Pandas assignment-1

March 19, 2022

[3]:
```
#import libraries
import pandas as pd
```

[15]:
```
#importing CSV file and reading into dataframe.
#Uploading the csv file to LMS and the reading
faa_data_df =pd.read_csv('faa_ai_prelim.csv')
```

[18]:
```
#view the dataset shape
faa_data_df.shape
```

[18]: (83, 42)

[67]:
```
#view the first 6 observations
faa_data_df.head(6)
```

[67]:

	UPDATED	ENTRY_DATE	EVENT_LCL_DATE	EVENT_LCL_TIME	LOC_CITY_NAME	LOC_STATE_NAME	LOC_CNTRY_NAME	RMK_TEXT	EVENT_TYPE_DESC
0	No	19-FEB-16	19-FEB-16	00:45:00Z	MARSHVILLE	North Carolina	NaN	AIRCRAFT CRASHED INTO TREES, THE 1 PERSON ON B…	Accident
1	No	19-FEB-16	18-FEB-16	23:55:00Z	TAVERNIER	Florida	NaN	AIRCRAFT ON LANDING WENT OFF THE END OF THE RU…	Incident
2	No	19-FEB-16	18-FEB-16	22:14:00Z	TRENTON	New Jersey	NaN	AIRCRAFT ON FINAL SUSTAINED A BIRD STRIKE, LAN…	Incident
3	No	19-FEB-16	18-FEB-16	17:10:00Z	ASHEVILLE	North Carolina	NaN	AIRCRAFT ON LANDING, GEAR COLLAPSED, ASHEVILLE…	Incident
4	No	19-FEB-16	18-FEB-16	00:26:00Z	TALKEETNA	Alaska	NaN	AIRCRAFT ON LANDING, NOSE GEAR COLLAPSED, TALK…	Incident
5	No	19-FEB-16	18-FEB-16	20:23:00Z	PEARL HARBOR	Hawaii	NaN		

```
5  N80918 BELL 206B ROTORCRAFT CRASHED INTO THE W…        Accident

                 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
5      FAA Honolulu FSDO-13  …          NaN           NaN              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
5           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            NaN           NaN

[6 rows x 42 columns]
```

[23]:
```
#view all the columns of dataset
faa_data_df.columns
```

[23]:
```
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')
```

[26]:
```
#create dataframe with required columns
required_df =␣
 ↪faa_data_df[['ACFT_MAKE_NAME','LOC_STATE_NAME','ACFT_MODEL_NAME','RMK_TEXT','FLT_PHASE',
```

```
       'EVENT_TYPE_DESC','FATAL_FLAG']]
```

[27]:
```
#type of new dataset i.e required_df
type(required_df)
```

[27]:
```
pandas.core.frame.DataFrame
```

[28]:
```
#view first five observations
required_df.head()
```

[28]:
```
  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

                                            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)

  EVENT_TYPE_DESC FATAL_FLAG
0        Accident        Yes
1        Incident        NaN
2        Incident        NaN
3        Incident        NaN
4        Incident        NaN
```

[30]:
```
#replace all NaN for Fatal flag with "No"
required_df['FATAL_FLAG'].fillna('No',inplace=True)
```

```
/usr/local/lib/python3.7/site-packages/pandas/core/series.py:4536:
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
  downcast=downcast,
```

[31]:
```
#Now view top 5 observations
required_df.head(5)
```

[31]:
```
  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

                                                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)

  EVENT_TYPE_DESC FATAL_FLAG
0        Accident        Yes
1        Incident         No
2        Incident         No
3        Incident         No
4        Incident         No
```

[33]:
```
required_df.shape
```

[33]:
```
(83, 7)
```

[49]:
```
#Find the aircraft types and their occurrences in the dataset(method-1)
required_df['ACFT_MAKE_NAME'].value_counts()
```

[49]:
```
CESSNA                  23
PIPER                   10
BEECH                    9
MOONEY                   4
BOEING                   3
BELL                     2
CHAMPION                 2
LANCAIR                  2
SOCATA                   2
EMBRAER                  1
AERO COMMANDER           1
ENSTROM                  1
AEROSTAR INTERNATIONAL   1
GREAT LAKES              1
CHRISTEN                 1
HUGHES                   1
SAAB                     1
SABRELINER               1
FAIRCHILD                1
AERONCA                  1
PITTS                    1
```

```
CONSOLIDATED VULTEE        1
NORTH AMERICAN             1
FLIGHT DESIGN              1
GLOBE                      1
MAULE                      1
GULFSTREAM                 1
AIRBUS                     1
VANS                       1
GRUMMAN                    1
Name: ACFT_MAKE_NAME, dtype: int64
```

[44]:
```
#Remove all the observations where aircraft names are not available
final_df = required_df.dropna(subset=['ACFT_MAKE_NAME'])
```

[45]:
```
final_df.shape
```

[45]:
```
(78, 7)
```

[50]:
```
#grouping of aircrafts by their names(method-2)
aircraft_groups = final_df.groupby('ACFT_MAKE_NAME')
aircraft_groups.size()
```

[50]:
```
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
```

[53]:
```
#grouping of aircrafts by their fatal flag
fatal_flag_df = final_df.groupby('FATAL_FLAG')
fatal_flag_df.size()
```

[53]:
```
FATAL_FLAG
No     71
Yes     7
dtype: int64
```

[65]:
```
#Display the observations where fatal flag is "Yes"
fatal_accidents_list = fatal_flag_df.get_group('Yes')
fatal_accidents_list
```

[65]:
```
    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
```