

!! CE PROJET EST TOUJOURS DANS SON ÉTAT
BROUILLON !!

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
```

```
df = pd.read_csv("C:\\Users\\shera\\Downloads\\cdc2.csv")
print(df)
```

\	_STATE	FMONTH	PVTRESID1	COLGHOUS	STATERE1	SEXVAR	GENHLTH
0	1.0	1.0	1.0	NaN	1.0	2.0	2.0
1	1.0	1.0	1.0	NaN	1.0	2.0	1.0
2	1.0	1.0	1.0	NaN	1.0	2.0	2.0
3	1.0	1.0	1.0	NaN	1.0	2.0	1.0
4	1.0	1.0	1.0	NaN	1.0	2.0	4.0
...
445127	78.0	11.0	NaN	NaN	NaN	2.0	3.0
445128	78.0	11.0	NaN	NaN	NaN	2.0	1.0
445129	78.0	11.0	NaN	NaN	NaN	2.0	5.0
445130	78.0	11.0	NaN	NaN	NaN	1.0	2.0
445131	78.0	11.0	NaN	NaN	NaN	1.0	2.0

	PHYSHLTH	MENTHLTH	PERSDOC3	...	USEMRJN4	TYP CNTR9	_AGE80
HTM4 \ 0	88.0	88.0	1.0	...	NaN	NaN	80.0
NaN 1	88.0	88.0	2.0	...	NaN	NaN	80.0
160.0 2	2.0	3.0	1.0	...	NaN	NaN	56.0
157.0 3	88.0	88.0	1.0	...	NaN	NaN	73.0
165.0 4	2.0	88.0	2.0	...	NaN	NaN	43.0
