**Assignment 01: Evaluate the FAA Dataset**

DESCRIPTION

**Problem:**

Analyze the Federal Aviation Authority (FAA) dataset using Pandas to do the following:

1. View

* aircraft make name
* state name
* aircraft model name
* text information
* flight phase
* event description type
* fatal flag

         2. Clean the dataset and replace the fatal flag NaN with “No”

3. Find the aircraft types and their occurrences in the dataset

4. Remove all the observations where aircraft names are not available

5. Display the observations where fatal flag is “Yes”

**Assessment:**

Note: For code refer the attached notebook in pdf format.

#### 1: View and import the dataset

* imported the Pandas library.
* Imported the FAA (Federal Aviation Authority) dataset using ‘read\_csv’ to read the ‘faa\_ai\_prelim.csv’ file.
* The dataset is named as ‘df\_FAA\_dataset’

#### 2: View and understand the dataset

* Using ‘shape’ method ‘df\_FAA\_dataset.shape’, the shape of the ‘df\_FAA\_dataset’ is obtained.

The dataset has 83 Rows and 42 columns.

* df\_FAA\_dataset.head() is used to view the first five observation in the dataset.
* df\_FAA\_dataset.columns is used to obtain all the columns present in the dataset.

#### 3: Extract the following attributes from the dataset:

1. Aircraft make name
2. State name
3. Aircraft model name
4. Text information
5. Flight phase
6. Event description type
7. Fatal flag

* A new dataframe ‘df\_analyse\_dataset’ is created, which contains above mentioned columns.
* The type of the dataset is obtained using ‘type(df\_analyse\_dataset)’, which returns pandas.core.frame.DataFrame as output.
* Again, The dataframe is checked whether it contains all the required attributes, here the column “FATAL\_FLAG” has number of missing values s ‘NaN’, which has to be cleaned.

#### 4. Clean the dataset and replace the fatal flag NaN with “No”

* Fillna() is used to replace all missing ‘NaN’ values with ‘No’ in Fatal\_Flag column.
* Verify the missing values are replaced as expected.
* Using shape method, the number of observation is determined as 83 rows and 7 Columns.

#### 5. Remove all the observations where aircraft names are not available.

* Using dropna(), the observations without aircraft names are dropped from the dataset.

#### 6. Find the aircraft types and their occurrences in the dataset

* Using shape method, the shape of the dataset is checked, now the new dataset is reduced to 78 rows and 7 columns.
* Using groupby(), the dataset is now grouped by aircraft name.
* Using size method, the number of times each aircraft type appears in the dataset is obtained.

#### 7: Display the observations where fatal flag is “Yes”

* Using groupby(), the dataset is now grouped by fatal flag.
* Using size method, the total number of fatal and non-fatal accidents are obtained,
* Non Fatal Accidents = 71, Fatal Accidents = 7.
* A new dataframe is created to view only the fatal accidents (Fatal Flag values = Yes)
* Using Shape method, the observation for Fatal Accident is obtained as (7, 7).