Title Page

# DATAANALYSISUSINGSQL–HRINSIGHTS

**Objective**

Theprimarygoalofthisprojectistoconductacomprehensiveanalysisofhuman resources data to gain valuable insights into employee information, recruitment trends, training and development, and employee engagement. The objective is to provide data-driven recommendations and improve HR processes and decision- making.

# Methodology

In the scope of this data analysis project, a series of meticulously planned steps employingSQLtoextractvaluableinsightsfromthedatasetisexecuted.Theproject involvesstagessuchasdatacollection,datacleaningandpreprocessing,exploratory data analysis, querying the database using SQL, aggregating and transforming data, and ultimately deriving meaningful conclusions supported by the data-driven findings.

# DatasetOverviewandData Processing

Theprovideddatasetconsistsofvariousattributesrelatedtoemployeeinformation, recruitment details, training records, and employee engagement survey scores. To ensure the reliability and accuracy of subsequent analyses, comprehensive data preprocessing is conducted. This involves tasks like handling missing values, removingduplicates,andstandardizingdataformatsacrosstables,ensuringthatthe data is consistent and ready for meaningful analysis.

# InsightsandFindings

Various SQLqueries are employed to execute the HR data analysis with efficiency. Thesequeriesincludeaggregatingemployeeinformation,calculatingkeyHRmetrics, categorizing data by different attributes, assess recruitment effectiveness, and understand engagement levels over time. Furthermore, these queries can aid in uncovering insights like the most common reasons for termination and patterns in employee demographics, helping HR professionals make informed decisions about workforce management and engagement strategies.

# KeyStepsPerformedusingSSMS

|  |  |
| --- | --- |
| **training\_and\_development\_data** | |
|  | Employee\_ID |
|  | Training\_Date |
|  | Training\_Program\_Name |
|  | Training\_Type |
|  | Training\_Outcome |
|  | Location |
|  | Trainer |
|  | Training\_Duration\_Days |
|  | Training\_Cost |
|  | |

1. DataRetrieval:UsedSELECTstatementstoextractrelevantdatafromtheHR tables.
2. DataFiltering:UsedWHEREclausetofilterdatabasedonspecificcriteria, such as employee status or training type.
3. DataTransformation:EmployedlogicaloperatorslikeAND,OR,andNOT for more advanced data manipulation.
4. DataSorting:UtilizedtheORDERBYclausetosortqueryresults,facilitating the identification of patterns or priorities.
5. DataIntegration:UsedJOINStocombinedatafrommultipletablesfor comprehensive analysis.
6. DataAggregation:UsedGROUPBYstatementstoaggregatedata,calculate metrics, and create reports for further analysis.

# Entity-RelationshipDiagram

AnEntity-RelationshipDiagram(ERD)iscreatedforthisHRdatasettovisually represent the structure and relationships between the various tables.

|  |  |
| --- | --- |
| **employee\_data** | |
|  | Employee\_ID |
|  | FirstName |
|  | LastName |
|  | StartDate |
|  | ExitDate |
|  | Title |
|  | Supervisor |
|  | ADEmail |
|  | BusinessUnit |
|  | EmployeeStatus |
|  | EmployeeType |
|  | PayZone |
|  | EmployeeClassificationType |
|  | TerminationType |
|  | TerminationDescription |
|  | DepartmentType |
|  | Division |
|  | DOB |
|  | State |
|  | JobFunctionDescription |
|  | GenderCode |
|  | LocationCode |
|  | RaceDesc |
|  | MaritalDesc |
|  | Performance\_Score |
|  | Current\_Employee\_Rating |
|  | |

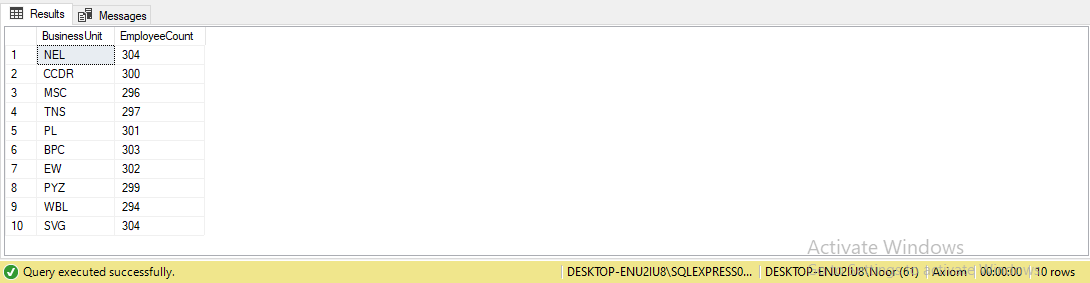
|  |  |
| --- | --- |
| **employee\_engagement\_survey\_data** | |
|  | Employee\_ID |
|  | Survey\_Date |
|  | Engagement\_Score |
|  | Satisfaction\_Score |
|  | Work\_Life\_Balance\_Score |
|  | |

|  |  |
| --- | --- |
| **recruitment\_data** | |
|  | Applicant\_ID |
|  | Application\_Date |
|  | First\_Name |
|  | Last\_Name |
|  | Gender |
|  | Date\_of\_Birth |
|  | Phone\_Number |
|  | Email |
|  | Address |
|  | City |
|  | State |
|  | Zip\_Code |
|  | Country |
|  | Education\_Level |
|  | Years\_of\_Experience |
|  | Desired\_Salary |
|  | Job\_Title |
|  | Status |
|  | |

# PERFORMEDQUERIES

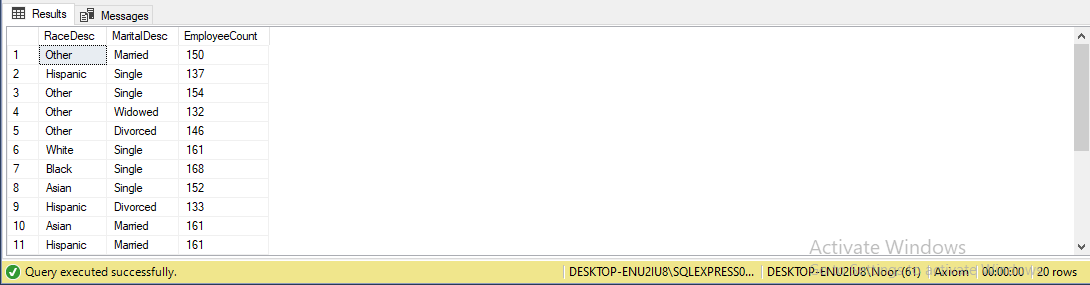
1. **CountofEmployeesbyBusinessUnit**

SELECTBusinessUnit,COUNT(\*)ASEmployeeCountFROM employee\_data

GROUPBYBusinessUnit

# CountofEmployeesbyRaceandMaritalStatus

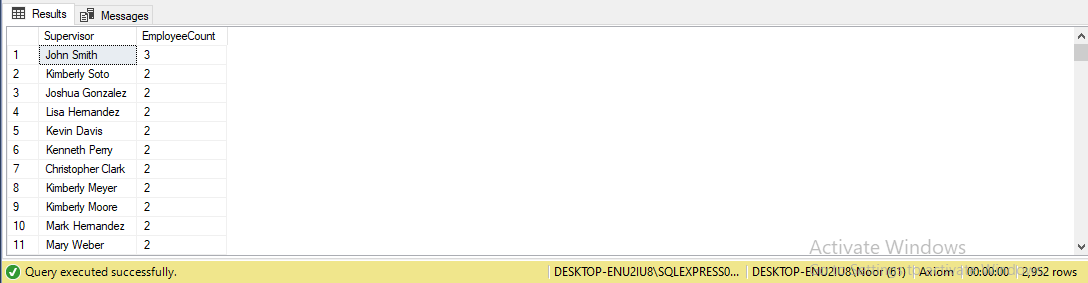
SELECTRaceDesc,MaritalDesc,COUNT(\*)ASEmployeeCountFROM employee\_data

