

# SQL PROJECT

## Airport Data Analysis



## PROBLEM STATEMENT-1

Which destination airport has the highest total number of passengers?

```
7 • select Destination_airport, Sum(Passengers) as Total_Passenger  
8   from airports2  
9   group by Destination_airport  
10  order by Total_Passenger DESC;  
11
```



Result Grid		Filter Rows:	Export:	Wrap Cell Content:
Destination_airport	Total_Passenger			
RDM	1430567			
AMW	21			

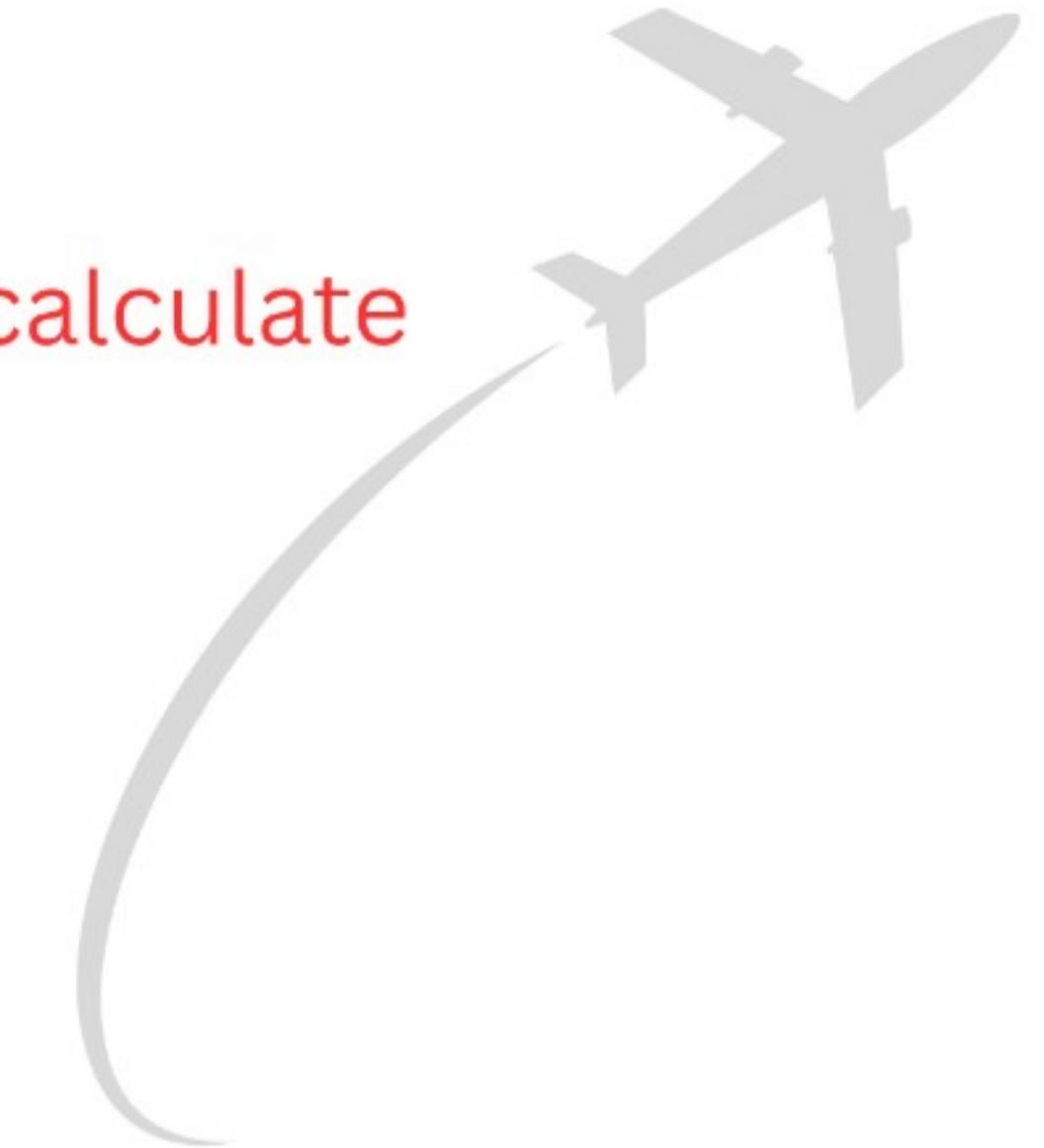
The query shows total passengers by destination airport, sorted in descending order.

