SQL Queries for Rapido Pune Data Analysis

Table Creation:

```sql
create database Rapido_Pune;
use Rapido_Pune;

CREATE TABLE rides (
    Date DATE NOT NULL,
    Time TIME NOT NULL,
    Booking_ID VARCHAR(15) PRIMARY KEY,
    Booking_Status VARCHAR(50) NOT NULL,
    Customer_ID VARCHAR(15) NOT NULL,
    Vehicle_Type VARCHAR(20) NOT NULL,
    Pickup_Location VARCHAR(50) NOT NULL,
    Drop_Location VARCHAR(50) NOT NULL,
    Avg_VTAT DECIMAL(5, 2) DEFAULT 0, -- Average Vehicle Time-to-Arrival
    Avg_CTAT DECIMAL(5, 2) DEFAULT 0, -- Average Customer Time-to-Arrival
    Cancelled_Rides_by_Customer INT DEFAULT 0,
    Reason_for_Cancelling_by_Customer VARCHAR(100) DEFAULT 'Not Applicable',
    Cancelled_Rides_by_Driver INT DEFAULT 0,
    Reason_for_Cancelling_by_Driver VARCHAR(100) DEFAULT 'Not Applicable',
    Incomplete_Rides INT DEFAULT 0,
    Incomplete_Rides_Reason VARCHAR(100) DEFAULT 'Not Applicable',
    Booking_Value DECIMAL(10, 2) DEFAULT 0,
    Ride_Distance DECIMAL(6, 2) DEFAULT 0,
    Driver_Ratings DECIMAL(3, 2) DEFAULT 0,
    Customer_Rating DECIMAL(3, 2) DEFAULT 0,
    Payment_Method VARCHAR(20) NOT NULL
);
```

Dataset Snippet:

Date	Time	Booking_ID	Booking_Status	Customer_ID	Vehicle_Type	Pickup_Location	Drop_Location	Avg_VTAT	Avg_CTAT	Cancelled_Ride
2024-01-19	13:38:39	CNR1000021256	Complete	CUST98377	eBike	Mundhwa	Nigdi	5.00	7.00	0
2024-01-10	19:07:44	CNR1000042329	Complete	CUST49643	Auto	Kalyani Nagar	Kondhwa	5.00	6.00	0
2024-01-23	20:19:15	CNR1000166244	Complete	CUST49630	Auto	Baner	Uruli Kanchan	3.00	10.00	0
2024-01-21	02:51:41	CNR1000221845	Cancelled by Driver	CUST12703	Mini	Sus	Chandni Chowk	0.00	0.00	0

SQL Questions:

1. How many bookings were made for each vehicle type each day?

```sql
-- 1. How many bookings were made for each vehicle type each day?
create view number_of_bookings_each_day as
select date(Date) as Date, Vehicle_Type, count(Booking_ID) as Number_of_Bookings
from rides
group by Vehicle_Type, date(Date);

select * from number_of_bookings_each_day;
```

Answer:

Date	Vehicle_Type	Number_of_Bookings
2024-01-19	eBike	178
2024-01-10	Auto	675
2024-01-23	Auto	676
2024-01-21	Mini	668
2024-01-02	Prime Sedan	478
2024-01-22	eBike	160
2024-01-25	Auto	706
2024-01-16	Bike	488
2024-01-12	Auto	632
2024-01-17	Prime Plus	339
2024-01-18	Mini	658
2024-01-20	Auto	686
2024-01-26	Prime Sedan	512
2024-01-24	Prime Sedan	490
2024-01-04	Mini	694
2024-01-07	Prime Plus	328
2024-01-28	Bike	487
2024-01-10	Prime Plus	340
2024-01-07	Prime Sedan	512
2024-01-01	Bike	498
2024-01-20	Bike	477
2024-01-03	Mini	657
2024-01-11	Prime Sedan	510

2. Which drop locations are most frequently chosen?

```sql
-- 2. Which drop locations are most frequently chosen?
create view frequent_drop_locations as
select Drop_Location, count(*) as Frequency
from rides
group by Drop_Location
order by Frequency desc
limit 5;

select * from frequent_drop_locations;
```

