# Employee reviews Analysis

****Objective****: To use Hive features for data analysis and sharing the actionable insights into the HR team for taking corrective actions.

****Problem Statement:****The HR team is surfing social media to gather current and ex-employee feedback or sentiments.  
This information gathered will be used to derive actionable insights and take corrective actions to improve the employer-employee relationship. The data is web-scraped from Glassdoor and contains detailed reviews of 67K employees from Google, Amazon, Facebook, Apple, Microsoft, and Netflix.

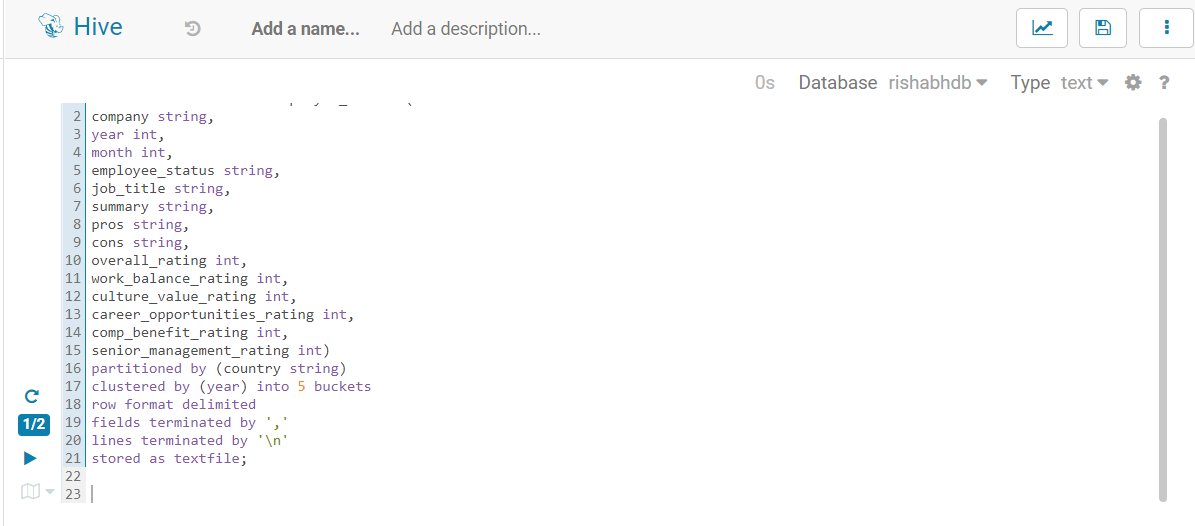
****Domain:**** Human Resource

****Analysis to be done:**** Exploratory analysis, to determine features and relationships impacting employee satisfaction and derive actionable insights by learning from the historical data

****Content:**** This data set contains **employee\_review\_data.csv** separated into the following categories:

1. *Index: Index number*
2. *Company: Name of the company being reviewed*
3. *Location: This dataset is global, as it includes the country&#39;s name in parenthesis [for example,"Toronto, ON(Canada)"]. However, if the location is the USA then it will only include the city and state [i.e. "Los Angeles, CA"].*
4. *Date Posted: Date posted will be in the following format MM DD; YYYY.*
5. *Job-Title: This string will also include whether the reviewer is a "Current" or "Former" Employee at the time of review. Both are fixed-length strings (“Current Employee “and “Former Employee “) followed by the role of reviewer.*
6. *Summary: Short summary of employee review.*
7. *Pros: Pros*
8. *Cons: Cons*
9. *Overall Rating: 1-5*
10. *Work/Life Balance Rating: 1-5*
11. *Culture and Values Rating: 1-5*
12. *Career Opportunities Rating: 1-5*
13. *Comp and Benefits Rating: 1-5*
14. *Senior Management Rating: 1-5*