- RDM had the highest traffic with 1,430,567 passengers.
- AMW had the lowest with only 21 passengers.

## PROBLEM STATEMENT-2

How can we analyze highest, lowest, and average seat occupancy and calculate seat utilization to improve flight operational efficiency?

```
27 • SELECT Origin_airport, Destination_airport,  
28     ROUND(AVG(CAST(Passengers AS FLOAT)/NULLIF(Seats,0))*100,2) AS Avg_Seat_Utilization  
29     FROM airports2  
30     GROUP BY  
31     Origin_airport, Destination_airport  
32     ORDER BY  
33     Avg_Seat_Utilization DESC  
34     LIMIT 3;
```



Result Grid		
Origin_airport	Destination_airport	Avg_Seat_Utilization
BIL	RDM	90
FAT	RDM	83.33
EKO	RDM	75.43

### Top Routes to RDM by Seat Utilization:

- BIL → RDM: 90%
- FAT → RDM: 83.33%
- EKO → RDM: 75.43%

## PROBLEM STATEMENT-3

Which are the Top 3 most frequently traveled routes based on total passenger volume?

```
44 •   SELECT Origin_airport, Destination_airport, SUM(Passengers) AS Total_Passengers  
45     FROM airports2  
46     GROUP BY Origin_airport, Destination_airport  
47     ORDER BY Total_Passengers DESC  
48     LIMIT 3;
```



Result Grid		
Origin_airport	Destination_airport	Total_Passengers
PDX	RDM	954459
SEA	RDM	313510
SFO	RDM	108356

### Most passengers traveling:

- Top Route to RDM: PDX → RDM (954,459 passengers)
- 2nd Route: SEA → RDM (313,510 passengers)
- 3rd Route: SFO → RDM (108,356 passengers)

## PROBLEM STATEMENT-4

Which origin cities have the highest flight activity and passenger traffic, and which ones can be considered as key hubs?

```
57 •  SELECT Origin_city, COUNT(Flights) as Total_Flight, SUM(Passengers) as Total_Passenger  
58  From airports2  
59  GROUP BY Origin_city  
60  ORDER BY Total_Passenger DESC  
61  LIMIT 3;
```

Result Grid | Filter Rows:  Export: Wrap Cell Content: Fetch rows:

Origin_city	Total_Flight	Total_Passenger
Portland, OR	509	954459
Seattle, WA	164	313542
San Francisco, CA	80	108356



### Key Hubs Based on Flight Activity & Passenger Traffic:

- Key Hubs: Portland (PDX), Seattle (SEA), San Francisco (SFO)
- They have the highest flight activity and passenger traffic.

## PROBLEM STATEMENT-5

Which airports cover the maximum total travel distance, and how can this help in future route planning?

```
68 • SELECT Origin_airport,Origin_city, SUM(Distance) AS Total_Distance  
69   FROM airports2  
70   GROUP BY Origin_airport, Origin_city|  
71   ORDER BY Total_Distance DESC  
72   LIMIT 5:
```

result Grid | Filter Rows:  Export: Wrap Cell Content: Fetch rows

Origin_airport	Origin_city	Total_Distance
PDX	Portland, OR	59044
SEA	Seattle, WA	37163
SFO	San Francisco, CA	36960
LMT	Klamath Falls, OR	14553
EKO	Elko, NV	8688

Airports with maximum total travel distance:

