Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

# Workforce Planning using Markov Chain

Garvit Vijai

August 19, 2022

Supervisor: Dr Bahman Honari

A dissertation submitted in partial fulfilment
of the requirements for the degree of
M.Sc. Computer Science (Data Science)

# Declaration

I hereby declare that this dissertation is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write.

Signed: ______________________ Date: ______________________

# Abstract

Workforce planning is how an organisation analyses its workforce and determines the actions that must be taken to prepare for future staffing requirements. Workforce planning is future-focused, helps human resources offices make good hiring decisions, increase employees' productivity, increase profit for an organisation, and identifies and saves any extra costs. If not done correctly, workforce planning can result in bad hiring decisions, high employee turnover, low productivity, and surplus cost to the company. Our research uses the Markov chain model to forecast probability distributions for different cohorts in the given organisation. We then use these probability distributions to find the corresponding employee headcounts. Markov chain is a stochastic random process based on the transitions between states based on transition probabilities. The dataset used in our analysis is a human resource dataset consisting of employee records in an organisation. The dataset fields are generic, so the proposed solution applies to the majority of the human resource datasets belonging to various organisations. Our analysis considers only the outflows but is flexible enough to incorporate the inflows. The results reflect the workforce distribution in an organisation for the next ten years, starting from 2018, based on department, age groups and gender. We found that the rate of resignation and layoffs are increasing in all the cases and hence require attention to draft relevant policies that address current shortcomings. The research output of this study could help HR departments form their respective organisations' hiring, training, retention and retirement policies.

***Keywords: Workforce Planning, Markov Chain, Shift Probability, Transition Probability Matrix, Transition Probability***

# Acknowledgements

I want to express my profound gratitude to my supervisor Dr Bahman Honari for his impeccable guidance throughout my dissertation. His advice and constructive feedback from time to time were beneficial in shaping my research. Without his assistance, I would not have finished my dissertation smoothly.

I am grateful to this esteemed institution, Trinity College Dublin, for providing a significant opportunity to pursue my master's degree. I am thankful to the Trinity College School of Computer Science and Statistics professors and members for their unlimited support during my master's programme. I also want to thank Trinity College library for providing resources and access to the scientific articles and required literature.

Lastly, I thank my friends and family for their constant support. Without them, this dissertation would not have been possible.

# Contents

# List of Figures

# List of Tables

# 1 Introduction

## 1.1 Overview

Workforce planning is analysing an organisation's current workforce, determining future workforce needs, identifying the gap between the present and the future, and enforcing solutions to ensure that its purpose, goals, and strategic plans are met. To accomplish an organisation's short and long-term goals, hiring the correct number of people with the right talents at the appropriate time, at a reasonable cost, and under the relevant contract is essential. Generally, large organisations have dedicated workforce planning teams. Other organisations may initiate the process in response to a specific event, such as a merger, acquisition, or a transformative change initiative. Comprehensive workforce planning can identify hurdles or unrealistic goals that may stymie strategic plans and propose solutions to manage risks to support future strategic objectives[1].


Figure 1.1: Steps in Workforce Planning
**Source:** Study on Workforce Planning - CIPD, Ireland

Managing demand and supply, identifying gaps in skills and studying employment trends are some of the dimensions of workforce planning, and the area of study depends on the end goal. A well-structured workforce planning can benefit organisations, individuals and the government directly or indirectly. A good workforce planning strategy is to construct a long-term plan to guarantee that the best talent is at the right place and to have a better knowledge of the type of future workforce that will be required. The HR leaders can then feed the roadmap into resourcing plans executed locally by line managers [2].

In this dissertation, we implemented the Markov chain in workforce planning. We also studied the advantages and disadvantages of the alternate forecasting models. As a result, we discussed the year-wise trend of various employee outflows for different cohorts for ten years. Moreover, we also learned how the Markov chain model could forecast the employee head-counts across different cohorts for the given HR data. At last, we discussed the negative consequences of the trends and possible solutions to ameliorate them.

## 1.2 Motivation

The COVID-19 pandemic has changed the world of work, especially in the IT sector. As per an Irish Times report, about half of the younger workforce plan to leave their current jobs in the next two years in search of a better work-life balance as part of the global "great resignation". 75% of generation Z and 77% of millennial employees desire hybrid or entirely remote work, while fewer than half presently have the choice. The pandemic has also caused a colossal staff shortage across different sectors, and therefore there is a dire need to implement workforce planning for organisations of all sizes. According to a report from ESRI, Hospitals in Ireland may need up to 15,000 more staff within 13 years, and the institute emphasises setting up various measures for workforce planning. Another report from BBC UK states that National Health Service (NHS) is facing the worst staffing crisis in history, and urgent actions are required to resolve this problem.

Since it is difficult to predict the future, organisations often find workforce planning challenging to implement. Some organisations implement improper workforce planning, which leads to severe consequences. Poor workforce planning could result in improperly skilled employees [3]. Effective workforce planning forecasts the training required to enhance the competencies of existing employees now and in the future, as well as what new talents the firm may obtain through hiring. Poor workforce planning fails to anticipate the skills and knowledge employees will require in the future, resulting in existing skills becoming quickly obsolete and providing declining value to the firm over time. Employees cannot operate wisely or efficiently without a well-thought-out workplace strategy that predicts staff training requirements.

It could lower the productivity of employees [3]. High productive employees can help a firm accomplish its growth objectives faster and more efficiently. On the other hand, poor workforce planning might result in a lack of attention to technology, training, and other resources that can increase staff productivity. Workforce planning ensures that teams and individuals are working on the right things. As per a study, an organisation increased its productivity by 6% after implementing a workforce plan, reorganising key responsibilities (KRAs) across different verticals [4].

Poor workforce planning could lead to reduced collaboration, and teamwork [3]. Organisation development and innovation go hand in hand, but a company will struggle to innovate if its teams are fragmented and not organised in a way that promotes good communication and collaboration. Workforce planning must create an organisational structure in which teams are appropriately aligned to communicate, share information and ideas, and work effectively. Working without a proper personnel strategy might result in groups that function in isolation and are so internally oriented that they inhibit innovation and organisational progress.

Poor workforce planning could also increase the expenditure [3]. Improper workforce planning can lead to bad hiring decisions resulting in increased expenses and low-grade productivity. Failure to appropriately prepare for and manage turnover further stifles corporate growth since replacing an employee takes time and money. In such cases, the HR department and managers spend more time and resources devising workarounds to fill talent gaps, leaving them less time to focus on activities that promote growth, such as training and developing existing employees.

Such problems can be addressed through a well-structured workforce planning process. Many studies are available on workforce planning, but a holistic study accounting for the staffing problems, consequences, and their solutions is rarely available. Exploring the staffing-related problems in an organisation, their impact, and determining possible resolutions to them is the main driving factor for this dissertation. We used the Markov chain model to carry out our study. The main advantage of the Markov chain is its simplicity and out-of-sample forecasting accuracy. The entire system is represented as states and transitions among the states. It is mathematically less rigorous than other models discussed in the next chapter. Another advantage of using the Markov chain model is its flexibility. It can handle both employee migration and employee retention situations. Another advantage of using the Markov chain model is that it makes predictions without looking too far into the past, which means we require less historical data to make predictions [5].

## 1.3 Thesis Objectives

The aim of this thesis is to perform a Markov chain analysis to forecast the year-wise trend of employees resigning, getting laid off, retiring and being promoted in an organisation for the next ten years. We will analyse these outflows on the basis of department, age group and gender. Based on the results, we will discuss the possible repercussions and their solutions.

While there are different dimensions of workforce planning, we are focusing only on the staffing need of an organisation. The outcome of this thesis could be helpful to the human resources department form their hiring, training, retention, and retirement policy diligently. Moreover, this thesis could also act as a groundwork for future research in the field of Workforce Planning.

## 1.4 Structure of Thesis

The outline of this thesis is as follows:

- Chapter 1 states the overview, motivation and objectives of this thesis.
- Chapter 2 discuss the literature review related to workforce planning, Markov chain and other forecasting models used in workforce planning.
- Chapter 3 discuss the dataset, theory involved in the Markov chain analysis and the steps involved in building our Markov chain model.
- Chapter 4 discuss the implementation of the strategy discussed in Chapter 3 to create our Markov chain model.
- Chapter 5 presents the predictions made by the model and the findings obtained from them.
- Chapter 6 concludes the analysis and further discusses the limitations of this study.
- Chapter 7 outlines the issues that could be addressed in future work.

# 2 Background and Related Work

## 2.1 Workforce Planning

The concept of workforce planning gained popularity in the late 1960s and early 1970s when economic stability increased, and unemployment was low. Organisations recognised the need for efficient labour utilisation and implemented various strategies. However, the 1980s recession caused many Human resource departments to believe that workforce planning was a failure, which led to the elimination of several initiatives. This failure occurred because the objectives were too narrow and hence misjudged[6].

However, most recent definitions of workforce planning contain more flexible target ranges, resulting in a more contextual understanding. This includes three variations of workforce planning methods: judgemental, statistical, or a combination of the two [7]. The judgmental methods use the skills and experience of humans to forecast an organisation's future demands. Depending on the input, prediction might be highly accurate or utterly incorrect. Judgmental procedures are helpful for short-term analysis in order to provide the flexibility required by the organisation[8]. The statistical methods provide forecasts based on previous data and have an accurate starting point. These methods are more suitable for long-term analysis. A good approach to developing a workforce model can be to use a combination of both techniques, i.e. the use of human judgement along with data-driven analysis[9].

With technological advancement, workforce planning capabilities have improved. Young and Hollander [2] talk about some of the recent advancements in workforce planning. Human resource management software can collect internal employee data based on job category, position, and skills, allowing companies insights beyond headcount segmentation (e.g., by tenure, level, location). This increased data availability can help explore new dimensions of working planning. Applications that integrate workforce data with enterprise data can facilitate cross-functional work, communications, and understanding. Artificial Intelligence and Machine Learning-based applications can deliver more detailed and accurate insights, which can help to fill gaps more efficiently. Since companies now may have more information than ever, they can make wiser decisions.

Three stakeholders that primarily get benefit directly or indirectly from properly implementing workforce planning are organisations, individuals and government[2]. Efficient workforce planning helps organisations easily manage the workforce's demand and supply. Increased availability of external workforce data, along with enhanced procedures and internal apps, enable organisations to update personnel data more regularly. This enables organisations to detect fluctuations in the internal workforce and external demand and supply. Going forward, based on employee profile, AI/ML-based solutions can predict the employee's expertise, gaps between demand and supply and steps required to address such gaps. Improved work-life balance, effective retention policies, reduced labour costs, relevant strategies for skill development, and improved productivity and output are some of the workforce planning benefits for organisations [2].

Managers can use workforce planning to manage their teams in several ways. They get a deeper insight into work and skill supply-and-demand patterns and how those patterns may affect their teams. They can get customised team insights such as gaps in team skills and ML-based recommendations on overcoming anticipated shortcomings. They can also have better visibility of the expertise of employees outside their team whom they could include in their team to fill a need [2]. Individuals such as applicants, employees or contingent workers will better understand where the employment market is heading, allowing them to better prepare for the future. Employees may get personalised recommendations based on their interests, career goals, or the requirement of their current team [2]. Workforce planning can provide detailed insights to government agencies about the industry, and employment trends, allowing them to manage supply and demand imbalances. Access to online professional profiles and advertisements that will provide them with a more comprehensive real-time snapshot of labour supply and demand can be obtained through surveys. They can facilitate workforce planning through national, state, and regional programs [2].

## 2.2 Forecasting Models for Workforce Planning

Workforce planning can be done using different prediction methodologies by investigating their logic, strengths, and limitations. The workforce models can be divided into three categories: supply-based, demand-based and need-based. Supply-based models aim to project the number of employees available at a given time based on training, retirement, and promotion or demotion. Demand-based models aim to forecast future service requirements and anticipated changes in demand. Need-based model predictions employ an exogenous benchmark to assess the sufficiency of the number of employees necessary to accomplish the objectives. It addresses the question, "How many resources do we need?" Depending on the goals, a workforce model can be constructed using one or more of these categories [10].

Qualitative methods are the simplest model that can be used in workforce planning. These approaches are efficient when high-quality experimental data is unavailable. They mainly depend on the expertise of engaged professionals. A discrete-choice experiment is an approach for eliciting individual preferences and transforming qualitative data into meaningful computing data. This strategy allows researchers to discover how individuals perceive specific traits by asking them to choose between various hypothetical situations. Although discrete-choice experiment methods are beneficial in hypothetical scenarios, they are complex to implement and need experience in multiple fields [11]. Another well-known consensus-based technique is the Delphi method, in which a panel of experts answers questions across multiple iterations. Following each iteration, an anonymous summary of the findings is processed and presented as the starting point for the following iteration, in which they may alter their previous judgements in light of the responses of the other experts. The idea is to arrive at a consensus, hopefully correct, solution [12]. The Delphi technique has been used to find the unknown variables of models in several workforce modelling projects, or it has been used hierarchically in conjunction with other modelling methodologies. The critics of the qualitative modelling approach point to its reliance on expert judgments, which may result in inaccurate estimates. Another disadvantage of qualitative modelling is that it can be a time-consuming procedure [13]. That being said, qualitative modelling is still used in some research problems.

