HR Data Analytics Project

By Juventus Chakma Paritosh

This is a data analytics project in respect of HR (Human Resources). In this project firstly SQL (Structured Query Language) has been used then the data has been presented with fine dashboard in Tableau. For SQL, PostgreSQL has been used.
A database has been created at first then a table was inserted into the database. Different SQL functions like SELECT, INSERT, COPY, WHERE, GROUP BY, ORDER BY, HAVING, CROSSTAB(), CAST(), ROUND(),SUBQUERIES, DATA TYPES etc. have been used. The main motive of this project is to find out the job satisfaction rating of employees in different department. It also tried to answer attrition rate in different department and by gender.

The queries of SQL are given below with few screenshots of the selective query outputs. At the end a screenshot of the Tableau Dashboard has been provided along with the link to Tableau Public where the Viz is present.

Queries:

```
-- at first we need to create a table with exact columns based on csv file
--
```

```
create table hrdata
 (
     emp_no int8 PRIMARY KEY,
     gender VARCHAR (50) NOT NULL,
     marital_status VARCHAR (50),
     age_band VARCHAR (50),
     age int8,
     department VARCHAR (50),
     education VARCHAR (50),
     education_field VARCHAR (50),
     job_role VARCHAR (50),
     business_travel VARCHAR (50),
     employee_count int8,
     attrition VARCHAR (50),
     attrition_label VARCHAR (50),
     job_satisfaction int8,
     active_employee int8
 )
```

```
-- let us check the table that has been created --
```

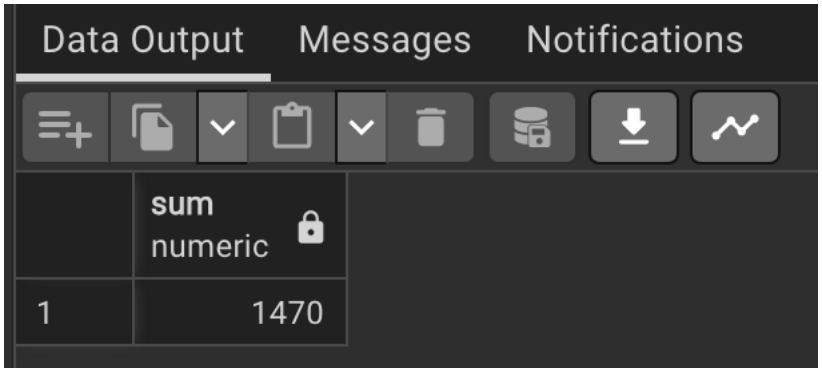
```
select *
    from hrdata; -- so far we can only see the structure with columns and
no data in it -
```

Data Output | Messages | Notifications

	emp_no [PK] bigint	gender character varying (50)	marital_status character varying (50)	age_band character varying (50)	age bigint	department character varying (50)	education character varying (50)	ed ch
1	10001	Female	Single	35 - 44	41	Sales	Associates Degree	Li
2	10002	Male	Married	45 - 54	49	R&D	High School	Li
3	10003	Male	Single	35 - 44	37	R&D	Associates Degree	Ot
4	10004	Female	Married	25 - 34	33	R&D	Master's Degree	Li
5	10005	Male	Married	25 - 34	27	R&D	High School	M
6	10006	Male	Single	25 - 34	32	R&D	Associates Degree	Li
7	10007	Female	Married	Over 55	59	R&D	Bachelor's Degree	M
8	10008	Male	Divorced	25 - 34	30	R&D	High School	Li
9	10009	Male	Single	35 - 44	38	R&D	Bachelor's Degree	Li
10	10010	Male	Married	35 - 44	36	R&D	Bachelor's Degree	M

```
/* let us add data into the table. simply right click on the table and
import the csv file.
after importing again run the query "select * from hrdata" */

-- writing a query to get the sum of employee count --
```

```
select sum (employee_count)
    from hrdata;                -- result shows as "1470" --
```

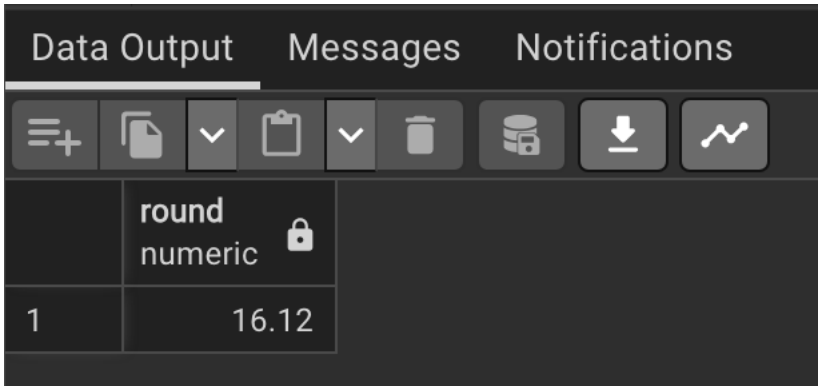



```
-- writing query to get sum of employee count only for education = high
school --
```

