Relevel Practice Round 3- SQL

Relevel
by Unacademy

Case Study 1

Case Study 1

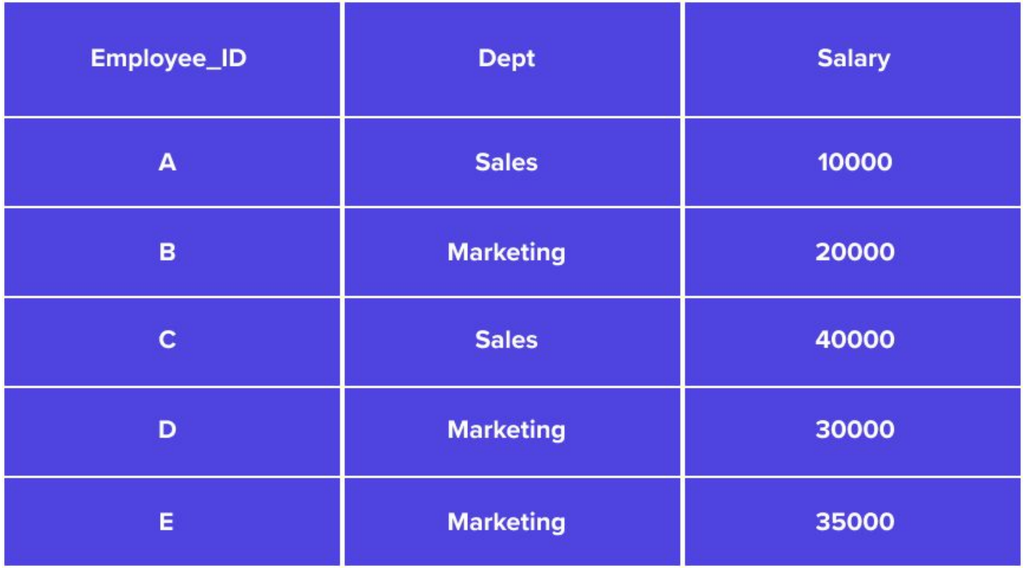
Employee_ID	Dept	Salary
A	Sales	10000
B	Marketing	20000
C	Sales	40000
D	Marketing	30000
E	Marketing	35000

Above is the data for an organization. The name of the table is **employee**.The table contains the information of the employee such as department and salary.The organization wants to maintain parity on the salary so that all the employees are satisfied

Case Study 1

Question 1.

Write a query to create a column Employee_ID_Last two digits of Year to identify which year data we are referring to for particular employee if we want to compare it with next year. Store this information in a tables called employee_v1



Employee_ID	Dept	Salary
A	Sales	10000
B	Marketing	20000
C	Sales	40000
D	Marketing	30000
E	Marketing	35000

Case Study 1

Solution:

{concepts used: string function (concat)}

```
Create table employee_v1 as
Select  concat ( EmployeeID , ‘_’, Right(YEAR(CURDATE()),2) ))  ,*
          from Employee
```

Case Study 1

Question - 2.

There is a rumor in the organization that the attrition rate for the third highest paid employee is the highest since they feel that their increment or their salary should be higher in comparison to the first and second highest paid employee as they performed at par or even better. To support this claim with data write a query to find the third highest paid employee on department level

Employee_ID	Dept	Salary
A	Sales	10000
B	Marketing	20000
C	Sales	40000
D	Marketing	30000
E	Marketing	35000

Case Study 1

Solution:

{Concepts used: Rank function(Windows function)}

```
Select *
 from (
Select * , rank() over (partition by dept order by salary desc) as rnm
 from employee) Base
where base.rnm =3
```

Case Study 1

Question - 3.

Also, Write a query to find the third highest paid employee on organization level using window function to further validate the claim

Employee_ID	Dept	Salary
A	Sales	10000
B	Marketing	20000
C	Sales	40000
D	Marketing	30000
E	Marketing	35000

Case Study 1

Solution:

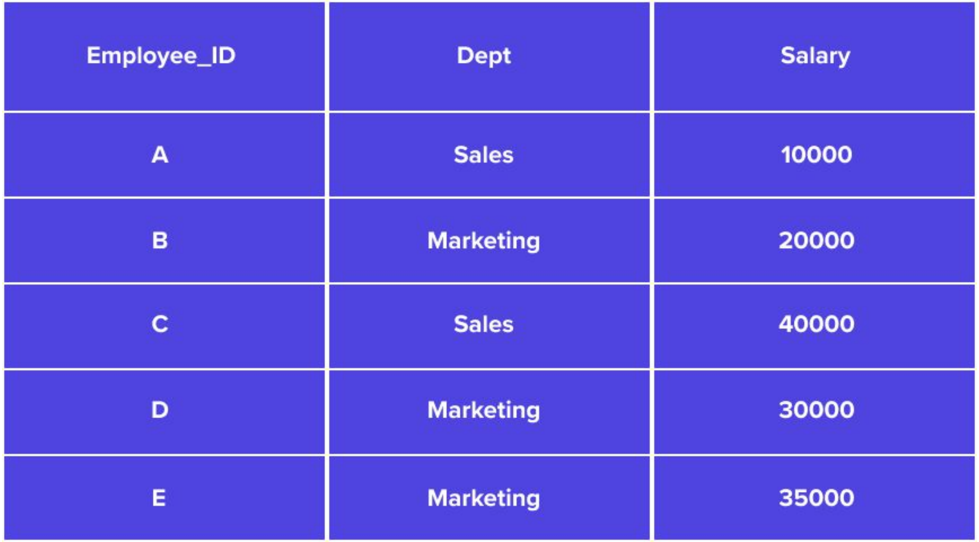
{Concepts used: subqueries, rank function}

```
Select *
 from (
Select * , rank() over (order by salary desc) as rnm
 from employee) Base
where base.rnm =3
```

Case Study 1

Question - 4.

The claim turned out to be true.To further validate the organization wants to compensate the fourth highest paid employee to avoid further attrition .Write a query to find the fourth highest paid employee on organisation level without using window function



Employee_ID	Dept	Salary
A	Sales	10000
B	Marketing	20000
C	Sales	40000
D	Marketing	30000
E	Marketing	35000

Case Study 1

Solution:

{Concepts used: subqueries,window function order by}

```
Select top1*
 from
(Select top 4*
 from employee
order by salary desc) base
order by salary )
```

Case Study 1

Question - 5.

There were some changes in the organization in terms of salary to avoid attrition and bandwidth related crunch but the organization wants to reflect back and see if the employees in all the teams are fairly compensated .Therefore wants to compare the average salary of dept with the average salary of organization. Write a query to create an additional column called flag. If the average salary of the department is greater than the average salary of the organisation then the flag is 1 else 0

Employee_ID	Dept	Salary
A	Sales	10000
B	Marketing	20000
C	Sales	40000
D	Marketing	30000
E	Marketing	35000

Case Study 1

Solution:

{Concepts used: average function , joins}

```
Select * ,
 case when as average_salary_dep> Average_Salary then 1
else 0 end as flag
(
(Select
 avg(salary) as average_salary_dep ,
 dept
from employee
group by dept) A
 join
(select avg(salary) as Average_Salary
from employee ) B) Base
```

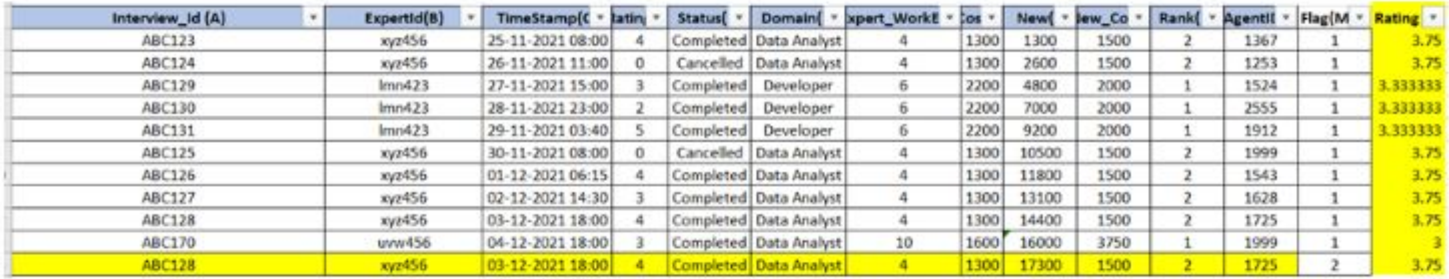
Case Study - 2

Case Study - 2

Given data is the data for an organisation which provides interviews as a service. It has data sorted in ascending order according to the timestamp of the interview.The interviewers are referred as experts.The data is for two domains Data Analyst and Developer.The extra row and column marked in yellow are only for the reference in the last question.The name of the table is **‘Interview’** and the name of columns are Interview_ID, Expert_ID, Timestamp, Rating, Status, Domain, Expert_WorkExperience, Cost, New_Cost_Rank, AgentID, Flag, Rating. The name of other table is ‘**New_Payout_Grid**’

Case Study - 2

Interview



Interview_Id (A)	ExpertId(B)	TimeStamp(C	Ratin	Status(	Domain(	xpert_WorkE	Cos	New(	lew_Co	Rank(	AgentI	Flag(M	Rating
ABC123	xyz456	25-11-2021 08:00	4	Completed	Data Analyst	4	1300	1300	1500	2	1367	1	3.75
ABC124	xyz456	26-11-2021 11:00	0	Cancelled	Data Analyst	4	1300	2600	1500	2	1253	1	3.75
ABC129	lmn423	27-11-2021 15:00	3	Completed	Developer	6	2200	4800	2000	1	1524	1	3.333333
ABC130	lmn423	28-11-2021 23:00	2	Completed	Developer	6	2200	7000	2000	1	2555	1	3.333333
ABC131	lmn423	29-11-2021 03:40	5	Completed	Developer	6	2200	9200	2000	1	1912	1	3.333333
ABC125	xyz456	30-11-2021 08:00	0	Cancelled	Data Analyst	4	1300	10500	1500	2	1999	1	3.75
ABC126	xyz456	01-12-2021 06:15	4	Completed	Data Analyst	4	1300	11800	1500	2	1543	1	3.75
ABC127	xyz456	02-12-2021 14:30	3	Completed	Data Analyst	4	1300	13100	1500	2	1628	1	3.75
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	14400	1500	2	1725	1	3.75
ABC170	uvw456	04-12-2021 18:00	3	Completed	Data Analyst	10	1600	16000	3750	1	1999	1	3
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	17300	1500	2	1725	2	3.75

