-- #1: Busiest City in US

-- a) Most Arrivals + Departures.   Hint: City, State: Count of Arrivals and Departures

USE dbda3;

SELECT

row_number() over (order by COUNT(ARR_TIME) + COUNT(DEP_TIME) desc) as Sno,

rank () over (order by COUNT(ARR_TIME) + COUNT(DEP_TIME) desc) as Rno,

dense_rank () over (order by COUNT(ARR_TIME) + COUNT(DEP_TIME) desc) as Dno,

   ORIGIN_CITY_NAME, DEST_CITY_NAME,

   COUNT(ARR_TIME) + COUNT(DEP_TIME) AS 'Most Arrivals + Departures'

FROM

   dbda3.ontime_reporting

GROUP BY ORIGIN_CITY_NAME

LIMIT 10;

Sno	Rno	Dno	ORIGIN_CITY_NAME	DEST_CITY_NAME	Most Arriv
1	1	1	Chicago, IL	New York, NY	76322
2	2	2	Atlanta, GA	Dallas/Fort Worth, TX	70838
3	3	3	Dallas/Fort Worth, TX	Atlanta, GA	55614
4	4	4	New York, NY	Minneapolis, MN	52888
5	5	5	Denver, CO	Los Angeles, CA	46594
6	6	6	Houston, TX	Atlanta, GA	40642
7	7	7	Charlotte, NC	Detroit, MI	40634
8	8	8	Los Angeles, CA	Seattle, WA	39632
9	9	9	Washington, DC	Atlanta, GA	36068
10	10	10	San Francisco, CA	Los Angeles, CA	33426

-- b) Most Delayed + Cancelled + Diverted   Hint: Carrier, Count of flights, CANCELLED, CANCELLATION_CODE, Diverted, Reasons ??

SELECT

row_number() over (order by count(if(arr_delay>=1,1,NULL)) + sum(cancelled) + sum(diverted) desc) as Sno,

rank () over (order by count(if(arr_delay>=1,1,NULL)) + sum(cancelled) + sum(diverted) desc) as Rno,

dense_rank () over (order by count(if(arr_delay>=1,1,NULL)) + sum(cancelled) + sum(diverted) desc) as Dno,

ORIGIN_CITY_NAME, DEST_CITY_NAME, OP_UNIQUE_CARRIER Carrier,

count(if(arr_delay>=1,1,NULL)) countdelay, sum(cancelled) cancelled, sum(diverted),

count(if(arr_delay>=1,1,NULL)) + sum(cancelled) + sum(diverted) as "Most Delay_Cancel_Diverted",

count(*) count_flights

from dbda3.ontime_reporting

group by ORIGIN_CITY_NAME, DEST_CITY_NAME, OP_UNIQUE_CARRIER

LIMIT 10;

Sno	Rno	Dno	ORIGIN_C	DEST_CITY	Carrier	countdelay	cancelled	sum(divert	Most Dela	count_fligh
1	1	1	New York,	Atlanta, G	DL	328	10	0	338	739
2	2	2	Seattle, W	Anchorage	AS	289	2	0	291	637
3	3	3	Atlanta, G	New York,	DL	260	9	8	277	739
4	4	4	San Franci	Newark, N	UA	244	5	12	261	447
5	5	5	Kahului, HI	Honolulu,	HA	258	1	0	259	815
6	6	6	Newark, N	San Franci	UA	239	9	0	248	445
7	7	7	Houston, T	Dallas, TX	WN	204	37	3	244	571
8	7	7	Chicago, IL	New York,	AA	212	29	3	244	465
9	9	8	Atlanta, G	Fort Laude	DL	236	1	1	238	383
10	9	8	Boston, M	Washingto	B6	232	5	1	238	410
11	11	9	Miami, FL	New York,	AA	219	11	7	237	446
12	12	10	San Franci	Seattle, W	AS	223	7	0	230	415
13	13	11	Dallas/For	Chicago, IL	AA	217	11	0	228	433
14	14	12	Chicago, IL	San Franci	UA	221	1	1	223	447
15	15	13	Chicago, IL	Dallas/For	AA	210	11	0	221	429
16	16	14	San Franci	Seattle, W	UA	220	0	0	220	420
17	17	15	Atlanta, G	Orlando, F	DL	217	0	1	218	494
18	18	16	Washingto	New York,	YX	178	37	2	217	428
19	18	16	San Franci	Chicago, IL	UA	216	0	1	217	445
20	18	16	Dallas/For	Los Angele	AA	209	7	1	217	453
21	21	17	Dallas/For	New York,	AA	193	16	7	216	464
22	22	18	New York,	Washingto	YX	178	36	0	214	428
23	23	19	Honolulu,	Kahului, HI	HA	204	0	0	204	808
24	24	20	New York,	Chicago, IL	DL	189	13	0	202	380
25	24	20	New York,	Chicago, IL	AA	174	28	0	202	464

