jupyter hr_analysis Last Checkpoint: 7 days ago

File Edit View Run Kernel Settings Help Trusted

Markdown

JupyterLab | Python 3 (ipykernel)

Diksha

diksharupanshi@gmail.com

HR DATA ASSESSMENT QUESTIONS

[1]:
```
%load_ext sql
%sql mysql+pymysql://root:admin@localhost/hr_analysis
```

```
The sql extension is already loaded. To reload it, use:
  %reload_ext sql
```

1. Retrieve the total number of employees in the dataset.

[2]:
```
%%cl select count(*) as Total_employees from general_data;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
1 rows affected.
```

[2]:

Total_employees
4410

2. List all unique job roles in the dataset.

[3]:
```
%%cl select distinct(JobRole) from general_data ;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
9 rows affected.
```

[3]:

JobRole
Healthcare Representative
Research Scientist
Sales Executive
Human Resources
Research Director
Laboratory Technician
Manufacturing Director
Sales Representative
Manager

3. Find the average age of employees.

[4]:
```
%%cl select avg(Age) as average_age from general_data;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
1 rows affected.
```

[4]:

average_age
36.92047619189509526

4. Retrieve the names and ages of employees who have worked at the company for more than 5 years.

[5]:
```
# Limit has been applied only to reduce the number of rows displayed in the result.
%%%cl select Emp_name, Age, YearsInCompany from general_data where YearsInCompany>5 limit 10 ;
print(a);
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
10 rows affected.
+----------+-----+----------------+
| Emp_name | Age | YearsInCompany |
+----------+-----+----------------+
| RENEE MARQUART | 38 | 8 |
| ALBERTO PEDRAZA | 31 | 9 |
| LEON ARETE | 46 | 7 |
| NATHAN HADDY | 31 | 9 |
| MARGARET BUCHANAN | 45 | 10 |
| VIRGINIA BARKER | 45 | 10 |
| DENNIS SUTTER | 36 | 15 |
| JOHN BROWN | 35 | 16 |
| KATHRYN BALLOU | 47 | 10 |
| DAVID WEBER | 38 | 10 |
+----------+-----+----------------+
```

5. Get a count of employees grouped by their department.

[6]:
```
%%cl select Department, count(*) as No_of_Employees  from general_data
group by Department;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
3 rows affected.
```

[6]:

Department	No_of_Employees
Sales	1336
Research & Development	2872
Human Resources	188

6. List employees who have 'High' Job Satisfaction.

[7]:
```
%%cl select * from general_data inner join employee_survey_data on general_data.EmployeeID= employee_survey_data.EmployeeID
where JobSatisfaction= 3 limit 10;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
10 rows affected.
```

[7]:

EmployeeID	Emp_name	Age	Attrition	business travel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole
7	DENNIS HERRERA	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	Male	2	Sales Executive
17	GEORGE FOURAS	21	No	Travel_Rarely	Research & Development	3	2	Life Sciences	1	Male	1	Laboratory Technician
23	OLLIE BANKS	53	No	Travel_Rarely	Research & Development	11	4	Life Sciences	1	Female	2	Research Scientist
29	CHRISTOPHER HAZEN	44	Yes	Travel_Frequently	Research & Development	1	2	Medical	1	Male	2	Research Scientist
32	CROCE CASSANO	18	No	Travel_Rarely	Research & Development	1	4	Life Sciences	1	Male	1	Sales Executive
44	GERARDO PINTO	18	No	Travel_Rarely	Sales	7	3	Life Sciences	1	Male	1	Research Scientist
49	ALAN HARVEY	32	No	Travel_Rarely	Sales	2	3	Marketing	1	Male	2	Laboratory Technician
51	BARBARA SCHLOSNER	38	No	Non-Travel	Research & Development	2	3	Life Sciences	1	Male	2	Manufacturing Director
59	BARRY BUCKNER	45	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	Male	3	Manufacturing Director
63	ARNOLD CHICK	27	No	Travel_Rarely	Research & Development	13	2	Medical	1	Female	4	Research Director

In the above cell limit has only been applied to reduce the number of the rows to be displayed

7. Find the highest Monthly Income in the dataset

[8]:
```
%%cl select max(MonthlyIncome) as highest_income from general_data;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
1 rows affected.
```

[8]:

highest_income
199990

8. List employees who have 'Travel_Rarely' as their BusinessTravel type.

[9]:
```
%%cl select * from general_data
where business_travel="Travel_Rarely" limit 10;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
10 rows affected.
```

[9]:

