REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Development of Analytical and KPI Datamart for Recruitment Analytics

Ashish Chandra Jha

SRN: R19MBA01

Date: 27th August 2022

MBA in Business Analytics

Capstone Project Presentation
Year: II

race.reva.edu.in

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Agenda

01 Introduction
Back Ground | Current status | Why this study

02 Literature Review
Seminal works | Summary | Research Gap

03 Problem Statement
Business Problem | Analytics Solution

04 Project Objectives
Primary & Secondary Objectives | Expected Outcome

05 Project Methodology
Conceptual Framework | Research Design

06 Business Understanding
Business Context | Monetary Impact

07 Data Understanding
Data Collection | Variables

08 Data Preparation
Pre-processing | Process | Techniques

09 Exploratory Data Analysis
Univariate | Bivariate | Hypothesis

10 Modeling
Machine Learning | Model Evaluation | Insights

11 Model Deployment
Applications | Demo

12 Suggestions and Conclusions
Insights | Next Step | Future Scope

13 Annexure
References | Publications | Plagiarism Score

Introduction

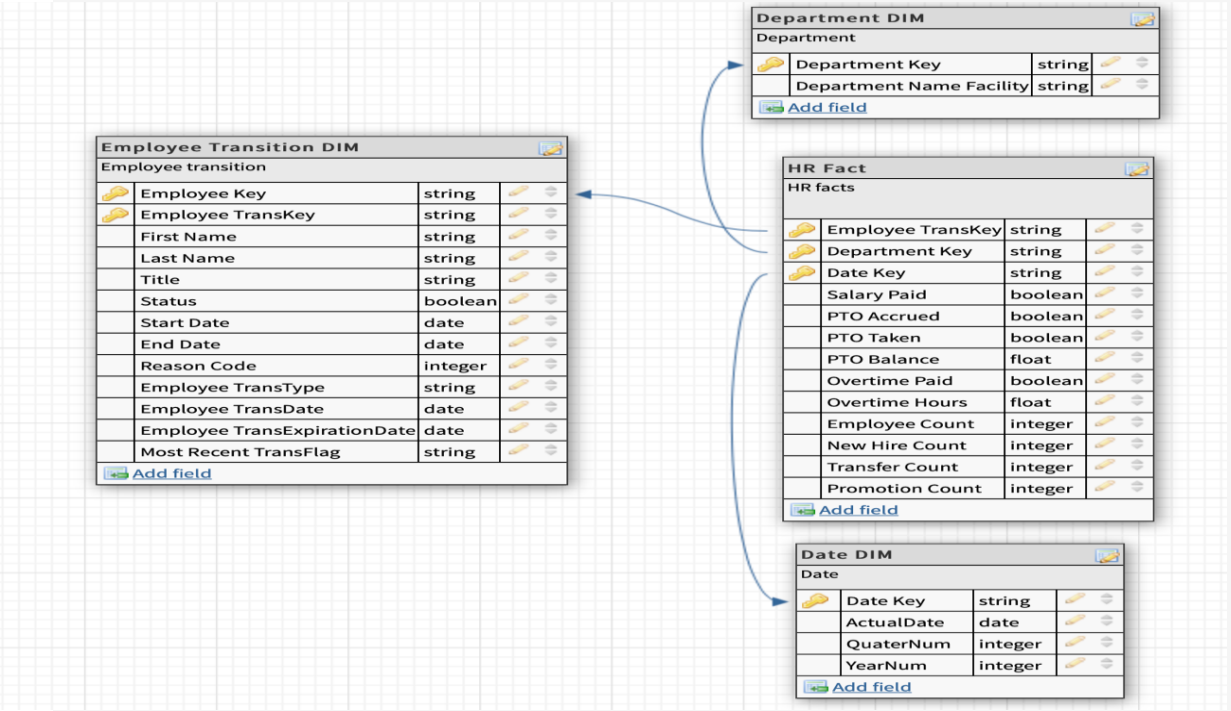
Background | Current status | Why this study

- The success and growth of any organization primarily depend on its employees
- 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 is responsible for almost all aspects concerning the workforce of an organization.