Optimisation models are another category of modelling techniques that can implement workforce planning. This technique minimises and maximises an objective function, subject to different constraints. The most commonly used optimisation methodologies are Linear programming, Goal programming, Mixed integer linear programming, and Dynamic programming [10]. Linear programming is a popular approach used in solving workforce optimisation problems. It aims to find the best value of a linear function given linear equality or inequality constraints. Srour and Mounir [14] developed a Linear programming-based model for managing construction workers' demand and supply while considering training, hiring, and allocation. Goal programming provides optimisation for multiple objectives. Niehaus [15] used Goal programming to simulate the demand and supply of a US Navy shipyard community experiencing widespread staffing cutbacks. Mixed integer linear programming finds a linear objective function and a combination of linear and discrete integer constraints. Júdice et al. [16] used a Mixed integer linear programming-based model to minimise the costs associated with the workforce size in a mail-processing centre. Dynamic programming is a multistage optimisation approach in which an enormous task is broken into smaller hierarchies. Wild and Schneewei [17] employed a dynamic programming technique to optimise the concurrent usage of floaters, overtime, and temporary workers in a German service organisation. Most of the above-discussed researches attempted optimal solutions based on linear models of the systems, diverging from the nonlinear dynamic systems that exist in reality [18]. However, nonlinear optimisation numerical solutions might be computationally demanding and challenging to comprehend.

Stock and flow is another popular modelling technique that can be used for workforce planning. It perceives workforce cohorts as stocks and transitions such as promotion, job rotation, and recruitments as flows. One advantage of stock and flow modelling is that it provides a systematic perspective of the entire workforce system that policymakers can readily grasp. This model also integrates well with the concept of time by representing changes in the system through time [10]. Wilson [19] applied the stock and flow model to construct a workforce planning model to address the demand and supply of the workforce. His model considers factors such as technological developments, governmental laws, and changes in social norms. The high dependence of the stock and flow model on computerised analytics is considered their major disadvantage. Nevertheless, stock and flow are still the most widely used modelling techniques.

Time series is another prominent approach for projecting future values based on historical data. Box–Jenkins, exponential smoothing and vector error correction are popular time series models used for workforce planning. Box-Jenkins use an autoregressive moving average to determine the best fit of a model for historical time series data [20]. Exponential smoothing is a weighted moving average(WMA) approach in which prior observations are allocated exponentially decreasing weights across a specific period. This time series modelling technique has been used to determine the workforce development in Taiwan's energy sector from 2009 to 2014 [21]. The Error correction models belong to the multiple time series model category and are fed by their present state and long-term dynamics dispersion. A major disadvantage of the time series models is that these are limited by the assumption that the recent trends will persist and overlook factors other than time. The requirement of too much historical data as input is another shortcoming of the time series model.

Statistics and regression are widely used for forecasting purposes. A statistical model uses uncertain parameters to forecast future events. A fitting procedure approximates the unknown parameters based on historical data. Variance, expected values, and other statistical methods are used to study the properties and different values obtained from the model [22]. Regression analysis is a statistical modelling method that analyses the influence of independent variables on dependent variables and is a helpful tool for examining variable interdependency. There are certain disadvantages associated with statistical modelling. Even though they are effective tools for describing the average behaviour of processes, they can only provide a single observation and a brief assessment of a workforce system. Hence, they cannot simulate highly dynamic systems(for example - a workforce system with high recruitment and attrition over time). Since regression models typically examine just one outcome, they are unsuitable for modelling systems with several outputs [10].

## 2.3 Markov Chain

Russian mathematician Andrey Markov defined the Markov chain in 1907. He defined a class of processes in which the likelihood of phenomena existing in a particular condition at a given moment is connected to that phenomenon's immediately preceding state. A Markov chain may be considered as a sequence of transitions between distinct states. The probabilities associated with each transition are determined by how the process got to that state. A Markov chain consists of a finite number of states, and the probabilities associated with state transitions do not alter over time [23].

### 2.3.1 Application of Markov Chain in Other Sectors

A vast amount of literature that deals with the application of the Markov chain exists. Markov chain can be applied for constructing forecasting models for resources across different sectors such as Education, Environment, Supply chain and Workforce planning. Nyandwaki and Kennedy [24] used the Markov model to study students' enrolment projection in both secondary and tertiary institutions. Their model consisted of three states in which a student can be in after an academic session: a student can move to a higher class, may repeat the same class or may, leave the institution as a successful graduate or drop out without attaining the maximum qualification. In another paper, Osagiede and Ekhosuehi [25] used the Markov model to forecast university students' intake. They discarded the previous authors' assumption of specific constant values for computing the intake rate and presented a better approach for determining the constant values. Ogbogbo et al. [26] used a single state absorbing Markov chain to define the proportion of staff recruited, promoted and withdrawn from various grades over the years at the polytechnic institution in Nigeria and also forecast the expected workforce structure of the institution.

Shugart Jr et al. [27] used the Markov process to model forest succession over large regions. Peden et al. [28], Cassell and Moser [29] also dealt with problems arising in forestry. Rao and Kshirsagar [30] investigated the dynamics of predator/prey systems, assuming that a predator's attack cycle consists of four distinct activities: search, chase, handling, and feed and digest. Similar Markov chain approaches are discussed in the field of geology and stratigraphy in [31][32]. Buongiorno and Michie [33] developed an interesting Markov model for selecting a forest. Their model predicts the long-term growth of undistributed and managed strands. The stochastic transition of trees between diameter classes and the recruitment of new trees into the strand are considered parameters of their model and taken from the North-Central region of the USA. Following that, a linear programming approach was utilized to predict the long-term usage of management regimes that would optimize the net present value of the periodic harvest. The approach calculates optimum harvests, residual stock, diameter distribution, and cutting cycles simultaneously. This forestry application of the Markov chain is similar to the Markov chain implementation of workforce planning.

A significant amount of work based on the Markov chain can be seen in the field of the supply chain. When establishing a supply chain, firms have fixed assets for each stage. As a result, the supply chain cannot serve customers with a limitless number of items or services. Because the supply chain market opportunity analysis will directly influence who the core business should pick as a partner and what sort of supply chain manufacturing processes they should design, it is necessary to estimate the market prospects [34]. In [34], the model forecasts the market opportunities before the formation of the supply chain, providing decision assistance for the pre-assessment of market potential. Their analysis shows that just because a product's current sales rate is high doesn't guarantee that it will continue to be such because it also depends on the consumers' preference (shift probability). In [35], the authors have focused on the closed-loop supply chain, which is less complex and uncertain than the traditional supply chain. The overall stability of the closed-loop supply chain is evaluated based on the transition matrix and the state probability distribution. In [36], authors have used the Markov model for inventory management which is a part of supply chain management. Their research aims to develop a discrete-time Markov model for an order-up-to-level inventory strategy with stochastic demand and lead time. Their model calculates the expected inventory on hand and the predicted shortage when unmet demand is believed to be lost. The above-discussed applications of the Markov chain model, be it in the field of education, environment or supply chain, have similar approaches to what we are looking for in workforce planning.

### 2.3.2 Implementation of Markov Chain in Workforce Planning

Numerous past studies in organizational management have used the Markov chain to describe promotions, demotions, or changes in different career development routes in order to validate an organization's current staffing demands or anticipate future staffing needs. The Markov chain model enables us to respond to policymakers' inquiries. For instance, it makes it simple to compute numerous statistics at individual and aggregate levels. At the individual level, it may be used to characterize the probability of advancement of a staff member at a specific career stage. At the aggregate level, it may extract information on overall continuation rates and separation behaviour, which are crucial inputs in building retention initiatives [37].

Collings and Wood [38] described workforce planning as "the set of ideas, tools, and processes that each company should use to monitor and control the mobility of employees in terms of both numbers and profiles." These employee moves, known as transitions, are frequently the result of promotions, transfers between segments, and recruitment into the system. As per [37], The planning of the process is one of the most important and complicated responsibilities, encompassing organizational development, managerial development, career planning, and succession planning. Workforce planning can be done in two phases: The first step involves careful planning of personnel requirements for all types and levels of employees during the planning period. The second stage focuses on workforce supply planning to provide the organization with the appropriate personnel to satisfy the planning criteria. Igboanugo and Onifade [39] proposed a Markov chain model to unravel the dynamics of workforce availability and supply in Nigerian universities to describe the existing workforce policy and point to its future direction. Parma et al. [40] proposed a better Markov model for managing workforce requirements and availability. In their system, vacancies are filled through promotions and recruitment. They proposed methods for determining transition probability for promotion and recruitment matrix using Markov chain theory and certain assumptions.

Effective workforce planning is critical for organizations such as big corporations, academic institutions, and administrative departments to implement because human resources are regarded as the most vital and possibly unpredictable resource an organization employs. For an established process, workforce planning necessitates a good understanding of those deployed in the establishment, along with employee entry, dropout, and promotion, to form a good strategy. Touama [41] used Markovian models and a transition probability matrix to examine worker migration in Jordan productive firms. To accomplish his goal, he gathered secondary data on employee migration from annual reports of Jordanian productivity firms: potash, phosphate and pharmaceutical. Kwon et al. [42] utilized a Markov chain model along with job coefficient to study the difference in workforce status between the US and Korean nuclear industries and to forecast future workforce requirements in Korea.

In [43], they developed an interesting model for academic and organizational hierarchical systems using an absorbing Markov chain in which they divided the wastages into retired and dropouts. Their model represents retired staff and dropouts as two separate absorbing states. This is in contrast to the models designed by Igboanugo and Edokpia [23], Ekhosuehi et al. [44], Ogbogbo et al. [26], Ezugwu and Ologun [37] where retired staff and dropouts were grouped together and no significant work has been done on separating them. De Feyter [45] presented an interesting time-dependent workforce planning model, which divides the system's population into multiple subgroups based on factors such as sex and number of children of individuals. Dimitriou and Tsantas [46] also addressed workforce planning by dividing the entire heterogeneous population into several homogeneous subgroups. Such an approach to forming subgroups provides an opportunity for a detailed investigation. It simplifies the prediction of workforce transformation since it becomes acceptable that everyone in the same group progresses identically.

Du and Li [47] used the Gray Markov model to forecast internal supply in the HR department so that organisations can reasonably predict their internal human resource supply and develop human resources strategies. [46] discussed the asymptotic behaviour of a generalised time-dependent Markov model. This approach, on the one hand, considers employees' need for continuous training, and on the other, its workforce requirement is met by internal mobility and two streams of recruitment, the first from the external environment and the second from a supplemental auxiliary system. Nilakantan and Raghavendra [48] proposed a "proportionality" strategy for a workforce planning system, in which recruitment to each level of the hierarchy (save the lowest level) is strictly proportionate to promotions into that level.

In [49], the author talks about challenges such as lack of sensitivity to problems, organisational complexity, rapid changes and inefficient communication that impede the successful implementation of workforce planning in organisations. It is anticipated that social media if appropriately explored, can improve communication and make important information available for better personnel planning in enterprises. The paper also examines several empirical studies on the workforce planning issue and finds that organisations should guarantee that they have the right number of people, in the right location, at the right time, and doing the right tasks by using the proper models for workforce planning.

# 3 Methodology

This chapter discusses the concepts used to complete the dissertation, such as Dataset and related preprocessing, Random variable, Stochastic process and Markov chain theory, in detail.

## 3.1 Dataset

Due to GDPR policies, very few public human resource datasets are available, which are insufficient to perform our analysis. The dataset used for our study is a fake dataset available on Kaggle [50]. It consists of the information of all the employees over the past ten years. The dataset contains information such as Hire Date, Termination Date, Department, Performance Score, Termination reason, Date of birth and Employment Status, based on which we projected the workforce count across different departments, age groups and gender. The further steps, such as pre-processing data, construction of transition matrices and the Markov chain model, are discussed further in the implementation chapter.

We have data only for outflows in the dataset. So, this dissertation only focuses on workforce planning when outflows are known. Based on the dataset, different outflows are Resignation, Retirement, Layoffs and Promotion. These factors are relevant to all employee databases across all kinds of industries. Therefore, to a great extent, the proposed model can be used by the Human Resources department at multiple organisations.

## 3.2 Data Preprocessing

Data Preprocessing refers to manipulating the data to make raw data suitable for further processing. It is the first and the crucial step in the data analysis process since the quality of the discovered metrics is heavily determined by the data quality [51]. We have used python libraries such as Pandas, NumPy, Seaborn and Matplotlib along with SQL and MS-Excel for the pre-processing purpose. For our analysis, data pre-processing is mainly done in three phases:

### 3.2.1 Data Collection:

This phase aims at finding a high-quality dataset since it directly affects the quality of the results. Some of the challenges that can be faced in this phase are [52]:

1. **Incorrect Data:** The dataset can contain information that is not relevant to the problem statement.
2. **Missing Data:** Some of the columns can have empty values, resulting in improper analysis.
3. **Data Imbalance:** Some of the columns in the dataset may have a disproportionately high or low number of samples. This may result in a risk of getting underrepresented in the analysis.

### 3.2.2 Data Cleaning

This phase involves detecting and correcting(or removing) incorrect values in the dataset, degrading the data quality. The real-world raw data could be incomplete, inconsistent and contain many errors. So, after collection, data should be cleaned properly before developing the model on top of it. In our analysis, we removed the records for which most of the necessary fields are missing, which means those records are unsuitable for our analysis. We also replaced missing values with "NA" for the remaining records to complete the dataset. These steps helped in efficiently filtering out the required data from the dataset.

