Project 11 Solutions

Kevin Choe

Collaborators: (Collaborators listed here. Include names, which part of the project you gave or sought help with, and how you helped or were helped.)

TA help: Summeth Guda

Online resources used: (List of links/resources (if any) here. Include web addresses, which part of the project the resource helped with, and how you were helped.)

Question 1

```
import pandas as pd
```

```
import pandas as pd
import numpy as np
accidents=pd.DataFrame()
for i in range(1975,1982):
  s=str(i)
  accidents=pd.concat([accidents,pd.read_csv("/class/datamine/data/fars/"+s+"/ACCIDENT.CSV")])

accidents.head
```

```
<bound method NDFrame.head of STATE COUNTY MONTH DAY ...  DRUNK_DR ST_CASE CITY
RAIL
0 1 95 1 4 ...  0 10001 400 *******
1 1 95 1 2 ...  0 10002 110 *******
2 1 35 1 1 ...  0 10003 0 *******
3 1 77 1 2 ...  0 10004 0 *******
4 1 97 1 5 ...  0 10005 2100 *******
...  ...  ...  ...  ...  ...  ...  ...  ...  ...
43995 56 21 12 18 ...  0 560228 0 0000000
43996 56 5 12 23 ...  0 560229 0 0000000
43997 56 45 12 28 ...  0 560230 0 0000000
43998 56 13 12 29 ...  1 560231 0 0000000
43999 56 25 12 30 ...  1 560232 0 0000000

[300059 rows x 45 columns]>
```

```
accidents['YEAR']="19"+accidents['YEAR'].astype(str)
accidents['YEAR']
```

```
0        1975
1        1975
2        1975
3        1975
4        1975
         ...
```

```
43995    1981
43996    1981
43997    1981
43998    1981
43999    1981
Name: YEAR, Length: 300059, dtype: object
```

Question 2

```
temp=accidents.loc[accidents.loc[:, 'SCH_BUS'] >0, :]
temp=temp.loc[temp.loc[:, 'DRUNK_DR'] >0, :]
temp.shape[0]
```

```
101
```

Question 3

```
temp=temp.groupby(['YEAR'])
a=temp['YEAR'].count()
a
```

```
YEAR
1977    12
1978    37
1979    17
1980    12
1981    23
Name: YEAR, dtype: int64
```

```
a.idxmax()
```

```
'1978'
```

Question 4

```
drk=[]
for i in range(0,7):
  tmp=accidents.loc[accidents.loc[:,'DRUNK_DR'] == i,:]
  count=tmp['PERSONS'].mean()
  drk.append(count)
drk
```

```
[2.61554022666993, 2.474079257044717, 3.6607109965002764, 5.197916666666667,
5.25, nan, 6.0]
```

Question 5

```
np.unique(accidents['HOUR'])
```

```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17, 18, 19, 20, 21, 22, 23, 24, 99])
```

```
accidents['bins']=pd.cut(accidents['HOUR'], bins=[0,6,12,18,24,99],labels=[1,2,3,4,5],include_lowest=Tru
acc=accidents.groupby(['bins'])
acc['FATALS'].sum()
```

```
bins
1     85589
2     47312
3     86919
4    117776
5       409
Name: FATALS, dtype: int64
```

```
acc['FATALS'].mean()
```

```
bins
1    1.134441
2    1.122947
3    1.127632
4    1.138669
5    1.105405
Name: FATALS, dtype: float64
```

Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

> As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together – We are Purdue.