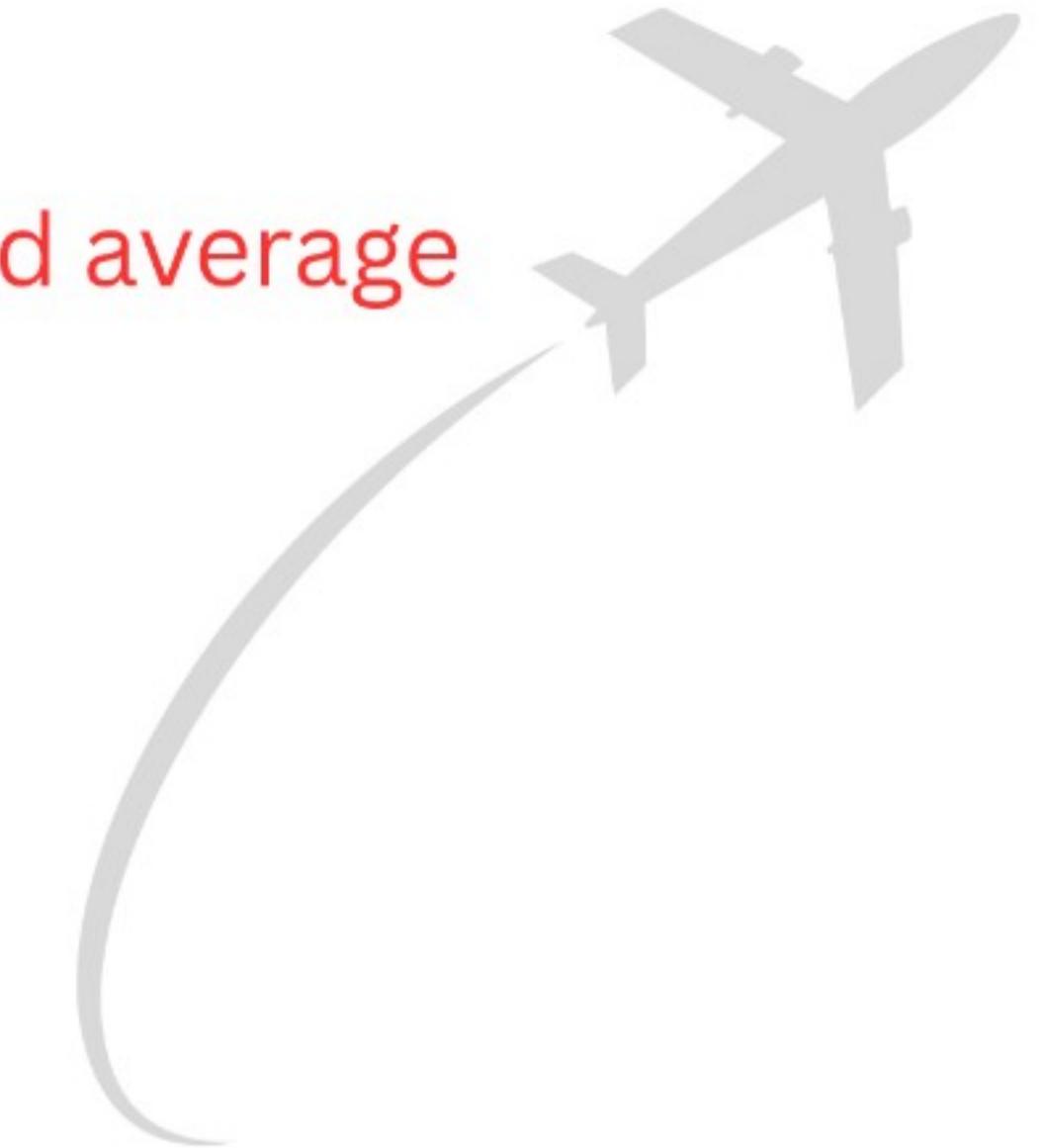
- Airports with maximum travel distance: Portland (PDX), Seattle (SEA), and San Francisco (SFO).
- Use in future planning: These airports can act as main hubs for adding new routes and more flights, while smaller airports can focus on local connections.



