1st EAI International Conference on Intelligent Systems and Machine Learning (ICISML-2022)
Organized
by
Department of Electronics & Communication Engineering,
Vardhman College of Engineering, (Autonomous), Hyderabad, India, 16th -17th December, 2022

Paper Title : Development of Analytical DataMart and Data Pipeline for Recruitment Analytics

Authors: Ashish Chandra Jha, REVA Academy for Corporate Excellence, REVA University, Bengaluru, India
Sanjeev Kumar Jha, REVA Academy for Corporate Excellence, REVA University, Bengaluru, India
J. B. Simha, REVA Academy for Corporate Excellence, Bengaluru, India

Paper id: 325167

Plan of Presentation

Introduction

Related work

Methodology

Results & Discussion

Future scope

Conclusion

References

Introduction

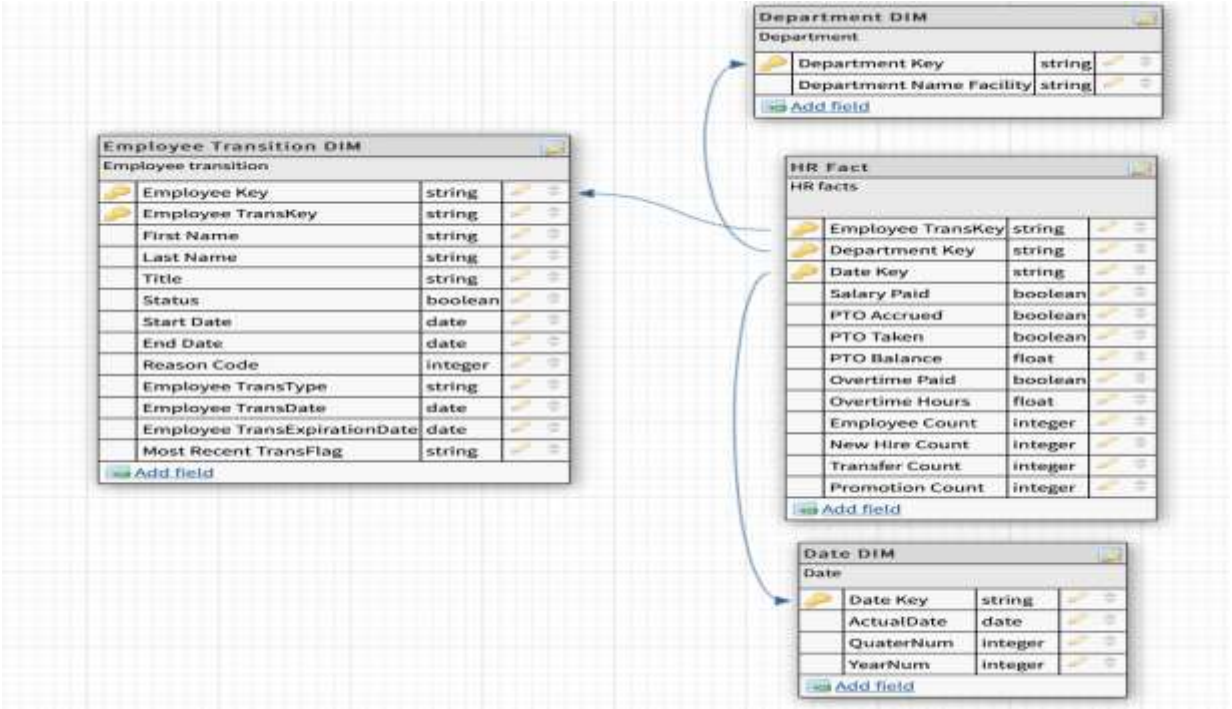
- The success and growth of any organization primarily depend on its employees
- The HR department handles all the data regarding the recruitment process while also analyzing them to select suitable candidates for the organization.
- The HR department is responsible for almost all aspects concerning the workforce of an organization.

Related work

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.