### 3.2.3 Data Reduction

This step helps reduce the data volume, making analysis easier and producing the same results. In our study, we dropped some columns such as marital status, marital description, and race description, which are not required. As discussed earlier in the introduction, the Markov chain makes predictions without looking too far in the past. Due to this property of the Markov chain, we are only interested in employees working in 2018, the latest year in the dataset, or terminated in 2018. Therefore, we eliminated all the employee data before the year 2018. This significantly reduced the input data required for developing our Markov chain model. Moreover, this reduction in data made our analysis work simpler and improved the efficiency of the SQL queries used for finding employee count across different criteria.

### 3.2.4 Data Transformation

This phase involves making changes to the format or structure of the dataset. We added an age column in our analysis to make our SQL age-based queries efficient. Employees' ages are derived from their respective dates of birth(DOB).

## 3.3 Random Variable

A random variable X is defined as the variable whose value is defined as the outcome of a random experiment. The range of a random variable X is the set of all possible values of X. Random variables are primarily of two types [53]:

1. **Discrete Random Variable:** A random variable X is discrete if its range is a countable set. A set S is said to be countable if either it is finite or its elements can be put in one-one correspondence with the natural numbers.

2. **Continuous Random Variable:** A random process X is said to be continuous if it has a range in the form of an interval or combination of non-overlapping intervals on the real line.

This dissertation focuses on using discrete random variables.

## 3.4 Stochastic Process

A stochastic or random process is the process in which the variables change at a random rate over time. Random walk, Bernoulli process, and Markov chain are some of the well-known stochastic processes [54]. The possible values of the random variables in a stochastic process are known as the **State Space** of the process.
There are two types of random processes:

1. **Discrete-Time Random Process:** A random process is said to be discrete-time if it is of form $\{ X(t), t \in J \}$, where J is a countable set such as $\mathbb{N}$ or $\mathbb{Z}$. [probability course bookmark]

2. **Continuous-Time Random Process:** A random process is said to be continuous-time if it is of form $\{ X(t), t \in J \}$, where J is an interval on a real number line such as $[0,\infty]$.

In this dissertation, we focused on the discrete-time random process.

## 3.5 Markov Chain

Markov Chains represent a class of stochastic processes that experiences transitions from one state to another according to certain probabilistic rules. It differs from a general stochastic process in that a Markov chain must be "memory-less" [54]. Markov Chain satisfies the Markov property, i.e. the probability of transitioning to any particular state depends solely on the current state and the time elapsed. A mathematical expression of the Markov property is shown below.

$$P\{X_{t+k} = j | X_0 = i_0, X_1 = i_1 .... X_t = i\}$$
$$= P\{X_{t+k} = j | X_n = i\}, k \geq 1$$

We represent a Markov chain as a sequence of states, as shown below.

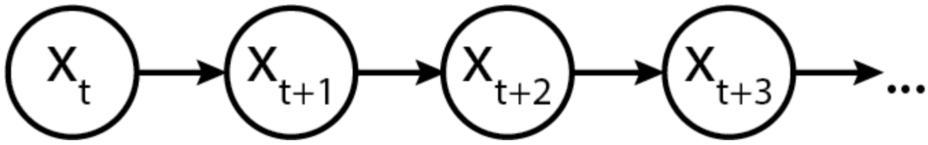


Figure 3.1: Markov chain representation

There are two types of Markov Chains [53]:

1. **Discrete-Time Markov Chain:** A discrete-time Markov chain is a discrete-time stochastic process in which the value of the following random variable is determined only by the value of the current random variable and does not depend on any previous variables.

2. **Continuous-Time Markov Chain:** A continuous-time Markov chain is a continuous-time stochastic process that changes states based on an exponential random variable and then moves to a different state based on the probabilities of a transition matrix. It is different from the discrete-time Markov chain in the way that it changes states continuously across time rather than discrete time steps.

This dissertation focuses on using the discrete-time Markov chain.

### 3.5.1 Initial Probability Distribution and Transition Probability Matrix

To define a Markov chain, we need to specify two things [55]:

1. **Initial Probability Distribution:** It is an NX1 vector representing probability distribution for N states at step 0.

$$P_{init} = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 & p_5 \end{bmatrix}$$

2. **Transition Probability Matrix:** A transition matrix is an NXN matrix that contains probabilities of transitions between N states in a system. The transition matrix row corresponds to the state at time 't' while the column represents the state at the time 't+1'. Since transitioning to the next state depends only on the current state and all

possible transitions are known, the sum of probabilities for each row equals [55]. The entries of the transition matrix represent the conditional probabilities. The conditional probability of an event A given another event B has already occurred is given by

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

Using the above equation, we obtain the transition matrix as shown below.

$$P_{transition} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix}$$

An initial probability distribution combined with the transition probability matrix defines the probability of all the events in the Markov chain [56].

### 3.5.2 K-step Transition Matrix

The shift probability reflects the probability in the state $j$ after the given process in the state $i$ at an instant $t$ has passed $k$ units of time [34].

$$P\{X_{t+k} = j | X_n = i\}, k \geq 1$$

The K-step shift probability is represented as $P_{ij}^{(k)}$. When k equals 1, it is called one-step shift probability. The relation between k-step shift probability matrix $P^{(k)}$ and one-step shift probability matrix P is given by

$$P^{(k)} = P^k$$

We used the above relation to find the K-step probability distribution discussed in the next chapter.

### 3.5.3 Stationary Distribution

The Markov chain can end up in stationary distribution, meaning the probability distribution remains unchanged as time progresses. An important property of stationary distribution is that it depends only on the transition matrix and not on the initial state vector [34]. A stationary distribution for an N state Markov chain with transition probability matrix P is a row vector $\pi$ that satisfies

$$\pi = \pi P \quad (1)$$

Suppose the probability distribution for 5 states in a steady state is $\pi = [\pi_1, \pi_3, \pi_3, \pi_4, \pi_5]$. Using transition matrix $P_{transition}$ discussed in section 3.5.1 and the steady state equation mentioned above, we get

$$[\pi_1, \pi_2, \pi_3, \pi_4, \pi_5] = [\pi_1, \pi_2, \pi_3, \pi_4, \pi_5] \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix}$$

$$\pi_1 p_{11} + \pi_2 p_{21} + \pi_3 p_{31} + \pi_4 p_{41} + \pi_5 p_{51} = \pi_1 \tag{2}$$

$$\pi_1 p_{12} + \pi_2 p_{22} + \pi_3 p_{32} + \pi_4 p_{42} + \pi_5 p_{52} = \pi_2 \tag{3}$$

$$\pi_1 p_{13} + \pi_2 p_{23} + \pi_3 p_{33} + \pi_4 p_{43} + \pi_5 p_{53} = \pi_3 \tag{4}$$

$$\pi_1 p_{14} + \pi_2 p_{24} + \pi_3 p_{34} + \pi_4 p_{44} + \pi_5 p_{54} = \pi_4 \tag{5}$$

$$\pi_1 p_{15} + \pi_2 p_{25} + \pi_3 p_{35} + \pi_4 p_{45} + \pi_5 p_{55} = \pi_5 \tag{6}$$

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1 \tag{7}$$

Solving eq. 2,3,4,5,6 & 7, we get steady state distribution, $\pi$.

## 3.6 State Space of the System

Based on the dataset, we defined five states where an employee can be present or transition: Working, Resign, Layoff, Retire and Promote.

1. **Working:** An employee can belong to a *working* state if he continues working at an organisation without getting promoted. From our dataset, we calculated working employees as

   Working Employees = Total Active Employees - Promoted Employees

2. **Resign:** The *resign* state is derived based on factors such as employee satisfaction score and termination reason. In our analysis, we have assumed that an employee can be in resign state if the employee has a satisfaction score less than equal to 3 or if the termination reasons are one amongst: Another position, career change, hours, maternity leave - did not return, medical issues, more money, relocation out of the area, return to school, unhappy.

3. **Layoff:** The *layoff* state is derived based on employee performance score. In our analysis, we have assumed that an employee can be in a layoff state if the employee has a performance score less than equal to 2.

4. **Retire:** An employee can be in *retire* state if the employee is retiring from the service. Generally, employees aged more than 50 years belong to this category.

5. **Promote:** The *promote* state is derived based on factors such as employee performance score. In our analysis, we have assumed that an employee can be in a promote state if the employee has a performance score value of 'Exceeds'(Performance score id > 3).

We define the state diagram for our Markov chain based on the states discussed above. The probability of transition from one state to another state depends on the transition matrix defined for a particular group.

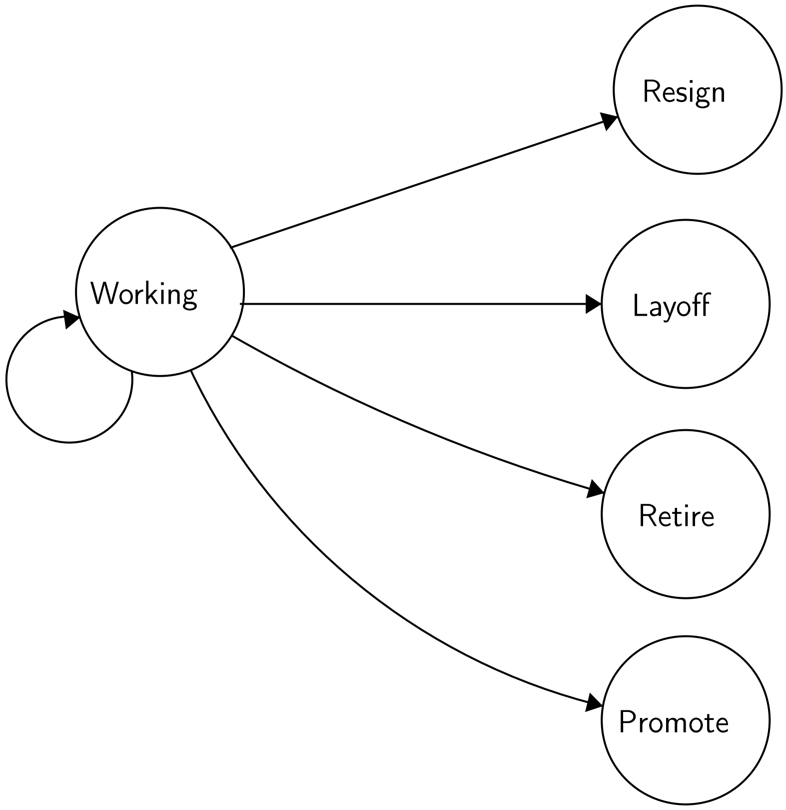


Figure 3.2: State Transition representation of the system.

The probability of transitioning from state Resign, Layoff, Retire and Promote to any other state is zero; hence, they are known as terminal states. The transition matrix of the system based on the above state diagram is shown below.

$$\begin{array}{c} \\ Work \\ Resign \\ Layoff \\ Retire \\ Promote \end{array} \begin{array}{c} \begin{array}{ccccc} Work & Resign & Layoff & Retire & Promote \end{array} \\ \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} \end{array}$$

To initialize the Markov chain, we define the initial probability vector as follows.

$$P_{init} = [1, 0, 0, 0, 0]$$

Initially, we assigned random probability for all the five states: 1 for working and 0 for Resign, Layoff, Retire and Promote states. The implementation of the above-mentioned initial probability matrix and transition probability to construct the Markov chain is discussed in the next chapter.

# 4 Implementation

We have seen earlier that the Markov chain is a sequence of transitions between distinct states in which the probabilities associated with each transition are solely determined by how the process got to that state. Keeping this idea of the Markov chain in mind, we considered employee records for the latest year, which turned out to be 2018.

After preprocessing the data, we dumped the dataset in the SQL database and ran SQL queries for different cohorts. The employee distribution across different states for the year 2018 is shown below.

| State | Number of Employees |
|---|---|
| Working | 1280 |
| Resign | 366 |
| Layoff | 239 |
| Retire | 46 |
| Promotion | 247 |
| **Total Employees** | **2178** |

Table 4.1: Employee distribution for the year 2018.

## 4.1 Based on Department

As discussed earlier, workforce planning helps organisations and their departments manage the demand and supply of the workforce. Department-level forecasting can help departments manage positions, skills, tenure, and budgetary more efficiently. The dataset consists of six departments: Admin Office(DeptID=1), Executive Office(DeptID=2), IT/IS(DeptID=3), Software Engineering(DeptID=4), Production(DeptID=5) and Sales department(DeptID=6).

### 4.1.1 Admin Office

This department deals with administrative work of an organisation such as budgetary, legal affairs, call for tenders, facilities and security. The employee distribution for the admin department is shown below.

| State | Number of Employees |
|---|---|
| Working | 63 |
| Resign | 7 |
| Layoff | 5 |
| Retire | 1 |
| Promotion | 0 |
| **Total Employees** | **76** |

Table 4.2: Number of employees in Admin department

### 4.1.2 Executive Office

This department includes key decision makers such as the CEO, CTO, CFO, President, Directors and Head of Departments. The employee distribution for the executive department is shown below.

| State | Number of Employees |
|---|---|
| Working | 20 |
| Resign | 4 |
| Layoff | 0 |
| Retire | 0 |
| Promotion | 0 |
| **Total Employees** | **24** |

Table 4.3: Number of employees in Executive department

### 4.1.3 IT/IS

This department deals with all the IT services and related security issues in the organisation. This department's positions include Network Engineer, Data Administrator and IT Support Engineer. The employee distribution for the IT/IS department is shown below.

| State | Number of Employees |
|---|---|
| Working | 272 |
| Resign | 60 |
| Layoff | 30 |
| Retire | 9 |
| Promotion | 39 |
| **Total Employees** | **410** |

Table 4.4: Number of employees in IT/IS department

### 4.1.4 Software Engineering

This department deals with the development, testing, deployment and maintenance of the software. It consists of many roles such as Software Engineer, Quality Assurance Engineer and DevOps Engineer. We got the following values for this department.

| State | Number of Employees |
|---|---|
| Working | 55 |
| Resign | 10 |
| Layoff | 8 |
| Retire | 0 |
| Promotion | 3 |
| **Total Employees** | **76** |

Table 4.5: Number of employees in Software Engineering department

### 4.1.5 Production

A production department is responsible for manufacturing goods. It can contain a few specialised functions with all other labour outsourced or a fully functional department that processes raw materials, assembles components into completed things, and packages them. Some of the positions in our dataset under this department are Production Technician and Production Manager. The employee distribution for the production department is shown below.

| State | Number of Employees |
|---|---|
| Working | 670 |
| Resign | 258 |
| Layoff | 180 |
| Retire | 31 |
| Promotion | 172 |
| **Total Employees** | **1311** |

Table 4.6: Number of employees in Production department

### 4.1.6 Sales

A sales department in an organisation sells products or services to customers. A sales team works collaboratively to create sales, boost profitability, and establish and maintain connections with consumers to drive repeat purchases. The sales department positions include Business Development Executive, Area Sales Manager and Inside Sales Specialist. The employee distribution for the sales department is shown below.

| State | Number of Employees |
|---|---|
| Working | 200 |
| Resign | 27 |
| Layoff | 16 |
| Retire | 5 |
| Promotion | 33 |
| **Total Employees** | **281** |

Table 4.7: Number of employees in Sales department

## 4.2 Based on Age Group

Age is another factor based on which we are making predictions. Age-based forecasting helps to take a more transparent look at the age of the workforce now and in future and therefore helps identify groups with high average age, identify departments having a high number of retirements in the future and form retirement policy [57]. We divided our employee dataset into three age groups: 26 to 40 years, 41 to 60 years and more than 60 years. Since no employee is less than 26 years for our target year 2018, we did not create age groups for employees who are less than 26 years. The employee distribution for the three age groups is shown below.

| State | Number of Employees |
|---|---|
| Working | 535 |
| Resign | 172 |
| Layoff | 119 |
| Retire | 0 |
| Promotion | 99 |
| **Total Employees** | **1162** |

Table 4.8: Number of employees in 26-40 age group

| State | Number of Employees |
|---|---|
| Working | 515 |
| Resign | 169 |
| Layoff | 96 |
| Retire | 34 |
| Promotion | 104 |
| **Total Employees** | **918** |

Table 4.9: Number of employees in 41-60 age group

| State | Number of Employees |
|---|---|
| Working | 51 |
| Resign | 18 |
| Layoff | 10 |
| Retire | 12 |
| Promotion | 7 |
| **Total Employees** | **98** |

Table 4.10: Number of employees in 60+ age group

## 4.3 Based on Gender

Workforce planning on a gender basis ensures that there is gender equality and diversity maintained in an organisation. The hiring team has broader options and a better chance of securing top talent. Henson [58] explored that employees over the age of 50 are committed to sacrifice and devotion, those in their 40s feel that hard effort would bring them success, those in their 30s aim to manage work and family, and new workers in their 20s are tech savvy with a realistic approach. Moreover, men and women bring different perspectives, and when those perspectives come together, they produce better results. According to Kellogg School of Management research, diversified workgroups generally outperform homogeneous workgroups [59]. The employee distribution on a gender basis is shown below.

| State | Number of Employees |
|---|---|
| Working | 821 |
| Resign | 166 |
| Layoff | 105 |
| Retire | 22 |
| Promotion | 97 |
| **Total Employees** | **1211** |

Table 4.11: Number of males employees

| State | Number of Employees |
|---|---|
| Working | 459 |
| Resign | 200 |
| Layoff | 134 |
| Retire | 24 |
| Promotion | 150 |
| **Total Employees** | **967** |

Table 4.12: Number of female employees

## 4.4 Generating Transition Matrix

The employee distributions calculated above for different cohorts are used to create transition matrices. As discussed in section 3.5, a transition matrix consists of transition probabilities between states in a system.

We will consider the Admin department to see how the transition matrix is created. By using the equation of conditional probability, we calculate all possible transition probabilities.

$$
\begin{array}{ll}
\text{P(Working/Working)} = 0.8289 & \text{P(Resign/Working)} = 0.0921 \\
\text{P(Layoff/Working)} = 0.0657 & \text{P(Retire/Working)} = 0.0131 \\
\text{P(Promote/Working)} = 0 &
\end{array}
$$

From the earlier discussed state transition diagram (Fig. 3.2) of our system, we can see that Resign, Layoff and Promote are terminal states. Hence, the probability of transitioning to any other states from these states is zero.

The probabilities of transitioning from *Resign* state to other states are

$$
\begin{array}{ll}
\text{P(Working/Resign)} = 0 & \text{P(Resign/Resign)} = 1 \\
\text{P(Layoff/Resign)} = 0 & \text{P(Retire/Resign)} = 0 \\
\text{P(Promote/Resign)} = 0 &
\end{array}
$$

The probabilities of transitioning from *Layoff* state to other states are

$$
\begin{array}{ll}
\text{P(Working/Layoff)} = 0 & \text{P(Resign/Layoff)} = 0 \\
\text{P(Layoff/Layoff)} = 1 & \text{P(Retire/Layoff)} = 0 \\
\text{P(Promote/Resign)} = 0 &
\end{array}
$$

The probabilities of transitioning from *Retire* state to other states are

$$
\begin{array}{ll}
\text{P(Working/Retire)} = 0 & \text{P(Resign/Retire)} = 0 \\
\text{P(Layoff/Retire)} = 0 & \text{P(Retire/Retire)} = 1 \\
\text{P(Promote/Retire)} = 0 &
\end{array}
$$

The probabilities of transitioning from *Promote* state to other states are

$$
\begin{array}{ll}
\text{P(Working/Promote)} = 0 & \text{P(Resign/Promote)} = 0 \\
\text{P(Layoff/Promote)} = 0 & \text{P(Retire/Promote)} = 0 \\
\text{P(Promote/Promote)} = 1 &
\end{array}
$$

Based on the above-mentioned probabilities, we construct the following transition matrix.

$$
\begin{array}{c}
\\
\textit{Working} \\
\textit{Resign} \\
\textit{Layoff} \\
\textit{Retire} \\
\textit{Promote}
\end{array}
\begin{array}{c}
\begin{array}{ccccc}
\textit{Working} & \textit{Resign} & \textit{Layoff} & \textit{Retire} & \textit{Promote}
\end{array} \\
\begin{bmatrix}
0.8289 & 0.0921 & 0.0657 & 0.0131 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

The first row and column in the above matrix represents *working* state, second row and column represents *resign* state, third row and column represents *layoff* state, fourth row and column represents *retire* state, and fifth row and column represents the *promotion* state. Similarly, we construct transition matrices for all the departments, age groups and genders which are discussed in the next chapter.

## 4.5 Calculating Probability Distribution

Earlier, we discussed the initial probability vector in section 3.6

$$P_{init} = [1, 0, 0, 0, 0]$$

We also discussed the concept of the K-steps transition matrix in section 3.5. We calculate probability distribution at K$^{th}$ step as

$$P_{t+k} = P_{init} * P^{k}_{transition}$$

Using the above two equations, we get

$$P_{t+1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.829 & 0.092 & 0.066 & 0.013 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.829 & 0.092 & 0.066 & 0.013 & 0 \end{bmatrix}$$

$$P_{t+2} = P_{t+1} * P_{transiton} = \begin{bmatrix} 0.687 & 0.168 & 0.120 & 0.024 & 0 \end{bmatrix}$$

$$P_{t+3} = P_{t+1} * P_{transiton} = \begin{bmatrix} 0.570 & 0.232 & 0.165 & 0.033 & 0 \end{bmatrix}$$

$$P_{t+4} = P_{t+1} * P_{transiton} = \begin{bmatrix} 0.472 & 0.284 & 0.203 & 0.041 & 0 \end{bmatrix}$$

.

.

*and so on...*

$$P_{t+10} = P_{t+9} * P_{transiton} = \begin{bmatrix} 0.153 & 0.456 & 0.325 & 0.065 & 0 \end{bmatrix}$$

### 4.5.1 Employee Headcounts

We calculate the employee distribution for a particular group for a specific year using the probability distribution of the group for the particular year. For example, to calculate the employee distribution after two years (t+2) from the current year(2018), we multiply each state's probability in the $(t+2)^{th}$ year distribution with the current employee strength of that group.

Number of employees in working state = 0.687 * 76 = 52.12 ≈ 52
Number of employees in resign state = 0.168 * 76 = 12.76 ≈ 13
Number of employees in layoff state = 0.120 * 76 = 9.12 ≈ 9
Number of employees in retire state = 0.024 * 76 = 1.824 ≈ 2
Number of employees in promotion state = 0 * 76 = 0

## 4.6 Data and Code Availability Statement

The analyses in this study were conducted in Python using Jupyter Notebook. The dataset and code is available on the Github repository (https://github.com/garvit1608/thesis). We used Python libraries for data manipulation and visualisation, such as NumPy (https://numpy.org), Pandas (https://pandas.pydata.org), Seaborn (https://seaborn.pydata.org) and Matplotlib (https://matplotlib.org). We also used MS-Excel (https://www.microsoft.com/en-ie/microsoft-365/excel) and SQL with MySQLWorkbench (https://www.mysql.com/products/workbench) to find insights from the data.

## 4.7 Summary

To summarize, we started by finding the employee distribution for each cohort. We then used these distributions to find the transition matrix for each cohort. For a particular group, we multiplied its transition matrix with the corresponding initial probability vector to get their probability distribution at time t+1. We then multiplied the probability distribution obtained at time t+1 with the transition matrix to obtain the probability distribution at time t+2, and we continued this step till we got the probability distribution for the 10th year, i.e. t+10. Lastly, we used the probability distributions to calculate exact headcounts for different years.

# 5 Results

We created the transition matrix for each of the scenarios. As discussed earlier, the rows of the transition matrix depict the state at the time 't' and columns depict the state at the time 't+1'. Therefore, for all the transition matrices, a row will represent the current year while the column will represent the state next year. Next, we derived each scenario's probability distribution for the next ten years using the initial probability distribution matrix, transition matrix, and the Markov chain process discussed in the previous chapter. In our system, 'resign'(Re), 'layoff'(Lo), 'retire'(Rt) and 'promote'(Pr) are terminal states; therefore, the probability of transitioning to any other states from these states is 0. Thus for all the transition matrices that we will see further, the transition probabilities for any states from these states is 0. We also looked for the stationary state distribution in each case.

## 5.1 Based on Department

### 5.1.1 Admin

The transition matrix for the admin department is shown in Fig. 5.1. The chances of an employee continuing work are very high(82.89%), while there is less than a 10% chance that an employee will resign(9.2%) and get laid off(6.5%) or retire(1%) in the next year. The probability of an employee getting promoted next year is 0, which means there is a rare possibility that an employee will get promoted next year.

$$
\begin{array}{c c}
 & \begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\
\begin{array}{c} Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} &
\begin{bmatrix}
0.829 & 0.092 & 0.066 & 0.013 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 5.1: Transition Matrix for Admin Department

Fig. 5.2 and Table 5.1 shows the probability distribution for the Admin department for the next 10 years. The probability of an employee continuing work without getting promoted decreases with time (blue line). The probability of an employee getting promoted in the next 10 years remains 0, but the chances of an employee resigning or being laid off increase. The chances of an employee getting retired are increasing, but it is still very low even after 10 years(6.5%). Fig. 5.2 shows that all the curves are approaching to form the flat lines hence the Markov chain will attain the stationary state sometime in future.

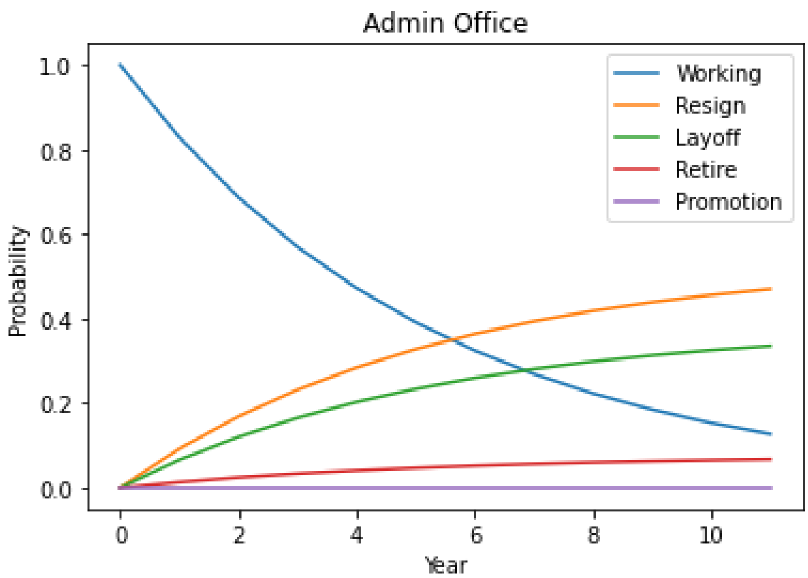


Figure 5.2: Forecasting probabilities of Admin department for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.829 | 0.092 | 0.066 | 0.013 | 0 |
| t+2 | 0.687 | 0.168 | 0.12 | 0.024 | 0 |
| t+3 | 0.57 | 0.232 | 0.165 | 0.033 | 0 |
| t+4 | 0.472 | 0.284 | 0.203 | 0.041 | 0 |
| t+5 | 0.391 | 0.328 | 0.234 | 0.047 | 0 |
| t+6 | 0.324 | 0.364 | 0.259 | 0.052 | 0 |
| t+7 | 0.269 | 0.394 | 0.281 | 0.056 | 0 |
| t+8 | 0.223 | 0.418 | 0.298 | 0.06 | 0 |
| t+9 | 0.185 | 0.439 | 0.313 | 0.063 | 0 |
| t+10 | 0.153 | 0.456 | 0.325 | 0.065 | 0 |

Table 5.1: Probability distribution for Admin department for next 10 years.