-- #2: Carriers

-- a) Most Operated Flight for Carrier   Hint: Carrier, Count of flights.

select row_number() over (order by count(FLIGHTS) desc) as sno,

rank() over (order by count(FLIGHTS) desc) as rno,

dense_rank() over (order by count(FLIGHTS) desc) as dno,

OP_UNIQUE_CARRIER Carrier, count(FLIGHTS) Count_flight

FROM dbda3.ontime_reporting

group by op_unique_carrier

limit 10;

sno	rno	dno	Carrier	Count_fligh
1	1	1	WN	116526
2	2	2	DL	91278
3	3	3	AA	82689
4	4	4	OO	74973
5	5	5	UA	56223
6	6	6	MQ	29931
7	7	7	YX	28997
8	8	8	B6	25518
9	9	9	OH	25452
10	10	10	AS	24688

-- b) Most Delayed + Cancelled + Diverted   Hint: As Above.

select

row_number() over (order by sum(DIVERTED) + sum(CANCELLED) + sum(CARRIER_DELAY) desc) as rno,

rank() over (order by sum(DIVERTED) + sum(CANCELLED) + sum(CARRIER_DELAY) desc) as rno,

dense_rank() over (order by sum(DIVERTED) + sum(CANCELLED) + sum(CARRIER_DELAY) desc) as dno,

OP_UNIQUE_CARRIER Carrier, sum(DIVERTED), sum(CANCELLED), sum(CARRIER_DELAY),

sum(DIVERTED) + sum(CANCELLED) + sum(CARRIER_DELAY) as TOTAL_DELAYS

from dbda3.ontime_reporting

group by OP_UNIQUE_CARRIER

limit 10;

rno	rno	dno	Carrier	sum(DIVER	sum(CANC	sum(CARR	TOTAL_DE
1	1	1	AA	263	1869	417206	419338
2	2	2	DL	224	216	397448	397888
3	3	3	OO	224	886	388829	389939
4	4	4	WN	254	1539	300938	302731
5	5	5	UA	203	517	236707	237427
6	6	6	B6	98	248	224054	224400
7	7	7	YV	49	730	127946	128725
8	8	8	EV	40	671	100815	101526
9	9	9	OH	89	956	96942	97987
10	10	10	9E	48	667	88995	89710

#3: Flights

-- a) Most Between Cities   Hint: City Pair, Count of flights.

use dbda3;

select row_number() over (order by count(*) desc) as sno,

ORIGIN_CITY_NAME, DEST_CITY_NAME, count(*) countofflights

FROM dbda3.ontime_reporting

group by origin, dest

limit 10;

sno	ORIGIN_C	DEST_CITY	countofflig
1	San Franci	Los Angele	1384
2	Los Angele	San Franci	1383
3	Chicago, Il	New York,	1315
4	New York,	Chicago, Il	1314
5	Seattle, W	San Franci	1092
6	San Franci	Seattle, W	1090
7	Los Angele	New York,	1077
8	New York,	Los Angele	1047
9	Las Vegas,	Los Angele	1046
10	Los Angele	Las Vegas,	1024

-- b)

-- Longest   Hint: Flight#, Sum of Distance.

select row_number() over (order by round(avg(distance),2) desc) as sno,

ORIGIN_CITY_NAME, dest_CITY_NAME, round(avg(distance),2) ,
count(*),round(avg(CRS_ELAPSED_TIME),2), round(avg(ACTUAL_ELAPSED_TIME),2)

from dbda3.ontime_reporting

group by ORIGIN_CITY_NAME, dest_CITY_NAME

limit 10;

sno	ORIGIN_C	dest_CITY	round(avg(	count(*)	round(avg(	round(avg(
1	Boston, M.	Honolulu,	5095	27	665	675.15
2	Honolulu,	Boston, M.	5095	27	590	578.7
3	Honolulu,	New York,	4983	31	585	577.23
4	New York,	Honolulu,	4983	31	650	657.19
5	Newark, N	Honolulu,	4962	31	656.45	650.32
6	Honolulu,	Newark, N	4962	31	568.45	559.81
7	Washingto	Honolulu,	4817	25	617.28	623.72
8	Honolulu,	Washingto	4817	25	549.6	539.92
9	Honolulu,	Atlanta, G.	4502	31	530.39	511.74
10	Atlanta, G.	Honolulu,	4502	31	580.45	560.39

