

Assignment28:

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights appears in DOT's monthly Air Travel Consumer Report, published about 30 days after the month's end, as well as in summary tables posted on this website. Summary statistics and raw data are made available to the public at the time the Air Travel Consumer Report is released.

You can download the datasets from the following links: Delayed_Flights.csv

Delayed_Flights.csv Datasets

There are 29 columns in this dataset.

Some of them have been mentioned below:

- Year: 1987 – 2008
- Month: 1 – 12
- FlightNum: Flight number
- Canceled: Was the flight canceled?
- CancellationCode: The reason for cancellation. For complete details, refer to this link.

Note: Source code with explanation is uploaded as separate file

Problem Statement 1:

Find out the top 5 most visited destinations.

Query used:

```
val destination = spark.sql("select Dest , count(Dest) from FlightDetails group by Dest order by count(Dest) desc").toDF().take(5)
println(s"Gives the top 5 most visited Destination by the user")
${destination.foreach(println)} "
```

Output:

```
[ORD,108984]
18/06/27 17:12:30 INFO SparkContext: Invoking stop() from shutdown hook
[ATL,106898]
[DFW,70657]
[DEN,63003]
[LAX,59969]
Gives the top 5 most visited Destination by the user {}
```

Problem Statement 2:

Which month has seen the most number of cancellations due to bad weather?

Query used:

```
-----  
val cancelled = spark.sql("select Month , count(cancelled) from FlightDetails where  
CancellationCode = 'B' and Cancelled = 1 group by Month order by count(cancelled)  
desc").toDF().take(1)  
println(s"Gives the month that has maximum cancellation due to bad weather  
${cancelled.foreach(println)} ")
```

Output:

```
-----  
[12,250]  
Gives the month that has maximum cancellation due to bad weather  ()
```

Problem Statement 3:

Which route (origin & destination) has seen the maximum diversion?

Query used:

```
-----  
val Route = spark.sql("select Origin, Dest , count(Diverted) from FlightDetails where  
Diverted = 1 group by Origin, Dest order by count(Diverted) desc").toDF().take(1)  
println(s"Gives the route that has seen the maximum diversions  
${Route.foreach(println)} ")
```

Output:

```
-----  
[ORD,LGA,39]  
Gives the route that has seen the maximum diversions  ()
```