## PROBLEM STATEMENT-6

What are the seasonal travel trends in terms of flights, passengers, and average distance traveled per month?

```
81 •  SELECT YEAR(Fly_date) as Years, MONTH(Fly_date) as Months,  
82      COUNT(Flights) as Total_Flights, SUM(Passengers) as Total_Passengers,  
83      AVG(Distance) as Avg_Distance  
84      FROM airports2  
85      GROUP BY Years,Months  
86      ORDER BY Years DESC,Months DESC  
87      LIMIT 3;
```



Result Grid				
Years	Months	Total_Flights	Total_Passengers	Avg_Distance
2008	10	1	21	254.0000
2005	12	7	13243	224.0000
2005	11	7	12895	240.0000

### Seasonal Trends:

- In **Dec 2005**, passenger traffic was highest (13,243 passengers, 7 flights).
- **Nov 2005** also had strong traffic (12,895 passengers, 7 flights).
- **Oct 2008** shows very few flights (1 flight, 21 passengers) with a higher average distance.

## PROBLEM STATEMENT-7

Which routes are underutilized based on passenger-to-seat ratio (less than 50%)?



```
97 •  SELECT Origin_airport, Destination_airport,  
98      SUM(Passengers) as Total_Passengers,  
99      SUM(Seats) as Total_Seats,  
100     ROUND((SUM(Passengers)*1.0/NULLIF(SUM(Seats),0)),2) as Passenger_to_seat_ratio  
101    From airports2  
102    GROUP BY Origin_airport, Destination_airport  
103    HAVING Passenger_to_seat_ratio < 0.5  
104    ORDER BY Passenger_to_seat_ratio DESC  
105    limit 3;
```

Result Grid | Filter Rows:  Export: Wrap Cell Content: Fetch rows:

Origin_airport	Destination_airport	Total_Passengers	Total_Seats	Passenger_to_seat_ratio
LWS	RDM	35	73	0.48
YKM	RDM	25	55	0.45
GEG	RDM	137	304	0.45

### Underutilized Routes (Passenger-to-Seat ratio < 50%)

- LWS → RDM – 35 passengers out of 73 seats (48% filled)
- YKM → RDM – 25 passengers out of 55 seats (45% filled)
- GEG → RDM – 137 passengers out of 304 seats (45% filled)

## PROBLEM STATEMENT-8

Which airports have the highest flight frequency, indicating the most active airports?



```
113 •  SELECT Origin_airport,  
114      SUM(Flights) AS Total_flight  
115      FROM airports2  
116      GROUP BY Origin_airport  
117      ORDER BY Total_flight DESC  
118      LIMIT 3;
```

Result Grid | Filter Rows: \_\_\_\_\_ | Export: Wrap Cell Content: Fetch rows:

Origin_airport	Total_flight
PDX	55147
SEA	12919
SFO	5849

Airports with highest flight frequency:

- Portland (PDX) – 55,147 flights
- Seattle (SEA) – 12,919 flights
- San Francisco (SFO) – 5,849 flights

## PROBLEM STATEMENT-9

Identify the origin airports with the longest flight distances to analyze extensive travel connections and operational reach.

```
L37 •  SELECT Origin_airport, MAX(Distance) AS Max_distance  
L38      FROM airports2  
L39      GROUP BY Origin_airport  
L40      ORDER BY Max_distance DESC  
L41      LIMIT 3;
```

Result Grid | Filter Rows:  Export: Wrap Cell Content:

Origin_airport	Max_distance
FLL	2601
GGG	1643
DFW	1514

Origin airports with the longest flight distances:

- FLL (Fort Lauderdale) – 2,601 miles
- GGG (Longview, TX) – 1,643 miles
- DFW (Dallas/Fort Worth) – 1,514 miles