Answer:

Drop_Location	Frequency
Nigdi	3893
Somwar Peth	2092
Deccan	2081
Uruli Kanchan	2077
Swargate	2074

3. What is the total number of completed rides?

```sql
-- 3. What is the total number of completed rides?
create view total_number_of_completed_rides as
select count(Booking_ID) as Total_Completed_Rides
from rides
where Booking_Status = 'Complete';

select * from total_number_of_completed_rides;
```

Answer:

Total_Completed_Rides
62115

4. What is the average fare amount for each vehicle type?

```
-- 4. What is the average fare amount for each vehicle type?
create view average_fare_amount_for_each_vehicle as
select Vehicle_Type, avg(Booking_Value) as Average_Fare_Amount
from rides
group by Vehicle_Type;

select * from average_fare_amount_for_each_vehicle;
```

Answer:

Vehicle_Type	Average_Fare_Amount
eBike	185.075524
Auto	185.527571
Mini	185.593690
Prime Sedan	480.503290
Bike	186.875359
Prime Plus	485.563963
Prime SUV	483.620399

5. How many rides were completed, canceled, and ongoing for each day?

```
-- 5. How many rides were completed, cancelled, and incomplete for each day?
create view rides_completed_canceled_ongoing as
select date(Date) AS date,
    sum(case when Booking_Status = 'Complete' then 1 else 0 end) as Completed_Rides,
    sum(case when Booking_Status IN ('Cancelled by customer', 'Cancelled by Driver') then 1 else 0 end) as Cancelled_Rides,
    sum(case when Booking_Status = 'Driver not found' then 1 else 0 end) as Incomplete_Rides
from rides
group by date(Date)
order by date(Date);

select * from rides_completed_canceled_ongoing;
```

Answer:

date	Completed_Rides	Cancelled_Rides	Incomplete_Rides
2024-01-01	2080	854	423
2024-01-02	2107	784	435
2024-01-03	2055	826	438
2024-01-04	2039	849	410
2024-01-05	2023	850	434
2024-01-06	2038	873	432
2024-01-07	2117	841	467
2024-01-08	2003	788	405
2024-01-09	2026	832	405

6. Which vehicle type has the highest cancellation rate?

```
-- 6. Which vehicle type has the highest cancellation rate?
create view highest_cancellation_rate_vehicle as
select Vehicle_Type,
sum(case when Booking_Status IN ('Cancelled by customer', 'Cancelled by Driver') then 1 else 0 end) / count(*) as Cancellation_Rate
from rides
group by Vehicle_Type
order by Cancellation_Rate desc
limit 1;

select * from highest_cancellation_rate_vehicle;
```

Answer:

	Vehicle_Type	Cancellation_Rate
▶	eBike	0.2535

7. What is the average ride distance for completed rides?

```
-- 7. What is the average ride distance for completed rides?
create view average_ride_distance as
select avg(Ride_Distance) as Avg_Ride_Distance
from rides
where Booking_Status = 'Complete';

select * from average_ride_distance;
```

Answer:

	Avg_Ride_Distance
▶	11.697220

8. How many bookings were made during peak hours (7 AM - 10 AM and 5 PM - 8 PM)?

```
-- 8. How many bookings were made during peak hours (7 AM - 10 AM and 5 PM - 8 PM)?
create view number_of_bookings_in_peak_hours as
select count(*) as Peak_Hour_Bookings
from rides
where (hour(Time) between 7 and 9) or (hour(Time) between 17 and 19);

select * from number_of_bookings_in_peak_hours;
```

	Peak_Hour_Bookings
▶	25337

9. What is the average fare amount for each pickup location?

```
-- 9. What is the average fare amount for each pickup location?
create view average_fare_amount_for_pickup as
select Pickup_Location, avg(Booking_Value) as average_fare
from rides
group by Pickup_Location;

select * from average_fare_amount_for_pickup;
```

Answer:

Pickup_Location	average_fare
Mundhwa	305.413605
Kalyani Nagar	298.986528
Baner	290.443452
Sus	299.945254
Bibvewadi	318.722166
Kothrud	316.884482
Fursungi	306.850323
Shivajinagar	311.301217
Parvati	293.382691

