databricks **Attrition_of_Employee**

(https://databricks.com)

```
--  EMPLOYEE ATTRITION DATASET - PROJECT
```

```
select * from employee_attrition;
```

Table

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Edu
1	41	Yes	Travel_Rarely	1102	Sales	1	2	Life
2	49	No	Travel_Frequently	279	Research & Development	8	1	Life
3	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Oth
4	33	No	Travel_Frequently	1392	Research & Development	3	4	Life
5	27	No	Travel_Rarely	591	Research & Development	2	1	Med
6	32	No	Travel_Frequently	1005	Research & Development	2	2	Life
7	59	No	Travel_Rarely	1324	Research & Development	3	3	Med

1,470 rows

EMPLOYEE COUNT

```
select count(*) as total_number_of_employee from employee_attrition
```

Table

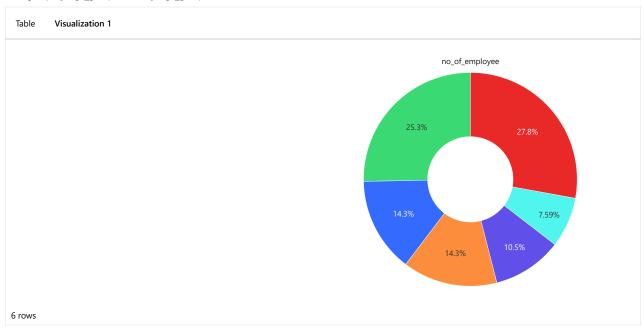|	total_number_of_employee
1	1470

1 row

Find out Attrition Division

```
select count(*) as no_of_employee , Attrition from employee_attrition group by Attrition
```

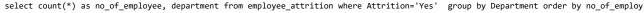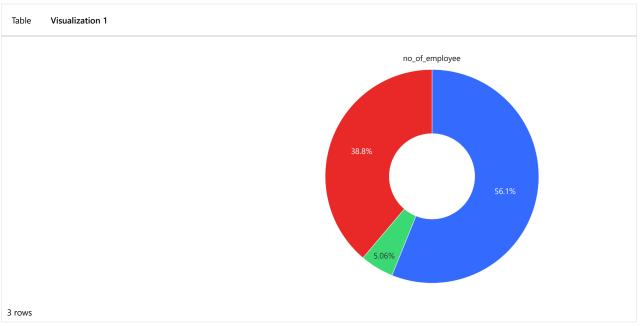| Table	**Visualization 1**

2 rows

AGE Analysis -- lets find out which particular age attrition is high (18-22,23-27,2

```
select
sum(EmployeeCount) as no_of_employee ,
case
  when age between 20 and 25 then '20-25'
  when age <=30 then '26-30'
  when age <=35 then '31-35'
  when age <=40 then '36-40'
  when age <=45 then '40-45'
else '46+' end age_group
 from employee_attrition where Attrition='Yes'
  group by age_group order by age_group
```

| Table | **Visualization 1** |

no_of_employee

6 rows

Attrition by Department

```
select count(*) as no_of_employee, department from employee_attrition where Attrition='Yes'  group by Department order by no_of_employ
```

| Table | **Visualization 1** |

no_of_employee

3 rows

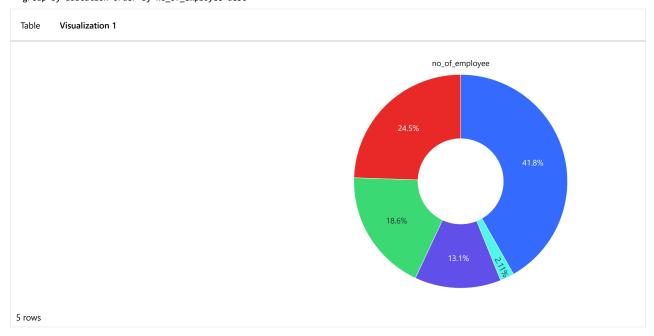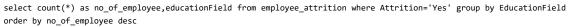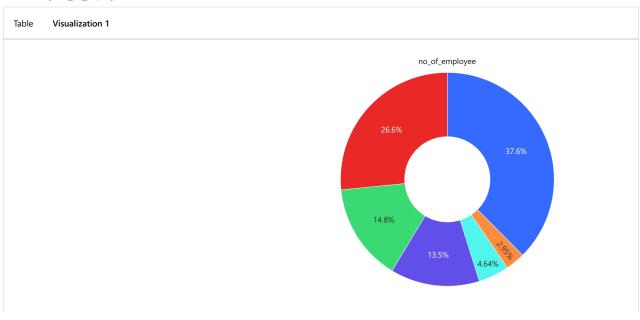# Attrition by education

```
-- 1-below college, 2-college,3-bachelor, 4-master, 5-doctor
select count(*)as no_of_employee,
case
  when Education=1 then 'Below College'
  when Education=2 then 'College'
  when Education=3 then 'Bachelor'
  when Education=4 then 'Master'
  else "Doctor"
  end Education
from employee_attrition where Attrition='Yes'
group by Education order by no_of_employee desc
```

| Table | **Visualization 1** |

no_of_employee

5 rows

Attrition by Eduction Field

```
select count(*) as no_of_employee,educationField from employee_attrition where Attrition='Yes' group by EducationField
order by no_of_employee desc
```

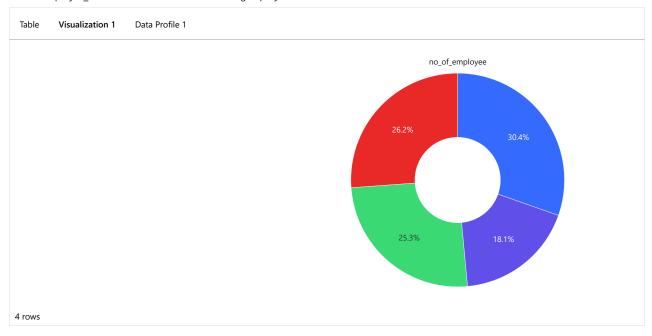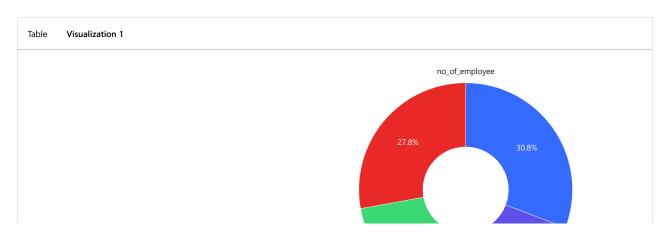| Table | **Visualization 1** |

no_of_employee

6 rows

Attrition by Environment Satisfaction

```
-- 1- low, 2-Medium, 3- High, 4 Highly Satisfied
select count(*)as no_of_employee,
case
  when EnvironmentSatisfaction =1 then 'Low'
  when EnvironmentSatisfaction =2 then 'Medium'
  when EnvironmentSatisfaction =3 then 'High'
  else 'Highly Satisfied'
end EnvironmentSatisfaction
from employee_attrition where Attrition='Yes' group by EnvironmentSatisfaction
```

| Table | **Visualization 1** | Data Profile 1 |

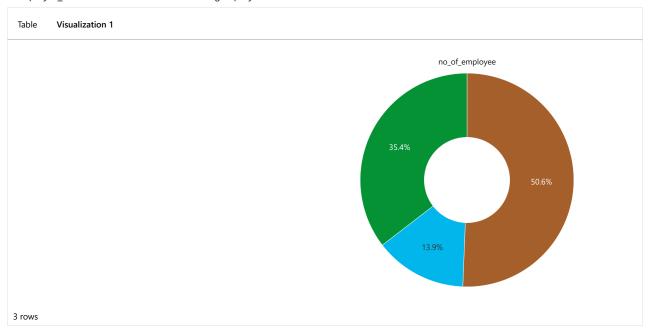4 rows

JobSatisfaction

```
select count(*) as no_of_employee,
case
  when JobSatisfaction =1 then 'Highly Satisfied'
  when JobSatisfaction =2 then 'High'
  when JobSatisfaction =3 then 'Medium'
  else 'Low'
end JobSatisfaction
from employee_attrition where Attrition='Yes'
 group by JobSatisfaction order by no_of_employee desc
```

| Table | **Visualization 1** |

21.9%   19.4%

4 rows

MaritalStatus

```
select count(*) as no_of_employee , MaritalStatus from
employee_attrition where Attrition='Yes' group by MaritalStatus
```

Table	**Visualization 1**

no_of_employee

35.4%

50.6%

13.9%

3 rows

```
-- Insight we got as of now
-- 27.8% people below to age group from 20-25 are leaving
-- Research & Development people are leaving
-- Life Sciences people are leaving
-- 41% from Batchlor Degree are leaving
-- people having low EnvironmentSatisfaction are leaving
-- 31.2% people having low JobSatisfaction are leaving
--50.6% single people are leaving

create table output_data(
  age_group varchar(50),
  department varchar(50),
  educationField varchar(50),
  Education_degree varchar(50),
  environmentSatisfaction varchar(50),
  jobSatisfaction varchar(50),
  MaritalStatus varchar(50)
)
```

OK

```
insert into output_data (age_group,department,educationField,Education_degree,environmentSatisfaction,jobSatisfaction,MaritalStatus) v
'Low','31.2%','50.6%')
```

Table

num_affected_rows ▲	**num_inserted_rows** ▲

1	1	1

1 row

Table

	age_group ▲	department ▲	educationField ▲	Education_degree ▲	environmentSatisfaction ▲	jobSatisfaction ▲
1	27.8%	Research and Development	Life Science	41%	Low	31.2%

1 row