Data Science Assignment 5

Dua Batool

db07098@st.habib.edu.pk

Part 1:

Creating Table EmployeeAttrition1:

```
CREATE TABLE employeeattrition1 (
    EmployeeNumber INT,
    Age INT,
    BusinessTravel VARCHAR (50),
    DailyRate INT,
    Department VARCHAR (50),
    DistanceFromHome INT,
    Education INT,
    EducationField  VARCHAR(50),
    EnvironmentSatisfaction INT,
    Gender VARCHAR (10),
    HourlyRate INT,
    JobInvolvement INT,
    JobLevel INT,
    JobRole VARCHAR (50),
    JobSatisfaction INT,
    MaritalStatus VARCHAR (10),
    MonthlyIncome INT,
    MonthlyRate INT,
    NumCompaniesWorked INT,
    PercentSalaryHike INT,
    PerformanceRating INT,
    RelationshipSatisfaction INT,
    StandardHours INT,
    StockOptionLevel INT,
    TotalWorkingYears INT,
    TrainingTimesLastYear INT,
    WorkLifeBalance INT,
    YearsAtCompany INT,
    YearsInCurrentRole INT,
    YearsSinceLastPromotion INT,
    YearsWithCurrManager INT
);
```

Creating Table EmployeeAttrition2:

```
CREATE TABLE employeeattrition2(
    EmployeeNumber INT,
    Over18 CHAR,
    OverTime VARCHAR (4),
    Attrition VARCHAR (4)
);
```

Showing that Data has been imported into employeeattrition1 from the csv file:

	employeenumber integer	age integer	businesstravel character varying (50)	dailyrate integer	department character varying (50)	distancefromhome integer	education integer	educationfield character varying (50)	enviro intege
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	
12	15	29	Travel_Rarely	153	Research & Development	15	2	Life Sciences	
13	16	31	Travel_Rarely	670	Research & Development	26	1	Life Sciences	
14	18	34	Travel_Rarely	1346	Research & Development	19	2	Medical	
15	19	28	Travel_Rarely	103	Research & Development	24	3	Life Sciences	
16	20	29	Travel_Rarely	1389	Research & Development	21	4	Life Sciences	
17	21	32	Travel_Rarely	334	Research & Development	5	2	Life Sciences	
18	22	22	Non-Travel	1123	Research & Development	16	2	Medical	
19	23	53	Travel_Rarely	1219	Sales	2	4	Life Sciences	
20	24	38	Travel_Rarely	371	Research & Development	2	3	Life Sciences	
21	26	24	Non-Travel	673	Research & Development	11	2	Other	
22	27	36	Travel_Rarely	1218	Sales	9	4	Life Sciences	
23	28	34	Travel_Rarely	419	Research & Development	7	4	Life Sciences	

Showing that Data has been imported into employeeattrition2 from the csv file:

	employeenumber integer	over18 character	overtime character varying (4)	attrition character varying (4)
1	1	Y	Yes	Yes
2	2	Y	No	No
3	4	Y	Yes	Yes
4	5	Y	Yes	No
5	7	Y	No	No
6	8	Y	No	No
7	10	Y	Yes	No
8	11	Y	No	No
9	12	Y	No	No
10	13	Y	No	No
11	14	Y	No	No
12	15	Y	Yes	No
13	16	Y	No	No
14	18	Y	No	No
15	19	Y	Yes	Yes
16	20	Y	No	No
17	21	Y	Yes	No
18	22	Y	Yes	No
19	23	Y	No	No
20	24	Y	Yes	No
21	26	Y	No	No
22	27	Y	No	Yes
23	28	Y	No	No

Interpreting results of EmployeeAttrition1

The count of total number of records in the table:

```
37  select count(*) from employeeattrition1
```

Data Output Messages Notifications

	count bigint
1	1470

The count of records for each JobRole in descending order of count:

```
39  select JobRole, count(*) as num
40  from employeeattrition1
41  group by JobRole
42  order by num desc;
```

Data Output | Messages | Notifications

	jobrole character varying (50)	num 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

The average MonthlyIncome and PercentSalaryHike for each JobRole in ascending order of JobRole:

```
44  select JobRole, avg(monthlyincome) as monthly_income, avg(percentsalaryhike) as salary_hike
45  from employeeattrition1
46  group by JobRole
47  order by count(JobRole) asc;
48
```

Data Output | Messages | Notifications

	jobrole character varying (50)	monthly_income numeric	salary_hike numeric
1	Human Resources	4235.7500000000000000	14.8076923076923077
2	Research Director	16033.550000000000	14.9500000000000000
3	Sales Representative	2626.0000000000000000	15.6746987951807229
4	Manager	17181.676470588235	15.1372549019607843
5	Healthcare Representative	7528.7633587786259542	15.4503816793893130
6	Manufacturing Director	7295.1379310344827586	15.5931034482758621
7	Laboratory Technician	3237.1698841698841699	15.0463320463320463
8	Research Scientist	3239.9726027397260274	15.4486301369863014
9	Sales Executive	6924.2791411042944785	14.8895705521472393

The average JobSatisfaction for each Gender and MaritalStatus:

```
49  select Gender, MaritalStatus, avg(JobSatisfaction) as avg_jobsatisfaction
50  from employeeattrition1
51  group by Gender, MaritalStatus;
52
```

Data Output | Messages | Notifications

	gender character varying (10)	maritalstatus character varying (10)	avg_jobsatisfaction numeric
1	Female	Divorced	2.5299145299145299
2	Female	Married	2.6838235294117647
3	Male	Single	2.7638376383763838
4	Male	Divorced	2.7904761904761905
5	Male	Married	2.7381546134663342
6	Female	Single	2.7738693467336683

The range (Min and Max) of Age and HourlyRate for each JobRole:

```
53  select JobRole, max(age) as max_age, min(age) as min_age,
54  max(HourlyRate) as max_HourlyRate, min(HourlyRate) as min_HourlyRate
55  from employeeattrition1
56  group by JobRole;
57
```

Data Output | Messages | Notifications

	jobrole character varying (50)	max_age integer	min_age integer	max_hourlyrate integer	min_hourlyrate integer
1	Manager	60	30	99	30
2	Research Scientist	59	18	100	30
3	Healthcare Representative	60	24	100	30
4	Human Resources	59	19	100	31
5	Laboratory Technician	59	18	100	30
6	Manufacturing Director	59	22	100	30
7	Sales Representative	53	18	100	30
8	Sales Executive	60	24	100	30
9	Research Director	58	27	99	30

Join two tables for EmployeeAttrition1.csv and EmployeeAttrition2.csv and display 20 records:

```
select e1.EmployeeNumber, e1.Age, e1.Gender, e1.JobRole, e2.OverTime, e2.Attrition
from employeeattrition1 e1, employeeattrition2 e2
where e1.EmployeeNumber = e2.EmployeeNumber
limit 20;
```

	employeenumber integer	age integer	gender character varying (10)	jobrole character varying (50)	overtime 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