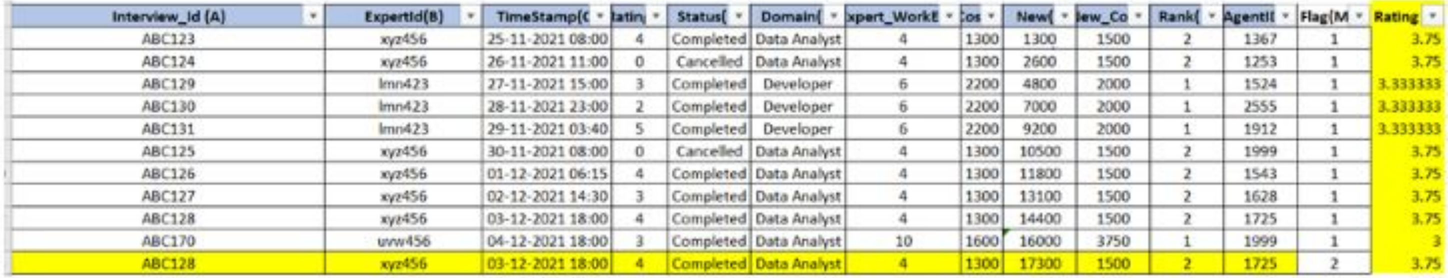
New_Payout_Grid

Years_of_Experience	2 to 4	4+ to 6	6+ To 9	9+
Cost	1500	2000	2500	3750

Case Study - 2

Question 1. The company wants to go on the path of profit earning and therefore, wants to track the expenses they have done.Write a query to calculate the column named 'New' i.e.the amount paid off for the interviews till that particular timestamp.

Interview



Interview_Id (A)	ExpertId(B)	TimeStamp(C	latin	Status(	Domain(	xpert_WorkE	Cos	New(	lew_Co	Rank(	AgentI	Flag(M	Rating
ABC123	xyz456	25-11-2021 08:00	4	Completed	Data Analyst	4	1300	1300	1500	2	1367	1	3.75
ABC124	xyz456	26-11-2021 11:00	0	Cancelled	Data Analyst	4	1300	2600	1500	2	1253	1	3.75
ABC129	lmn423	27-11-2021 15:00	3	Completed	Developer	6	2200	4800	2000	1	1524	1	3.333333
ABC130	lmn423	28-11-2021 23:00	2	Completed	Developer	6	2200	7000	2000	1	2555	1	3.333333
ABC131	lmn423	29-11-2021 03:40	5	Completed	Developer	6	2200	9200	2000	1	1912	1	3.333333
ABC125	xyz456	30-11-2021 08:00	0	Cancelled	Data Analyst	4	1300	10500	1500	2	1999	1	3.75
ABC126	xyz456	01-12-2021 06:15	4	Completed	Data Analyst	4	1300	11800	1500	2	1543	1	3.75
ABC127	xyz456	02-12-2021 14:30	3	Completed	Data Analyst	4	1300	13100	1500	2	1628	1	3.75
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	14400	1500	2	1725	1	3.75
ABC170	uvw456	04-12-2021 18:00	3	Completed	Data Analyst	10	1600	16000	3750	1	1999	1	3
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	17300	1500	2	1725	2	3.75

New_Payout_Grid

Years_of_Experience	2 to 4	4+ to 6	6+ To 9	9+
Cost	1500	2000	2500	3750

Case Study - 2

Solution:

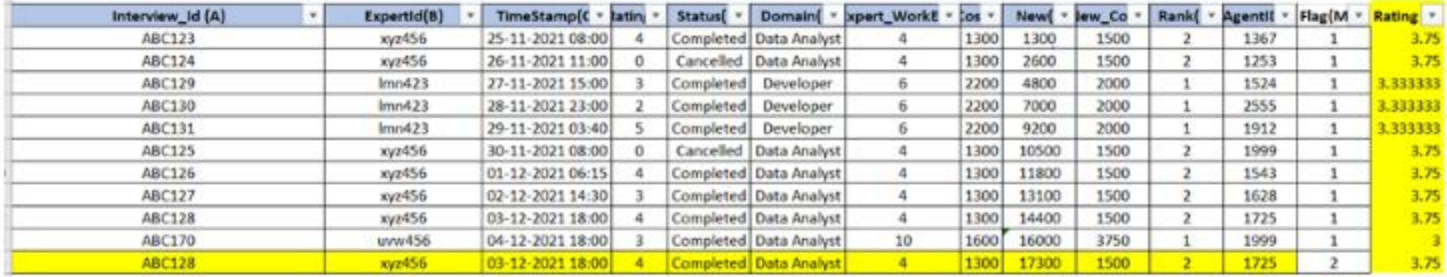
{Concepts used: sum function and timestamp}

```
Select *, sum(cost) over( order by timestamp) as New
 from
Interview
```

Case Study - 2

Question 2. The organisation wants to investigate which interviewers or interviews have led to most incurred cost to the organisation to track for any pattern if exists,Rank the interviews for every domain.Rank the interview for which the cost spent by the organisation for that particular domain is highest as 1.Refer to the column name 'Rank' for more clarity.

Interview



Interview_Id (A)	ExpertId(B)	TimeStamp(C	Ratin	Status(	Domain(	xpert_WorkE	Cos	New(	ew_Co	Rank(	AgentI	Flag(M	Rating
ABC123	xyz456	25-11-2021 08:00	4	Completed	Data Analyst	4	1300	1300	1500	2	1367	1	3.75
ABC124	xyz456	26-11-2021 11:00	0	Cancelled	Data Analyst	4	1300	2600	1500	2	1253	1	3.75
ABC129	lmn423	27-11-2021 15:00	3	Completed	Developer	6	2200	4800	2000	1	1524	1	3.333333
ABC130	lmn423	28-11-2021 23:00	2	Completed	Developer	6	2200	7000	2000	1	2555	1	3.333333
ABC131	lmn423	29-11-2021 03:40	5	Completed	Developer	6	2200	9200	2000	1	1912	1	3.333333
ABC125	xyz456	30-11-2021 08:00	0	Cancelled	Data Analyst	4	1300	10500	1500	2	1999	1	3.75
ABC126	xyz456	01-12-2021 06:15	4	Completed	Data Analyst	4	1300	11800	1500	2	1543	1	3.75
ABC127	xyz456	02-12-2021 14:30	3	Completed	Data Analyst	4	1300	13100	1500	2	1628	1	3.75
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	14400	1500	2	1725	1	3.75
ABC170	uvw456	04-12-2021 18:00	3	Completed	Data Analyst	10	1600	16000	3750	1	1999	1	3
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	17300	1500	2	1725	2	3.75

New_Payout_Grid

Years_of_Experience	2 to 4	4+ to 6	6+ To 9	9+
Cost	1500	2000	2500	3750

Case Study - 2

Solution:

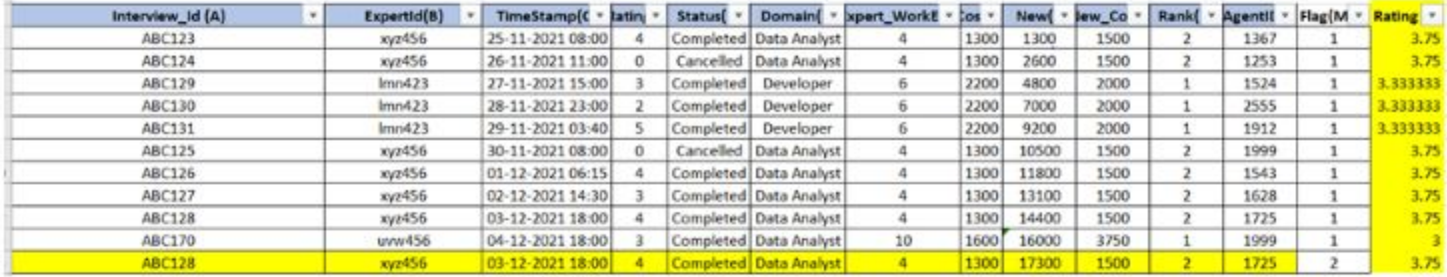
{Concepts used: window functions(rank, partition by)}

```
Select *, rank() over( partition by domain order by cost desc) as Rank
 from
Interview
```

Case Study - 2

Question 3. After tracking the costs, a new payout plan has been introduced for the interviewers where payouts for an interview will be done on the number of years of experience that the interviewer holds.Refer to the new payout grid to calculate the new cost of the expert .Write a query to calculate column ‘New_cost’.

Interview



Interview_Id (A)	ExpertId(B)	TimeStamp(C	latin	Status(	Domain(	xpert_WorkE	Cos	New(	lew_Co	Rank(	AgentI(	Flag(M	Rating
ABC123	xyz456	25-11-2021 08:00	4	Completed	Data Analyst	4	1300	1300	1500	2	1367	1	3.75
ABC124	xyz456	26-11-2021 11:00	0	Cancelled	Data Analyst	4	1300	2600	1500	2	1253	1	3.75
ABC129	lmn423	27-11-2021 15:00	3	Completed	Developer	6	2200	4800	2000	1	1524	1	3.333333
ABC130	lmn423	28-11-2021 23:00	2	Completed	Developer	6	2200	7000	2000	1	2555	1	3.333333
ABC131	lmn423	29-11-2021 03:40	5	Completed	Developer	6	2200	9200	2000	1	1912	1	3.333333
ABC125	xyz456	30-11-2021 08:00	0	Cancelled	Data Analyst	4	1300	10500	1500	2	1999	1	3.75
ABC126	xyz456	01-12-2021 06:15	4	Completed	Data Analyst	4	1300	11800	1500	2	1543	1	3.75
ABC127	xyz456	02-12-2021 14:30	3	Completed	Data Analyst	4	1300	13100	1500	2	1628	1	3.75
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	14400	1500	2	1725	1	3.75
ABC170	uvw456	04-12-2021 18:00	3	Completed	Data Analyst	10	1600	16000	3750	1	1999	1	3
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	17300	1500	2	1725	2	3.75

New_Payout_Grid

Years_of_Experience	2 to 4	4+ to 6	6+ To 9	9+
Cost	1500	2000	2500	3750

Case Study - 2

Solution:

