


Cleansing of Group csv file

jupyter Hospital Claim Management Last Checkpoint: 2 hours ago (autosaved)  Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [1]: `import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns`

In [12]: `df=pd.read_csv('raw_Group.csv')`

In [13]: `df.head()`

Out[13]:

	Group ID	Group Name	Effective Date	Term(M)	Address
0	76891	Accident	12/12/2021	12	NaN
1	76892	Industrial Disease	3/1/2022	36	NaN
2	76893	Critical Illness	4/11/2021	30	NaN
3	76894	Surgeries	18/5/2022	24	NaN
4	76895	Genetic Illness	14/5/2022	26	NaN


In [14]: `del df['Address']`

In [15]: `df.head()`

Out[15]:

	Group ID	Group Name	Effective Date	Term(M)
0	76891	Accident	12/12/2021	12

Cleansing of Subgroup1 csv file

jupyter Hospital Claim Management Last Checkpoint: 2 hours ago (autosaved)  Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [15]: `df.head()`

Out[15]:

	Group ID	Group Name	Effective Date	Term(M)
0	76891	Accident	12/12/2021	12
1	76892	Industrial Disease	3/1/2022	36
2	76893	Critical Illness	4/11/2021	30
3	76894	Surgeries	18/5/2022	24
4	76895	Genetic Illness	14/5/2022	26

In [18]: `df.to_csv('Group.csv',index=False)`

In [19]: `df=pd.read_csv('raw_Subgroup1.csv')`

In [21]: `df.head(7)`

Out[21]:

	Subgroup ID	Group ID	Subgroup Name	Effective Date	Term(M)
0	56431	76891	Road-traffic accident	12/4/2021	32
1	56432	76891	Fractures caused by falls	22/9/2021	23
2	56433	76891	Gun shot	5/5/2022	15
3	56434	76891	Burn Cases	27/4/2022	26
4	56435	76891	Current Shock	1/1/2022	10

jupyter Hospital Claim Management Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

0	56431	76891	Road-traffic accident	12/4/2021	32
1	56432	76891	Fractures caused by falls	22/9/2021	23
2	56433	76891	Gun shot	5/5/2022	15
3	56434	76891	Burn Cases	27/4/2022	26
4	56435	76891	Current Shock	1/1/2022	10
5	56436	76891	Broken Arm	NaN	8

```
In [40]: df=df.dropna()
```

```
In [41]: df
```

```
Out[41]:
```

	Subgroup ID	Group ID	Subgroup Name	Effective Date	Term(M)
0	56431	76891	Road-traffic accident	12/4/2021	32
1	56432	76891	Fractures caused by falls	22/9/2021	23
2	56433	76891	Gun shot	5/5/2022	15
3	56434	76891	Burn Cases	27/4/2022	26
4	56435	76891	Current Shock	1/1/2022	10

```
In [42]: df = df.reset_index(drop = True)
```

```
In [43]: df
```

Cleansing of Subgroup2 csv file

jupyter Hospital Claim Management Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

```
In [43]: df
```

```
Out[43]:
```

	Subgroup ID	Group ID	Subgroup Name	Effective Date	Term(M)
0	56431	76891	Road-traffic accident	12/4/2021	32
1	56432	76891	Fractures caused by falls	22/9/2021	23
2	56433	76891	Gun shot	5/5/2022	15
3	56434	76891	Burn Cases	27/4/2022	26
4	56435	76891	Current Shock	1/1/2022	10


```
In [44]: df.to_csv('Subgroup1.csv',index=False)
```

```
In [45]: df=pd.read_csv('raw_Subgroup2.csv')
```

```
In [46]: df
```

```
Out[46]:
```

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56441	76892	Industrial deafness	23/1/2022	32.0
1	56442	76892	Respiratory Problems	22/9/2021	23.0
2	56443	76892	Breathing Problems	5/5/2022	15.0
3	56448	76892	NaN	26/1/2022	NaN
4	56444	76892	Dermatitis	27/4/2022	26.0
5	56445	76892	Musculoskeletal disorder	1/1/2022	10.0

jupyter Hospital Claim Management Last Checkpoint: 2 hours ago (autosaved)  Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [47]: `df=df.dropna()`

In [48]: `df`

Out[48]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56441	76892	Industrial deafness	23/1/2022	32.0
1	56442	76892	Respiratory Problems	22/9/2021	23.0
2	56443	76892	Breathing Problems	5/5/2022	15.0
4	56444	76892	Dermatitis	27/4/2022	26.0
5	56445	76892	Musculoskeletal disorder	1/1/2022	10.0


In [49]: `df = df.reset_index(drop = True)`

In [50]: `df`

Out[50]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56441	76892	Industrial deafness	23/1/2022	32.0
1	56442	76892	Respiratory Problems	22/9/2021	23.0
2	56443	76892	Breathing Problems	5/5/2022	15.0
3	56444	76892	Dermatitis	27/4/2022	26.0
4	56445	76892	Musculoskeletal disorder	1/1/2022	10.0

Cleansing of Subgroup3 csv file

jupyter Hospital Claim Management Last Checkpoint: a few seconds ago (autosaved)  Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [51]: `df.to_csv('Subgroup2.csv',index=False)`

