



REVA
UNIVERSITY
Bengaluru, India

A Project Report on
Development of Analytical and KPI Datamart for
Recruitment Analytics

Submitted in Partial Fulfilment for Award of Degree of
Master of Business Administration
In Business Analytics

Submitted By
Ashish Chandra Jha
R19MBA01

Under the Guidance of
Dr. J. B. Simha
Chief Mentor
RACE, REVA
University

REVA Academy for Corporate Excellence - RACE
REVA University
Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064
race.reva.edu.in

August, 2022



Candidate's Declaration

I, **Ashish Chandra Jha** hereby declare that I have completed the project work towards the second year of **Master of Business Administration in Business Analytics** at, REVA University on the topic entitled **Development of Analytical and KPI Datamart for Recruitment Analytics** under the supervision of **Dr. J. B. Simha**, Chief Mentor. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year 2022.

A handwritten signature in blue ink, appearing to read "Ashish Chandra Jha", is placed over a light gray rectangular background.

Place: Bengaluru

Date: 27th August 2022

Name of the Student: Ashish Chandra Jha

Signature of Student



Certificate

This is to Certify that the project work entitled **Development of Analytical and KPI Datamart for Recruitment Analytics** carried out by Ashish Chandra Jha with R19MBA01, is a bonafide student of REVA University, is submitting the second year project report in fulfilment for the award of Master of Business Administration in Business Analytics during the academic year 2022. The project report has been tested for plagiarism, and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of Development of Analytical Datamart and Data Pipeline for Recruitment Analytics work prescribed for the said degree.

A handwritten signature in blue ink, appearing to read "Dr. J. B. Simha".

Signature of the Guide

Dr. J. B. Simha
Guide

A handwritten signature in blue ink, appearing to read "Dr. Shinu Abhi".

Dr. Shinu Abhi
Director

External Viva

Names of the Examiners

1. Dr. Sai Hareesh, Research Expert (AI), SAP Labs
2. Pradeeptha Mishra, Director of Artificial Intelligence, Fosfor

Place: Bengaluru

Date: 27th August 2022



Acknowledgement

I would like to thank my mentor Dr. J. B. Simha and to the director of RACE Dr. Shinu Abhi who gave me a golden opportunity to work on this project. I would also like to express my gratitude to Prof. Mithun Dolthody Jayaprakash and Dr. Rashmi Agarwal wholeheartedly.

I must also thank my parents and friends for the immense support and help during this project. Without their help, completing this project would have been very difficult.

I would like to acknowledge the support provided by Honourable Chancellor, Dr. P Shayma Raju, Honourable Vice Chancellor, Dr. M. Dhanamjaya, and Registrar, Dr. N. Ramesh, as a standard protocol.

Place: Bengaluru

Date: 27th August 2022



Similarity Index Report

This is to certify that this project report titled **Development of Analytical and KPI Datamart for Recruitment Analytics** was scanned for similarity detection. Process and outcome is given below.

Software Used: **Turnitin**

Date of Report Generation: **26th Aug 2022**

Similarity Index in %: **5%**

Total word count: **7131**

Name of the Guide: **Dr. J.B.Simha**

A handwritten signature in blue ink, appearing to read "Ajitha", with a diagonal line through it.

Place: Bengaluru

Date: 27th August 2022

Name of the Student: Ashish Chandra Jha

Signature of Student

Verified by: Dincy Dechamma

Signature

Dr. Shinu Abhi,
Director, Corporate Training

List of Abbreviations

| Sl. No | Abbreviation | Long Form |
|---------------|---------------------|---------------------------|
| 1 | ATS | Applicant Tracking System |
| 2 | KPI | Key Performance Indicator |

List of Figures

| No. | Name | Page No. |
|----------------|---|-----------------|
| Figure No. 1.1 | Recruitment Funnel | 11 |
| Figure No. 2.1 | Kimball HR Datamart | 13 |
| Figure No. 2.2 | Oracle HR Datamart | 14 |
| Figure No. 2.3 | SAP HANA Datamart | 16 |
| Figure No. 5.1 | Dimension Table | 21 |
| Figure No.5.2 | Fact Table | 21 |
| Figure No.5.3 | Dimension and Fact Tables in SQL light | 22 |
| Figure No.5.4 | Dimensional Datamart | 22 |
| Figure No.5.5 | Analytical Datamart | 23 |
| Figure No. 7.1 | Dimension Table Example | 24 |
| Figure No. 7.2 | Fact Table Example | 24 |
| Figure No. 8.1 | Location wise candidate joined or not joined | 25 |
| Figure No. 8.2 | Candidate who has extended joining date and joined the company or not | 25 |
| Figure No. 8.3 | Candidate who has actually relocated and joined the company or not | 26 |
| Figure No. 8.4 | Ratio of Male and Female candidates joined and not joined | 26 |
| Figure No. 8.5 | Ratio of people who joined and different Notice Period | 27 |
| Figure No. 9.1 | Star Schema Example | 28 |
| Figure No. 9.2 | Analytical Datamart | 29 |
| Figure No. 9.3 | Time Dimension Table | 30 |
| Figure No. 9.4 | Elbow Method | 31 |
| Figure No. 9.5 | Historic Dashboard | 31 |
| Figure No. 9.6 | Predictive Dashboard | 32 |

Abstract

The HR department handles all the data regarding the recruitment process while also analyzing them to select suitable candidates for the organization. The HR department interacts with the data regarding the recruitment process using many tools for data analysis and interpretation. Such data has to be organized in a way to ensure proper analysis as well as the screening of candidates.

Each system gives their own report in their own format. Data is not integrated and aggregated at proper granularity suitable for analysis. This can be mitigated by the use of Dimensional datamarts and analytical datamarts.

The dimensional datamarts contains transitional data modified for analysis with dimensions and facts. These datamarts can be used for any adhoc analysis like drill down/roll up, slice and dice, drill through, comparative analysis etc. The analytical datamart contains the aggregated data in one row per employee format. This will be task specific in the sense that the standard tabular structure/table containing multiple attributes/independent variables for a specific problem like Quality of Hire modeling, Cost of Hire, Time of Hire, demand prediction etc.

This project has focused on creating a recruitment datamart for the Human Resource Department with the assistance of the data modeling technique. This Project has also discussed the exploratory data analysis and model building in relation to the datamarts for the recruitment

Keywords:- HR Analytics, Recruitment, Datamart, KPI, KPI Datamart, Analytical Datamart

Table of Contents

| Contents | |
|---|----|
| CANDIDATE'S DECLARATION | 2 |
| CERTIFICATE | 3 |
| ACKNOWLEDGEMENT | 4 |
| SIMILARITY INDEX REPORT..... | 5 |
| LIST OF ABBREVIATIONS..... | 6 |
| LIST OF FIGURES | 6 |
| ABSTRACT..... | 7 |
| TABLE OF CONTENTS..... | 8 |
| CHAPTER 1: INTRODUCTION..... | 9 |
| CHAPTER 2: LITERATURE REVIEW | 14 |
| CHAPTER 3: PROBLEM STATEMENT | 19 |
| CHAPTER 4: OBJECTIVES OF THE STUDY | 20 |
| CHAPTER 5: PROJECT METHODOLOGY | 21 |
| CHAPTER 6: BUSINESS UNDERSTANDING | 25 |
| CHAPTER 7: DATA UNDERSTANDING..... | 28 |
| CHAPTER 8: DATA PREPARATION | 31 |
| CHAPTER 9: MODELING | 35 |
| CHAPTER 10: EVALUATION | 39 |
| CHAPTER 11: DEPLOYMENT | 42 |
| CHAPTER 12: ANALYSIS AND RESULTS | 43 |
| CHAPTER 13: CONCLUSIONS AND FUTURE SCOPE | 44 |
| BIBLIOGRAPHY | 45 |
| APPENDIX..... | 46 |
| PLAGIARISM REPORT | 46 |
| PUBLICATIONS IN A JOURNAL/CONFERENCE PRESENTED/WHITE PAPER..... | 49 |
| ANY ADDITIONAL DETAILS | 49 |

Chapter 1: Introduction

The success and growth of any organization primarily depend on its employees. Without employees, no organization can run as they are the workforce for the organization. Various employees take up different roles in the organization and perform the duties assigned to their positions. The employees of an organization are crucial for the organization and therefore the department that majorly deals with them becomes crucial as well. This department is the Human Resource Department of an organization. The HR department is responsible for almost all aspects concerning the workforce of an organization. The importance of employees for an organization makes the HR department very important. The Human Resource Department is responsible for the effective administration of the employees of an organization to help in its overall growth. Moreover, an organization can only succeed if competent people are recruited in it. This department needs to carefully examine and select suitable persons for the various roles in the organization.

1.1 Recruitment Process

Recruitment can be defined as the overall process of choosing the right candidates for job positions at an organization based on several criteria. The recruitment process is usually handled primarily by the Human Resource department of an organization. This process might differ for every organization but there are certain general steps involved in it that most organizations seem to follow. These are Job Description Preparation, Sourcing, Screening, Selecting, Hiring, and Onboarding.

1.1.1 JD Preparation

Job description refers to the details of the job position for which the recruitment is taking place. It contains all the necessary information regarding the job such as-

- Qualifications required for the job
- Skills required to perform the necessary functions of the job
- Different responsibilities that the selected candidate has to take care of
- Suitable characteristics for the candidates for the job such as age, experience, etc.
- Salary expectations concerning the job
- Working hours

The job description is an important aspect of the recruitment process as it notifies the interested candidates with all the necessary information about the job. It is the primary source of attracting candidates for the job position.

1.1.2 Sourcing

Sourcing is the process of collection of data relevant to the position for which recruitment is being done. In other words, this process involves the collection of resumes or data of various candidates who are suitable for the concerned position. The HR department scrambles through multiple portals for data collection such as Naukri.com, Linkedin, monsterindia.com, etc. This process does not involve the selection of the candidates, it only involves the collection of the information of the relevant candidates for further screening and selection.

1.1.3 Screening

Screening is the first assessment of the candidates with the necessary qualifications for the job position. All the resumes and data that are collected in the sourcing process are reviewed and screened by the HR department. This stage finds out those candidates who will be advancing to the next level of selection. Screening the resumes of candidates and selecting the top ones from the whole lot is the usual activity in this process.

1.1.4 Selecting

Selecting is the main stage of choosing the most suitable candidate for the job for which the recruitment is being done. Those candidates who passed the screening stage have to go through this phase of selection which might involve multiple tests. Personal tests, group discussions, aptitude tests are some of the most common tests that are used to select the most suitable candidate for the job post. The candidate who is selected also gets to negotiate his/her salary with the HR department in this phase itself. The best candidate is selected for the job and no more candidates are assessed after this selection stage.

1.1.5 Hiring

The candidate who is finally selected in the selecting stage is hired by the company. This occurs in this phase called hiring. The concerned candidate is presented with the offer letter from the organization's HR department. All the terms and conditions of the job along with other necessary information regarding the job are presented to the chosen candidate at this

stage. If the organization and the chosen candidate agree on all the terms and conditions set forward, the candidate is hired into the organization.

1.1.6 Onboarding

Onboarding is the last phase in the recruitment process and by this stage, a suitable candidate is already selected and hired for the concerned job position. In this phase, the new employee is welcomed into the organization by the HR department. Onboarding is the process of introducing the new employee to the organization, its cultures, rules & regulations, his/her role, other employees, and the way of working. It is a way of familiarizing the new employee about the various aspects of the organization and the job role.

1.2 Recruitment Funnel Analysis

The process of recruitment can be quite a tough process for an organization. Selecting the most suitable candidate/candidates from a whole lot of individuals can be gruesome. Even more, the multiple phases in recruitment can be hard to manage. This brings up the need for a recruitment funnel. A recruitment funnel is a structure that can help to ease the process of recruitment for an organization. It breaks down the entire recruitment process into multiple stages through goals and tasks which brings down the complexity of the process.

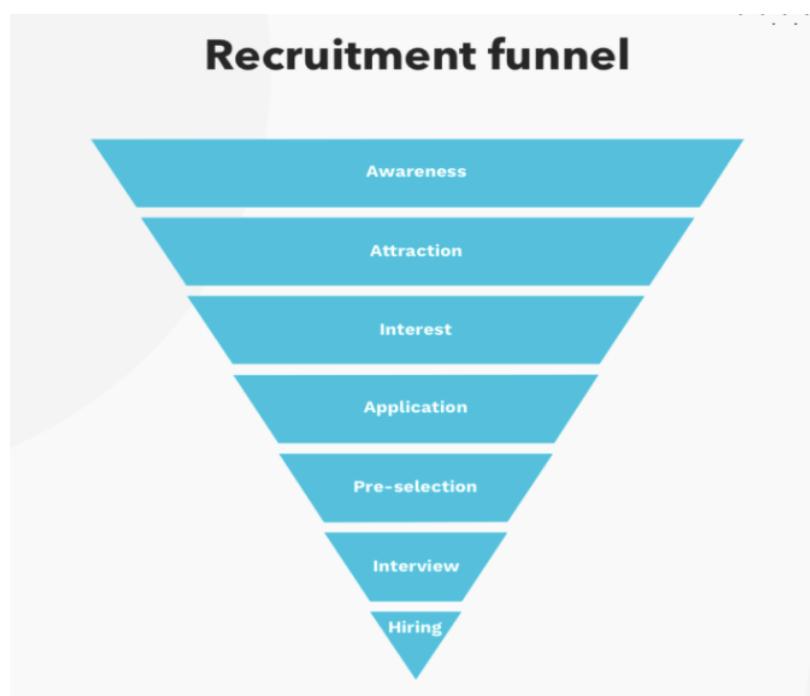


Figure No. 1.1 Recruitment Funnel

When an organization breaks down the recruitment process using a well-designed recruitment funnel, the task is significantly simplified without having any negative results (Alexandra, 2019). The recruitment funnel is shown Figure No. 1.1.

1.2.2 Recruitment Funnel

Various layers of recruitment are:-

1. **Awareness** – This is the stage where an organization builds awareness about its brand image to attract the attention of the interested candidates. Employees tend to work at places about which they know about before joining officially. Therefore it is imperative that organizations build their brand image well. Social media can work wonders to create awareness about vacant job positions and build a good brand image as well. Sometime companies do cost marketing activities to create the awareness.
2. **Attraction** – In this stage, it need to be able to market the recruitment process in the most attractive manner possible. Ideal candidates will only choose to be a part of the recruitment process if it attracts their attention. The first pillar of attraction is always the job description. It should always be written in a way that conveys proper information about the job. Organizations should also target their recruitment strategies towards specific areas where they are most likely to find the best people for the job. Organizations can also share the experiences of their current employees to build their image and attract suitable candidates. Making good use of technology can also be of great help to achieve the goal of attraction such as maintaining the organization's website properly.
3. **Interest** – Here we have to build up the interest of the candidates who are interested in the job. Organizations must be ready to answer the queries of the candidates. A well-designed chatbot might be of great use in this case. Also, organizations should look to have a good rating of themselves on employer sites to further ingest more interest into the candidates.
4. **Application** – This is the stage where the candidates actually apply for the concerned job position. Organizations must ensure that the application process is easy. To make the process even simpler, organizations should make sure that candidates can apply

through mobile devices. The most important part about the application part is that organizations should always follow up on the applications with an e-mail or any form of communication.