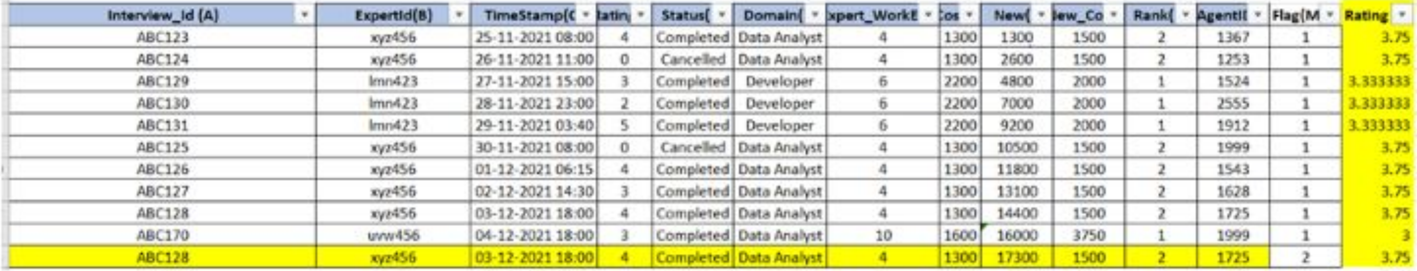
{Concepts used: case statements}

```
Select
case when Expert_WorkExperience >=2 and Expert_WorkExperience<=4 then 1500
when Expert_WorkExperience >=4 and Expert_WorkExperience<=6 then 2000
when Expert_WorkExperience >=6 and Expert_WorkExperience<=9 then 2500
else 3750 end as new_cost
from interview
```

Case Study - 2

Question 4. The organisation further wants to go to a step ahead and wants to track the cost incurred on agent which tag the interview questions. For that they want to extract the AgentID from an extra table which has two columns Interviewed,AgentId .Please note that there is are different tables for each domain and the sheet is named after the domain. Write a query to map the AgentID for each Interview ID.

Interview



Interview_Id (A)	ExpertId(B)	TimeStamp(C	latin	Status(	Domain(	xpert_WorkE	Cos	New(	lew_Co	Rank(	AgentI	Flag(M	Rating
ABC123	xyz456	25-11-2021 08:00	4	Completed	Data Analyst	4	1300	1300	1500	2	1367	1	3.75
ABC124	xyz456	26-11-2021 11:00	0	Cancelled	Data Analyst	4	1300	2600	1500	2	1253	1	3.75
ABC129	lmn423	27-11-2021 15:00	3	Completed	Developer	6	2200	4800	2000	1	1524	1	3.333333
ABC130	lmn423	28-11-2021 23:00	2	Completed	Developer	6	2200	7000	2000	1	2555	1	3.333333
ABC131	lmn423	29-11-2021 03:40	5	Completed	Developer	6	2200	9200	2000	1	1912	1	3.333333
ABC125	xyz456	30-11-2021 08:00	0	Cancelled	Data Analyst	4	1300	10500	1500	2	1999	1	3.75
ABC126	xyz456	01-12-2021 06:15	4	Completed	Data Analyst	4	1300	11800	1500	2	1543	1	3.75
ABC127	xyz456	02-12-2021 14:30	3	Completed	Data Analyst	4	1300	13100	1500	2	1628	1	3.75
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	14400	1500	2	1725	1	3.75
ABC170	uvw456	04-12-2021 18:00	3	Completed	Data Analyst	10	1600	16000	3750	1	1999	1	3
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	17300	1500	2	1725	2	3.75

New_Payout_Grid

Years_of_Experience	2 to 4	4+ to 6	6+ To 9	9+
Cost	1500	2000	2500	3750

Case Study - 2

Solution:

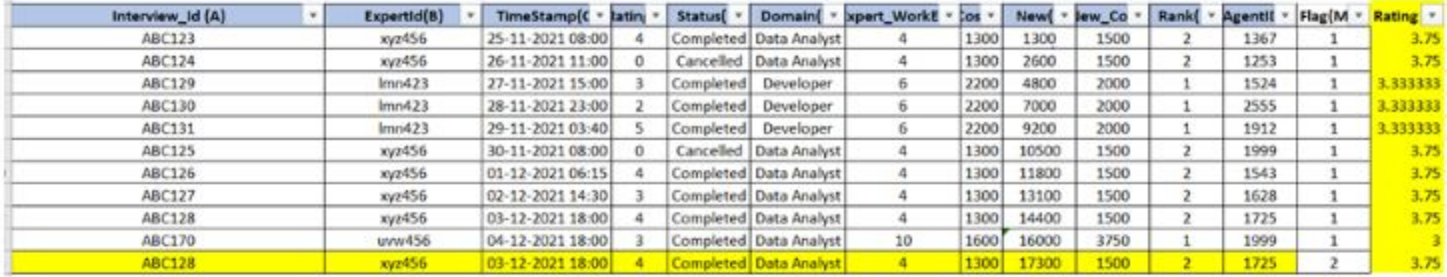
{Concepts used: joins and UNION}

```
Select I.*,D. AgentID
From
 Interview I
Left join
Data Analyst D
On I.Interview ID= D. Interview ID
Where domain = ‘Data Analyst’
Union
Select I.*,D. AgentID
From
 Interview I
Left join
Developer  D
On I.Interview ID= D. Interview ID
Where domain = ‘Developer’
```

Case Study - 2

Question 5. The organization wants to ensure that along with cost effectiveness they are not impacting the quality of the interviews. They want to calculate the average rating of the interviewer which they may use to find a correlation to the old cost.For this write a query to calculate the feedback score of the expert.Refer to the columns highlighted in yellow for the same. Please ensure that the duplicate rows are removed.

Interview



Interview_Id (A)	ExpertId(B)	TimeStamp(C	latin	Status(	Domain(	xpert_WorkE	os	New(	lew_Co	Rank(	AgentI(	Flag(M	Rating
ABC123	xyz456	25-11-2021 08:00	4	Completed	Data Analyst	4	1300	1300	1500	2	1367	1	3.75
ABC124	xyz456	26-11-2021 11:00	0	Cancelled	Data Analyst	4	1300	2600	1500	2	1253	1	3.75
ABC129	lmn423	27-11-2021 15:00	3	Completed	Developer	6	2200	4800	2000	1	1524	1	3.333333
ABC130	lmn423	28-11-2021 23:00	2	Completed	Developer	6	2200	7000	2000	1	2555	1	3.333333
ABC131	lmn423	29-11-2021 03:40	5	Completed	Developer	6	2200	9200	2000	1	1912	1	3.333333
ABC125	xyz456	30-11-2021 08:00	0	Cancelled	Data Analyst	4	1300	10500	1500	2	1999	1	3.75
ABC126	xyz456	01-12-2021 06:15	4	Completed	Data Analyst	4	1300	11800	1500	2	1543	1	3.75
ABC127	xyz456	02-12-2021 14:30	3	Completed	Data Analyst	4	1300	13100	1500	2	1628	1	3.75
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	14400	1500	2	1725	1	3.75
ABC170	uvw456	04-12-2021 18:00	3	Completed	Data Analyst	10	1600	16000	3750	1	1999	1	3
ABC128	xyz456	03-12-2021 18:00	4	Completed	Data Analyst	4	1300	17300	1500	2	1725	2	3.75

New_Payout_Grid

Years_of_Experience	2 to 4	4+ to 6	6+ To 9	9+
Cost	1500	2000	2500	3750

Case Study - 2

Solution:

{Concepts used: subqueries, case statements partition by functions}

```
Select sum(case when status = ‘Completed’ then Rating end ) as Rating/
Count(case when status = ‘Completed’ then InterviewID end ) as Interview as Avg_Ratings,
ExpertID
From
(

Select
 ExpertID,
row_number() over(partition by interviewid) as rnm
From interview)
Base
Where rnm = 1
```

Case Study- 3

Case Study- 3

When unisex perfumes had been lauched in the market, they succeeded in evacuating the myth that **“a man must smell like musky cream wood whereas the lady should have the fragrance like a blooming flower”.** Calvin Klein was the first to launch unisex perfumes in the market. Now, there are huge varieties & brands of unisex perfumes which are equally worn be both the genders. The unisex fragrance is made in such a balanced way that it is neither too strong nor too light or feminine.

Case Study- 3

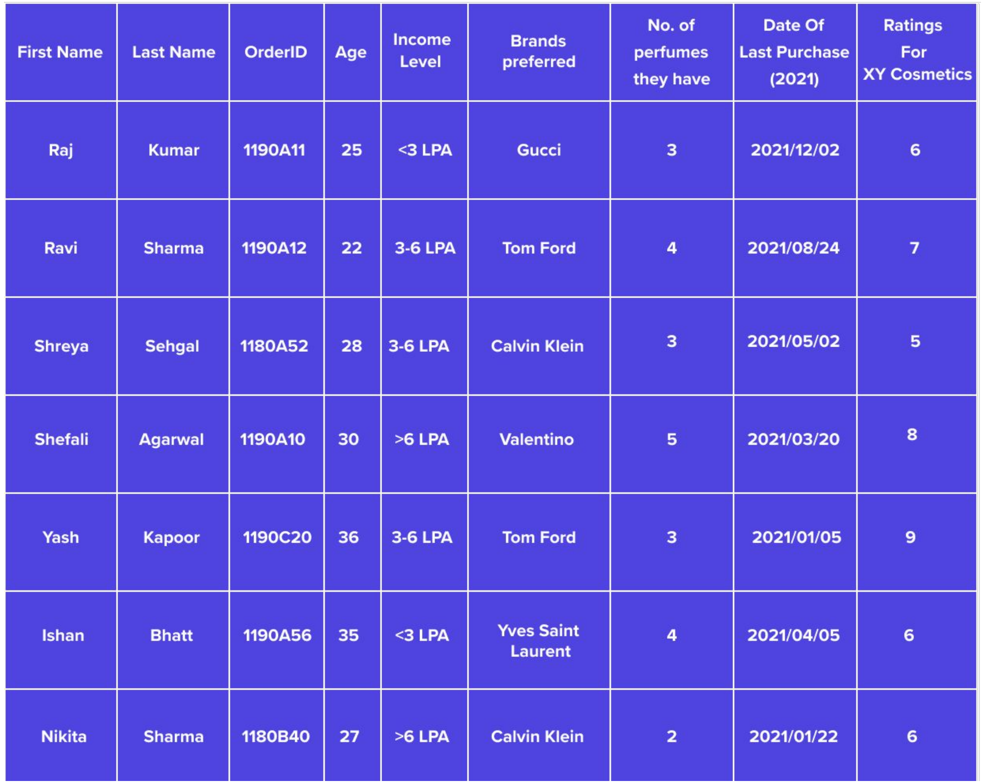
Well, this is not the single reason which have boosted the growth of unisex perfumes. There are many other reasons. Some of them are-

- Economical (1 unisex perfume can be used by both men and women, no need to spend on different fragrance)
- Easy to carry (Can always carry single perfumes in your bags)
- Remedy for forgetfulness (In case you forget to carry your own perfume, you can use your partner’s)
- Best for all occasion (Unisex perfumes comes up with special blend of fragrances which are highly demanded for special occasions)

In this case study, you will get some dataset which would help you to know what is the frequency in which people wear/use perfumes, how many people prefer unisex perfume varieties and what is the pattern of their purchase, which unisex brands they prefer.