Literature Review

HR Datamart by Ralph Kimball

Seminal works | Summary | Research Gap

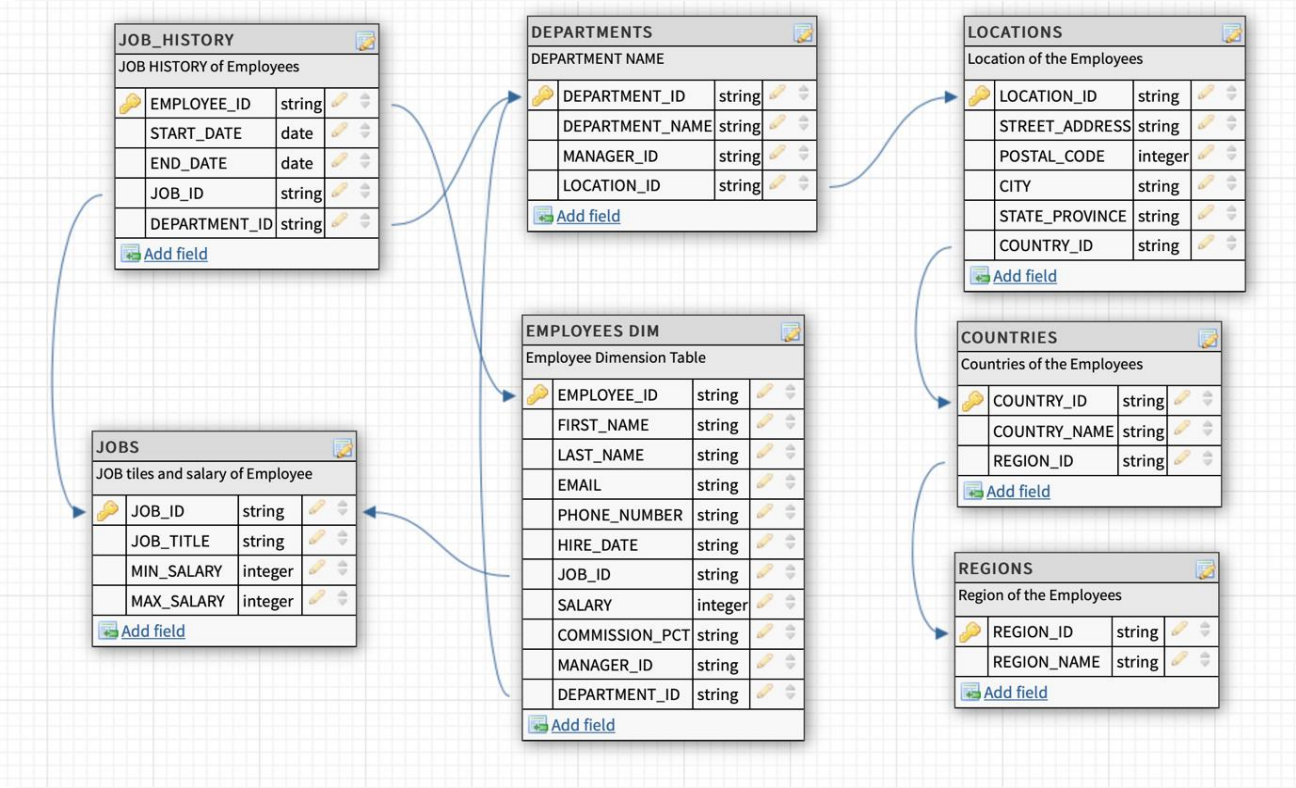


- The Kimball HR Datamart is based on the ideology of Ralph Kimball according to whom the data warehouses should be model using dimensional models such as the star schema or snowflake schema.
- A star schema is a tool for dimensional modeling of data by organizing it to allow analytical operations to run on it.

Literature Review

HR Datamart by Oracle

Seminal works | Summary | Research Gap



- 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.

Problem Statement

Business Problem | Analytics Solution

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

Project Objectives

Primary & Secondary Objectives | Expected Outcome

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.

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Project Methodology

Conceptual Framework | Research Design

01

Planning of Dimension and Fact Table

In this process, we have explored for the dimension and facts of the data

02

Building Dimensional and Analytical Datamart

Dimensional and Analytical datamart will be build up

03

Building Star Schema

Star schema has been build using the dimension and fact tables

04

Historical and Analytical Dashboards

Building Historical and Analytical Dashboards

05

Model Building and Evaluation

Segmentation Model Building and Evaluated

06

Deployment

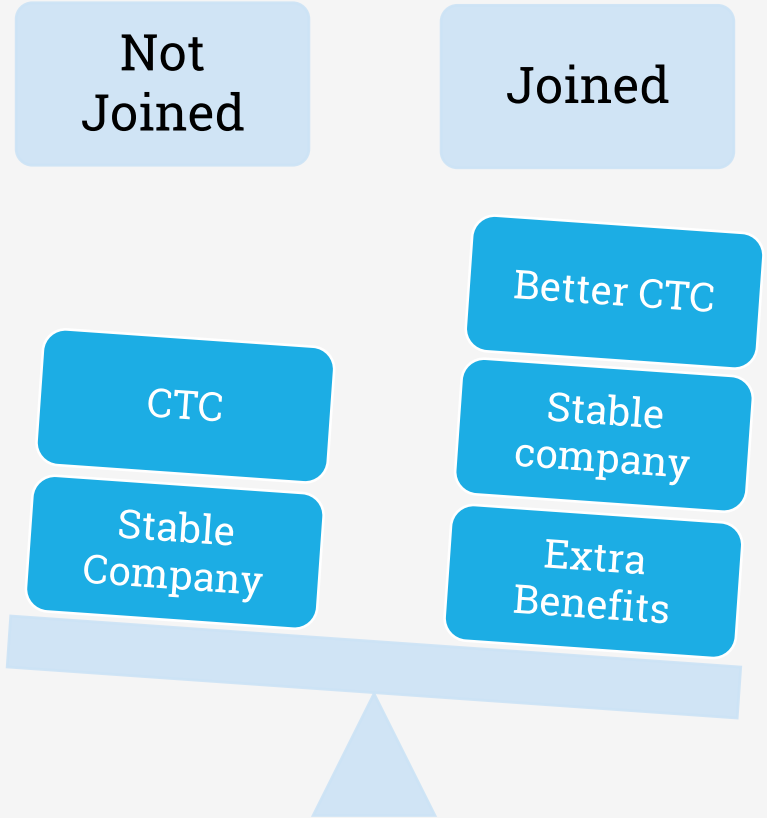
Deploying the model

Business Understanding

Business Impact | Challenges | Monetary Impact

Company Wise

Candidate Wise



Business Understanding

Business Impact | Challenges | Monetary Impact

Data Challenges

Data scattered over multiple sources

No single source of truth, Multiple metrices

Management by spreadsheet

Inconsistent Business Process

Business Understanding

Business Impact | Challenges | Monetary Impact

Business Challenges

No accurate view of workforce profile

Lack of visibility into the effectiveness of HR programs

Poor alignment of the talent management strategy with corporate strategy

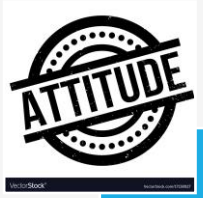
Time wasted gathering data to manage and report

