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 [2]:
```
#Import necessary libraries
import pandas as pd
```

In [6]:
```
#Import the FAA (Federal Aviation Authority) dataset
df = pd.read_csv('faa_ai_prelim/faa_ai_prelim.csv')
```

2: View and understand the dataset

In [57]:
```
#View the dataset shape
df.shape
```

Out[57]:
```
(83, 42)
```

In [58]:
```
#View the first five observations
df.head()
```

Out[58]:

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STAT
0	No	19-FEB-16	19-FEB-16	00:45:00Z	MARSHVILLE	North
1	No	19-FEB-16	18-FEB-16	23:55:00Z	TAVERNIER	
2	No	19-FEB-16	18-FEB-16	22:14:00Z	TRENTON	Ne
3	No	19-FEB-16	18-FEB-16	17:10:00Z	ASHEVILLE	North
4	No	19-FEB-16	18-FEB-16	00:26:00Z	TALKEETNA	

5 rows × 42 columns

In [59]:
```
#View all the columns present in the dataset
df.columns
```

Out[59]:
```
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_L
VL',
       '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_MIN
OR',
       'PAX_INJ_SERIOUS', 'PAX_INJ_FATAL', 'PAX_INJ_UNK', 'GRND_INJ_NONE',
       'GRND_INJ_MINOR', 'GRND_INJ_SERIOUS', 'GRND_INJ_FATAL', 'GRND_INJ_UN
K'],
      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 [60]:
```
#Create a new dataframe with only the required columns
req_df = df[['ACFT_MAKE_NAME','LOC_STATE_NAME','ACFT_MODEL_NAME','RMK_TEXT
req_df
```

Out[60]:

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHAS
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOW (UN
1	VANS	Florida	RV7	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDIN (LD
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROAC (AP
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDIN (LD
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDIN (LD
...	...	...	...	...	
78	AERONCA	Texas	O58B	AIRCRAFT ON LANDING, GROUND LOOPED, BULVERDE A...	LANDIN (LD
79	NORTH AMERICAN	Arizona	F51	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOW (UN
80	CHAMPION	California	8KCAB	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOW (UN
81	BEECH	California	35	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOW (UN
82	CESSNA	Alabama	182	N784CP AIRCRAFT CRASHED INTO A WOODED AREA NEA...	UNKNOW (UN

83 rows × 7 columns

In [33]:
```
#View the type of the object
req_df.dtypes
```

Out[33]:
```
ACFT_MAKE_NAME      object
LOC_STATE_NAME      object
ACFT_MODEL_NAME     object
RMK_TEXT            object
FLT_PHASE           object
EVENT_TYPE_DESC     object
FATAL_FLAG          object
dtype: object
```

In [61]:
```
#Check if the dataframe contains all the required attributes
req_df.columns
```

Out[61]:
```
Index(['ACFT_MAKE_NAME', 'LOC_STATE_NAME', 'ACFT_MODEL_NAME', 'RMK_TEXT',
       'FLT_PHASE', 'EVENT_TYPE_DESC', 'FATAL_FLAG'],
      dtype='object')
```

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

In [64]:
```
#Replace all Fatal Flag missing values with the required output
req_df[['FATAL_FLAG']] = req_df[['FATAL_FLAG']].fillna(value= 'NO')
req_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 83 entries, 0 to 82
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   ACFT_MAKE_NAME   78 non-null     object
 1   LOC_STATE_NAME   83 non-null     object
 2   ACFT_MODEL_NAME  79 non-null     object
 3   RMK_TEXT         83 non-null     object
 4   FLT_PHASE        82 non-null     object
 5   EVENT_TYPE_DESC  83 non-null     object
 6   FATAL_FLAG       83 non-null     object
dtypes: object(7)
memory usage: 4.7+ KB
```

In [66]:
```
#Verify if the missing values are replaced
req_df.FATAL_FLAG
```

Out[66]:
```
0     Yes
1      NO
2      NO
3      NO
4      NO
     ...
78     NO
79    Yes
80    Yes
81    Yes
82    Yes
Name: FATAL_FLAG, Length: 83, dtype: object
```

In [70]:
```
#Check the number of observations
```

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

In [69]:
```
#Drop the unwanted values/observations from the dataset
req_df.dropna(subset=['ACFT_MAKE_NAME'],inplace = True )
req_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 78 entries, 0 to 82
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   ACFT_MAKE_NAME    78 non-null     object
 1   LOC_STATE_NAME    78 non-null     object
 2   ACFT_MODEL_NAME   77 non-null     object
 3   RMK_TEXT          78 non-null     object
 4   FLT_PHASE         77 non-null     object
 5   EVENT_TYPE_DESC   78 non-null     object
 6   FATAL_FLAG        78 non-null     object
dtypes: object(7)
memory usage: 4.9+ KB
```

```
<ipython-input-69-16832a57904b>:2: 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
  req_df.dropna(subset=['ACFT_MAKE_NAME'],inplace = True )
```

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

In [71]:
```
#Check the number of observations now to compare it with the original datas
req_df.shape
```

Out[71]:
```
(78, 7)
```

In [74]:
```
#Group the dataset by aircraft name
group_aircraft_name = req_df.groupby(['ACFT_MAKE_NAME'])
group_aircraft_name
```

Out[74]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x11ae42f40>
```

In [75]:
```
#View the number of times each aircraft type appears in the dataset (Hint:
group_aircraft_name.size()
```

Out[75]:
```
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 [79]:
```
#Group the dataset by fatal flag
fatal = req_df.groupby(['FATAL_FLAG'])
```

In [80]:
```
#View the total number of fatal and non-fatal accidents
fatal.size()
```

Out[80]:
```
FATAL_FLAG
NO     71
Yes     7
dtype: int64
```

In [83]:
```
#Create a new dataframe to view only the fatal accidents (Fatal Flag value
fatal_DF =fatal.get_group('Yes')
```

In [84]:
```
fatal_DF
```

Out[84]:

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHAS
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOW (UN
53	PIPER	Florida	PA28	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES. ...	UNKNOW (UN
55	FLIGHT DESIGN	California	CTLS	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES A...	UNKNOW (UN
79	NORTH AMERICAN	Arizona	F51	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOW (UN
80	CHAMPION	California	8KCAB	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOW (UN
81	BEECH	California	35	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOW (UN
82	CESSNA	Alabama	182	N784CP AIRCRAFT CRASHED INTO A WOODED AREA NEA...	UNKNOW (UN

In [ ]: