

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 [1]:

```
#Import necessary libraries
import pandas as pd
```

In [14]:

```
#Import the FAA (Federal Aviation Authority) dataset
df=pd.read_csv("faa_ai_prelim.csv")
df
```

Out[14]:

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LOC_CNTRY_NAME
0	No	19-FEB-16	19-FEB-16	00:45:00Z	MARSHVILLE	North Carolina	NaN
1	No	19-FEB-16	18-FEB-16	23:55:00Z	TAVERNIER	Florida	NaN
2	No	19-FEB-16	18-FEB-16	22:14:00Z	TRENTON	New Jersey	NaN
3	No	19-FEB-16	18-FEB-16	17:10:00Z	ASHEVILLE	North Carolina	NaN
4	No	19-FEB-16	18-FEB-16	00:26:00Z	TALKEETNA	Alaska	NaN
...
78	No	08-FEB-16	31-DEC-15	17:00:00Z	SAN ANTONIO	Texas	NaN
79	No	08-FEB-16	05-FEB-16	11:17:00Z	MARICOPA	Arizona	NaN

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LOC_CNTRY_NAME
--	---------	------------	----------------	----------------	---------------	----------------	----------------

81	No	08-FEB-16	05-FEB-16	23:02:00Z	SAN PEDRO	California	NaN
----	----	-----------	-----------	-----------	-----------	------------	-----

82	Yes	02-FEB-16	02-FEB-16	01:52:00Z	MOBILE	Alabama	NaN
----	-----	-----------	-----------	-----------	--------	---------	-----

83 rows x 42 columns

2: View and understand the dataset

In [15]:

```
#View the dataset shape
df.shape
```

Out[15]:

(83, 42)

In [16]:

```
#View the first five observations
df.head()
```

Out[16]:

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LOC_CNTRY_NAME
--	---------	------------	----------------	----------------	---------------	----------------	----------------

0	No	19-FEB-16	19-FEB-16	00:45:00Z	MARSHVILLE	North Carolina	NaN
---	----	-----------	-----------	-----------	------------	----------------	-----

1	No	19-FEB-16	18-FEB-16	23:55:00Z	TAVERNIER	Florida	NaN
---	----	-----------	-----------	-----------	-----------	---------	-----

2	No	19-FEB-16	18-FEB-16	22:14:00Z	TRENTON	New Jersey	NaN
---	----	-----------	-----------	-----------	---------	------------	-----

3	No	19-FEB-16	18-FEB-16	17:10:00Z	ASHEVILLE	North Carolina	NaN
---	----	-----------	-----------	-----------	-----------	----------------	-----

4	No	19-FEB-16	18-FEB-16	00:26:00Z	TALKEETNA	Alaska	NaN
---	----	-----------	-----------	-----------	-----------	--------	-----

5 rows x 42 columns

In [17]:

```
#View all the columns present in the dataset
df.columns
```

Out[17]:

```
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 [19]:

```
#Create a new dataframe with only the required columns
newdatadf=df[["ACFT_MAKE_NAME","LOC_STATE_NAME","ACFT_MODEL_NAME","RMK_TEXT","FLT_PHASE",
,"EVENT_TYPE_DESC","FATAL_FLAG"]]
newdatadf
```

Out[19]:

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DESC	FATA
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accident	
1	VANS	Florida	RV7	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDING (LDG)	Incident	
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incident	
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incident	
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incident	
...
78	AERONCA	Texas	O58B	AIRCRAFT ON LANDING, GROUND LOOPED, BULVERDE A...	LANDING (LDG)	Accident	

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

83 rows x 7 columns



```
In [20]:  
  
#View the type of the object  
type(newdatadf)  
  
Out[20]:  
  
pandas.core.frame.DataFrame  
  
In [21]:  
  
#Check if the dataframe contains all the required attributes  
newdatadf.columns  
  
Out[21]:  
  
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 [22]:  
  
#Replace all Fatal Flag missing values with the required output  
newdatadf["FATAL_FLAG"].fillna(value="NO",inplace=True)  
  
c:\users\dell\appdata\local\programs\python\python39\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_g  
uide/indexing.html#returning-a-view-versus-a-copy  
return super().fillna()
```

```
In [23]:  
  
#Verify if the missing values are replaced  
newdatadf  
  
Out[23]:
```

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

AIRCRAFT ON

1	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DESC	FATAL_FLAG
				LANDING WENT OFF THE END OF THE RU...	LANDING (LDG)		
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incident	
3	LANCAIR	North Carolina	235	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incident	
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incident	
...
78	AERONCA	Texas	O58B	AIRCRAFT ON LANDING, GROUND LOOPED, BULVERDE A...	LANDING (LDG)	Accident	
79	NORTH AMERICAN	Arizona	F51	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOWN (UNK)	Accident	
80	CHAMPION	California	8KCAB	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	Accident	
81	BEECH	California	35	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	Accident	
82	CESSNA	Alabama	182	N784CP AIRCRAFT CRASHED INTO A WOODED AREA NEA...	UNKNOWN (UNK)	Accident	

83 rows x 7 columns



In [24]:

```
#Check the number of observations
newdatadf.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	NO
2	CESSNA	New Jersey	172	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH (APR)	Incident	NO

ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT AIRCRAFT	FLT_PHASE	EVENT_TYPE_DESC	FATAL_FLAG	
3	LANCAIR	North Carolina	235	ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING (LDG)	Incident	NO
4	CESSNA	Alaska	172	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING (LDG)	Incident	NO

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

In [26]:

```
#Drop the unwanted values/observations from the dataset
newdatadf=newdatadf.dropna(subset=["ACFT_MAKE_NAME"])
```

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

In [27]:

```
#Check the number of observations now to compare it with the original dataset and see how many values have been dropped
newdatadf.shape
```

Out[27]:

(78, 7)

In [28]:

```
#Group the dataset by aircraft name
aircraftftype=newdatadf.groupby("ACFT_MAKE_NAME")
```

In [29]:

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

Out[29]:

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 [30]:

```
#Group the dataset by fatal flag
fatalaccidents=newdatadf.groupby("FATAL_FLAG")
```

In [31]:

```
#View the total number of fatal and non-fatal accidents
fatalaccidents.size()
```

Out[31]:

FATAL_FLAG
NO 71
Yes 7
dtype: int64

In [34]:

```
#Create a new dataframe to view only the fatal accidents (Fatal Flag values = Yes)
fatalityaccidents=fatalaccidents.get_group("Yes")
```

In [35]:

```
#view the accidents with fatality
fatalityaccidents
```

Out[35]:

	ACFT_MAKE_NAME	LOC_STATE_NAME	ACFT_MODEL_NAME	RMK_TEXT	FLT_PHASE	EVENT_TYPE_DESC	FATA
0	BEECH	North Carolina	36	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN (UNK)	Accident	
53	PIPER	Florida	PA28	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES. ...	UNKNOWN (UNK)	Accident	
55	FLIGHT DESIGN	California	CTLS	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES A...	UNKNOWN (UNK)	Accident	
79	NORTH AMERICAN	Arizona	F51	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOWN (UNK)	Accident	
80	CHAMPION	California	8KCAB	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN (UNK)	Accident	
81	BEECH	California	35	N9872R, BEECH M35 AIRCRAFT, AND N5057G,	UNKNOWN (UNK)	Accident	