Data Understanding

Data Collection | Variables

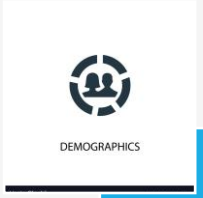
Behavioural Data

- Percentage Salary Hike
- Feedback from Hiring Manager



Attitudinal

- Job Satisfaction
- Number of companies worked for
- Tenure in each company



Demographical

- Age
- Gender
- Location

Data Understanding

Data Understanding of Dimension & Fact Tables

Dimension Table

Candidate_Ref_Id	integer	
Manager_Id	integer	
Candidate_Experience	integer	
Candidate_Gender	text	
Candidate_Age	integer	
Candidate_Location	text	
Candidate_Source	text	
Job_Role	text	
Job_Category	text	
Salary_Band	integer	
Department	text	

Add field

Fact Table

Job_Status	integer	
Cost_Of_Hire	integer	
Quality_Of_Hire	integer	
Time_To_Hire	integer	
Candidate_Source	text	

Add field

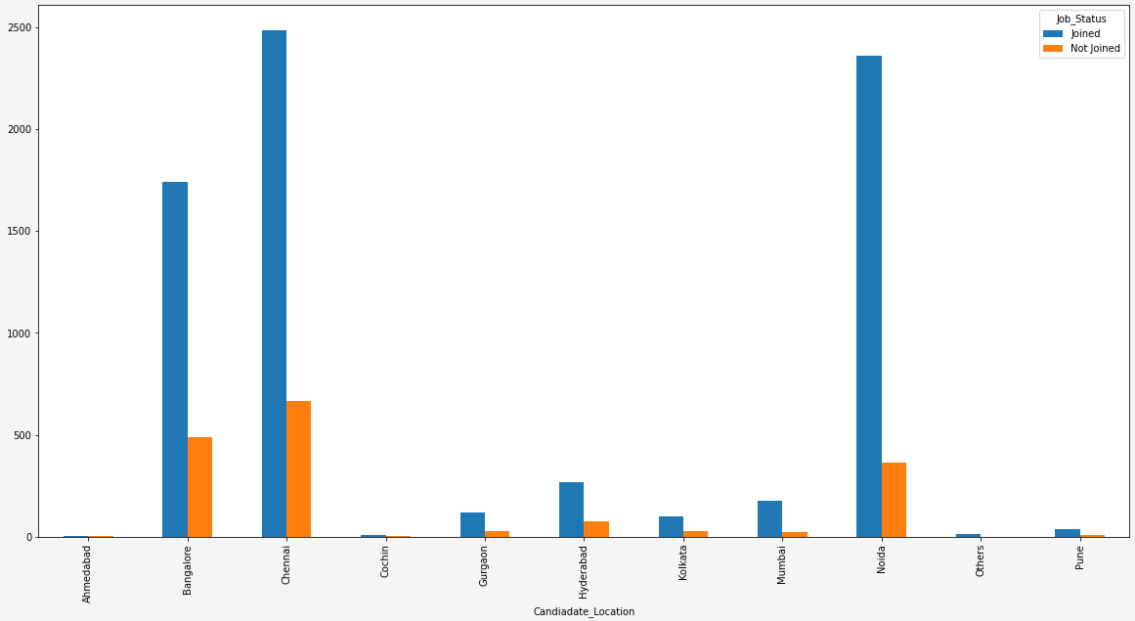
Data Preparation

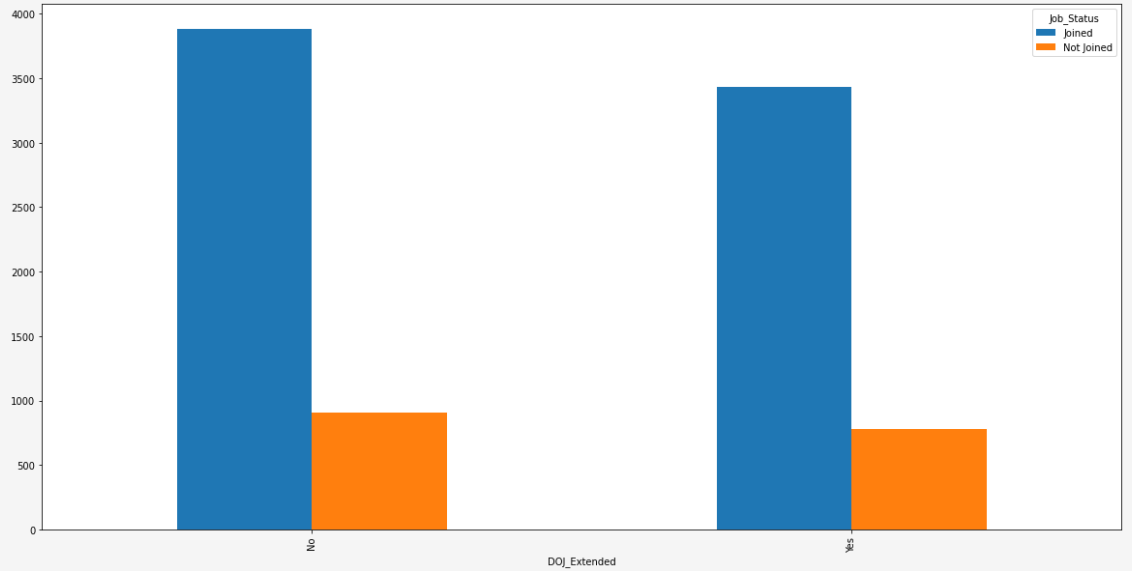
Pre-processing | Techniques

- Calculating Cost of Hire (COH) - Selection of Profile, Scheduling the interview and Operations
- Calculating Quality of Hire (QOH) – Feedback from Manager to the candidate
- Calculating Time to Hire (TTH) – Date the interview selection started till the position is closed

Exploratory Data Analysis

Pre-processing | Techniques



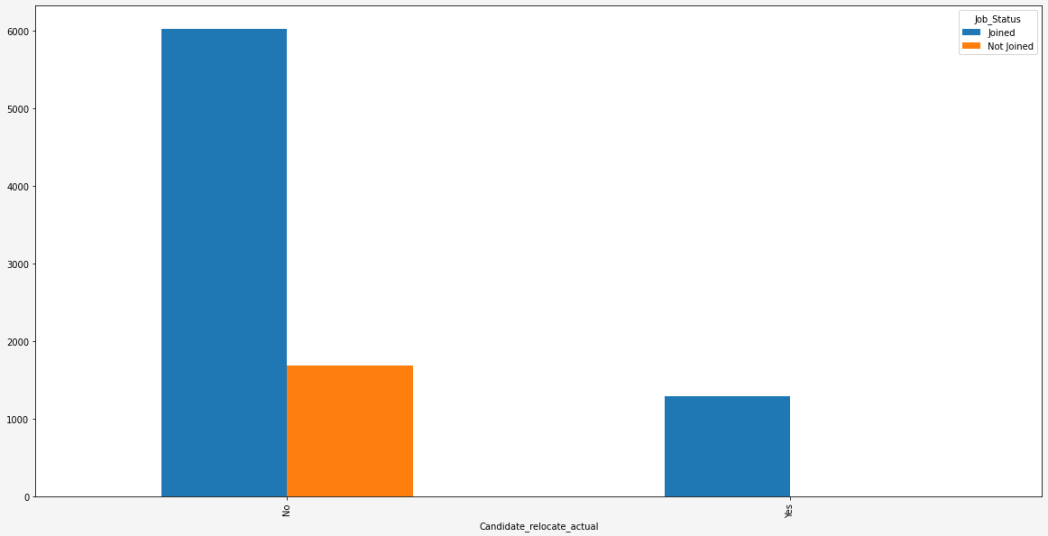


Ratio of candidate location wise who has joined the company or not.

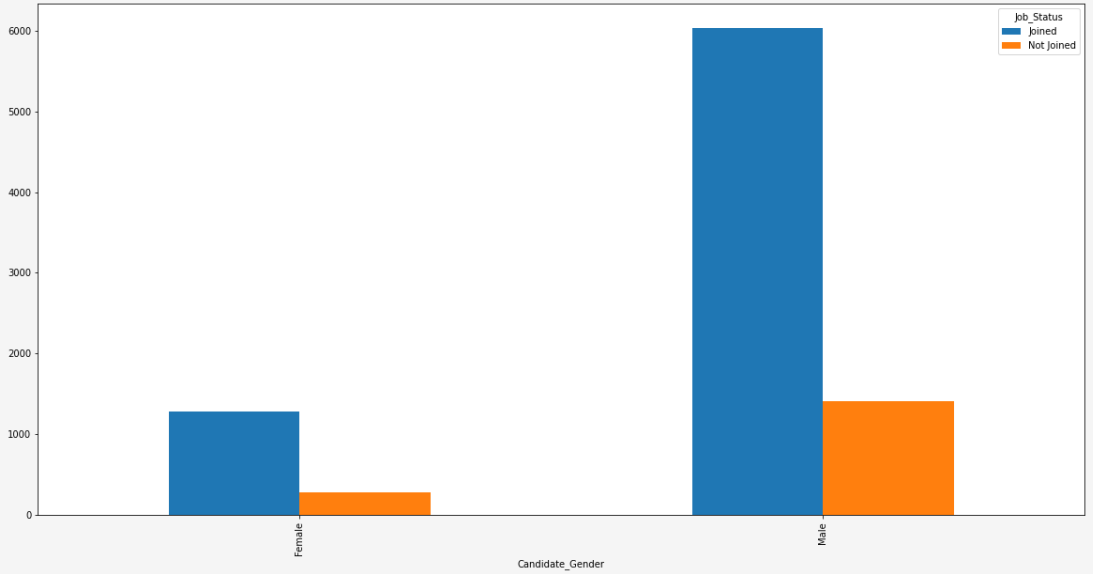
Ratio of candidate who has extended joining date and joined the company or not.