GROUPBYRaceDesc, MaritalDesc

# NumberofEmployeesperSupervisor

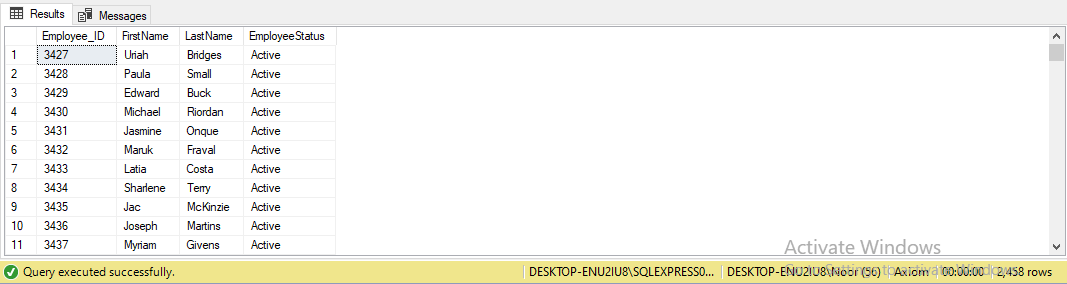
SELECTSupervisor,COUNT(\*)ASEmployeeCountFROM employee\_data

GROUPBYSupervisor

ORDERBYEmployeeCountDESC

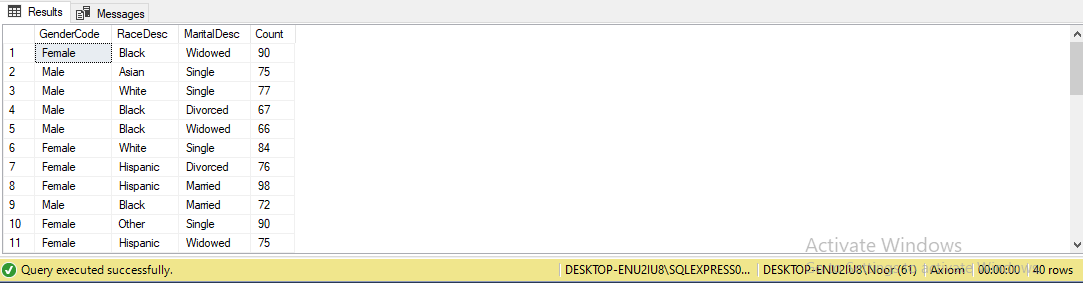
# ListofActiveEmployees

SELECTEmp**loyee\_**ID,FirstName,LastName,EmployeeStatusFROM employee\_data

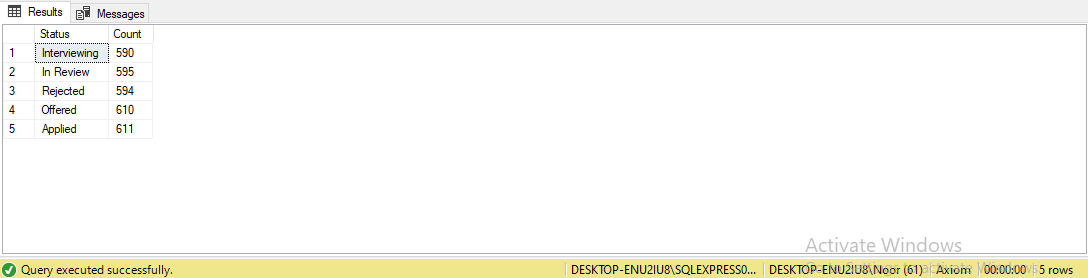
WHEREEmployeeStatus='Active'

# EmployeeDemographics

SELECTGenderCode,RaceDesc,MaritalDesc,COUNT(\*)ASCount FROM employee\_data

GROUPBYGenderCode,RaceDesc,MaritalDesc

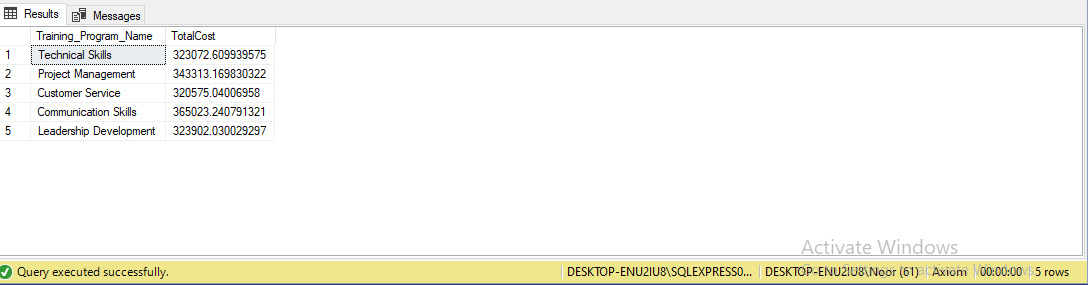
1. **Recruitment Status Breakdown** SELECTStatus,COUNT(\*)ASCount FROM recruitment\_data

GROUPBYStatus

# TrainingCostsbyProgram

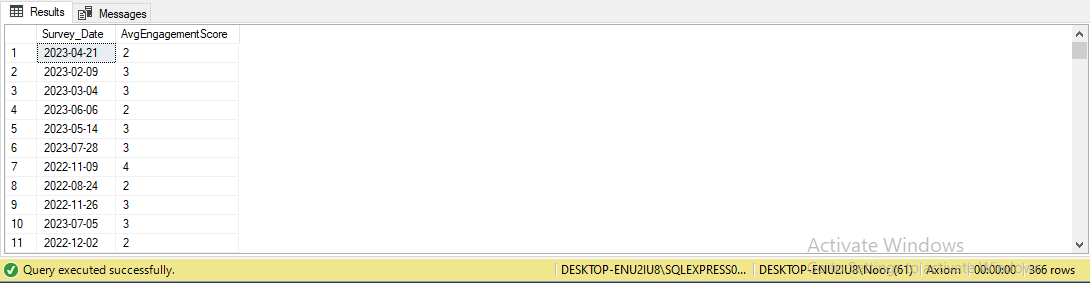
SELECTTraining\_Program\_Name,SUM(Training\_Cost)ASTotalCostFROM training\_and\_development\_data

GROUPBYTraining\_Program\_Name



# EmployeeEngagementScores

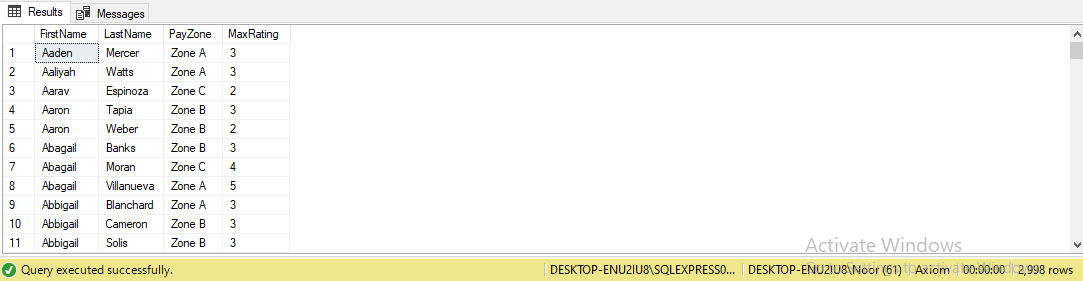
SELECTSurvey\_Date,AVG(Engagement\_Score)ASAvgEngagementScoreFROM employee\_engagement\_survey\_data

