

CASE STUDY #3

A. Customer Journey

Based on the 8 sample customers provided in the sample from the `subscriptions` table, write a brief description about each customer's onboarding journey.

```
select customer_id, plan_id, plan_name, start_date  
from subscriptions  
join plans using (plan_id)  
where customer_id in(1,2,11,13,15,16,18,19);
```

	customer_id integer	plan_id integer	plan_name character varying (13)	start_date date
1	1	0	trial	2020-08-01
2	1	1	basic monthly	2020-08-08
3	2	0	trial	2020-09-20
4	2	3	pro annual	2020-09-27
5	11	0	trial	2020-11-19
6	11	4	churn	2020-11-26
7	13	0	trial	2020-12-15
8	13	1	basic monthly	2020-12-22
9	13	2	pro monthly	2021-03-29
10	15	0	trial	2020-03-17
11	15	2	pro monthly	2020-03-24
12	15	4	churn	2020-04-29
13	16	0	trial	2020-05-31
14	16	1	basic monthly	2020-06-07
15	16	3	pro annual	2020-10-21
16	18	0	trial	2020-07-06
17	18	2	pro monthly	2020-07-13
18	19	0	trial	2020-06-22
19	19	2	pro monthly	2020-06-29
20	19	3	pro annual	2020-08-29

B. Data Analysis Questions

–1) How many customers has Foodie-Fi ever had?

```
SELECT COUNT(DISTINCT customer_id) AS total_customers  
FROM subscriptions;
```

	total_customers	bigint
1	1000	

– 2) What is the monthly distribution of trial plan start_date values for our dataset - use the start of the month as the group by value

```
SELECT date_part('month',start_date), count(customer_id) as no_of_customers from  
subscriptions  
join plans using (plan_id)  
where plan_id = 0  
group by date_part('month',start_date)  
order by date_part('month',start_date)
```

	date_part	no_of_customers
1	1	88
2	2	68
3	3	94
4	4	81
5	5	88
6	6	79
7	7	89
8	8	88
9	9	87
10	10	79
11	11	75
12	12	84

-- 3 What plan `start_date` values occur after the year 2020 for our dataset?
Show the breakdown by count of events for each `plan_name`

```
SELECT plan_id, plan_name,
       COUNT(customer_id) AS num_of_event
  FROM subscriptions
 JOIN plans using (plan_id)
 WHERE start_date >= '2021-01-01'
 GROUP BY plan_id, plan_name
 ORDER BY plan_id;
```

	plan_id	plan_name	num_of_event
	integer	character varying (13)	bigint
1	1	basic monthly	8
2	2	pro monthly	60
3	3	pro annual	63
4	4	churn	71

- 4) What is the customer count and percentage of customers who have churned rounded to 1 decimal place?

```
select
count(*) as total_count,
round(count(customer_id)*100.0/(select count(*) from subscriptions),1 )as percentage
from subscriptions
where plan_id=4
```

	total_count	percentage
	bigint	numeric
1	307	11.6

-5) How many customers have churned straight after their initial free trial - what percentage is this rounded to the nearest whole number?

```
with rank_plan as(
    select customer_id, plan_id, plan_name, start_date,
    lead(plan_name) over (
        partition by customer_id
        order by start_date
    ) as next_plan
    from subscriptions join plans using (plan_id)
)

select count(*) as churn_after_trial,
round(100 * count(customer_id)/(select count(DISTINCT customer_id) from subscriptions),0) as
percentage
from rank_plan
where plan_name ='trial'
and next_plan='churn'
```

	churn_after_trial	percentage
1	92	9

-6) What is the number and percentage of customer plans after their initial free trial?

```
with rank_plan as(
    select customer_id, plan_id, plan_name, start_date,
    lead(plan_name) over (
        partition by customer_id
        order by start_date
    ) as next_plan
    from subscriptions join plans using (plan_id)
)

select count(*) as churn_after_trial,
round(100 * count(distinct customer_id)/(select count(DISTINCT customer_id) from
subscriptions),1) as percentage
from rank_plan
```

```
where next_plan is not null and plan_name ='trial'  
group by next_plan
```

	churn_after_trial bigint	percentage numeric
1	546	54.0
2	92	9.0
3	37	3.0
4	325	32.0

-- 7) What is the customer count and percentage breakdown of all 5 **plan_name** values at **2020-12-31**?

```
select plan_name,count(distinct customer_id) as total_customers,  
round(100 * count(distinct customer_id)/(select count(*) from subscriptions),1) as percentage  
from subscriptions join plans using (plan_id)  
where start_date <='2020-12-31'  
group by plan_name
```

	plan_name character varying (13)	total_customers bigint	percentage numeric
1	basic monthly	538	20.0
2	churn	236	8.0
3	pro annual	195	7.0
4	pro monthly	479	18.0
5	trial	1000	37.0

-- 8) How many customers have upgraded to an annual plan in 2020?

```
with rank_plan as(
    select customer_id, plan_id, plan_name, start_date,
    lead(plan_name) over (
        partition by customer_id
        order by start_date
    ) as next_plan
    from subscriptions join plans using (plan_id)
)

select count(*) as churn_after_trial,
round(100 * count(distinct customer_id)/(select count(DISTINCT customer_id) from
subscriptions),1) as percentage
from rank_plan
where next_plan = 'pro annual'
group by next_plan
```

	churn_after_trial	percentage
1	258	25.0

-- 9) How many days on average does it take for a customer to make an annual plan from the day they join Foodie-Fi?

```
with trial as(
    select customer_id, start_date as trial_date from subscriptions
    where plan_id = 0),
annual as(
    select customer_id, start_date as annual_date from subscriptions
    where plan_id = 3 )
select round(avg(annual_date - trial_date))
as avg_time from trial join annual using (customer_id)
```

	avg_time
1	105

– 10) Can you further breakdown this average value into 30 day periods (i.e. 0-30 days, 31-60 days etc)

WITH annual AS (

```
SELECT DISTINCT customer_id, start_date
FROM subscriptions
INNER JOIN plans USING(plan_id)
WHERE plan_name LIKE '%annual'
),
```

trial AS(

```
SELECT DISTINCT customer_id, start_date
FROM subscriptions
INNER JOIN plans USING(plan_id)
WHERE plan_name = 'trial'
```

),

diff AS (

```
SELECT customer_id, (a.start_date - t.start_date) date_diff
FROM annual a
INNER JOIN trial t USING(customer_id)
)
```

-- width_bucket(date_diff, MIN(date_diff), MAX(date_diff), MAX(date_diff)/30)

```
SELECT (date_diff / 30 + 1) thirty_days_period_number, COUNT(*) no_of_customers
FROM diff
GROUP BY thirty_days_period_number
```

	thirty_days_period_number	no_of_customers
1	1	48
2	2	25
3	3	33
4	4	35
5	5	43
6	6	35
7	7	27
8	8	4
9	9	5
10	10	1
11	11	1
12	12	1

-- 11) How many customers downgraded from a pro monthly to a basic monthly plan in 2020?

```
with rank_plan as(
    select customer_id,plan_id,plan_name, start_date,
    lag(plan_name) over (
        partition by customer_id
        order by start_date
    ) as last_plan
    from subscriptions join plans using (plan_id)
)

select count(*) as downgraded,
round(100 * count(distinct customer_id)/(select count(DISTINCT customer_id) from
subscriptions),1) as percentage
from rank_plan
where plan_id=1 and last_plan = 'pro monthly'
```

	downgrade_pro_to_basic
1	0