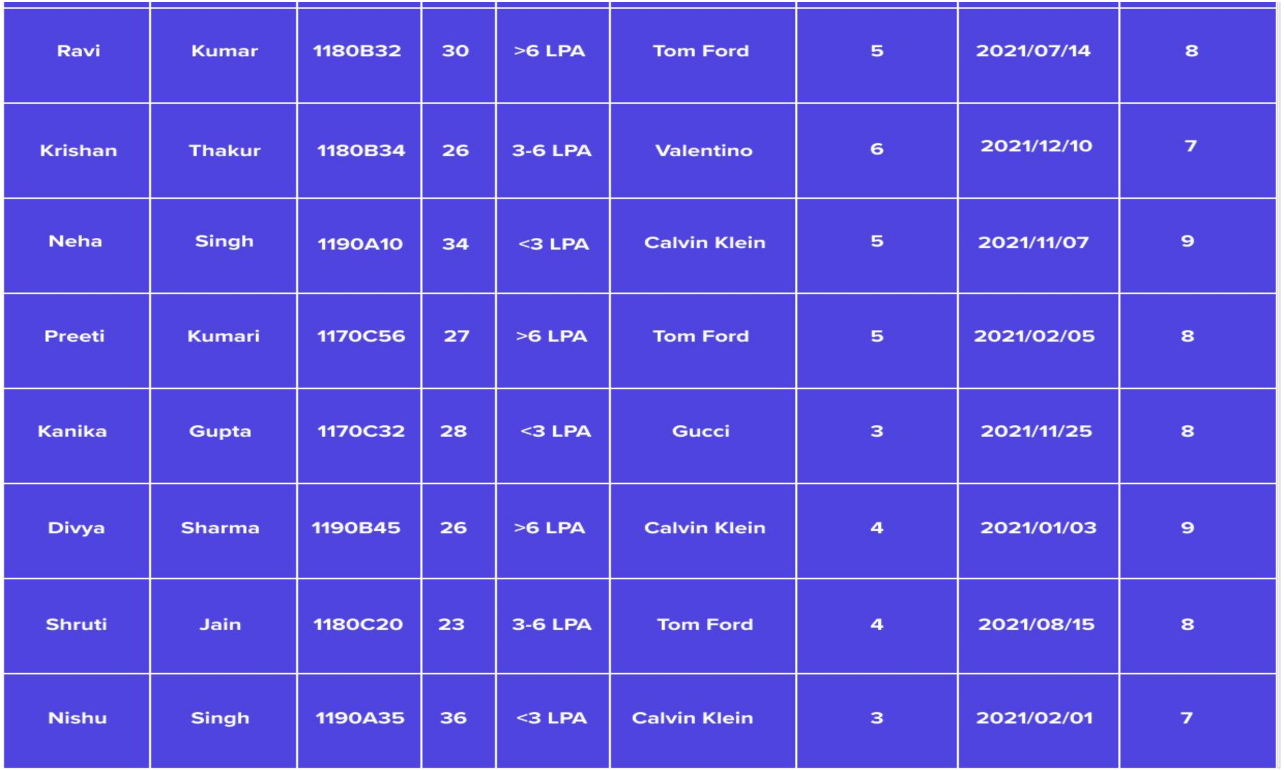
The sample is collected from top loyal 15 customers ordering unisex perfumes from XY Cosmetics in 2021.
The name of the table is **customers**

Case Study- 3



First Name	Last Name	OrderID	Age	Income Level	Brands preferred	No. of perfumes they have	Date Of Last Purchase (2021)	Ratings For XY Cosmetics
Raj	Kumar	1190A11	25	<3 LPA	Gucci	3	2021/12/02	6
Ravi	Sharma	1190A12	22	3-6 LPA	Tom Ford	4	2021/08/24	7
Shreya	Sehgal	1180A52	28	3-6 LPA	Calvin Klein	3	2021/05/02	5
Shefali	Agarwal	1190A10	30	>6 LPA	Valentino	5	2021/03/20	8
Yash	Kapoor	1190C20	36	3-6 LPA	Tom Ford	3	2021/01/05	9
Ishan	Bhatt	1190A56	35	<3 LPA	Yves Saint Laurent	4	2021/04/05	6
Nikita	Sharma	1180B40	27	>6 LPA	Calvin Klein	2	2021/01/22	6

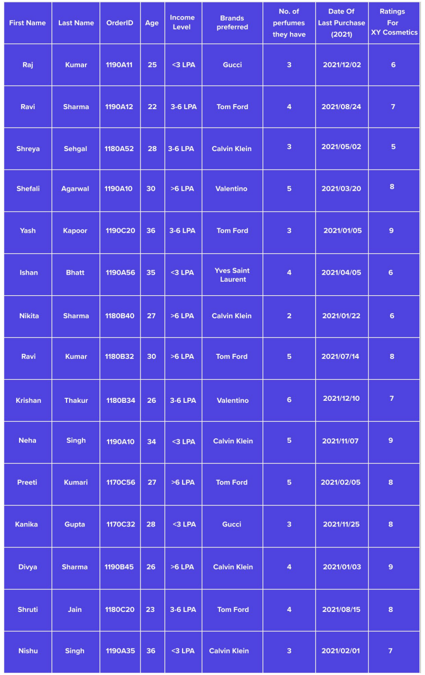
Case Study- 3



Ravi	Kumar	1180B32	30	>6 LPA	Tom Ford	5	2021/07/14	8
Krishan	Thakur	1180B34	26	3-6 LPA	Valentino	6	2021/12/10	7
Neha	Singh	1190A10	34	<3 LPA	Calvin Klein	5	2021/11/07	9
Preeti	Kumari	1170C56	27	>6 LPA	Tom Ford	5	2021/02/05	8
Kanika	Gupta	1170C32	28	<3 LPA	Gucci	3	2021/11/25	8
Divya	Sharma	1190B45	26	>6 LPA	Calvin Klein	4	2021/01/03	9
Shruti	Jain	1180C20	23	3-6 LPA	Tom Ford	4	2021/08/15	8
Nishu	Singh	1190A35	36	<3 LPA	Calvin Klein	3	2021/02/01	7

Case Study- 3

Question 1. Let’s begin by some data cleaning. Identify and remove duplicates. Sort the list in alphabetical order of order ids. Also add a column named “Full_Name” by using the First Name and Last Name columns by creating new table as **perfume_customers.**



First Name	Last Name	OrderID	Age	Income Level	Brands preferred	No. of perfumes they have	Date Of Last Purchase (2021)	Ratings For XY Cosmetics
Raj	Kumar	1190A11	25	<3 LPA	Gucci	3	2021/12/02	6
Ravi	Sharma	1190A12	22	3-6 LPA	Tom Ford	4	2021/08/24	7
Shreya	Sehgal	1180A52	28	3-6 LPA	Calvin Klein	3	2021/05/02	5
Shefali	Agarwal	1190A10	30	>6 LPA	Valentino	5	2021/03/20	8
Yash	Kapoor	1190C20	36	3-6 LPA	Tom Ford	3	2021/01/05	9
Ishan	Bhatt	1190A56	35	<3 LPA	Yves Saint Laurent	4	2021/04/05	6
Nikita	Sharma	1180B40	27	>6 LPA	Calvin Klein	2	2021/01/22	6
Ravi	Kumar	1180B32	30	>6 LPA	Tom Ford	5	2021/07/14	8
Krishan	Thakur	1180B34	26	3-6 LPA	Valentino	6	2021/12/10	7
Neha	Singh	1190A10	34	<3 LPA	Calvin Klein	5	2021/11/07	9
Preeti	Kumari	1170C56	27	>6 LPA	Tom Ford	5	2021/02/05	8
Kanika	Gupta	1170C32	28	<3 LPA	Gucci	3	2021/11/25	8
Divya	Sharma	1190B45	26	>6 LPA	Calvin Klein	4	2021/01/03	9
Shruti	Jain	1180C20	23	3-6 LPA	Tom Ford	4	2021/08/15	8
Nishu	Singh	1190A35	36	<3 LPA	Calvin Klein	3	2021/02/01	7

Case Study- 3

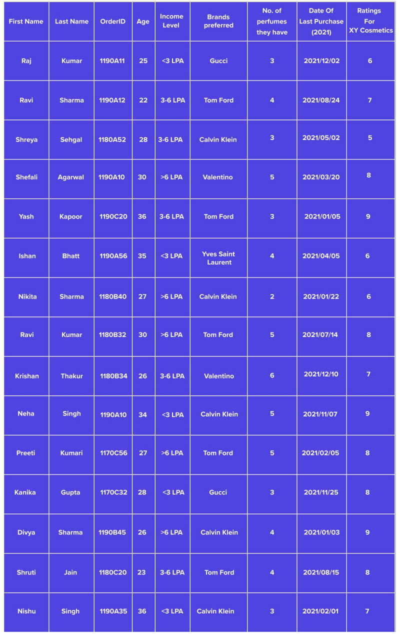
Solution

{Concepts used:concat string function,subqueries,count function(aggregate funcs) and order by statements}

```
Create table perfume_customers as
Select * from
(Select * ,
count(orderid) over (order by orderid) as  cnt,
concat( FirstName , ‘ ’, LastName) as Full_Name
From
customers)
Base
Where cnt =1
Order by orderids
```

Case Study- 3

Question 2. A. In order to have a successful business it is important to understand your customers and this is what XY Cosmetics in 2021 is trying to do.Please help XY Cosmetics by writing a query to calculate the average age of the top 15 customers.
Note: From here you can use(optional) the new generated table Perfume_customers from question 1 to design queries.