5. **Pre-selection** – In this stage organizations narrow down a whole lot of candidates to a smaller group for further selection. The candidates are chosen on the basis of the required skills for the job position. Organizations should ensure that there is no biasness in any of the tests during this stage. All the tests should give equal opportunities to all individuals whatsoever. Making the entire assessment program unique to the organization can also work to the benefit of the employers.
6. **Interview** – This is the stage of the final process of selection. The group of candidates to have passed the previous stage have to go through a personal interview with the employers. Interviews should be carefully designed to select the best candidate while also giving all the candidates an equal opportunity as well. The interview should be able to assess all the desired skills of a candidate. These days most of the interviews are happening online. Which is good for the organization in terms of saving interview conduct cost.
7. **Hiring** – In this stage, the candidate who is selected in the interview is finally hired into the organization. The HR department of the organization should negotiate with the selected individual regarding the necessary terms and conditions for the concerned job position. The candidate should be well satisfied with the job offer.

HR Analytics is defined as a HR process which uses the HR data (e.g. recruitment data) and domain knowledge to do start predicting the performance and cost involved of the people on the basis of their work (Udhay Kailash, 2020).

Chapter 2: Literature Review

According to the Gartner, human resource leader has begun analysing the HR data at each level to improve the HR efficiency, recruitment efficiency and enhance employee experience (Gartner, 2019). In the fast growing world with their emerging new challenges, it is very difficult to hire skilled people which will be in the organization for a decent span of time. To create such system, we need good quality of data backed by domain knowledge to provide analytical suggestion to hire a candidate. To create such analytical system, the business data should be extensively available to the hiring manager to work align to find the suitable candidate (Alysson Prado, 2010).

For a smooth functioning of an organization, their strategic decision should be backed by data. Data helps us to find the insights or patterns among the people or work. This data should be properly created and organized in such a way that it can be further used to perform any analytical tasks. The success of human resource department with the help of analytics can be measured as the success of the organization. HR department tries to understand the business requirement and ensure the finding of right candidate according to their skills and qualifications. To hire such a niche level people we need to create a precise datamart and use its data for analytical view. Human resource (HR) datamart is the base stone for building an enterprise data warehouse. The paper presents the implementation process of HR datamart starting from implementing datamart schema to online analytical processing (OLAP) reports (Alaa Khalaf Hamoud, 2020).

There are quite a few datamarts available in the market these days. Among them, quite popular are the Kim Kimball HR Datamart which is the base for many available datamarts, Oracle HR Datamart, SAP HR Datamart.

HR Datamart by Ralph Kimball

The Kimball HR Datamart is based on the ideology of Ralph Kimball according to whom the data warehouses should be model using dimensional models such as the star schema or snowflake schema. Kimball's approach was the bottom-up approach that involved creating datamarts first to allow quick analysis and interpretation of the data in them. These small

datamarts can later be combined to form a normal data warehouse. Kimball focused on the use of dimensional models to improve the performance of the users of the data warehouse (George, 2012).

In order to understand Kimball's HR datamart, we have to look at the star schema dimension model. A star schema is a tool for dimensional modeling of data by organizing it to allow analytical operations to run on it. The star schema presents the data in the form of fact tables and dimension tables. The fact tables contain the primary data regarding the business process, in our case, the recruitment data, which are termed as 'facts'. The other tables or dimension tables contain the data associated with the primary fact tables. The dimension table contains 'dimensions' which sort of describe the data in the facts table (Ralph Kimball, 2016).

The dimensional tables surround the fact table depicting the descriptive nature of the facts. The multiple dimension tables around the fact table make this model look something like a star due to which it is known as the star schema.

Kimball's HR datamart focused on doing the harder things in the beginning while keeping the easy things for the end. Below is the Employees tracking datamart from Kimball. I could not find anything specific to only recruitment in Kimball datamart, which we will be discussing inferring Kimball datamart concept (Ralph Kimball, 2016). Kimball HR Datamart is shown in Figure No. 2.1.

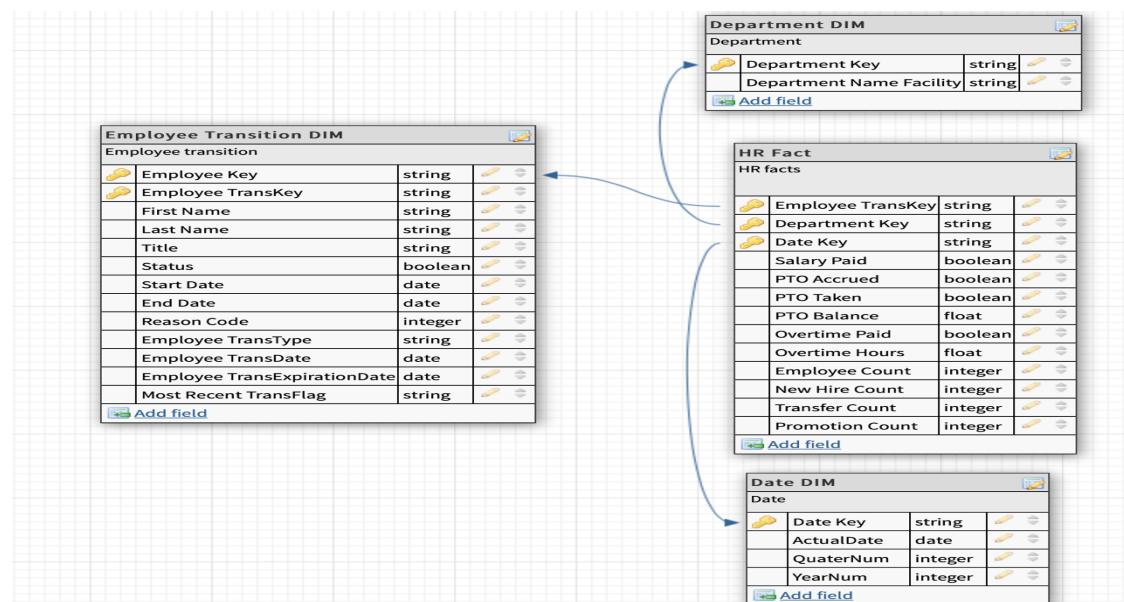


Figure No. 2.1 Kimball HR Datamart (Ralph Kimball, 2016)

HR Datamart by Oracle

Then we have Oracle Datamart. Oracle Recruitment datamart is a data storehouse of enlistment also, staffing drives, status, costs, and results. This information store contains more outlined data and all current and memorable well-known enrolment drives including open positions, orders, candidates, applications, results, and enrolment cost. Recruitment measurements can be reported and investigated by a wide assortment of credits, including specialty unit, division, work code, area, status, status reason, reference class, and subcategory, etc. The Recruiting information store furnishes staff and the board with data expected to settle on informed choices with respect to current and future recruitment drives.

Oracle gives a total and self-administration arrangement that permits business groups to get the profound, reliable, information-driven experiences they need to settle on speedy choices. Business groups can rapidly join all vital information across various sources and organizations, including spatial and diagram, in a combined data set to drive secure cooperation around a solitary wellspring of truth given by information stores. Analysts can without much of a stretch influence self-administration information devices and implanted AI—with zero coding needed—to speed up information stacking, change, and readiness, consequently find examples and patterns, make expectations, and gain experiences dependent on information with straightforward ancestry (Oracle, Recruiting Data Mart Dimension Tables, n.d.) (Oracle, Installing Human Resources Schema on Autonomous Database, 2020).

Oracle Datamart has been shown in Figure No.2.2.

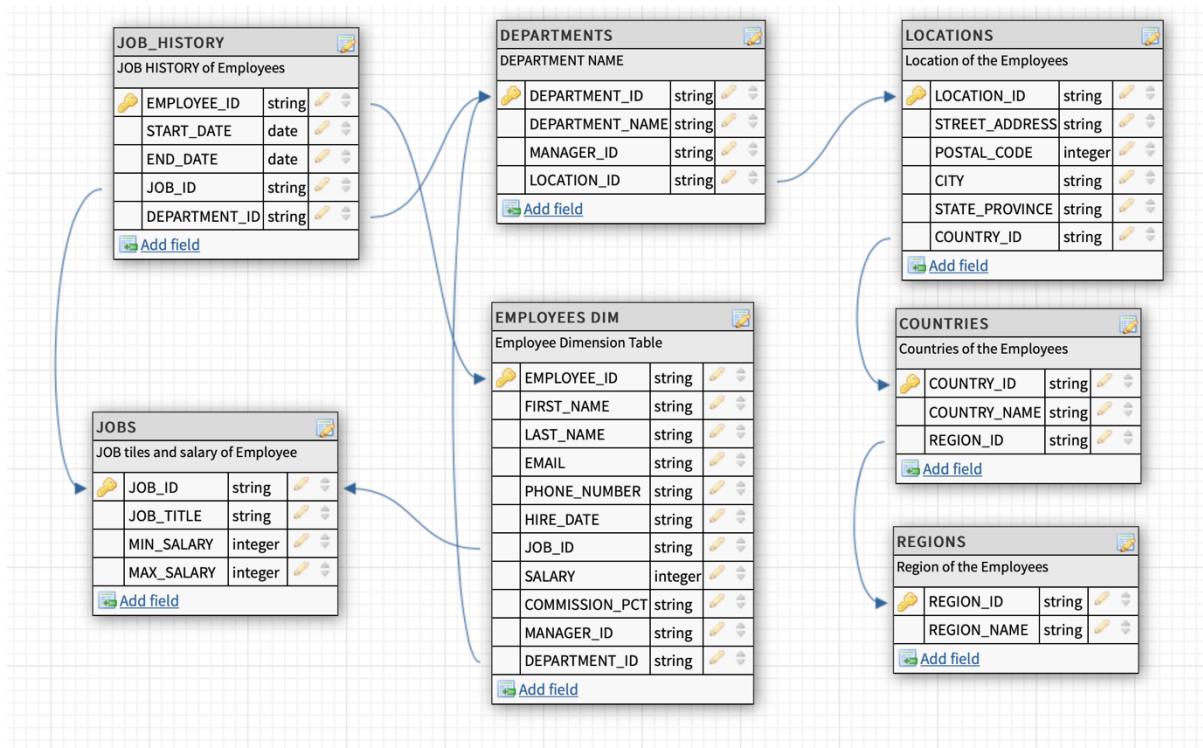


Figure No. 2.2 Oracle HR Data Mart (Oracle, Recruiting Data Mart Dimension Tables, n.d.)

HR Datamart by SAP

SAP HANA is a database platform that offers customers excellent analysis of large chunks of data in real-time. It is available both as a service deployable on-site or as a cloud-based service. The SAP HANA HR datamart offers real-time analytics and reporting on data replicated from the system's database.

Filling in as a platform for big business asset arranging (ERP) programming and other business applications, SAP HANA can be set on-premises, in the cloud, or both, in a mixture cloud framework.

SAP HANA coordinates information from various regions inside an association, for instance:

- Customary business records – including agreements and accounting pages.
- User Experience/User Interface – including site structures, messages, and other client associations.
- Versatile – data from the cell phones of clients and their labour force.
- Internet of Things – information from the numerous sensors that disagreement each part of a business, from stockrooms and trucks to stores and workplaces.

The SAP HANA datamart not just coordinates the entirety of this information; it can likewise apply AI and AI to break down it right away and profoundly, speeding up ongoing dynamics by giving key bits of knowledge into an organization's tasks. It's in-memory, multi-model information management motor exploits the abilities of its equipment to limit information developments, along these lines speeding up and nimbleness as it examines ongoing information.

Contingent upon the requirements of an enterprise, SAP HANA can be conveyed on-premises, in the cloud, or as a mixture structure, mixing the security and control of an on-premises system with the lower cost, more prominent memory, and expanded admittance of the cloud. Its capacity to effectively handle huge measures of information makes it effectively versatile to suit a developing business without forfeiting security or dependability (SAP, 2018). SAP HANA datamart has been shown in Figure No. 2.3.

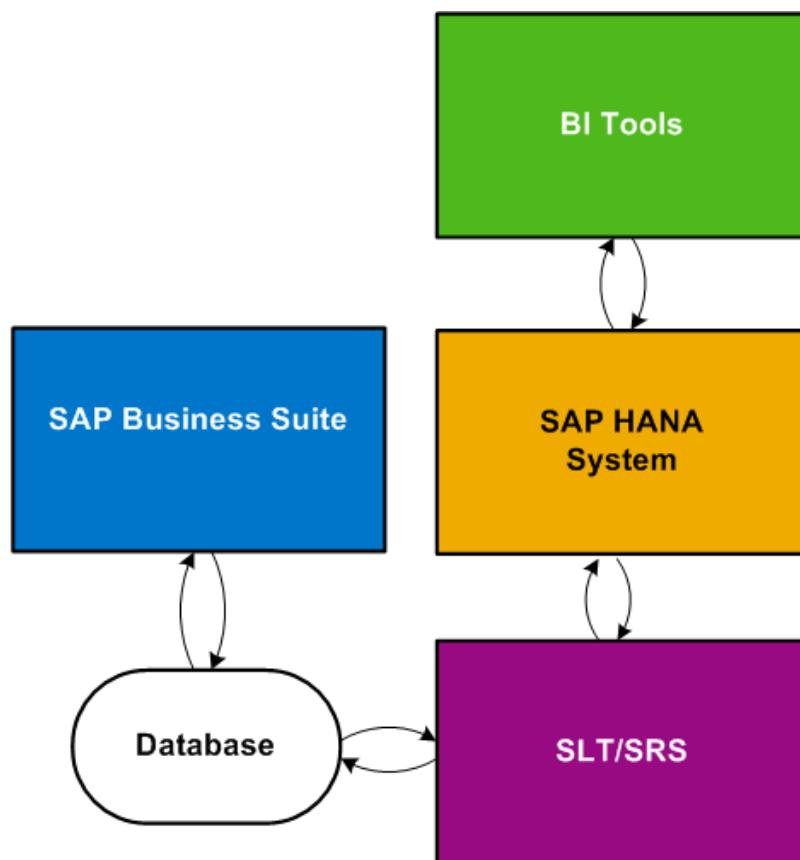


Figure No.2.3. SAP HANA Datamart (SAP, 2018)

Chapter 3: Problem Statement

Recruitment data comes from various data sources. Many tools exist to help spotters and business pioneers track measurements and sort out recruitment information. The popular HR examination tools are Applicant Tracking System (ATS), Background screening technology, Candidate sourcing platforms, Social media, Google Analytics, Survey tools etc.

None of these sources provide the same information or data. If we use more than one tool to analyse the same data, there are chances we will get different insights. This proves the fact that till now there is no single source of truth.

1. There is no concrete KPI tracking, comparison, deviation detection in a single port.
2. To come up with a comprehensive analytical Datamart for Human Resource departments with minimal drawbacks and disadvantages.
3. To develop an efficient data pipeline for the analysis of recruitment data.
4. To develop a proper system for analysis and interpretation of recruitment data and assist in the recruitment process.

Chapter 4: Objectives of the Study

The objectives of this study can be summarised in the following points-

1. Define and develop dimensional/analytical datamarts for recruitment.
2. Define the source data structure required to populate the proposed recruitment datamart.
3. Develop data integration engine to process and populate dimensional and analytical datamart.
4. Implement end to end reference case to demonstrate the feasibility of the proposed solution.

Chapter 5: Project Methodology

The project methodology comprises of Planning, Building Dimensional and Analytical Datamart, Building Star Schema, Historical and Analytical Dashboards, Model Building, Model Evaluation and Model Deployment.

