Azure Lab Clinic

Assignment Day 2

1. Enter the following commands in the terminal window to create a "flights" subdirectory in the "notebooks" directory and download a dataset from Azure blob storage into the "flights" subdirectory:
   - `cd notebooks`
   - `mkdir flights`
   - `cd flights`
   - `curl https://topcs.blob.core.windows.net/public/FlightData.csv --output flightdata.csv`
2. Now Launch Jupyter and in the browser window that opens, click **flights** to open the "flights" directory.
3. Confirm that **flightdata.csv** is present in the "flights" directory. Then click the **New** button and choose **Python 3 Spark - local** from the drop-down list to create a new Jupyter notebook with a Python 3 kernel.

4. In the first cell of the notebook, enter the following Python code to load **flightdata.csv** and create a Pandas DataFramefrom it.

```
import pandas as pd

df = pd.read_csv('flightdata.csv')
     df.head()
```

5. Select the **Run Cells** command from the **Cell** menu (or press **Ctrl+Enter**) to execute the Python code. Confirm that the output resembles the output below.

Now using the Flight data. try and execute the following: -

- Use Pandas to identify columns in a dataset with missing values
- Use Pandas to replace missing values with real values
- Use Pandas to filter columns in a dataset
- Use Pandas to quantize values in a column
- Use Pandas to create indicator columns representing categorical data