GROUPBYSurvey\_Date

# Highest-PaidEmployees

SELECTFirstName,LastName,PayZone,MAX(Current\_Employee\_Rating) AS MaxRating

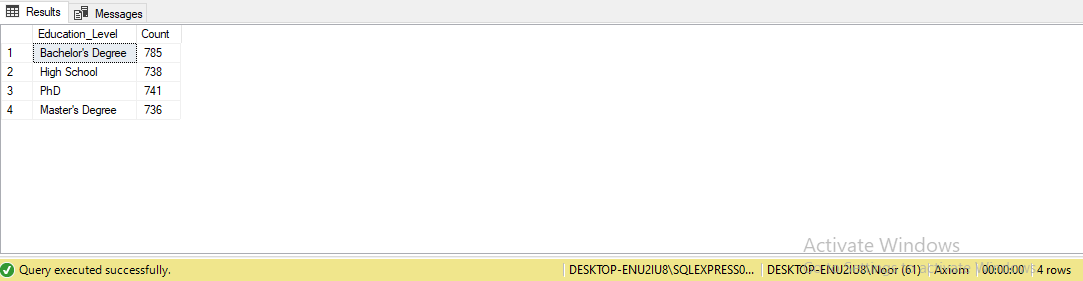
FROMemployee\_data

GROUPBYFirstName,LastName,PayZone

# ApplicantsbyEducationLevel

SELECTEducation\_Level,COUNT(\*)ASCount FROM recruitment\_data

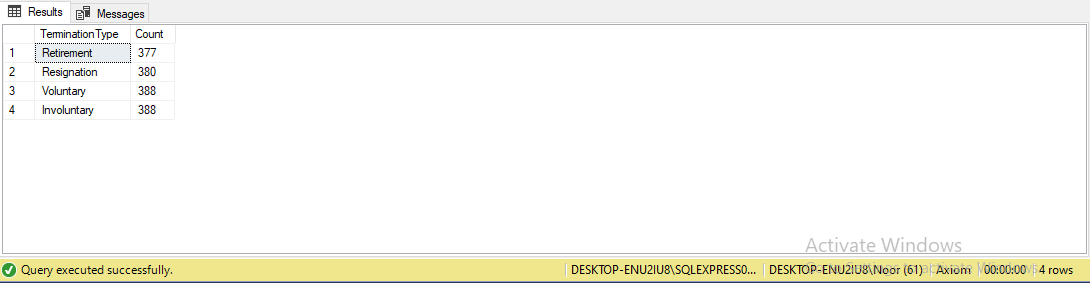
GROUPBYEducation\_Level



# EmployeeExit Reasons

SELECTTerminationType,COUNT(\*)ASCount FROM employee\_data

WHEREExitDateISNOTNULL

GROUPBYTerminationType

# AverageDesiredSalarybyJobTitle

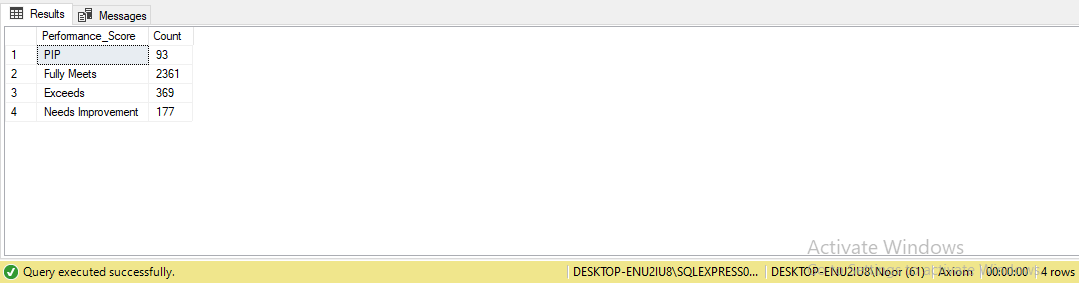
SELECTJob\_Title,AVG(Desired\_Salary)ASAvgSalaryFROM recruitment\_data

GROUPBYJob\_Title

# EmployeePerformance Distribution

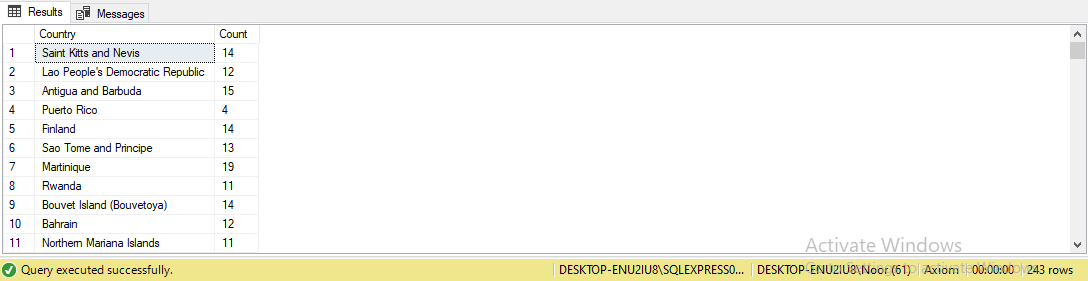
SELECTPerformance\_Score,COUNT(\*)ASCount FROM employee\_data

GROUPBYPerformance\_Score



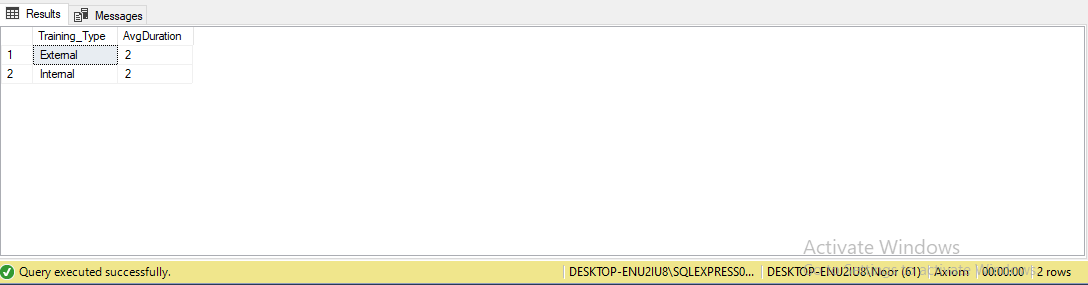
# Applicantsby Country

SELECTCountry,COUNT(\*)ASCount FROM recruitment\_data

GROUPBYCountry

# TrainingDurationbyType

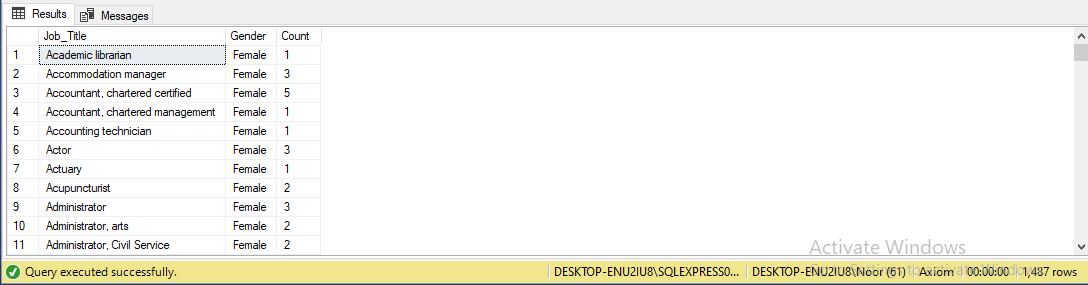
SELECTTraining\_Type,AVG(Training\_Duration\_Days)ASAvgDurationFROM training\_and\_development\_data

GROUPBYTraining\_Type

# RecruitmentbyJobTitleandGender

SELECTJob\_Title,Gender,COUNT(\*)ASCount FROM recruitment\_data

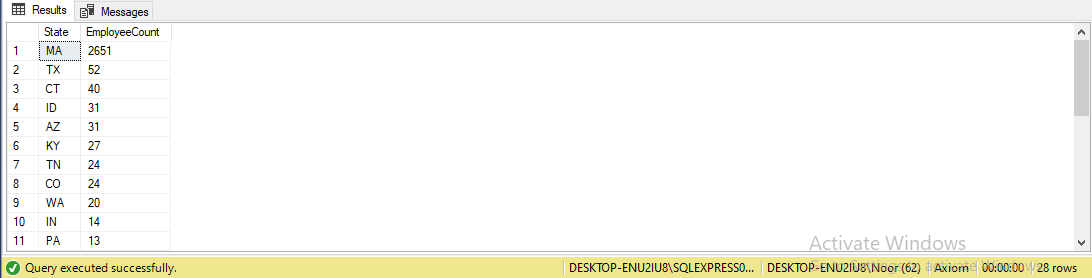
GROUPBYJob\_Title,Gender



# EmployeeCountby State

SELECTState,COUNT(\*)ASEmployeeCountFROM employee\_data

GROUPBYState

ORDERBYEmployeeCountDESC

# TrainingCostsbyLocation

SELECTLocation,SUM(Training\_Cost)ASTotalCostFROM training\_and\_development\_data

