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 [139]:
```
#Import necessary libraries
import numpy as np
import pandas as pd
```

In [140]:
```
#Import the FAA (Federal Aviation Authority) dataset
faa = pd.read_csv('../faa_ai_prelim/faa_ai_prelim.csv')
```

2: View and understand the dataset

In [141]:
```
#View the dataset shape
faa.shape
```

Out[141]: `(83, 42)`

In [142]:
```
#View the first five observations
faa.head(5)
```

Out[142]:

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LOC_CNTRY_NAME	RMK_TEX
0	No	19-FEB-16	19-FEB-16	00:45:00Z	MARSHVILLE	North Carolina	NaN	AIRCRAF CRASHE INTO TREES THE PERSON O B.
1	No	19-FEB-16	18-FEB-16	23:55:00Z	TAVERNIER	Florida	NaN	AIRCRAF ON LANDIN WENT OF THE END O THE RU.
2	No	19-FEB-16	18-FEB-16	22:14:00Z	TRENTON	New Jersey	NaN	AIRCRAF ON FINA SUSTAINE A BIR STRIKE LAN.
3	No	19-FEB-16	18-FEB-16	17:10:00Z	ASHEVILLE	North Carolina	NaN	AIRCRAF O LANDIN GEA COLLAPSED ASHEVILLE.
4	No	19-FEB-16	18-FEB-16	00:26:00Z	TALKEETNA	Alaska	NaN	AIRCRAF O LANDIN NOSE GEA COLLAPSED TALK.

5 rows × 42 columns

In [143]:
```
#View all the columns present in the dataset
faa.columns
```

Out[143]:
```
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')
```

In [144]:
```
# Aircraft make name = ACFT_MAKE_NAME
# State name = LOC_STATE_NAME
# Aircraft model name = ACFT_MODEL_NAME
# RMK_TEXT = information
# Flight phase = FLT_PHASE
# Event description type =EVENT_TYPE_DESC
# Fatal flag = FATAL_FLAG
```

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 [145]:
```
#Create a new dataframe with only the required columns
df_faa=faa[['ACFT_MAKE_NAME','LOC_STATE_NAME','ACFT_MODEL_NAME','RMK_TEXT','FLT_PHASE','EVENT_TYPE_DES
C','FATAL_FLAG']]
```

In [154]:
```
#View the type of the object
type(df_faa)
```

Out[154]: `pandas.core.frame.DataFrame`

In [147]:
```
#Check if the dataframe contains all the required attributes
print('the number of columns is '+ str(df_faa.columns.value_counts().sum()))
print(df_faa.columns)
```

```
the number of columns is 7
Index(['ACFT_MAKE_NAME', 'LOC_STATE_NAME', 'ACFT_MODEL_NAME', 'RMK_TEXT',
       'FLT_PHASE', 'EVENT_TYPE_DESC', 'FATAL_FLAG'],
      dtype='object')
```

In [155]:
```
df_faa.head()
```

Out[155]:

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DESC	FATAL_FLAG
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accident	Yes
1	VANS	Florida	RV7	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDING (LDG)	Incident	No
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incident	No
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incident	No
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incident	No

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

In [156]:
```
#Replace all Fatal Flag missing values with the required output
df_faa['FATAL_FLAG'].fillna(value='No', inplace = True)
```

```
C:\Users\Mohannad\Anaconda3\lib\site-packages\pandas\core\generic.py:6287: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/indexin
g.html#returning-a-view-versus-a-copy
  self._update_inplace(new_data)
```

In [160]:
```
#Verify if the missing values are replaced
df_faa
```

Out[160]:

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DESC	FATAL_FLAG
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accident	Yes
1	VANS	Florida	RV7	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDING (LDG)	Incident	No
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incident	No
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incident	No
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incident	No
...	...	...	...	...	...	...	...
78	AERONCA	Texas	O58B	AIRCRAFT ON LANDING, GROUND LOOPED, BULVERDE A...	LANDING (LDG)	Accident	No
79	NORTH AMERICAN	Arizona	F51	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOWN (UNK)	Accident	Yes
80	CHAMPION	California	8KCAB	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	Accident	Yes
81	BEECH	California	35	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	Accident	Yes
82	CESSNA	Alabama	182	N784CP AIRCRAFT CRASHED INTO A WOODED AREA NEA...	UNKNOWN (UNK)	Accident	Yes

