databricks

Employee Attrition (SQL)

Import Notebook

```
SELECT * FROM emp_table;
```

Table

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfa
1	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2
2	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	3
3	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4
4	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	4
5	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	1
6	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	8	4
7	59	No	Travel_Rarely	1324	Research & Development	3	3	Medical	1	10	3

1,470 rows

```
-- To see the total number of employees
select sum(employeeCount) from emp_table
```

Table

	sum(employeeCount)
1	1470

1 row

```
-- to see the division of YES and NO amongst the employees

select sum(employeeCount),attrition from emp_table
group by 2
```

Table

	sum(employeeCount)	attrition
1	1233	No
2	237	Yes

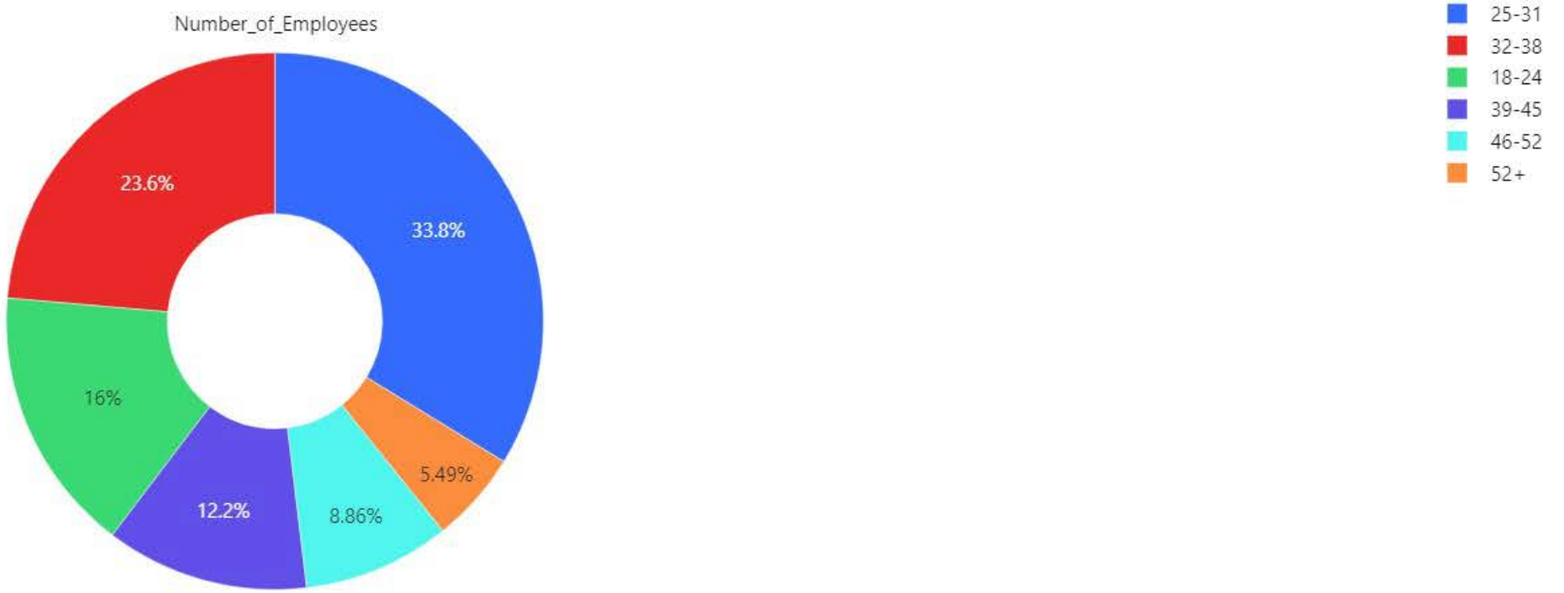
2 rows

```
-- AGE analysis (to see the distribution between the age groups - 18-24,25-31,32-28,39-45,46-52,52+)

select sum(employeeCount) AS Number_of_Employees,
CASE when age between 18 and 24 then '18-24'
when age between 25 and 31 then '25-31'
when age between 32 and 38 then '32-38'
when age between 39 and 45 then '39-45'
when age between 46 and 52 then '46-52'
else '52+' END AS Age_Group from emp_table
where attrition='Yes'
group by 2
order by 1 desc
```