## PROBLEM STATEMENT-10

Identify the months with the highest and lowest flight frequencies to analyze seasonal travel trends over multiple years.

```
# USING WITH KEYWORD
-- Step 1: Creating a Common Table Expression (CTE) named 'Monthly_flights'
WITH Monthly_flights AS
(
  SELECT MONTH(Fly_date) AS Months,
  COUNT(Flights) AS Total_Flight
  FROM airports2
  GROUP BY Months
  ORDER BY Months DESC, Total_Flight DESC
)
-- Step 2: Selecting and labeling the most and least busy months
SELECT Months,
Total_Flight,
CASE
WHEN Total_Flight=(SELECT MAX(Total_Flight) FROM Monthly_flights) THEN 'Most Busy'
WHEN Total_Flight=(SELECT MIN(Total_Flight) FROM Monthly_flights) THEN 'Least Busy'
ELSE NULL
END AS Status
FROM Monthly_flights
WHERE
Total_Flight=(SELECT MAX(Total_Flight) FROM Monthly_flights) OR
Total_Flight=(SELECT MIN(Total_Flight) FROM Monthly_flights);
```

result Grid | Filter Rows:

Months	Total_Flight	Status
9	74	Least Busy
1	95	Most Busy



## PROBLEM STATEMENT-11

Analyze origin airports to identify those with the highest weighted passenger-to-seat utilization ratio, factoring in total flight operations, to evaluate operational efficiency and optimize flight performance.

```
287 • WITH Utilization_Ratio AS
288   (
289     SELECT Origin_airport, SUM(Passengers) as Total_passenger, SUM(Seats) as Total_seats,
290     COUNT(Flights) AS Total_flights, SUM(Passengers)*1.0/SUM(Seats) AS Passenger_seat_ratio
291     FROM airports2
292     GROUP BY Origin_airport
293     ORDER BY Total_seats DESC, Total_passenger DESC
294   )
295   SELECT
296     Origin_airport, Total_passenger, Total_seats, Total_flights, Passenger_seat_ratio,
297     (Passenger_seat_ratio*Total_flights)/SUM(Total_flights)
298   OVER () AS Weighted_Utilization
299   FROM Utilization_Ratio
300   LIMIT 3;
```

Result Grid					
Filter Rows: <input type="text"/> Export: <input type="button"/> Wrap Cell Content: <input type="checkbox"/>					
Origin_airport	Total_passenger	Total_seats	Total_flights	Passenger_seat_ratio	Weighted_Utilization
PDX	954459	1596892	509	0.59770	0.303925374
SEA	313510	477707	163	0.65628	0.106866773
SFO	108356	175454	80	0.61757	0.049356243



## PROBLEM STATEMENT-12

Identify underperforming origin-to-destination routes with a minimum of 3 flights and less than 50% seat utilization to optimize operational efficiency.

```
377 • with flight_stats as (SELECT
378     Origin_airport, Destination_airport,
379     Count(Flights) as total_flights,
380     Sum(Passengers) as Total_passengers,
381     Sum(Seats) as total_seats,
382     (Sum(Passengers)/nullif(Sum(Seats),0)) as Avg_seat_utilization
383
384     FROM
385     Airports2
386     group by
387     Origin_airport, Destination_airport)
388
389     Select Origin_airport, Destination_airport, total_flights, Total_passengers, total_seats,
390     Round((Avg_seat_utilization * 100),2) as Avg_seat_utilization_percentage
391     From
392     flight_stats
393     where
394     total_flights >= 10 and |
395     Round((Avg_seat_utilization * 100),2) < 50
396     order by Avg_seat_utilization_percentage
397     LIMIT 3;
```

Result Grid | Filter Rows: \_\_\_\_\_ | Export: \_\_\_\_\_ | Wrap Cell Content: \_\_\_\_\_

Origin_airport	Destination_airport	total_flights	Total_passengers	total_seats	Avg_seat_utilization_percentage
EUG	RDM	31	1300	6937	18.74
LMT	RDM	99	13317	54131	24.60
MFR	RDM	19	248	678	36.58



## PROBLEM STATEMENT-13

Identify origin-to-destination routes with the longest average flight distances to analyze long-haul travel patterns and support airline strategic planning

```
405 • with distance_stats as (select Origin_airport,      #Temporary Table  
406   Destination_airport,  
407   Round(Avg(Distance),2) as Avg_flight_distance  
408   from Airports2  
409   group by Origin_airport, Destination_airport)  
410  
411   Select Origin_airport,  
412   Destination_airport,  
413   round(Avg_flight_distance,2) as Avg_flight_distance  
414   from distance_stats  
415   order by Avg_flight_distance desc  
416   LIMIT 3;  
417  
418   # PROBLEM STATEMENTS - 19
```

