**Data Aggregation Activity**

**Table of Contents**

[Section 1: Queries Used 2](#_Toc44969145)

[***1.1 Loading the “employee” Table 2***](#_Toc44969146)

[***1.2 Data Transformation Steps 3***](#_Toc44969147)

[***1.3 Data Aggregation Steps 3***](#_Toc44969148)

[Count of Employees with Attrition 3](#_Toc44969149)

[Count of Employees without Attrition 4](#_Toc44969150)

[Income Statistics for All Employees 4](#_Toc44969151)

[Income Statistics for Employees with Attrition 4](#_Toc44969152)

[Income Statistics for Employees without Attrition 5](#_Toc44969153)

[Count by Monthly Income for Salespeople with Attrition 5](#_Toc44969154)

[Count by Monthly Income for Salespeople without Attrition 6](#_Toc44969155)

[Section 2: Screenshots 7](#_Toc44969156)

[***2.1 Verifying Data Transformation Steps 7***](#_Toc44969157)

[***2.2 Data Aggregation Steps 8***](#_Toc44969158)

[Count of Employees with Attrition 8](#_Toc44969159)

[Count of Employees without Attrition 9](#_Toc44969160)

[Income Statistics for All Employees 10](#_Toc44969161)

[Income Statistics for Employees with Attrition 11](#_Toc44969162)

[Income Statistics for Employees without Attrition 12](#_Toc44969163)

[Count by Monthly Income for Salespeople with Attrition 13](#_Toc44969164)

[Count by Monthly Income for Salespeople without Attrition 14](#_Toc44969165)

# Section 1: Queries Used

## 1.1 Loading the “employee” Table

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");

## 1.2 Data Transformation Steps

CREATE EXTERNAL TABLE employee\_sales

(

Attrition string,

Department string,

JobSatisfaction int,

MonthlyIncome int

);

INSERT OVERWRITE TABLE employee\_sales

SELECT Attrition, Department, JobSatisfaction, MonthlyIncome

FROM employee;

INSERT OVERWRITE TABLE employee\_sales

SELECT Attrition, Department, JobSatisfaction, ROUND(MonthlyIncome, -3) as MonthlyIncome

FROM employee\_sales;

INSERT OVERWRITE TABLE employee\_sales

SELECT \*

FROM employee\_sales

WHERE Department LIKE "%Sales%";

INSERT OVERWRITE TABLE employee\_sales

SELECT \*

FROM employee\_sales

ORDER BY JobSatisfaction DESC;

SELECT \* FROM employee\_sales LIMIT 20;

## 1.3 Data Aggregation Steps

### Count of Employees with Attrition

CREATE EXTERNAL TABLE employee\_attrition

(

attrition int

);

INSERT OVERWRITE TABLE employee\_attrition

SELECT COUNT(Attrition) AS attrition

FROM employee\_sales

WHERE Attrition LIKE "%Yes%";

SELECT \* FROM employee\_attrition;

### Count of Employees without Attrition

CREATE EXTERNAL TABLE employee\_non\_attrition

(

non\_attrition int

);

INSERT OVERWRITE TABLE employee\_non\_attrition

SELECT COUNT(Attrition) AS non\_attrition

FROM employee\_sales

WHERE Attrition LIKE "%No%";

SELECT \* FROM employee\_non\_attrition;

### Income Statistics for All Employees

CREATE EXTERNAL TABLE all\_employee\_income\_stats

(

average\_monthly\_income double,

min\_monthly\_income int,

max\_monthly\_income int

);

INSERT OVERWRITE TABLE all\_employee\_income\_stats

SELECT AVG(MonthlyIncome) AS average\_monthly\_income, MIN(MonthlyIncome) AS min\_monthly\_income, MAX(MonthlyIncome) AS max\_monthly\_income

FROM employee\_sales;

SELECT \* FROM all\_employee\_income\_stats;

### Income Statistics for Employees with Attrition

CREATE EXTERNAL TABLE attrition\_employee\_income\_stats

(

average\_monthly\_income double,

min\_monthly\_income int,

max\_monthly\_income int

);

INSERT OVERWRITE TABLE attrition\_employee\_income\_stats

SELECT AVG(MonthlyIncome) AS average\_monthly\_income, MIN(MonthlyIncome) AS min\_monthly\_income, MAX(MonthlyIncome) AS max\_monthly\_income

FROM employee\_sales

WHERE Attrition LIKE "%Yes%";

SELECT \* FROM attrition\_employee\_income\_stats;