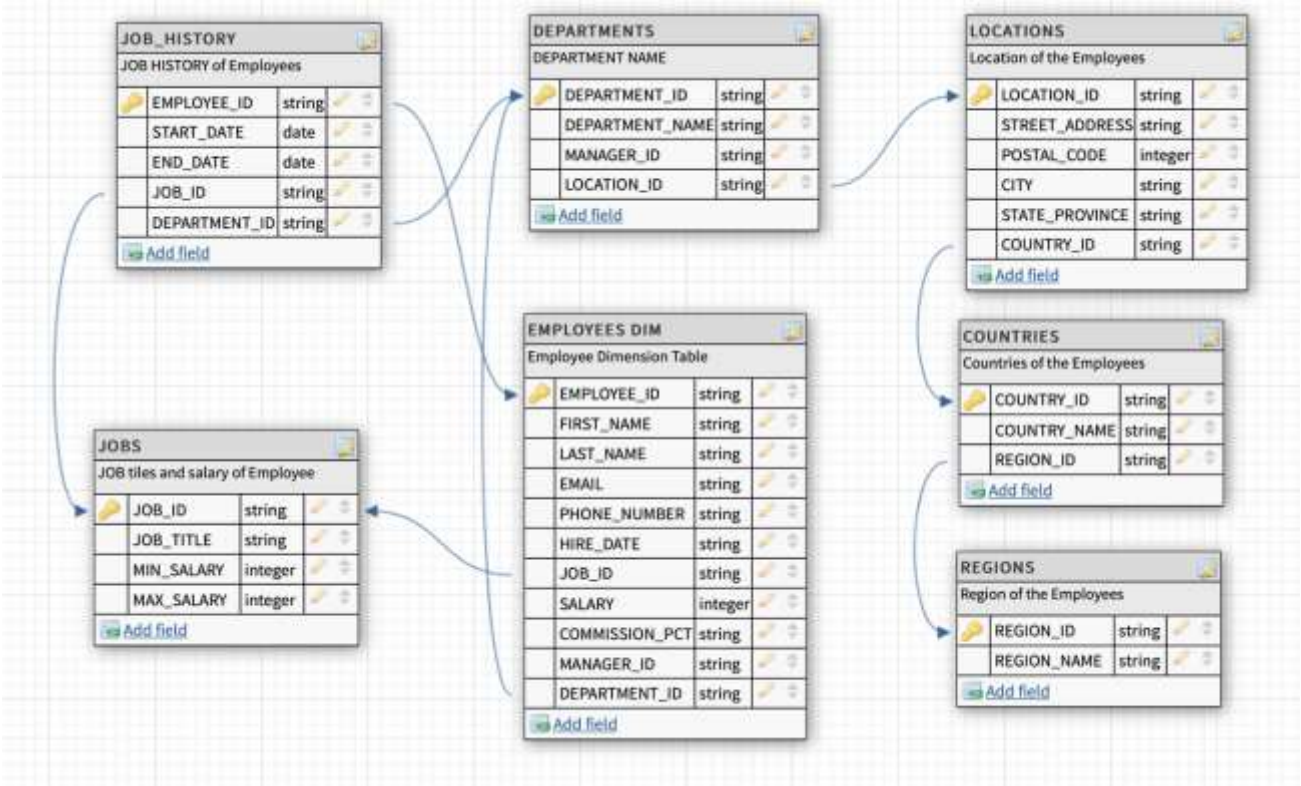
- A star schema is a tool for dimensional modeling of data by organizing it to allow analytical operations to run on it.



Related work

HR Datamart by Ralph Kimball

- 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.



Methodology

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

05
Model Building and Evaluation
Segmentation Model Building and Evaluated

06
Deployment
Deploying the model

Methodology

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 1.

Fact Table:- A fact table contains the numeric measures such as metrices produced by an operational measurement. Fact Table is shown in Figure 2.