First Name	Last Name	OrderID	Age	Income Level	Brands preferred	No. of perfumes they have	Date Of Last Purchase (2021)	Ratings For XY Cosmetics
Raj	Kumar	1190A11	25	<3 LPA	Gucci	3	2021/12/02	6
Ravi	Sharma	1190A12	22	3-6 LPA	Tom Ford	4	2021/08/24	7
Shreya	Sehgal	1180A52	28	3-6 LPA	Calvin Klein	3	2021/05/02	5
Shefali	Agarwal	1190A10	30	>6 LPA	Valentino	5	2021/03/20	8
Yash	Kapoor	1190C20	36	3-6 LPA	Tom Ford	3	2021/01/05	9
Ishan	Bhatt	1190A56	35	<3 LPA	Yves Saint Laurent	4	2021/04/05	6
Nikita	Sharma	1180B40	27	>6 LPA	Calvin Klein	2	2021/01/22	6
Ravi	Kumar	1180B32	30	>6 LPA	Tom Ford	5	2021/07/14	8
Krishan	Thakur	1180B34	26	3-6 LPA	Valentino	6	2021/12/10	7
Neha	Singh	1190A10	34	<3 LPA	Calvin Klein	5	2021/11/07	9
Preeti	Kumari	1170C56	27	>6 LPA	Tom Ford	5	2021/02/05	8
Kanika	Gupta	1170C32	28	<3 LPA	Gucci	3	2021/11/25	8
Divya	Sharma	1190B45	26	>6 LPA	Calvin Klein	4	2021/01/03	9
Shruti	Jain	1180C20	23	3-6 LPA	Tom Ford	4	2021/08/15	8
Nishu	Singh	1190A35	36	<3 LPA	Calvin Klein	3	2021/02/01	7

Case Study- 3

Solution:

{Concepts used: average function}

```
Select avg(age) from
Perfume_customers
```

Question 2 B.XY cosmetics has also decided to choose some lucky customers by organizing a lucky draw.OrderID number 1180B40 is picked up at random. XY cosmetics wants to give a free tester of the preferred perfume to the customer with that order id.Find the brand which is preferred by the customer with the order ID number 1180B40.

{Concepts used are dml commands}

```
select Brands_preferred
 from
Perfume_customers
where
OrderID = '1180B40'
```

Case Study- 3

Question 3. The lucky draw trick kind of worked and to check this further a pot was filled with some random questions and customers were asked to pick those chits up.

The questions in the chits were below:

Who has made the latest purchase?

Who among the customers have purchased the perfume in the second week of the year?

Who are the customers above the age of 26 years have more than 3 perfumes and prefer Calvin Klein?

The answer to these questions will be the customers who will be eligible for next free testers to make the perfume purchase more interesting.

Note: From here you can use(optional) the new generated table Perfume_customers from question 1 to design queries.

First Name	Last Name	OrderID	Age	Income Level	Brands preferred	No. of perfumes they have	Date Of Last Purchase (2021)	Ratings For XY Cosmetics
Raj	Kumar	1190A11	25	<3 LPA	Gucci	3	2021/12/02	6
Ravi	Sharma	1190A12	22	3-6 LPA	Tom Ford	4	2021/08/24	7
Shreya	Sehgal	1180A52	28	3-6 LPA	Calvin Klein	3	2021/05/02	5
Shefali	Agarwal	1190A10	30	>6 LPA	Valentino	5	2021/03/20	8
Yash	Kapoor	1190C20	36	3-6 LPA	Tom Ford	3	2021/01/05	9
Ishan	Bhatt	1190A56	35	<3 LPA	Yves Saint Laurent	4	2021/04/05	6
Nikita	Sharma	1180B40	27	>6 LPA	Calvin Klein	2	2021/01/22	6
Ravi	Kumar	1180B32	30	>6 LPA	Tom Ford	5	2021/07/14	8
Krishan	Thakur	1180B34	26	3-6 LPA	Valentino	6	2021/12/10	7
Neha	Singh	1190A10	34	<3 LPA	Calvin Klein	5	2021/11/07	9
Preeti	Kumari	1170C56	27	>6 LPA	Tom Ford	5	2021/02/05	8
Kanika	Gupta	1170C32	28	<3 LPA	Gucci	3	2021/11/25	8
Divya	Sharma	1190B45	26	>6 LPA	Calvin Klein	4	2021/01/03	9
Shruti	Jain	1180C20	23	3-6 LPA	Tom Ford	4	2021/08/15	8
Nishu	Singh	1190A35	36	<3 LPA	Calvin Klein	3	2021/02/01	7

Case Study- 3

Solution:

{Concepts used: Top function,order by statement, AND operations}

```
Select
Top1 * from
            Perfume_customers
Order by Date of last purchase (2021)

Select  *
from
 Perfume_customers
Where week(Date of last purchase (2021)) = 2

Select *
from
 Perfume_customers
Where age >26
And Brands_prefered = ‘Calvin Klein’
And No. of perfumes they have > 3
```

Case Study- 3

Question 4. In the latest month an issue while billing a customer occurred.The discount which the customer was eligible for was not appearing correctly and the customer highlighted it leading to a poor customer experience. To prevent it in future XY cosmetics has decided to standardize the process. They have come up with the new approach of giving out discounts-There will be 4 different offers for their customers based on their age. Add this column in the dataset and name it ‘XY offer’ and show which of the following offers they belong to in order to avoid confusions in the future.

Note: From here you can use(optional) the new generated table Perfume_customers from question 1 to design queries.

Age	Offer Name
21-25	Offer A (Buy 1 Perfume Get 1 Deodorant Free)
26-30	Offer B (Buy 2 Perfumes Get 1 Deodorant And 1 Face Wash Free)
31-35	Offer C (Buy 1 Perfume And Get 2 Face Wash Free)
36-40	Offer D (Buy 2 Perfumes And Get 2 Deodorants Free)

First Name	Last Name	OrderID	Age	Income Level	Brands preferred	No. of perfumes they have	Date Of Last Purchase (2021)	Ratings For XY Cosmetics
Raj	Kumar	1190A11	25	<3 LPA	Gucci	3	2021/12/02	6
Ravi	Sharma	1190A12	22	3-6 LPA	Tom Ford	4	2021/08/24	7
Shreya	Sehgal	1180A52	28	3-6 LPA	Calvin Klein	3	2021/05/02	5
Shefali	Agarwal	1190A10	30	>6 LPA	Valentino	5	2021/03/20	8
Yash	Kapoor	1190C20	36	3-6 LPA	Tom Ford	3	2021/01/05	9
Ishan	Bhatt	1190A56	35	<3 LPA	Yves Saint Laurent	4	2021/04/05	6
Nikita	Sharma	1180B40	27	>6 LPA	Calvin Klein	2	2021/01/22	6
Ravi	Kumar	1180B32	30	>6 LPA	Tom Ford	5	2021/07/14	8
Krishan	Thakur	1180B34	26	3-6 LPA	Valentino	6	2021/12/10	7
Neha	Singh	1190A10	34	<3 LPA	Calvin Klein	5	2021/11/07	9
Preeti	Kumari	1170C56	27	>6 LPA	Tom Ford	5	2021/02/05	8
Kanika	Gupta	1170C32	28	<3 LPA	Gucci	3	2021/11/25	8
Divya	Sharma	1190B45	26	>6 LPA	Calvin Klein	4	2021/01/03	9
Shruti	Jain	1180C20	23	3-6 LPA	Tom Ford	4	2021/08/15	8
Nishu	Singh	1190A35	36	<3 LPA	Calvin Klein	3	2021/02/01	7

Case Study- 3

Solution:

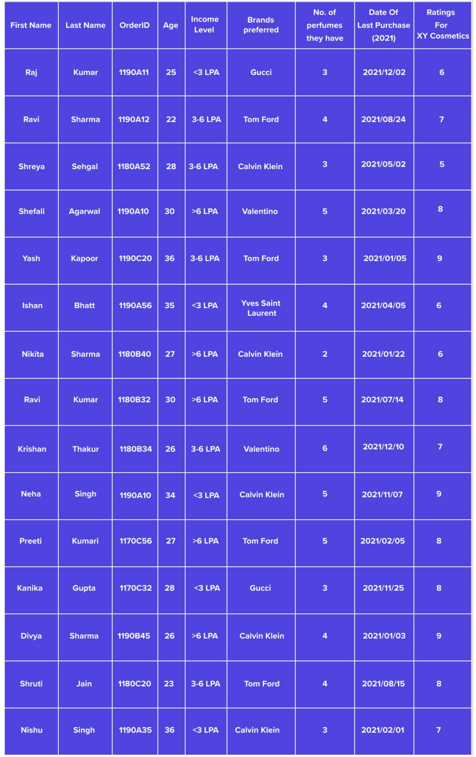
{Concepts used: case statements, create table}

```
Create table XY_Offer as
Select
*,
Case when age>= 21 and age<=25 then ‘Offer A (Buy 1 perfume get 1 Deodorant free)’
Case when age>= 26 and age<=30 then ‘Offer B (Buy 2 perfumes get 1 Deodorant and 1 Face wash free)’
Case when age>= 31 and age<=35 then ‘Offer C (Buy 1 perfume and get 2 Face wash free)’
Case when age>= 36 and age<=40 then ‘Offer D (Buy 2 perfumes and get 2 Deodorants free)’ else end as Offer name


 from

Perfume_customers
```

Case Study- 3

Question 5. To be double sure on the discounts that have been offered , XY cosmetics has decided to understand the customer's interest in the perfume. They want to understand the perfumes possessed by them based on the age group they fall in . help XY cosmetics find the possessed_perfumes for each age group mentioned in the above question

Note: From here you can use(optional) the new generated table Perfume_customers from question 1 to design queries.