GROUP BY Location ORDERBYTotalCostDESC

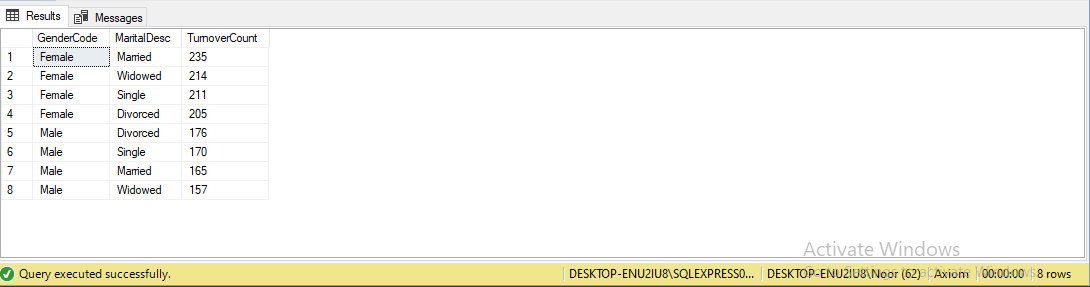


# EmployeeTurnoverbyGenderandMaritalStatus

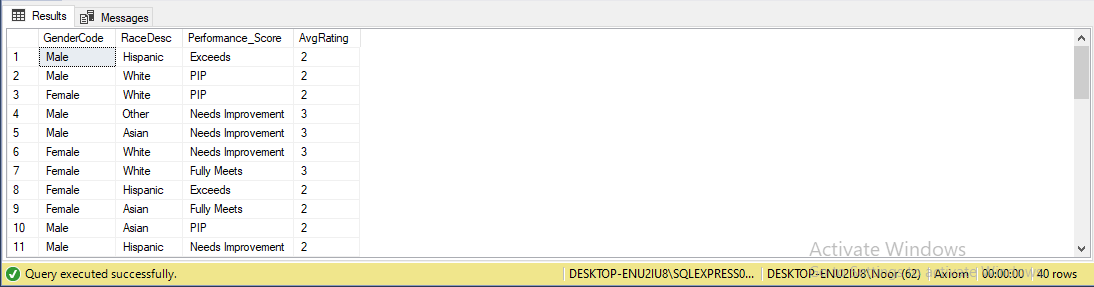
SELECTGenderCode,MaritalDesc,COUNT(\*)ASTurnoverCountFROM employee\_data

WHEREExitDateISNOTNULL

GROUPBYGenderCode,MaritalDescORDER BY TurnoverCountDESC



1. **Employee Demographics and Performance**SELECTGenderCode,RaceDesc,Performance\_Score, AVG(Current\_Employee\_Rating) AS AvgRatingFROM employee\_data

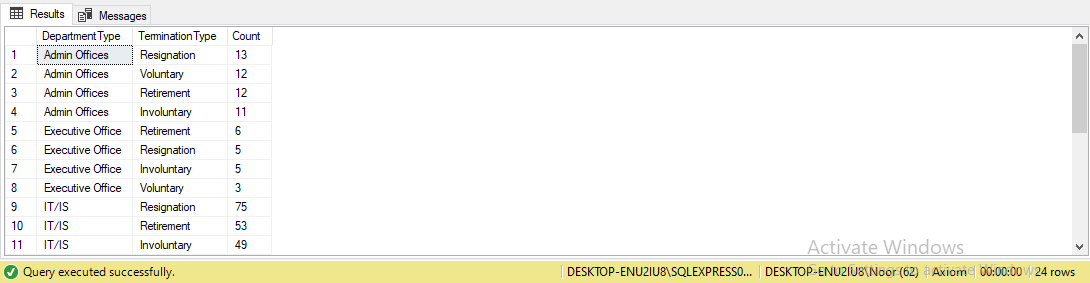
GROUPBYGenderCode,RaceDesc,Performance\_Score

# EmployeeExitReasonsbyDepartment

SELECTDepartmentType,TerminationType,COUNT(\*)ASCount FROM employee\_data

WHEREExitDateISNOTNULL

GROUPBYDepartmentType,TerminationTypeORDER BY DepartmentType, Count DESC



# AverageDesiredSalarybyEducation Level

SELECTEducation\_Level,AVG(Desired\_Salary)ASAvgSalaryFROM recruitment\_data

GROUP BY Education\_LevelORDERBYAvgSalaryDESC

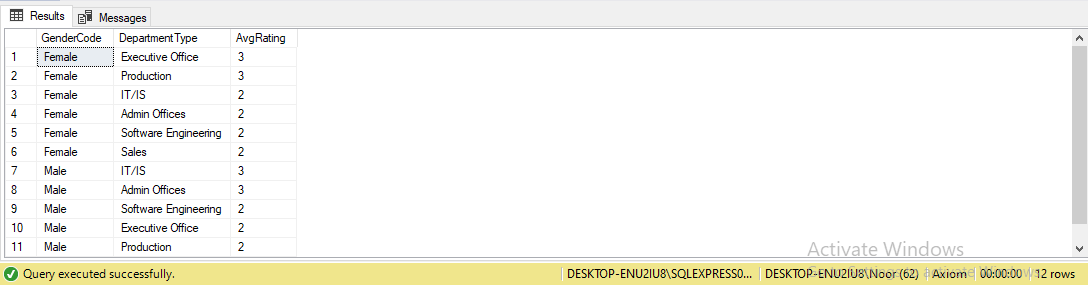


# EmployeePerformancebyGenderand Department

SELECTGenderCode,DepartmentType,AVG(Current\_Employee\_Rating) AS AvgRating

FROMemployee\_data

GROUP BY GenderCode, DepartmentTypeORDERBYGenderCode,AvgRatingDESC

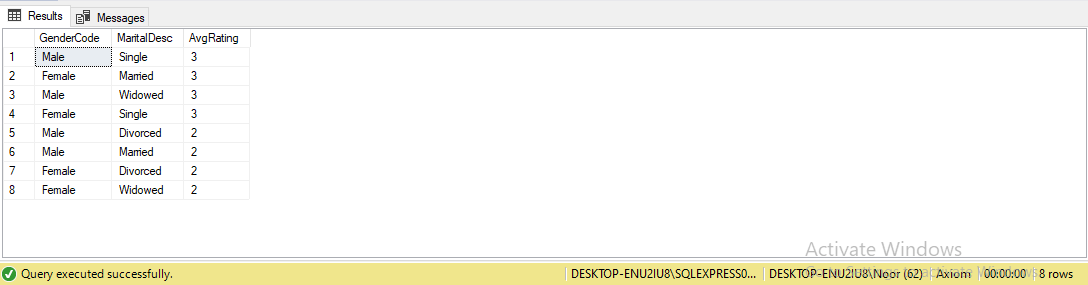


# AverageEmployeeRatingbyGenderandMaritalStatus

SELECTGenderCode,MaritalDesc,AVG(Current\_Employee\_Rating)AS AvgRating

FROMemployee\_data

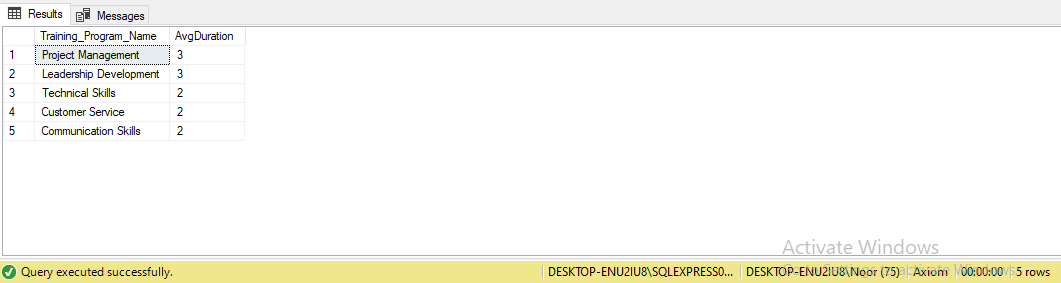
GROUPBYGenderCode,MaritalDescORDER BY AvgRatingDESC



# AverageTrainingDurationbyTrainingProgramType

SELECTTraining\_Program\_Name,AVG(Training\_Duration\_Days)AS AvgDuration

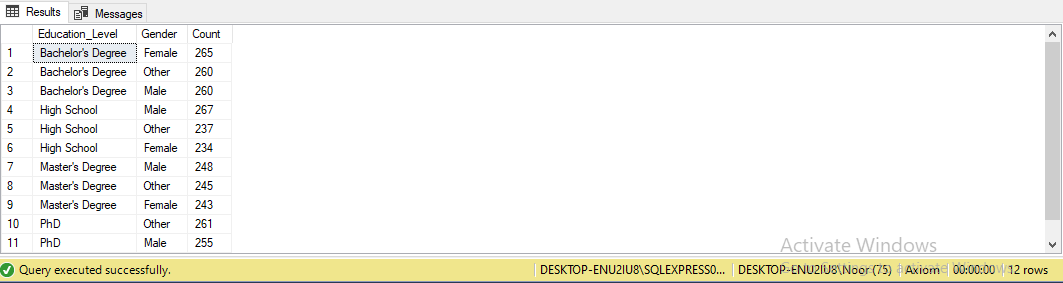
FROMtraining\_and\_development\_dataGROUP BY Training\_Program\_NameORDER BY AvgDurationDESC