Table Visualization 1

Table | Visualization 1



6 rows

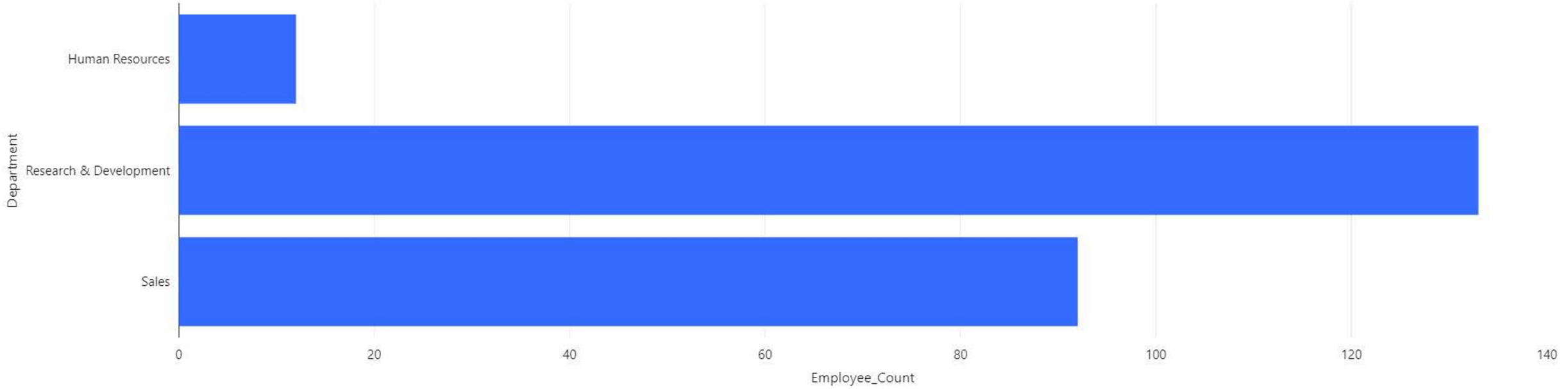
```
-- Department analysis

Select sum(EmployeeCount) As Employee_Count, Department
from emp_table
where attrition='Yes'
Group by 2
```

```sql
-- Department analysis

Select sum(EmployeeCount) As Employee_Count, Department
from emp_table
where attrition='Yes'
Group by 2
```

Table | Visualization 1



3 rows

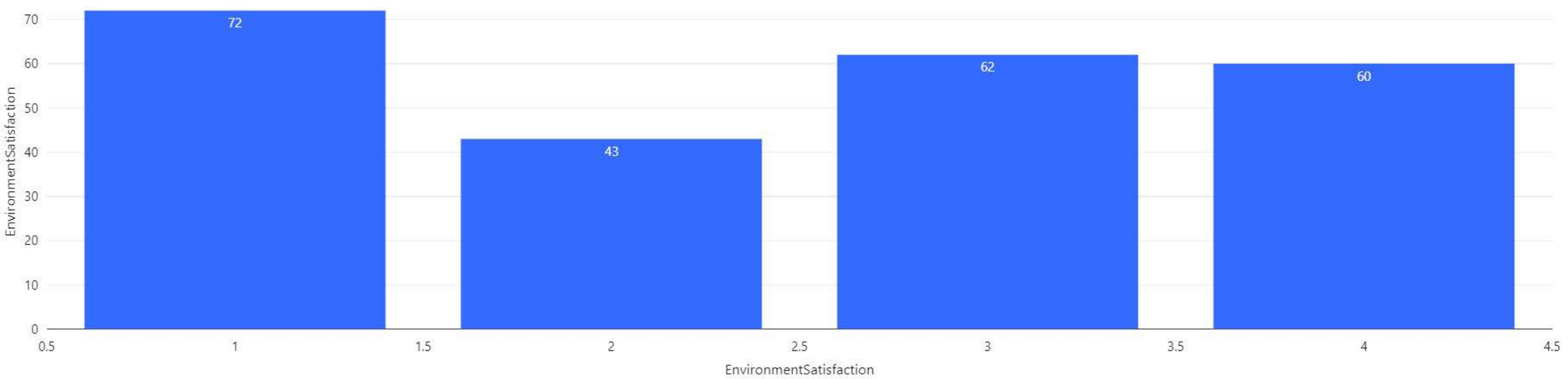
```
-- Education Analysis (1-School, 2-College, 3-Bachelors, 4-Masters, 5-PhD)

Select sum(EmployeeCount) As Employee_Count,
CASE WHEN education=1 then 'School Graduate'
WHEN education=2 then 'College Graduate'
WHEN education=3 then 'Bachelors Degree'
WHEN education=4 then 'Masters Degree'
ELSE 'Doctoral Degree' END AS Degree_Name
from emp_table
where attrition='Yes'
Group by 2
```