### 5.1.2 Executive Office

Fig. 5.3 shows the transition matrix for the Executive Office. The chances of an employee continuing work next year are very high(83.3%), while the chance of an employee being laid off, retired or promoted is rare. The possibility of an employee resigning next year is low(16.6%).

$$\begin{array}{c} \\ Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} \begin{array}{c} \begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\ \begin{bmatrix} 0.833 & 0.1666 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Figure 5.3: Transition matrix for Executive Office Department

Fig. 5.4 and Table 5.2 show the Executive department's probability distribution for the next ten years. The probability of an employee continuing work without getting promoted decreases with time, while the likelihood that an employee will resign in the future increases with time.

In fact, the rate of resignation becomes highest over a period of time. The probability of an employee getting laid off, retired or promoted in the next ten years remains 0. Fig. 5.2 shows that 'working' and 'resign' curves are approaching to form flat lines, and the lines for 'resign', 'layoff' and 'retire' states are straight from the beginning(probability = 0). Hence, it is evident that the Markov chain will attain the stationary state around 20 years into the future.

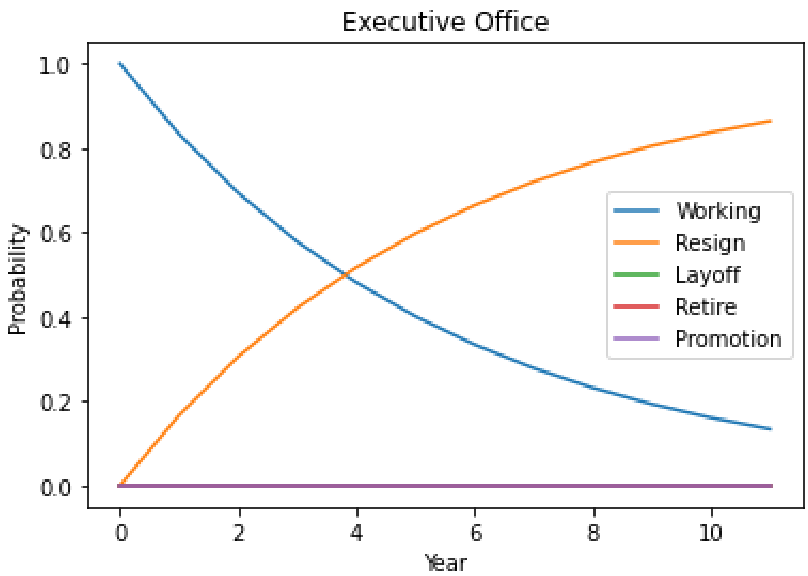


Figure 5.4: Forecasting probabilities of Executive department for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.833 | 0.167 | 0 | 0 | 0 |
| t+2 | 0.694 | 0.305 | 0 | 0 | 0 |
| t+3 | 0.578 | 0.421 | 0 | 0 | 0 |
| t+4 | 0.481 | 0.517 | 0 | 0 | 0 |
| t+5 | 0.401 | 0.597 | 0 | 0 | 0 |
| t+6 | 0.334 | 0.664 | 0 | 0 | 0 |
| t+7 | 0.278 | 0.720 | 0 | 0 | 0 |
| t+8 | 0.232 | 0.766 | 0 | 0 | 0 |
| t+9 | 0.193 | 0.805 | 0 | 0 | 0 |
| t+10 | 0.161 | 0.837 | 0 | 0 | 0 |

Table 5.2: Probability distribution for Executive department for next 10 years.

### 5.1.3 IT/IS

Fig. 5.5 represents the transition matrix for the IT/IS department. The chances of an employee being in a working state are high(67.3%) for the next year compared to other states of resignation (14.8%), layoff(7.4%), retirement (2.2%) and promotion (8.1%).

$$\begin{array}{c c} & \begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\ \begin{array}{c} Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} & \begin{bmatrix} 0.673 & 0.148 & 0.074 & 0.022 & 0.081 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Figure 5.5: Transition matrix for IT/IS Department

Fig. 5.6 and Table 5.3 shows the IT/IS department's probability distribution for the next ten years. The probability of an employee continuing work without getting promoted decreases with time, while the likelihood that an employee will resign, be laid off, retire or be promoted in the future increases. Moreover, the possibility of resignation becomes highest after 4$^{th}$ year. From Fig. 5.2, we found that all curves approach to form flat lines. Hence, it is evident that the Markov chain will attain a stationary state sometime in the future.

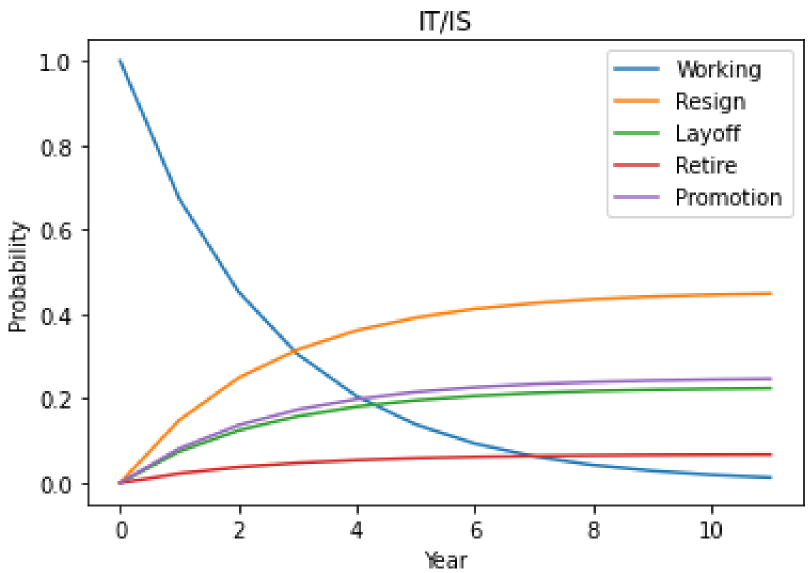


Figure 5.6: Forecasting probabilities of IT/IS department for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.673 | 0.148 | 0.074 | 0.022 | 0.082 |
| t+2 | 0.453 | 0.248 | 0.124 | 0.037 | 0.137 |
| t+3 | 0.305 | 0.316 | 0.158 | 0.047 | 0.174 |
| t+4 | 0.205 | 0.361 | 0.18 | 0.054 | 0.198 |
| t+5 | 0.138 | 0.392 | 0.196 | 0.059 | 0.215 |
| t+6 | 0.093 | 0.412 | 0.206 | 0.062 | 0.226 |
| t+7 | 0.063 | 0.426 | 0.213 | 0.064 | 0.234 |
| t+8 | 0.042 | 0.435 | 0.217 | 0.065 | 0.239 |
| t+9 | 0.028 | 0.442 | 0.221 | 0.066 | 0.243 |
| t+10 | 0.019 | 0.446 | 0.223 | 0.067 | 0.245 |

Table 5.3: Probability distribution for IT/IS department for next 10 years.

### 5.1.4 Software Engineering

Fig. 5.7 shows the transition matrix for Software Engineering Department. There are very high chances that an employee will continue to work(73.3%) in the next year, while there is extremely less possibility that an employee will retire(P(Retire/Working)=0) in the next year. The chances that an employee will resign or get laid off next year are almost equal(13.3% and 10% respectively). The possibility of getting a promotion next year seems significantly less(2.7%).

$$\begin{array}{c c} & \begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\ \begin{array}{c} Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} & \begin{bmatrix} 0.733 & 0.133 & 0.107 & 0 & 0.027 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Figure 5.7: Transition Matrix for Software Engineering Department

Fig. 5.8 and Table 5.4 shows the Software Engineering department's probability distribution for the next ten years. The probability of an employee continuing work without getting promoted decreases with time, while the likelihood that an employee will resign, be laid off or be promoted in the future increases with time. The probability that an employee will get retire over the next ten years is 0. In the $4^{th}$ year, we can see that the probability of an employee continuing to work and being laid off becomes equal(0.28), which is higher than the probability of promotion(0.06). An interesting point to notice is that the rate of resignation and layoff becomes considerably high than other factors rate beyond $4^{th}$ year. Fig. 5.2 shows that all curves are approaching to form flat lines. Hence, it is clear that the Markov chain will attain a stationary state sometime in the future.

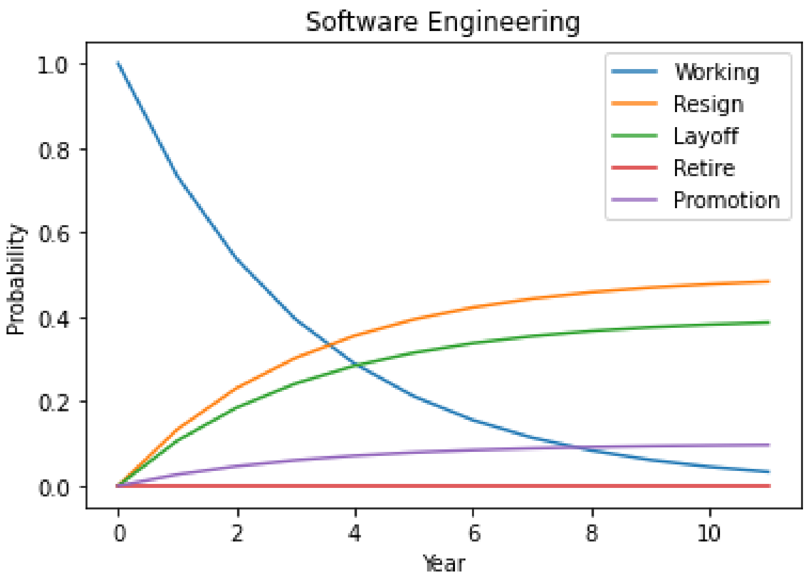


Figure 5.8: Forecasting probabilities of Software Engineering department for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.733 | 0.133 | 0.107 | 0 | 0.027 |
| t+2 | 0.538 | 0.231 | 0.185 | 0 | 0.046 |
| t+3 | 0.394 | 0.303 | 0.242 | 0 | 0.061 |
| t+4 | 0.289 | 0.355 | 0.284 | 0 | 0.071 |
| t+5 | 0.212 | 0.394 | 0.315 | 0 | 0.079 |
| t+6 | 0.155 | 0.422 | 0.338 | 0 | 0.084 |
| t+7 | 0.114 | 0.443 | 0.354 | 0 | 0.089 |
| t+8 | 0.084 | 0.458 | 0.366 | 0 | 0.092 |
| t+9 | 0.061 | 0.469 | 0.375 | 0 | 0.094 |
| t+10 | 0.045 | 0.477 | 0.382 | 0 | 0.095 |

Table 5.4: Probability distribution for Software Engineering department for next 10 years.

### 5.1.5 Production

Fig. 5.9 shows the transition matrix for Production Department. There are high chances that an employee will continue to work(52.3%), while there is extremely less possibility that an employee will retire(2.4%) in the next year. The chances that an employee will resign next year is 20% which is comparatively less than the value for the working state. The possibility of getting a promotion(11%) and being laid off(14%) next year seems less.

$$
\begin{array}{c|ccccc}
 & Wk & Re & Lo & Rt & Pr \\
\hline
Wk & 0.523 & 0.201 & 0.140 & 0.024 & 0.110 \\
Re & 0 & 1 & 0 & 0 & 0 \\
Lo & 0 & 0 & 1 & 0 & 0 \\
Rt & 0 & 0 & 0 & 1 & 0 \\
Pr & 0 & 0 & 0 & 0 & 1
\end{array}
$$

Figure 5.9: Transition Matrix for Production Department

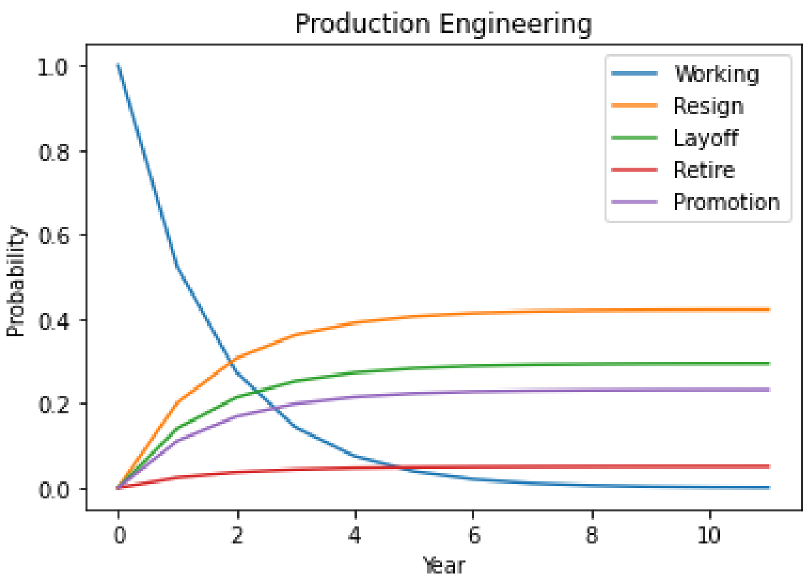


Figure 5.10: Forecasting probabilities of Production Engineering department for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.523 | 0.201 | 0.141 | 0.024 | 0.111 |
| t+2 | 0.274 | 0.307 | 0.214 | 0.037 | 0.169 |
| t+3 | 0.143 | 0.362 | 0.252 | 0.043 | 0.199 |
| t+4 | 0.075 | 0.391 | 0.273 | 0.047 | 0.215 |
| t+5 | 0.039 | 0.406 | 0.283 | 0.049 | 0.223 |
| t+6 | 0.02 | 0.414 | 0.289 | 0.05 | 0.228 |
| t+7 | 0.011 | 0.418 | 0.291 | 0.05 | 0.23 |
| t+8 | 0.006 | 0.42 | 0.293 | 0.05 | 0.231 |
| t+9 | 0.003 | 0.421 | 0.294 | 0.051 | 0.232 |
| t+10 | 0.002 | 0.422 | 0.294 | 0.051 | 0.232 |

Table 5.5: Probability distribution for Production Engineering department for next 10 years.

