

Assignment_07_01_Completed

July 19, 2020

1 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

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

```
[4]: #Import the FAA (Federal Aviation Authority) dataset
FAA_DataFrame = pd.read_csv("faa_ai_prelim.csv")
FAA_DataFrame.describe()
```

```
[4]:
```

	ACFT_MISSING_FLAG	FAR_PART	FLT_CRW_INJ_NONE	FLT_CRW_INJ_MINOR	\
count	0.0	1.0	64.000000	6.000000	
mean	NaN	91.0	1.218750	1.333333	
std	NaN	NaN	0.700765	0.516398	
min	NaN	91.0	1.000000	1.000000	
25%	NaN	91.0	1.000000	1.000000	
50%	NaN	91.0	1.000000	1.000000	
75%	NaN	91.0	1.000000	1.750000	
max	NaN	91.0	6.000000	2.000000	

	FLT_CRW_INJ_SERIOUS	FLT_CRW_INJ_FATAL	FLT_CRW_INJ_UNK	\
count	1.0	8.000000	0.0	
mean	3.0	1.500000	NaN	
std	NaN	0.534522	NaN	
min	3.0	1.000000	NaN	
25%	3.0	1.000000	NaN	

50%	3.0	1.500000	NaN
75%	3.0	2.000000	NaN
max	3.0	2.000000	NaN

	CBN_CRW_INJ_NONE	CBN_CRW_INJ_MINOR	CBN_CRW_INJ_SERIOUS	...	\
count	0.0	0.0	0.0	...	
mean	NaN	NaN	NaN	...	
std	NaN	NaN	NaN	...	
min	NaN	NaN	NaN	...	
25%	NaN	NaN	NaN	...	
50%	NaN	NaN	NaN	...	
75%	NaN	NaN	NaN	...	
max	NaN	NaN	NaN	...	

	PAX_INJ_NONE	PAX_INJ_MINOR	PAX_INJ_SERIOUS	PAX_INJ_FATAL	\
count	17.000000	1.0	3.0	1.0	
mean	1.823529	1.0	1.0	1.0	
std	1.236694	NaN	0.0	NaN	
min	1.000000	1.0	1.0	1.0	
25%	1.000000	1.0	1.0	1.0	
50%	2.000000	1.0	1.0	1.0	
75%	2.000000	1.0	1.0	1.0	
max	6.000000	1.0	1.0	1.0	

	PAX_INJ_UNK	GRND_INJ_NONE	GRND_INJ_MINOR	GRND_INJ_SERIOUS	\
count	0.0	0.0	0.0	0.0	
mean	NaN	NaN	NaN	NaN	
std	NaN	NaN	NaN	NaN	
min	NaN	NaN	NaN	NaN	
25%	NaN	NaN	NaN	NaN	
50%	NaN	NaN	NaN	NaN	
75%	NaN	NaN	NaN	NaN	
max	NaN	NaN	NaN	NaN	

	GRND_INJ_FATAL	GRND_INJ_UNK
count	0.0	1.0
mean	NaN	2.0
std	NaN	NaN
min	NaN	2.0
25%	NaN	2.0
50%	NaN	2.0
75%	NaN	2.0
max	NaN	2.0

[8 rows x 22 columns]

2: View and understand the dataset

```
[5]: #View the dataset shape
```

```
FAA_DataFrame.shape
```

```
[5]: (83, 42)
```

```
[6]: #View the first five observations
```

```
FAA_DataFrame.head(5)
```

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

```
   LOC_STATE_NAME LOC_CNTRY_NAME \
0  North Carolina           NaN
1         Florida           NaN
2    New Jersey           NaN
3  North Carolina           NaN
4         Alaska           NaN
```

```
   RMK_TEXT EVENT_TYPE_DESC \
0  AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...    Accident
1  AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...    Incident
2  AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...    Incident
3  AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...    Incident
4  AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...    Incident
```

```
   FSDO_DESC ... PAX_INJ_NONE PAX_INJ_MINOR PAX_INJ_SERIOUS \
0  FAA Charlotte FSDO-68 ...      NaN      NaN      NaN
1  FAA Miami FSDO-19 ...      NaN      NaN      NaN
2  FAA Philadelphia FSDO-17 ...      NaN      NaN      NaN
3  FAA Charlotte FSDO-68 ...      NaN      NaN      NaN
4  FAA Anchorage FSDO-03 ...      NaN      1.0      NaN
```

```
   PAX_INJ_FATAL PAX_INJ_UNK GRND_INJ_NONE GRND_INJ_MINOR GRND_INJ_SERIOUS \
0           NaN      NaN      NaN      NaN      NaN
1           NaN      NaN      NaN      NaN      NaN
2           NaN      NaN      NaN      NaN      NaN
3           NaN      NaN      NaN      NaN      NaN
4           NaN      NaN      NaN      NaN      NaN
```

```
   GRND_INJ_FATAL GRND_INJ_UNK
0           NaN      NaN
```

1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

[5 rows x 42 columns]

```
[7]: #View all the columns present in the dataset
FAA_DataFrame.columns
```

```
[7]: 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

```
[9]: #Create a new dataframe with only the required columns
FAA_New_df =
    ↪FAA_DataFrame[['ACFT_MAKE_NAME', 'LOC_STATE_NAME', 'ACFT_MODEL_NAME', 'RMK_TEXT', 'FLT_PHASE', 'FATAL_FLAG']]
FAA_New_df
```