### Income Statistics for Employees without Attrition

CREATE EXTERNAL TABLE non\_attrition\_employee\_income\_stats

(

average\_monthly\_income double,

min\_monthly\_income int,

max\_monthly\_income int

);

INSERT OVERWRITE TABLE non\_attrition\_employee\_income\_stats

SELECT AVG(MonthlyIncome) AS average\_monthly\_income, MIN(MonthlyIncome) AS min\_monthly\_income, MAX(MonthlyIncome) AS max\_monthly\_income

FROM employee\_sales

WHERE Attrition LIKE "%No%";

SELECT \* FROM non\_attrition\_employee\_income\_stats;

### Count by Monthly Income for Salespeople with Attrition

CREATE EXTERNAL TABLE attrition\_income\_count

(

monthlyincome int,

count int

);

INSERT OVERWRITE TABLE attrition\_income\_count

SELECT MonthlyIncome AS monthlyincome, COUNT(MonthlyIncome) AS count

FROM employee\_sales

WHERE Attrition LIKE "%Yes%"

GROUP BY monthlyincome;

SELECT \* FROM attrition\_income\_count;

### Count by Monthly Income for Salespeople without Attrition

CREATE EXTERNAL TABLE non\_attrition\_income\_count

(

monthlyincome int,

count int

);

INSERT OVERWRITE TABLE non\_attrition\_income\_count

SELECT MonthlyIncome AS monthlyincome, COUNT(MonthlyIncome) AS count

FROM employee\_sales

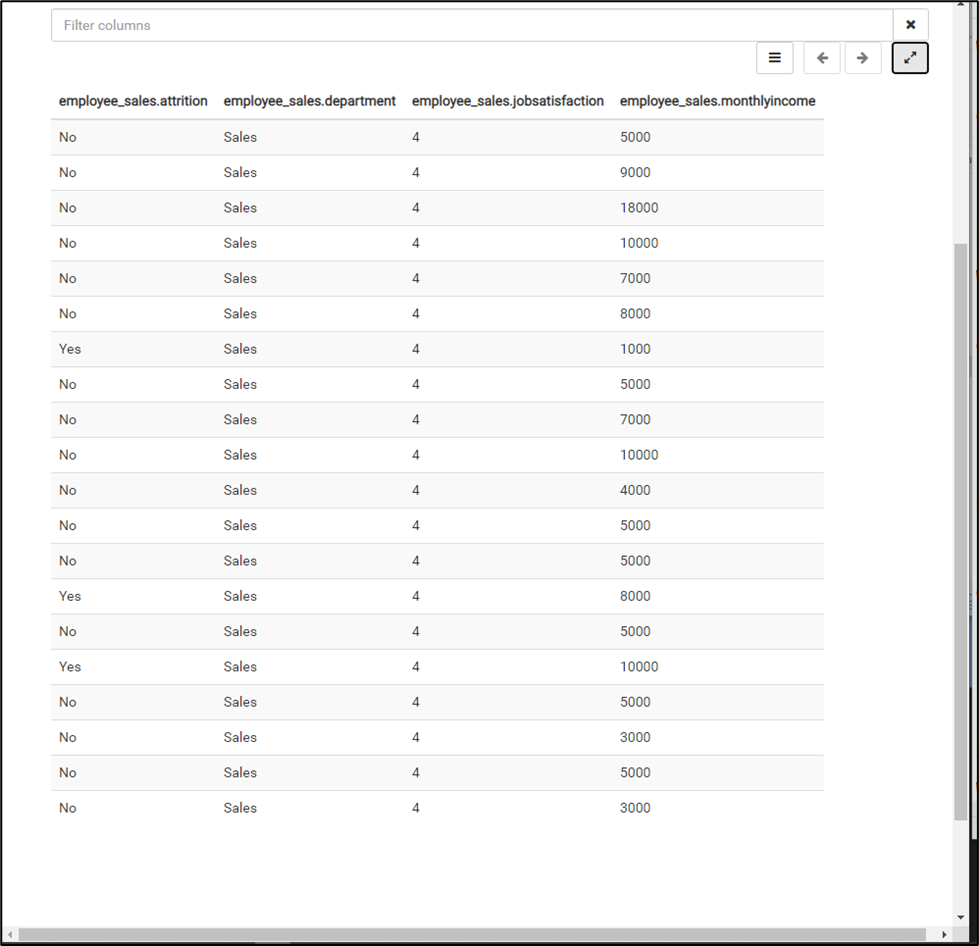
WHERE Attrition LIKE "%No%"

