SQL MAANG Questionnaire

1) **[Medium] Given the table R, compute the correlation coefficient of X1 and X2 columns.**

X1	X2
1	34
2	34
3	4
10	5
…	…

Solution:

Given two variables, finding at what extent X1 and X2 are related/correlated to each other is correlation in simple language.

The formula is: Cov(X1, X2)/[Std(X1)*Std(X2)]

Here,

1. **Cov is Covariance = Avg(X1- $X1_{mu}$) * (X2 - $X2_{mu}$)**
2. **Var is Variance = Avg((X1 - $X1_{mu}$)**2)**
3. **Std is Standard Deviation = Sqrt(Avg(X1 - $X1_{mu}$)**2))**

Step 1: Calculate the mean

```
With mean as(
Select X1, X2,
Avg(X1) OVER() as mean_X1,
Avg(X2) OVER() as mean_X2,
FROM R);
```

Step 2: Calculate the variance

```
With mean as(
Select X1, X2,
Avg(X1) OVER() as mean_X1,
```

```
Avg(X2) OVER() as mean_X2,
FROM R
),
Variance as(
Select
Avg(POWER(X1 - mean_X1, 2)) as var_X1,
Avg(POWER(X2 - mean_X2, 2)) as var_X2
From mean
);
```

Step 3: Calculate the standard deviation

```
With mean as(
Select X1, X2,
Avg(X1) OVER() as mean_X1,
Avg(X2) OVER() as mean_X2,
FROM R
),
Variance as(
Select
Avg(POWER(X1 - mean_X1, 2) as var_X1,
Avg(POWER(X2 - mean_X2, 2) as var_X2
From mean
),
StdDev as(
Select
POWER(var_X1, 0.5) as std_X1,
POWER(var_X2, 0.5) as std_X2
From Variance
);
```

Step 4: Calculate the covariance

```
With mean as(
Select X1, X2,
Avg(X1) OVER() as mean_X1,
Avg(X2) OVER() as mean_X2,
FROM R
),
Variance as(
Select
Avg(POWER(X1 - mean_X1, 2) as var_X1,
Avg(POWER(X2 - mean_X2, 2) as var_X2
From mean
),
StdDev as(
Select
POWER(var_X1, 0.5) as std_X1,
```

```
POWER(var_X2, 0.5) as std_X2
From Variance
),
Covariance as(
Select
AVG((X1 - mean_X1)*(X2 - mean_X2)) as cov_X1_X2
From mean
);
```

Step 4: Calculate the correlation coefficient

```
With mean as(
Select X1, X2,
Avg(X1) OVER() as mean_X1,
Avg(X2) OVER() as mean_X2,
FROM R
),
Variance as(
Select
Avg(POWER(X1 - mean_X1, 2) as var_X1,
Avg(POWER(X2 - mean_X2, 2) as var_X2
From mean
),
StdDev as(
Select
POWER(var_X1, 0.5) as std_X1,
POWER(var_X2, 0.5) as std_X2
From Variance
),
Covariance as(
Select
AVG((X1 - mean_X1)*(X2 - mean_X2)) as cov_X1_X2
From mean
)
Select
cov_X1_X2 / (std_X1 * std_X2) as corr_X1_X2
From Covariance, StdDev;
```

Q: [Medium]For the below given relations:

google_gmail_emails(
id int
from_user varchar
to_user varchar
day int
)

google_gmail_labels(
email_id int
label varchar
)
Find the number of emails received by each user under each built-in email label. The email labels are:

1. **Promotion**
2. **Social**
3. **Shopping**

Output the user along with the number of promotion, social, and shopping mails count.

Q: **[Easy]** Find the total costs of each customer's orders. Output the customer's id, first name, and the total order cost. Order records by customer's first name are alphabetical.

The relation schema given is
customers(
id int,
first_name varchar,
last_name varchar,
city varchar,
address varchar,
phone_number varchar
)

orders(
id int,
cust_id int,
order_date datetime,
order_details varchar,
total_order_cost int
)

Q: [Hard] You have a table of in-app purchases by user. Users that make their first in-app purchase are placed in a marketing campaign where they see call-to-actions for more in-app purchases. Find the number of users that made additional in-app purchases due to the marketing campaign's success.

The marketing campaign doesn't start until one day after the initial in-app purchase so users that only made one or multiple purchases on the first day do not count, nor do we count users that over time purchase only the products they purchased on the first day.
The relation given below is:
marketing_campaign(
user_id int,
created_at datetime,
product_id int,
quantity int,
price int
);