# ANALYSIS ON FLIGHT TO SEATTLE

1. *Find the number of flights took-off from NYC airports to different destination and display the top 5 destinations*

```
mysql> select dest,count(*) as cnt from flight group by dest order by cnt desc limit 5;
+------+-------+
| dest | cnt   |
+------+-------+
| ATL  | 16837 |
| ORD  | 16566 |
| LAX  | 16026 |
| BOS  | 15022 |
| MCO  | 13967 |
+------+-------+
5 rows in set (0.20 sec)
```

2. *Find the count of flights took-off from NYC in 2013 to Seattle*

```
mysql> select count(*) from flight where dest = 'SEA' and year = '2013';
+----------+
| count(*) |
+----------+
|     3885 |
+----------+
1 row in set (0.13 sec)
```

3. *Total number of flights flown in 2013*

```
mysql> select year, count(col) from flight group by year;
+------+------------+
| year | count(col) |
+------+------------+
| 2013 |     327346 |
+------+------------+
1 row in set (0.18 sec)
```

4. *Number of airlines flown from NYC and its count*

```
mysql> select carrier, count(*) as cnt from flight group by carrier order by cnt desc;
+---------+-------+
| carrier | cnt   |
+---------+-------+
| UA      | 57782 |
| B6      | 54049 |
| EV      | 51108 |
| DL      | 47658 |
| AA      | 31947 |
| MQ      | 25037 |
| US      | 19831 |
| 9E      | 17294 |
| WN      | 12044 |
| VX      |  5116 |
| FL      |  3175 |
| AS      |   709 |
| F9      |   681 |
| YV      |   544 |
| HA      |   342 |
| OO      |    29 |
+---------+-------+
16 rows in set (0.20 sec)
```

5. *Number of distinct flights from NYC to SEATTLE*

```
mysql> select count(distinct tailnum) from flight where dest = 'SEA';
+-------------------------+
| count(distinct tailnum) |
+-------------------------+
|                     933 |
+-------------------------+
1 row in set (0.13 sec)
```

6. *Average arrival delay by carriers for flights from NYC to SEATTLE*

```
mysql> select carrier, avg(arr_delay) as aver from flight where dest = 'SEA' group by carrier order by aver;
+---------+---------+
| carrier | aver    |
+---------+---------+
| AS      | -9.9309 |
| DL      | -5.8860 |
| AA      | -1.4750 |
| UA      |  5.8274 |
| B6      |  7.7212 |
+---------+---------+
5 rows in set (0.14 sec)
```

7. *What proportion of flights come to Seattle from NYC airport?*

```
mysql> create view tab1 as select origin, count(col) as cnt from flight where dest = 'SEA' group by origin;

mysql> select origin,cnt,round((cnt*100)/(select sum(cnt) as s from tab1),2) as '% DISTRIBUTION' from tab1;
+--------+------+----------------+
| origin | cnt  | % DISTRIBUTION |
+--------+------+----------------+
| EWR    | 1810 |          46.59 |
| JFK    | 2075 |          53.41 |
+--------+------+----------------+
```

*Premjith Augustine – SQL queries on NYC flights and Weather database*

# ANALYSIS ON FLIGHT DELAY

1. *Top 3 days which are having the longest average departure delay?*

```
mysql> select year,month,day, avg(dep_delay) as avg_delay from flight group by year, month, day order by avg_delay desc limit 3;
+------+-------+------+-----------+
| year | month | day  | avg_delay |
+------+-------+------+-----------+
| 2013 |     3 |    8 |   83.6479 |
| 2013 |     7 |    1 |   56.2235 |
| 2013 |     9 |    2 |   53.0583 |
+------+-------+------+-----------+
```

2. *Top 3 days which are having the longest average arrival delay?*

```
mysql> select year, month, day, avg (arr_delay) as avg_arr_delay from flight group by year, month, day order by avg_arr_delay desc limit 3;
+------+-------+------+---------------+
| year | month | day  | avg_arr_delay |
+------+-------+------+---------------+
| 2013 |     3 |    8 |       85.8622 |
| 2013 |     6 |   13 |       63.7537 |
| 2013 |     7 |   22 |       62.7634 |
+------+-------+------+---------------+
3 rows in set (0.30 sec)
```