**Create a hive table partitioned by country and bucketed by year and also load the review.csv file.**



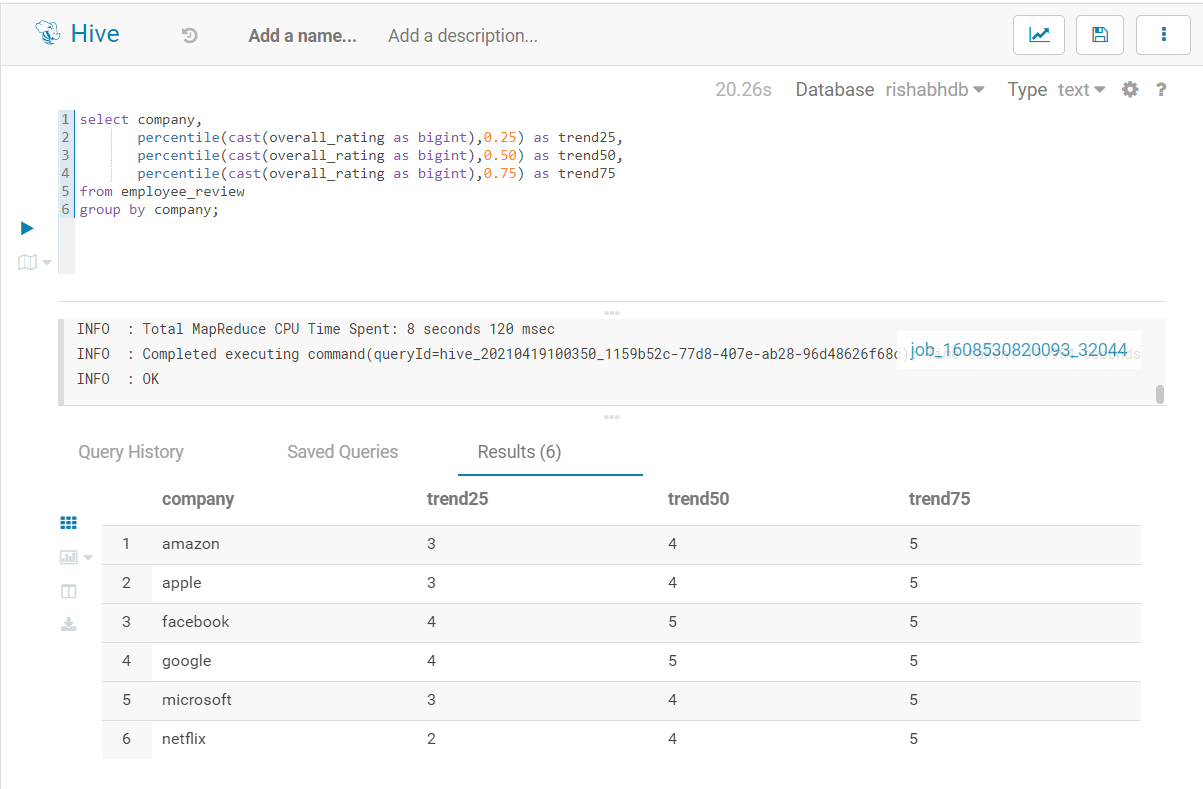
**Impute the missing value (none) for all rating columns with a numerical value between 0 and 5.**