EmployeeID	Emp_name	Age	Attrition	business travel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole
1	ALBERTO PEDRAZA	31	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	1	Healthcare Representative
5	HARVEY EDWIN	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	1	Sales Executive
6	LEON ARETE	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	Female	4	Research Director
7	DENNIS HERRERA	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	Male	2	Sales Executive
8	DONALD BRYANT	29	No	Travel_Rarely	Research & Development	18	3	Life Sciences	1	Male	2	Sales Executive
9	NATHAN HADDY	31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	Male	3	Laboratory Technician
11	KRISTEN BARLOW	45	No	Travel_Rarely	Research & Development	17	2	Medical	1	Male	2	Laboratory Technician
12	DENNIS SUTTER	36	No	Travel_Rarely	Research & Development	28	1	Life Sciences	1	Male	1	Laboratory Technician
13	JOHN BROWN	35	No	Travel_Rarely	Research & Development	14	4	Life Sciences	1	Female	1	Sales Executive
15	DONALD FRITZ	28	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	Male	1	Manufacturing Director

In the above cell limit has only been applied to reduce the number of the rows to be displayed

9. Retrieve the distinct MaritalStatus categories in the dataset.

[10]:
```
%%cl select distinct(MaritalStatus) from general_data;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
3 rows affected.
```

[10]:

MaritalStatus
Married
Single
Divorced

10. Get a list of employees with more than 2 years of work experience but less than 4 years in their current role.

[11]:
```
%%cl select * from general_data
where YearsSinceLastPromotion= 2 and YearsInCompany= 4 limit 10;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
10 rows affected.
```

[11]:

EmployeeID	Emp_name	Age	Attrition	business travel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole
7	DENNIS HERRERA	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	Male	2	Sales Executive
8	DONALD BRYANT	29	No	Travel_Rarely	Research & Development	18	3	Life Sciences	1	Male	2	Sales Executive
17	GEORGE FOURAS	21	No	Travel_Rarely	Research & Development	3	2	Life Sciences	1	Male	1	Laboratory Technician
29	CHRISTOPHER HAZEN	44	Yes	Travel_Frequently	Research & Development	1	2	Medical	1	Male	2	Research Scientist
30	PIERRE FRANCOIS	38	No	Travel_Rarely	Sales	2	3	Marketing	1	Female	1	Manager
31	ROSELYN ROLDAN	29	No	Travel_Rarely	Research & Development	4	3	Medical	1	Male	3	Research Scientist
33	COLLEEN RILEY	49	No	Travel_Frequently	Research & Development	1	1	Medical	1	Female	2	Research Scientist
41	TIMOTHY SULLIVAN	36	No	Travel_Frequently	Research & Development	8	3	Other	1	Female	3	Sales Executive
53	STEPHEN TACCHINI	31	No	Travel_Rarely	Sales	7	4	Marketing	1	Male	2	Sales Executive
55	GREGORY MCFARLAND	31	No	Travel_Rarely	Sales	16	4	Marketing	1	Female	2	Research Scientist

In the above cell limit has only been applied to reduce the number of the rows to be displayed

12. Find the average distance from home for employees in each department.

[12]:
```
%%cl select Department, round(avg(DistanceFromHome),2) as average_distance from general_data
group by Department;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
3 rows affected.
```

[12]:

Department	average_distance
Sales	9.23
Research & Development	9.21
Human Resources	8.29

13. Retrieve the top 5 employees with the highest MonthlyIncome

[13]:
```
%%cl select * from general_data
order by MonthlyIncome desc limit 5;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
5 rows affected.
```

[13]:

EmployeeID	Emp_name	Age	Attrition	business travel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole	Ma
3326	LAWRENCE LIU	42	No	Travel_Rarely	Research & Development	18	1	Technical Degree	1	Female	2	Research Scientist	1
386	KEVIN LABARKEWICZ	42	No	Travel_Rarely	Research & Development	18	1	Technical Degree	1	Female	2	Research Scientist	1
1056	DAVID KUCIA	42	No	Travel_Rarely	Research & Development	18	1	Technical Degree	1	Female	2	Research Scientist	1
2412	NATHAN SADOFSKI	36	No	Travel_Rarely	Research & Development	2	2	Technical Degree	1	Male	3	Sales Executive	1
3942	KEVIN MCNAUGHTON	36	No	Travel_Rarely	Research & Development	2	2	Technical Degree	1	Male	3	Sales Executive	1

14. Calculate the percentage of employees who have had a promotion in the last year.

[14]:
```
%%cl select(select count(EmployeeID) from general_data where YearsSinceLastPromotion=1)/(select(EmployeeID) *100,2)
                                  as Percentage from general_data;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
1 rows affected.
```