Fig. 5.10 and Table 5.5 shows the Production department's probability distribution for the next ten years. The probability of an employee continuing work without getting promoted decreases with time, while the likelihood that an employee will resign, be laid off or be promoted in the future increases. From the 2$^{nd}$ year onwards, the probability of an employee quitting and layoff passes over working state probability. Such probability distribution shows that the production engineering department needs some attention. Fig. 5.2 shows that all curves form flat lines from the 7$^{th}$ year onwards. Thus, it is clear that the Markov chain will attain a stationary state in the future.

### 5.1.6 Sales

Fig. 5.11 represents the transition matrix for the sales department. There are high chances that an employee will continue to work(71.6%) in the next year, while there is significantly less possibility that an employee will retire(1.8%) and be laid off(5.7%) in the next year. The chances that an employee will get promoted is 11.1% which is comparatively less than the value for the working state.

$$
\begin{array}{c c}
 & \begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\
\begin{array}{c} Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} &
\begin{bmatrix}
0.716 & 0.096 & 0.057 & 0.018 & 0.111 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 5.11: Transition Matrix for Sales Department

Fig. 5.12 and Table 5.6 represents the probability distribution for the sales department for the next ten years. The probability of an employee continuing work is highest at the start(0.71), but it decreases with time and reduces to the minimum(0.036) amongst all by the end of the 10th year. The probability of promotion increases with time, which means more employees will get promoted in the coming years. The probability of resignation and layoff increases with time, but these remain low(under 20%) throughout the prediction period. Fig. 5.12 shows that all curves are approaching to form flat lines. Hence, it is clear that the Markov chain will attain a stationary state in the future.

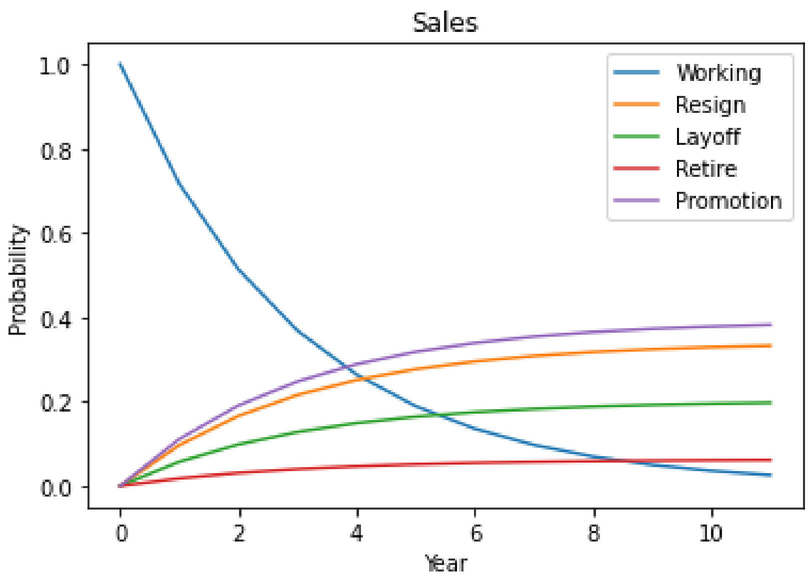


Figure 5.12: Forecasting probabilities of the Sales department for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.717 | 0.097 | 0.057 | 0.018 | 0.111 |
| t+2 | 0.514 | 0.166 | 0.098 | 0.031 | 0.191 |
| t+3 | 0.368 | 0.216 | 0.128 | 0.04 | 0.248 |
| t+4 | 0.264 | 0.251 | 0.149 | 0.047 | 0.289 |
| t+5 | 0.189 | 0.277 | 0.164 | 0.051 | 0.318 |
| t+6 | 0.136 | 0.295 | 0.175 | 0.055 | 0.339 |
| t+7 | 0.097 | 0.308 | 0.183 | 0.057 | 0.354 |
| t+8 | 0.07 | 0.318 | 0.188 | 0.059 | 0.365 |
| t+9 | 0.05 | 0.324 | 0.192 | 0.06 | 0.373 |
| t+10 | 0.036 | 0.329 | 0.195 | 0.061 | 0.378 |

Table 5.6: Probability distribution for Sales department for next 10 years.

## 5.2 Based on Age Group

### 5.2.1 Age group 26-40

Fig. 5.13 represents the transition matrix for the youngest age group of an organization. There are high chances that an employee will continue to work(57.8%), while there is less possibility that an employee will resign(18.6%), be laid off(12.9%) or get promoted(10.7%) in the next year. The probability that an employee will retire next year is 0. This also makes sense since the employees in this age group are young and far from retirement.

$$
\begin{array}{c c}
 & \begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\
\begin{array}{c} Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} &
\begin{bmatrix}
0.578 & 0.186 & 0.129 & 0 & 0.107 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 5.13: Transition matrix for 26-40 age group

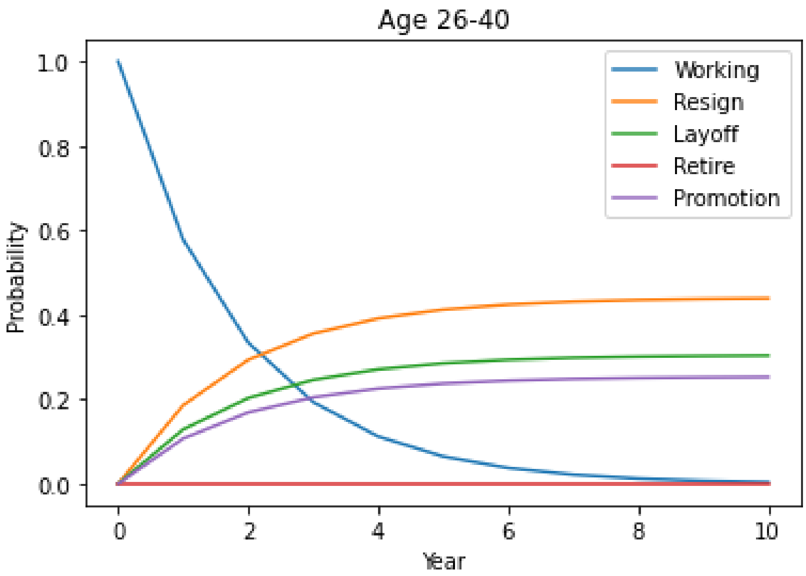


Figure 5.14: Forecasting probabilities of 26-40 age group for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.578 | 0.186 | 0.129 | 0 | 0.107 |
| t+2 | 0.334 | 0.293 | 0.203 | 0 | 0.169 |
| t+3 | 0.193 | 0.356 | 0.246 | 0 | 0.205 |
| t+4 | 0.112 | 0.392 | 0.271 | 0 | 0.225 |
| t+5 | 0.065 | 0.412 | 0.285 | 0 | 0.237 |
| t+6 | 0.037 | 0.424 | 0.294 | 0 | 0.244 |
| t+7 | 0.022 | 0.431 | 0.298 | 0 | 0.248 |
| t+8 | 0.013 | 0.435 | 0.301 | 0 | 0.251 |
| t+9 | 0.007 | 0.438 | 0.303 | 0 | 0.252 |
| t+10 | 0.004 | 0.439 | 0.304 | 0 | 0.253 |

Table 5.7: Probability distribution for 26-40 age group for next 10 years.

Fig. 5.14 and Table 5.7 represents the probability distribution for the 26-40 age group for the next ten years. The probability of an employee continuing work is highest at the start(0.578), but it decreases with time and reduces to the minimum(0.004) amongst all by the end of the 10th year. The probability of promotion increases with time, which means more employees will get promoted in the coming years. On the contrary, the probability of resigning increases and becomes considerably large(around 0.4) at the end of the 4th year. The probability of layoff also increases with time, but it is considerably less than resign state probabilities. Fig. 5.12 shows that all curves are approaching to form flat lines. Hence, it is clear that the Markov chain will attain a stationary state sometime in the future.

### 5.2.2 Age group 41-60

Fig. 5.15 represents the transition matrix for an organisation's 41 to 60 age group. There are high chances that an employee will continue to work(56.1%), while there is less possibility that an employee will resign(18.4%), be laid off(10.4%) or get promoted(11.3%) in the next year. There is very less chance that an employee will retire next year(3.7%).

$$\begin{array}{c} \\ Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} \begin{array}{c} \begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\ \begin{bmatrix} 0.561 & 0.184 & 0.104 & 0.037 & 0.113 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Figure 5.15: Transition matrix for 41-60 age group

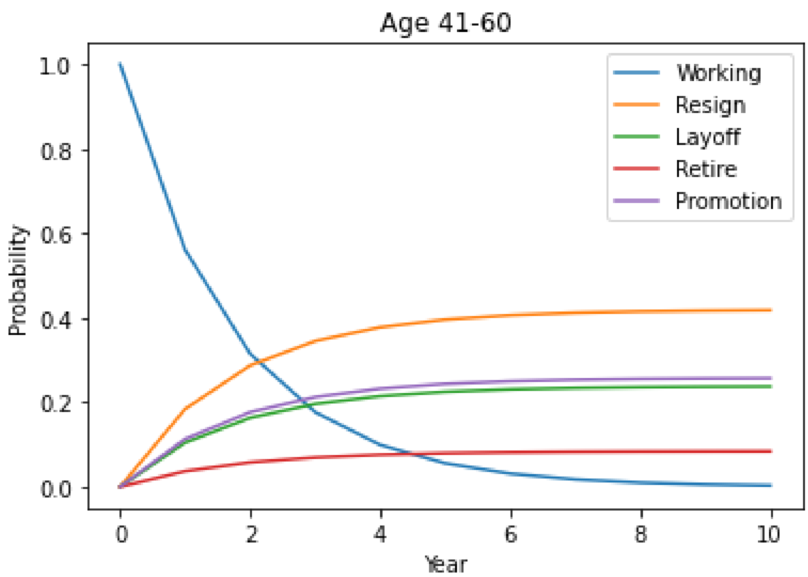


Figure 5.16: Forecasting probabilities of 41-60 age group for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.561 | 0.184 | 0.104 | 0.037 | 0.113 |
| t+2 | 0.315 | 0.287 | 0.163 | 0.058 | 0.177 |
| t+3 | 0.177 | 0.345 | 0.196 | 0.069 | 0.212 |
| t+4 | 0.099 | 0.378 | 0.214 | 0.076 | 0.232 |
| t+5 | 0.056 | 0.396 | 0.225 | 0.08 | 0.244 |
| t+6 | 0.031 | 0.406 | 0.231 | 0.082 | 0.25 |
| t+7 | 0.017 | 0.412 | 0.234 | 0.083 | 0.253 |
| t+8 | 0.01 | 0.415 | 0.236 | 0.083 | 0.255 |
| t+9 | 0.006 | 0.417 | 0.237 | 0.084 | 0.256 |
| t+10 | 0.003 | 0.418 | 0.237 | 0.084 | 0.257 |

Table 5.8: Probability distribution for 41-60 age group for next 10 years.

Fig. 5.16 and Table 5.8 represents the probability distribution for the 41-60 age group for the next ten years. We can see that the probability of an employee continuing work is highest at the start(0.561), but it decreases with time and reduces to the minimum(0.003) amongst all by the end of the 10th year. The probability of promotion increases with time, which means more employees will get promoted in the coming years. The probability of resigning increases and becomes considerably high 4th year onwards. The probability of layoff and promotion increases at almost the same rate, but it is considerably less than resign state probabilities. Fig. 5.12 shows that all curves are approaching to form flat lines around $10^{th}$ year. Hence, it is clear that the Markov chain will attain a stationary state in the future.