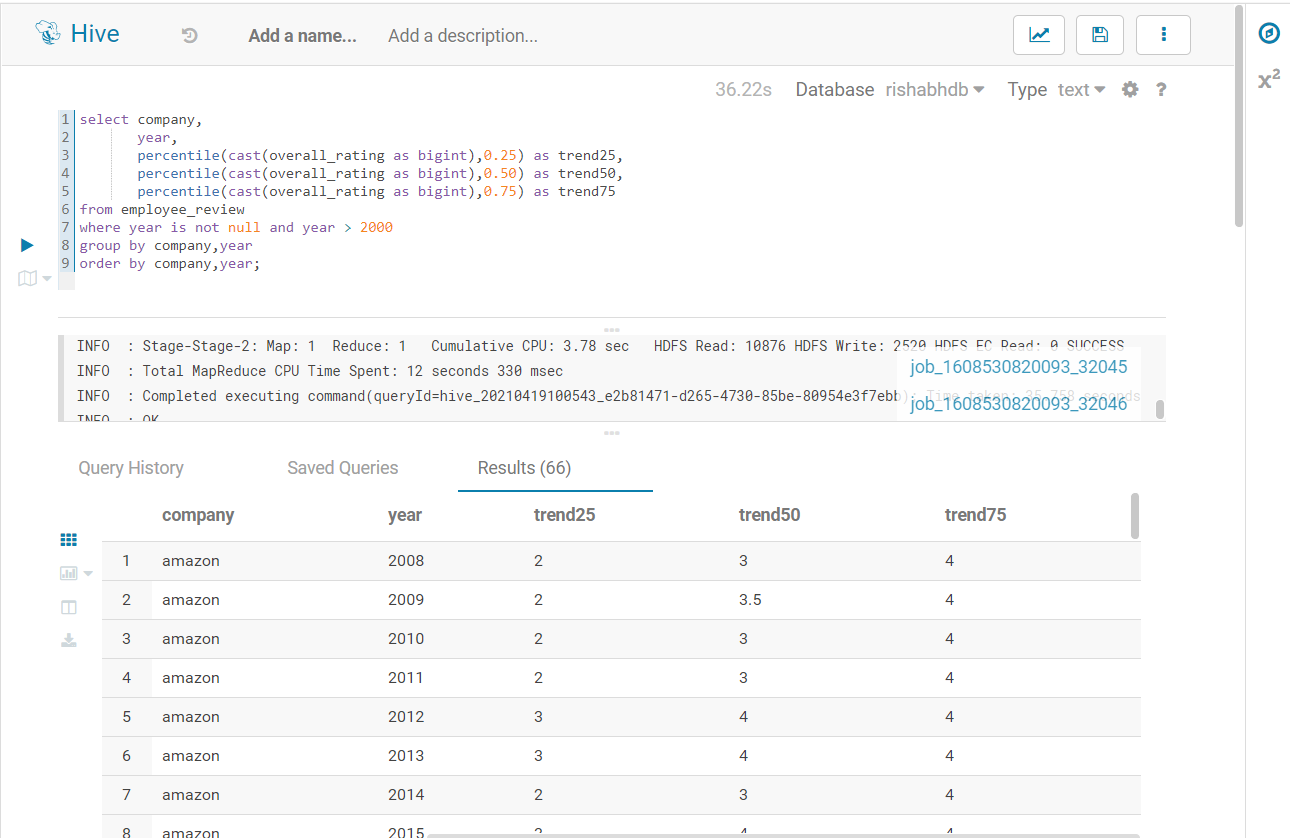
**Write the final relation schema to review.csv file in your HDFS home directory.**



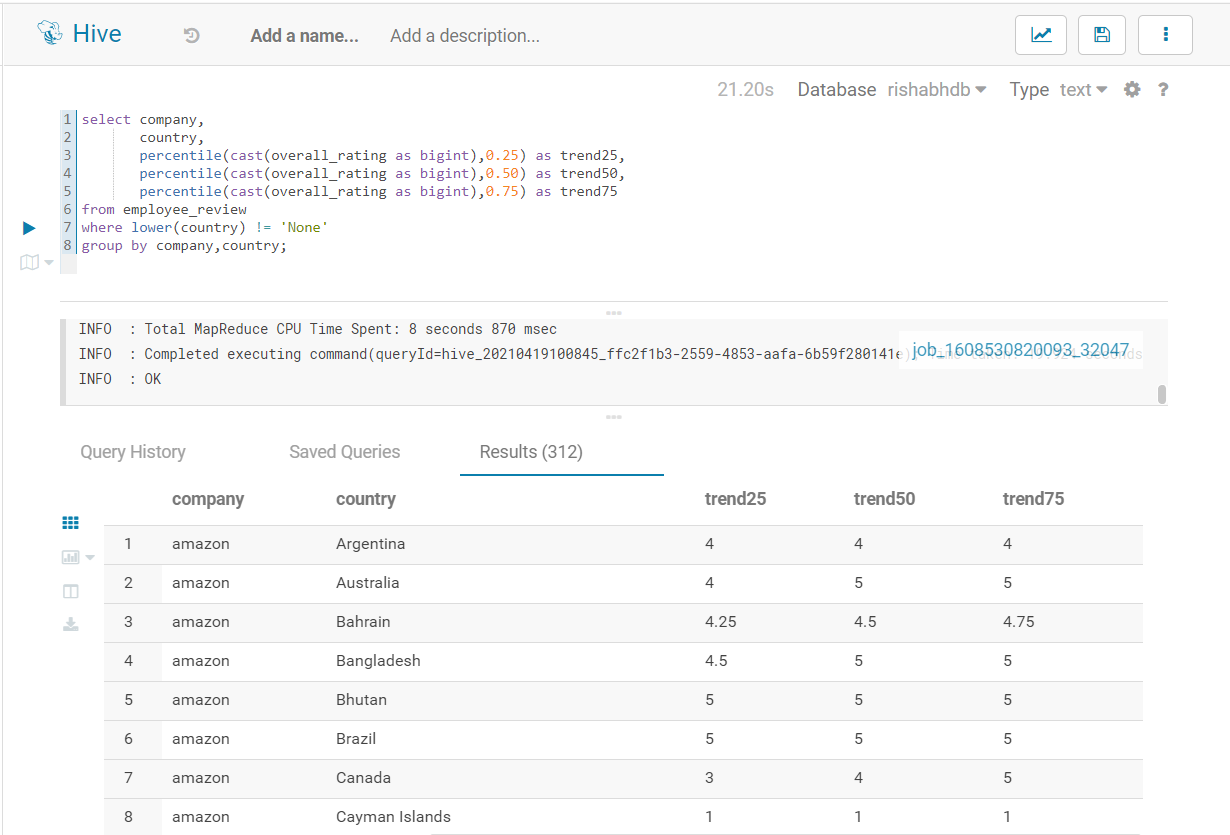
**Using the over-all rating fields display trend globally by company**