# RecruitmentbyEducationLeveland Gender

SELECTEducation\_Level,Gender,COUNT(\*)ASCount FROM recruitment\_data

GROUP BY Education\_Level, Gender ORDERBYEducation\_Level,CountDESC

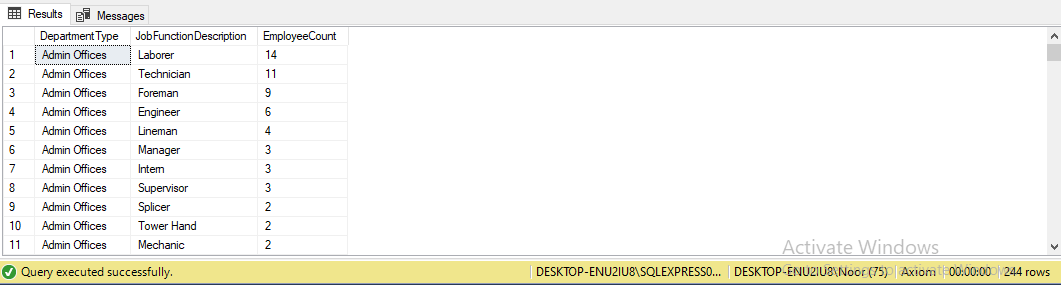


# EmployeeCountbyDepartmentandJob Title

SELECTDepartmentType,JobFunctionDescription,COUNT(\*)AS EmployeeCount

FROMemployee\_data

GROUPBYDepartmentType,JobFunctionDescriptionORDER BY DepartmentType, EmployeeCountDESC



# TotalTrainingCostsbyBusinessUnit

SELECTed.BusinessUnit,SUM(tdd.Training\_Cost)ASTotalCostFROM employee\_data ed

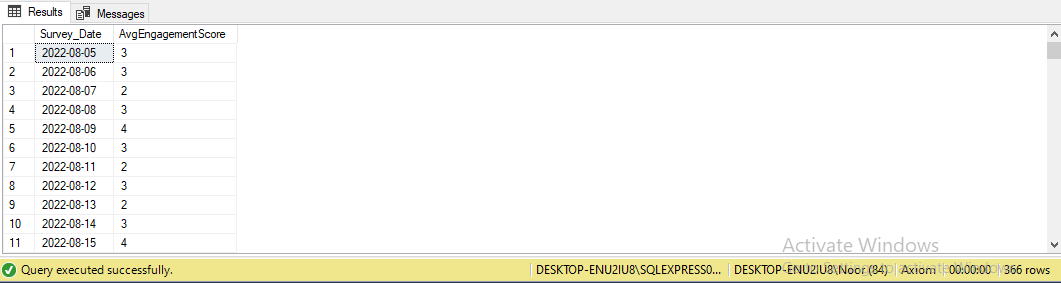
JOINtraining\_and\_development\_datatddONed.EmpID=tdd.Employee\_IDGROUP BY ed.BusinessUnit

ORDERBYTotalCostDESC

# EmployeeEngagement TrendsOver Time

SELECTSurvey\_Date,AVG(Engagement\_Score)ASAvgEngagementScoreFROM employee\_engagement\_survey\_data

GROUPBYSurvey\_DateORDER BYSurvey\_Date



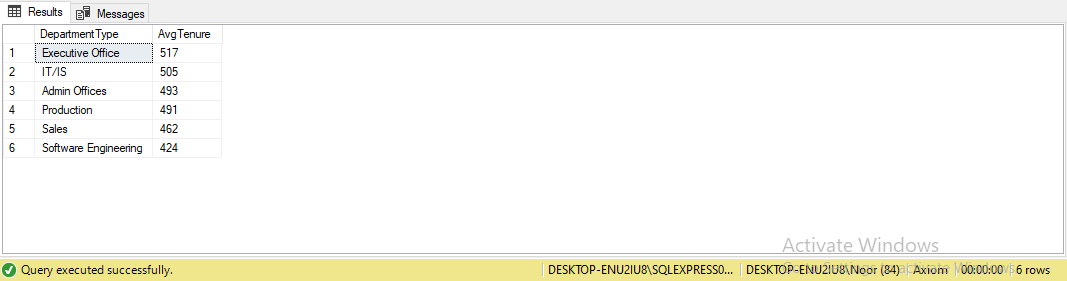
# AverageTenurebyDepartment

SELECTDepartmentType,AVG(DATEDIFF(day,StartDate,ExitDate))AS AvgTenure

FROMemployee\_data

WHEREExitDateISNOTNULL

GROUP BY DepartmentTypeORDERBYAvgTenureDESC



# RecruitmentbyCityand State

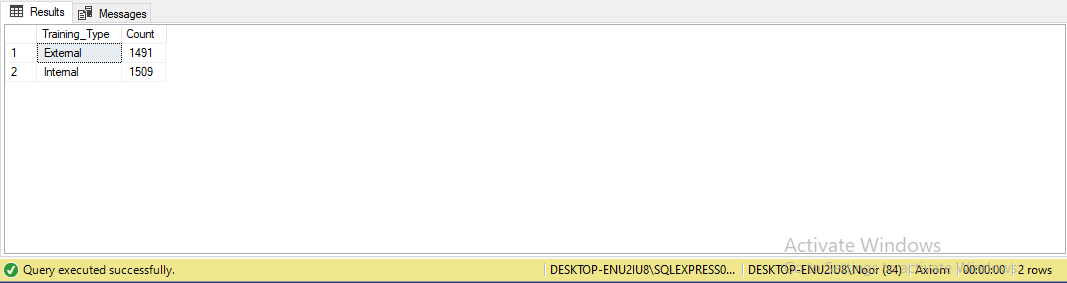
SELECTCity,State,COUNT(\*)ASCount FROM recruitment\_data

GROUP BY City, State ORDERBYCountDESC



# TrainingTypeDistribution

SELECTTraining\_Type,COUNT(\*)ASCount FROM training\_and\_development\_dataGROUP BY Training\_Type