Exploratory Data Analysis

Pre-processing | Techniques



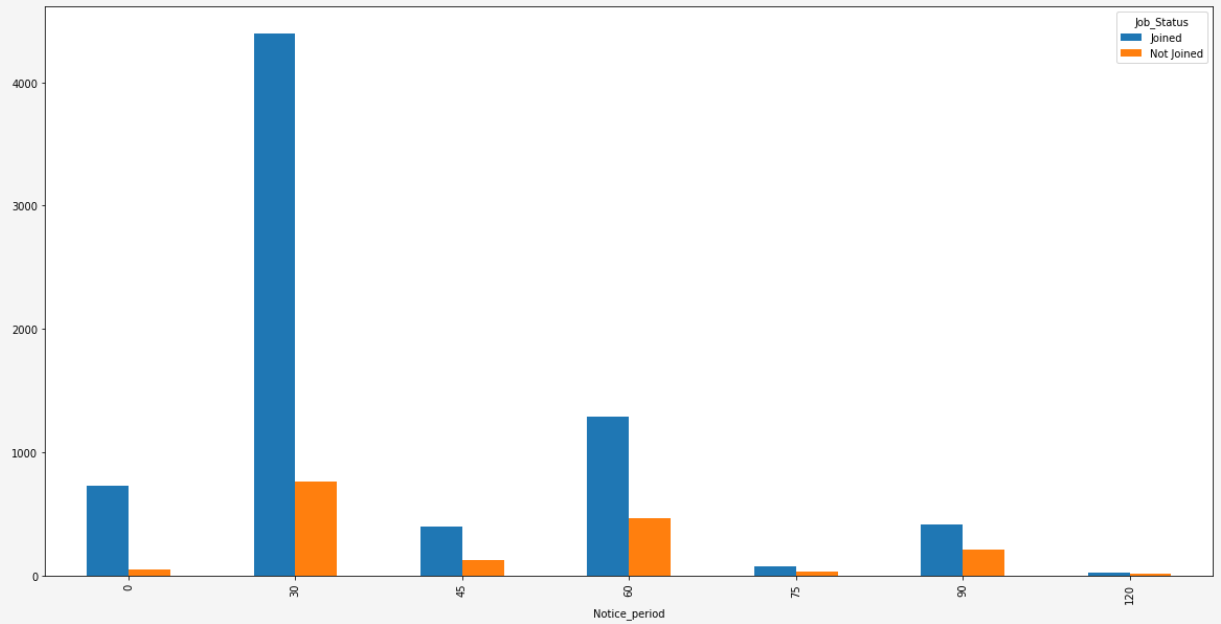
Ratio of candidate who has actually relocated and joined the company or not



Ratio of Male and Female candidates joined and not joined

Exploratory Data Analysis

Pre-processing | Techniques



Ratio of people who joined or not joined and different Notice Period

Modeling

Step 1:- Planning of Dimension and Fact Table

Dimension Table

Candidate_Ref_Id	integer	
Manager_Id	integer	
Candidate_Experience	integer	
Candidate_Gender	text	
Candidate_Age	integer	
Candidate_Location	text	
Candidate_Source	text	
Job_Role	text	
Job_Category	text	
Salary_Band	integer	
Department	text	

Add field

Fact Table

Job_Status	integer	
Cost_Of_Hire	integer	
Quality_Of_Hire	integer	
Time_To_Hire	integer	
Candidate_Source	text	

Add field

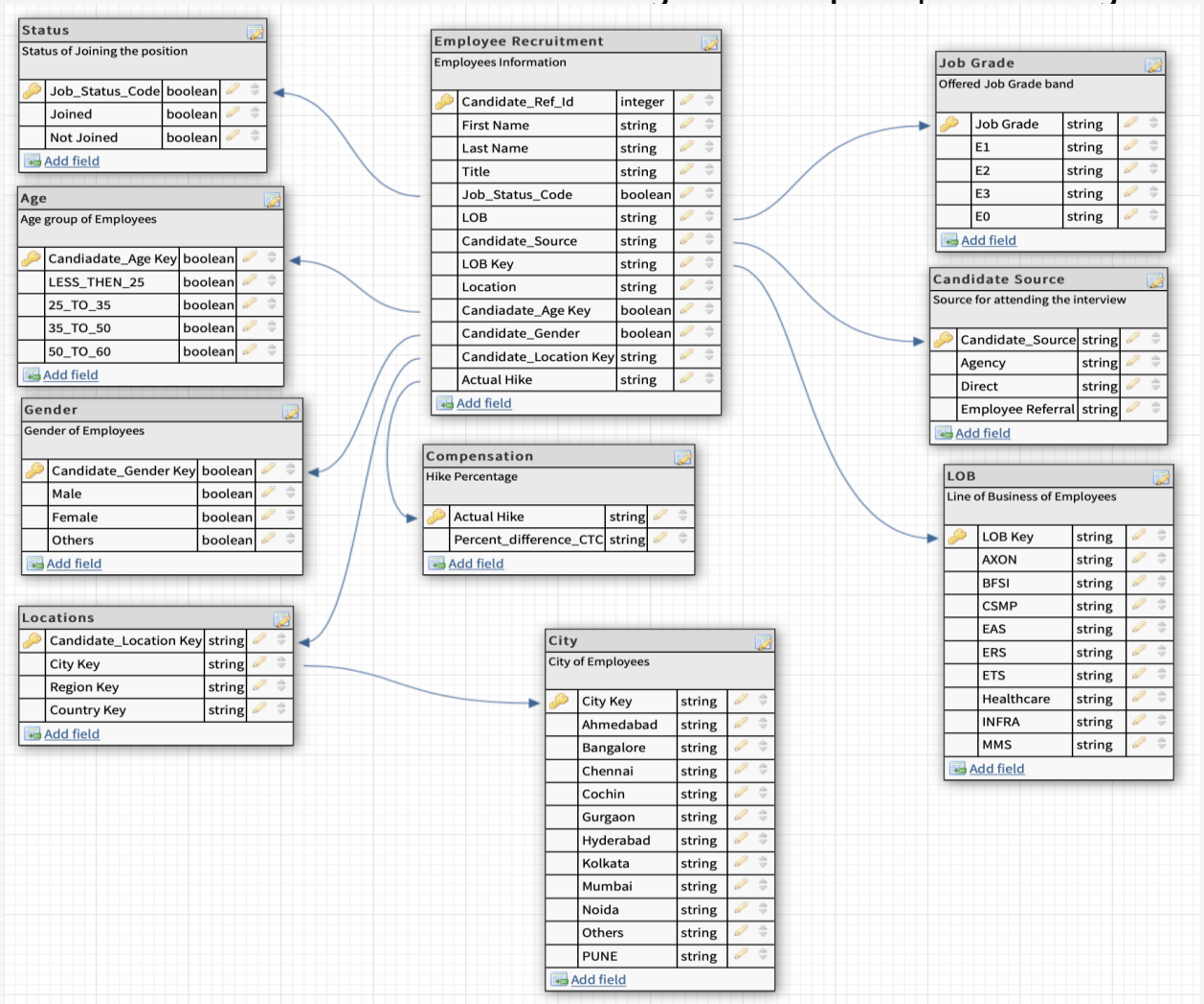
Step 2:- Building Dimensional and Analytical Datamart

