

```
import pandas as pd

from google.colab import drive
drive.mount('/content/gdrive')

Mounted at /content/gdrive
```

Q1. Create a file "Person.txt" with the following data:

<i>Age</i>	<i><u>agegroup</u></i>	<i>height</i>	<i>status</i>	<i><u>yearsmarried</u></i>
21	adult	6.0	single	-1
2	child	3	married	0
18	adult	5.7	married	20
221	elderly	5	widowed	2
34	child	-7	married	3

i) Read the data from the file "Person.txt".

```
data=pd.read_table("/content/gdrive/MyDrive/people.txt",delim_whitespace=True)
print(data)
```

```

      Age agegroup height status yearsmarried
0    21    adult   6.0   single           -1
1     2    child   3.0  married            0
2    18    adult   5.7  married           20
3   221  elderly   5.0  widowed            2
4    34    child   7.0  married            3
```

ii) Create a rule set E that contain rules to check for the following conditions :

1. The age should be in the range 0-150.

```
def check_age_range(data):
    age_range = lambda r: r in range(151) #r>=0 and r<=150
    return data['Age'].apply(age_range).rename('Check Age Range')
```

2. The age should be greater than years married.

```
def check_age(data):
```

```
age_limit = lambda r: r[0] > r[4]
return data.apply(age_limit,axis=1).rename("Check Age")
```

3. The status should be married or single or widowed.

```
def check_status(data):
    status_value= lambda r: r in data["status"].values
    return data['status'].apply(status_value).rename("Check status")
```

4. If age is less than 18 the age group should be child, if age is between 18 and 65 the age group should be adult, if age is more than 65 the age group should be elderly.

```
def check_age_group(data):
    age_group = lambda x: (x[0] in range(18) and x[1]=="child") or (x[0] in range(18,66) and x[
    return data[['Age', 'agegroup']].apply(age_group, axis =1).rename("Check age group")
```

```
E={"check_age_range":check_age_range, "check_age_limit":check_age, "check_status_value":check
result=list()
```

iii) Check whether rule set E is violated by the data in the file people.txt.

```
for i in E.keys():
    #print(E[i](data))
    result.append(E[i](data))
#print(result)
print("Number of voilations by peoples:")
print(len(result)-sum(result))
```

```
Number of voilations by peoples:
0    0
1    0
2    3
3    3
4    3
dtype: int64
```

iv) Summarize the results obtained in part(iii)

```
#result=pd.DataFrame(result)
#print(result)
print(result.describe())
```

```
↳
```

	0	1	2	3	4
count	4	4	4	4	4

unique	1	1	2	2	2
top	True	True	True	True	True
freq	4	4	3	3	3

v) Visualize the results obtained in part(iii)

```
import matplotlib.pyplot as plt
result.astype(int).plot(kind='bar')
plt.title("Visualization")
plt.xlabel('Rules')
plt.ylabel('True/False')
plt.legend(['Person1', 'Person2', 'Person3', 'Person4', 'Person5'])
```

<matplotlib.legend.Legend at 0x7f338638e0b8>



