eds.s.ebscohost.com/eds/ebookviewer/ebook/ZTAyMG13d19fMjIwNDIyNV9fQU41?sid=12b53d14-610b-4f61-880e-acda26e9a4a9@redis&vid=4&format=EB&lpid=lp\_vii&rid=0 (Accessed: 16 September 2023).

Djoundourian, S.S. (2017) ‘Assessment of learning in business education: Standardized or homegrown?’, *Journal of Education for Business*, 92(5), pp. 238–244. Available at: https://doi.org/10.1080/08832323.2017.1339662.

Dong, F. (2022) ‘Construction of Enterprise Human Resource Intelligent Scheduling Model Based on Fuzzy Relationship’, *Mobile Information Systems*, pp. 1–13. Available at: https://doi.org/10.1155/2022/5342176.

Eberhardt, F. (2017) ‘Introduction to the foundations of causal discovery’, *International Journal of Data Science and Analytics*, 3(2), pp. 81–91. Available at: https://doi.org/10.1007/s41060-016-0038-6.

Edwards, M.R. and Edwards, K. (2019) *Preditice HR Analytics*. 2nd edn.

Ferrar, J. and Green, D. (2021) *Excellence in People Analytics*. 2nd edn. Kogan Page Ltd.

Gefferth, A. (2023) *One Hot Encoding scikit vs pandas*, *Medium*. Available at: https://towardsdatascience.com/one-hot-encoding-scikit-vs-pandas-2133775567b8 (Accessed: 29 August 2023).

*General Data Protection Regulation (GDPR) – Official Legal Text* (no date) *General Data Protection Regulation (GDPR)*. Available at: https://gdpr-info.eu/ (Accessed: 16 September 2023).

‘Handbook of Learning Analytics - Second edition’ (2022) *Society for Learning Analytics Research (SoLAR)*. Available at: https://www.solaresearch.org/publications/hla-22/ (Accessed: 16 September 2023).

Huselid, M.A., Beatty, R.W. and Becker, B.E. (2005) ‘A Players or A Positions?’, *Harvard Business Review*, 83(12), pp. 110–117. Available at: https://search.ebscohost.com/login.aspx?direct=true&db=bsh&AN=18916545&site=eds-live (Accessed: 14 May 2023).

Hussain, M. *et al.* (2018) ‘Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores’, *Computational Intelligence and Neuroscience*, 2018, p. e6347186. Available at: https://doi.org/10.1155/2018/6347186.

‘Hyperparameter tuning using GridSearchCV and KerasClassifier’ (2020) *GeeksforGeeks*, 25 November. Available at: https://www.geeksforgeeks.org/hyperparameter-tuning-using-gridsearchcv-and-kerasclassifier/ (Accessed: 3 September 2023).

*Importance of Succession Planning (With Benefits and Tips) | Indeed.com Canada* (no date). Available at: https://ca.indeed.com/career-advice/career-development/importance-of-succession-planning (Accessed: 16 September 2023).

Karakolis, E. *et al.* (2022) ‘Bridging the Gap between Technological Education and Job Market Requirements through Data Analytics and Decision Support Services’, *Applied Sciences*, 12(14), p. 7139. Available at: https://doi.org/10.3390/app12147139.

Katrina Sin and Loganathan Muthu (2015) ‘Application of Big Data in Education Data Mining and Learning Analytics – a Literature Review’, *ICTACT Journal on Soft Computing*, 5(4), pp. 1035–1049.

Kohnke, L., Foung, D. and Chen, J. (2022) ‘Using Learner Analytics to Explore the Potential Contribution of Multimodal Formative Assessment to Academic Success in Higher Education’, *SAGE Open*, 12(2), p. 21582440221089957. Available at: https://doi.org/10.1177/21582440221089957.

Kokoç, M. and Altun, A. (2021) ‘Effects of learner interaction with learning dashboards on academic performance in an e-learning environment’, *Behaviour & Information Technology*, 40(2), pp. 161–175. Available at: https://doi.org/10.1080/0144929X.2019.1680731.