### 5.2.3 Age group 60+

Fig. 5.17 represents the transition matrix for the 60+ age group of an organization. There are high chances that an employee will continue to work(52.0%), while there is less possibility that an employee will resign(18.4%) or be laid off(10.2%) in the next year. The likelihood of an employee retiring is 12.2%, while there is a significantly less chance that an employee will be promoted next year(7.1%).

$$
\begin{array}{cc}
 & \begin{array}{ccccc} \quad Wk & \quad Re & \quad Lo & \quad Rt & \quad Pr \end{array} \\
\begin{array}{c} Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array} &
\begin{bmatrix}
0.520 & 0.184 & 0.102 & 0.122 & 0.071 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 5.17: Transition matrix for 60+ age group

Fig. 5.18 and Table 5.9 represents the probability distribution for the 60+ age group for the next ten years. We can see that the probability of an employee continuing work is highest at the start(0.52), but it decreases with time and reduces to the minimum(0.001) amongst all by the end of the $10^{th}$ year. The probability of resigning increases and becomes considerably high $4^{th}$ year onwards. This is the only case amongst all cases where the probability of an employee getting retired increases at a considerable rate to become 0.25 at the end of 10 years. This also makes sense since the employees in this age group are closer to the retirement age. The probability of layoff and promotion increases at a slow rate and remains under 0.25 under the observation period. Fig. 5.18 shows that all curves are approaching to form flat lines at the end of the 8th year. Hence, it is clear that the Markov chain will attain a stationary state in the future.

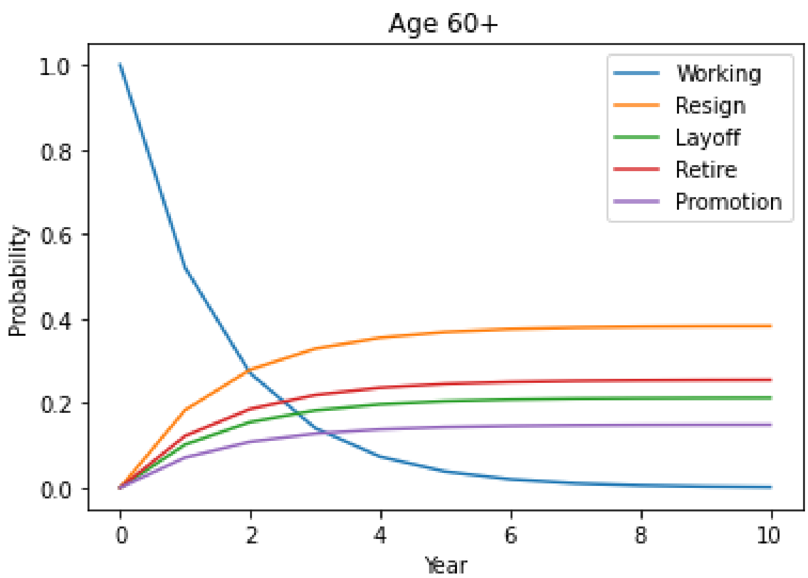


Figure 5.18: Forecasting probabilities of 60+ age group for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.520 | 0.184 | 0.102 | 0.122 | 0.071 |
| t+2 | 0.271 | 0.279 | 0.155 | 0.186 | 0.109 |
| t+3 | 0.141 | 0.329 | 0.183 | 0.219 | 0.128 |
| t+4 | 0.073 | 0.355 | 0.197 | 0.236 | 0.138 |
| t+5 | 0.038 | 0.368 | 0.205 | 0.245 | 0.143 |
| t+6 | 0.02 | 0.375 | 0.209 | 0.25 | 0.146 |
| t+7 | 0.01 | 0.379 | 0.211 | 0.253 | 0.147 |
| t+8 | 0.005 | 0.381 | 0.212 | 0.254 | 0.148 |
| t+9 | 0.003 | 0.382 | 0.212 | 0.254 | 0.148 |
| t+10 | 0.001 | 0.382 | 0.212 | 0.255 | 0.149 |

Table 5.9: Probability distribution for 60+ age group for next 10 years.

## 5.3 Based on Gender

### 5.3.1 Male

Fig. 5.19 represents the transition matrix for the male employees of an organisation. There are high chances that an employee will continue to work(67.7%), while there is less possibility that an employee will resign(13.7%), be laid off(8.7%), or promote(8%) in the next year. The likelihood of an employee retiring in the next year is least among all (1.8%).

$$
\begin{array}{c} \\ Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array}
\begin{array}{c}
\begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\
\begin{bmatrix}
0.677 & 0.137 & 0.087 & 0.018 & 0.080 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 5.19: Transition matrix for Male employees

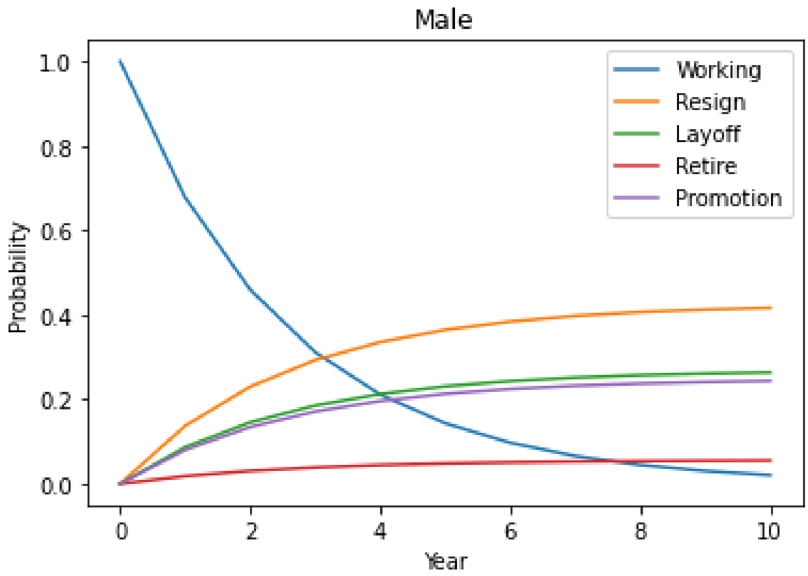


Figure 5.20: Forecasting probabilities of Male group for next 10 years.

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.678 | 0.137 | 0.087 | 0.018 | 0.08 |
| t+2 | 0.46 | 0.23 | 0.145 | 0.03 | 0.134 |
| t+3 | 0.312 | 0.293 | 0.185 | 0.039 | 0.171 |
| t+4 | 0.211 | 0.336 | 0.212 | 0.044 | 0.196 |
| t+5 | 0.143 | 0.364 | 0.231 | 0.048 | 0.213 |
| t+6 | 0.097 | 0.384 | 0.243 | 0.051 | 0.224 |
| t+7 | 0.066 | 0.397 | 0.251 | 0.052 | 0.232 |
| t+8 | 0.045 | 0.406 | 0.257 | 0.054 | 0.237 |
| t+9 | 0.03 | 0.412 | 0.261 | 0.054 | 0.241 |
| t+10 | 0.02 | 0.417 | 0.264 | 0.055 | 0.243 |

Table 5.10: Probability distribution for Male group for next 10 years.

Fig. 5.20 and Table 5.10 represents the probability distribution for the male employees for the next ten years. We can see that the probability of an employee continuing work is highest at the start(0.678), but it decreases with time and reduces to the minimum(0.02) amongst all by the end of the 10th year. The probability of resigning increases and becomes considerably high(around 0.4) at the end of the 8th year. The probability of layoff and promotion increases at almost the same rate, but it is considerably less than resign state probabilities. Fig. 5.12 shows that all curves are approaching to form flat lines. Hence, it is clear that the Markov chain will attain a stationary state sometime in the future.

### 5.3.2 Female

Fig. 5.21 represents the transition matrix for the female employees of an organisation. There are high chances that an employee will continue to work(47.5%), while there is relatively less possibility that an employee will resign(20.7%), be laid off(13.9%), or get promoted(15%) in the next year. The likelihood of an employee retiring in the next year is least among all (2.5%).

$$
\begin{array}{c} \\ Wk \\ Re \\ Lo \\ Rt \\ Pr \end{array}
\begin{array}{c}
\begin{array}{ccccc} Wk & Re & Lo & Rt & Pr \end{array} \\
\begin{bmatrix}
0.475 & 0.207 & 0.139 & 0.025 & 0.155 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 5.21: Transition matrix for Female employees

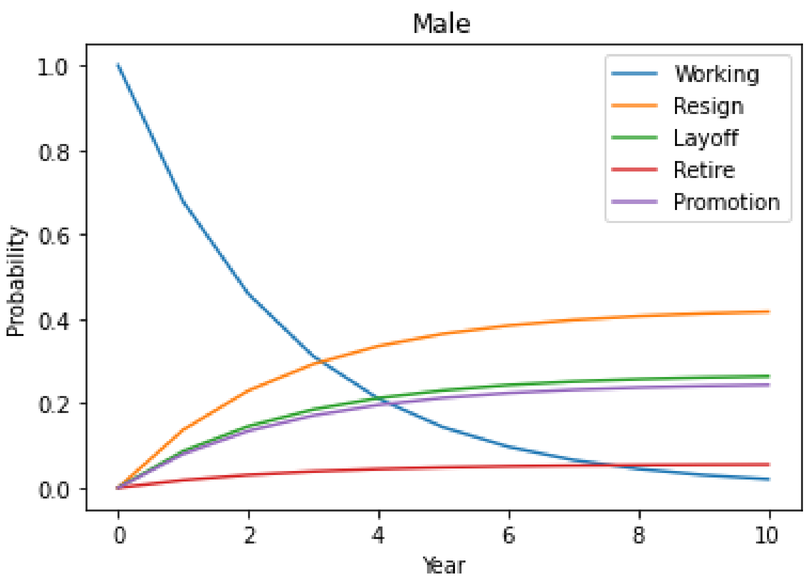


Figure 5.22: Forecasting probabilities of Female group for next 10 years.

Fig. 5.22 and Table 5.11 represents the probability distribution for the female employees for the next ten years. The probability of an employee continuing work is highest at the start(0.475), but it decreases with time and reduces to the minimum(0.001) amongst all by

| Year | Working | Resign | Layoff | Retire | Promote |
|---|---|---|---|---|---|
| t = 0 | 1 | 0 | 0 | 0 | 0 |
| t+1 | 0.475 | 0.207 | 0.139 | 0.025 | 0.155 |
| t+2 | 0.225 | 0.305 | 0.204 | 0.037 | 0.229 |
| t+3 | 0.051 | 0.374 | 0.25 | 0.045 | 0.28 |
| t+4 | 0.024 | 0.384 | 0.257 | 0.046 | 0.288 |
| t+5 | 0.011 | 0.389 | 0.261 | 0.047 | 0.292 |
| t+6 | 0.005 | 0.391 | 0.262 | 0.047 | 0.294 |
| t+7 | 0.003 | 0.393 | 0.263 | 0.047 | 0.294 |
| t+8 | 0.001 | 0.393 | 0.263 | 0.047 | 0.295 |
| t+9 | 0.001 | 0.393 | 0.263 | 0.047 | 0.295 |
| t+10 | 0.001 | 0.393 | 0.263 | 0.047 | 0.295 |

Table 5.11: Probability distribution for Female group for next 10 years.

the end of the 10th year. The probability of resigning increases and becomes considerably high(around 0.4) at the end of the 5th year. The probability of layoff and promotion increases at almost the same rate, but it is considerably less than resign state probabilities. Fig. 5.12 shows that all curves are approaching to form flat lines. Hence, it is clear that the Markov chain will attain a stationary state sometime in the future.

## 5.4 Calculating Employees Headcounts

As discussed in the methods of calculating probability distributions, the probabilities can be converted to an exact number by multiplying the probability for a particular state and year with the total number of employees in that particular category. For example, consider Table 5.11, which shows probability distributions for female employees for ten years. From Table 4.12, we know that the total number of female employees in the organisation is 967. Let us say we want to calculate the employee strength across different states for the year 't+3'.

Number of female employees continuing working = $0.051 * 967 = 49.31 \approx 49$
Number of female employees resigning = $0.374 * 967 = 361.65 \approx 362$
Number of female employees laid off = $0.25 * 967 = 241.75 \approx 242$
Number of employees retiring = $0.045 * 967 = 43.51 \approx 43$
Number of employees being promoted = $0.28 * 967 = 270.76 \approx 271$

Similarly, we can find out the number of employees for any year and any of the discussed scenarios.

# 6 Discussion

The results project the year-wise trend of employees resigning, getting laid off, retiring and being promoted for the given organisation for ten years starting from 2018 across different departments, age groups and gender. For all the cases, we found that the probability of continuing to work decreases with time while the probability of resignation, layoff, retirement and promotion increases. In most cases, the rate of resignation is higher than the rate for other factors. Also, there is a considerable growth rate in a layoff in all cases. These two factors are necessary to address to maintain employee strength. The promotion rate remains less than 20% in almost every case over the span of ten years. The retirement rate remained the lowest (less than 10%) of all in most of the cases. We can see that all the cases will attain the stationary state sometime in future.

At an organisation level, this means that there is a high chance that more employees will be resigning in the future. More employees leaving implies that the employees are not satisfied with their work or the department's culture, and hence the HR department should form necessary policies to retain such employees. Low wage rates, inability to cope with early career work stress and frustration due to stagnation in the current role are some of the reasons for employee turnover [60]. Layoffs are also a matter of concern for an organisation. One reason for layoffs is that the employees are not performing well, so the HR department should focus on organising skills-related training sessions for employees. Another reason for the layoffs can be to improve the financial performance of the organisation [61]. Again, a thorough introspection and specific policies need to be implemented to ensure that such crises do not arise again. According to Campbell [62], promotion in organisations serves two purposes: assigning individuals into roles for which their talents and capabilities are best suited and rewarding prior performance with increased pay and rank in the company. Decent growth in the promotion rate will imply that both purposes are met. The HR department should ensure that deserving employees get promoted and implement necessary performance review policies.

