CS-457 HW No. 5
Owais Waheed (ow07611)
October 16, 2023

Task 1:

Create a SQL database and separate tables for both datasets EmployeeAttrition1.csv and EmployeeAttrition2.csv using an RDBMS (PostgreSQL preferred).

Creating table 1 named as EmployeeAttrition1 with the following query,

```
CREATE TABLE EmployeeAttrition1 (
Employee_Number INT,
Age INT,
Business_Travel varchar(255),
Daily_Rate INT,
Department varchar(255),
Distance_from_Home INT,
Education INT,
Education_Field varchar(255),
Environment_Satisfaction INT,
Gender varchar(255),
Hourly_Rate INT,
Job_Involvement INT,
Job_Level INT,
Job_Role varchar(255),
Job_Satisfaction INT,
Marital_Status varchar(255),
Monthly_Income INT,
Monthly_Rate INT,
Num_Companies_Worked INT,
Percent_Salary_Hike INT,
Performance_Rating INT,
Relationship_Satisfaction INT,
Standard_Hours INT,
Stock_Option_Level INT,
Total_Working_Years INT,
Training_Times_Last_Year INT,
Work_Life_Balance INT,
Years_At_Company INT,
Years_In_CurrentRole INT,
Years_Since_Last_Promotion INT,
Years_With_Curr_Manager INT
);
```

Creating table 2 named as EmployeeAttrition2 with the following query,

```
CREATE TABLE EmployeeAttrition2 (
Employee_Number INT,
```

```
Over_18 varchar(4),
Over_Time varchar(4),
Attrition varchar(4)
);
```

Task 2:

Load/Import the dataset into the table.

The table were populated using the UI interface provided by pgadmin4.
Following screenshots verify the successful import of data:

Query | Query History

```
1  select * from employeeattrition1
2
```

Messages | Data Output | Notifications

Copy with headers

	em		business_travel character varying (255)	daily_rate integer	department character varying (255)	distance_from_home integer	education integer	education_field character varying
1	1	41	Travel_Rarely	1102	Sales	1	2	Life Sciences
2	2	49	Travel_Frequently	279	Research & Development	8	1	Life Sciences
3	4	37	Travel_Rarely	1373	Research & Development	2	2	Other
4	5	33	Travel_Frequently	1392	Research & Development	3	4	Life Sciences
5	7	27	Travel_Rarely	591	Research & Development	2	1	Medical
6	8	32	Travel_Frequently	1005	Research & Development	2	2	Life Sciences
7	10	59	Travel_Rarely	1324	Research & Development	3	3	Medical
8	11	30	Travel_Rarely	1358	Research & Development	24	1	Life Sciences
9	12	38	Travel_Frequently	216	Research & Development	23	3	Life Sciences
10	13	36	Travel_Rarely	1299	Research & Development	27	3	Medical
11	14	35	Travel_Rarely	809	Research & Development	16	3	Medical

Figure 1: Table EmployeeAttrition1

Process completed

Copying table data 'public.employeeattrition2' on database 'postgres' and server 'PostgreSQL 16 (localhost:5432)'

View Processes

Task 3:

Query the database table for EmployeeAttrition1.csv and interpret the results, displaying:

1. the count of total number of records in the table

```
select count(*) as TotalRecords from employeeattrition1
```

	totalrecords bigint
1	1470

2. the count of records for each JobRole in descending order of count

```
select Job_role, count(job_role) from employeeattrition1 group by Job_role order by count(job_role) DESC
```

```
14  select Job_role, count(job_role)
```

Messages | Data Output | Notifications

	job_role character varying (255)	count bigint
1	Sales Executive	326
2	Research Scientist	292
3	Laboratory Technician	259
4	Manufacturing Director	145
5	Healthcare Representative	131
6	Manager	102
7	Sales Representative	83
8	Research Director	80
9	Human Resources	52

3. the average MonthlyIncome and PercentSalaryHike for each JobRole in ascending order of JobRole

```
SELECT Job_Role as JobRole, avg(Monthly_Income) as Average_Monthly_Income, avg(Percent_Salary_Hike) as Salary_Hike FROM EmployeeAttrition1 GROUP BY Job_Role ORDER BY Job_Role ASC;
```