Kulala, H. and Rani, K. (2017) ‘Advancements in Multi-Layer Perceptron Training to Improve Classification Accuracy’, *International Journal on Recent and Innovation Trends in Computing and Communication*, 5, pp. 353–357.

Kumar, R. (2011) *Resarch Methodology*. 3rd edn. SAGE.

Kuzilek, J., Hlosta, M. and Zdrahal, Z. (2017) ‘Open University Learning Analytics dataset’, *Scientific Data*, 4(1), p. 170171. Available at: https://doi.org/10.1038/sdata.2017.171.

*Learning Analytics – A Growing Field and Community Engagement* (2015) *OpenAIRE - Explore*. Available at: https://explore.openaire.eu/search/publication?pid=10.18608%2Fjla.2015.21.1 (Accessed: 14 May 2023).

Loftus, J.R. *et al.* (2018) ‘Causal Reasoning for Algorithmic Fairness’. arXiv. Available at: https://doi.org/10.48550/arXiv.1805.05859.

Losey, M.R., Meisinger, S.R. and Ulrich, D. (eds) (2005) *The future of human resource management: 64 thought leaders explore the critical HR issues of today and tomorrow*. Alexandria, Va. : Hoboken, N.J: Society for Human Resource Management ; Wiley.

Malinsky, D. and Danks, D. (2018) ‘Causal discovery algorithms: A practical guide’, *Philosophy Compass*, 13(1), p. e12470. Available at: https://doi.org/10.1111/phc3.12470.

Martin, L. (2019) ‘Leading Practices to Upskill HRBPs as Ambassadors for People Analytics’, *Workforce Solutions Review*, 10(3), pp. 24–27. Available at: https://search.ebscohost.com/login.aspx?direct=true&db=bsh&AN=140284586&site=eds-live (Accessed: 14 May 2023).

Mattox II, J.R., Parskey, P. and Hall, C. (2020) *Learning Analytics: Using Talent Data to Improve Business Outcomes*. 2nd edn. Kogan Page Ltd.

M.Sc, G.A. (2023a) *10 Must-Know Python Libraries for Every Programmer*, *Medium*. Available at: https://levelup.gitconnected.com/10-must-know-python-libraries-for-every-programmer-bc4f18427e38 (Accessed: 29 August 2023).

M.Sc, G.A. (2023b) *Let’s Talk About the Real MVPs of Python — Libraries*, *Medium*. Available at: https://levelup.gitconnected.com/lets-talk-about-the-real-mvps-of-python-libraries-9fbb03dc1d1e (Accessed: 29 August 2023).

‘Multiple deprivation index’ (2023) *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=Multiple\_deprivation\_index&oldid=1172805397 (Accessed: 9 September 2023).

Muslim, A., Chatti, M.A. and Guesmi, M. (2020) ‘Open Learning Analytics: A Systematic Literature Review and Future Perspectives’, in N. Pinkwart and S. Liu (eds) *Artificial Intelligence Supported Educational Technologies*. Cham: Springer International Publishing (Advances in Analytics for Learning and Teaching), pp. 3–29. Available at: https://doi.org/10.1007/978-3-030-41099-5\_1.

Mustafa Yağcı (2022) ‘Educational data mining: prediction of students’ academic performance using machine learning algorithms’, *Smart Learning Environments*, 9(1), pp. 1–19. Available at: https://doi.org/10.1186/s40561-022-00192-z.

*One-way ANOVA - An introduction to when you should run this test and the test hypothesis | Laerd Statistics* (no date). Available at: https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php (Accessed: 18 September 2023).

Peng, J. *et al.* (2021) ‘DataPrep.EDA: Task-Centric Exploratory Data Analysis for Statistical Modeling in Python’, *Research Data Science & Engineering Track Paper* [Preprint].

