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!

Analyse the Federal Aviation Authority Dataset using Pandas

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"

1: View and import the dataset

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

In [2]:
```
#Import the FAA (Federal Aviation Authority) dataset
df_FAA_dataset = pd.read_csv('D:\\NIPUN_SC_REC\\3_Practice_Project\Course_5_Data Science with Python\\Practice_
```

2: View and understand the dataset

In [3]:
```
#View the dataset shape
df_FAA_dataset.shape
```

Out[3]:
```
(83, 42)
```

In [4]:
```
#View the first five observations
df_FAA_dataset.head()
```

Out[4]:

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LOC_CNTRY_NAME	RMK_TEXT	EVENT_TY
0	No	19-FEB-16	19-FEB-16	00:45:00Z	MARSHVILLE	North Carolina	NaN	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	
1	No	19-FEB-16	18-FEB-16	23:55:00Z	TAVERNIER	Florida	NaN	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	
2	No	19-FEB-16	18-FEB-16	22:14:00Z	TRENTON	New Jersey	NaN	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	
3	No	19-FEB-16	18-FEB-16	17:10:00Z	ASHEVILLE	North Carolina	NaN	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	
4	No	19-FEB-16	18-FEB-16	00:26:00Z	TALKEETNA	Alaska	NaN	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	

5 rows × 42 columns

In [5]:
```
#View all the columns present in the dataset
df_FAA_dataset.columns
```

Out[5]:
```
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 [6]:
```
#Create a new dataframe with only the required columns
df_analyse_dataset=df_FAA_dataset[['ACFT_MAKE_NAME','LOC_STATE_NAME','ACFT_MODEL_NAME','RMK_TEXT','FLT_PHASE','
```

In [7]:
```
#View the type of the object
type(df_analyse_dataset)
```

Out[7]:
```
pandas.core.frame.DataFrame
```

In [8]:
```
#Check if the dataframe contains all the required attributes
df_analyse_dataset
```

Out[8]:

	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
...	...	...	...	...	...	...	...
78	AERONCA	Texas	O58B	AIRCRAFT ON LANDING, GROUND LOOPED, BULVERDE A...	LANDING (LDG)	Accident	NaN
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

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

In [9]:
```
#Replace all Fatal Flag missing values with the required output
df_analyse_dataset['FATAL_FLAG'].fillna(value='NO',inplace=True)
```

```
C:\Users\Sohaib\anaconda3\lib\site-packages\pandas\core\series.py:4463: 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#ret
urning-a-view-versus-a-copy
  return super().fillna(
```

In [10]:
```
#Verify if the missing values are replaced
df_analyse_dataset
```

Out[10]:

	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 [11]:
```
#Check the number of observations
df_analyse_dataset.shape
```

Out[11]:
```
(83, 7)
```

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

In [12]:
```
#Drop the unwanted values/observations from the dataset
df_analyse_dataset['ACFT_MAKE_NAME']
```

Out[12]:
```
0              BEECH
1               VANS
2             CESSNA
3            LANCAIR
4             CESSNA
           ...
78           AERONCA
79    NORTH AMERICAN
80          CHAMPION
81             BEECH
82            CESSNA
Name: ACFT_MAKE_NAME, Length: 83, dtype: object
```

In [13]:
```
df_final_dataset = df_analyse_dataset.dropna(subset=['ACFT_MAKE_NAME'])
df_final_dataset
```

Out[13]:

	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

78 rows × 7 columns

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

In [14]:
```
#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[14]:
```
(78, 7)
```

In [15]:
```
#Group the dataset by aircraft name
aircraftname = df_final_dataset.groupby('ACFT_MAKE_NAME')
```

In [16]:
```
#View the number of times each aircraft type appears in the dataset (Hint: use the size() method)
aircraftname.size()
```

Out[16]:
```
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 [17]:
```
#Group the dataset by fatal flag
fatal_flag = df_final_dataset.groupby('FATAL_FLAG')
```

In [18]:
```
#View the total number of fatal and non-fatal accidents
fatal_flag.size()
```

Out[18]:
```
FATAL_FLAG
NO     71
Yes     7
dtype: int64
```

In [19]:
```
#Create a new dataframe to view only the fatal accidents (Fatal Flag values = Yes)
fatal_accidents=fatal_flag.get_group('Yes')
fatal_accidents
```

Out[19]:

	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 [20]:
```
fatal_accidents.shape
```

Out[20]:
```
(7, 7)
```