Step 1:- Planning

The structure of data comprises of the source and target. These source and target need to be identified first. Source of data will be candidate_ref_id, job_id etc. and target will be recruitment datamart, dimensional datamart etc.

Dimension Table:- Dimension table comprises of a primary key column which is embedded as a foreign key in any associated fact table where the dimension row's descriptive context is exactly correct for that fact table row. Dimension Table is shown in Figure No. 5.1.

| Dimension Table | | | |
|---------------------------|----------------------|---------|------|
| key | Candidate_Ref_Id | integer | edit |
| | Manager_Id | integer | edit |
| | Candidate_Experience | integer | edit |
| | Candidate_Gender | text | edit |
| | Candidate_Age | integer | edit |
| | Candidate_Location | text | edit |
| | Candidate_Source | text | edit |
| | Job_Role | text | edit |
| | Job_Category | text | edit |
| | Salary_Band | integer | edit |
| | Department | text | edit |
| Add field | | | |

Figure No.5.1. Dimension Table

Fact Table:- A fact table contains the numeric measures such as metrics produced by an operational measurement. Fact Table is shown in Figure No. 5.2.

| Fact Table | | |
|---------------------------|---------|--|
| Job_Status | integer | |
| Cost_Of_Hire | integer | |
| Quality_Of_Hire | integer | |
| Time_To_Hire | integer | |
| Candidate_Source | text | |
| Add field | | |

Figure No. 5.2. Fact Table

Step 2:- Building Dimensional and Analytical Datamart

Before designing the dimensional model, we need to find the dimension and fact tables.

Creating the Dimension and Fact Table in SQL. Here, SQL light has been used to build the dimension and fact tables. Dimension and Fact Tables in SQL light is shown in Figure 5.3.

| Name | Type | Schema |
|----------------------|---------|--|
| Tables (2) | | |
| Dimension Table | | |
| Candidate_Ref_Id | INTEGER | CREATE TABLE "Dimension Table" ("Candidate_Ref_Id" INTEGER, "Manager_Id" INTEGER, "Candidate_Experience" INTEGER, "Candidat |
| Manager_Id | INTEGER | "Candidate_Ref_Id" INTEGER |
| Candidate_Experience | INTEGER | "Manager_Id" INTEGER |
| Candidate_Gender | TEXT | "Candidate_Experience" INTEGER |
| Candidate_Age | INTEGER | "Candidate_Gender" TEXT |
| Candidate_Source | TEXT | "Candidate_Age" INTEGER |
| Job_Role | TEXT | "Candidate_Source" TEXT |
| Job_Category | TEXT | "Job_Role" TEXT |
| Salary_Band | TEXT | "Job_Category" TEXT |
| Department | TEXT | "Salary_Band" TEXT |
| | | "Department" TEXT |
| Fact Table | | CREATE TABLE "Fact Table" ("Job_Status" BLOB, "Cost_of_Hire" INTEGER, "Quality_of_Hire" TEXT, "Time_to_Hire" INTEGER) |
| Job_Status | BLOB | "Job_Status" BLOB |
| Cost_of_Hire | INTEGER | "Cost_of_Hire" INTEGER |
| Quality_of_Hire | TEXT | "Quality_of_Hire" TEXT |
| Time_to_Hire | INTEGER | "Time_to_Hire" INTEGER |
| Indices (0) | | |
| Views (0) | | |
| Triggers (0) | | |

Figure No.5.3. Dimension and Fact Tables in SQL light

Dimensional Datamart

The dimensional datamarts will contain transitional data modified for analysis with dimensions and facts. These datamarts can be used for any adhoc analysis like drill down/roll up, slice and dice, drill through, comparative analysis etc. Dimensional Datamart has been shown in Figure No. 5.4.

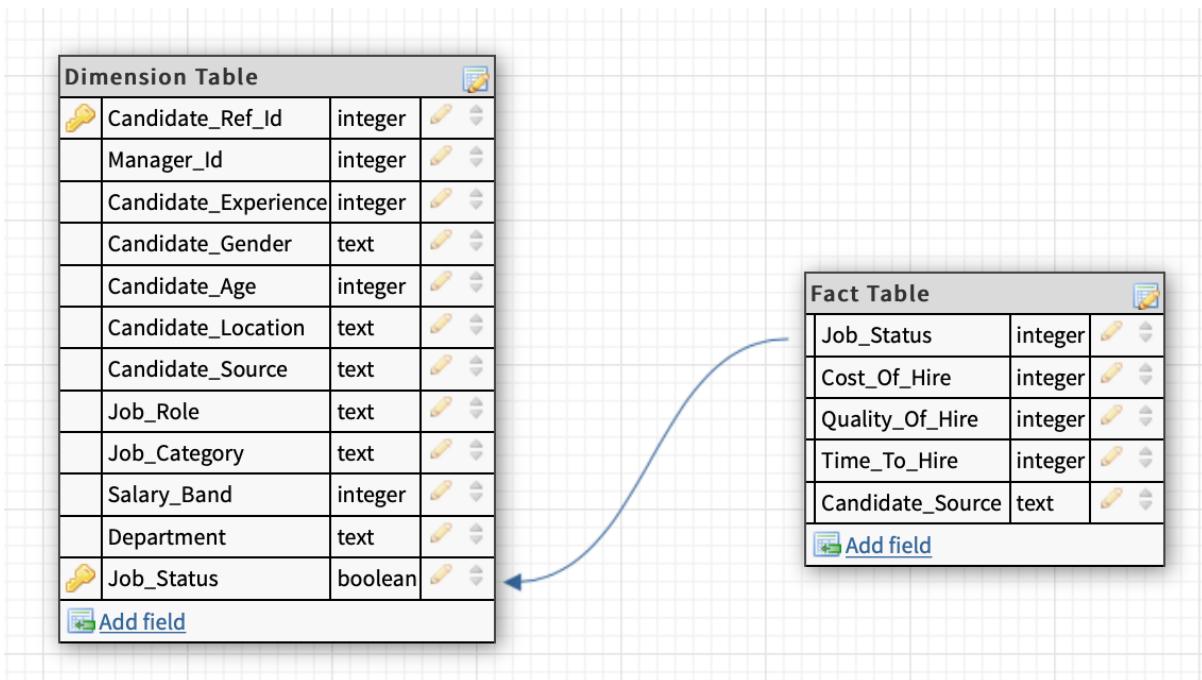


Figure No. 5.4. Dimensional Datamart

Analytical Datamart

The analytical datamart will contain the aggregated data in one row per employee format. This will be task specific in the sense that the standard tabular structure/table containing multiple attributes/independent variables for a specific problem like Quality of Hire modeling, Cost of Hire, Time to Hire, demand prediction etc. Analytical Datamart has been shown in Figure No. 5.5.

| Name | Type | Schema |
|---------------------|------|---|
| Tables (1) | | |
| Analytical_Datamart | | <pre>CREATE TABLE "Analytical_Datamart" ("Candidate_Ref_Id" INTEGER, "Candidate_Gender" TEXT, "Candidadate_Age" INTEGER, "Candidate_Ref_Id" INTEGER "Candidate_Gender" TEXT "Candidadate_Age" INTEGER "Rex_in_Yrs" INTEGER "LOB" TEXT "Candidate_Location" TEXT "Candidate_relocate_actual" TEXT "Notice_period" INTEGER "Job_ID" INTEGER "Department" TEXT "Job_Title" TEXT "Candidate_Source" TEXT "DOJ_Extended" TEXT "Duration_to_accept_offer" INTEGER "Offered_band" TEXT "Prospective_Salary_Band" INTEGER "Percent_hike_expected_in_CTC" INTEGER "Percent_hike_offered_in_CTC" INTEGER "Percent_difference_CTC" INTEGER "Joining_Bonus" TEXT "Job_Status" TEXT "Job_Status_Code" INTEGER "Job_Opening_Date" TEXT "Job_Closing_Date" TEXT "Hiring_Time" INTEGER "Recruiter" TEXT "Job_Portal_Cost" INTEGER "Marketing_Cost" INTEGER "Agency_Cost" INTEGER</pre> |

Figure No. 5.5. Analytical Data Mart

Step 3:- Building Star Schema

The clean data will be used to find the dimension and fact tables. Further star schema datamart will be created.

Step 4:- Historical and Analytical Dashboards

Creating analytical dashboard and historic dashboard will be created.

Step 5:- Model Building

A segmentation model will be created to segment the candidates as per their profiles.

Step 6:- Model Evaluation

Fine Tuning of the model to get a better segments.

Step 7:- Model Deployment

Segmentation model will be deployed.

Chapter 6: Business Understanding

Recruitment is the basic core of any organization. If any organization is able to organize a better recruitment or say human resource system, they will always grow. A better recruitment process backed by proper analytics will solve the problem of poor hire. Any organization wants have a better developer, better analyst, better consultant, better manager etc. And they are ready to pay for the good guys. Even though they don't have surety that the candidate will join the organization after getting selected. In case, they opt out of the offer by organization, they have to go through again rehiring cost and rehiring hire will be involved.

Organization are working on recruitment related problems to smoothen their requirement process. The major consequences of a better requirement process are end up selecting right candidate, less fear of rehire, optimize the cost of recruitment, improve the quality of hire, track the recruiter performance, reducing the time to hire etc. If any organization succeed in establishing a good requirement process, they will get benefited in terms of money, time, quality of their work.

To establish a good requirement process, data plays a vital role. Organizing different source of recruitment data, gathering it at one place, then organizing it into such a way that it can be easily used for any kind of analytical work like prediction or profiling of the candidates. So, the data must be saved in proper accessing format like building a datamart or a star schema.

To guarantee the effectiveness and versatility of the venture Datamart, follow these information distribution centre plan tips:-

1. Characterize the Scope of Datamart

Prior to leaping to the execution period of the undertaking information shop, it's fundamental to have a fool-proof arrangement that thinks about all the business needs and needs of the end clients. Start by laying out the extent of the venture, featuring all dangers and constraints. It'll assist with setting the right assumptions and gauge costs. It might need to change the necessities for the assets (like human, specialized, and monetary assets) to stay aware of the

arranged fruition date. In the light of this degree, foster the rundown of principle expectations and assign obligations to the group.

2. Focus on the Logical Datamart Model

A consistent information store model is a hypothetical, theoretical plan that sorts out information as far as sensible relations known as elements and qualities. A substance is an information thing, though a characteristic characterizes the selectiveness of the element. When planning a consistent model, centre around the business needs. Guide source information to subject-arranged data in the objective information store diagram. The source information model and end-client necessities are the fundamental components used to plan an information shop pattern. It might need to change the actual execution of the consistent information model dependent on the framework boundaries, for example, the computer size, number of administrators, plate stockpiling, network type, and programming.

3. Recognize Relevant Data

For the most part, information components are recognized dependent on the business prerequisites. Be that as it may, it may frequently need to look past the end client asks for and anticipate impending necessities. A decent tip is in the first place the business factors applicable to the branch of knowledge and basic to the specialization. Then, produce a rundown of basic information fields dependent on the requirements set forward by the information shop administrators. For example, a few fields of revenue in the showcasing information store could be item names, advancement attributes, regions, and nations. It ought to likewise isolate the information into numeric measurements (called facts) and illustrative records (called dimensions).

4. Limit Down the Data Sources

Whenever the data has been recorded all measurements and realities expected to plan the information store, the following stage is to recognize the sources that will take care of the storehouse. These sources can incorporate information bases, Excel documents, delimited records, and so on. Then, continue to plan measurements to query tables in the functional framework though, and realities can be planned to exchange tables. It may likewise discover that a portion of the necessary information can't be planned. This regularly happens when fields in the source framework aren't reliable with the necessary information bunches in the information shop.

5. Plan the Star Schema

While making a star composition, it's fundamental to portray the connection between reality and measurement tables. This is finished utilizing keys that incorporate single or various sections, making the column inside a table select. An essential key that incorporates a few segments is known as a composite or connected key. To interface current realities and the measurements, it's great to utilize substitute keys rather than the essential key of the real source table. It permits the information store supervisor to control the keys inside the information shop climate, regardless of whether the keys change in the functional framework. A substitute key is a framework-made series of whole numbers that can be remembered for the measurement table alongside the essential key. It offers more advantages when contrasted with an essential key in light of the fact that the last option is frequently an extensive series of characters. Though an essential key incorporates numbers, so it further develops the inquiry reaction time. By following the five accepted procedures referenced above, it can decrease the odds of mistakes and accelerate the planning system.

Chapter 7: Data Understanding

The dataset is recruitment dataset for an organization. It contains following columns:-

- Candidate_Ref_Id – Unique id assigned to each candidate
- Candidate_Gender – Gender of the candidate
- Candiadate_Age – Age of the candidate
- Rex_in_Yrs – Years of experience
- LOB – Line (category) of Business
- Candiadate_Location – location of the candidate
- Candidate_relocate_actual – Candidate actually relocated or not
- Notice_period - Notice period of the candidate
- Job_ID – Unique Id assigned to each job
- Department – Department of the job
- Job_Title – title of the job
- Candidate_Source – source from which candidate has applied for the job
- DOJ_Extended – joining date extended or not
- Duration_to_accept_offer – number of days taken to accept the job offer
- Offered_band – band offered to the selected candidate
- Prospective_Salary_Band – candidate expectation band
- Percent_hike_expected_in_CTC – candidate expected hike in CTC
- Percent_hike_offered_in_CTC – candidate offered hike in CTC
- Percent_difference_CTC – difference between expected and offered hike of the candidate
- Joining_Bonus – joining bonus given or not
- Job_Status – job offered joined or not
- Job_Status_Code – job offered joined or not code
- Job_Opening_Date – opening date of the job
- Job_Closing_Date – Closing date of the job
- Hiring_Time – time taken for the hiring
- Recruiter – recruiter name

- Job_Portal_Cost – cost of the job portal
- Marketing_Cost – marketing cost to advertise the job
- Agency_Cost – agency cost of the job
- Referral_Cost – referral cost of the job
- Recruitment_Cost – recruitment cost of the job
- Interview_Selection_Cost – cost involved in taking the interview
- Operation_Cost – cost involved in operations like sending bag, bottle etc to the selected candidate
- Rehire_Cost – If selected candidate not joined, cost of rehiring new candidate
- Hiring_Cost – total cost of hiring a candidate
- Recruiter_Performance_Selectionwise – recruiter performance selection wise
- Source_Performance – source of candidate performance
- Interview_to_Hire_Ratio_Job_Id_Wise – ratio of interview taken to candidate selected job id wise
- Candidate_Feedback – feedback from the candidate
- Manager_Feedback – feedback of the manager
- Quality_of_Hire – quality of the hire

Dimension Table

Dimension table the tables of a star schema which store the attributes that describes the object. Dimension Table Example has been shown in Figure No. 7.1.

| | | | | | Dimensions |
|---------------------|-------------------|--------------|--------|----------------|---------------|
| EMPLOYEES | JOB_ROLE | AGE_GROUP | GENDER | DEPARTMENT | JOB_HISTORY |
| EMPLOYEE_ID | EXECUTIVE | LESS_THAN_25 | MALE | HR/RECRUITMENT | EMPLOYEE_ID |
| FIRST_NAME | SR._EXECUTIVE | 25_TO_35 | FEMALE | | START_DATE |
| LAST_NAME | ASSISTANT_MANAGER | 35_TO_50 | OTHERS | | END_DATE |
| EMAIL | MANAGER | 50_TO_60 | | | JOB_ID |
| PHONE_NUMBER | DIRECTOR | | | | DEPARTMENT_ID |
| HIRE_DATE | VP | | | | |
| JOB_ID | CEO | | | | |
| YEARS_OF_EXPERIENCE | | | | | |

Figure No. 7.1. Dimension Table Example

Facts Table

Facts are the numeric dimensions of the business. They support numerical computations used to provide details regarding and investigate the business. Some numeric information is a dimension in camouflage, regardless of whether they appear to be realities. In case it are not

inspired by an outline of a specific thing, the thing may really be a measurement. Data set size and generally, execution will improve if it order marginal fields as measurements. Fact Table Example has been shown in Figure No. 7.2.

| Fact_Table | | | |
|------------------------------|----------------------|---------------|-----------------------|
| REQUIREMENT_EXP DOJ_EXTENDED | DURATION_TO_NOTICE_F | JOINING_BONUS | MATCHING_EXPECTED_CTC |
| NUM_OF_DAYS_FOI YES | 1_DAY | AVAILABL YES | YES |
| NO | 2_DAY | 7_WEEK NO | NO |
| | 3_DAY | 15_DAYS | |
| | 4_DAY | 30_DAYS | |
| | MORE_THAN_4 | 45_DAYS | |
| | | 60_DAYS | |
| | | 90_DAYS | |

Figure No. 7.2. Fact Table Example

Chapter 8: Data Preparation

Data has been cleaned properly using data preparation techniques like checking for missing values, standardization, normalization, categorical data encoding.