157.0
...							

445127	88.0	3.0	3.0	...	NaN	NaN	19.0
165.0							
445128	2.0	2.0	2.0	...	NaN	NaN	51.0
170.0							
445129	30.0	30.0	3.0	...	NaN	NaN	65.0
170.0							
445130	88.0	88.0	2.0	...	NaN	NaN	73.0
183.0							
445131	88.0	1.0	3.0	...	NaN	NaN	42.0
168.0							

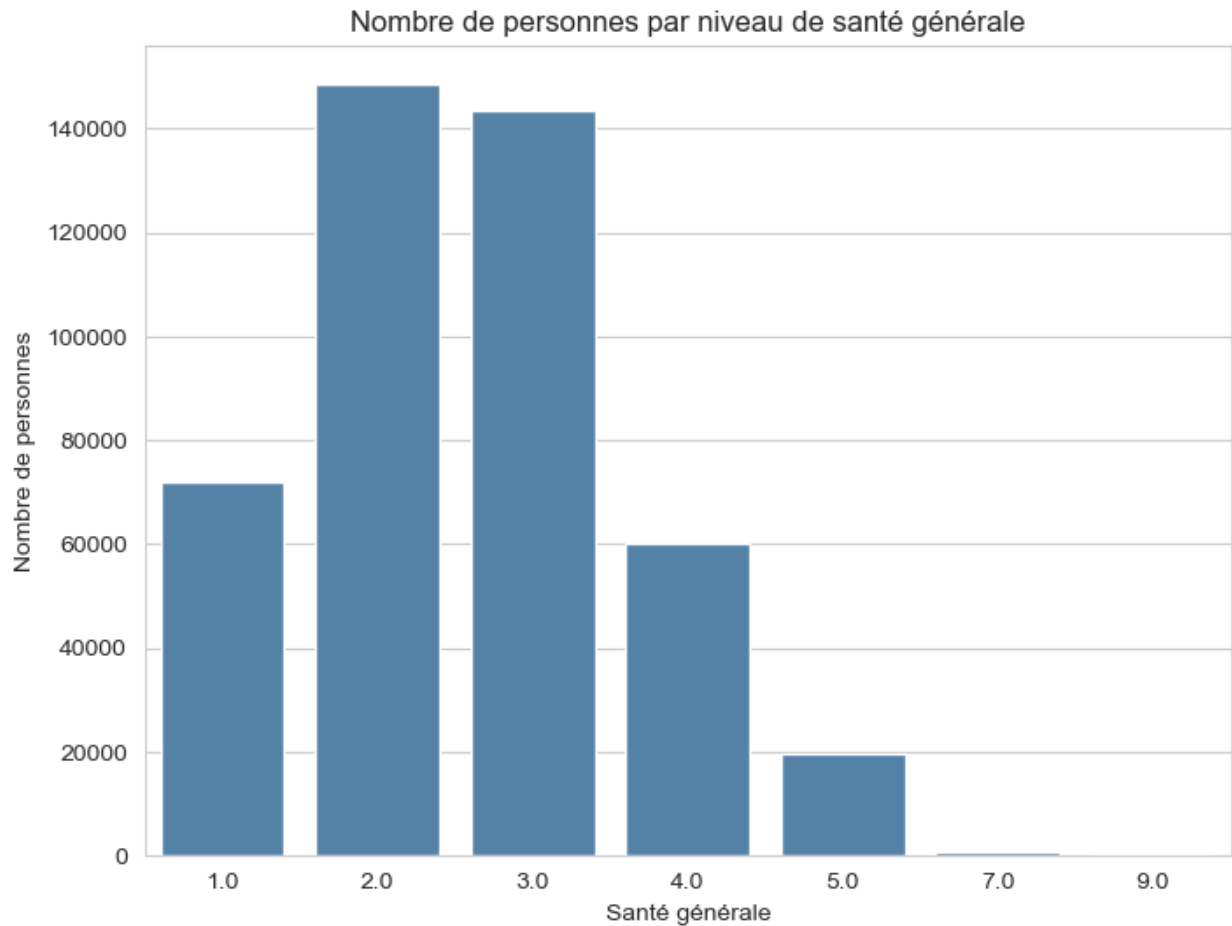
	WTKG3	_BMI5	_YRSSMOK	_PACKDAY	_PACKYRS	_DRNKWK2
0	NaN	NaN	NaN	NaN	NaN	5.397605e-79
1	6804.0	2657.0	NaN	NaN	NaN	5.397605e-79
2	6350.0	2561.0	NaN	NaN	NaN	5.397605e-79
3	6350.0	2330.0	56.0	0.1	6.0	5.397605e-79
4	5398.0	2177.0	NaN	NaN	NaN	1.400000e+02
...
445127	6985.0	2563.0	NaN	NaN	NaN	9.990000e+04
445128	8301.0	2866.0	NaN	NaN	NaN	5.397605e-79
445129	4990.0	1723.0	44.0	1.0	44.0	9.990000e+04
445130	10886.0	3255.0	NaN	NaN	NaN	5.397605e-79
445131	6350.0	2260.0	NaN	NaN	NaN	9.990000e+04

[445132 rows x 73 columns]