First Name	Last Name	OrderID	Age	Income Level	Brands preferred	No. of perfumes they have	Date Of Last Purchase (2021)	Ratings For XY Cosmetics
Raj	Kumar	1190A11	25	<3 LPA	Gucci	3	2021/12/02	6
Ravi	Sharma	1190A12	22	3-6 LPA	Tom Ford	4	2021/08/24	7
Shreya	Sehgal	1180A52	28	3-6 LPA	Calvin Klein	3	2021/05/02	5
Shefali	Agarwal	1190A10	30	>6 LPA	Valentino	5	2021/03/20	8
Yash	Kapoor	1190C20	36	3-6 LPA	Tom Ford	3	2021/01/05	9
Ishan	Bhatt	1190A56	35	<3 LPA	Yves Saint Laurent	4	2021/04/05	6
Nikita	Sharma	1180B40	27	>6 LPA	Calvin Klein	2	2021/01/22	6
Ravi	Kumar	1180B32	30	>6 LPA	Tom Ford	5	2021/07/14	8
Krishan	Thakur	1180B34	26	3-6 LPA	Valentino	6	2021/12/10	7
Neha	Singh	1190A10	34	<3 LPA	Calvin Klein	5	2021/11/07	9
Preeti	Kumari	1170C56	27	>6 LPA	Tom Ford	5	2021/02/05	8
Kanika	Gupta	1170C32	28	<3 LPA	Gucci	3	2021/11/25	8
Divya	Sharma	1190B45	26	>6 LPA	Calvin Klein	4	2021/01/03	9
Shruti	Jain	1180C20	23	3-6 LPA	Tom Ford	4	2021/08/15	8
Nishu	Singh	1190A35	36	<3 LPA	Calvin Klein	3	2021/02/01	7

Case Study- 3

Solution:

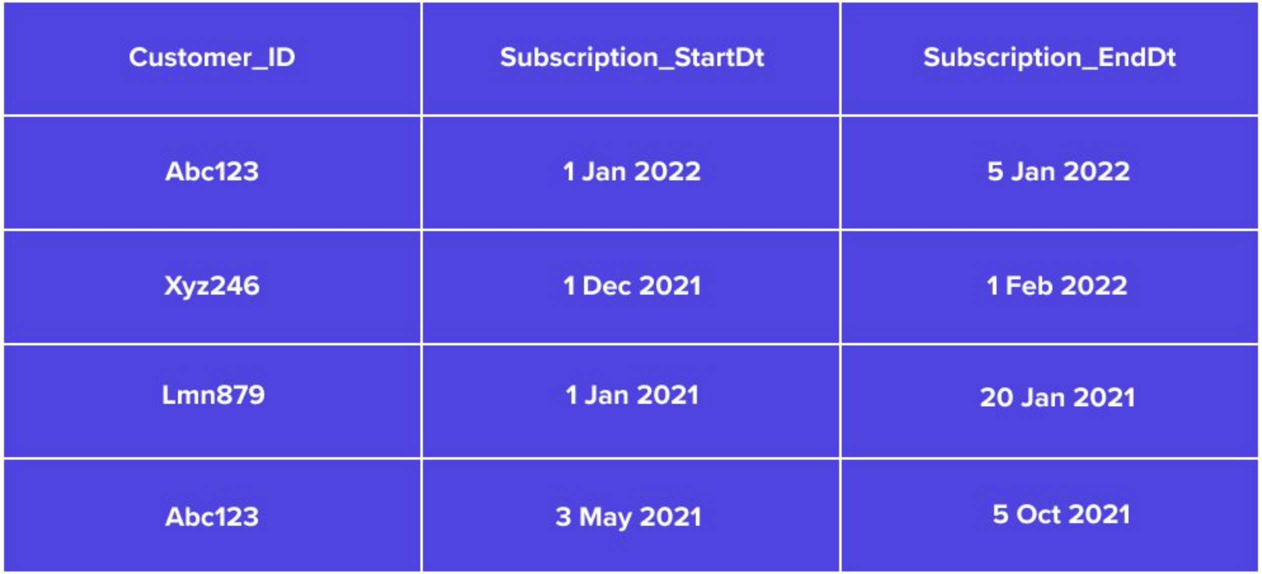
{Concepts used are case statements}

```
Select
Case when age>= 21 and age<=25 then ‘Group1’
Case when age>= 26 and age<=30 then ‘Group2’
Case when age>= 31 and age<=35 then ‘Group3’
Case when age>= 36 and age<=40 then ‘Group4’
 else end as Age_Brackets, sum(No. of perfumes they have) as possessed_perfumes

 from

Perfume_customers
Group by
Case when age>= 21 and age<=25 then ‘Group1’
Case when age>= 26 and age<=30 then ‘Group2’
Case when age>= 31 and age<=35 then ‘Group3’
Case when age>= 36 and age<=40 then ‘Group4’
 else end
```

Case Study - 4

Case Study - 4

Table 1:Amazon Subscription Data
Table Name : **Subscription**
Column Names:Customer_Id , Subscription_StartDt , Subscription_EndDt



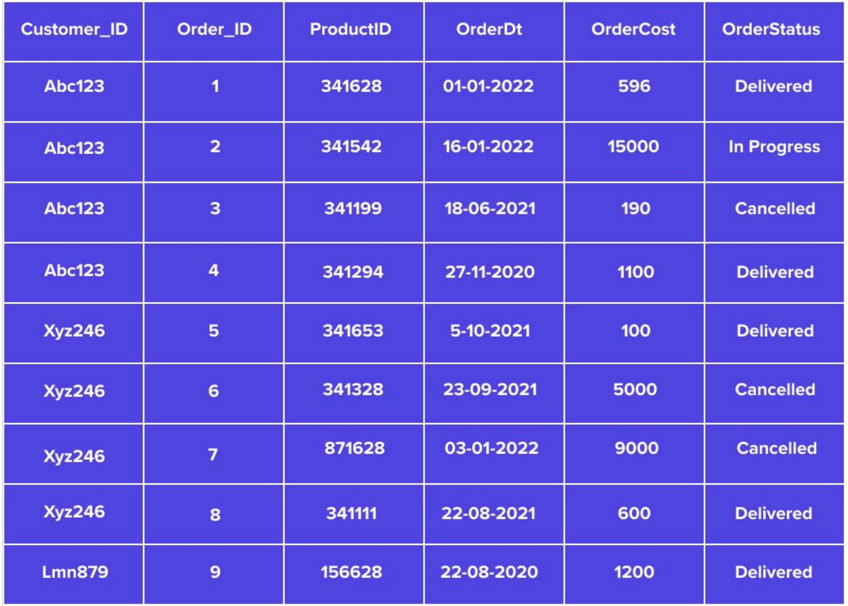
Customer_ID	Subscription_StartDt	Subscription_EndDt
Abc123	1 Jan 2022	5 Jan 2022
Xyz246	1 Dec 2021	1 Feb 2022
Lmn879	1 Jan 2021	20 Jan 2021
Abc123	3 May 2021	5 Oct 2021

Case Study - 4

Table 2:Amazon Orders Table

Table Name : **Order**

Column Names:Customer_Id , Order_ID, OrderDt, Order Cost, OrderStatus



Customer_ID	Order_ID	ProductID	OrderDt	OrderCost	OrderStatus
Abc123	1	341628	01-01-2022	596	Delivered
Abc123	2	341542	16-01-2022	15000	In Progress
Abc123	3	341199	18-06-2021	190	Cancelled
Abc123	4	341294	27-11-2020	1100	Delivered
Xyz246	5	341653	5-10-2021	100	Delivered
Xyz246	6	341328	23-09-2021	5000	Cancelled
Xyz246	7	871628	03-01-2022	9000	Cancelled
Xyz246	8	341111	22-08-2021	600	Delivered
Lmn879	9	156628	22-08-2020	1200	Delivered

We have this data for analysis purposes.We want to generate more revenue out of the product

Case Study - 4

Question 1 Write a query to check for duplicates in ‘Order’ Table.Write the query to find the duplicate rows of this table.

Customer_ID	Subscription_StartDt	Subscription_EndDt
Abc123	1 Jan 2022	5 Jan 2022
Xyz246	1 Dec 2021	1 Feb 2022
Lmn879	1 Jan 2021	20 Jan 2021
Abc123	3 May 2021	5 Oct 2021

Customer_ID	Order_ID	ProductID	OrderDt	OrderCost	OrderStatus
Abc123	1	341628	01-01-2022	596	Delivered
Abc123	2	341542	16-01-2022	15000	In Progress
Abc123	3	341199	18-06-2021	190	Cancelled
Abc123	4	341294	27-11-2020	1100	Delivered
Xyz246	5	341653	5-10-2021	100	Delivered
Xyz246	6	341328	23-09-2021	5000	Cancelled
Xyz246	7	871628	03-01-2022	9000	Cancelled
Xyz246	8	341111	22-08-2021	600	Delivered
Lmn879	9	156628	22-08-2020	1200	Delivered

Case Study - 4

Solution:

{Concepts used: count function and group by statements}

```
SELECT Order_ID, COUNT(Order_ID)
FROM Orders
GROUP BY Order_ID
HAVING COUNT(Order_ID) > 1
```

Case Study - 4

Question 2. Identify the outliers from the data .Identify 10 customers who are outliers.The outlier customer is defined as the one who has maximum difference of the order value of the orders placed by him to the overall average value of orders placed by all customers accounted together.

Customer_ID	Subscription_StartDt	Subscription_EndDt
Abc123	1 Jan 2022	5 Jan 2022
Xyz246	1 Dec 2021	1 Feb 2022
Lmn879	1 Jan 2021	20 Jan 2021
Abc123	3 May 2021	5 Oct 2021

Customer_ID	Order_ID	ProductID	OrderDt	OrderCost	OrderStatus
Abc123	1	341628	01-01-2022	596	Delivered
Abc123	2	341542	16-01-2022	15000	In Progress
Abc123	3	341199	18-06-2021	190	Cancelled
Abc123	4	341294	27-11-2020	1100	Delivered
Xyz246	5	341653	5-10-2021	100	Delivered
Xyz246	6	341328	23-09-2021	5000	Cancelled
Xyz246	7	871628	03-01-2022	9000	Cancelled
Xyz246	8	341111	22-08-2021	600	Delivered
Lmn879	9	156628	22-08-2020	1200	Delivered

Case Study - 4

Solution:

{Concepts used: joins,subqueries and base table concept}

```
Select
 customer_id
from  (
Select
customer_id,rank() over( order by abs(average_for_customer,overall_average desc) as rnk
 from
(
( Select
 avg(ordercost)  average_for_customer,
 Customer_Id
From order
Group by Customer_Id ) A
 join
(Select
avg(ordercost) as overall_average
 From order) B)
Base
) base1 where rnk <=10
```

Case Study - 4

Question 3. Amazon is distributing gift cards to random customers as a promotional event.The customer for the month of May is defined as the one who has placed 5th highest orders in that month for 2021.Write a query to do so without using window/partition function.