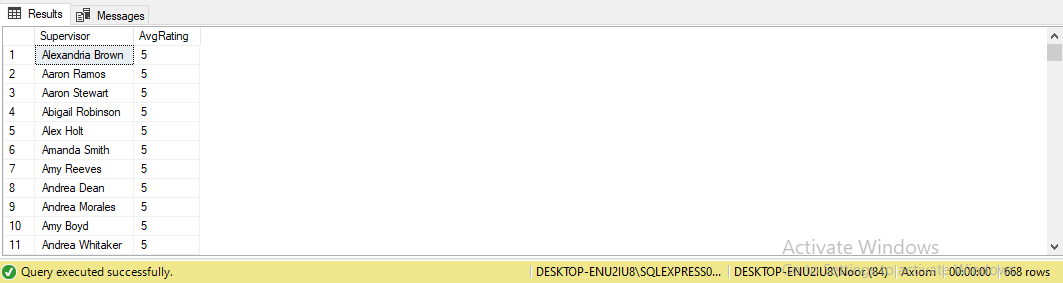


# Top-performingSupervisorsbyEmployeeRating

SELECTSupervisor,AVG(Current\_Employee\_Rating)ASAvgRatingFROM employee\_data

GROUPBYSupervisor

HAVINGAVG(Current\_Employee\_Rating)>=4.0 ORDER BY AvgRatingDESC



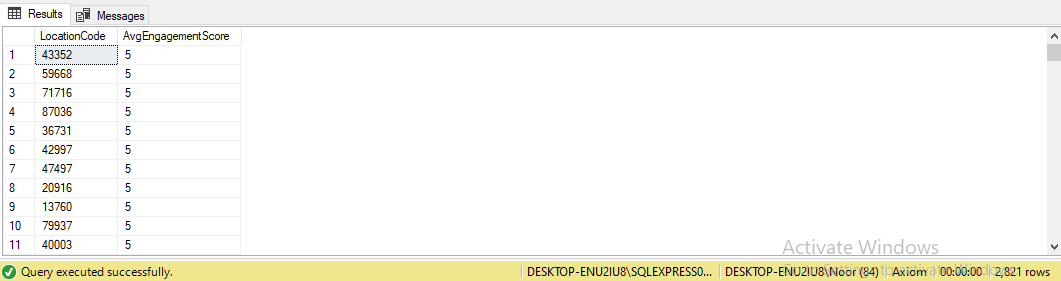
# EmployeeEngagement byLocation

SELECTe.LocationCode,AVG(ee.Engagement\_Score)AS AvgEngagementScore

FROMemployee\_datae

JOINemployee\_engagement\_survey\_dataeeONe.Employee\_ID= ee.Employee\_ID

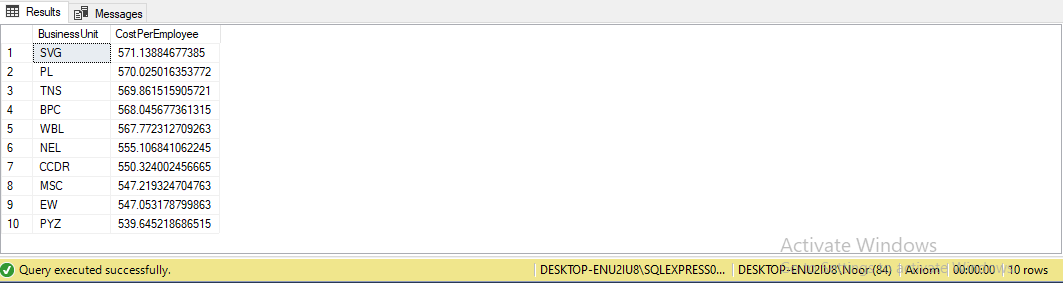
GROUPBYe.LocationCode

ORDERBYAvgEngagementScoreDESC

1. **Training Cost per Employee by Business Unit** SELECTed.BusinessUnit,SUM(tdd.Training\_Cost)/ COUNT(ed.Employee\_ID) AS CostPerEmployeeFROM employee\_data ed

JOINtraining\_and\_development\_datatddONed.Employee\_ID= tdd.Employee\_ID

GROUP BY ed.BusinessUnitORDERBYCostPerEmployeeDESC



# AverageDesiredSalarybyGenderandEducationLevel

SELECTGender,Education\_Level,AVG(Desired\_Salary)ASAvgSalaryFROM recruitment\_data

GROUPBYGender,Education\_LevelORDER BY AvgSalaryDESC



# EmployeesandTheirTrainingPrograms

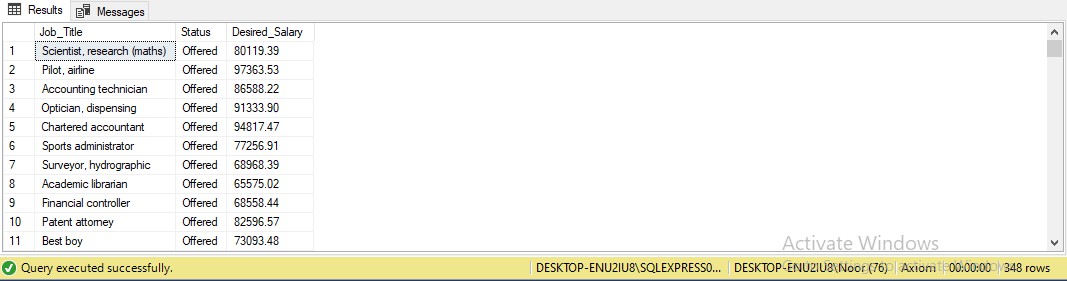
SELECTed.FirstName,ed.LastName,tdd.Training\_Program\_NameFROM employee\_data ed

JOINtraining\_and\_development\_datatddONed.Employee\_ID= tdd.Employee\_ID

WHEREtdd.Training\_Program\_NameISNOTNULL

1. **RecruitmentStatusandDesiredSalarybyJobTitle** SELECT rd.Job\_Title, rd.Status, rd.Desired\_SalaryFROM recruitment\_datard

WHERErd.Status='Offered'ANDrd.Desired\_Salary>=60000

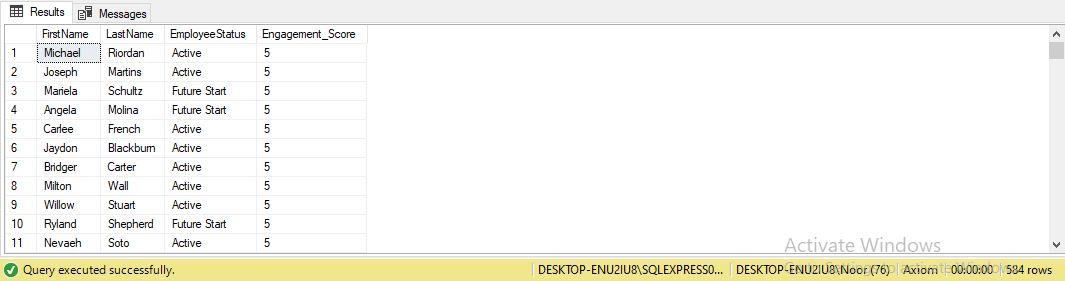


# EmployeeswithHighEngagementScores

SELECTed.FirstName,ed.LastName,ed.EmployeeStatus, ees.Engagement\_Score

FROMemployee\_dataed

JOINemployee\_engagement\_survey\_dataeesONed.Employee\_ID= ees.Employee\_ID

WHEREees.Engagement\_Score>=4.5

# EmployeesandTheirTrainingCosts

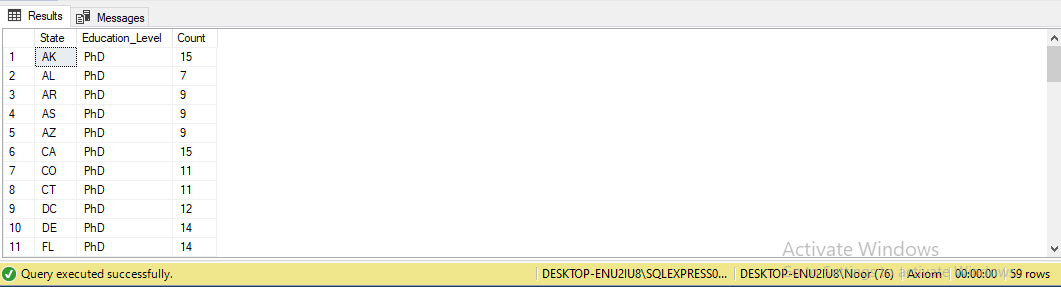
SELECTed.FirstName,ed.LastName,tdd.Training\_CostFROM employee\_data ed

JOINtraining\_and\_development\_datatddONed.Employee\_ID= tdd.Employee\_ID

WHEREtdd.Training\_Cost>500

1. **Recruitment by State and Highest Education Level** SELECTrd.State,rd.Education\_Level,COUNT(\*)ASCount FROM recruitment\_datard

WHERE rd.Education\_Level= 'PhD' GROUPBYrd.State,rd.Education\_LevelORDER BY State, Count DESC

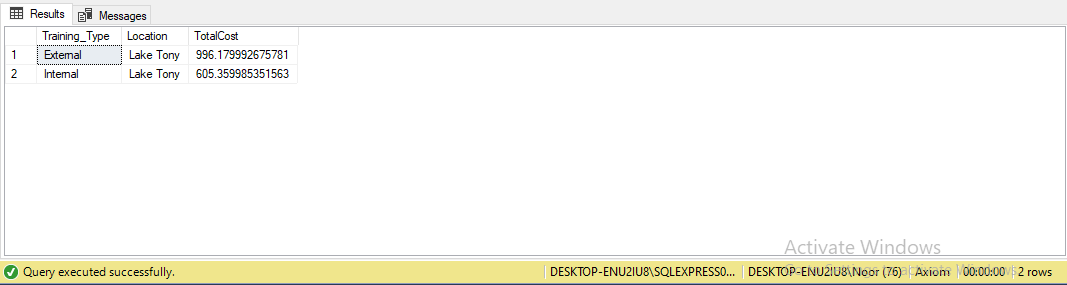


# TrainingCostsbyTrainingTypeandSpecificLocation

SELECTtdd.Training\_Type,tdd.Location,SUM(tdd.Training\_Cost)AS TotalCost

FROMtraining\_and\_development\_datatddWHERE tdd.Location= 'Lake Tony'

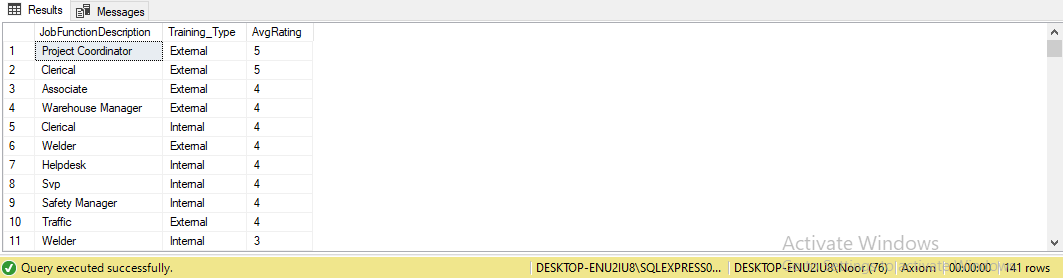
GROUPBYtdd.Training\_Type,tdd.LocationORDER BY TotalCostDESC