Table | Visualization 1

```sql
-- Environment Satisfaction Analysis
Select sum(EmployeeCount) As Employee_Count, EnvironmentSatisfaction
from emp_table
where attrition='Yes'
Group by 2
```

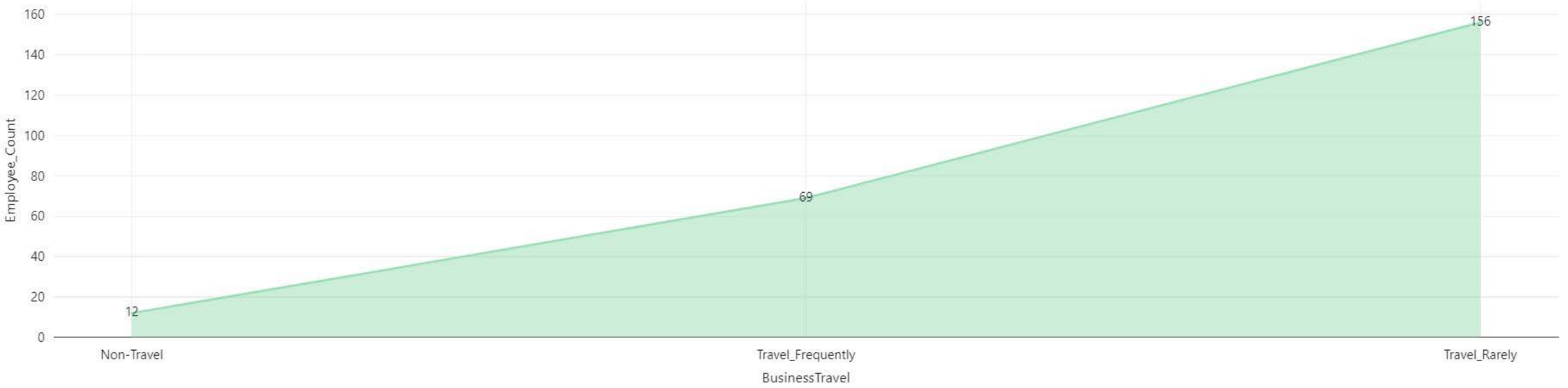
Table | Visualization 1



4 rows

```
-- Business Travel dependency Analysis on attrition
Select sum(EmployeeCount) As Employee_Count, BusinessTravel
from emp_table
where attrition='Yes'
Group by 2
```

Table | Visualization 1



3 rows

BusinessTravel
3 rows

INSIGHTS DERIVED FROM THE DATASET USING SPARK SQL

1. Employees falling in the age group of '25-31' tend to leave the company the most.
2. Employees under the 'Research and Developement' department are more likely to leave the department. This can mean there is some issues internally within that department.
3. 41.77% of the employees who have left have a Bachelors degree.
4. 72 employees were not satisfied with the working condition environment.
5. The employees who travelled rarely because of business reasons tend to leave the company more often.