Customer_ID	Subscription_StartDt	Subscription_EndDt
Abc123	1 Jan 2022	5 Jan 2022
Xyz246	1 Dec 2021	1 Feb 2022
Lmn879	1 Jan 2021	20 Jan 2021
Abc123	3 May 2021	5 Oct 2021

Customer_ID	Order_ID	ProductID	OrderDt	OrderCost	OrderStatus
Abc123	1	341628	01-01-2022	596	Delivered
Abc123	2	341542	16-01-2022	15000	In Progress
Abc123	3	341199	18-06-2021	190	Cancelled
Abc123	4	341294	27-11-2020	1100	Delivered
Xyz246	5	341653	5-10-2021	100	Delivered
Xyz246	6	341328	23-09-2021	5000	Cancelled
Xyz246	7	871628	03-01-2022	9000	Cancelled
Xyz246	8	341111	22-08-2021	600	Delivered
Lmn879	9	156628	22-08-2020	1200	Delivered

Case Study - 4

Solution:

{Concepts used: joins,subqueries,rank function and base table concept}

```
Select
customer_id from
(
Select *,
rank() over(order by number_of_orders desc) as rnk
From
(Select
count(order_id) number_of_orders,
Customer_id
From
Order
Where
month(OrderDt) = 5
year(OrderDt) = 2021) base)
base1)
base 2
 where rnk =5
```

Case Study - 4

Question 4. Suppose Amazon has some logistics issues for duration- Jan 15’22 -Feb 14’22 due to which they have to outsource the order delivery process.They cannot exploit the help that they have got,therefore cannot deliver all the orders but they also want to complete the monthly minimum target of Rs 10 lac revenue .What are the productIDs that they should dispatch such that they have to take as minimum help as possible and also complete their target at the same time.

Customer_ID	Subscription_StartDt	Subscription_EndDt
Abc123	1 Jan 2022	5 Jan 2022
Xyz246	1 Dec 2021	1 Feb 2022
Lmn879	1 Jan 2021	20 Jan 2021
Abc123	3 May 2021	5 Oct 2021

Customer_ID	Order_ID	ProductID	OrderDt	OrderCost	OrderStatus
Abc123	1	341628	01-01-2022	596	Delivered
Abc123	2	341542	16-01-2022	15000	In Progress
Abc123	3	341199	18-06-2021	190	Cancelled
Abc123	4	341294	27-11-2020	1100	Delivered
Xyz246	5	341653	5-10-2021	100	Delivered
Xyz246	6	341328	23-09-2021	5000	Cancelled
Xyz246	7	871628	03-01-2022	9000	Cancelled
Xyz246	8	341111	22-08-2021	600	Delivered
Lmn879	9	156628	22-08-2020	1200	Delivered

Case Study - 4

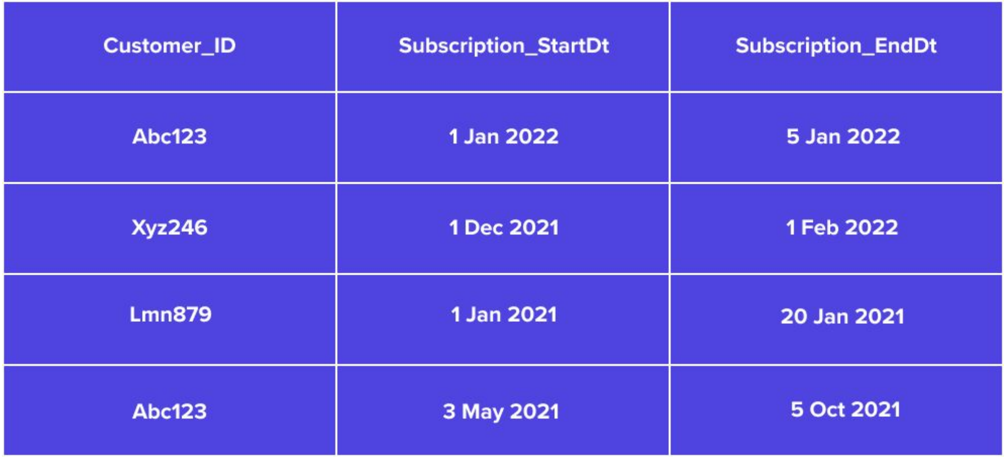
Solution:

{Concepts used: sum function,order by statements,sub queries and base table concept}

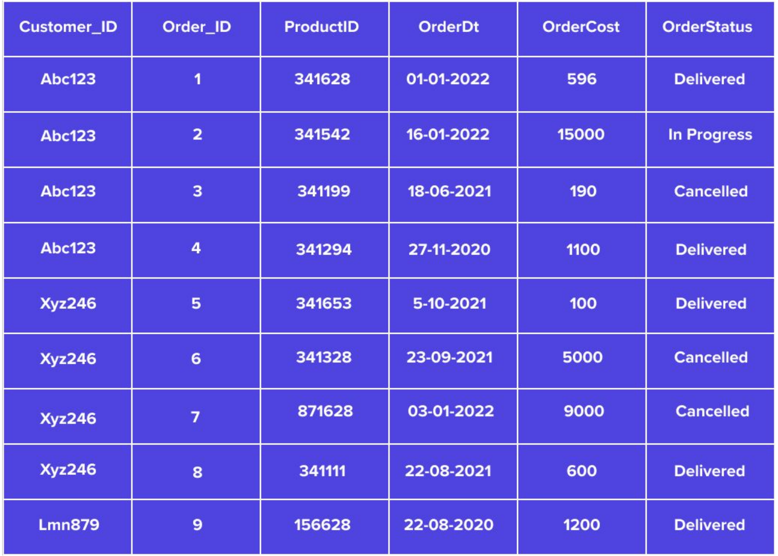
```
Select * from (
SELECT ProductID customer_ID,
SUM (ordercost) OVER ( ORDER BY ordercost) AS RunningTotal
FROM order
Where orderdt =>’15-01-2022’
And orderdt <=’14-02-2022’

And orderstatus = ‘In Progress’) base
Where RunningTotal<= 1000000
```

Case Study - 4

Question 5. Suppose to promote the product,it has been decided to distribute free membership to some customers. For this promotional event, write a query to identify the customers which have overlapping subscription periods.



Customer_ID	Subscription_StartDt	Subscription_EndDt
Abc123	1 Jan 2022	5 Jan 2022
Xyz246	1 Dec 2021	1 Feb 2022
Lmn879	1 Jan 2021	20 Jan 2021
Abc123	3 May 2021	5 Oct 2021



Customer_ID	Order_ID	ProductID	OrderDt	OrderCost	OrderStatus
Abc123	1	341628	01-01-2022	596	Delivered
Abc123	2	341542	16-01-2022	15000	In Progress
Abc123	3	341199	18-06-2021	190	Cancelled
Abc123	4	341294	27-11-2020	1100	Delivered
Xyz246	5	341653	5-10-2021	100	Delivered
Xyz246	6	341328	23-09-2021	5000	Cancelled
Xyz246	7	871628	03-01-2022	9000	Cancelled
Xyz246	8	341111	22-08-2021	600	Delivered
Lmn879	9	156628	22-08-2020	1200	Delivered

Case Study - 4

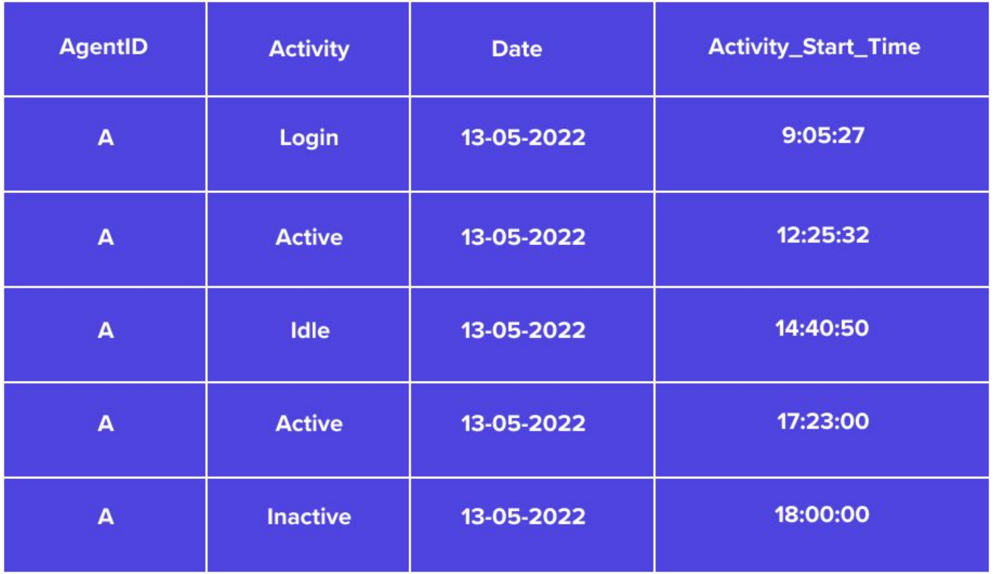
Solution:

{Concepts used: joins AND operations}

```
SELECT *
FROM new_schema.tab1 AS s1
LEFT JOIN new_schema.tab1 AS s2
    ON s1.Customer_ID != s2.Customer_ID
        AND s1.Subscription_StartDt <= s2.Subscription_EndDt
        AND s1.Subscription_EndDt >= s2.Subscription_StartDt
```

Case Study - 5

Case Study - 5



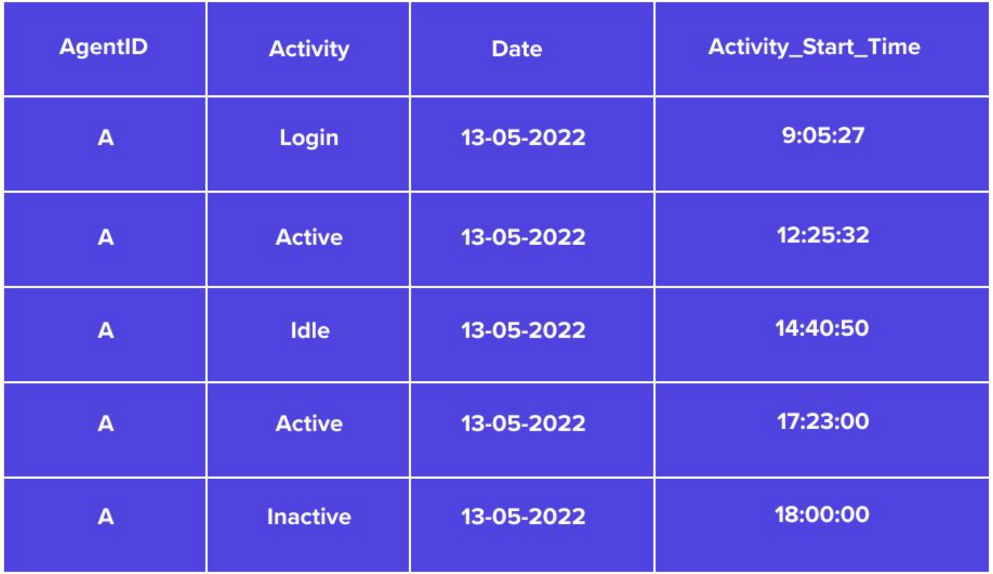
AgentID	Activity	Date	Activity_Start_Time
A	Login	13-05-2022	9:05:27
A	Active	13-05-2022	12:25:32
A	Idle	13-05-2022	14:40:50
A	Active	13-05-2022	17:23:00
A	Inactive	13-05-2022	18:00:00

Below is the Agent table which stores the agent's timesheet.It is required to track the agent performance for improvement in the customer service and therefore, necessary to avoid customer churn

Table Name - **Agent**

Case Study - 5

Question 1. Before doing analysis it is important to clean data. An important step towards data cleaning is to identify the duplicate rows in the table. Write a query for the same.