There are some assumption which has been taken while preparing the data.

1. All the interview has been conducted online so interview_selection cost is not considered.
2. In Prospective_Salary_Band, salary of E0 has been taken from NTPC site and other following bands are having 50% hike on their previous band.
3. Recruiter salary cost has not been added in the hiring cost due to unavailability of data.

Figure No. 8.1 to 8.5 shows the few results from Exploratory Data Analysis. Details of the figures are mentioned below:-

- Figure No. 8.1. shows the ratio of candidate location wise who has joined the company or not.
- Figure No. 8.2. shows the ratio of candidate who has extended joining date and joined the company or not.
- Figure No. 8.3. shows the ratio of candidate who has actually relocated and joined the company or not.
- Figure No. 8.4. shows the ratio of Male and Female candidates joined and not joined.
- Figure No. 8.5. shows the ratio of people who joined or not joined with different Notice Period.

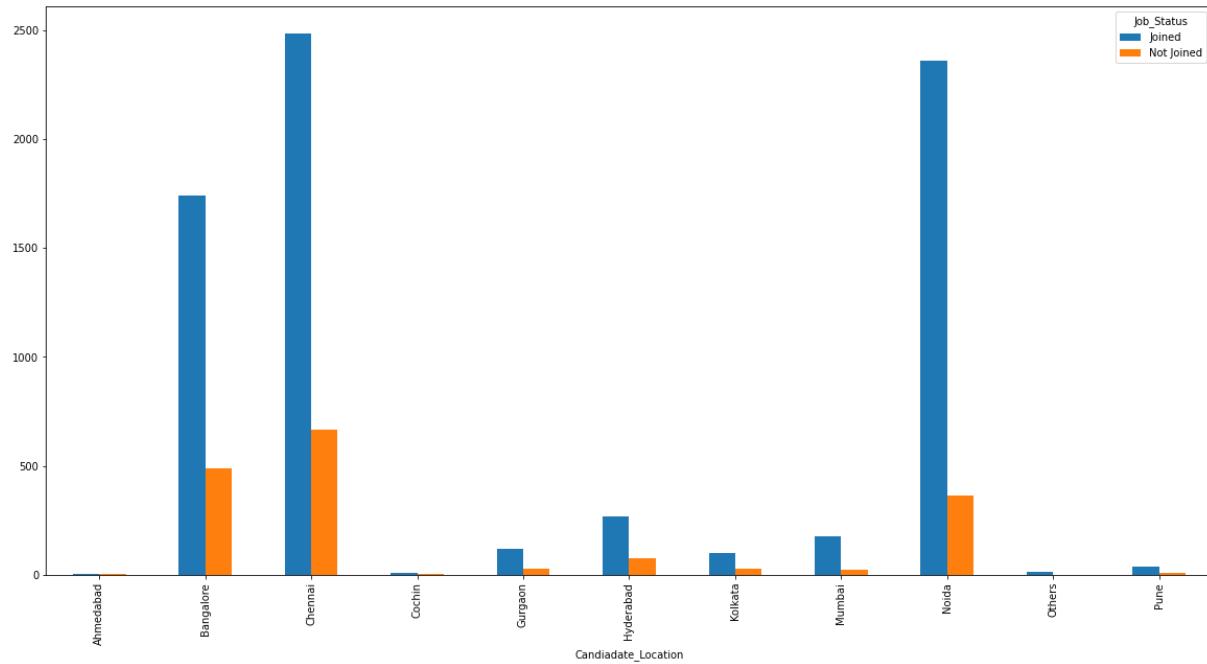


Figure No. 8.1. Location wise candidate joined or not joined

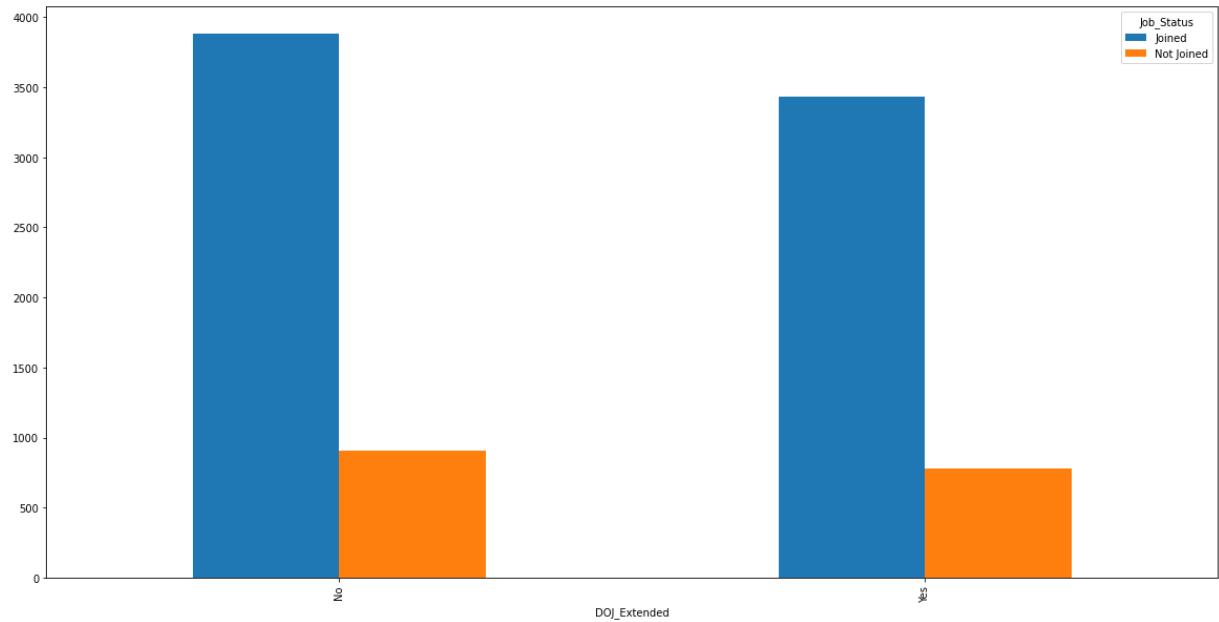


Figure No. 8.2. Candidate who has extended joining date and joined the company or not

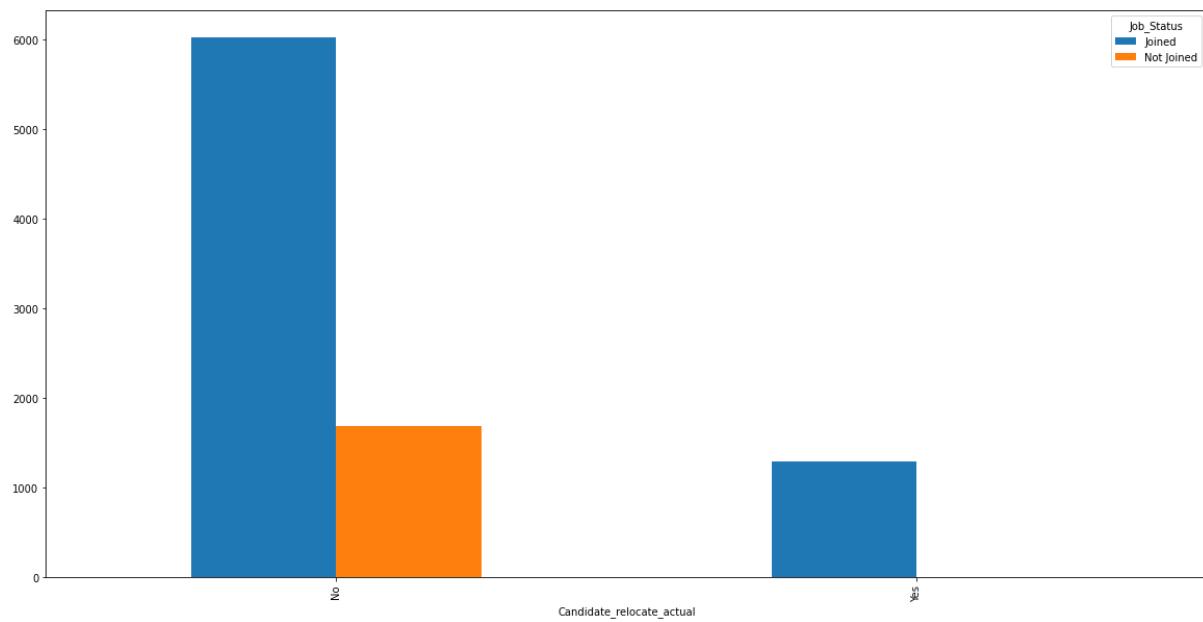


Figure No. 8.3. Candidate who has actually relocated and joined the company or not

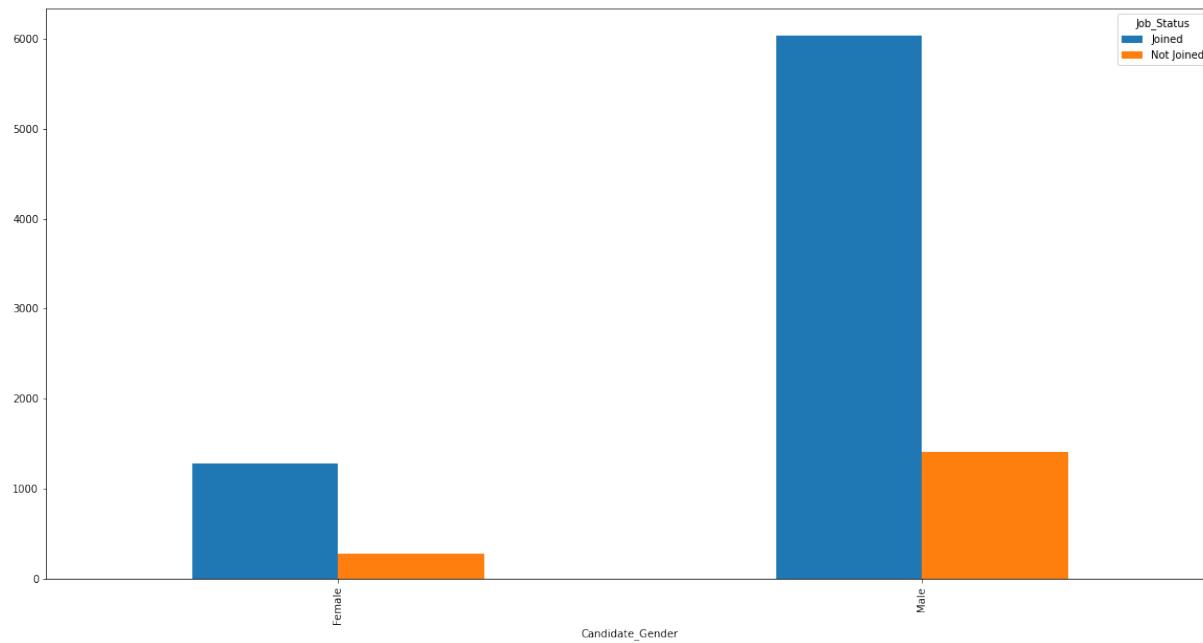


Figure No. 8.4. Ratio of Male and Female candidates joined and not joined

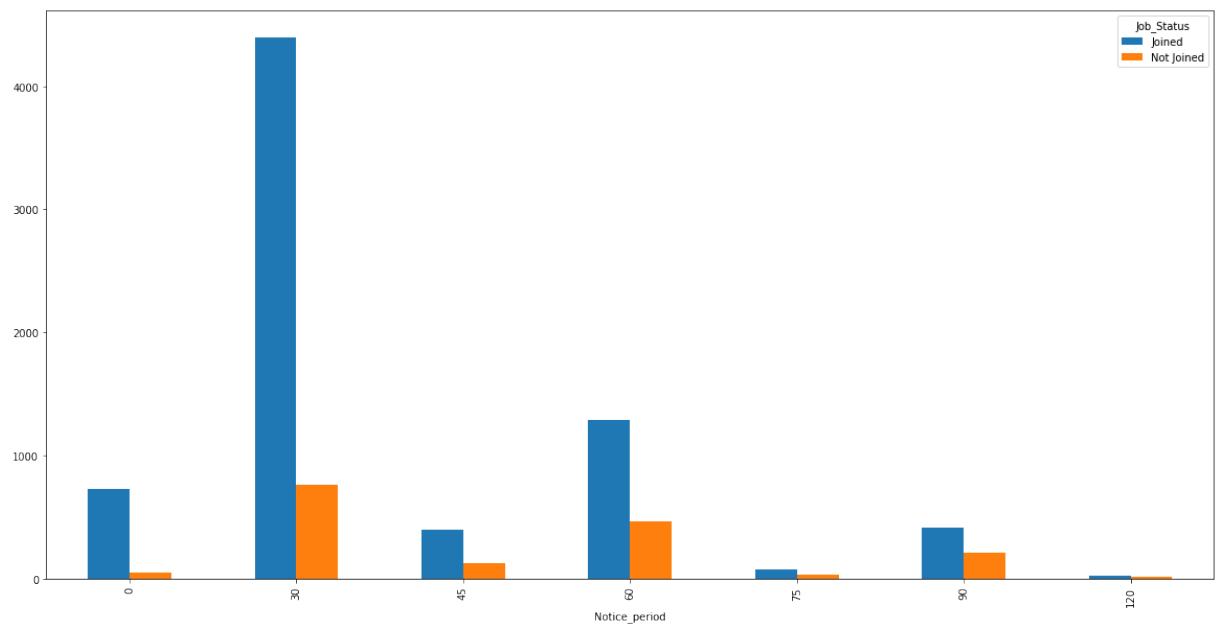


Figure No. 8.5. Ratio of people who joined and different Notice Period

Chapter 9: Modeling

Recruitment Data can be analysed and modelled for 3 popular datamarts namely, KPI Datamart, Analytical Datamart and star schema. These datamarts are mentioned below:-

Star Schema:- Ralph Kimball created the "Dimensional Data Model," sometimes known as the "Star Schema," in the 1980s to address these business objectives. This method of data organization has stood the test of time and is the preferred method for business inquiry and analysis. The Fact and the Dimension are the two main table types in the Star Schema. The Dimension contains classification information, while the Fact contains quantitative measures. With the appearance of a star, each Fact is surrounded by Dimensions that provide context. (Inuwa, 2015). Star Schema Example for recruitment has been shown in Figure No. 9.1.