```
[9]:   ACFT_MAKE_NAME  LOC_STATE_NAME  ACFT_MODEL_NAME  \
0          BEECH  North Carolina             36
1          VANS   Florida             RV7
2        CESSNA   New Jersey            172
3        LANCAIR  North Carolina            235
4          CESSNA     Alaska            172
..          ...             ...             ...
78        AERONCA     Texas            058B
79  NORTH AMERICAN  Arizona            F51
```

80	CHAMPION	California	8KCAB
81	BEECH	California	35
82	CESSNA	Alabama	182

		RMK_TEXT	FLT_PHASE \
0	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...	UNKNOWN	(UNK)
1	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU...	LANDING	(LDG)
2	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN...	APPROACH	(APR)
3	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE...	LANDING	(LDG)
4	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK...	LANDING	(LDG)
..
78	AIRCRAFT ON LANDING, GROUND LOOPED, BULVERDE A...	LANDING	(LDG)
79	AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...	UNKNOWN	(UNK)
80	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN	(UNK)
81	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	UNKNOWN	(UNK)
82	N784CP AIRCRAFT CRASHED INTO A WOODED AREA NEA...	UNKNOWN	(UNK)

	EVENT_TYPE_DESC	FATAL_FLAG
0	Accident	Yes
1	Incident	NaN
2	Incident	NaN
3	Incident	NaN
4	Incident	NaN
..
78	Accident	NaN
79	Accident	Yes
80	Accident	Yes
81	Accident	Yes
82	Accident	Yes

[83 rows x 7 columns]

```
[10]: #View the type of the object
type(FAA_New_df)
```

```
[10]: pandas.core.frame.DataFrame
```

```
[11]: #Check if the dataframe contains all the required attributes
FAA_New_df.columns
```

```
[11]: 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”

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

```
/usr/local/lib/python3.7/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
```

```
self._update_inplace(new_data)
```

```
[19]: #Verify if the missing values are replaced
FAA_New_df.groupby('FATAL_FLAG').describe()
```

```
[19]:
```

	ACFT_MAKE_NAME			LOC_STATE_NAME \		
	count	unique	top freq	count	unique	
FATAL_FLAG						
No	71	28	CESSNA	22	75	32
Yes	7	6	BEECH	2	8	5

	ACFT_MODEL_NAME			... \		
	top freq	count	unique	...		
FATAL_FLAG						
No	California	13	72	52	...	
Yes	California	3	7	7	...	

	RMK_TEXT		FLT_PHASE \	
	top freq	count		
FATAL_FLAG				
No	AIRCRAFT, N704SL CESSNA 150, AND N57BC AVIAT ...	2	74	
Yes	N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...	2	8	

	EVENT_TYPE_DESC					
	unique	top freq	count	unique	top freq	
FATAL_FLAG						
No	6	LANDING (LDG)	47	75	2	Incident 56
Yes	2	UNKNOWN (UNK)	7	8	1	Accident 8

[2 rows x 24 columns]

```
[17]: #Check the number of observations
FAA_New_df.shape
```

```
[17]: (83, 7)
```

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

```
[23]: #Drop the unwanted values/observations from the dataset
df_FAA_final = FAA_New_df.dropna(subset=['ACFT_MAKE_NAME'])
```

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

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

```
[24]: (78, 7)
```

```
[26]: #Group the dataset by aircraft name
aircraft_type = df_FAA_final.groupby('ACFT_MAKE_NAME')
aircraft_type
```

```
[26]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f2b09f2e990>
```

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

```
[27]: 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”

```
[28]: #Group the dataset by fatal flag
fatal_flag_type = df_FAA_final.groupby('FATAL_FLAG')
```

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

```
[29]: FATAL_FLAG
No      71
Yes      7
dtype: int64
```

```
[31]: #Create a new dataframe to view only the fatal accidents (Fatal Flag values = 1
      ↪ Yes)
df_fatal_Accidents_faa = fatal_flag_type.get_group('Yes')
df_fatal_Accidents_faa
```

```
[31]:   ACFT_MAKE_NAME  LOC_STATE_NAME  ACFT_MODEL_NAME  \
0          BEECH  North Carolina          36
53         PIPER      Florida          PA28
55  FLIGHT DESIGN  California          CTLS
79  NORTH AMERICAN      Arizona          F51
80         CHAMPION  California          8KCAB
81         BEECH    California          35
82         CESSNA    Alabama          182

      RMK_TEXT      FLT_PHASE  \
0  AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B...  UNKNOWN (UNK)
53  AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES. ...  UNKNOWN (UNK)
55  AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES A...  UNKNOWN (UNK)
79  AIRCRAFT CRASHED UNDER UNKNOWN CIRCUMSTANCES, ...  UNKNOWN (UNK)
80  N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...  UNKNOWN (UNK)
81  N9872R, BEECH M35 AIRCRAFT, AND N5057G, BELLAN...  UNKNOWN (UNK)
82  N784CP AIRCRAFT CRASHED INTO A WOODED AREA NEA...  UNKNOWN (UNK)

   EVENT_TYPE_DESC  FATAL_FLAG
0      Accident      Yes
53      Accident      Yes
55      Accident      Yes
```


79	Accident	Yes
80	Accident	Yes
81	Accident	Yes
82	Accident	Yes