The analysis of this study could help the Human resource department to find the gaps related to promotion, retirement, layoffs and resignations and hence can help them to form their policies diligently. However, there are certain limitations to this thesis. Firstly, our Markov chain model does not take inflows into account. We are projecting future employee strength based on current employee strength without taking factors such as hiring, return from sabbatical and return from maternity leave into account. Secondly, there could have been more features in the dataset that we could have considered to get some better results. Next, in our research, we are projecting the absolute probability distribution of different states for over ten years, but these distributions might change due to certain external factors. Hence, the proposed model might not be robust to accommodate the reasonable and fluctuating probabilities caused by external factors.

# 7 Future Work

As discussed in section 1.3, the scope of our study was limited to projecting future workforce trends for ten years, and we have adequately covered the scope. This thesis could be a starting point for someone interested in exploring workforce forecasting using Markov chain analysis. There is an evident amount of work available on workforce planning, but a holistic study of workforce planning considering inflows, outflows and sensitivity analysis is rarely available.

Using the supply(inflow) data to develop the workforce strategy is the first future work that could be taken into consideration. The dataset used in our analysis lacked information about the inflow of employees. As per the suggestion of Dr Honari, I tried looking for alternate ways to incorporate the inflows, but due to limited time, I was not able to make progress in that direction. If given more time, I would like to explore the possibilities of incorporating inflows in this study.

We could also do the perturbation or sensitivity analysis of the required workforce planning strategy due to variation on both supply and demand sides. The perturbation in the parameters could take place due to certain exogenous factors. Matrix calculus is an appealing method for determining the sensitivity and elasticity of parameter perturbations for Markov chain models [63]. The sensitivity of the steady-state performance can also be studied using realization factors, performance potentials, and the group inverse of the infinitesimal generator [64]. Sensitivity analysis can increase the reliability and confidence intervals of predictions of our model.

# Bibliography


[1] Chartered Institute of Personnel and Development. Workforce planning, 2020. URL `https://www.cipd.ie/news-resources/practical-guidance/factsheets/workforce-planning`.

[2] Mary Young and Seth Hollander. Looking back at strategic workforce planning—and peering ahead. URL `https://www.shrm.org/executive/resources/people-strategy-journal/Fall2019/Pages/young-hollander-feature.aspx`.

[3] Cameron Nouri. 6 reasons why poor workforce planning stunts growth. URL `https://pingboard.com/blog/6-reasons-why-poor-workforce-planning-stunts-growth`.

[4] Julia Howes. Strategic workforce planning - latest trends and leading practice examples, 05 2015. URL `https://tinyurl.com/yvuc5wet`.

[5] Phillip E Pfeifer and Robert L Carraway. Modeling customer relationships as markov chains. *Journal of interactive marketing*, 14(2):43–55, 2000.

[6] J. Sullivan. Why you need workforce planning. 81(12):46–50, 2022.

[7] Martin Kunc. Achieving a balanced organizational structure in professional services firms: some lessons from a modeling project. *System Dynamics Review: The Journal of the System Dynamics Society*, 24(2):119–143, 2008.

[8] Gareth H Rees, Peter Crampton, Robin Gauld, and Stephen MacDonell. New zealand's health workforce planning should embrace complexity and uncertainty. *NZ Med J*, 131 (1477):109–15, 2018.

[9] Scott Shpak. Hr forecasting statistical vs judgmental techniques. URL `https://smallbusiness.chron.com/hr-forecasting-statistical-vs-judgmental-techniques-76166.html`.

[10] Anahita Safarishahrbijari. Workforce forecasting models: a systematic review. *Journal of Forecasting*, 37(7):739–753, 2018.

[11] Mylene Lagarde and Duane Blaauw. A review of the application and contribution of discrete choice experiments to inform human resources policy interventions. *Human resources for health*, 7(1):1–10, 2009.

[12] Don R Bryant, Michael J Maggard, and Robert P Taylor. Manpower planning models and techniques: A descriptive survey. *Business Horizons*, 16(2):69–78, 1973.

[13] Brian Parker and David Caine. Holonic modelling: human resource planning and the two faces of janus. *International journal of manpower*, 1996.

[14] Srour and Issam Mounir. *A linear programming approach to optimize strategic investment in the construction workforce*. The University of Texas at Austin, 2005.

[15] RJ Niehaus. Evolution of the strategy and structure of a human resource planning dss application. *Decision Support Systems*, 14(3):187–204, 1995.

[16] Joaquim Júdice, Pedro Martins, and Jacinto Nunes. Workforce planning in a lotsizing mail processing problem. *Computers & operations research*, 32(11):3031–3058, 2005.

[17] Bernhard Wild and Christoph Schneewei. Manpower capacity planning—a hierarchical approach. *International Journal of Production Economics*, 30:95–106, 1993.

[18] Jay W Forrester. System dynamics, systems thinking, and soft or. *System dynamics review*, 10(2-3):245–256, 1994.

[19] B Wilson. Workforce planning for future requirements. *International Journal of Workforce*, 8(3):3–8, 1987.

[20] Jacques JF Commandeur and Siem Jan Koopman. *An introduction to state space time series analysis*. Oxford University Press, 2007.

[21] Chien-Chin Hsu, Shun-Hsing Chen, and Ben-Chang Hsien. The manpower forecast model of the energy technology industry. *Journal of Statistics and Management Systems*, 15(4-5):499–517, 2012.

[22] NC Cary. User's guide. *SAS Institute Inc*, 2010.

[23] AC Igboanugo and OR Edokpia. A markovian study of manpower planning in the soft-drink industry in nigeria. *Nigerian Journal of Technology*, 33(4):547–552, 2014.

[24] Mose Job Nyandwaki and J Kennedy. Statistical modeling of kenyan secondary school students enrollment: An application of markov chain model. *IOSR Journal of Mathematics*, 12(2):11–18, 2016.

[25] AA Osagiede and VU Ekhosuehi. Markovian approach to school enrolement projection process. *Global Journal of Mathematical Sciences*, 5(1):1–7, 2006.

[26] GO Ogbogbo, GU Ebuh, and CO Aronu. Prediction of academic manpower system of a polytechnic institution in nigeria. *Science Journal of applied mathematics and statistics*, 1(5):54–61, 2013.

[27] HH Shugart Jr, TR Crow, and JM Hett. Forest succession models: a rationale and methodology for modeling forest succession over large regions. *Forest Science*, 19(3): 203–212, 1973.

[28] L.M Peden, J.S. William, and W.E. Frayer. A markov model for stand-alone projection. *Forest Science*, 19:303–314, 1973.

[29] Robert F Cassell and John W Moser. programmed markov model for predicting diameter distribution and species composition in uneven-aged forests. 1974.

[30] Chennupati R Rao and AM Kshirsagar. A semi-markovian model for predator-prey interactions. *Biometrics*, pages 611–619, 1978.

[31] AE Ajayi and AA Olufayo. The sequence of wet and dry days at ibadan and onne (sub-humid zone of nigeria). *Nigerian journal of technology*, 21(1):38–45, 2002.

[32] J. W. Harbaugh and O. Bonham-Carter. Computer simulation in geology, london, chichester. 1970.

[33] Joseph Buongiorno and Bruce R Michie. A matrix model of uneven-aged forest management. *Forest science*, 26(4):609–625, 1980.

[34] Jian-qiang Luo and Yan-ping Zhao. Research on the supply chain product market forecasting based on markov chain. In *2010 International Conference on E-Product E-Service and E-Entertainment*, pages 1–3, 2010. doi: 10.1109/ICEEE.2010.5660723.

[35] Xiangzhou Yin, Xinwei Tian, and Yun Chen. A study on stability of closed-loop supply chain based on markov chain. In *2010 International Conference on Management and Service Science*, pages 1–4, 2010. doi: 10.1109/ICMSS.2010.5575601.

[36] Nguyen Trong Tri Duc, Pham Duc Tai, and Jirachai Buddhakulsomsiri. Approximating measures of performance of a periodic review inventory system by using markov chain. In *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, pages 543–547, 2020. doi: 10.1109/ICIEA49774.2020.9102069.

[37] VO Ezugwu and S Ologun. Markov chain: A predictive model for manpower planning. *Journal of Applied Sciences and Environmental Management*, 21(3):557–565, 2017.

[38] David G Collings and Geoffrey Wood. A critical approach. *Human resource management: A critical approach*, page 1, 2009.

[39] AC Igboanugo and MK Onifade. Markov chain analysis of manpower data of a nigerian. *Journal of innovative research in engineering and science*, 2:2, 2011.

[40] D Parma, S Raisinghani, and P Makwana. Application of markovian theory in manpower planning: A case study. *Global research analysis*, 2(2):122–124, 2013.

[41] HY Touama. Application of markovian models and transition propbabilities' matrix to analyze the workforce movement in jordanian productivity companies. *Paripex-Indian Journal of Research*, 4(6):215–218, 2015.

[42] Hyuk Kwon, J Byung, L Eui-Jin, and Y ByungHoon. Prediction of human resource supply/demand in nuclear industry using markov chain model and job coefficient. In *Transactions of Korean Nuclear Society Autumn Meeting, Gyeongin, Korea*, 2006.

[43] VO Ezugwu and LI Igbinosun. Analysis of manpower system using multi-absorbing states markov chain. 2020.

[44] Virtue U Ekhosuehi, Augustine A Osagiede, and Wilfred A Iguodala. A procedure for distributing recruits in manpower systems. *Yugoslav Journal of Operations Research*, 25 (3):445–456, 2015.

[45] Tim De Feyter. Modelling heterogeneity in manpower planning: dividing the personnel system into more homogeneous subgroups. *Applied stochastic models in business and industry*, 22(4):321–334, 2006.

[46] V.A. Dimitriou and N. Tsantas. Evolution of a time dependent markov model for training and recruitment decisions in manpower planning. *Linear Algebra and its Applications*, 433(11):1950–1972, 2010. ISSN 0024-3795. doi: https://doi.org/10.1016/j.laa.2010.07.001. URL https://www.sciencedirect.com/science/article/pii/S0024379510003587.

[47] Wan-yin Du and Shou Li. Application of markov model in human resource supply forecasting in enterprises. In *2015 International Conference on Computational Science and Engineering*, pages 151–156. Atlantis Press, 2015.

[48] Kannan Nilakantan and BG Raghavendra. Control aspects in proportionality markov manpower systems. *Applied Mathematical Modelling*, 29(1):85–116, 2005.

[49] Helen E. O. Lucent-Iwhiwh and Simon Ayo Adekunle. Conceptual approach to manpower planning in organizations. 2014.

[50] DAVIDE POLIZZI. Human resources data set, 2020. URL https://www.kaggle.com/datasets/davidepolizzi/hr-data-set-based-on-human-resources-data-set.

[51] Sanjay Kumar Dwivedi and Bhupesh Rawat. A review paper on data preprocessing: A critical phase in web usage mining process. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pages 506–510, 2015. doi: 10.1109/ICGCIoT.2015.7380517.

[52] Qualcomm. Data collection and pre-processing techniques. URL https://developer.qualcomm.com/software/qualcomm-neural-processing-sdk/learning-resources/ai-ml-android-neural-processing/data-collection-pre-processing.

[53] Hossein Pishro-Nik. Introduction to probability, statistics, and random processes. 2016.

[54] Jeremy Jackson Henry Maltby, Worranat Pakornrat. Markov chains. URL https://brilliant.org/wiki/markov-chains.

[55] Ericmjl. Markov models from the bottom up, with python. URL https://ericmjl.github.io/essays-on-data-science/machine-learning/markov-models/.

[56] David Schön Myers, Lisa Wallin, and Petter Wikström. An introduction to markov chains and their applications within finance. *MVE220 Financial Risk: Reading Project*, 2017.

[57] Gerhard Naegele and Alan Walker. A guide to good practice in age management. 2006.

[58] Row Henson. Hr in the 21st century: Challenges and opportunities. *IHRIM Journal*, 6 (6):28–32, 2002.

[59] Kellogg School of Management at Northwestern University. Better decisions through diversity, 10 2010. URL https://insight.kellogg.northwestern.edu/article/better_decisions_through_diversity.

[60] KUS Somarathna. An agent-based approach for modeling and simulation of human resource management as a complex system: Management strategy evaluation. *Simulation Modelling Practice and Theory*, 104:102118, 2020.

[61] Kenneth P De Meuse, Paul A Vanderheiden, and Thomas J Bergmann. Announced layoffs: Their effect on corporate financial performance. *Human Resource Management*, 33(4):509–530, 1994.

[62] Dennis Campbell. Nonfinancial performance measures and promotion-based incentives. *Journal of Accounting Research*, 46(2):297–332, 2008.

[63] Hal Caswell. Sensitivity analysis of discrete markov chains via matrix calculus. *Linear Algebra and its Applications*, 438(4):1727–1745, 2013.

[64] Xi-Ren Cao and Han-Fu Chen. Perturbation realization, potentials, and sensitivity analysis of markov processes. *IEEE Transactions on Automatic Control*, 42(10): 1382–1393, 1997. doi: 10.1109/9.633827.