Result Grid | Filter Rows: \_\_\_\_\_ | Export: Wrap Cell Content:

Origin_airport	Destination_airport	Avg_flight_distance
FLL	RDM	2601.00
GGG	RDM	1643.00
DFW	RDM	1514.00



## PROBLEM STATEMENT-14

Analyze annual trends in total flights and passenger traffic, including year-over-year growth percentages, to provide insights for strategic planning and performance evaluation.

```
427 • with Yearly_summary as (select Year(Fly_date) as Year,    #Temporary table  
428     count(Flights) as Total_flight,  
429     sum(Passengers) as Total_passengers  
430     from Airports2  
431     group by Year  
432     order by Total_passengers desc),  
433  
434     Yearly_growth as (Select Year,  
435         Total_flight,  
436         Total_passengers,  
437             -- LAG function: gets the previous row's value  
438         lag(Total_flight) over (order by Year) as Prev_flights,    -- order rows by Year  
439         lag(Total_passengers) over (order by Year) as Prev_passengers  
440         from Yearly_summary)  
441  
442         select Total_flight,Total_passengers,  
443             round(((Total_flight-Prev_flights) / nullif(Prev_flights,0)*100),2) as Flights_growth_percentage,  
444             round(((Total_passengers-Prev_passengers) / nullif(Prev_passengers,0)*100),2) as Passenger_growth_percentage  
445         from Yearly_growth  
446         order by Year desc  
447         limit 3;
```

Result Grid			
Total_flight	Total_passengers	Flights_growth_percentage	Passenger_growth_percentage
1	21	-98.89	-99.99
90	173468	4.65	17.39
86	147775	-9.47	7.28



## PROBLEM STATEMENT-15

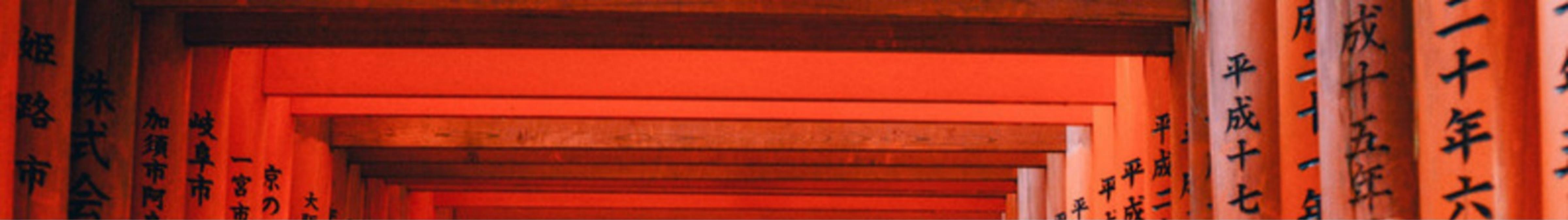
Identify the most significant routes by combining flight frequency and distance, using weighted distance to highlight top origin-to-destination pairs.

```
456 • with Route_distance as(Select Origin_airport, Destination_airport,  
457     Sum(Flights) as Total_flights, Sum(Distance) as Total_distance  
458     from Airports2  
459     group by Origin_airport, Destination_airport),  
460  
461     Weighted_route as  
462     (select Origin_airport,  
463         Destination_airport, Total_flights, Total_distance,  
464         Total_distance * Total_flights as Weighted_distance  
465     from Route_distance)  
466  
467     Select Origin_airport, Destination_airport, Total_flights, Total_distance, Weighted_distance  
468     from Weighted_route  
469     order by Weighted_distance desc  
470     Limit 3;
```

Result Grid | Filter Rows: \_\_\_\_\_ | Export: | Wrap Cell Content:

	Origin_airport	Destination_airport	Total_flights	Total_distance	Weighted_distance
▶	PDX	RDM	55147	59044	3256099468
	SEA	RDM	12919	37163	480108797
	SFO	RDM	5849	36960	216179040





# THANK YOU