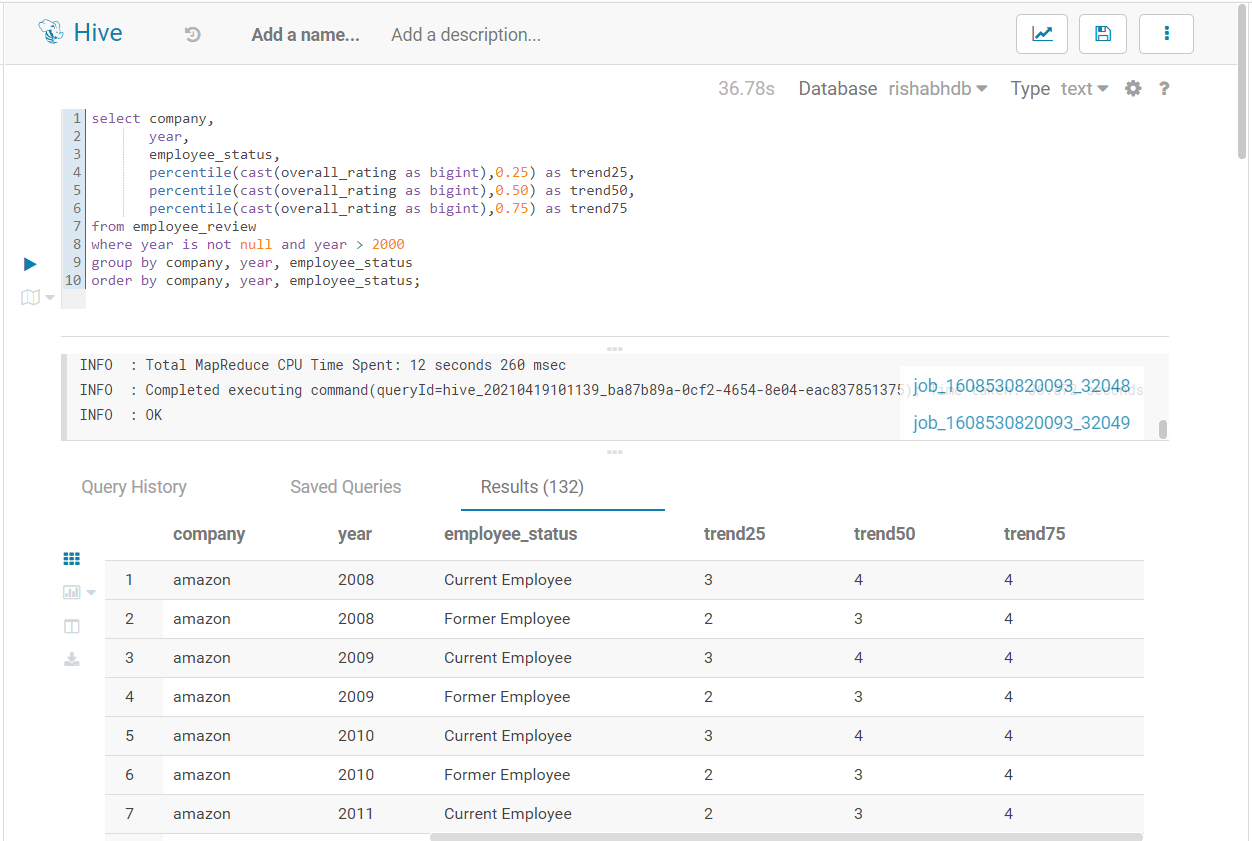
**Using the over-all rating fields display trend globally by company per year**



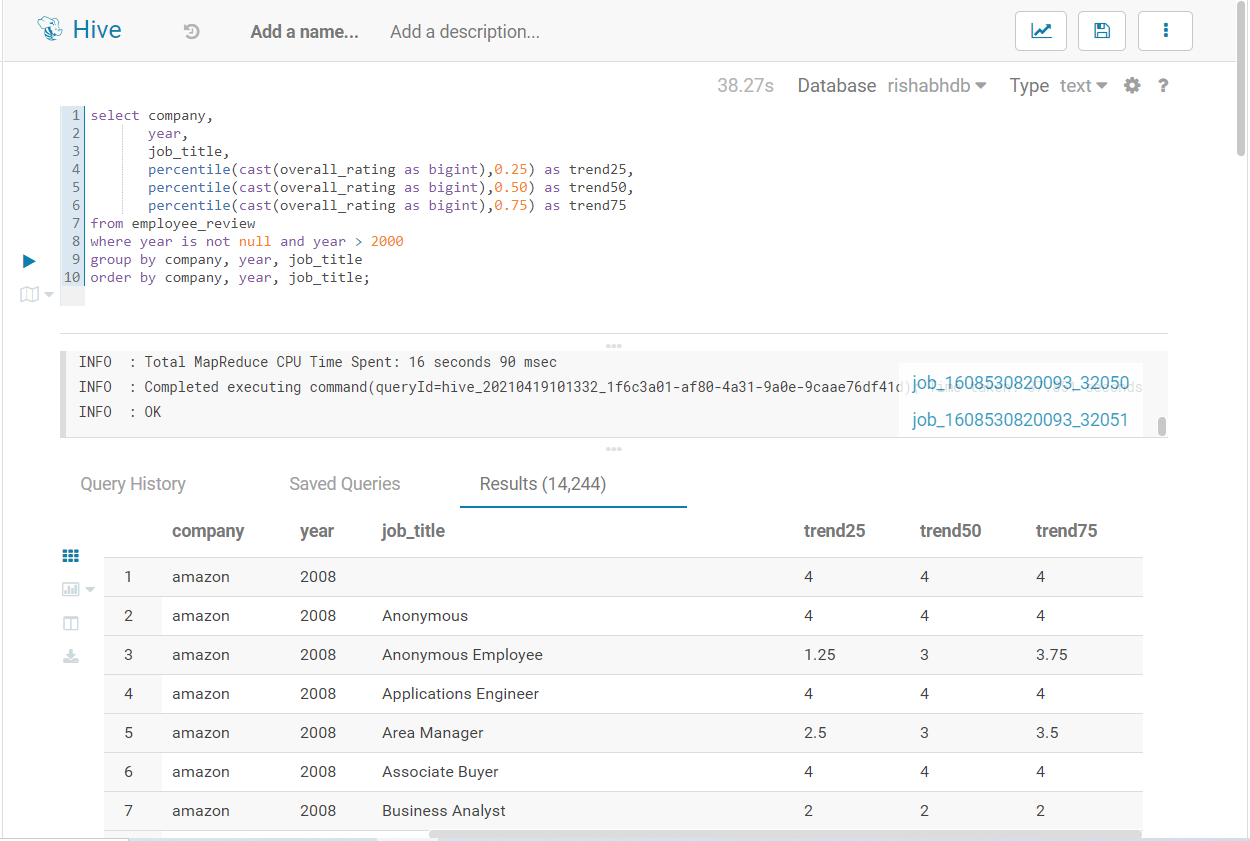
**Using the over-all rating fields display trend globally by company by country**