Name	Type	Schema
Tables (2)		
Dimension Table		CREATE TABLE "Dimension Table" ("Candidate_Ref_Id" INTEGER,"Manager _Id" INTEGER,"Candidate_Experience" INTEGER,"Candidat
Candidate_Ref_Id	INTEGER	"Candidate_Ref_Id" INTEGER
Manager _Id	INTEGER	"Manager _Id" INTEGER
Candidate_Experience	INTEGER	"Candidate_Experience" INTEGER
Candidate_Gender	TEXT	"Candidate_Gender" TEXT
Candidate_Age	INTEGER	"Candidate_Age" INTEGER
Candidate_Source	TEXT	"Candidate_Source" TEXT
Job_Role	TEXT	"Job_Role" TEXT
Job_Category	TEXT	"Job_Category" TEXT
Salary_Band	TEXT	"Salary_Band" TEXT
Department	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)		

Modeling

Modeling Techniques | Modeling Process | Model Building

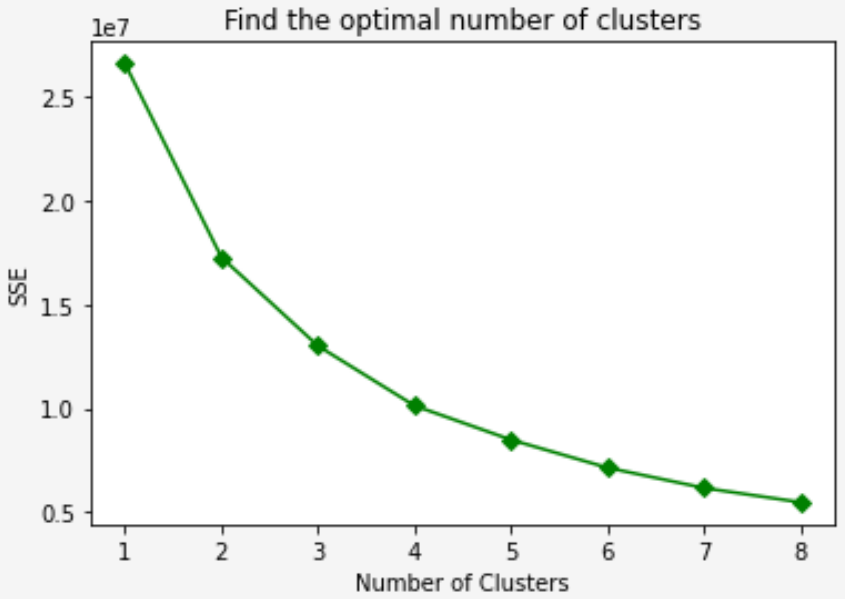
Step 3:- Building Star Schema



Modeling

Step 4:- Building Segmentation Model

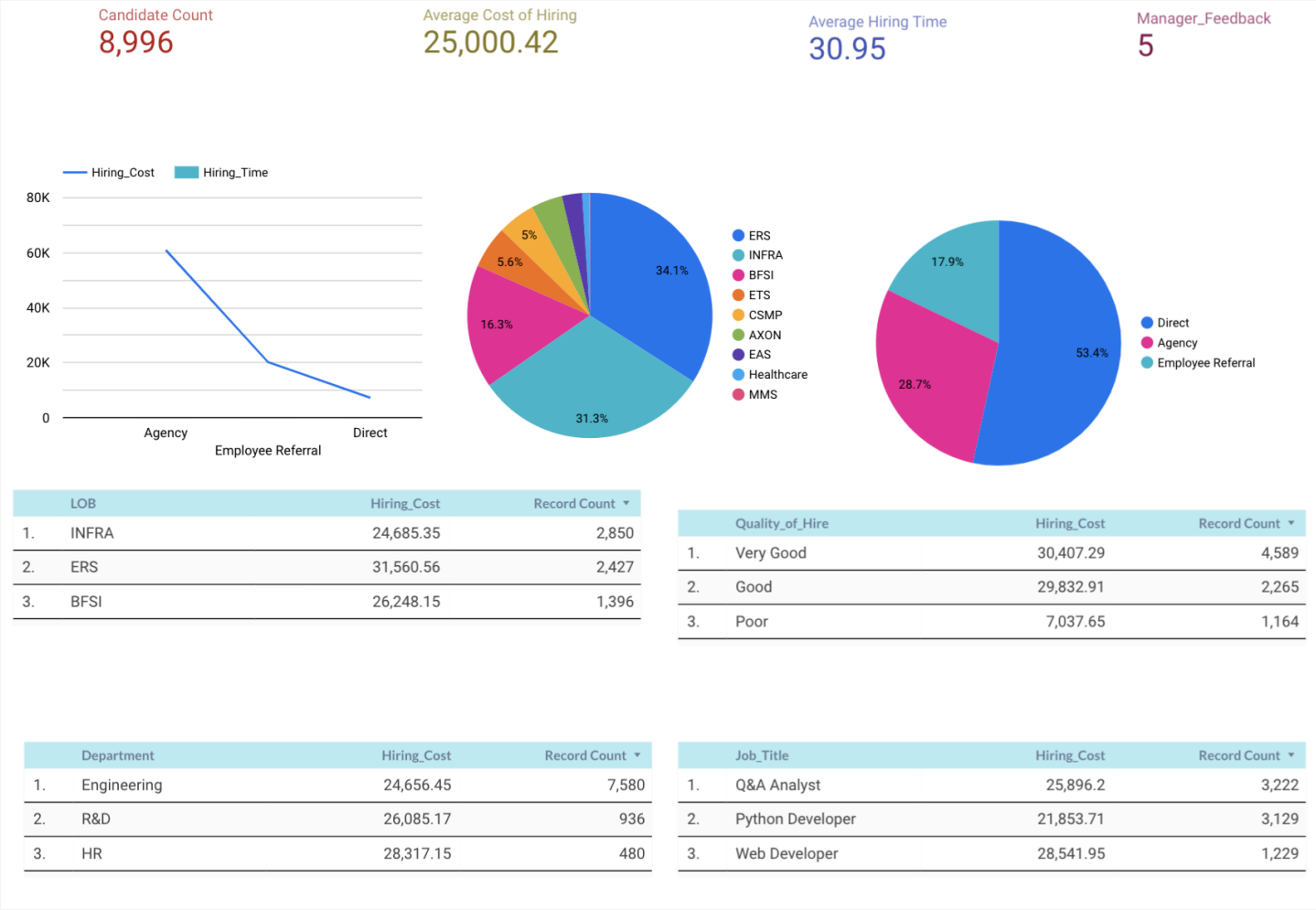
After data modelling by building Datamart, segmentation has been done using K-Mean Algorithm.