GROUP BY monthlyincome;

SELECT \* FROM non\_attrition\_income\_count;

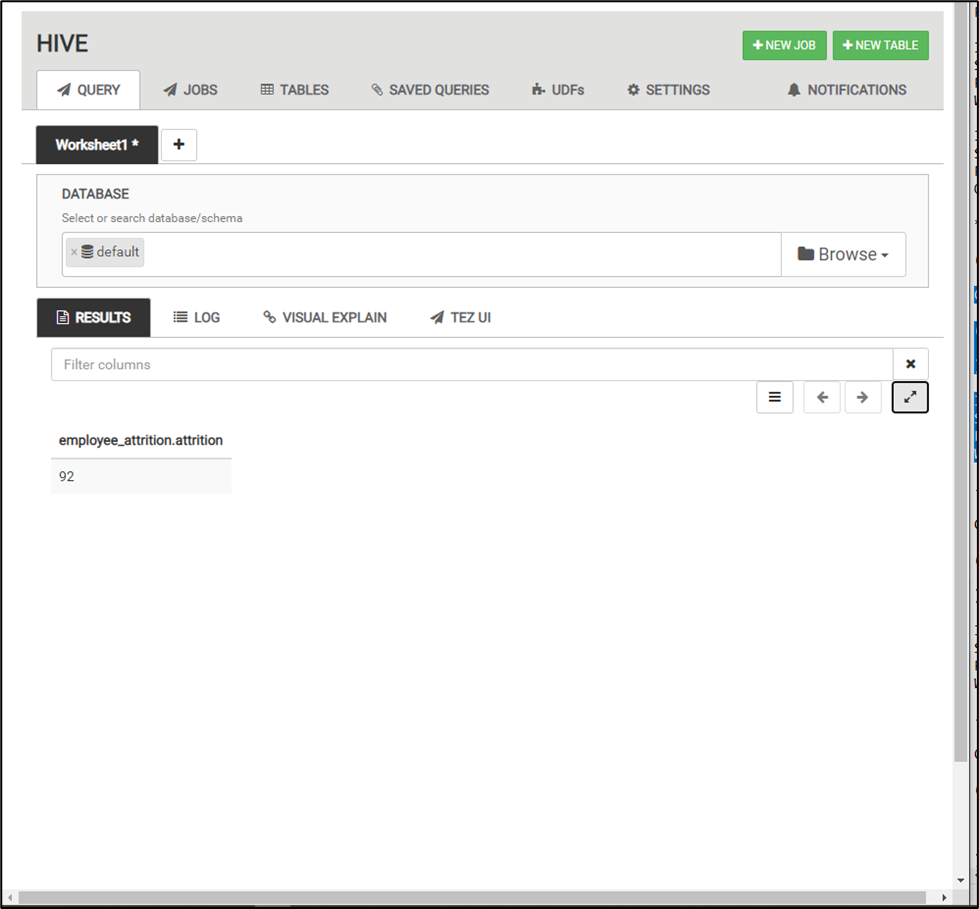
# Section 2: Screenshots

## 