```
--num_orders by distinct customer ID desc
select
	distinct customer_id,
	count(distinct delivery_id) num_orders
from
Skillful_Data.Project_Data
group by customer_id
order by 2 desc;

--total # of distinct users -> 3,192
select
	count(distinct customer_id) total_customers,
from Skillful_Data.Project_Data;

--total # distinct couriers -> 578
select
	count (distinct Courier_id)
from Skillful_Data.Project_Data;

--num trips by distinct courier desc
select
	distinct Courier_id,
	count (distinct delivery_id) num_deliveries
from Skillful_Data.Project_Data
group by 1
order by 2 desc;

--span of dataset in days -> 29 days
Select
	timestamp_diff((max(when_the_delivery_started)),(min(when_the_delivery_started)),day
) data_span_days
from Skillful_Data.Project_Data;

--top performing nyc zones by revenue desc
select
	zone_name,
	count(*) as Num_Orders,
	sum(Total) as Total_Rev,
	sum(Total)/count(*) Avg_Order
FROM Skillful_Data.Project_Data
JOIN Skillful_Data.Zone_Information ON cast(Zone_Information.zone_id as int64) =
cast(Project_Data.Pickup_Zone_ID as int64)
GROUP BY zone_name
order by total_rev desc;
```

```
--variety of cuisine breakdown by num_orders desc
select
distinct place_category,
        count(distinct delivery_id) as Num_Orders,
        avg(Total) avg_order_total,
        avg(item_quantity) avg_item_quantity
FROM Skillful_Data.Project_Data
group by place_category
order by num_orders desc;

--avg trip duration -> 46.08 min [does this take into account cancellations?]
select
        avg(timestamp_diff(when_the_Courier_arrived_at_dropoff,when_the_delivery_started,
minute)) avg_trip_duration,
from Skillful_Data.Project_Data;

--num_orders by zone
select
        zd.zone_name,
        count(distinct pd.delivery_id) num_orders,
        avg(pd.total) avg_rev
from Skillful_Data.Project_Data pd
left join Skillful_Data.Zone_Information zd on cast(zd.zone_id as int64) = pd.Pickup_Zone_ID
group by 1
order by 2 desc;

--trip duration (ascending), delivery time, and total rating
select
        distinct pd.delivery_id,
        timestamp_diff(pd.when_the_Courier_arrived_at_dropoff,pd.when_the_delivery_started,
        minute) trip_duration,
        dr.delivery_time,
        Dr.total_rating
from Skillful_Data.Project_Data pd
right join Skillful_Data.Ratings dr on dr.delivery_id = pd.delivery_id
where total_rating is not null
group by 1,2,3,4
order by 2 asc;

--avg delivery time, avg delivery rating, avg total rating, num orders by customer (desc)
[could be used in customer retention analysis]
select
        distinct pd.customer_id,
```

```
        avg(timestamp_diff(pd.when_the_Courier_arrived_at_dropoff,pd.when_the_delivery_star
        ted,minute)) avg_deliv_time,
        avg(dr.delivery_time) avg_deliv_time_rating,
        avg(dr.total_rating) avg_total_rating,
        count(distinct pd.delivery_id) customer_num_orders
from Skillful_Data.Project_Data pd
right join Skillful_Data.Ratings dr on dr.delivery_id = pd.delivery_id
where total_rating is not null
group by 1
order by 5 desc;

-- ARPU by distinct customer ID
select
        distinct customer_id,
        count(distinct delivery_id) num_orders,
        sum(Total)/(count(distinct delivery_id)) avg_rev_user
from
Skillful_Data.Project_Data
group by 1
order by 2 desc;

--avg orders per customer -> 1.63
select
count(distinct delivery_id)/count(distinct customer_id)
from Skillful_Data.Project_Data;

--ordering experience rating count [for a histogram]
select
        distinct Ordering_Experince,
        count(Ordering_Experince)
from Skillful_Data.Ratings
where Ordering_Experince is not null
group by 1;

--delivery time rating count [for a histogram]
Select
        distinct Delivery_Time,
        count(Delivery_Time)
from Skillful_Data.Ratings
where delivery_time is not null
group by 1;
```