3. *Which was the worst day to fly out from NYC in year 2013, if you dislike delayed flight?*

```
mysql> select year, month, day, count(dep_delay) as cnt_delayed_flight
    -> from flight
    -> where dep_delay > 0
    -> group by year, month, day
    -> order by cnt_delayed_flight
    -> desc
    -> limit 1;
+------+-------+------+--------------------+
| year | month | day  | cnt_delayed_flight |
+------+-------+------+--------------------+
| 2013 |    12 |   23 |                673 |
+------+-------+------+--------------------+
```

Considering the choice of delayed flights, the most obvious day not to fly out from NY would Dec 23rd which had the maximum number of delayed flights.

4. *Check whether autumn (Sept, Oct, Nov) period is worse than summer (June, July, August) for delayed flights from NYC*

```
mysql> select avg(avg_delay) as avg_autumn_delay from (select month, avg(dep_delay) as avg_delay from flight where month in (9,10,11) group by month) as temp;
+------------------+
| avg_autumn_delay |
+------------------+
|       6.09460000 |
+------------------+
1 row in set (0.15 sec)

mysql> select avg(avg_delay) as avg_summer_delay from (select month, avg(dep_delay) as avg_delay from flight where month in (6,7,8) group by month) as temp;
+------------------+
| avg_summer_delay |
+------------------+
|      18.27276667 |
+------------------+
```

Summer period has more delay time than autumn season

5. *Find out the departure delay from NY airports on an hourly basis.*

There are few instances where hours are recorded as both 24 and 0, hence it required to merge and read it as either 24 or 0.

```
mysql> select
    -> (case when hour = 24 then 0 else hour end)
    -> as group_hour,
    -> avg(dep_delay) as avg_delay
    -> from flight
    -> group by group_hour;
+-------------+-------------+
| group_hour  | avg_delay   |
+-------------+-------------+
|          0  |   127.2232  |
|          1  |   206.7557  |
|          2  |   236.2540  |
|          3  |   304.7273  |
|          4  |    -5.5541  |
|          5  |    -4.3563  |
|          6  |    -1.5218  |
|          7  |     0.2147  |
|          8  |     1.0923  |
|          9  |     4.2341  |
|         10  |     5.5111  |
|         11  |     5.6133  |
|         12  |     7.5173  |
|         13  |     9.3639  |
|         14  |     8.0518  |
|         15  |    10.5933  |
|         16  |    13.5572  |
|         17  |    16.6557  |
|         18  |    18.4747  |
|         19  |    21.3102  |
|         20  |    28.0876  |
|         21  |    41.8441  |
|         22  |    67.9586  |
|         23  |    96.6384  |
+-------------+-------------+
24 rows in set (0.25 sec)
```

Flights during the mid-night period has the lengthiest delay, majorly due to passengers preferred timing and availability of cheaper flights during the period

# ANALYSIS ON FASTEST FLIGHT

1.  *Which is the fastest flight from NYC*

```
mysql> select year, month, day, carrier, tailnum, flight, origin, dest,
    -> (distance/air_time) as speed_minute
    -> from flight
    -> order by speed_minute desc
    -> limit 1;
+------+-------+------+---------+---------+--------+--------+------+--------------+
| year | month | day  | carrier | tailnum | flight | origin | dest | speed_minute |
+------+-------+------+---------+---------+--------+--------+------+--------------+
| 2013 |     5 |   25 | DL      | N666DN  | 1499   | LGA    | ATL  |      11.7231 |
+------+-------+------+---------+---------+--------+--------+------+--------------+
1 row in set (0.27 sec)
```

Flight 1499 from LGA to ATL is the fastest with miles/minute at 11.72.