1. **EmployeePerformancebyJobTitleandTrainingType** SELECT ed.JobFunctionDescription, tdd.Training\_Type, AVG(ed.Current\_Employee\_Rating) AS AvgRatingFROM employee\_data ed

JOINtraining\_and\_development\_datatddONed.Employee\_ID= tdd.Employee\_ID

GROUPBYed.JobFunctionDescription,tdd.Training\_TypeORDER BY AvgRatingDESC



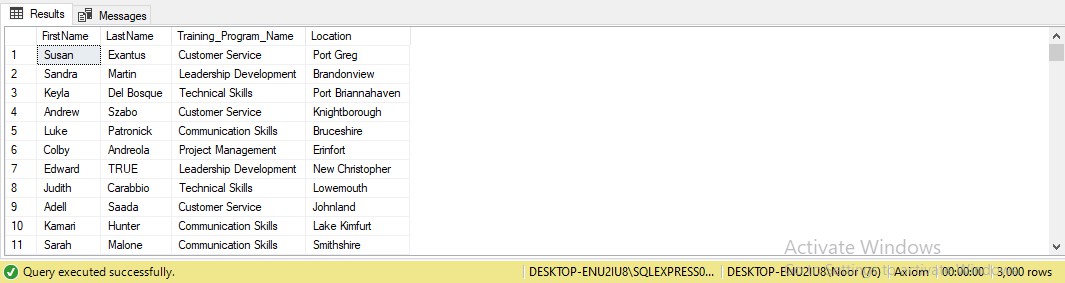
# TrainingProgramsbyEmployeeandLocation

SELECTed.FirstName,ed.LastName,tdd.Training\_Program\_Name, tdd.Location

FROMemployee\_dataed

JOINtraining\_and\_development\_datatddONed.Employee\_ID= tdd.Employee\_ID

WHEREtdd.Training\_Program\_NameISNOTNULLANDtdd.LocationIS NOT NULL



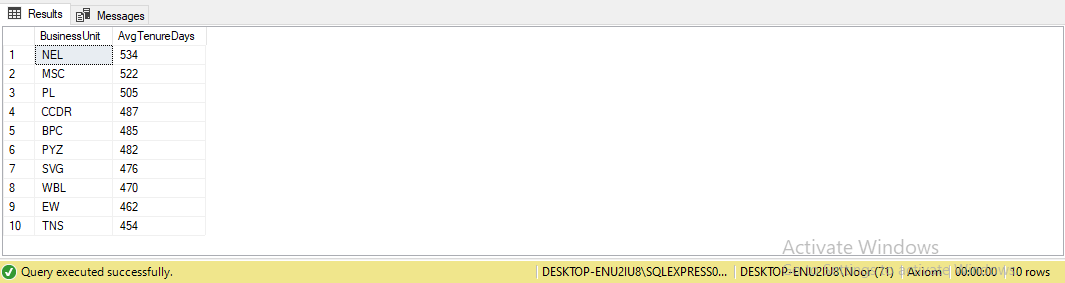
# EmployeeTenureDistributionbyBusinessUnit

SELECTBusinessUnit,AVG(DATEDIFF(day,StartDate,ExitDate))AS AvgTenureDays

FROMemployee\_data

WHEREExitDateISNOTNULL

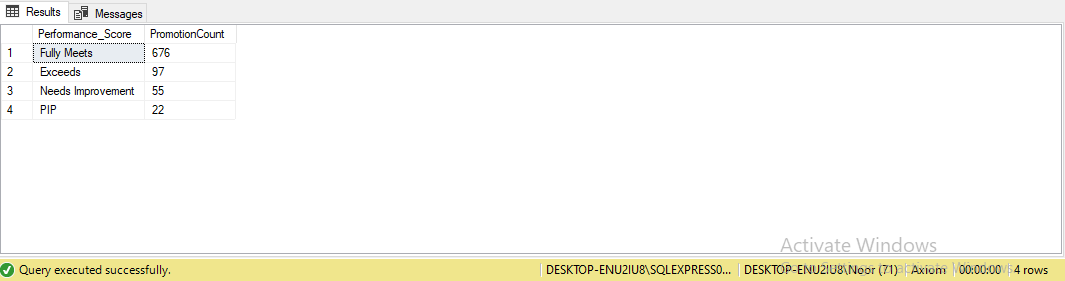
GROUPBYBusinessUnit

ORDERBYAvgTenureDaysDESC

# EmployeePromotionAnalysis

SELECTPerformance\_Score,COUNT(\*)ASPromotionCountFROM employee\_data

WHEREEmployeeStatus='Active'ANDEmployeeType='Full-Time' GROUP BY Performance\_Score

ORDERBYPromotionCountDESC

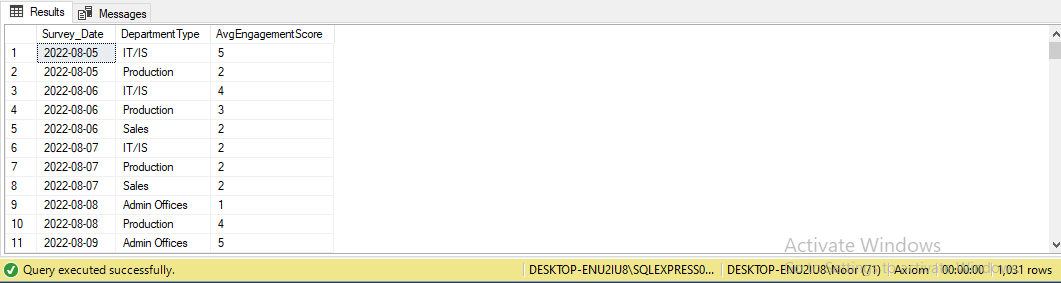
# EmployeeEngagementTrendbyDepartment

SELECTee.Survey\_Date,ed.DepartmentType,AVG(ee.Engagement\_Score) AS AvgEngagementScore

FROMemployee\_dataed

JOINemployee\_engagement\_survey\_dataeeONed.Employee\_ID= ee.Employee\_ID

GROUPBYee.Survey\_Date,ed.DepartmentTypeORDER BY ee.Survey\_Date

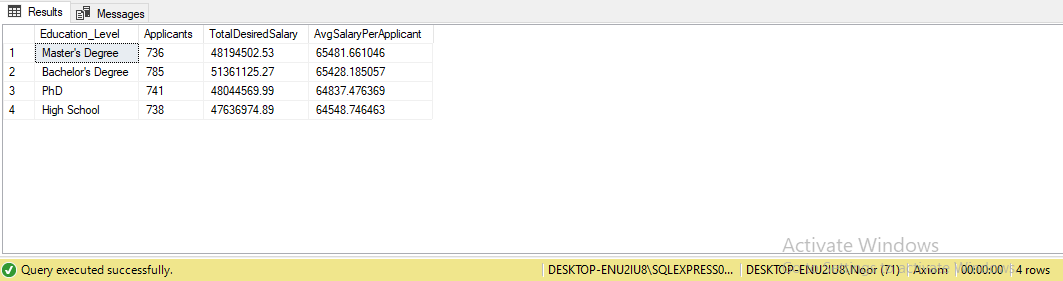


# RecruitmentCostAnalysis

SELECT rd.Education\_Level, COUNT(\*) AS Applicants, SUM(Desired\_Salary)ASTotalDesiredSalary,SUM(Desired\_Salary)/ COUNT(\*) AS AvgSalaryPerApplicant