In [52]: `df=pd.read_csv('raw_Subgroup3.csv')`

In [53]: `df`

Out[53]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56451	768923	Cancer	23/1/2022	30.0
1	56452	768923	Major organ transplant	22/9/2021	27.0
2	56459	768923	NaN	7/7/2021	20.0
3	56453	768923	Kidney failure	5/5/2022	15.0
4	56454	768923	Paralysis	27/4/2022	26.0
5	56460	768923	Stroke	NaN	NaN
6	56455	768923	First heart attack	1/1/2022	10.0

In [54]: `df=df.dropna()`

In [55]: `df`

Out[55]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56451	768923	Cancer	23/1/2022	30.0

jupyter Hospital Claim Management Last Checkpoint: a few seconds ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [55]: df

Out[55]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56451	768923	Cancer	23/1/2022	30.0
1	56452	768923	Major organ transplant	22/9/2021	27.0
3	56453	768923	Kidney failure	5/5/2022	15.0
4	56454	768923	Paralysis	27/4/2022	26.0
6	56455	768923	First heart attack	1/1/2022	10.0

In [56]: df = df.reset_index(drop = True)

In [57]: df

Out[57]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56451	768923	Cancer	23/1/2022	30.0
1	56452	768923	Major organ transplant	22/9/2021	27.0
2	56453	768923	Kidney failure	5/5/2022	15.0
3	56454	768923	Paralysis	27/4/2022	26.0
4	56455	768923	First heart attack	1/1/2022	10.0

In [58]: df.to_csv('Subgroup3.csv', index=False)

Cleansing of Subgroup4 csv file

jupyter Hospital Claim Management Last Checkpoint: 8 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [59]: df=pd.read_csv('raw_Subgroup4.csv')

In [60]: df

Out[60]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56461	768924	Neurological Surgery	23/1/2022	30
1	56462	768924	Oncology	22/9/2021	27
2	56463	768924	Chemotherapy	4/4/2022	20
3	56472	768924	NaN	NaN	19
4	56464	768924	Hystoscopy	27/4/2022	15
5	56465	768924	Nasal Concha	1/1/2022	10

In [61]: df=df.dropna()

In [62]: df

Out[62]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56461	768924	Neurological Surgery	23/1/2022	30
1	56462	768924	Oncology	22/9/2021	27
2	56463	768924	Chemotherapy	4/4/2022	20
4	56464	768924	Hystoscopy	27/4/2022	15

jupyter Hospital Claim Management Last Checkpoint: 8 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [62]: df
```

Out[62]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56461	768924	Neurological Surgery	23/1/2022	30
1	56462	768924	Oncology	22/9/2021	27
2	56463	768924	Chemotherapy	4/4/2022	20
4	56464	768924	Hystoscopy	27/4/2022	15
5	56465	768924	Nasal Concha	1/1/2022	10

```
In [63]: df = df.reset_index(drop = True)
```

```
In [64]: df
```

Out[64]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56461	768924	Neurological Surgery	23/1/2022	30
1	56462	768924	Oncology	22/9/2021	27
2	56463	768924	Chemotherapy	4/4/2022	20
3	56464	768924	Hystoscopy	27/4/2022	15
4	56465	768924	Nasal Concha	1/1/2022	10

```
In [65]: df.to_csv('Subgroup4.csv',index=False)
```

Cleansing of Subgroup5 csv file

jupyter Hospital Claim Management Last Checkpoint: 11 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [66]: df=pd.read_csv('raw_Subgroup5.csv')
```

```
In [67]: df
```

Out[67]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56470	768925	Autism	NaN	25
1	56471	768925	Down syndrome	23/1/2022	30
2	56472	768925	FragileX syndrome	22/9/2021	27
3	56473	768925	Turner syndrome	4/4/2022	20
4	56474	768925	Trisomy 18	27/4/2022	28
5	56475	768925	Trisomy 13	1/1/2022	14

```
In [68]: df=df.dropna()
```

```
In [69]: df
```

Out[69]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
1	56471	768925	Down syndrome	23/1/2022	30
2	56472	768925	FragileX syndrome	22/9/2021	27
3	56473	768925	Turner syndrome	4/4/2022	20
4	56474	768925	Trisomy 18	27/4/2022	28



Out[69]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
1	56471	768925	Down syndrome	23/1/2022	30
2	56472	768925	FragileX syndrome	22/9/2021	27
3	56473	768925	Turner syndrome	4/4/2022	20
4	56474	768925	Trisomy 18	27/4/2022	28
5	56475	768925	Trisomy 13	1/1/2022	14

In [70]: `df = df.reset_index(drop = True)`In [71]: `df`

Out[71]:

	Subgroup ID	Goup ID	Subgroup Name	Effective Date	Term(M)
0	56471	768925	Down syndrome	23/1/2022	30
1	56472	768925	FragileX syndrome	22/9/2021	27
2	56473	768925	Turner syndrome	4/4/2022	20
3	56474	768925	Trisomy 18	27/4/2022	28
4	56475	768925	Trisomy 13	1/1/2022	14

In [72]: `df.to_csv('Subgroup5.csv', index=False)`