2.1 Verifying Data Transformation Steps

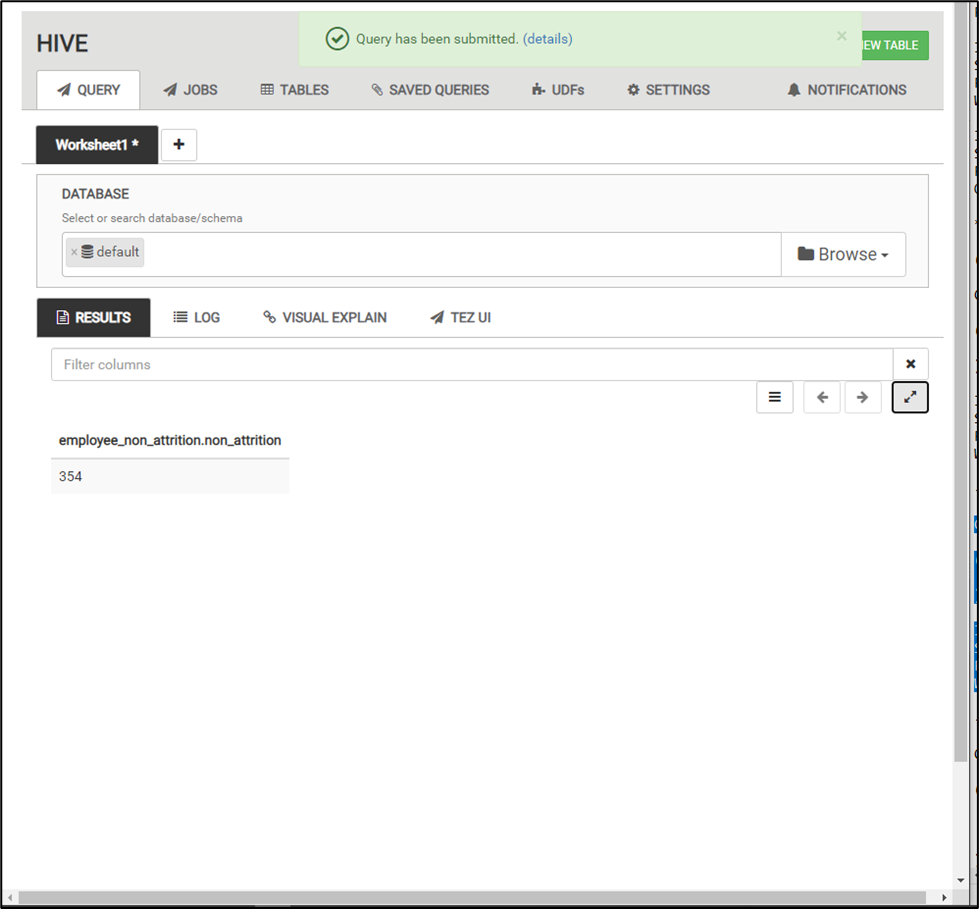


## 2.2 Data Aggregation Steps

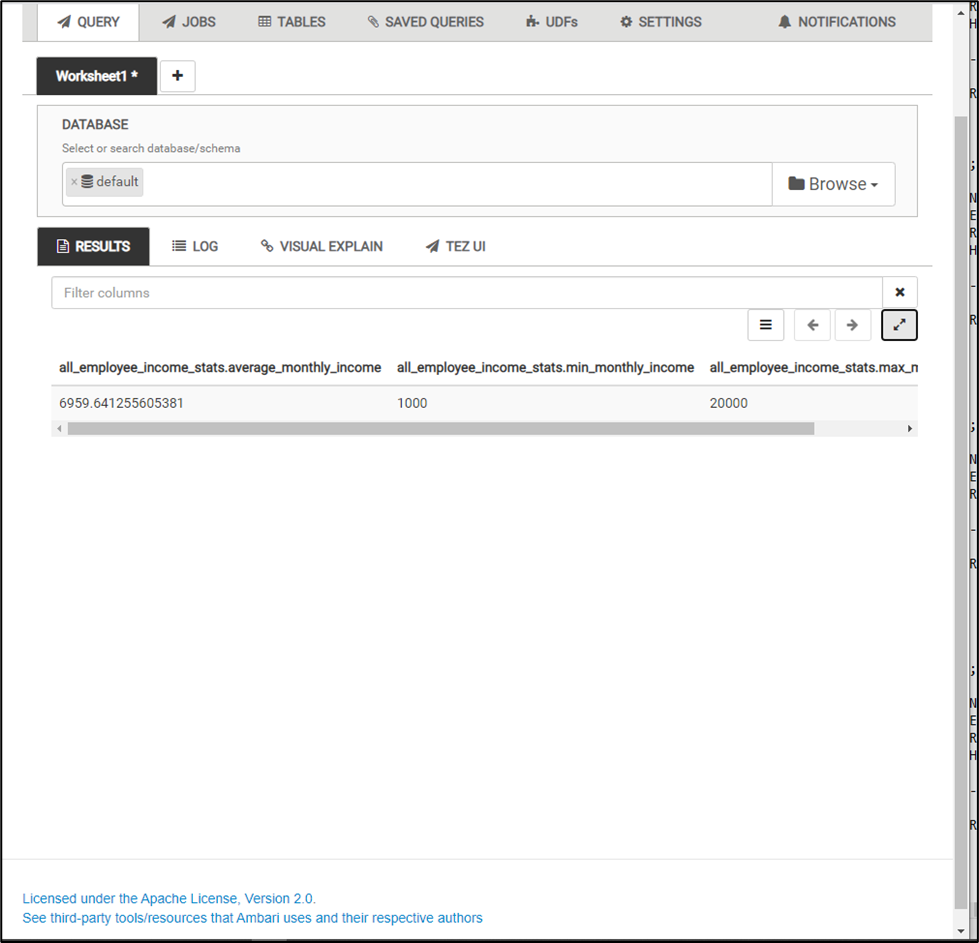