10. Which pickup locations have the highest cancellation rates?

```
-- 10. Which pickup locations have the highest cancellation rates?
create view pickup_location_with_highest_cancellation_rates as
select Pickup_Location, sum(case when Cancelled_Rides_by_Customer = 1 then 1 else 0 end) as cancellation_rates
from rides
group by Pickup_Location
order by cancellation_rates desc
limit 5;

select * from pickup_location_with_highest_cancellation_rates;
```

Answer:

Pickup_Location	cancellation_rates
Nigdi	282
Pashan	168
Warje	159
Wakad	159
Vadgaon Sheri	159

11. What is the average distance traveled for each vehicle type?

```
-- 11. What is the average distance traveled for each vehicle type?
create view average_distance_travelled as
select Vehicle_Type, avg(Ride_Distance) as average_distance
from rides
group by Vehicle_Type;

select * from average_distance_travelled;
```

Answer:

Vehicle_Type	average_distance
eBike	5.243871
Auto	5.247730
Mini	5.268521
Prime Sedan	10.176703
Bike	5.305166
Prime Plus	10.336229
Prime SUV	10.294492

12. What is the average revenue generated per day?

```
-- 12. What is the average revenue generated per day?
create view average_revenue_per_day as
select date(Date) as Date, avg(Booking_Value) as Average_Revenue_Per_Day
from rides
group by date(Date)
order by date(Date);

select * from average_revenue_per_day;
```

Answer:

Date	Average_Revenue_Per_Day
2024-01-01	305.353590
2024-01-02	307.173782
2024-01-03	305.590841
2024-01-04	304.363857
2024-01-05	299.161778

13. What is the average CTAT and VTAT for rides canceled by customers?

```
-- 14. What is the average CTAT and VTAT for rides canceled by customers?
create view average_ctat_vtat_canceled_rides as
select avg(Avg_CTAT) as Average_CTAT, avg(Avg_VTAT) as Average_VTAT
from rides
where
    Booking_Status In ('Cancelled by customer', 'Cancelled by Driver');

select * from average_ctat_vtat_canceled_rides;
```

Answer:

	Average_CTAT	Average_VTAT
▶	0.000000	0.000000

14. Which drop locations have the highest customer ratings?

```
-- 15. Which drop locations have the highest customer ratings?
create view drop_locations_highest_customer_ratings as
select Drop_Location, sum(case when Customer_Rating > 3 then 1 else 0 end) as highest_cust_rating
from rides
group by Drop_Location
order by highest_cust_rating desc
limit 5;

select * from drop_locations_highest_customer_ratings;
```

Answer:

	Drop_Location	highest_cust_rating
▶	Nigdi	1507
	Fursungi	825
	Dhole Patil Road	817
	Deccan	796
	Uruli Kanchan	794

15. How does the number of canceled rides vary across vehicle types?

```
-- 16. How does the number of canceled rides vary across vehicle types?
create view number_of_canceled_rides as
select Vehicle_Type, sum(case when Booking_Status IN ('Cancelled by customer', 'Cancelled by Driver') then 1 else 0 end)
as canceled_rides
from rides
group by Vehicle_Type;

select * from number_of_canceled_rides;
```

Answer:

	Vehicle_Type	canceled_rides
▶	eBike	1282
	Auto	5103
	Mini	5015
	Prime Sedan	3801
	Bike	3724
	Prime Plus	2427
	Prime SUV	3693

17. What is the correlation between ride distance and fare amount across vehicle types?

```
-- 17. What is the correlation between ride distance and fare amount across vehicle types?
create view correlations as
select (avg(Ride_Distance * Booking_Value) - avg(Ride_Distance) * avg(Booking_Value)) /
(sqrt(avg(Ride_Distance * Ride_Distance) - avg(Ride_Distance) * avg(Ride_Distance)) *
(sqrt(avg(Booking_Value * Booking_Value) - avg(Booking_Value) * avg(Booking_Value)))) as correlation
from rides;

select * from correlations;
```