83 rows × 7 columns

In [159]:
```
df_faa['FATAL_FLAG'].isnull().sum()
```

Out[159]: `0`

In [150]:
```
#Check the number of observations
df_faa.shape
```

Out[150]: `(83, 7)`

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

In [162]:
```
#Drop the unwanted values/observations from the dataset
df_faa.dropna(subset = ['ACFT_MAKE_NAME'], inplace=True);
```

```
C:\Users\Mohannad\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/indexin
g.html#returning-a-view-versus-a-copy
```

In [163]:
```
df_faa.ACFT_MAKE_NAME.values
```

Out[163]:
```
array(['BEECH', 'VANS', 'CESSNA', 'LANCAIR', 'CESSNA', 'BELL', 'PIPER',
       'CESSNA', 'BOEING', 'CESSNA', 'CESSNA', 'CESSNA', 'BEECH',
       'CESSNA', 'CHRISTEN', 'PIPER', 'PIPER', 'CESSNA', 'HUGHES',
       'CESSNA', 'BOEING', 'GLOBE', 'PIPER', 'GREAT LAKES', 'CESSNA',
       'CONSOLIDATED VULTEE', 'CESSNA', 'BEECH', 'CESSNA', 'FAIRCHILD',
       'SOCATA', 'SAAB', 'EMBRAER', 'MAULE', 'ENSTROM', 'BEECH', 'MOONEY',
       'PITTS', 'CESSNA', 'CESSNA', 'CESSNA', 'AEROSTAR INTERNATIONAL',
       'CESSNA', 'CHAMPION', 'BELL', 'AIRBUS', 'GRUMMAN', 'GULFSTREAM',
       'SOCATA', 'CESSNA', 'PIPER', 'CESSNA', 'FLIGHT DESIGN', 'PIPER',
       'PIPER', 'PIPER', 'BEECH', 'BEECH', 'CESSNA', 'CESSNA', 'CESSNA',
       'PIPER', 'LANCAIR', 'BEECH', 'MOONEY', 'AERO COMMANDER', 'CESSNA',
       'MOONEY', 'BEECH', 'SABRELINER', 'MOONEY', 'PIPER', 'BOEING',
       'AERONCA', 'NORTH AMERICAN', 'CHAMPION', 'BEECH', 'CESSNA'],
      dtype=object)
```

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

In [164]:
```
#Check the number of observations now to compare it with the original dataset and see how many values
 have been dropped
df_faa.shape
```

Out[164]: `(78, 7)`

In [165]:
```
#Group the dataset by aircraft name
aircraft= df_faa.groupby('ACFT_MAKE_NAME')
```

In [166]:
```
#View the number of times each aircraft type appears in the dataset (Hint: use the size() method)
aircraft.size()
```

Out[166]:
```
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
PIPER                     10
PITTS                      1
SAAB                       1
SABRELINER                 1
SOCATA                     2
VANS                       1
dtype: int64
```

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

In [167]:
```
#Group the dataset by fatal flag
fatal= df_faa.groupby('FATAL_FLAG')
```

In [168]:
```
#View the total number of fatal and non-fatal accidents
fatal.size()
```

Out[168]:
```
FATAL_FLAG
No     71
Yes     7
dtype: int64
```

In [183]:
```
#Create a new dataframe to view only the fatal accidents (Fatal Flag values = Yes)
yes = fatal.get_group('Yes')
```

In [184]:
```
yes.head()
```

Out[184]:

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DESC	FATAL_FLAG
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accident	Yes
53	PIPER	Florida	PA28	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES. ...	UNKNOWN (UNK)	Accident	Yes
55	FLIGHT DESIGN	California	CTLS	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES A...	UNKNOWN (UNK)	Accident	Yes
79	NORTH AMERICAN	Arizona	F51	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOWN (UNK)	Accident	Yes
80	CHAMPION	California	8KCAB	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	Accident	Yes