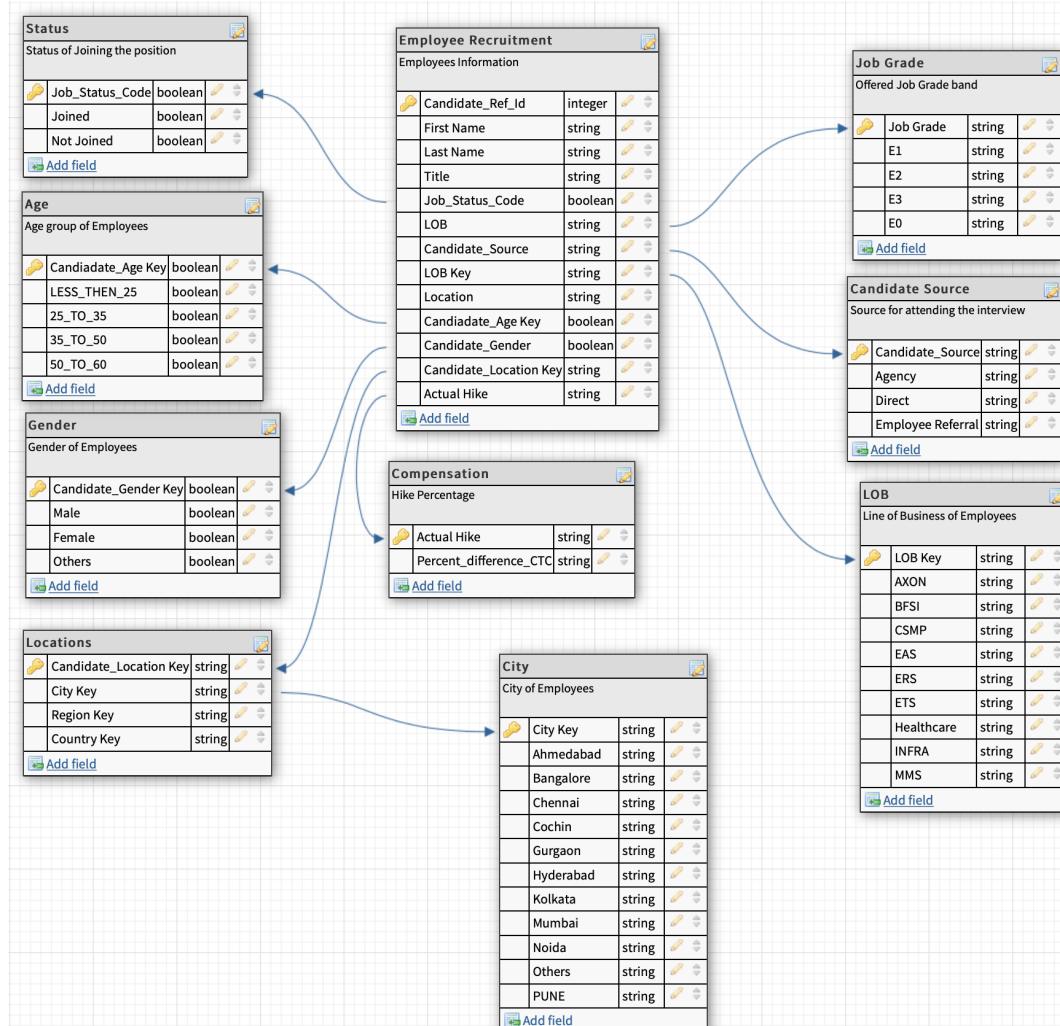


Figure No. 9.1 Star Schema Example

KPI Datamart:- KPI Datamart stands for Key Performance Indicator Datamart. It include tables which will help data engineers and scientists to understand and analyse the key areas of recruitment. Below are few of the tables from KPI datamart for recruitment (Gabcanova, 2012) (Vulpen, n.d.) (Datapine, n.d.):-

- Application Filled Rate:- Number of application filled up for the position
- Qualified Candidate per Opening – Number of resumes matching with the recruitments
- Interview Clearance Rate:- It is the total number of people cleared the interview to total number of people
- Application Completion Rate :- It is the ratio of people completed the form filling to the people seen the opening and filled up the form
- Time to Hire : Average time taken to hire a candidate
- Quality of Hire :- It can be measured after few months of recruitment, mostly based on the matching skill and how the candidate has performed for few months
- Cost Per Hire :- Total cost which includes time taken by the recruiter, tools used by the recruiter, time taken by the hiring manager and salary of the candidate
- Offer Acceptance Rate :- It is the ratio of offer accepted by candidate by total number of offer letter released
- Candidate Satisfaction Rate:- It is the value depends on how the interview process has been conducted and how soon and polite the candidate received the communication.
- Hiring Manager Satisfaction Rate:- It is the value which depends on what quality of resumes or candidates received for the interview.

Analytical Datamart:- Analytical Datamart helps us to create decision management platforms where data scientists can use the analytical DataMart for their analysis. Analytical Datamart Example has been shown in Figure No. 9.2.

- Application Tracking
 - Number of Candidate seen the application
 - Number of candidate filled up the application
- Interview Tracking
 - Number of interview schedule
 - Number of interview rescheduled
 - Number of interview taken
 - Number of interviews completed
 - Number of selections
- Recruitment Source Analytics
 - Application Sourced Directly
 - Application Sourced by Employee Referral
 - Application Sourced by Agency
- Notice Period Tracking
 - Notice period
 - Joined

Figure No. 9.2. Analytical Datamart Example

After data modelling by building Datamart, segmentation has been done using K-Mean Algorithm. To check the number of clusters, elbow method has been used which is shown in Figure No. 9.3.

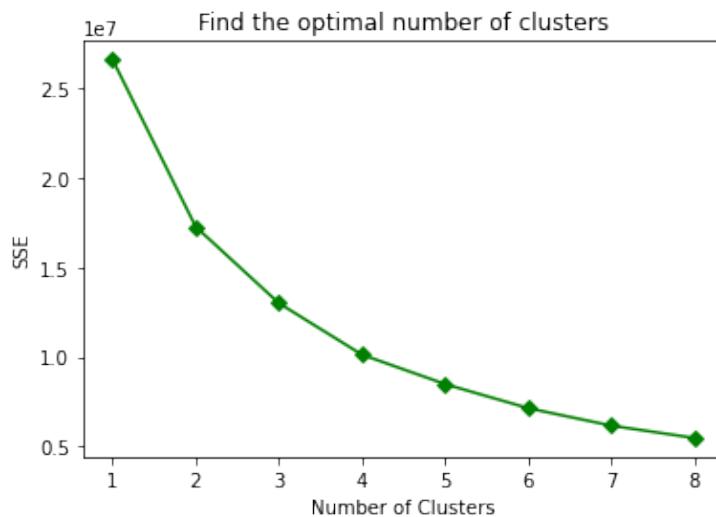


Figure No. 9.3. Elbow Method

Once the data is segmented in clusters, historic and predictive dashboard has been build on data studio tool. The Historic Dashboard and Predictive Dashboard has been shown below in Figure 9.4 and Figure 9.5 respectively. Historic Dashboard shows the historic trend of the recruitment referring to other features like cost of hire, hiring cost and hiring time. Predictive Dashboard shows the prediction trend of the recruitment referring to other features like cost of hire, hiring cost and hiring time by their segments.

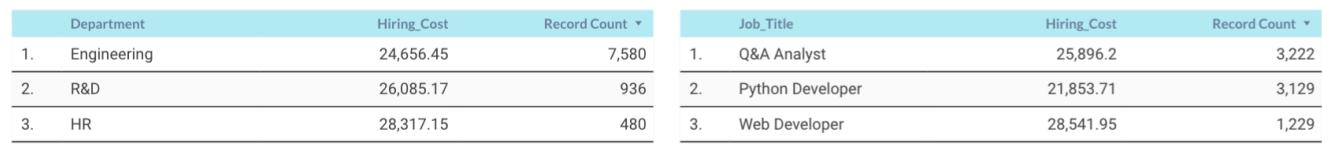
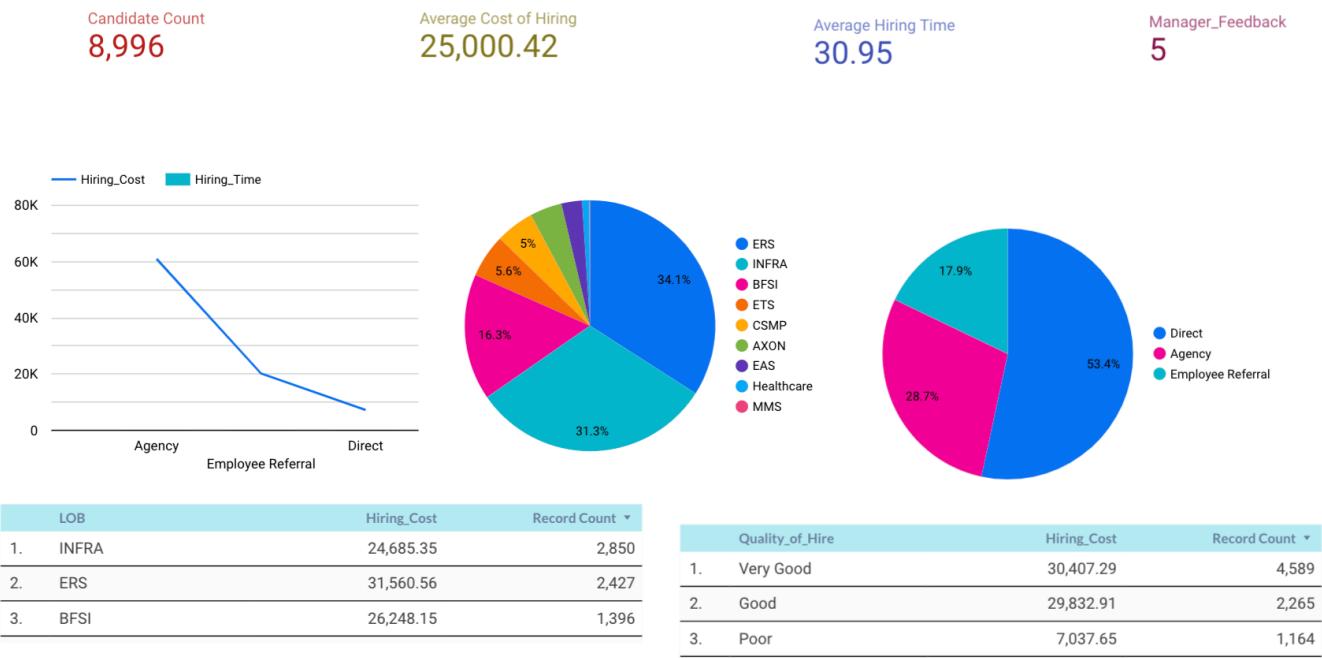


Figure No. 9.4 Historic Dashboard

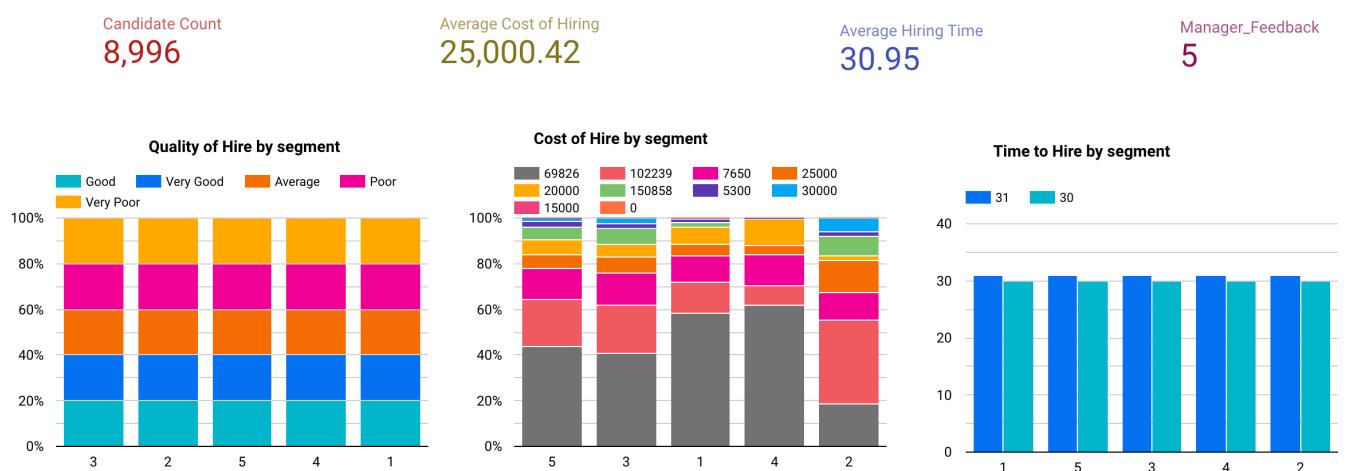


Figure No. 9.5. Predictive Dashboard

Chapter 10: Evaluation

After the datamart is made, it need to make sure if it is fit to work. The certain things it have to keep in mind are -

1. Large-Scale Performance:

Querying and ETL/ELT operations can become tiresome and irritating as data diversity and size expand rapidly. Consider the use-cases while analysing the performance of cloud data warehouses. When it comes to customer-facing analytics, performance may make or destroy the business. However, even in internal BI use cases, the flexibility to perform more queries rather than waiting for data, as well as having fresh data available for querying fast, can be critical to success. Consider this: Would it still use Google if it took a few minutes (or hours) to acquire the search results? The problem is that delayed user experiences cause us to abandon products.

Performance is held back by two major bottlenecks:

Storage bottleneck: Data lakes are designed for indefinite storage, but when vast amounts of data need to be scanned and sent to the compute layer for querying, they are extremely slow.

Compute bottleneck: for today's data sets, century-old data processing techniques are insufficient. Queries that cannot be scaled out have a negative impact on the end-user experience.

RACE Recruitment Datamart is however being tested by a limited set of users.

2. Elasticity

Users of traditional data warehouses will notice that performance is no longer matching their business needs as usage or data grows. Decoupling storage and computing allows for seamless scaling up and down to meet any workload, amount of data, and number of concurrent users. For handling ever-changing resource requirements on demand, the ability to easily scale nodes without expensive and time-consuming re-sharing / re-clustering, vacuuming, fragmentation, and other heavy lifting operations is critical. Different teams can start/stop compute engines for different workloads/tasks like ETL, heavy querying, and exploration while maintaining the performance they require by isolating resources. Choosing

a cloud data warehouse that gives it complete control over-allocating the correct resources to the right tasks can help it save money and have a more efficient and smooth experience.

RACE Recruitment Datamart is quite elastic in nature.

3. Ease of use

Modern cloud data warehouses are improving and becoming far easier to handle than what has been used to, and this should be a major factor in the decision. Efforts should be made to replace time and resources spent on non-productive tasks with time and resources spent on data analysis and development.