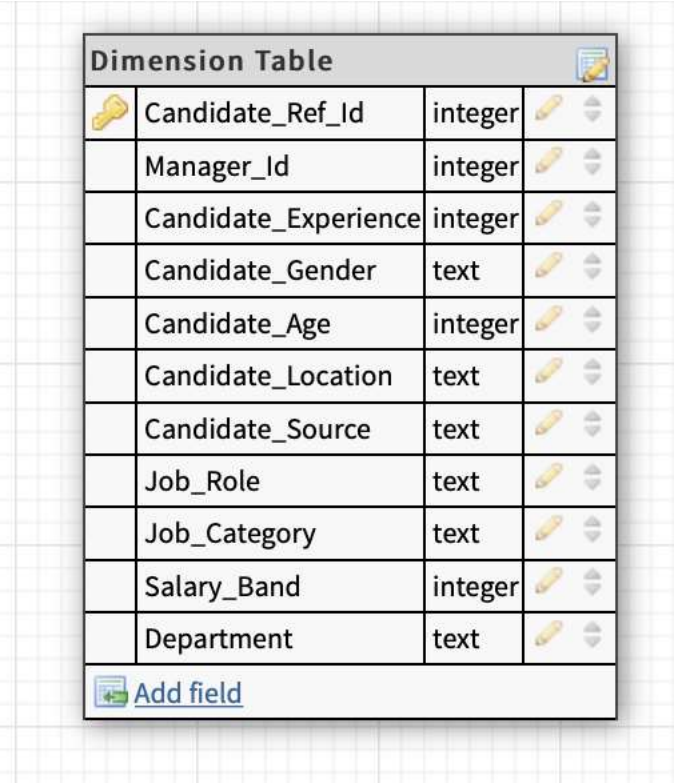


Figure 1

Figure 2

Methodology

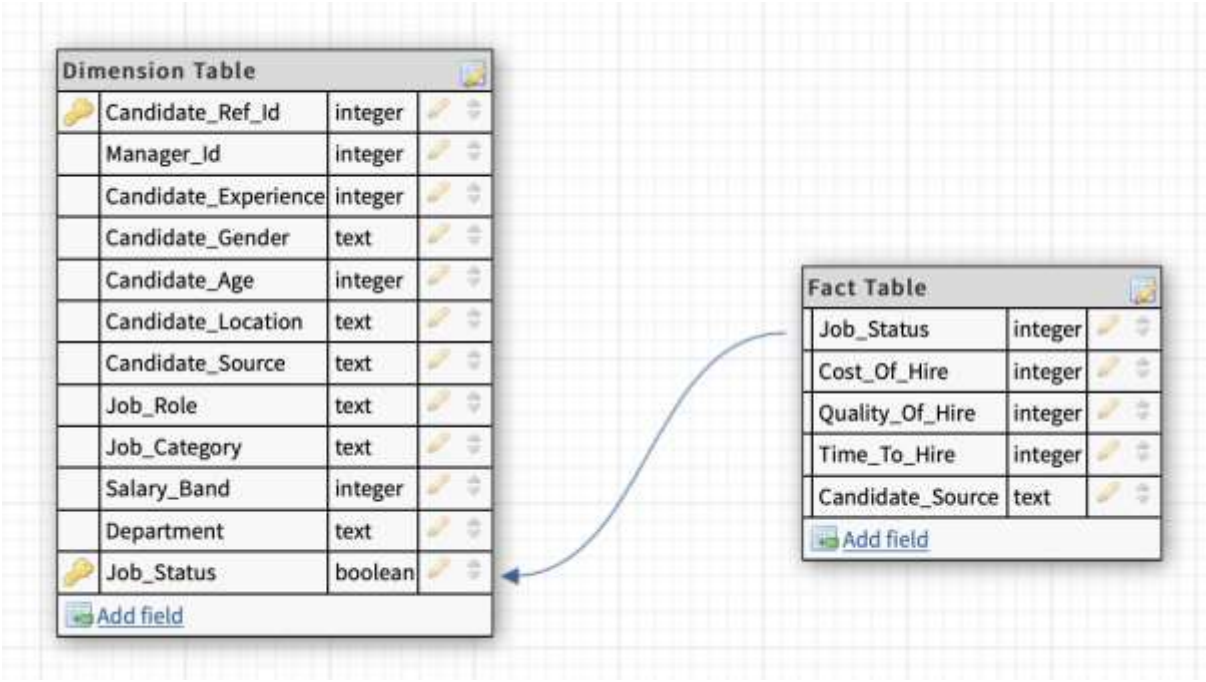
Step 2:- Building Dimensional and Analytical Datamart

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.

Dimensional Datamart

The dimensional datamart 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.