Answer:

	correlation
▶	0.808935458868383

18. How does the average VTAT and CTAT vary by time of day?

```
-- 18. How does the average VTAT and CTAT vary by time of day?
create view average_vtat_ctat_vary_by_time_of_day as
select hour(Time) as Time_of_day, avg(Avg_VTAT) as average_vtat, avg(Avg_CTAT) as average_ctat
from rides
group by Time_of_day
order by Time_of_day;

select * from average_vtat_ctat_vary_by_time_of_day;
```

Answer:

Time_of_day	average_vtat	average_ctat
0	4.091712	6.264867
1	4.050249	6.222637
2	4.036690	6.228023
3	4.012533	6.207614
4	3.971177	6.059355
5	3.996844	6.133042
6	4.005383	6.032298

19. Which pickup and drop location pairs generate the highest revenue?

```
-- 19. Which pickup and drop location pairs generate the highest revenue?
create view pickup_and_drop_location_pairs as
select Pickup_Location, Drop_Location, sum(Booking_Value) as revenue
from rides
group by Pickup_Location, Drop_Location
order by revenue desc
limit 5;
```

Answer:

Pickup_Location	Drop_Location	revenue
Kharadi	Nigdi	36018.00
Kothrud	Nigdi	34379.00
Nigdi	Sadashiv Peth	33526.00
Shivajinagar	Nigdi	32973.00
Nigdi	Kalyani Nagar	32788.00

20. What is the cancellation rate for rides with high VTAT or CTAT?

```
-- 20. What is the cancellation rate for rides with high VTAT or CTAT?
create view cancellation_rate_for_rides_with_high_vtat_or_ctat as
select sum(case when Booking_Status IN ('Cancelled by customer', 'Cancelled by Driver') then 1 else 0 end) / count(*)
as cancellation_rate,  avg(Avg_VTAT) as avg_vtat,
avg(Avg_CTAT) as avg_ctat
from rides
where Avg_VTAT > 10 or  Avg_CTAT > 10;

select * from cancellation_rate_for_rides_with_high_vtat_or_ctat;
```

Answer:

cancellation_rate	avg_vtat	avg_ctat
0.0000	6.482566	13.004367

21. How does the average driver rating vary with the number of rides completed?

```
-- 21. How does the average driver rating vary with the number of rides completed?
create view driver_rating_vary as
select avg(Driver_Ratings) as driver_rating, count(Booking_ID) as completed_rides
from rides
where Booking_Status = 'Complete'
group by Booking_Status;

select * from driver_rating_vary;
```

Answer:

driver_rating	completed_rides
3.005538	62115

22. What is the average fare amount for rides canceled due to "Driver is not moving towards pickup location"?

```sql
-- 22. What is the average fare amount for rides canceled due to "Driver is not moving towards pickup location"?
create view average_fare_amount_for_rides_canceled as
select avg(Booking_Value) as average_amount, count(Booking_ID) as Total_cancelled_rides
from rides
where Reason_for_Cancelling_by_Customer = 'Driver is not moving towards pickup location';

select * from average_fare_amount_for_rides_canceled;
```

	average_amount	Total_cancelled_rides
▶	0.000000	1706

23. How do peak hours impact the average CTAT and VTAT?

```sql
-- 23. How do peak hours impact the average CTAT and VTAT?
select time_format(Time, '%H') as Hour, avg(Avg_CTAT) as average_ctat,
avg(Avg_VTAT) as average_vtat
from rides
group by time_format(Time, '%H')
order by average_ctat desc, average_vtat desc;
```

Answer:

	Hour	average_ctat	average_vtat
▶	16	6.392349	4.113833
	09	6.329203	4.121814
	15	6.284675	4.022066
	22	6.278817	4.060291
	18	6.274556	4.048877
	14	6.267496	4.095830
	00	6.264867	4.091712

24. Which vehicle type generates the highest revenue per kilometer?

```sql
-- 24. Which vehicle type generates the highest revenue per kilometer?
create view vehicle_type_generates_the_highest_revenue as
select Vehicle_Type, sum(Booking_Value / Ride_Distance) as revenue_per_km
from rides
group by Vehicle_Type
order by revenue_per_km desc
limit 1;

select * from vehicle_type_generates_the_highest_revenue;
```