[14]:

Percentage
24.27

15. List the employees with the highest and lowest EnvironmentSatisfaction.

[15]:
```
%%cl select general_data.Emp_name, general_data.EmployeeID,employee_survey_data.JobSatisfaction, employee_survey_data.EnvironmentSatisfaction
from general_data inner join employee_survey_data
on general_data.EmployeeID= employee_survey_data.EmployeeID
where employee_survey_data.EnvironmentSatisfaction= (select min(EnvironmentSatisfaction) from employee_survey_data) or
employee_survey_data.EnvironmentSatisfaction= (select max(EnvironmentSatisfaction) from employee_survey_data) limit 10 ;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
10 rows affected.
```

[15]:

Emp_name	EmployeeID	JobSatisfaction	EnvironmentSatisfaction
RENEE MARQUART	4	4	1
HARVEY EDWIN	5	1	4
DENNIS HERRERA	7	3	1
DONALD BRYANT	8	2	1
JOHN BROWN	13	1	4
KATHRYN BALLOU	14	2	1
DONALD FRIZZI	15	4	4
GEORGE FOURAS	17	3	4
MARTIN LALOR JR	18	4	1
DARCY KELLER	20	1	1

In the above cell limit has only been applied to reduce the number of the rows to be displayed

16. Find the employees who have the same JobRole and MaritalStatus.

[16]:
```
%%cl SELECT * FROM general_data
WHERE (JobRole, MaritalStatus) IN (
    SELECT JobRole, MaritalStatus
    FROM general_data
    GROUP BY JobRole, MaritalStatus
    HAVING COUNT(*) > 1) limit 10;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
10 rows affected.
```

[16]:

EmployeeID	Emp_name	Age	Attrition	business travel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole
1	ALBERTO PEDRAZA	31	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	1	Healthcare Representative
2	LAWRENCE LEE	31	No	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	1	Research Scientist
3	DWAYNE CLARK	32	No	Travel_Rarely	Research & Development	17	4	Other	1	Male	4	Sales Executive
4	RENEE MARQUART	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	3	Human Resources
5	HARVEY EDWIN	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	1	Sales Executive
6	LEON ARETE	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	Female	4	Research Director
7	DENNIS HERRERA	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	Male	2	Sales Executive
8	DONALD BRYANT	29	No	Travel_Rarely	Research & Development	18	3	Life Sciences	1	Male	2	Sales Executive
9	NATHAN HADDY	31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	Male	3	Laboratory Technician
10	SUSAN BUCHENDER	45	No	Non-Travel	Research & Development	7	4	Medical	1	Female	4	Laboratory Technician

In the above cell limit has only been applied to reduce the number of the rows to be displayed

17. List the employees with the highest TotalWorkingYears who also have a PerformanceRating of 4

[17]:
```
%%cl select * from general_data inner join manager_survey_data
on general_data.EmployeeID= manager_survey_data.EmployeeID
where general_data.YearsSinceLastPromotion= (select max(YearsSinceLastPromotion) from general_data)
and manager_survey_data.PerformanceRating=4;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
0 rows affected.
```

[17]:

EmployeeID	Emp_name	Age	Attrition	business travel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole	Marital

Note:
There is no such employee with the highest TotalWorkingYears who also have a performance rating of 4

18. Calculate the Average Age and JobSatisfaction for each BusinessTravel type.

[18]:
```
%%cl select general_data.business_travel,round(avg(general_data.Age)) as average_age,round(avg(employee_survey_data.JobSatisfaction),2) as average_JobSat
from general_data inner join employee_survey_data on general_data.EmployeeID= employee_survey_data.EmployeeID
group by general_data.business_travel;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
3 rows affected.
```

[18]:

business travel	Average age	Average_JobSatisfaction
Travel_Rarely	37.0	2.7
Travel_Frequently	36.0	2.79
Non-Travel	37.0	2.78

19. Retrieve the most common EducationField among employees.

[19]:
```
%%cl select EducationField,count(EducationField) as count from general_data
group by EducationField
order by count desc limit 1 ;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
1 rows affected.
```

[19]:

EducationField	count
Life Sciences	1812

20. List the employees who have worked for the company the longest but haven't had a promotion

[20]:
```
%%cl select * from general_data
where YearsSinceLastPromotion= (select max(YearsSinceLastPromotion) from general_data) and YearsSinceLastPromotion=0;
```

```
 * mysql+pymysql://root:***@localhost/hr_analysis
0 rows affected.
```

[20]:

EmployeeID	Emp_name	Age	Attrition	business travel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole	Marital

Note: There is no such employee.