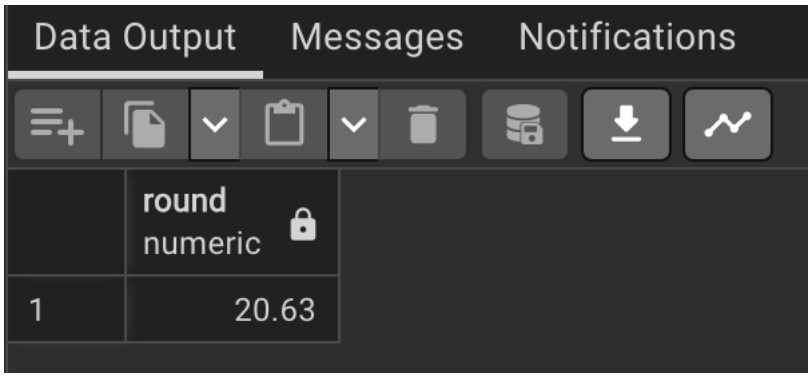
```
select sum (employee_count)
    from hrdata
    where education = 'High School'; -- result shows as "170" --

-- writing query to get sum of employee count only for department = sales
--

select sum (employee_count)
    from hrdata
    where department = 'Sales'; -- result shows as "446" --

-- writing query to get sum of employee count only for department = R&D --

select sum (employee_count)
    from hrdata
    where department = 'R&D';  -- result shows as "961" --
```

```
-- writing query to get sum of employee count only for education_field =
Medical --

select sum (employee_count) as employee_count
    from hrdata
    where education_field = 'Medical';  -- result shows as "464" --

-- writing a query to get the count of attrition from hrdata but attrition
= Yes --

select count (attrition)
    from hrdata
    where attrition = 'Yes'; -- result shows as "237" --

/* writing a query to get the count of attrition from hrdata but attrition
= Yes and
education = 'Doctoral Degree' */

select count (attrition)
    from hrdata
    where attrition = 'Yes' and education = 'Doctoral Degree'; -- result
shows as "5" --

/* writing a query to get the count of attrition from hrdata but attrition
= Yes and
department = 'R&D' */

select count (attrition)
    from hrdata
    where attrition = 'Yes' and department = 'R&D'; -- result shows as
"133"

/* writing a query to get the count of attrition from hrdata but attrition
= Yes and
department = 'R&D' and education_field = 'Medical'*/

select count (attrition)
    from hrdata
    where attrition = 'Yes' and department = 'R&D' and education_field =
'Medical'; -- result shows as "47"

/* writing a query to get the count of attrition from hrdata but attrition
= Yes and
department = 'R&D' and education_field = 'Medical' and education = 'High
School' */

select count (attrition)
    from hrdata
    where attrition = 'Yes' and department = 'R&D' and education_field =
'Medical'
    and education = 'High School'; -- result shows as "9" --
```

```
-- writing a query to get attrition rate --

select round (((select count (attrition) from hrdata where attrition =
'Yes') /
        sum (employee_count)) * 100, 2) from hrdata; -- result shows as
"16.12" --
```




```
-- writing a query to get attrition rate and from department sales --

select round (((select count (attrition) from hrdata where attrition =
'Yes' and department = 'Sales') /
        sum (employee_count)) * 100, 2) from hrdata
        where department = 'Sales'; -- result shows as "20.63" --
```




```
-- writing query to get active employees --

select sum (employee_count) - (select count (attrition) from hrdata where
attrition = 'Yes')
    from hrdata; -- result shows as "1233" --

-- writing query to get active employees but gender = male --

select sum (employee_count) - (select count (attrition) from hrdata where
attrition = 'Yes' and gender = 'Male')
    from hrdata where gender = 'Male'; -- result shows as "732" --

-- writing query to get average age --
```

```
select round (avg (age),0) as Aveg_age from hrdata; -- result shows as
"37" --
```

```
-- writing query to get attrition by gender --
```

```
select gender, count  (attrition) from hrdata
    where attrition = 'Yes'
    group by gender; -- result shows as "female = 87 and male = 150" --

select gender, count  (attrition) from hrdata
    where attrition = 'Yes'
    group by gender
    order by count (attrition) desc;
```

Data Output | Messages | Notifications

	gender character varying (50)	count bigint
1	Male	150
2	Female	87

```
-- writing query to get attrition by gender and education = high school --
```

```
select gender, count  (attrition) from hrdata
    where attrition = 'Yes' and education = 'High School'
    group by gender
    order by count (attrition) desc; -- result shows as "female = 11 and
male = 20" --
```

```
-- writing query to get attrition by department --
```

```
select department, count  (attrition) from hrdata
    where attrition = 'Yes'
    group by department; -- result shows as "HR = 12, Sales = 92 and R&D =
133" --

select department, count  (attrition) from hrdata
    where attrition = 'Yes'
    group by department
    order by count  (attrition) desc;
```

```
-- writing query to get percentage of attriton by department –
```

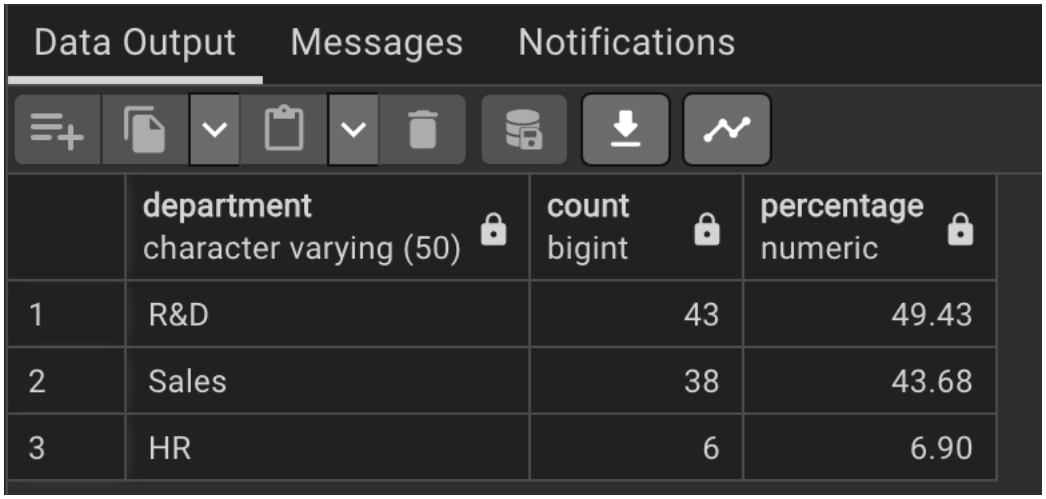
```
select department, count  (attrition),
    round ((cast (count (attrition) as numeric) / -- here changing the
data type to numeric with "cast" --
           (select count (attrition) from hrdata where attrition =
'Yes')) * 100, 2) as percentage
    from hrdata
    where attrition = 'Yes'
    group by department
    order by count  (attrition) desc; -- result shows as "HR = 5.06%,
Sales = 38.82% and R&D = 56.12%" --
```

Data Output | Messages | Notifications

	department character varying (50)	count bigint	percentage numeric
1	R&D	133	56.12
2	Sales	92	38.82
3	HR	12	5.06

```
-- writing query to get percentage of attriton by department and by gender
--
```

```
select department, count  (attrition),
    round ((cast (count (attrition) as numeric) /
         (select count (attrition) from hrdata where attrition = 'Yes' and
gender = 'Female')) * 100, 2) as percentage
    from hrdata
    where attrition = 'Yes' and gender = 'Female'
    group by department
    order by count  (attrition) desc; -- result shows as "HR = 6.90%,
Sales = 43.68% and R&D = 49.43%" --
```



	department character varying (50)	count bigint	percentage numeric
1	R&D	43	49.43
2	Sales	38	43.68
3	HR	6	6.90

```
-- writing query to get number of employee by age group --
```

```
select age, sum (employee_count)
    from hrdata
    group by age
    order by age;
```

```
-- writing query to get number of employee by age group and department =
R&D --
```

```
select age, sum (employee_count)
    from hrdata
    where department = 'R&D'
    group by age
    order by age;
```

```
-- writing query to get atrrition by education_field --
```

```
select education_field, count (attrition)
    from hrdata
    where attrition = 'Yes'
    group by education_field
    order by count (attrition) desc;
```

```
-- writing query to get atrrition by education_field and department =
'Sales' --
```

```
select education_field, count (attrition)
    from hrdata
    where attrition = 'Yes' and department = 'Sales'
    group by education_field
    order by count (attrition) desc;
```

```
-- writing query to get attrition rate by different age group --

select age_band, gender, count (attrition),
    round ((cast (count (attrition) as numeric) /
    (select count (attrition) from hrdata where attrition = 'Yes')) * 100,
2) as percentage
    from hrdata
    where attrition = 'Yes'
    group by age_band, gender
    order by age_band, gender;

-- if crosstab extension is not existing --

create extension if not exists tablefunc;

-- writng query to get rating for job satisfaction --

select *
from crosstab (
    'select job_role, job_satisfaction, sum (employee_count)
    from hrdata
    group by job_role, job_satisfaction
    order by job_role, job_satisfaction')
    as ct (job_role varchar (50), one numeric, two numeric, three numeric,
four numeric)
order by job_role;
```

Data Output | Messages | Notifications

	job_role character varying (50)	one numeric	two numeric	three numeric	four numeric
1	Healthcare Representative	26	19	43	43
2	Human Resources	10	16	13	13
3	Laboratory Technician	56	48	75	80
4	Manager	21	21	27	33
5	Manufacturing Director	26	32	49	38
6	Research Director	15	16	27	22
7	Research Scientist	54	53	90	95
8	Sales Executive	69	54	91	112
9	Sales Representative	12	21	27	23

Tableau Dashboard

Link to Tableau Public:

https://public.tableau.com/views/DataAnalyticsHR/AnalyticsHR?:language=en-US&:display_count=n&:origin=viz_share_link