**My SQL Case Study on Flights**

**GRADED ASSIGNMENT - II**

**SQL**

**Dataset:** Flights\_Delay.csv

Dataset Description:

ID: Rows ID

YEAR: 2015

MONTH: 1-12

DAY: 1-31

DAY\_OF\_WEEK: 1 (Monday) - 7 (Sunday)

AIRLINE: Airline CODE

FLIGHT\_NUMBER: Flight Number

TAIL\_NUMBER: Flight’s tail number

ORIGIN\_AIRPORT: Origin IATA airport code

DESTINATION\_AIRPORT: Destination IATA airport code

SCHEDULED\_DEPARTURE: Actual departure time (local, hhmm)

DEPARTURE\_TIME: Scheduled departure time (local, hhmm)

DEPARTURE\_DELAY: Departure delay, in minutes

TAXI\_OUT: Taxi out time in minutes

WHEELS\_OFF:

SCHEDULED\_TIME: Scheduled arrival time (local, hhmm)

ELAPSED\_TIME: in Minutes

AIR\_TIME: in Minutes

DISTANCE: in Miles

WHEELS\_ON:

TAXI\_IN: Taxi in time, in minutes

SCHEDULED\_ARRIVAL: Scheduled arrival time (local, hhmm)

ARRIVAL\_TIME: Actual arrival time (local, hhmm)

ARRIVAL\_DELAY: Arrival delay, in minutes

DIVERTED: 1 = yes, 0 = no

CANCELLED: 1 = yes, 0 = no

1. Create external table “flights” using Database “airline\_delayDB”
2. Describe the table schema & show top 10 rows of Dataset
3. Find duplicates rows present in dataset.

Write SQL to show following analysis

1. Average arrival delay caused by airlines
2. Days of months with respected to average of arrival delays
3. Arrange weekdays with respect to the average arrival delays caused
4. Arrange Days of month as per cancellations done in Descending
5. Finding busiest airports with respect to day of week
6. Finding airlines that make the maximum number of cancellations
7. Find and order airlines in descending that make the most number of diversions
8. Finding days of month that see the most number of diversions
9. Calculating mean and standard deviation of departure delay for all flights in minutes
10. Calculating mean and standard deviation of arrival delay for all flights in minutes
11. Create a partitioning table “flights partition” using suitable partitioned by schema.
12. Finding all diverted Route from a source to destination Airport & which route is the most diverted
13. Write a query to show Top 3 airlines from each airport making most Delays. (Use Dense Rank/ Rank)
14. Write a query to show Top 10 airlines from each week making most Delays. Find its Ranking.
15. Create a materialized view for client to show Top 10 airlines with highest Delay.
16. Create a new column named ‘Delay\_Comaprison’ showing if flights making higher or lower than average flight delay.
17. Finding AIRLINES with its total flight count, total number of flights arrival delayed by more than 30 Minutes, % of such flights delayed by more than 30 minutes when it is not Weekends with minimum count of flights from Airlines by more than 10. Also Exclude some of Airlines 'AK', 'HI', 'PR', 'VI' and arrange output in descending order by % of such count of flights.

Negative delays can be considered, but ignore the null values

1. Finding AIRLINES with its total flight count with total number of flights departure delayed by less than 30 Minutes, % of such flights delayed by less than 30 minutes when it is Weekends with minimum count of flights from Airlines by more than 10. Also Exclude some of Airlines 'AK', 'HI', 'PR', 'VI' and arrange output in descending order by % of such count of flights.
2. When is the best time of day/day of week/time of a year to fly with minimum delays?
3. Suggest reasons of airlines delays and suggest, build solutions for it.
4. Create a stored procedure to find weeks with maximum flights delays count.