Source of Dataset : <https://www.kaggle.com/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018>

References : <https://openedx.atlassian.net/wiki/spaces/OpenOPS/pages/25559308/How+to+use+the+S3+command-line+tool+Mac+OSX+Linux>

Creating the Tables in the database

CREATE TABLE Airline (Serial\_Number INT, Flight\_Date Date, Carrier varchar(5), Flight\_Number varchar(5), Origin varchar(5), Destination varchar(5), Scheduled\_Departure Time, Actual\_Departure Time, Departure\_Delay INT, Scheduled\_Arrival Time, Actual\_Arrival Time, Arrival\_Delay INT, Cancelled INT,  
Cancellation\_Code varchar(3), Diverted INT, Distance INT, Carrier\_Delay INT, Weather\_Delay INT, NAS\_Delay INT, Security\_Delay INT, Late\_Aircraft\_Delay INT, PRIMARY KEY (Serial\_Number));

CREATE TABLE TopRoutesOTP (Carrier varchar(5), Origin varchar(5), Destination varchar(5), Total\_Flights INT, Flights\_On\_Time INT, OTP\_Percentage float);

Example Loading files from S3 to Linux through PuTTY:

aws s3 cp “s3://airlinedata622/part5\_4.csv” “part5\_4.csv”

Example of Loading csv file into Table from Linux:

load data local infile "/home/centos/part5\_4.csv" into table Airline fields terminated by ',' lines terminated by '\n' ignore 1 lines (Serial\_Number, Flight\_Date, Carrier, Flight\_Number, Origin, Destination, Scheduled\_Departure, Actual\_Departure, Departure\_Delay, Scheduled\_Arrival, Actual\_Arrival, Arrival\_Delay, Cancelled, Cancellation\_Code, Diverted, Distance, Carrier\_Delay, Weather\_Delay, NAS\_Delay, Security\_Delay, Late\_Aircraft\_Delay);