Make sure the cloud data warehouse makes the following tasks easier:

- Servers, clusters, installations, and hardware: consider a SaaS platform that handles infrastructure while allowing it to maintain control over the customer experience and prices.
- Performance: Ensure that performance at scale is as efficient as feasible, without the use of pre-aggregations, costly scaling-up methods, or unending complex optimization initiatives.
- It want to be able to support any data type, including semi-structured data. Ascertain that the data warehouse supports semi-structured data and SQL queries. Semi-structured data should be evaluated quickly and easily, without the use of time-consuming ETL operations that flatten and increase the size and expense of data sets.
- Updates and deletions: make it easier to handle merging data updates without having to rewrite tables.

RACE Recruitment Datamart is quite easy to use as it has been designed with reference of other popular datamarts.

4. Cost-effectiveness

Understanding cloud data warehouse pricing models are complicated because they are based on a variety of factors such as speed, scale, and usage.

RACE Recruitment Datamart is very cost effective, since most of the work has been done using open sources

5. It can handle both structured and semi-structured data

Data no longer comes in predictable and structured formats; information instead comes in a variety of formats and from a variety of sources.

Although semi-structured data can improve the analytics, most typical data warehouses aren't designed to handle it. Flattening/unnesting/exploding, which doubles the number of rows with the number of cells in the arrays, wastes time and money. As a result, it'll have a considerably larger table, a lot more useless expenditures, and much worse performance. With a modern datamart, it may query semi-structured data using ordinary SQL rather than expensive ETL operations that flatten and blow up data set sizes and prices. This can be accomplished using native array manipulation techniques while maintaining speed and efficiency.

RACE Recruitment Datamart currently accepts the csv file. Unstructured data is in future scope.

6. Concurrency

Concurrency limits are imposed by some vendors, who enable customers to submit only one query at a time and limit the number of concurrent inquiries per account. Furthermore, even without restrictions, concurrency can degrade performance, which gets me back to my earlier argument regarding flexibility. It's critical to be able to quickly deploy resources to support business expansion without sacrificing performance.

Consider a cloud data warehouse that does not impose concurrent limits or performance penalties. Also, keep in mind that the company will expand. The requirements today are not the same as the requirements tomorrow.

RACE Recruitment Datamart has not been tested with multiple users working at the same time yet. Although concurrency will be taken care in near future.

Chapter 11: Deployment

The Datamart has been designed on dbdesigner which generates the SQL code which can be further developed in any database tool like Microsoft SQL server management studio or oracle database or can be built on any cloud like Amazon Web Services, Google Cloud Platform, Microsoft Azure. Various dashboards like historic and predictive dashboards have been made on Datastudio which is a free tool from Google. Alternately one can use Power BI or Tableau as a substitute of Datastudio. For segmentation (modeling), K-Means algorithm has been used to build the model. The model has not been deployed as of now.

Chapter 12: Analysis and Results

A datamart is the access layer of a data warehouse used to offer data to users in business intelligence. Small chunks of the data warehouse are often referred to as datamart. Data warehouses typically hold enterprise-wide data, whereas datamart typically store data unique to a department or team. The primary goal of datamart is to offer the most relevant data for BI to the business user in the shortest time feasible. Millions of records can be stored in a datamart, which necessitates terabytes of storage. The following are some of the benefits of employing a datamart:

- Allows users to have access to the precise type of data they require, which improves end-user response time.
- A data warehouse that has been compacted and narrowed in scope.
- Each one is devoted to a single item or function.
- Implementing a full data warehouse is more expensive.
- Contains a lot of information.
- It is less crowded and simply contains necessary company information and data.
- Assists in the integration of all data sources

The building and use of a datamart result in a significant amount of data summarization.

In this project, a datamart has been build up referring to Kimball HR datamart and Oracle recruitment datamart. A star schema with facts and dimension tables has been built up to develop KPI datamart and analytical datamart. Further, segmentation model has been created using K-Mean algorithm to cluster the similar kind of candidates. As per elbow method, five clusters has been created to segment the candidates. Later, historic and predictive dashboard has been developed to analyse the candidate profiles in details.

Chapter 13: Conclusions and Future Scope

Different human resource Datamart like Kimball HR Datamart, Oracle HR Datamart, SAP-HANA Datamart has been studied and inferred. Based on inferences from popular Datamart, industry expert experience and through research done, a new Datamart for recruitment has been created. This Datamart will be a good option to start building the recruitment Datamart. The K-Mean algorithm has been used to build a model on recruitment data. Candidates are segmented under 5 classes with different features. Historic and predictive analytical dashboard has been created on recruitment dataset.

Future work will be extending recruitment process to other human resource process like Employee Lifecycle, Employee Exit (Retention) etc. Will be creating Datamart, create dashboards and build model for the same process.

Bibliography

- Alexandra. (2019, August 27). *25 Practical Tips For Building And Managing A Solid Recruitment Funnel*. Retrieved from <https://harver.com/blog/recruitment-funnel/>
- George, S. (2012, April 14). *Inmon or Kimball: Which approach is suitable for the data warehouse*. Retrieved from <https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-the-data-warehouse>
- Ralph Kimball, M. R. (2016). Data Warehousing, Business Intelligence, and Dimensional Modeling Primer. In *Kimball The Data Warehouse Toolkit 3rd Edition* (pp. 7-16). Wiley.
- Ralph Kimball, M. R. (2016). Employees Tracking for Profile. In *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition* (pp. 263-265). John Wiley & Sons, Inc.
- Oracle. (n.d.). *Recruiting Datamart Dimension Tables*. Retrieved from https://docs.oracle.com/cd/E41507_01/epm91pbr3/eng/epm/phcw/concept_HCMWarehouseStructure-399b81.html
- SAP. (2018, 1 24). *SAP HANA as Datamart*. Retrieved from https://help.sap.com/doc/e95f6750b0fd10148ea5c6be75016694/1.0.12/en-US/SAP_HANA_Master_Guide_en.pdf
- Gartner. (2019, 06). *Gartner Identifies Three Most Common AI Use Cases in HR and Recruiting*. Retrieved from <https://www.gartner.com/en/newsroom/press-releases/2019-06-19-gartner-identifies-three-most-common-ai-use-cases-in-hr-and-recruiting>
- tttt. (n.d.). Retrieved from Gartner Identifies Three Most Common AI Use Cases in HR and Recruiting
- google. (n.d.). Retrieved from www.google.com
- Alysson Prado, C. F. (2010). Using OLAP Tools for e-HRM: A case study. *International Journal of Technology and Human Interaction* .
- Udhay Kailash, M. P. (2020). HR Analytics Methodical Measurement of HR Processes. *International Journal of Innovative Science and Research Technology* .
- Alaa Khalaf Hamoud, M. A. (2020). Improve HR Decision-Making Based On Datamart and OLAP. *Journal of Physics*.
- Inuwa, I. (2015). Design of a Data Warehouse Model for a University Decision Support System. *Journal of Information & Knowledge Management* , 5.
- Gabanova, I. (2012). Human Resources Key Performance Indicators. *Journal of Competitiveness* .
- Oracle. (2020). *Installing Human Resources Schema on Autonomous Database*. Retrieved from https://docs.oracle.com/en/cloud/paas/autonomous-data-warehouse-cloud/dw_hr_analytics/run-analytics/run-analytics.html#setting-up-the-human-resource-analytics-project-in-oac
- Vulpen, E. v. (n.d.). *Human Resources KPIs: An In-depth Explanation with Metrics & Examples*. Retrieved from <https://www.aihr.com/blog/human-resources-key-performance-indicators-hr-kpis/>
- Datapine. (n.d.). *19 KPIs THAT EVERY HR MANAGER SHOULD USE*. Retrieved from <https://www.datapine.com/kpi-examples-and-templates/human-resources>

Appendix

Plagiarism Report¹

Development of Analytical Datamart and Data Pipeline for Recruitment Analytics

by Ashish Chandra Jha

Submission date: 26-Aug-2022 06:50PM (UTC+0530)

Submission ID: 1887440805

File name: nd_Data_Pipeline_for_Recruitment_Analytics_AshishChandraJha.docx (6.51M)

Word count: 7004

Character count: 38359

¹ Turnitin report to be attached from the University.

Development of Analytical Datamart and Data Pipeline for Recruitment Analytics

ORIGINALITY REPORT



PRIMARY SOURCES

| | | |
|---|---|------|
| 1 | www.firebolt.io Internet Source | 1 % |
| 2 | Submitted to Academy of Information Technology Student Paper | 1 % |
| 3 | Submitted to Trident University International Student Paper | 1 % |
| 4 | Submitted to De Montfort University Student Paper | <1 % |
| 5 | Submitted to La Trobe University Student Paper | <1 % |
| 6 | www.researchgate.net Internet Source | <1 % |
| 7 | Submitted to University of Greenwich Student Paper | <1 % |
| 8 | docplayer.net Internet Source | <1 % |
| Submitted to University of the Western Cape | | |

9

Student Paper

<1 %

10

Submitted to S.P. Jain Institute of
Management and Research, Mumbai

Student Paper

<1 %

Exclude quotes On

Exclude matches < 10 words

Exclude bibliography On

Publications in a Journal/Conference Presented/White Paper²



भारतीय प्रबंध संस्थान बैंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE



Paper Presentation

This is to certify that the paper titled

RESTOQ - ASPECT BASED SENTIMENT ANALYSIS

authored by

**ASHISH CHANDRA JHA, LALIT AGARWAL &
SANJEEV KUMAR JHA**

was presented at the

"Seventh International Conference on Business Analytics and Intelligence"

5 - 7 December, 2019

A handwritten signature in black ink, appearing to read "U Dinesh Kumar".

U Dinesh Kumar
Conference Chair

INDIAN INSTITUTE OF MANAGEMENT BANGALORE, BANNERGHATTA ROAD, BANGALORE 560076, INDIA

Any Additional Details

² URL of the white paper/Paper published in a Journal/Paper presented in a Conference/Certificates to be provided.

RestoQ – Aspect Based Sentiment Analysis

Ashish Chandra Jha ^{1,†}
REVA University
Bangalore, India
Ashish.ba05@rev.edu.in

Lalit Agarwal
REVA University
Bangalore, India
Lalit.ba05@rev.edu.in

Sanjeev Kumar Jha ^{2,†}
REVA University
Bangalore, India
Sanjeev.ba05@rev.edu.in

†A.C.J. and S.K.J. contributed equally to this work.

Abstract – People love experimenting on their food with different tastes. And when it's about visiting the restaurant or ordering food online, they will definitely look for the reviews which will talk about the aspects like services, ambience and cost along with food quality. This food problem is not a single day problem, it's getting repeated everyday. Some end up with positive reviews and few end up with negative or neutral reviews. In this work a framework is developed called 'RestoQ', which uses text analytics for sentiment analysis at the aspect level to discover and rank the restaurants. The framework analyzes the reviews for the sentiments across four aspects – price, food quality, service quality and ambience. Unsupervised lexicon-based classifier and a naïve Bayesian classifier are used to evaluate and score the sentiments at aspect level. The final score will be a combined sum of each score for the review, which requires further work rank the aspects based on reviews. Surprisingly unsupervised method out performs the supervised method. It is proposed to extend the work with context based methods using word2vect and LSTM.

I. INTRODUCTION

Before ordering food or booking a table in any restaurant consumers generally check the reviews of the places. Online food ordering sites like Zomato, Food Panda and UberEats do sentiment analysis of the reviews given by the customers and give a rating for these restaurants. The majority of current sentiment analysis approaches try to detect the overall polarity of the reviews or sentence regardless of the target entities (e.g. restaurants) and their aspects (e.g. services, ambience and cost along with food quality). Aspect Based Sentiment Analysis is fine grained sentiment analysis. A sentence may contain multiple opinions about different entities and we need to find each of them. This has to be analysed by model and should give insights. In this work, the research findings of such a system are presented.

II. LITERATURE SURVEY

Sentiment analysis is one of the fastest growing research areas in computer science, making it challenging to keep track of all the activities in the area. It is a case of natural language processing which could mark the emotion or mood of the people about any specific product by analysis. It is a process of automatic extraction of features by mode of notions of others about specific product, services or experience. [1]

Customers as well as ecommerce companies (online food ordering in this case) are looking for the reviews of the restaurants to order food or to check their customer satisfaction ratio. A lot of research has been done on Sentiment analysis on restaurants and their reviews. Reviews are considered to be positive, negative or neutral on the overall score of the sentence. To some extent it is very useful and many customers are using it before ordering their food on daily basis [2][3].

Unlike document level sentiment classification task, aspect based sentiment analysis is a more fine-grained classification task. It aims at identifying the sentiment polarity (e.g. positive, negative and neutral) of one specific

aspect in its context sentence. For example, given a sentence "great food but the service was dreadful" the sentiment polarity for aspects "food" and "service" are positive and negative respectively [4].

Aspect Based Sentiment Analysis (ABSA) was introduced as a shared task for the first time in the context of SemEval in 2014; SemEval2014 Task 41 (SE-ABSA14) provided datasets of English reviews annotated at the sentence level with aspect terms (e.g., "mouse", "pizza") and their polarity for the laptop and restaurant domains, as well as coarser aspect categories (e.g., "food") and their polarity only for restaurants (Pontiki et al., 2014). SemEval-2015 Task 122 (SE-ABSA15) built upon SE-ABSA14 and consolidated its subtasks into a unified framework in which all the identified constituents of the expressed opinions (i.e., aspects, opinion target expressions and sentiment polarities) meet a set of guidelines and are linked to each other within sentence-level tuples (Pontiki et al., 2015) [5][6][7].

Aspect Based Sentiment Analysis poses several challenges in processing text data, is a popular area of research in this direction. Several challenges which has not been addressed and people are trying to do some research are implicit aspect detection, mapping aspect words to categories, resolving anaphora references etc. Researchers combine techniques from common sense rules, unsupervised supervised and semi supervised techniques to perform these tasks.

Aspect Based Sentiment Analysis has been done for this particular topic by various researchers [8] [9] [10]. In this paper, the four aspects depending on which the comments will be reviewed. It has been seen that there are lots of aspects which affects the overall sentiment of the review. For example, in restaurants, people give review based on food quality, services, ambience and price. In this work restaurants will categorized based on the customer reviews. The goal is to determine the sentiment expressed toward each aspect on restaurant of Bangalore in English language.

The problem of aspect-based sentiment analysis deals with classifying sentiments (negative, neutral, positive) for a given aspect in a sentence. A traditional sentiment classification task involves treating the entire sentence as a text document and classifying sentiments based on all the words [11].

Labeling of data is a little difficult task to perform automatically. Most of the researcher who are working on new dataset used to label the data manually. The lack of labeled data has led to several researchers to explore unsupervised learning techniques to learn both aspects and their sentiments expressed in plain text. Particularly the fact that aspects are normally described by opinion words and opinion words in turn will have a target aspect can be used to iteratively expand the sentiment and aspect lexicon. The expansion is done with the help of rules to associate aspects and sentiment [12][13].

In this paper, we are trying to do aspect based sentiment analysis on restaurant reviews data from an online food delivery site (Zomato). We have introduce a system based on Text Analytics on the reviews using Supervised Machine Learning with a Naïve Bayes algorithm and unsupervised Machine Learning with Lexicon based algorithm for scoring sentiments

III. RESEARCH METHODOLOGY

In this section, we will explore the different techniques, methods, and features used in this experiment. We will divide the section into two sections: data exploration and pre-processing and model building. Model building is further divided into supervised ML and Unsupervised ML.

Data access: 2000 restaurant across Bangalore along with their reviews has been collected. The data is from an online food ordering company i.e. Zomato. Labeling of data is the hard part of any new research and is done manually. Here the reviews has been labeled based on restaurants name, aspects and sentiment. Positive, negative and neutral sentiments have been used as the three classes.

Data Exploration: Data has the solution to every problem. But one must know how to use that data. Data exploration gives the ability to summarize the main characteristics of a data set, including its size, accuracy, initial patterns, null values, outlier values and missing values. It can use a combination of manual methods and automated tools such as data visualizations, charts, and initial reports to explore the data.

Data Preprocessing: It is the most vital part of any analysis. Considering few important preprocessing steps, below mentioned techniques has been used

- Stopwords Removal - Stopwords are the meaningless and repeated words which do not contribute to the semantic of the statement. It should be removed.
- Symbol Removal – Reviews generally contains symbols like @,#,\$ with no contribution towards analyzing the sentiments. So, it should be removed.
- Contractions and Annotation Removal – The contractions and annotation like shouldn't should be removed with 'should not'
- Normalization - Normalization stands for making the word or sentence case insensitive. Data should be normalized.
- Exploration – It is to check the word frequency of the corpus. It gives the idea of what the document is about. We check the word frequency by TF-IDF model. Words with high frequency can be seen using word cloud. In addition conditional exploration, based on sentiments and aspects word cloud has been made.
- For sentiments three word cloud, one for positive, one negative and one for neutral emotions. For aspect four word cloud has been made based on food quality, services, ambience and cost.

Model Building: The model will be created four times with different strategies. Here combination of 90%-10%, 80%-20%, 70%-30% and 60%-40% train/test split along with 10 fold cross validation has been used.

Starting with the supervised learning model then tried unsupervised learning model has been created to compare which algorithm will perform better.

In Supervised Machine Learning stage, label data is used for building classifier using Naive Bayes algorithm. Naïve Bayesian algorithm is a probabilistic ML algorithm, which assume independence among the features.

In Unsupervised Machine Learning stage, label data is used for building classifier using Lexicon based algorithm. The lexicon based approach is based on the assumption that the contextual sentiment orientation is the sum of the sentiment orientation of each word or phrase.

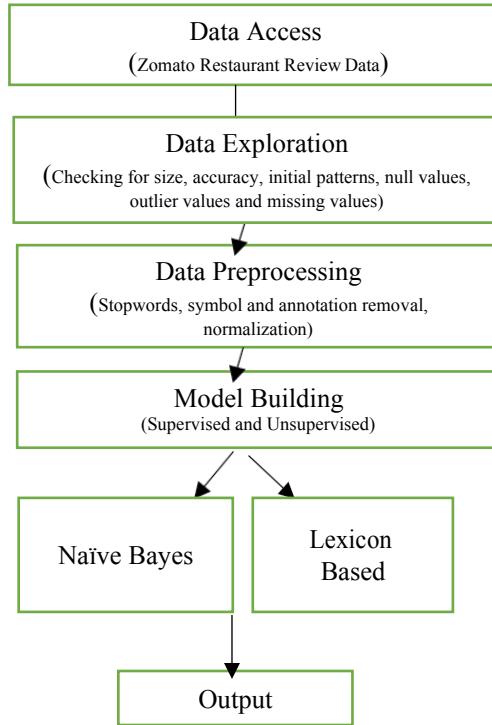


FIG.1: PROPOSED APPROACH

IV. RESULTS AND DISCUSSION

Supervised (Lexicon based model) and Unsupervised (Naive Bayes model) has been created at different train test split. For each instant, accuracy has been captured and report is mentioned below:

Table 1. Experimental results at review level

| | | Accuracy | |
|----------|------|----------|-------------|
| Training | Test | Lexicon | Naive Bayes |
| 90% | 10% | 71% | 72% |
| 80% | 20% | 68% | 65% |
| 70% | 30% | 70% | 65% |
| 60% | 40% | 68% | 63% |

The classification accuracy of all the models are consistent with the results published in literature and hence support the methodology used in this research. From the result, it clearly shows, the unsupervised learning lexicon-based model performs better than supervised learning technique using Naive Bayes. Accuracy at different train test split given an idea of optimum split scoring highest accuracy.

V. CONCLUSIONS

This paper covers the Aspect Based Sentiment Analysis on restaurants reviews dataset for Bangalore restaurants. The ABSA task consists of four aspects namely food quality, services, ambience and cost. For each aspect, sentiments have been analyzed. Supervised and Unsupervised machine learning has been used.

The proposed approaches achieved very good results. The algorithm successfully able to analyze the aspects of the sentiments. Further the restaurants are ranked based on the over all score and the positive score, which can be used by consumers for selection of restaurants. It is proposed to carry out a context-based analysis of the sentiments using word2vec and LSTM to test the improvements in the accuracies.

REFERENCES

- [1] "A Study on Sentiment Analysis: Methods and Tools by Abhishek Kaushik1, Anchal Kaushik2, Sudhanshu Naithani3 <https://pdfs.semanticscholar.org/c151/dfad8c1bf88b0afc716758c77d533ded7dd0.pdf>"
- [2] "Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews <https://arxiv.org/ftp/arxiv/papers/1709/1709.08698.pdf>"
- [3] "SENTIMENT ANALYSIS OF RESTAURANT REVIEWS USING HYBRID CLASSIFICATION METHOD

[http://iraj.in/journal/journal_file/journal_pdf/4-54-140014488817-23.pdf"](http://iraj.in/journal/journal_file/journal_pdf/4-54-140014488817-23.pdf)

- [4] "Sethia, A., & Bhattacharyya, P. Aspect Based Sentiment Analysis-A Survey. Accessed on August 2019; http://www.cfilt.iitb.ac.in/resources/surveys/aspect-based-sentiment-analysis_survey.pdf"
- [5] "Bhoi, A., & Joshi, S. (2018). Various Approaches to Aspect-based Sentiment Analysis. arXiv preprint arXiv:1805.01984"
- [6] "Brychcín, T., Konkol, M., & Steinberger, J. (2014, August). Uwb: Machine learning approach to aspect-based sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 817-822) "<http://alt.qcri.org/semeval2014/cdrom/pdf/SemEval2014145.pdf>
- [7] "Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks Binxuan Huang, Yanglan Ou and Kathleen M. Carley <https://arxiv.org/pdf/1804.06536.pdf>"
- [8] 8."Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., & Hoste, V. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30). <https://www.aclweb.org/anthology/S15-2082>"
- [9] "SemEval-2016 Task 5: Aspect Based Sentiment Analysis Maria Pontiki*, Dimitrios Galanis1 , Haris Papageorgiou1 , Ion Androutsopoulos1,2 , Suresh Manandhar3 , Mohammad AL-Smadi4 , Mahmoud Al-Ayyoub4 , Yanyan Zhao5 , Bing Qin5 , Orphée De Clercq6 , Véronique Hoste6 , Marianna Apidianaki7 , Xavier Tannier7 , Natalia Loukachevitch8 , Evgeny Kotelnikov9 , Nuria Bel10 , Salud María Jiménez-Zafra11 , Gülsen Eryiğit12 <https://www.aclweb.org/anthology/S16-1002.pdf>"
- [10] "SemEval-2015 Task 12: Aspect Based Sentiment Analysis Maria Pontiki*, Dimitrios Galanis*, Haris Papageorgiou*, Suresh Manandhar*, Ion Androutsopoulos* <https://www.aclweb.org/anthology/S15-2082.pdf>"
- [11] "Various Approaches to Aspect-based Sentiment Analysis Amlaan Bhoi Department of Computer Science University of Illinois at Chicago Chicago, IL, USA abhoi3@uic.edu Sandeep Joshi Department of Computer Science University of Illinois at Chicago Chicago, IL, USA sjoshi37@uic.edu . <https://arxiv.org/pdf/1805.01984.pdf>"
- [12] "http://UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis Toma's Brychcín NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 306 14 Plzeň Czech Republic brychcin@kiv.zcu.cz Michal Konkol NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 306 14 Plzeň Czech Republic konkol@kiv.zcu.cz Josef Steinberger Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 306 14 Plzeň Czech Republic jstein@kiv.zcu.cz <https://www.aclweb.org/anthology/S14-2145.pdf>"
- [13] "Semi-supervised Aspect Based Sentiment Analysis for Movies using Review Filtering Deepa Ananda Deepan Naorema <https://www.sciencedirect.com/science/article/pii/S1877050916300850>"

Development of Analytical DataMart and Data Pipeline for Recruitment Analytics

1st Ashish Chandra Jha
Business Analysis Scholar
REVA University
Bengaluru, India
ashish.ba05@rev.edu.in

2nd Sanjeev Kumar Jha
Business Analysis Scholar
REVA University
Bengaluru, India
sanjeev.ba05@rev.edu.in

3rd Dr. J. B. Simha
Business Analysis Chief Mentor
REVA University
Bengaluru, India
jb.simha@rev.edu.in

Abstract—The HR department handles all the data regarding the recruitment process while also analyzing them to select suitable candidates for the organization. The HR department interacts with the data regarding the recruitment process using many tools for data analysis and interpretation. Such data has to be organized in a way to ensure proper analysis as well as the screening of candidates.

Each system gives their own report in their own format. Data is not integrated and aggregated at proper granularity suitable for analysis. This can be mitigated by the use of Dimensional DataMarts and analytical DataMarts.

The dimensional DataMarts will contain transitional data modified for analysis with dimensions and facts. These DataMarts can be used for any adhoc analysis like drill down/roll up, slice and dice, drill through, comparative analysis etc.

The analytical DataMart will contain the aggregated data in one row per employee format. This will be task specific in the sense that the standard tabular structure/table containing multiple attributes/independent variables for a specific problem like Quality of Hire modeling, Cost of Hire, Time to Hire, demand prediction etc.

This project has focused on creating a recruitment DataMart for the Human Resource Department with the assistance of the data modeling technique. This Project has also discussed the exploratory data analysis and model building in relation to the DataMarts for the recruitment.

Keywords— *HR Analytics, Recruitment, DataMart, KPI, KPI DataMart, Analytical DataMart*

I. INTRODUCTION

The success and growth of any organization primarily depend on its employees. Without employees, no organization can run as they are the workforce for the organization. Various employees take up different roles in the organization and perform the duties assigned to their positions. The employees of an organization are crucial for the organization and therefore the department that majorly deals with them becomes crucial as well. This department is the Human Resource Department of an organization. The HR department is responsible for almost all aspects concerning the workforce of an organization. The importance of employees for an organization makes the HR department very important. The Human Resource Department is responsible for the effective administration of the employees of an organization to help in its overall growth. Moreover, an organization can only succeed if competent people are recruited in it. This

department needs to carefully examine and select suitable persons for the various roles in the organization.

A. Recruitment Process

Recruitment can be defined as the overall process of choosing the right candidates for job positions at an organization based on several criteria. The recruitment process is usually handled primarily by the Human Resource department of an organization. This process might differ for every organization but there are certain general steps involved in it that most organizations seem to follow. These are Job Description Preparation, Sourcing, Screening, Selecting, Hiring, and Onboarding.

1) JD Preparation

Job description refers to the details of the job position for which the recruitment is taking place. It contains all the necessary information regarding the job such as-

- Qualifications required for the job
- Skills required to perform the necessary functions of the job
- Different responsibilities that the selected candidate has to take care of
- Suitable characteristics for the candidates for the job such as age, experience, etc.
- Salary expectations concerning the job
- Working hours

The job description is an important aspect of the recruitment process as it notifies the interested candidates with all the necessary information about the job. It is the primary source of attracting candidates for the job position.

2) Sourcing

Sourcing is the process of collection of data relevant to the position for which recruitment is being done. In other words, this process involves the collection of resumes or data of various candidates who are suitable for the concerned position. The HR department scrambles through multiple portals for data collection such as Naukri.com, LinkedIn, monsterindia.com, etc. This process does not involve the selection of the candidates, it only involves the collection of the information of the relevant candidates for further screening and selection.

3) Screening

Screening is the first assessment of the candidates with the necessary qualifications for the job position. All the resumes and data that are collected in the sourcing process are

reviewed and screened by the HR department. This stage finds out those candidates who will be advancing to the next level of selection. Screening the resumes of candidates and selecting the top ones from the whole lot is the usual activity in this process.

4) Selecting

Selecting is the main stage of choosing the most suitable candidate for the job for which the recruitment is being done. Those candidates who passed the screening stage have to go through this phase of selection which might involve multiple tests. Personal tests, group discussions, aptitude tests are some of the most common tests that are used to select the most suitable candidate for the job post. The candidate who is selected also gets to negotiate his/her salary with the HR department in this phase itself. The best candidate is selected for the job and no more candidates are assessed after this selection stage.

5) Hiring

The candidate who is finally selected in the selecting stage is hired by the company. This occurs in this phase called hiring. The concerned candidate is presented with the offer letter from the organization's HR department. All the terms and conditions of the job along with other necessary information regarding the job are presented to the chosen candidate at this stage. If the organization and the chosen candidate agree on all the terms and conditions set forward, the candidate is hired into the organization.

6) Onboarding

Onboarding is the last phase in the recruitment process and by this stage, a suitable candidate is already selected and hired for the concerned job position. In this phase, the new employee is welcomed into the organization by the HR department. Onboarding is the process of introducing the new employee to the organization, its cultures, rules & regulations, his/her role, other employees, and the way of working. It is a way of familiarizing the new employee about the various aspects of the organization and the job role.

HR Analytics is defined as a HR process which uses the HR data (e.g. recruitment data) and domain knowledge to do start predicting the performance and cost involved of the people on the basis of their work. [1]

II. LITERATE REVIEW

According to the Gartner, human resource leader has begun analysing the HR data at each level to improve the HR efficiency, recruitment efficiency and enhance employee experience [2]. In the fast growing world with their emerging new challenges, it is very difficult to hire skilled people which will be in the organization for a decent span of time. To create such system, we need good quality of data backed by domain knowledge to provide analytical suggestion to hire a candidate. To create such analytical system, the business data should be extensively available to the hiring manager to work align to find the suitable candidate [3].

For a smooth functioning of an organization, their strategic decision should be backed by data. Data helps us to find the insights or patterns among the people or work. This data should be properly created and organized in such a way that it can be further used to perform any analytical tasks. The success of human resource department with the help of analytics can be measured as the success of the organization. HR department tries to understand the business requirement

and ensure the finding of right candidate according to their skills and qualifications. To hire such a niche level people we need to create a precise DataMart and use it's data for analytical view. Human resource (HR) DataMart is the base stone for building an enterprise data warehouse. The paper presents the implementation process of HR DataMart starting from implementing DataMart schema to online analytical processing (OLAP) reports [4].

There are quite a few DataMarts available in the market these days. Among them, quite popular are the Kim Kimball HR DataMart which is the base for many available DataMarts, Oracle HR DataMart, SAP HR DataMart.

A. HR DataMart by Ralph Kimball

The Kimball HR DataMart is based on the ideology of Ralph Kimball according to whom the data warehouses should be model using dimensional models such as the star schema or snowflake schema. Kimball's approach was the bottom-up approach that involved creating DataMarts first to allow quick analysis and interpretation of the data in them. These small DataMarts can later be combined to form a normal data warehouse. Kimball focused on the use of dimensional models to improve the performance of the users of the data warehouse [5].

