databricks Employee_Attrition

(https://databricks.com)

Running Spark in SQL.

```
SELECT * FROM employee
```

Table

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

1,470 rows

```
select sum(EmployeeCount) from employee
```

Table

	sum(EmployeeCount)
1	1470

1 row

Find out attrition division:

```
select sum(EmployeeCount),attrition from employee
group by 2
```

Table

	sum(EmployeeCount)	attrition
1	1233	No
2	237	Yes

2 rows

Age analysis- Checking which age group has higher attrition.

```
select sum(EmployeeCount),Age from employee
where Attrition='Yes'
group by Age
```

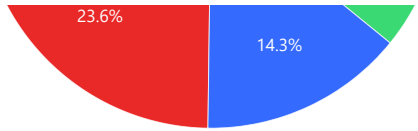
Table

	sum(EmployeeCount)	Age
1	18	31
2	2	53
3	9	34
4	14	28
5	12	26
6	3	27
7	6	44

39 rows

```
select sum(EmployeeCount),
case when age between 20 and 25 then '20-25' when age between 26 and 32 then '26-32' when age between 33 and 40 then '33-40' else '40+' end age_group
from employee
where Attrition='Yes'
group by 2
```

Table **Visualization 1**

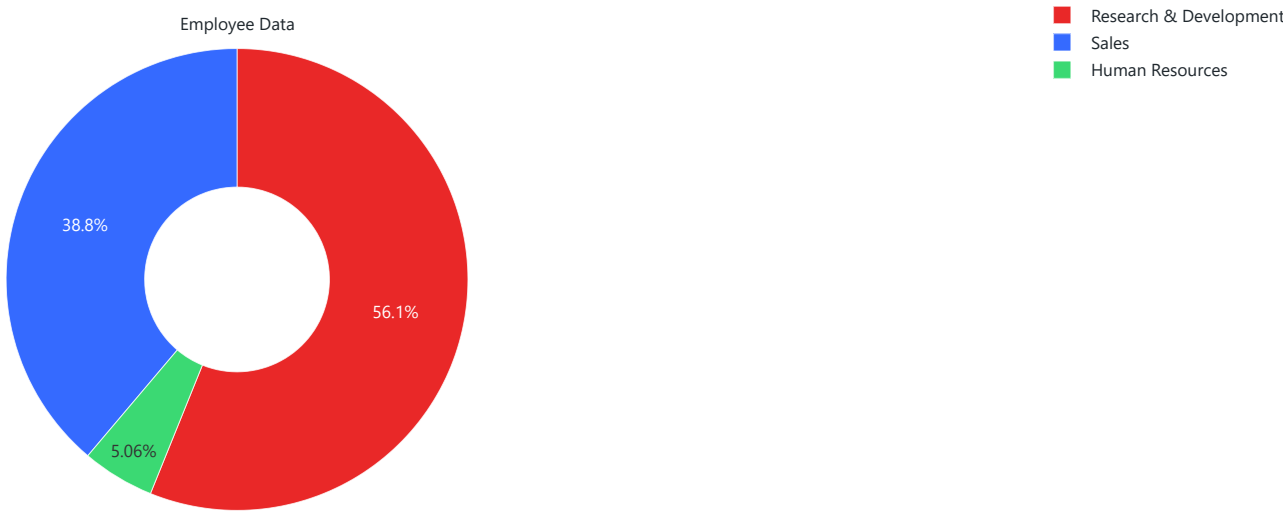


4 rows

attrition based on departments:

```
SELECT sum(EmployeeCount),department from employee
where attrition='Yes'
GROUP BY department
```