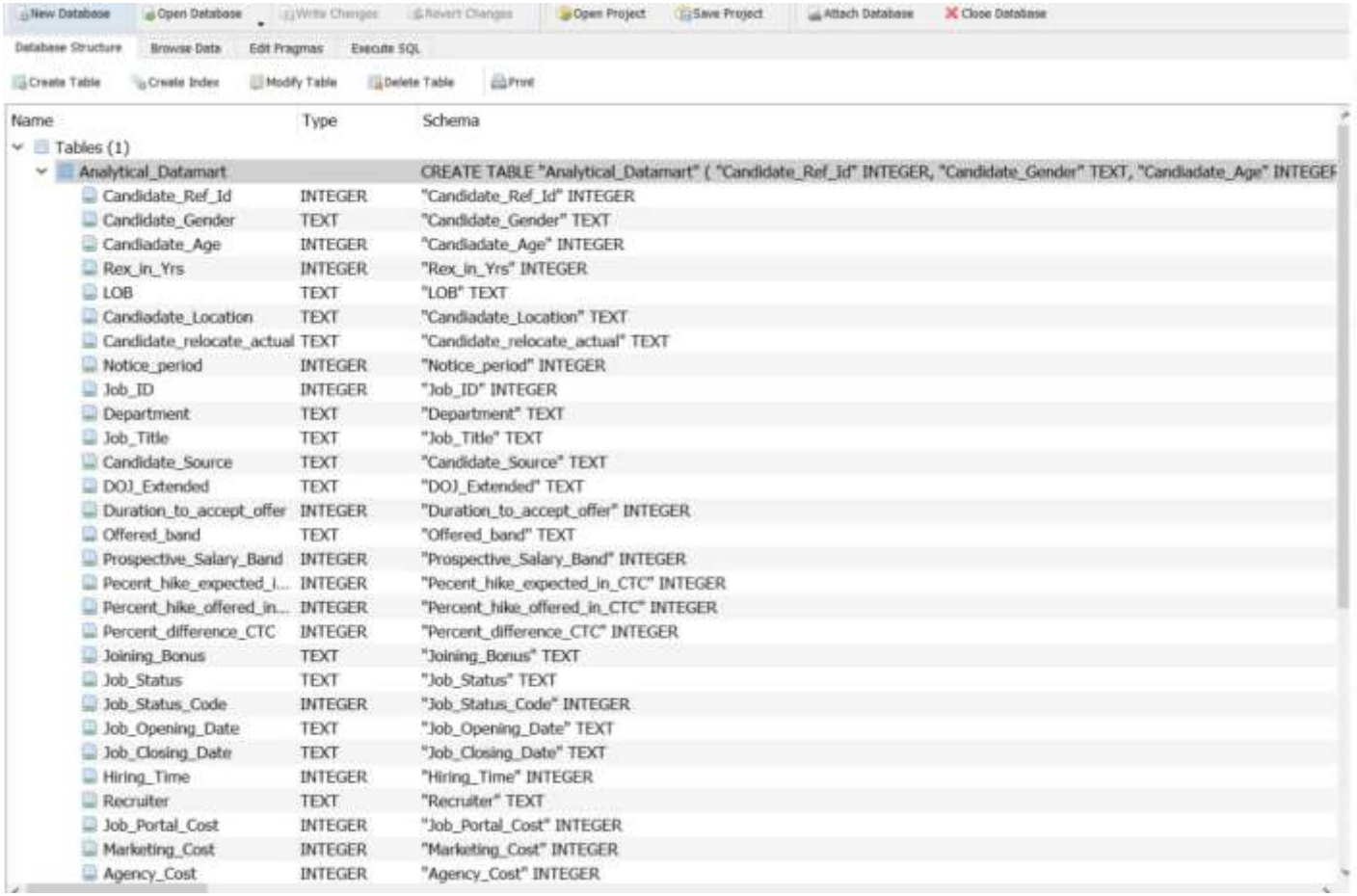
Methodology

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

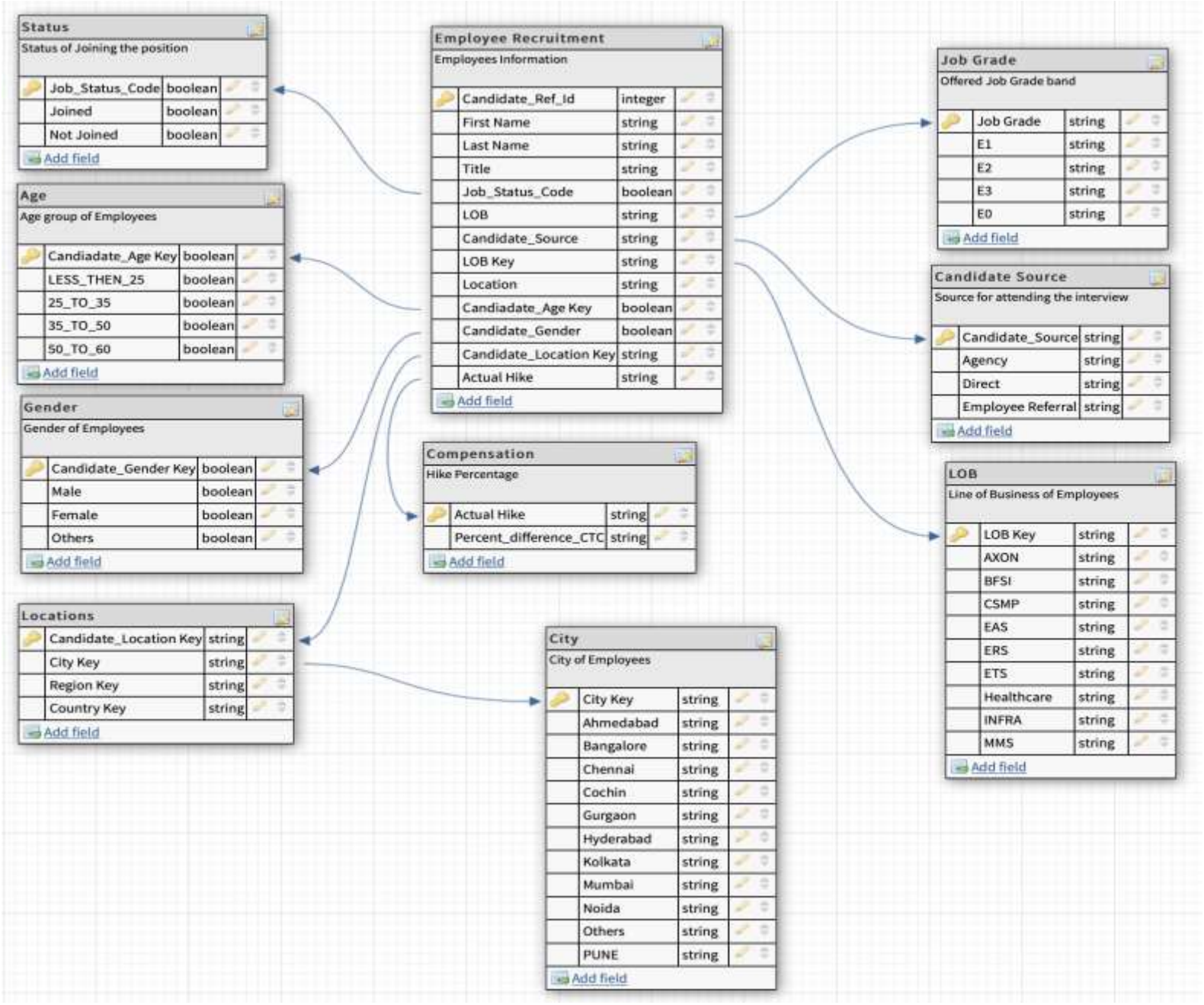
Analytical Datamart has been shown in Figure.



Methodology

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.



Methodology

Step 4:- Historical and Analytical Dashboards

Creating analytical dashboard and historic dashboard will be created.

	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

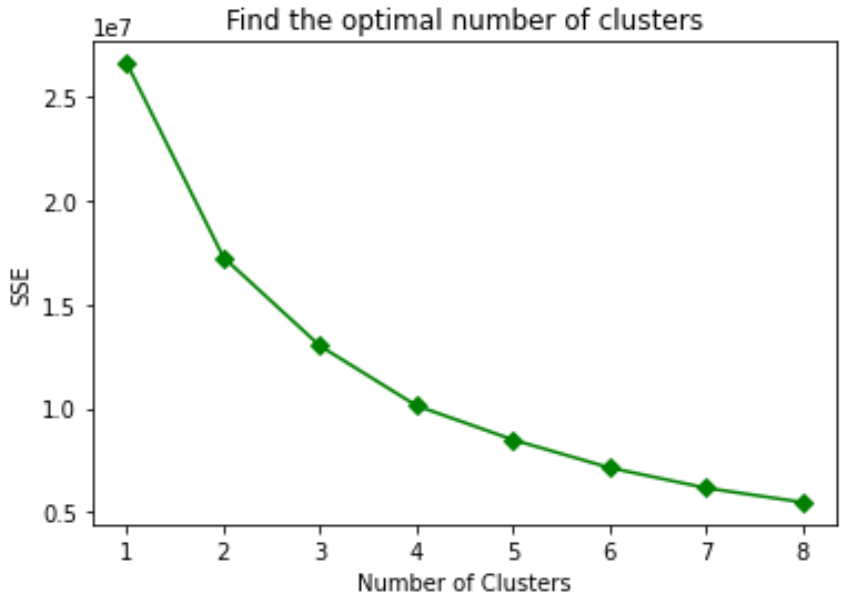
	Job_Title	Hiring_Cost	Record Count
1.	Q&A Analyst	25,896.2	3,222
2.	Python Developer	21,853.71	3,129
3.	Web Developer	28,541.95	1,229

Methodology

Step 5:- Model Building

A segmentation model will be crated to segment the candidates as per their profiles. Fine Tuning of the model to get a better segments.

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



Methodology

Step 5:- Model Evaluation

- **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

Methodology

Step 6:- Model Deployment

Segmentation model will be deployed.

- 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 & Discussion

- 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.

Future scope

- 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.

Conclusion

- 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.

References

- 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

Thank you