### Count of Employees with Attrition



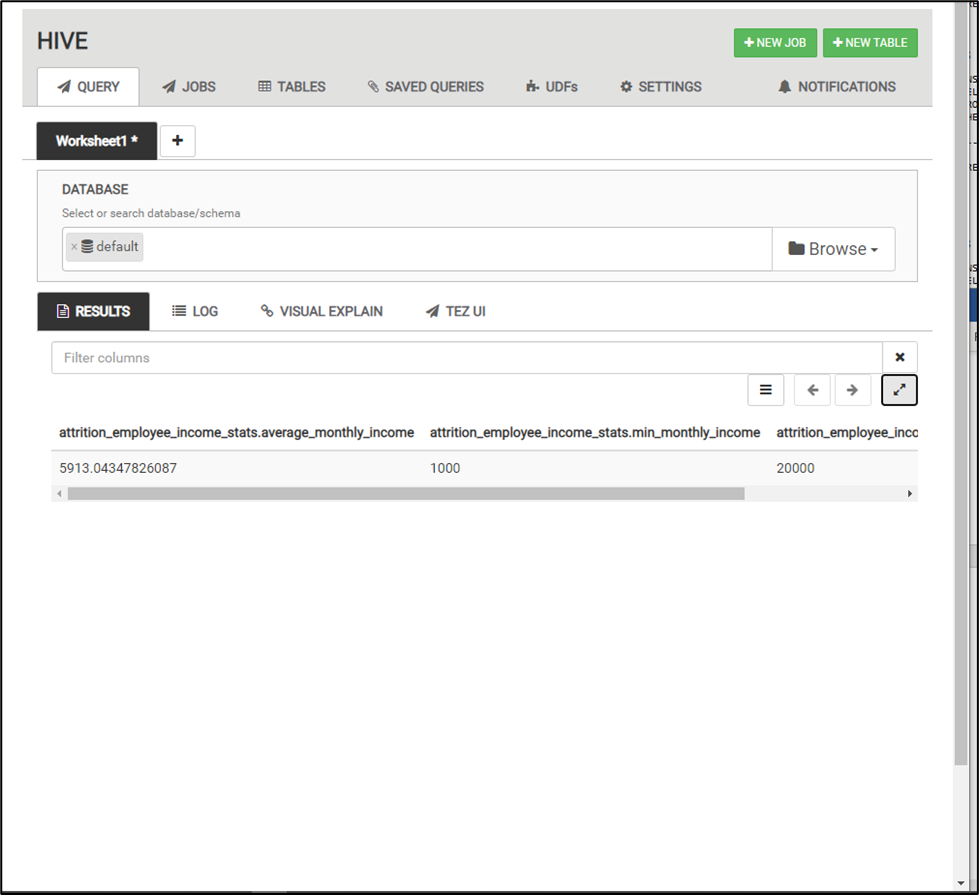
### Count of Employees without Attrition



