Python Data Analysis Review Exercise

Start a new notebook and save it with a suitable name
Import pandas, numpy and matplotlib.pyplot as usual

USA Flights Data

Create a data frame from the following URL:
https://raw.githubusercontent.com/ismayc/pnwflights14/master/data/flights.csv

- Show how many elements this data has, the column names and the data types for each column
- Show the statistical summaries for the numeric columns in the dataset (use df.describe())
- Group the data into separate data frames for each airline
- Show the minimum and mean departure delay for each airline
- Use the .agg() method to aggregate min, mean and max delays for departure and arrival

In the 'flights' data set, find

- The number of flight records
- The number of unique airlines
- How many unique aircraft are represented
- The greatest recorded delay (show the related data)
- The mean values for just the first 50 records in the dataset, then for all delays of 15 or more units

Generate the following

- A data frame containing all AA flights that have no missing data members
- A data frame containing all flights departing from LHA
- A data frame sorted by increasing flight duration

Decide how you can address the following challenges:

- On average, which airline is the most punctual
- Which airline arrives ahead of time most often
- Does flight duration appear to affect arrival punctuality more for some airlines than others
- Does time of day appear to affect departure punctuality across all flights
- Are the standard deviations of flight duration close to the standard deviations of flight distance

Optional

Explore the datasets available at https://www.kaggle.com/datasets
Sign up for free and download a few data sets to analyse
e.g. find statistical relationships between wealth and life expectancy