AgentID	Activity	Date	Activity_Start_Time
A	Login	13-05-2022	9:05:27
A	Active	13-05-2022	12:25:32
A	Idle	13-05-2022	14:40:50
A	Active	13-05-2022	17:23:00
A	Inactive	13-05-2022	18:00:00

Case Study - 5

Solution:

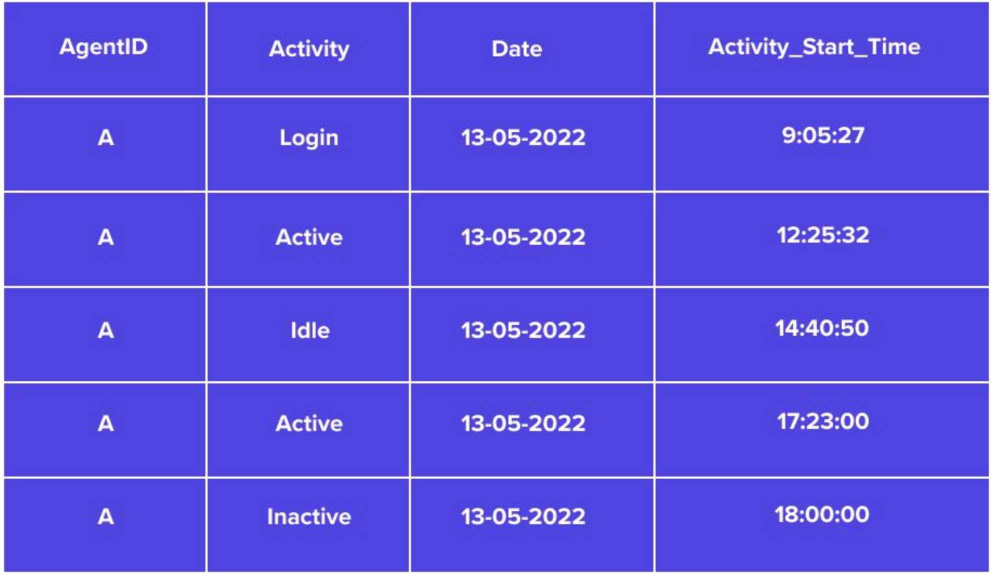
{Concepts used: aggregation function(count)}

```
Select Agentid,activity,date,activity_start_time, count(*)
from Agent
group by Agentid,activity,date,activity_start_time
having count(*) > 1
```

Case Study - 5

Question 2.

For analysis purpose , it is important to create a key to tag the activity to the particular agent who did this activity. Create the column key as (‘AgentID’ – ‘Activity’). Write a query for the same



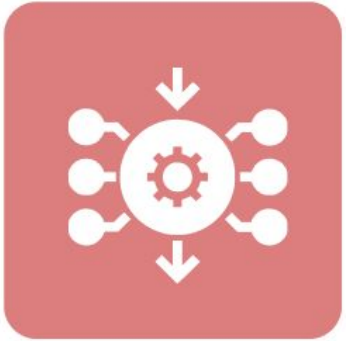
AgentID	Activity	Date	Activity_Start_Time
A	Login	13-05-2022	9:05:27
A	Active	13-05-2022	12:25:32
A	Idle	13-05-2022	14:40:50
A	Active	13-05-2022	17:23:00
A	Inactive	13-05-2022	18:00:00

Case Study - 5

Solution:

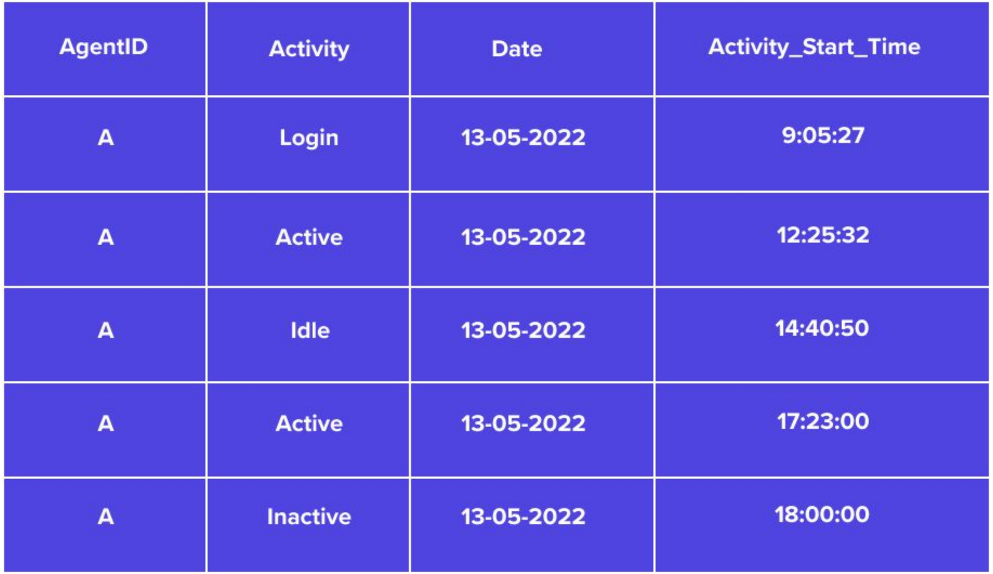
{Concepts used: concat string function}

```
Select * , concat( AgentID , ‘ – ‘ , Activity ) from Agent
```



Case Study - 5

Question 3. The agents are believed to login for 9 hours still the efficiency has decreased. Is it because the number of active hours for the analyst which is less? Write a query to find the number of ‘Active’ hours of the Agent



AgentID	Activity	Date	Activity_Start_Time
A	Login	13-05-2022	9:05:27
A	Active	13-05-2022	12:25:32
A	Idle	13-05-2022	14:40:50
A	Active	13-05-2022	17:23:00
A	Inactive	13-05-2022	18:00:00

Case Study - 5

Solution:

{Concepts used: subqueries}

```
Select *, sum(diff) as active_hours
from (
Select * , datediff(hour, activity_start_time - next_activity_timestamp) as diff
 from (
Select Agentid,activity,date,activity_start_time, lead(activity_start_time)
Over ( partition of agentid,date order by activity_start_time) as next_activity_timestamp
FROM Agent
) base
 where Activity = ‘Active’
)base1
```

Case Study - 5

Question 4. A study shows that agent logging in on time perform better.Agents might be logging in for 9 hours but are they logging in on time.Write a query to create an extra column called flag which will show if the agent has logged in on time . If the Agent Logins in <= 9:00:00 then the flag will be 'YES' else 'NO'

AgentID	Activity	Date	Activity_Start_Time
A	Login	13-05-2022	9:05:27
A	Active	13-05-2022	12:25:32
A	Idle	13-05-2022	14:40:50
A	Active	13-05-2022	17:23:00
A	Inactive	13-05-2022	18:00:00

Case Study - 5

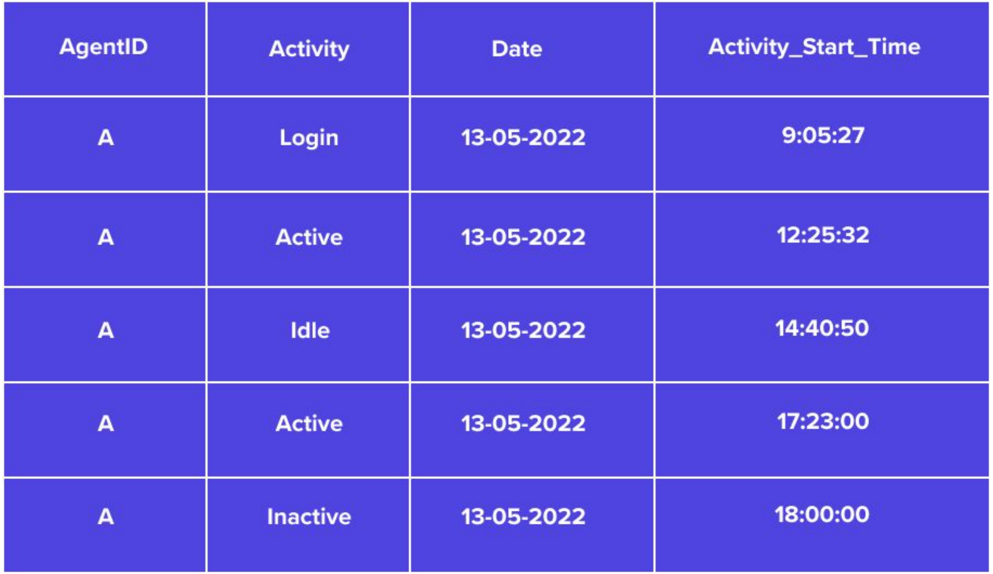
Solution:

{Concepts used: case statements}

```
Select * , case when Activity = ‘Login’ and Activity_Start_Time <= 9:00:00 then ‘YES’
case when Activity = ‘Login’ and Activity_Start_Time > 9:00:00 then ‘NO’
else null end as flag from Agent
```

Case Study - 5

Question 5. It was observed that people logged in on time in the month of may but had average logging hours to be lesser than 9.Write a query to find out number of distinct agent for each day in the month of May who logged out before 18:00:00



AgentID	Activity	Date	Activity_Start_Time
A	Login	13-05-2022	9:05:27
A	Active	13-05-2022	12:25:32
A	Idle	13-05-2022	14:40:50
A	Active	13-05-2022	17:23:00
A	Inactive	13-05-2022	18:00:00

Case Study - 5

Solution:

{Concepts used: distinct function}

```
Select distinct(AgentID) , date
 from Agent
where Activity = ‘Logout’
And Activity_Start_Timestamp < ’18:00:00’
And month(Date) = 5
Group by date
```

THANK YOU