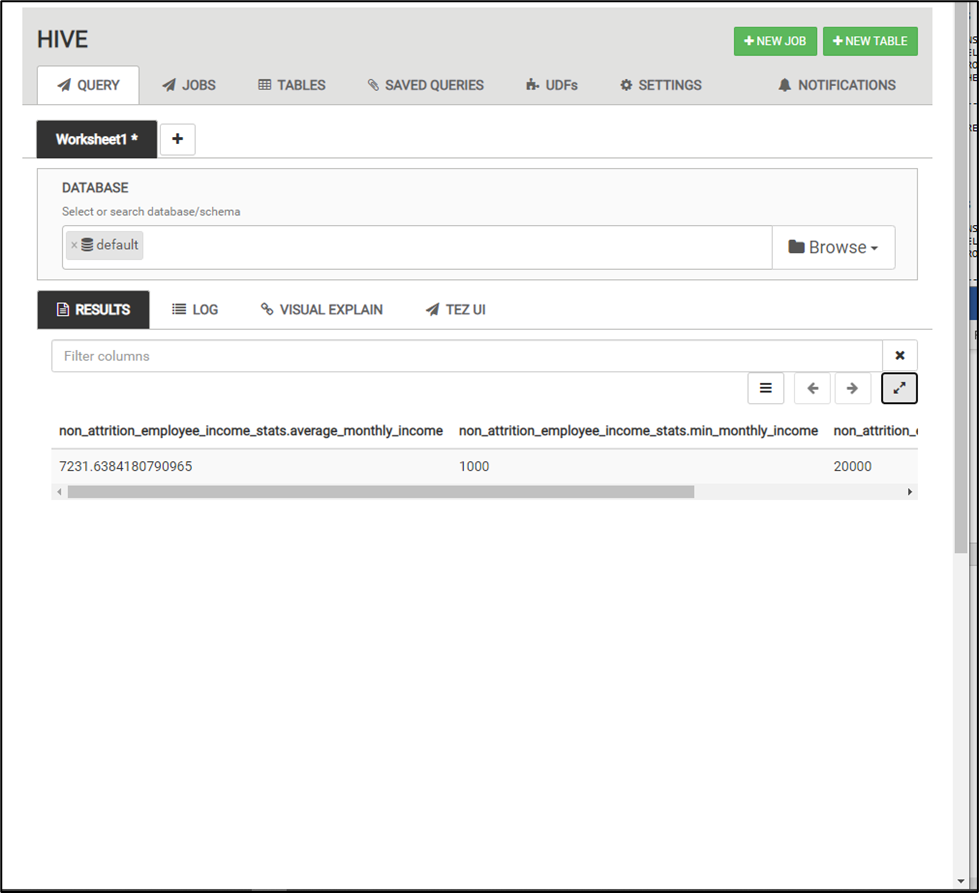
### Income Statistics for All Employees



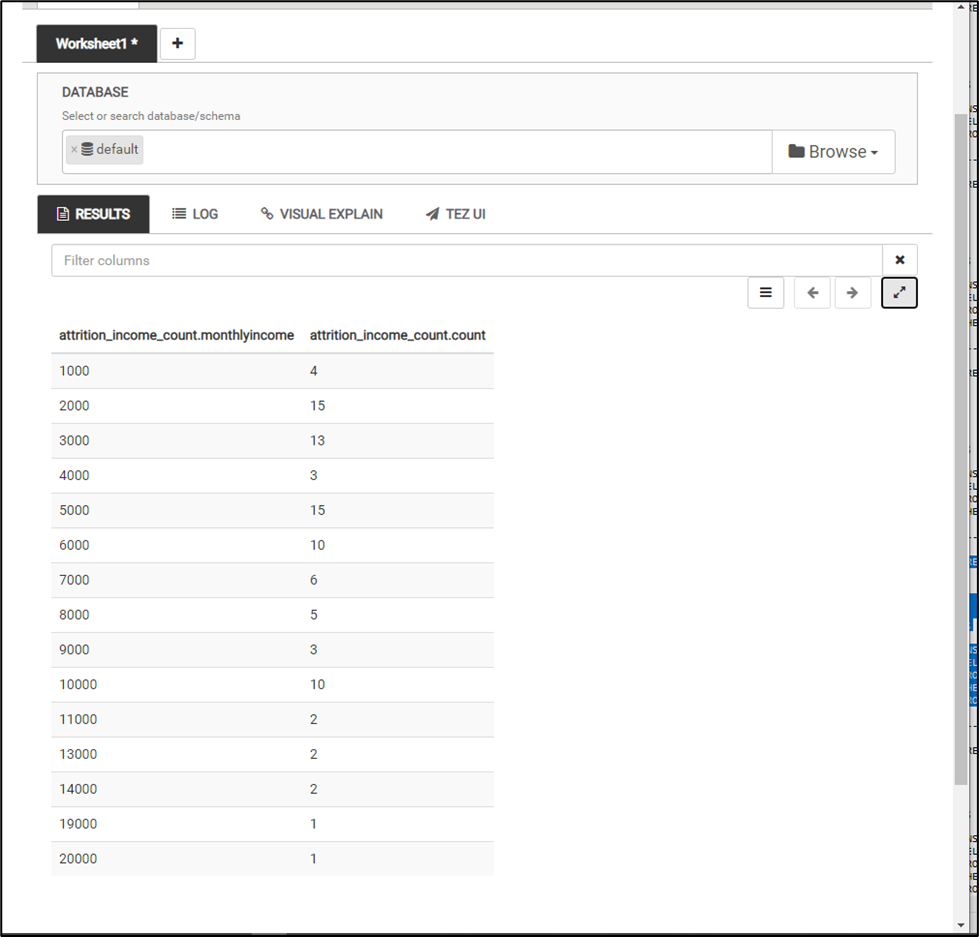
### Income Statistics for Employees with Attrition



### Income Statistics for Employees without Attrition



### Count by Monthly Income for Salespeople with Attrition



### Count by Monthly Income for Salespeople without Attrition

