library('ProjectTemplate); load.project()

\*1) Create 2 tables. The first table shows how many salespeople have attrition. Store this information in a column called “attrition”. The second table counts how many salespeople have no attrition. Store this information in a column called “Non Attrition”.

CREATE EXTERNAL TABLE employee

(

Age int,

Attrition string,

BusinessTravel string,

DailyRate int,

Department string,

DistanceFromHome int,

Education int,

EducationField string,

EmployeeCount int,

EmployeeNumber int,

EnvironmentSatisfaction int,

Gender string,

HourlyRate int,

JobInvolvement int,

JobLevel int,

JobRole string,

JobSatisfaction int,

MaritalStatus string,

MonthlyIncome int,

MonthlyRate int,

NumCompaniesWorked int,

Over18 string,

OverTime string,

PercentSalaryHike int,

PerformanceRating int,

RelationshipSatisfaction int,

StandardHours int,

StockOptionLevel int,

TotalWorkingYears int,

TrainingTimesLastYear int,

WorkLifeBalance int,

YearsAtCompany int,

YearsInCurrentRole int,

YearsSinceLastPromotion int,

YearsWithCurrManager int

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED AS TEXTFILE LOCATION '/employee'

TBLPROPERTIES("skip.header.line.count"="1");

**\*ATTRITION , YES.** #Create a new table

CREATE TABLE Attrition

(

Attrition int

)

\*a) Count “attrition”

INSERT OVERWRITE TABLE attrition

SELECT COUNT(Attrition) AS Attrition

FROM employee

WHERE Attrition = "Yes" AND JobRole LIKE "%Sales%";

\*b) NON ATTRITION

CREATE EXTERNAL TABLE Non Attrition

(

Non Attrition string,

)

\*c) Count “Non Attrition”

INSERT OVERWRITE TABLE Non\_Attrition

SELECT COUNT(Attrition) AS Non\_Attrition

FROM employee

WHERE Attrition = "No" AND JobRole LIKE "%Sales%";

\*2) Create 3 statistics tables (average, min, max) showing the monthly income for: all salespeople, those with attrition and those without attrition. For each table, label the columns “average\_monthly\_income”, “min\_monthly\_income”, and “max\_monthly\_income”.

\*a) AVERAGE

CREATE TABLE Average

(

Attrition string,

average\_monthly\_income float

)

INSERT OVERWRITE TABLE Average

SELECT Attrition, AVG(MonthlyIncome) AS average\_monthly\_income

FROM employee

WHERE JobRole LIKE "%Sales%"

GROUP BY Attrition;

\*b) MIN

CREATE TABLE Min

(

Attrition string,

average\_monthly\_income float

)

INSERT OVERWRITE TABLE Min

SELECT Attrition, MIN(MonthlyIncome) AS min\_monthly\_income

FROM employee

WHERE JobRole LIKE "%Sales%"

GROUP BY Attrition;

\*c) MAX

CREATE TABLE Max

(

Attrition string,

average\_monthly\_income float

)

INSERT OVERWRITE TABLE Max

SELECT Attrition, MAX(MonthlyIncome) AS max\_monthly\_income

FROM employee

WHERE JobRole LIKE "%Sales%"

GROUP BY Attrition;

\*3) Create 2 tables containing the monthly income and count of salespeople making that monthly income for: salespeople with attrition and salespeople without attrition. Label the columns as: “monthlyincome” and “count”.

\*a) ATTRITION

CREATE TABLE monthly\_income\_attrition

(

MonthlyIncome int,

Count int

)

INSERT OVERWRITE TABLE monthly\_income\_attrition

SELECT MonthlyIncome, COUNT(MonthlyIncome) AS count

FROM employee

WHERE Attrition = "Yes" AND JobRole LIKE "%Sales%"

GROUP BY MonthlyIncome;

\*b) NON ATTRITION

CREATE TABLE monthly\_income\_nonattrition

(

MonthlyIncome int,

Count int

)

INSERT OVERWRITE TABLE monthly\_income\_nonattrition

SELECT MonthlyIncome, COUNT(MonthlyIncome) AS count

FROM employee

WHERE Attrition = "No" AND JobRole LIKE "%Sales%"

GROUP BY MonthlyIncome;