Answer:

	Vehicle_Type	revenue_per_km
▶	Auto	584294.247825

25. What are the most common cancellation reasons for specific vehicle types?

```
-- 25. What are the most common cancellation reasons for specific vehicle types?
create view most_common_cancellation_reason as
select Vehicle_Type, Reason_for_Cancelling_by_Customer, count(*) as cancellation_count
from rides
where Reason_for_Cancelling_by_Customer != 'Not Applicable'
group by Vehicle_Type, Reason_for_Cancelling_by_Customer
order by count(Reason_for_Cancelling_by_Customer) desc
limit 5;
```

Answer:

	Vehicle_Type	Reason_for_Cancelling_by_Customer	cancellation_count
▶	Mini	Wrong Address	479
	Auto	Wrong Address	435
	Auto	Driver is not moving towards pickup location	363
	Mini	Driver is not moving towards pickup location	348
	Prime Sedan	Wrong Address	326

26. How does customer rating vary with ride distance and fare amount?

```sql
-- 26. How does customer rating vary with ride distance and fare amount?
create view customer_rating_vary_with_ride_distance_fare_amount as
select avg(Customer_Rating) as average_customer_rating,  count(*) as Ride_Count,
case when Ride_Distance <= 5 then '0-5 KM'
     when Ride_Distance > 5 and Ride_Distance <= 10 then '5-10 KM'
     when Ride_Distance > 10 and Ride_Distance <= 20 then '10-20 KM'
     when Ride_Distance > 20 then '20+ KM'
     end as Distance_Range,
case  when Booking_Value <= 100 then '0-100 INR'
     when Booking_Value > 100 and Booking_Value <= 200 then '100-200 INR'
     when Booking_Value > 200 and Booking_Value <= 400 then '200-400 INR'
     when Booking_Value > 400 then '400+ INR'
     end as Fare_Range
from rides
where Customer_Rating is not null
group by Distance_Range, Fare_Range
order by Distance_Range, Fare_Range;
```

Answer:

average_customer_rating	Ride_Count	Distance_Range	Fare_Range
0.001952	37905	0-5 KM	0-100 INR
3.619294	2125	0-5 KM	100-200 INR
3.551835	4360	0-5 KM	200-400 INR
3.621178	2191	0-5 KM	400+ INR
3.615385	39	10-20 KM	0-100 INR
3.572243	3509	10-20 KM	100-200 INR
3.576918	7027	10-20 KM	200-400 INR

27. What percentage of total bookings come from the top 5 pickup locations?

```sql
-- 27. What percentage of total bookings come from the top 5 pickup locations?
create view percentage_of_total_bookings_from_top_pickups as
select (Booking_count  / total_bookings) * 100 as percentage_total_bookings, Pickup_Location
from (select count(Booking_ID) as Booking_count, Pickup_Location, (select count(*) from rides) as total_bookings
from rides
group by Pickup_Location
) as subquery
order by percentage_total_bookings desc
limit 5;

select * from percentage_of_total_bookings_from_top_pickups;
```

Answer:

percentage_total_bookings	Pickup_Location
4.0080	Nigdi
2.1250	Wakad
2.0810	Deccan
2.0660	Balewadi
2.0580	Lavale

28. How does ride distance vary by time of day for different vehicle types?

```
-- 28. How does ride distance vary by time of day for different vehicle types?
create view ride_distance_vary_by_time_of_day as
select hour(Time) as hour_of_day, Vehicle_Type, avg(Ride_Distance) as average_ride_dist
from rides
group by hour_of_day, Vehicle_Type
order by hour_of_day;

select * from ride_distance_vary_by_time_of_day;
```

hour_of_day	Vehicle_Type	average_ride_dist
4	Bike	5.179239
4	eBike	5.587678
4	Mini	4.929500
4	Prime Plus	10.422414
4	Prime Sedan	10.375439
4	Prime SUV	10.054984
5	Auto	4.866255