



Hadoop: Pig Assignment - 2

Problem Statement:

The US 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.

Dataset:

<https://intellipaate-course-attachments.s3.ap-south-1.amazonaws.com/Hadoop/Hadoop+Datasets-20200609T120700Z-001.zip>

Dataset Description:

The data format is comma-separated values.

These are two different datasets, i.e., **flight_details.csv** and **airports.csv**.

There are 29 rows in the **flight_details.csv** dataset:

Name	Description
Year	1987-2008
Month	1-12
DayofMonth	1-31
DayOfWeek	1 (Monday) - 7 (Sunday)
DepTime	Actual departure time (local, hhmm)
CRSDepTime	Scheduled departure time (local, hhmm)
ArrTime	Actual arrival time (local, hhmm)
CRSArrTime	Scheduled arrival time (local, hhmm)
UniqueCarrier	Unique carrier code
FlightNum	Flight number
TailNum	Plane tail number
ActualElapsedTime	In minutes
CRSElapsedTime	In minutes
AirTime	In minutes
ArrDelay	Arrival delay, in minutes

DepDelay	Departure delay, in minutes
Origin	Origin IATA airport code
Dest	Destination IATA airport code
Distance	In miles
TaxiIn	Taxi in time, in minutes
TaxiOut	Taxi out time in minutes
Canceled	Was the flight canceled?
CancellationCode	Reason for cancellation (A = carrier, B = weather, C = NAS, D = security)
Diverted	1 = yes, 0 = no
CarrierDelay	In minutes
WeatherDelay	In minutes
NASDelay	In minutes
SecurityDelay	In minutes
LateAircraftDelay	In minutes

For **airport.csv**:

- **iaa**: the international airport abbreviation code
- name of the airport
- city and country in which the airport is located
- **lat** and **long**: the latitude and longitude of the airport

Tasks To Be Performed:

- Find out the top five most visited destinations.
- Which month has seen the greatest number of cancellations due to bad weather?
- Find out the top ten origins with the highest AVG departure delay.
- Which route (origin and destination) has seen the maximum diversion?