**Project in Python and SQLite**

Capstone Project 2016 june Coursera Python for everybody

The data are available at:

**http://stat-computing.org/dataexpo/2009/the-data.html**

[The data](http://stat-computing.org/dataexpo/2009/the-data.html) consists of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008.

I use the data from 2007 and 2008. This is a moderate dataset: there are nearly 8 million records in total. I use SQLite, a sql database, to import the data and use python to extract data from the tables for analysis. First I used 10.000 records from the 2008 Data as a test to see if the code works.

## The project

The aim of the data expo is to provide a summary of important features of the data set. I focus on different aspects of the data:

* Which airports (Top 10) have the most departure delays?
* Which airports (Top 10) have the most arrival delays?
* When is the best time of day/day of week/time of year to fly to minimise departure and/or arrival delays?

The purpose is to extract data from the SQLite tables and use python for data analysis.