Modeling

Modeling Techniques | Modeling Process | Model Building

Step 5:- Creating Historical and Analytical Dashboard



	LOB	Hiring_Cost	Record Count
1.	INFRA	24,685.35	2,850
2.	ERS	31,560.56	2,427
3.	BFSI	26,248.15	1,396

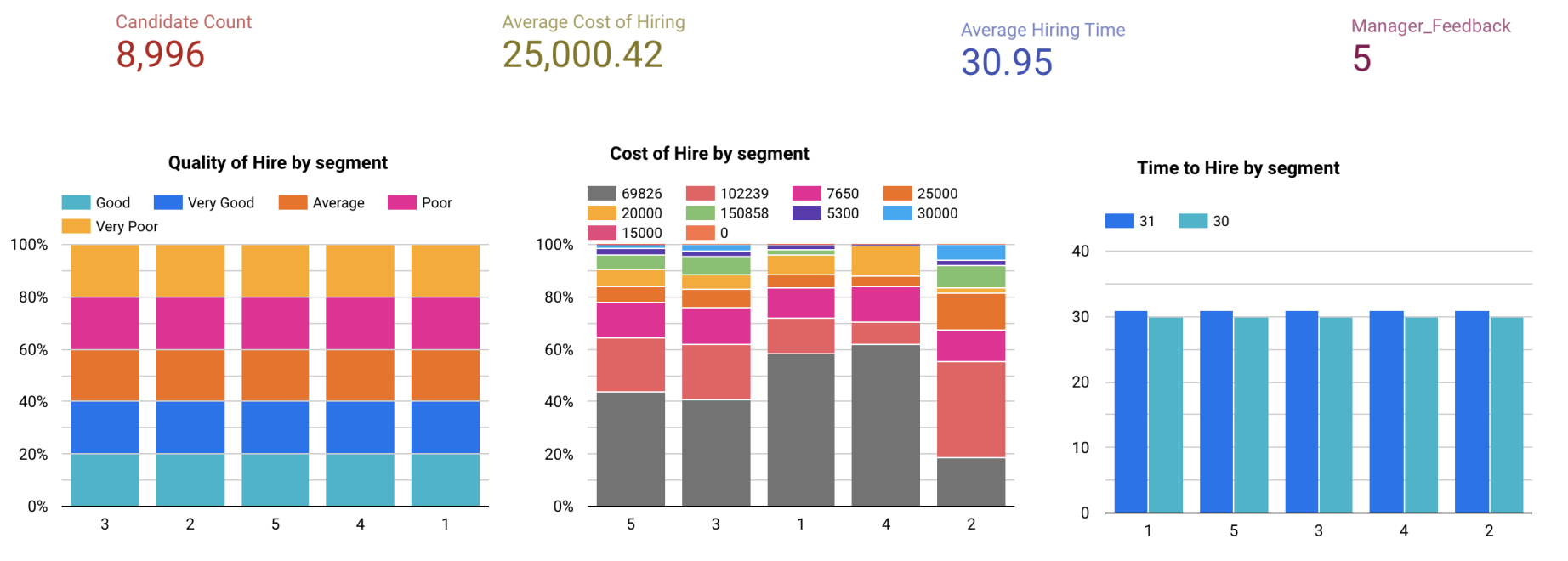
	Quality_of_Hire	Hiring_Cost	Record Count
1.	Very Good	30,407.29	4,589
2.	Good	29,832.91	2,265
3.	Poor	7,037.65	1,164

	Department	Hiring_Cost	Record Count
1.	Engineering	24,656.45	7,580
2.	R&D	26,085.17	936
3.	HR	28,317.15	480

	Job_Title	Hiring_Cost	Record Count
1.	Q&A Analyst	25,896.2	3,222
2.	Python Developer	21,853.71	3,129
3.	Web Developer	28,541.95	1,229

Modeling

Step 5:- Creating Historical and Analytical Dashboard



Model Evaluation

Results | Interpretation | Insights

- **Large-Scale Performance:** RACE Recruitment Datamart is however being tested by a limited set of users
- **Elasticity** - RACE Recruitment Datamart is quite elastic in nature
- **Ease of use –** It is quite easy to use as it has been designed with reference of other popular datamarts
- **Cost-effectiveness** - It is very cost effective, since most of the work has been done using open sources
- **Data Source Supported** - RACE Recruitment Datamart currently accepts the csv file. Unstructured data is in future scope
- **Concurrency** – It has not been tested with multiple users working at the same time yet. Although concurrency will be taken care in near future

Model Deployment

Demonstration

- 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, **K-Means** algorithm has been used to build the model. The model has not been deployed as of now.

