Assignment 01: Evaluate the FAA Dataset

The comments/sections provided are your cues to perform the assignment. You don't need to limit yourself to the number of rows/cells provided. You can add additional rows in each section to add more lines of code.

If at any point in time you need help on solving this assignment, view our demo video to understand the different steps of the code.

Happy coding!

1: VIew and import the dataset

In [106…
```
#Import necessary libraries
import pandas as pd
```

In [107…
```
#Import the FAA (Federal Aviation Authority) dataset
df = pd.read_csv('DataSets/faa_ai_prelim.csv')
```

2: View and understand the dataset

In [108…
```
#View the dataset shape
df.shape
```

Out[108… (83, 42)

In [109…
```
#View the first five observations
df.head()
```

Out[109…

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LOC
0	No	19-FEB-16	19-FEB-16	00:45:00Z	MARSHVILLE	North Carolina	

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LO
1	No	19-FEB-16	18-FEB-16	23:55:00Z	TAVERNIER	Florida	
2	No	19-FEB-16	18-FEB-16	22:14:00Z	TRENTON	New Jersey	
3	No	19-FEB-16	18-FEB-16	17:10:00Z	ASHEVILLE	North Carolina	
4	No	19-FEB-16	18-FEB-16	00:26:00Z	TALKEETNA	Alaska	

5 rows × 42 columns

In [110…
```
#View all the columns present in the dataset
df.columns
```

Out[110…
```
Index(['UPDATED', 'ENTRY_DATE', 'EVENT_LCL_DATE', 'EVENT_LCL_TIME',
       'LOC_CITY_NAME', 'LOC_STATE_NAME', 'LOC_CNTRY_NAME', 'RMK_TEXT',
       'EVENT_TYPE_DESC', 'FSDO_DESC', 'REGIST_NBR', 'FLT_NBR', 'ACFT_OPRTR',
       'ACFT_MAKE_NAME', 'ACFT_MODEL_NAME', 'ACFT_MISSING_FLAG',
       'ACFT_DMG_DESC', 'FLT_ACTIVITY', 'FLT_PHASE', 'FAR_PART', 'MAX_INJ_LVL',
       'FATAL_FLAG', 'FLT_CRW_INJ_NONE', 'FLT_CRW_INJ_MINOR',
       'FLT_CRW_INJ_SERIOUS', 'FLT_CRW_INJ_FATAL', 'FLT_CRW_INJ_UNK',
       'CBN_CRW_INJ_NONE', 'CBN_CRW_INJ_MINOR', 'CBN_CRW_INJ_SERIOUS',
       'CBN_CRW_INJ_FATAL', 'CBN_CRW_INJ_UNK', 'PAX_INJ_NONE', 'PAX_INJ_MINOR',
       'PAX_INJ_SERIOUS', 'PAX_INJ_FATAL', 'PAX_INJ_UNK', 'GRND_INJ_NONE',
       'GRND_INJ_MINOR', 'GRND_INJ_SERIOUS', 'GRND_INJ_FATAL', 'GRND_INJ_UNK'],
      dtype='object')
```

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

In [111…
```
#Create a new dataframe with only the required columns
df_new_req = df.filter(['ACFT_MAKE_NAME', 'LOC_STATE_NAME','ACFT_MODEL_NAME', 'RMK_TEXT
df_new_req.head()
```

Out[111…

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DE
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accid
1	VANS	Florida	RV7	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDING (LDG)	Incid
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incid
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incid
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incid

In [112…
```
#View the type of the object
type(df_new_req)
```

Out[112… `pandas.core.frame.DataFrame`

In [113…
```
#Check if the dataframe contains all the required attributes
df_new_req.head()
```

Out[113…

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DE
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accid

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DE
1	VANS	Florida	RV7	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDING (LDG)	Incid
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incid
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incid
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incid

4. Clean the dataset and replace the fatal flag NaN with "No"

In [114…
```
#Replace all Fatal Flag missing values with the required output
df_new_req['FATAL_FLAG'].fillna(value='No', inplace=True)
```

In [115…
```
#Verify if the missing values are replaced
df_new_req.head()
```

Out[115…

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DE
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accid
1	VANS	Florida	RV7	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDING (LDG)	Incid

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DE
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incid
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incid
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incid

In [116…
```
#Check the number of observations
df_new_req.shape
```

Out[116…
```
(83, 7)
```

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

In [117…
```
#Drop the unwanted values/observations from the dataset
df_new_modified_new = df_new_req.dropna(subset = ['ACFT_MAKE_NAME'])
```

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

In [118…
```
#Check the number of observations now to compare it with the original dataset and see h
df_new_modified_new.shape
```

Out[118…
```
(78, 7)
```

In [119…
```
#Group the dataset by aircraft name
df_new_modified_new.groupby(['ACFT_MAKE_NAME'])
```

Out[119…
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001DB30DE6430>
```

In [58]:
```
#View the number of times each aircraft type appears in the dataset (Hint: use the size
df_new_modified_new.groupby(['ACFT_MAKE_NAME']).size()
```

Out[58]:
```
ACFT_MAKE_NAME
AERO COMMANDER             1
AERONCA                    1
AEROSTAR INTERNATIONAL     1
AIRBUS                     1
BEECH                      9
BELL                       2
BOEING                     3
CESSNA                    23
```

```
CHAMPION                 2
CHRISTEN                 1
CONSOLIDATED VULTEE      1
EMBRAER                  1
ENSTROM                  1
FAIRCHILD                1
FLIGHT DESIGN            1
GLOBE                    1
GREAT LAKES              1
GRUMMAN                  1
GULFSTREAM               1
HUGHES                   1
LANCAIR                  2
MAULE                    1
MOONEY                   4
NORTH AMERICAN           1
No                       5
PIPER                   10
PITTS                    1
SAAB                     1
SABRELINER               1
SOCATA                   2
VANS                     1
dtype: int64
```

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

In [120…
```
#Group the dataset by fatal flag
df_fatalflag_yes = df_new_modified_new[df_new_modified_new['FATAL_FLAG']=='Yes']
```

In [121…
```
#View the total number of fatal and non-fatal accidents
df_fatalflag_no = df_new_req[df_new_req['FATAL_FLAG']=='No']
print("No of Fatal Accidents: ", df_fatalflag_yes.shape[0])
print("No of Non-Fatal Accidents: ", df_fatalflag_no.shape[0])
```

```
No of Fatal Accidents:  7
No of Non-Fatal Accidents:  75
```

In [122…
```
#Create a new dataframe to view only the fatal accidents (Fatal Flag values = Yes)
df_fatalflag_yes
```

Out[122…

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_T
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	
53	PIPER	Florida	PA28	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES. ...	UNKNOWN (UNK)	
55	FLIGHT DESIGN	California	CTLS	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES A...	UNKNOWN (UNK)	

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TY
79	NORTH AMERICAN	Arizona	F51	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOWN (UNK)	
80	CHAMPION	California	8KCAB	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	
81	BEECH	California	35	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	
82	CESSNA	Alabama	182	N784CP AIRCRAFT CRASHED INTO A WOODED AREA NEA...	UNKNOWN (UNK)	

In [ ]: