CS 457 - Homework Assignment 5: SQL

rk06451

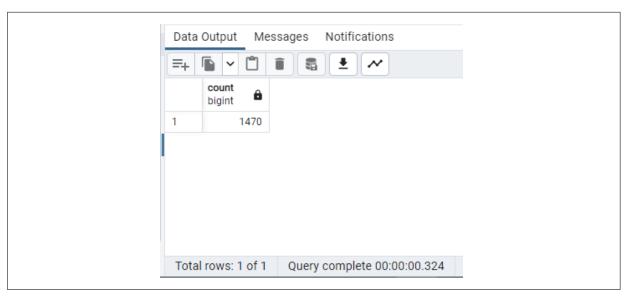
CS 457 Data Science Habib University Spring 2023

1. Create a SQL database and separate tables for both datasets EmployeeAttrition1.csv and EmployeeAttrition2.csv using a RDBMS (PostgreSQL preferred).



- 2. Query the database table for EmployeeAttrition1.csv and interpret the results, displaying:
 - (a) the count of total number of records in the table

Solution:
SQL Query:
select count(EmployeeNumber) from employeeattrition1;

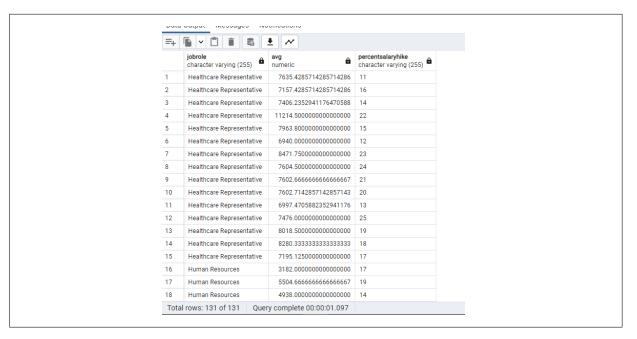


(b) the count of records for each JobRole in descending order of count



(c) the average MonthlyIncome and PercentSalaryHike for each JobRole in ascending order of JobRole

Solution: select JobRole, avg(MonthlyIncome), PercentSalaryHike from employeeattrition1 group by JobRole, PercentSalaryHike Order by JobRole ASC;



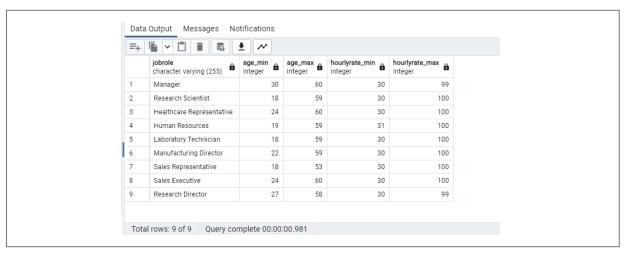
(d) the average JobSatisfaction for each Gender and MaritalStatus



(e) the range (Min and Max) of Age and HourlyRate for each JobRole

Solution:

select JobRole, min(Age) as Age(Min), Max(Age) as Age_Max , min(HourlyRate) as $HourlyRate_Min$, Max(HourlyRate) as $HourlyRate_Max$ from employeeattrition1 group by JobRole;



(f) Join two tables for EmployeeAttrition1.csv and EmployeeAttrition2.csv and display 20 records with the following columns EmployeeNumber, Age, Gender, JobRole, OverTime and Attrition