2.  *Find the avg speed of carriers from NY airport and arrange in descending order*

```
mysql> select carrier,
    -> avg(distance/air_time) as miles_minute
    -> from flight
    -> group by carrier
    -> order by miles_minute desc;
+---------+--------------+
| carrier | miles_minute |
+---------+--------------+
| HA      |   8.00596462 |
| VX      |   7.43624848 |
| AS      |   7.39465331 |
| F9      |   7.08620015 |
| UA      |   7.01473118 |
| DL      |   6.97438076 |
| AA      |   6.95788094 |
| WN      |   6.67553191 |
| B6      |   6.66619308 |
| FL      |   6.57263915 |
| MQ      |   6.14004828 |
| OO      |   6.10533793 |
| EV      |   6.04906057 |
| 9E      |   5.75717496 |
| US      |   5.69899333 |
| YV      |   5.53282813 |
+---------+--------------+
16 rows in set (0.33 sec)
```

# REGULAR FLIGHTS

3.  *Top 3 Routine flights from NYC*

```
mysql> create view new_tab as select concat (year,'-',month,'-',day) as new_days, concat (carrier,'-',flight,'-',dest) as new_fl from flight group by new_fl,new_days;
Query OK, 0 rows affected (0.04 sec)

mysql> select new_fl, count(distinct new_days) as cnt from new_tab group by new_fl order by cnt desc limit 3;
+-------------+-----+
| new_fl      | cnt |
+-------------+-----+
| B6-1783-MCO | 365 |
| B6-703-SJU  | 364 |
| DL-2159-MCO | 364 |
```

*Premjith Augustine – SQL queries on NYC flights and Weather database*

Flight B6 – 1783 to MCO (Orlando International Airport) has 365 days service from NYC

## RESEARCH QUESTIONS

1. *Which two airlines flying to Seattle were most reliable in terms of having the minimum average departure delay in 2013?*

```
mysql> select carrier,
    -> round(avg(dep_delay),2) as avg_delay,
    -> count(dep_delay) as count
    -> from flight where dest = 'SEA'
    -> group by carrier
    -> order by avg_delay
    -> limit 2;
+---------+-----------+-------+
| carrier | avg_delay | count |
+---------+-----------+-------+
| AS      |      5.83 |   709 |
| DL      |      6.98 |  1202 |
+---------+-----------+-------+
```

2. *Find if there is any link between visibility and temperature on flight delay?*

   Let's check the effect of temperature and visibility on flights which had departure delay by more than 120 minutes.

```
mysql> create view left_table as
    -> select w.year,w.month,w.day,w.origin,w.temp,w.visib,
    -> f.dep_delay,f.carrier,f.tailnum,f.flight,f.dest
    -> from whe w left join flight f on
    -> (w.year = f.year and w.month = f.month and
    -> w.day = f.day and w.origin = f.origin);
Query OK, 0 rows affected (0.05 sec)

mysql> select carrier,tailnum,flight,
    -> round(avg(dep_delay),2) as avg_delay,
    -> round(avg(temp),2) as avg_temp,
    -> round(avg(visib),2) as avg_visb
    -> from left_table
    -> where dep_delay>120
    -> group by carrier,flight,tailnum
    -> order by avg_delay desc
    -> limit 20;

+---------+---------+--------+-----------+----------+----------+
| carrier | tailnum | flight | avg_delay | avg_temp | avg_visb |
+---------+---------+--------+-----------+----------+----------+
| MQ      | N517MQ  | 3695   |   1126.00 |    44.19 |    10.00 |
| AA      | N5DMAA  | 172    |    896.00 |    52.61 |     5.44 |
| MQ      | N523MQ  | 3744   |    878.00 |    57.53 |    10.00 |
| DL      | N375NC  | 1223   |    849.00 |    32.34 |     9.63 |
| AA      | N5EMAA  | 172    |    845.00 |    25.66 |     8.13 |
| DL      | N990AT  | 2042   |    798.00 |    48.72 |    10.00 |
| DL      | N348NW  | 575    |    786.00 |    40.52 |     8.14 |
| DL      | N943DL  | 502    |    702.00 |    58.70 |    10.00 |
| DL      | N338NB  | 2343   |    592.00 |    38.12 |    10.00 |
| AA      | N3EMAA  | 1895   |    580.00 |    81.22 |    10.00 |
| EV      | N12163  | 4711   |    548.00 |    52.61 |     5.44 |
| DL      | N310DE  | 2343   |    545.00 |    40.87 |    10.00 |
| EV      | N909EV  | 5714   |    503.00 |    26.96 |    10.00 |
| DL      | N360NB  | 2042   |    502.00 |    79.21 |    10.00 |
| B6      | N661JB  | 517    |    502.00 |    36.06 |     6.36 |
| US      | N180US  | 1745   |    486.00 |    64.10 |    10.00 |
| EV      | N13970  | 4300   |    483.00 |    80.01 |     9.43 |
| DL      | N322US  | 1157   |    467.00 |    76.34 |     9.63 |
| DL      | N322NB  | 926    |    454.00 |    79.21 |    10.00 |
| EV      | N29917  | 4326   |    443.00 |    31.96 |     9.95 |
+---------+---------+--------+-----------+----------+----------+
20 rows in set (7.08 sec)
```