Messages Data Output Notifications

	job_role character varying (255)	avg_monthly_income numeric	avg_salary_hike numeric
1	Healthcare Representative	7528.7633587786259542	15.4503816793893130
2	Human Resources	4235.7500000000000000	14.8076923076923077
3	Laboratory Technician	3237.1698841698841699	15.0463320463320463
4	Manager	17181.676470588235	15.1372549019607843
5	Manufacturing Director	7295.1379310344827586	15.5931034482758621
6	Research Director	16033.550000000000	14.9500000000000000
7	Research Scientist	3239.9726027397260274	15.4486301369863014
8	Sales Executive	6924.2791411042944785	14.8895705521472393
9	Sales Representative	2626.0000000000000000	15.6746987951807229

4. the average JobSatisfaction for each Gender and MaritalStatus

Select avg(job_satisfaction) as AVG_jobsatisfaction , gender , marital_status from employeeattrition1 group by Marital_status,gender

Messages Data Output Notifications

	avg_jobsatisfaction numeric	gender character varying (255)	marital_status character varying (255)
1	2.6838235294117647	Female	Married
2	2.5299145299145299	Female	Divorced
3	2.7738693467336683	Female	Single
4	2.7638376383763838	Male	Single
5	2.7381546134663342	Male	Married
6	2.7904761904761905	Male	Divorced

5. the range (Min and Max) of Age and HourlyRate for each JobRole

```
SELECT job_role as jobrole , min(Age) as min_age, Max(Age) as max_age,
min(Hourly_Rate)
as min_hourly_rate, Max(Hourly_Rate) as max_hourly_rate
FROM EmployeeAttrition1
GROUP BY job_role;
```

Messages | Data Output | Notifications

	jobrole character varying (255)	min_age integer	max_age integer	min_hourly_rate integer	max_hourly_rate integer
1	Manager	30	60	30	99
2	Research Scientist	18	59	30	100
3	Healthcare Representative	24	60	30	100
4	Human Resources	19	59	31	100
5	Laboratory Technician	18	59	30	100
6	Manufacturing Director	22	59	30	100
7	Sales Representative	18	53	30	100
8	Sales Executive	24	60	30	100
9	Research Director	27	58	30	99

6. Join two tables for EmployeeAttrition1.csv and EmployeeAttrition2.csv and display 20 records with the following columns
   - EmployeeNumber, Age, Gender, JobRole, OverTime and Attrition

select ea1.employee_number, age , gender , job_role , over_time ,attrition from employeeattrition1 as ea1 , employeeattrition2 as ea2 where ea1.employee_number = ea2.employee_number limit 20

No limit

Query | Query History

Messages | Data Output | Notifications

	employee_number integer	age integer	gender character varying (255)	job_role character varying (255)	over_time character varying (4)	attrition character varying (4)
1	1	41	Female	Sales Executive	Yes	Yes
2	2	49	Male	Research Scientist	No	No
3	4	37	Male	Laboratory Technician	Yes	Yes
4	5	33	Female	Research Scientist	Yes	No
5	7	27	Male	Laboratory Technician	No	No
6	8	32	Male	Laboratory Technician	No	No
7	10	59	Female	Laboratory Technician	Yes	No
8	11	30	Male	Laboratory Technician	No	No
9	12	38	Male	Manufacturing Director	No	No
10	13	36	Male	Healthcare Representative	No	No
11	14	35	Male	Laboratory Technician	No	No
12	15	29	Female	Laboratory Technician	Yes	No
13	16	31	Male	Research Scientist	No	No
14	18	34	Male	Laboratory Technician	No	No
15	19	28	Male	Laboratory Technician	Yes	Yes
16	20	29	Female	Manufacturing Director	No	No
17	21	32	Male	Research Scientist	Yes	No
18	22	22	Male	Laboratory Technician	Yes	No
19	23	53	Female	Manager	No	No
20	24	38	Male	Research Scientist	Yes	No