Poudyal, S., Mohammadi-Aragh, M.J. and Ball, J.E. (2022) ‘Prediction of Student Academic Performance Using a Hybrid 2D CNN Model’, *Electronics*, 11(7), p. 1005. Available at: https://doi.org/10.3390/electronics11071005.

Python, R. (no date) *Logistic Regression in Python – Real Python*. Available at: https://realpython.com/logistic-regression-python/ (Accessed: 30 August 2023).

Rasmussen, T. and Ulrich, D. (2015) ‘Learning from practice: how HR analytics avoids being a management fad’, *Organizational Dynamics*, 44(3), pp. 236–242. Available at: https://doi.org/10.1016/j.orgdyn.2015.05.008.

Ryan, L. (2016) *Ten Reasons Successful People Change Jobs More Often*, *Forbes*. Available at: https://www.forbes.com/sites/lizryan/2016/10/28/ten-reasons-successful-people-change-jobs-more-often/ (Accessed: 16 May 2023).

Saunders, M., Lewis, P. and Thornhill, A. (2012) *Research Methods for Business Students*. 6th edn. London: Pearson.

*scikit-learn: machine learning in Python — scikit-learn 1.3.0 documentation* (no date). Available at: https://scikit-learn.org/stable/index.html (Accessed: 29 August 2023).

*SciPy* (no date). Available at: https://scipy.org/ (Accessed: 29 August 2023).

*sklearn.ensemble.RandomForestClassifier* (no date) *scikit-learn*. Available at: https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (Accessed: 31 August 2023).

*sklearn.preprocessing.OneHotEncoder* (no date) *scikit-learn*. Available at: https://scikit-learn/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html (Accessed: 29 August 2023).

Tambe, P., Cappelli, P. and Yakubovich, V. (2019) ‘Artificial Intelligence in Human Resources Management: Challenges and a Path Forward’, *California Management Review*, 61(4), pp. 15–42. Available at: https://doi.org/10.1177/0008125619867910.

Team, K. (no date) *Keras documentation: About Keras*. Available at: https://keras.io/about/ (Accessed: 17 September 2023).

*The LMS Guidebook : Learning Management Systems Demystified* (2018). Available at: https://eds.p.ebscohost.com/eds/ebookviewer/ebook/ZTAyMG13d19fMTY0OTQyMl9fQU41?sid=ab4dacb2-3bac-43a1-ac3f-4d6851c29243%40redis&vid=4&format=EK&rid=1 (Accessed: 14 May 2023).

Topno, H. (2012) ‘Evaluation of Training and Development: An Analysis of Various Models’, *IOSR Journal of Business and Management*, 5(2), pp. 16–22. Available at: https://doi.org/10.9790/487X-0521622.

Tsui, A.S. and Gutek, B.A. (1999) *Demographic Differences in Organizations: Current Research and Future Directions*. Lexington Books.

Vowels, M.J., Cihan Camgoz, N. and Bowden, R. (2023) ‘D’ya Like DAGs? A Survey on Structure Learning and Causal Discovery’, *ACM Computing Surveys*, 55(4), pp. 1–36. Available at: https://doi.org/10.1145/3527154.

Wang, X., Guo, B. and Shen, Y. (2022) ‘Predicting the At-Risk Online Students Based on the Click Data Distribution Characteristics’, *Scientific Programming*, pp. 1–12. Available at: https://doi.org/10.1155/2022/9938260.

Wilson, C. (2013) *Interview Techniques for UX Practitioners : A User-Centered Design Method*. Available at: https://eds.p.ebscohost.com/eds/ebookviewer/ebook/ZTI1MHh3d19fNTE2MjAwX19BTg2?sid=70cde69d-863a-4c78-a479-4619b4cbc1e5%40redis&vid=17&format=EB&rid=1 (Accessed: 13 May 2023).

Zhang, J. *et al.* (2023) ‘E-Learning Behavior Categories and Influencing Factors of STEM Courses: A Case Study of the Open University Learning Analysis Dataset (OULAD)’, *Sustainability*, 15(10), p. 8235. Available at: https://doi.org/10.3390/su15108235.