-- Shortest

select row_number() over (order by round(avg(distance),2)) as sno,

 ORIGIN_CITY_NAME, dest_CITY_NAME, round(avg(distance),2), count(*),round(avg(CRS_ELAPSED_TIME),2), round(avg(ACTUAL_ELAPSED_TIME),2)

from dbda3.ontime_reporting

group by ORIGIN_CITY_NAME, dest_CITY_NAME

limit 10;

sno	ORIGIN_C	dest_CITY_	round(avg(	count(*)	round(avg(	round(avg(
1	Petersburg	Wrangell, /	31	31	20	23.74
2	Wrangell, /	Petersburg	31	31	20	23.87
3	Juneau, AK	Gustavus, /	41	26	35	32.81
4	Gustavus, /	Juneau, AK	41	26	25	27.46
5	Cape Girar	Paducah, K	45	57	47.84	35.04
6	Paducah, K	Cape Girar	45	58	48.4	31.24
7	San Francis	Santa Rosa	66	31	54.35	41.48
8	Santa Rosa	San Francis	66	31	50.61	42.77
9	Chicago, IL	Milwaukee	67	145	51.63	47.68
10	Milwaukee	Chicago, IL	67	145	58.82	53.43

-- #4: Day of Week

-- a) Most Operated   Hint: Day#, Count of flights

```sql
select
  row_number() over (order by count(*) desc) as Sno,
  case day_of_week
    when 1 then "Monday" when 2 then "Tuesday" when 3 then "Wednesday" when 4 then "Thursday"
    when 5 then "Friday" when 6 then "Saturday" when 7 then "Sunday" when 9 then "Unknown"
   END as DayOfWeek,
    count(*) numflights, min(arr_delay) minarrdelay,
    max(arr_delay) maxarrdelay, sum(arr_delay) sumarrdelay,
    round(avg(arr_delay),2) avgarrdelay, min(dep_delay) mindepdelay,
    max(dep_delay) maxdepdelay, sum(dep_delay) sumdepdelay,
    round(avg(dep_delay),2) avgdepdelay
from dbda3.ontime_reporting
group by DAY_OF_WEEK;
```

Sno	DayOfWeek	numflights	minarrdela	maxarrdel	sumarrdela	avgarrdela	mindepdel	maxdepde	sumdepde	avgdepdela
1	Friday	110812	-67	1658	791408	7.14	-34	1667	1363132	12.3
2	Thursday	110762	-67	1561	1317379	11.89	-36	1580	1752470	15.82
3	Saturday	90300	-83	1647	55547	0.62	-35	1667	688618	7.63
4	Monday	88625	-66	1222	398511	4.5	-36	1247	908663	10.25
5	Wednesda	86418	-56	1516	927777	10.74	-37	1503	1269294	14.69
6	Sunday	85819	-73	1453	547425	6.38	-33	1471	1050010	12.24
7	Tuesday	85725	-66	1505	800355	9.34	-47	1527	1161267	13.55

-- b) Most Delayed + Cancelled + Diverted   Hint: As Above.

SELECT   row_number() over (order by count(if(arr_delay>=1,1,NULL)) + sum(cancelled) + sum(diverted) desc) as Sno,

case day_of_week

   when 1 then "Monday" when 2 then "Tuesday" when 3 then "Wednesday" when 4 then "Thursday"

   when 5 then "Friday" when 6 then "Saturday" when 7 then "Sunday" when 9 then "Unknown"

  END as DayOfWeek,

count(if(arr_delay>=1,1,NULL)) countdelay, sum(cancelled) cancelled, sum(diverted),

count(if(arr_delay>=1,1,NULL)) + sum(cancelled) + sum(diverted) as "Most Delayed + Cancelled + Diverted",

count(*) count_flights

from dbda3.ontime_reporting

group by DAY_OF_WEEK;

Sno	DayOfWeek	countdelay	cancelled	sum(divert	Most Dela	count_fligh
1	Thursday	43533	2082	358	45973	110762
2	Friday	40197	1640	303	42140	110812
3	Wednesda	32088	2557	375	35020	86418
4	Tuesday	31301	1820	227	33348	85725
5	Sunday	28082	1425	196	29703	85819
6	Monday	27957	1173	218	29348	88625
7	Saturday	26036	601	135	26772	90300