FROM recruitment\_datardGROUPBYrd.Education\_Level

ORDERBYAvgSalaryPerApplicantDESC



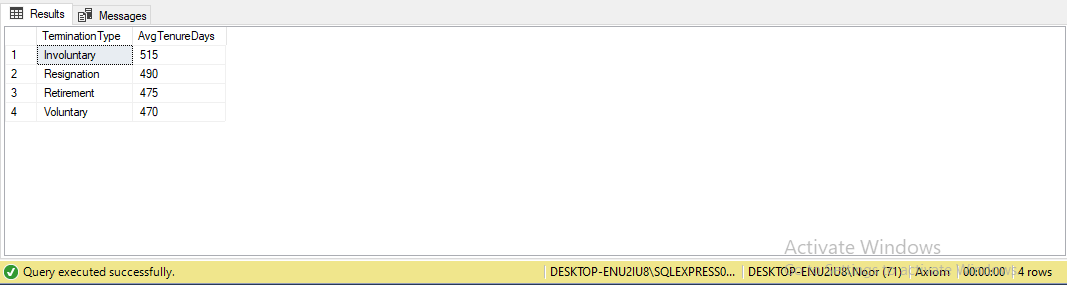
# EmployeeExitPatternsbyTenure

SELECTTerminationType,AVG(DATEDIFF(day,StartDate,ExitDate))AS AvgTenureDays

FROMemployee\_data

WHEREExitDateISNOTNULL

GROUP BY TerminationTypeORDERBYAvgTenureDaysDESC

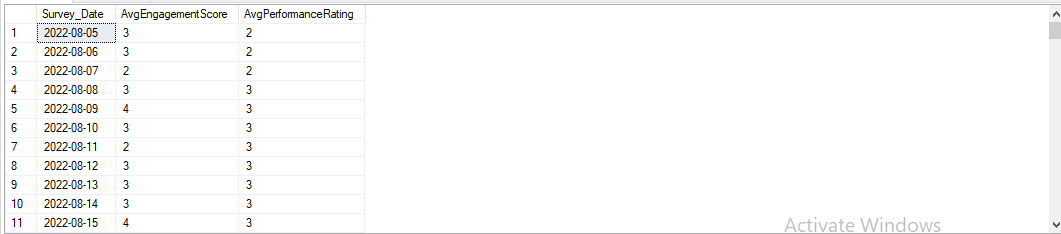


1. **EmployeeEngagementandPerformanceCorrelationOverTime** WITH EngagementPerformanceAS (SELECT e.Employee\_ID, e.Survey\_Date, e.Engagement\_Score, ed.Current\_Employee\_RatingFROM employee\_engagement\_survey\_data e

JOINemployee\_dataedONe.Employee\_ID=ed.Employee\_IDWHERE ed.EmployeeStatus= 'Active')

SELECTSurvey\_Date,AVG(Engagement\_Score)ASAvgEngagementScore, AVG(Current\_Employee\_Rating) AS AvgPerformanceRating

FROMEngagementPerformanceGROUP BY Survey\_DateORDER BY Survey\_Date



# DATAANALYSISSUMMARY

The provided SQL queries encompass a wide range of HR-related data analyses, sheddinglightonvariousaspectsofanorganization'sworkforce.Thesequeriesallow HR professionals and decision-makers to gain valuable insights into employee demographics, engagement, performance, recruitment, training, turnover, and more. Below is a summary of the key findings and insights derived from these queries:

# EmployeeDemographicsandDistribution

The organization's workforce is divided into different business units, and the employeecountvariesacrosstheseunits.Demographicssuchasgender,race,and marital status have been analyzed to provide a comprehensive view of the workforce's diversity.

# EmployeeEngagementandPerformance

Employee engagement scores have been tracked over time, showing trends in employeesatisfaction.Theanalysisofperformancescoresbygender,department, and job title allows for a deeper understanding of performance trends within the organization.

# RecruitmentandApplicantAnalysis

Recruitmentdataincludesinformationonapplicants'educationlevels,desired salaries, and statuses. The analysis provides insights into the educational qualifications and salary expectations of job applicants.

# TrainingandDevelopment

Trainingprogramshavebeencategorizedbytype,andthecostsassociated with eachprogramhavebeencalculated.Trainingdurationandcostsareanalyzedby program, location, and business unit.

# EmployeeTurnoverandExitPatterns

Employee turnover has been examined by gender, marital status, tenure, and department.Exitreasonshavebeencategorizedtoidentifycommonreasonsfor employee departures.

# PerformanceandPromotion

Employeeperformancescoreshavebeenanalyzed,andthenumberofpromotions for active, full-time employees has been determined. This analysis helps in understanding the relationship between performance and promotions.

# GeographicAnalysis

Employeecountsandrecruitmentdatahavebeenbrokendownbystate,city,and country. Training costs and employee tenure are also analyzed by location.

# Supervisorand Leadership

Theanalysisidentifiestop-performingsupervisorsbasedonemployeeratings, potentially highlighting effective leadership.

# CostAnalysis

Various cost-related analyses have been performed, including training costs by program,businessunit,andlocation.Recruitmentcostshavealsobeenanalyzed in terms of desired salary expectations.

# EngagementandPerformanceCorrelation

Acorrelationbetweenemployeeengagementscoresandperformanceratingsover time has been established.

In summary, these SQL queries enable HR professionals to make data-driven decisions, optimize workforce management, enhance employee engagement and performance,andaddressrecruitmentandtrainingneedseffectively.Theinsights gained from these analyses can inform HR strategies, improve organizational efficiency, and contribute to a more engaged and productive workforce.

# SUMMARYOFHRDECISION-MAKINGINSIGHTS

The extensive analysis of HR data through 50 SQLqueries has yielded valuable insightsthatcansignificantlyimpactHRdecision-makingprocesses.Theseinsights span various facets of human resources, offering actionable knowledge to enhance workforcemanagement,recruitmentstrategies,traininganddevelopmentprograms, employee engagement initiatives, and overall organizational performance. Here are the key decision-making insights derived from the analysis:

# WorkforceManagement

1. EmployeeDemographics:Understandingthedemographiccompositionofthe workforce is crucial for fostering diversity and inclusion efforts, addressing disparities, and tailoring HR policies to specific groups.
2. EmployeePerformanceDistribution:Identifyingperformancescore distributions allow HR to recognize top performers, address underperformance, and implement targeted improvement plans.
3. EmployeePromotionAnalysis:Recognizinghigh-performingemployeesfor potential promotions can boost morale and retain talent.
4. EmployeeExitPatternsbyTenure:Recognizingpatternsinemployeeexits concerning tenure can help in designing retention strategies and reducing turnover.
5. AverageTenurebyDepartment:Knowingtheaveragetenurebydepartment can inform succession planning and workforce stability.
6. Highest-PaidEmployees:Identifyinghighperformerswhoarealsowell- compensated can help in retaining and incentivizing top talent.
7. Employee Engagement and Performance Correlation Over Time: Understandingtherelationshipbetweenengagementandperformancescores over time can help HR tailor engagement strategies to boost performance.

# RecruitmentStrategies

1. ApplicantsbyEducationLevel:Insightintotheeducationalbackgroundof applicants assists in tailoring job requirements and recruitment strategies.
2. RecruitmentbyJobTitleandGender:Identifyinggenderdisparitiesinjob applicationshighlightspotential areas forgender diversity improvements.
3. RecruitmentbyCityandState:Understandinggeographicalpatternsin recruitment aids in targeting specific regions for talent acquisition.
4. RecruitmentCostAnalysis:Analyzingrecruitmentcostsanddesiredsalaries by education level can optimize recruitment budgets and compensation packages.

# TrainingandDevelopment

1. TrainingCostsbyProgram:Identifyingthecostofdifferenttrainingprograms enables efficient allocation of resources.
2. TrainingDurationbyType:Knowingthedurationofvarioustraining programs helps in scheduling and planning employee development.
3. TrainingCostsbyLocation:Understandingtrainingcostsbylocationcanaid in optimizing training budgets and selecting cost-effective venues.

# EmployeeEngagement

1. EmployeeEngagementScores:Monitoringengagementscoresovertime provides insights into employee satisfaction and morale.
2. Employees with High Engagement Scores: Recognizing highly engaged employeescanserveasanexampleofbestpracticesandinformengagement strategies for others.
3. EmployeeEngagementTrendsOverTime:Trackingengagementtrendshelps HR assess the effectiveness of engagement initiatives and adapt strategies as needed.

# Cross-FunctionalInsights

1. Employee Engagement and Performance Correlation Over Time: The correlationbetween engagementandperformance scoreshighlightsthe importanceofengagementinitiativesinenhancingproductivityandjob satisfaction.

These insights underscore the significance of data-driven decision-making in HR. By leveragingthefindingsfromtheseSQLqueries,HRprofessionalscanmakeinformed choices to optimize workforce management, improve recruitment processes, refine training programs, boost employee engagement, and ultimately contribute to the overall success of the organization. These insights also emphasize the value of regularly monitoring and analyzing HR data to adapt to changing workforce dynamics and continually improve HR strategies and practices.