Let's check the effect of temperature and visibility on flights which had departed on or before scheduled time

```
mysql> create view left_table1 as
    -> select w.year,w.month,w.day,w.origin,w.temp,w.visib,
    -> f.dep_delay,f.carrier,f.tailnum,f.flight,f.dest
    -> from whe w left join flights f on
    -> (w.year = f.year and w.month = f.month and
    -> w.day = f.day and w.origin = f.origin);
Query OK, 0 rows affected (0.03 sec)

mysql> select carrier,tailnum,flight,
    -> round(avg(dep_delay),2) as avg_delay,
    -> round(avg(temp),2) as avg_temp,
    -> round(avg(visib),2) as avg_visb
    -> from left_table1
    -> where dep_delay<1
    -> group by carrier,flight,tailnum
    -> order by avg_delay desc
    -> limit 20;
+---------+---------+--------+-----------+----------+----------+
| carrier | tailnum | flight | avg_delay | avg_temp | avg_visb |
+---------+---------+--------+-----------+----------+----------+
| EV      | N612QX  | 5203   |      0.00 |    39.90 |     9.79 |
| UA      | N415UA  | 375    |      0.00 |    60.25 |    10.00 |
| EV      | N15983  | 5801   |      0.00 |    52.38 |     6.40 |
| UA      | N37413  | 1492   |      0.00 |    25.08 |    10.00 |
| EV      | N16561  | 4612   |      0.00 |    39.34 |    10.00 |
| UA      | N14230  | 1668   |      0.00 |    79.29 |     9.87 |
| UA      | N15712  | 1623   |      0.00 |    31.05 |     9.54 |
| UA      | N816UA  | 363    |      0.00 |    25.87 |    10.00 |
| UA      | N445UA  | 499    |      0.00 |    33.48 |    10.00 |
| UA      | N41140  | 1592   |      0.00 |    64.43 |    10.00 |
| UA      | N77296  | 1645   |      0.00 |    75.59 |    10.00 |
| US      | N151UW  | 1437   |      0.00 |    87.58 |     9.79 |
| WN      | N474WN  | 16     |      0.00 |    58.70 |    10.00 |
| UA      | N68805  | 1493   |      0.00 |    35.83 |     4.60 |
| WN      | N737JW  | 1078   |      0.00 |    38.12 |    10.00 |
| WN      | N485WN  | 329    |      0.00 |    63.05 |    10.00 |
| B6      | N705JB  | 1004   |      0.00 |    17.96 |    10.00 |
| EV      | N14203  | 4660   |      0.00 |    38.48 |    10.00 |
| EV      | N14143  | 4372   |      0.00 |    39.35 |    10.00 |
| UA      | N481UA  | 234    |      0.00 |    44.26 |     8.67 |
+---------+---------+--------+-----------+----------+----------+
20 rows in set (2 min 6.88 sec)
```

There's no proper evidence to prove that there is relation between temperature, visibility and departure delay.