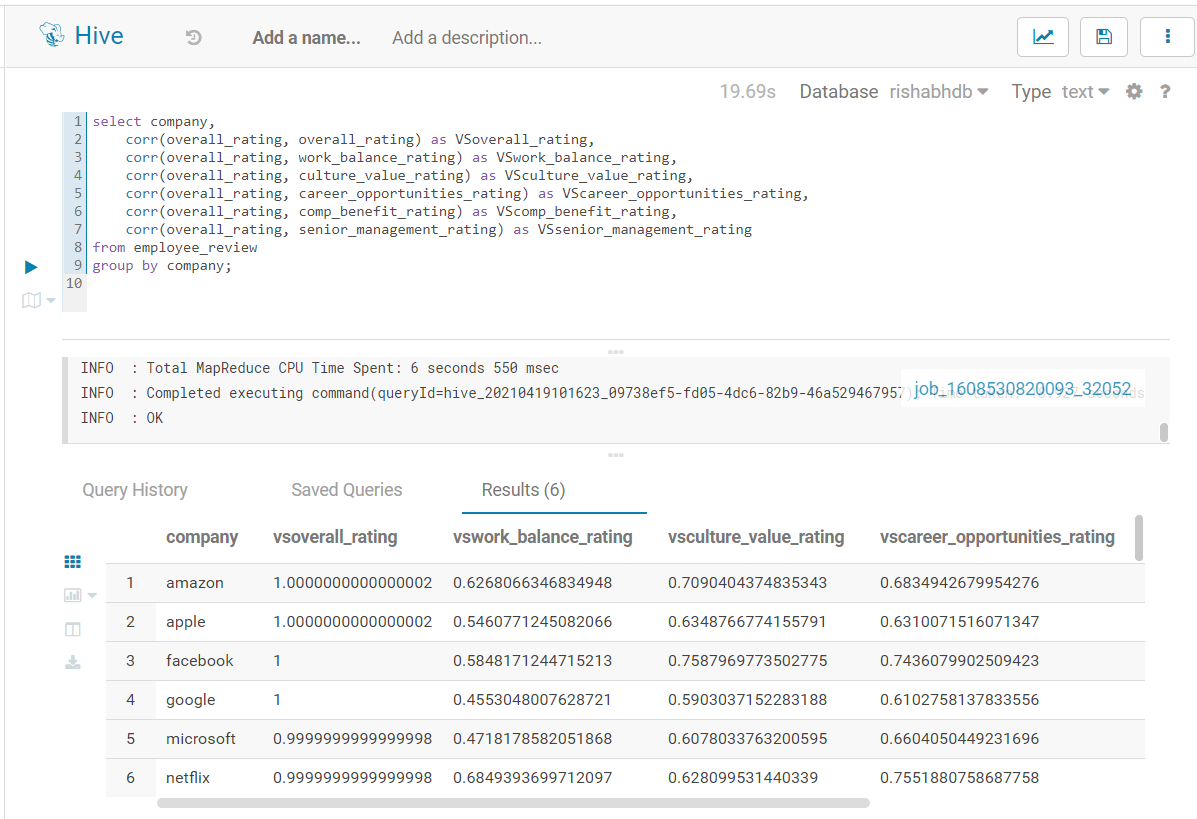
**Display the impact of employee status on rating a company using the overall-ratings field by the company by year.**



**Display the impact of job role on rating a company using the overall-ratings field by the company by year.**



**Display the relationship between the overall rating score vs. the rest of the rating field scores by company. Also, document your findings.**



Observation: Looking at the correlation between overall rating and other rating, I can clearly say that for most of current and former employees of these companies, following were the most and least relevant factors

Fields with high rating

* Career Opportunities
* Senior Management

Fields with low rating

* Work balance
* Company Culture
* Compensation and Benefits