In order to understand Kimball's HR DataMart, we have to look at the star schema dimension model. A star schema is a tool for dimensional modeling of data by organizing it to allow analytical operations to run on it. The star schema presents the data in the form of fact tables and dimension tables. The fact tables contain the primary data regarding the business process, in our case, the recruitment data, which are termed as 'facts'. The other tables or dimension tables contain the data associated with the primary fact tables. The dimension table contains 'dimensions' which sort of describe the data in the facts table.

The dimensional tables surround the fact table depicting the descriptive nature of the facts. The multiple dimension tables around the fact table make this model look something like a star due to which it is known as the star schema.

Kimball's HR DataMart focused on doing the harder things in the beginning while keeping the easy things for the end. Below is the Employees tracking DataMart from Kimball. I could not find anything specific to only recruitment in Kimball DataMart, which we will be disusing inferring Kimball DataMart concept [6]. Kimball HR DataMart is shown in below Figure.

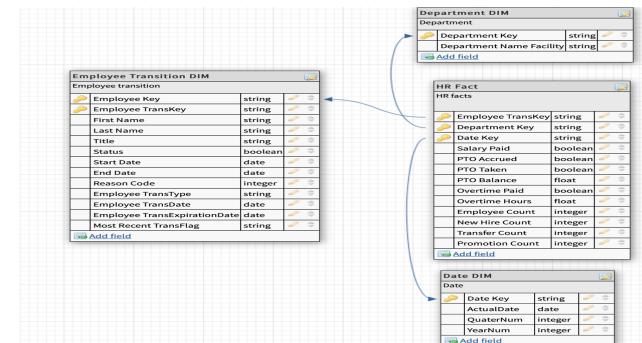


Figure - Kimball HR DataMart [1]

B. HR DataMart by Oracle

Then we have Oracle DataMart. Oracle Recruitment DataMart is a data storehouse of enlistment also, staffing drives, status, costs, and results. This information store contains more outlined data and all current and memorable well-known enrolment drives including open positions, orders, candidates, applications, results, and enrolment cost. Recruitment measurements can be reported and investigated by a wide assortment of credits, including specialty unit, division, work code, area, status, status reason, reference class, and subcategory, etc. The Recruiting information store furnishes staff and the board with data expected to settle on informed choices with respect to current and future recruitment drives [7].

Oracle gives a total and self-administration arrangement that permits business groups to get the profound, reliable, information-driven experiences they need to settle on speedy choices. Business groups can rapidly join all vital information across various sources and organizations, including spatial and diagram, in a combined data set to drive secure cooperation around a solitary wellspring of truth given by information stores. Analysts can without much of a stretch influence self-administration information devices and implanted AI—with zero coding needed—to speed up information stacking, change, and readiness, consequently find examples and patterns, make expectations, and gain experiences dependent on information with straightforward ancestry [8] [9]. Oracle DataMart has been shown in Figure No.2.2.

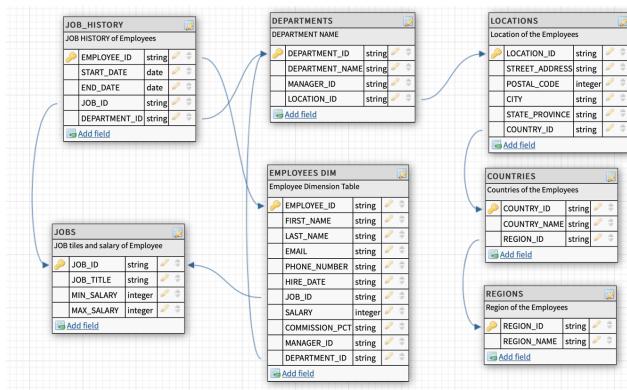


Figure - Oracle HR DataMart [2]

III. RESEARCH METHODOLOGY

The following point consists of the project methodology:-

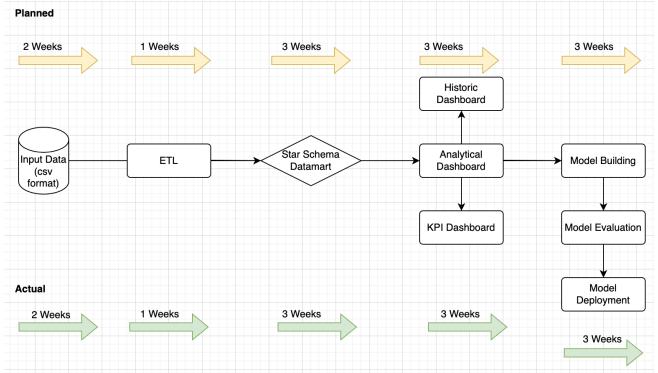


Figure - Project Pipeline

1) Stage 1:-

Load the Data :- At this stage, the data will be load to the system. Here we are accepting only csv file as input data for now.

2) Stage 2:-

ETL and finding fact and dimension tables:- Data exploration and further cleaning as per the requirement. Ingest the clean data into system. Dimension table the tables of a star schema which store the attributes that describes the object. Here dimension tables are Candidate Reference Id, Manager Id, Job Role, Job Category, Source, Salary Band, Experience etc.

Facts are the numeric dimensions of the business. They support numerical computations used to provide details regarding and investigate the business. Here fact tables are Cost of Hire, Quality of Hire, Time to Hire.

Below figure shows the dimension and fact tables.

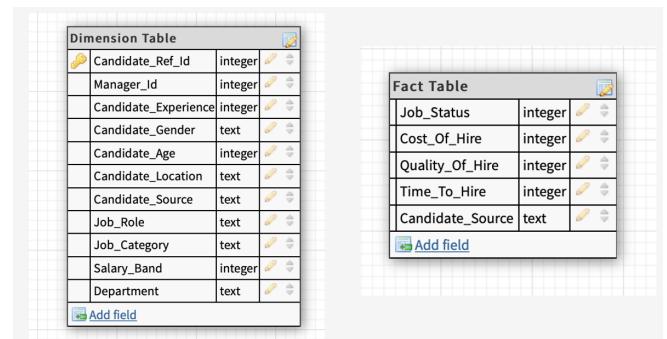


Figure – Dimension & Fact Tables

3) Stage 3:-

Star Schema DataMart:- The clean data will be used to find the dimension and fact tables. Further star schema DataMart will be created.

Below figure shows the star schema for recruitment process. Snowflake. Star schema has been taken under consideration for this study. Snowflake schema is just to understanding in detail purpose.

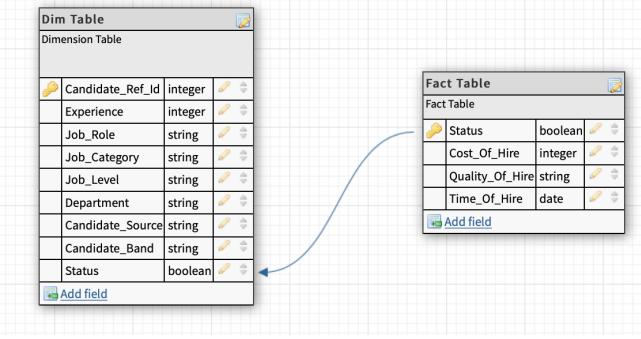


Figure - Star Schema

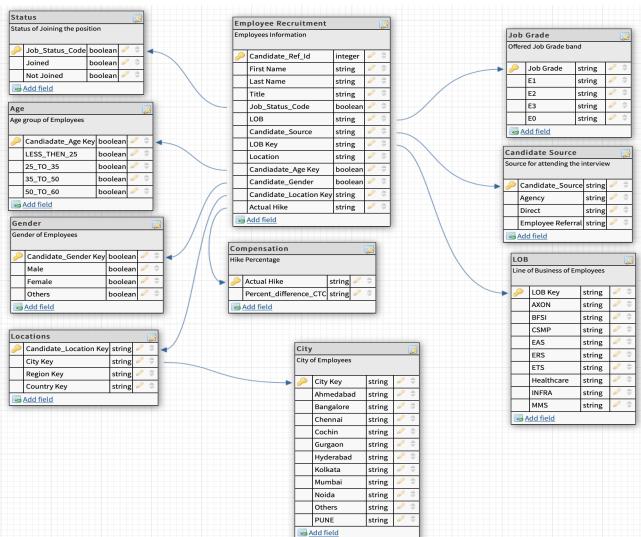


Figure - Snowflake

4) Stages 4:-

Creating Analytical Dashboards:- Analytical dashboard, historic dashboard and KPI dashboard will be created. Analytical DataMart helps us to create decision management platforms where data scientists can use the analytical DataMart for their analysis. Analytical DataMart Example has been shown in below figure.

- Application Tracking
 - Number of Candidate seen the application
 - Number of candidate filled up the application
- Interview Tracking
 - Number of interview schedule
 - Number of interview rescheduled
 - Number of interview taken
 - Number of interviews completed
 - Number of selections
- Recruitment Source Analytics
 - Application Sourced Directly
 - Application Sourced by Employee Referral
 - Application Sourced by Agency
- Notice Period Tracking
 - Notice period
 - Joined

Figure - Analytical DataMart

5) Stage 5:-

Model Building:- After data modelling by building DataMart, segmentation has been done using K-Mean Algorithm. To check the number of clusters, elbow method has been used which is shown in below figure.

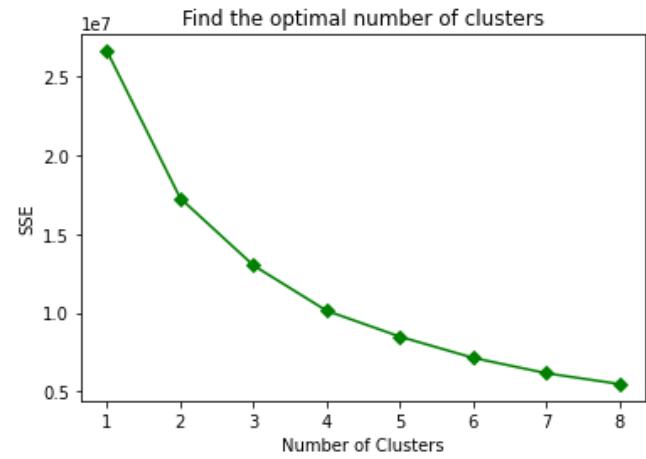


Figure – Elbow Method for K-Mean Clustering

Once the data is segmented in clusters, historic dashboard has been built on data studio tool. The Historic Dashboard shows the historic trend of the recruitment referring to other features like cost of hire, hiring cost and hiring time.

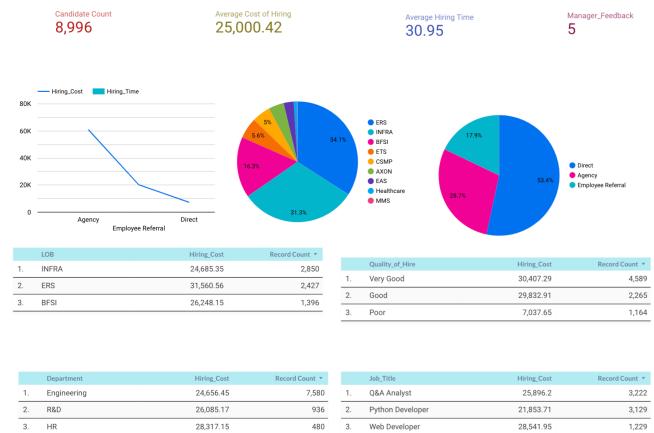


Figure - Historic Dashboard

6) Step 6:-

Deployment:- The DataMart has been designed on dbdesigner which generates the SQL code which can be further developed in any database tool like Microsoft SQL server management studio or oracle database or can be built on any cloud like Amazon Web Services, Google Cloud Platform, Microsoft Azure. Various dashboards like historic and predictive dashboards have been made on Datastudio which is a free tool from Google. Alternately one can use Power BI or Tableau as a substitute of Datastudio. For segmentation (modeling), K-Means algorithm has been used to build the model. The model has not been deployed as of now.

IV. RESULTS AND DISCUSSION

After A DataMart is the access layer of a data warehouse used to offer data to users in business intelligence. Small chunks of the data warehouse are often referred to as DataMart. Data warehouses typically hold enterprise-wide data, whereas DataMart typically store data unique to a department or team. The primary goal of DataMart is to offer the most relevant data for BI to the business user in the shortest

time feasible. Millions of records can be stored in a DataMart, which necessitates terabytes of storage. The following are some of the benefits of employing a DataMart:

- Allows users to have access to the precise type of data they require, which improves end-user response time.
- A data warehouse that has been compacted and narrowed in scope.
- Each one is devoted to a single item or function.
- Implementing a full data warehouse is more expensive.
- Contains a lot of information.
- It is less crowded and simply contains necessary company information and data.
- Assists in the integration of all data sources
- The building and use of a DataMart result in a significant amount of data summarization.

In this project, a DataMart has been build up referring to Kimball HR DataMart and Oracle recruitment DataMart. A star schema with facts and dimension tables has been built up to develop KPI DataMart and analytical DataMart. Further, segmentation model has been created using K-Mean algorithm to cluster the similar kind of candidates. As per elbow method, five clusters has been created to segment the candidates. Later, historic dashboard has been developed to analyse the candidate profiles in details.

V. REFERENCE

- [1] M. P. U. Kailash, " "HR Analytics Methodical Measurement of HR Processes," International Journal of Innovative Science and Research Technology , 2020".
- [2] S. George, "Inmon or Kimball: Which approach is suitable for your data warehouse," 14 April 2012. [Online]. Available: <https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-your-data-warehouse>.
- [3] M. R. R. Kimball, "Data Warehousing, Business Intelligence, and Dimensional Modeling Primer," in *Kimball The Data Warehouse Toolkit 3rd Edition*, Wiley, 2016, pp. 7-16.
- [4] M. R. R. Kimball, "Employees Tracking for Profile," in *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition*, John Wiley & Sons, Inc, 2016, pp. 263-265.
- [5] Oracle, "Recruiting Data Mart Dimension Tables," [Online]. Available: https://docs.oracle.com/cd/E41507_01/epm91pbr3/eng/epm/phcw/concept_HCMWarehouseStructure-399b81.html..
- [6] M. R. R. Kimball, "Data Warehousing," in *Business Intelligence, and Dimensional Modeling Primer*, in *Kimball The Data Warehouse Toolkit 3rd Edition*, Wiley, 2016, pp. 7-16.
- [7] M. R. Ralp Kimball, "Employees Tracking for Profile," in *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition*, John Wiley & Sons, Inc, 2016, pp. 263-265.
- [8] Oracle, "Recruiting Data Mart Dimension Tables," [Online]. Available: https://docs.oracle.com/cd/E41507_01/epm91pbr3/eng/epm/phcw/concept_HCMWarehouseStructure-399b81.html..
- [9] C. F. T. R. S. A. Prado, " Using OLAP Tools for e-HRM: A case study," *International Journal of Technology and Human Interaction*, 2010.
- [1] M. A. U. H. N. H. Z. A. M. G. M. S. A. K. Hamoud, [0] "Improve HR Decision-Making Based On Data Mart and OLAP," *Journal of Physics*, 2020.
- [1] Oracle, "Recruiting Data Mart Dimension Tables," [Online]. Available: https://docs.oracle.com/cd/E41507_01/epm91pbr3/eng/epm/phcw/concept_HCMWarehouseStructure-399b81.html..
- [1] M. P. Udhay Kailash, " "HR Analytics Methodical Measurement of HR Processes," International Journal of Innovative Science and Research Technology , 2020.