CREATE EXTERNAL TABLE employee\_sales

(

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,

MartialStatus 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");

SELECT \* FROM employee\_sales LIMIT 10;

**How many salespeople have attrition**

CREATE TABLE employees

(

Salespeople string

);

INSERT OVERWRITE TABLE employees

SELECT Attrition

FROM employee\_sales;

INSERT OVERWRITE TABLE employees

SELECT \*

FROM employees

WHERE Salespeople LIKE "Yes";

SELECT "Yes", COUNT (salespeople) AS attrition

FROM employees

GROUP BY salespeople;

**How many salespeople have no attrition**

INSERT OVERWRITE TABLE employees

SELECT Attrition

FROM employee\_sales;

(did this one again to reset the data in my employees table)

INSERT OVERWRITE TABLE employees

SELECT \*

FROM employees

WHERE Salespeople LIKE "No";

SELECT "No", COUNT (salespeople) AS Non\_attrition

FROM employees

GROUP BY salespeople;

**Average, Min, Max**

CREATE TABLE employee\_income

(

Attrition string,

MonthlyIncome int

);

INSERT OVERWRITE TABLE employee\_income

SELECT Attrition, MonthlyIncome

FROM employee\_sales;

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

FROM employee\_income

GROUP BY Attrition;

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

FROM employee\_income

GROUP BY Attrition;

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

FROM employee\_income

GROUP BY Attrition;

**Monthly income for salespeople with attrition**

INSERT OVERWRITE TABLE employee\_income

SELECT \*

FROM employee\_income

WHERE Attrition LIKE "Yes";

SELECT monthlyincome, COUNT (MonthlyIncome) AS amountofincome

FROM employee\_income

GROUP BY monthlyincome;

**Monthly income for salespeople without attrition**

INSERT OVERWRITE TABLE employee\_income

SELECT Attrition, MonthlyIncome

FROM employee\_sales;

(did this one again to reset the data in my employee\_income table)

INSERT OVERWRITE TABLE employee\_income

SELECT \*

FROM employee\_income

WHERE Attrition LIKE "No";

(then ran the same query again to get the count of No’s)

SELECT monthlyincome, COUNT (MonthlyIncome) AS amountofincome

FROM employee\_income

GROUP BY monthlyincome;