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: Vlew and import the dataset

In [1]:
```
#Import necessary libraries
import pandas as pd
```

In [15]:
```
#Import the FAA (Federal Aviation Authority) dataset
df_faa_dataset = pd.read_csv(r'C:\Users\ctoqu\Desktop\faa_ai_prelim\faa_ai_prelim.csv', engine= 'python')
```

2: View and understand the dataset

In [16]:
```
#View the dataset shape
df_faa_dataset.shape
```

Out[16]: (83, 42)

In [18]:
```
#View the first five observations
df_faa_dataset.head()
```

Out[18]:

STATE_NAME	LOC_CNTRY_NAME	RMK_TEXT	EVENT_TYPE_DESC	FSDO_DESC	...	PAX_INJ_NONE	PAX_INJ_MINOR	PAX_INJ_SERIOUS
North Carolina	NaN	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	Accident	FAA Charlotte FSDO-68	...	NaN	NaN	NaN
Florida	NaN	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	Incident	FAA Miami FSDO-19	...	NaN	NaN	NaN
New Jersey	NaN	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	Incident	FAA Philadelphia FSDO-17	...	NaN	NaN	NaN
North Carolina	NaN	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	Incident	FAA Charlotte FSDO-68	...	NaN	NaN	NaN
Alaska	NaN	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	Incident	FAA Anchorage FSDO-03	...	NaN	1.0	NaN

In [20]:
```
#View all the columns present in the dataset
df_faa_dataset.columns
```

Out[20]:
```
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 [22]:
```
#Create a new dataframe with only the required columns
df_analyze_dataset=df_faa_dataset[['ACFT_MAKE_NAME', 'LOC_STATE_NAME', 'ACFT_MODEL_NAME', 'RMK_TEXT','FLT_PHASE'
                                   'EVENT_TYPE_DESC','FATAL_FLAG']]
```

In [23]:
```
#View the type of the object
type(df_analyze_dataset)
```

Out[23]:
```
pandas.core.frame.DataFrame
```

In [24]:
```
#Check if the dataframe contains all the required attributes
df_analyze_dataset.head()
```

Out[24]:

	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	NaN
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incident	NaN
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incident	NaN
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incident	NaN

In [25]:
```
```

```
C:\Users\ctoqu\anaconda3\lib\site-packages\pandas\core\generic.py:6245: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

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

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

In [26]:
```
#Replace all Fatal Flag missing values with the required output
df_analyze_dataset['FATAL_FLAG'].fillna(value='No',inplace=True)
```

In [28]:
```
#Verify if the missing values are replaced
df_analyze_dataset.head()
```

Out[28]:

	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

In [30]:
```
#Check the number of observations
df_analyze_dataset.shape
```

Out[30]: (83, 7)

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

In [31]:
```
#Drop the unwanted values/observations from the dataset
#Remove all observations with aircraft names are not available

df_final_dataset = df_analyze_dataset.dropna(subset=['ACFT_MAKE_NAME'])
```

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

In [32]:
```
#Check the number of observations now to compare it with the original dataset and see how many values have been
df_final_dataset.shape
```

Out[32]: (78, 7)

In [33]:
```
#Group the dataset by aircraft name
aircraftType =df_final_dataset.groupby('ACFT_MAKE_NAME')
```

In [35]:
```
#View the number of times each aircraft type appears in the dataset (Hint: use the size() method)
aircraftType.size()
```

Out[35]:
```
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 [39]:
```
#Group the dataset by fatal flag
fatalAccidents = df_final_dataset.groupby('FATAL_FLAG')
```

In [40]:
```
#View the total number of fatal and non-fatal accidents
fatalAccidents.size()
```

Out[40]:
```
FATAL_FLAG
No     71
Yes     7
dtype: int64
```

In [41]:
```
#Create a new dataframe to view only the fatal accidents (Fatal Flag values = Yes)
accidents_with_fatality = fatalAccidents.get_group('Yes')
```

In [42]:
```
accidents_with_fatality
```

Out[42]:

	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
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

In [ ]:
```
```