Table **Visualization 1**

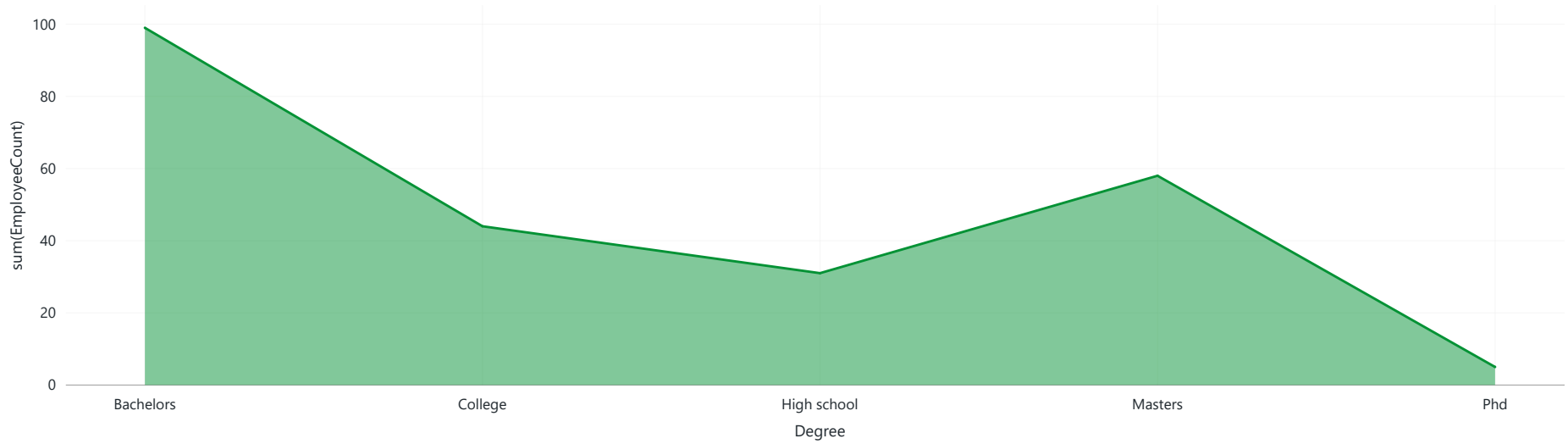


3 rows

Attrition based on Education: 1-High School, 2-College, 3-Bachelor, 4-Masters, 5-Phd.

```
SELECT sum(EmployeeCount),
case when Education=1 then 'High school' when Education=2 then 'College' when Education='3' then 'Bachelors' when Education=4 then 'Masters' else 'Phd' end Degree
from employee
where attrition='Yes'
GROUP BY 2
```

Table **Visualization**

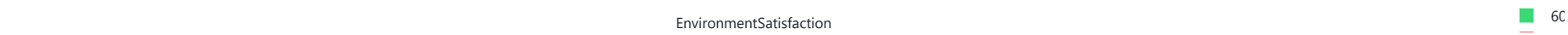


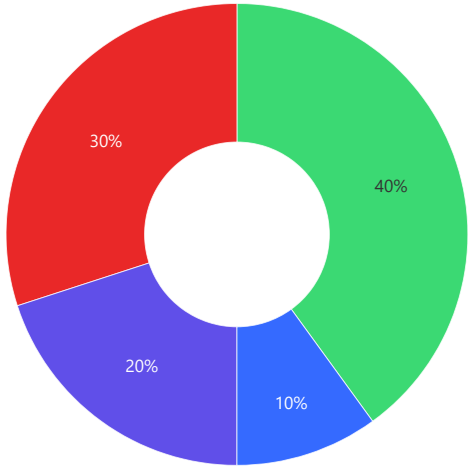
5 rows

Attrition based on team enviroment: 1-low, 2-medium, 3-satisfied, 4-Very satisfied:

```
SELECT sum(EmployeeCount),
EnvironmentSatisfaction
from employee
where attrition='Yes'
GROUP BY 2
```

Table **Visualization 1**



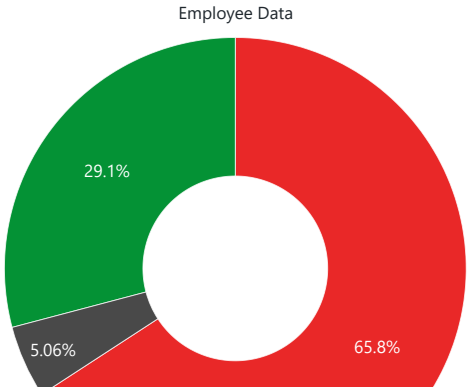


4 rows

Attrition based on travelling:

```
SELECT sum(EmployeeCount),
BusinessTravel
from employee
where attrition='Yes'
GROUP BY 2
```

Table **Visualization 1**



3 rows

Job Role with the Highest Attrition Rate:

```
SELECT JobRole, (COUNT(*) / SUM(COUNT(*)) OVER ()) AS AttritionRate
FROM employee
WHERE Attrition = 'Yes'
GROUP BY JobRole
ORDER BY AttritionRate DESC
LIMIT 1
```

Table

	JobRole	AttritionRate
1	Laboratory Technician	0.2616033755274262

1 row

Attrition based on Marital Status:

```
SELECT MaritalStatus, Attrition, COUNT(*) AS Count
FROM employee
GROUP BY MaritalStatus, Attrition
```

Table

	MaritalStatus	Attrition	Count
1	Divorced	Yes	33
2	Married	Yes	84
3	Single	Yes	120
4	Divorced	No	294
5	Single	No	350
6	Married	No	589

6 rows

Average Years at Company for Attrited Employees:

```
SELECT AVG(YearsAtCompany) AS AvgYearsAtCompany
FROM employee
WHERE Attrition = 'Yes'
```

Table

	AvgYearsAtCompany
1	5.1308016877637135

1 row

Average Monthly Income by Gender:

```
SELECT Gender, AVG(MonthlyIncome) AS AvgMonthlyIncome
FROM employee
GROUP BY Gender
```

Table

	Gender	AvgMonthlyIncome
1	Female	6686.566326530612
2	Male	6380.507936507936

2 rows

Distribution of employee tenure among attrited and non-attrited employees:

```
SELECT
  Attrition,
  AVG(YearsAtCompany) AS AvgYearsAtCompany,
  MEDIAN(YearsAtCompany) AS MedianYearsAtCompany,
  MIN(YearsAtCompany) AS MinYearsAtCompany,
  MAX(YearsAtCompany) AS MaxYearsAtCompany
FROM employee
GROUP BY Attrition
```

Table

	Attrition	AvgYearsAtCompany	MedianYearsAtCompany	MinYearsAtCompany	MaxYearsAtCompany
1	No	7.369018653690187	6	0	37
2	Yes	5.1308016877637135	3	0	40

2 rows

Attrition by Work-Life Balance:

Table

	WorkLifeBalance	Attrition	Count
1	3	Yes	127
2	2	No	286
3	2	Yes	58
4	1	No	55
5	3	No	766
6	4	No	126
7	4	Yes	27

8 rows