Results and Insights

Key Findings | Suggestions

- 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.
- Segmentation model has been created using K-Mean algorithm to cluster the similar kind of candidates
- Historic and predictive dashboard has been developed to analyse the candidate profiles in details.

Conclusion and Future Work

Proposed solutions | Scope for future work

- 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.

References

Bibliography | Webliography

- 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.
- Ralp Kimball, M. R. (2016). Employees Tracking for Profile. In *The Data Warehouse Toolkit: The Defi nitive 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-
- ttttt. (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*.
- Gabcanova, 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

Annexure

Additional Information | Plagiarism score

Development of Analytical Datamart and Data Pipeline for Recruitment Analytics

ORIGINALITY REPORT

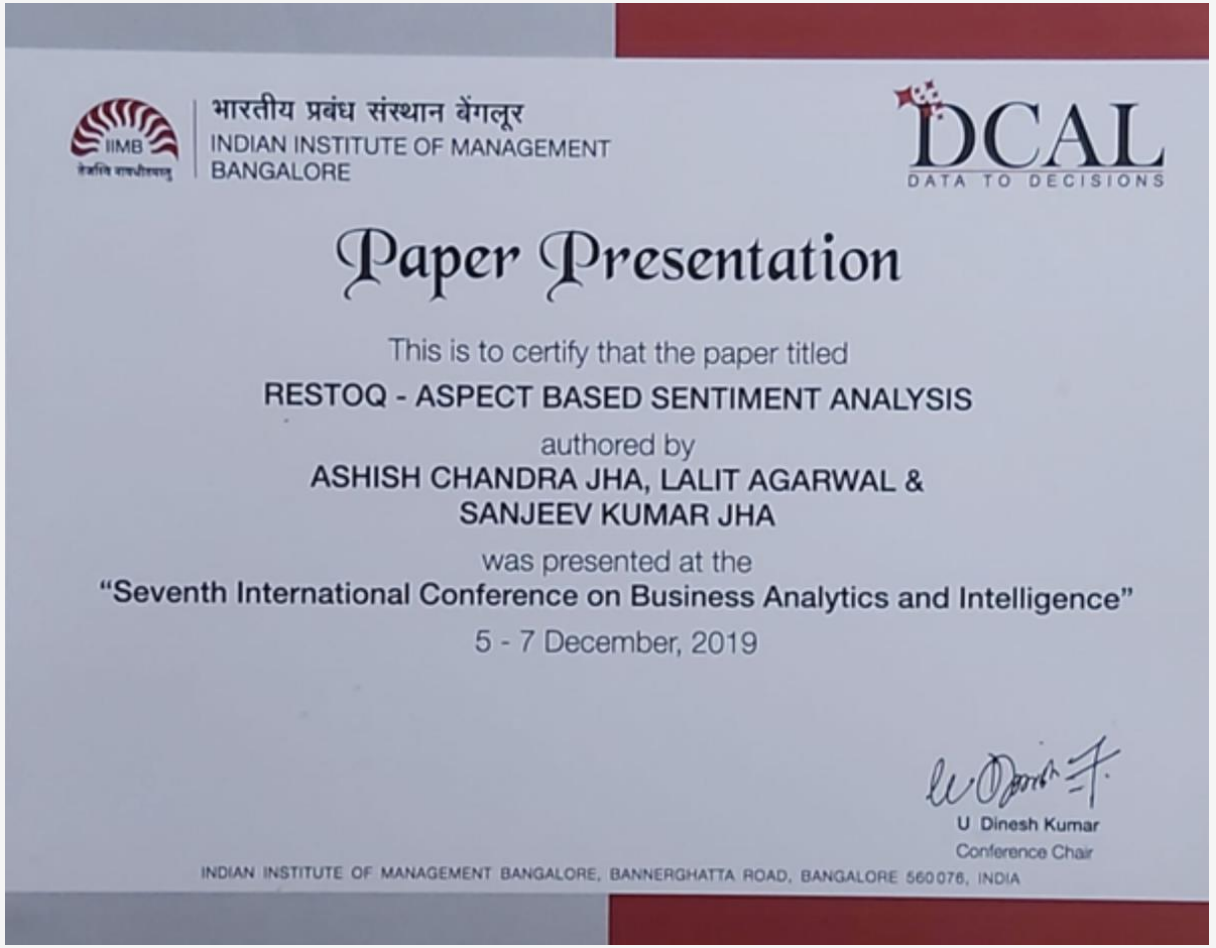
5%	2%	1%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

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	

Annexure

Publications | Conferences



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

IIMB

DCAL
DATA TO DECISIONS

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

U Dinesh Kumar
Conference Chair

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

REVA UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Special thanks to Dr. J. B. Simha, Dr. Shinu Abhi Mr. Vishal Thakur and my wife or their guidance and all the support on this initiative.

Thank You!