```

health_counts = df['GENHLTH'].value_counts()
sns.set_style("whitegrid")
plt.figure(figsize=(8, 6))
sns.barplot(x=health_counts.index, y=health_counts.values,
color='steelblue')
plt.title("Nombre de personnes par niveau de santé générale")
plt.xlabel("Santé générale")
plt.ylabel("Nombre de personnes")
plt.show()

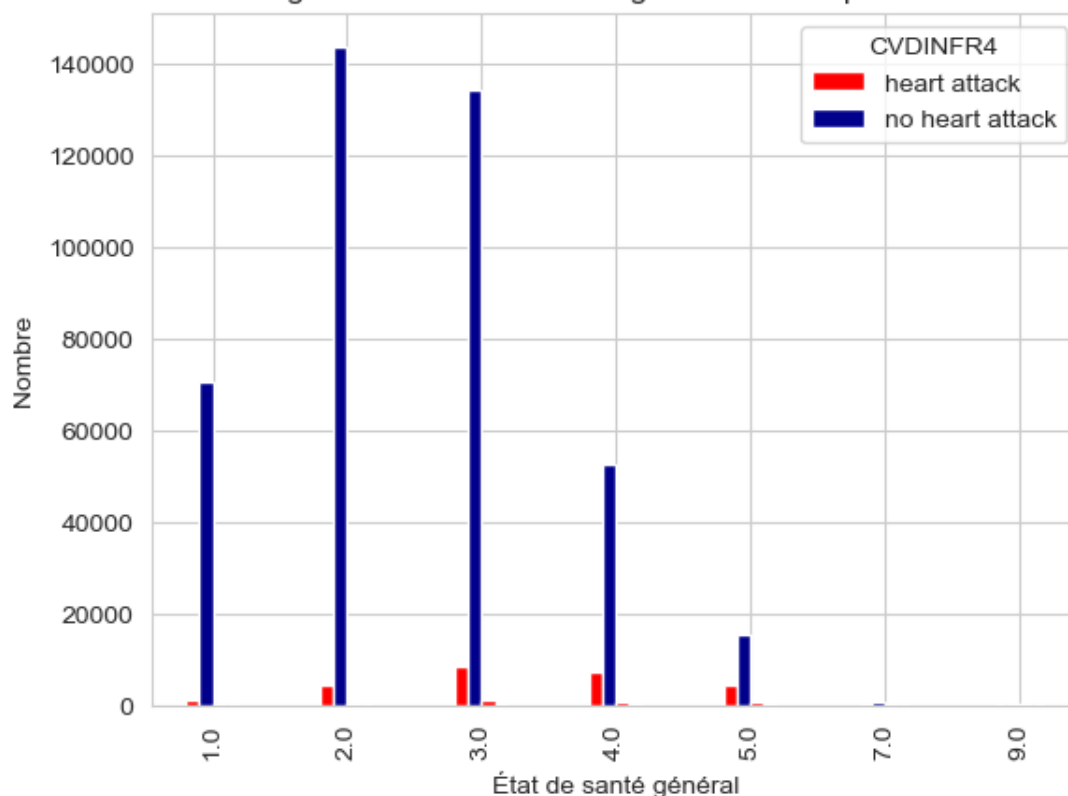
```



```
combien_de_malades = df.groupby('GENHLTH')
['CVDINFR4'].value_counts().unstack(fill_value=0) # Fréquence de
chaque diagnostic par rapport à leur état de santé général
plt.figure(figsize=(10, 6))
combien_de_malades.plot(kind='bar', color=['red', 'darkblue'],
width=0.4)
plt.title("Relation entre le diagnostic et l'état de santé général
estimé par les individus eux meme")
plt.xlabel("État de santé général")
plt.ylabel("Nombre")
plt.legend(title='CVDINFR4', labels=["heart attack", "no heart
attack"])
plt.show()
print(combien_de_malades)
```

<Figure size 1000x600 with 0 Axes>

Relation entre le diagnostic et l'état de santé général estimé par les individus eux meme



CVDINFR4	1.0	2.0	7.0	9.0
GENHLTH				
1.0	1097	70526	207	48
2.0	4214	143753	409	68
3.0	8295	134255	948	98
4.0	7228	52300	702	43
5.0	4161	15143	416	21
7.0	86	688	34	2
9.0	27	293	15	50

```
import numpy as np

categorical = df.select_dtypes(include=[object])
print(categorical.columns)

Index([], dtype='object')

numerical = df.select_dtypes(include=[np.number])
print(numerical.columns)

Index(['_STATE', 'FMONTH', 'PVTRES1', 'COLGHOUS', 'STATERE1',
'SEXVAR',
      'GENHLTH', 'PHYSHLTH', 'MENTHLTH', 'PERSDOC3', 'MEDCOST1',
'CHECKUP1',
      'EXERANY2', 'SLEPTIM1', 'LASTDEN4', 'RMVTETH4', 'CVDINFR4',
```

```

'CVDCRHD4',
    'CVDSTRK3', 'ASTHMA3', 'ASTHNOW', 'CHCSCNC1', 'CHCOCNC1',
'CHCCOPD3',
    'ADDEPEV3', 'CHCKDNY2', 'HAVARTH4', 'DIABETE4', 'DIABAGE4',
'MARITAL',
    'EDUCA', 'NUMPHON4', 'CPDEMO1C', 'VETERAN3', 'EMPLOY1',
'CHILDREN',
    'INCOME3', 'PREGNANT', 'DEAF', 'BLIND', 'DIFFWALK', 'SMOKDAY2',
    'USENOW3', 'ECIGNOW2', 'LCSNUMCG', 'ALCDAY4', 'DIABTYPE',
'INSULIN1',
    'COPDSMOK', 'CNCRAGE', 'CNCRTYP2', 'CIMEMLoS', 'ACEDEPRS',
'ACEDRINK',
    'ACEDRUGS', 'ACEHURT1', 'LSATISFY', 'EMTSUPRT', 'SDHISOLT',
'SDHFOOD1',
    'SDHBILLS', 'SDHSTRE1', 'MARIJAN1', 'USEMRJN4', 'TYPCNTR9',
'_AGE80',
    'HTM4', 'WTKG3', '_BMI5', '_YRSSMOK', '_PACKDAY', '_PACKYRS',
    '_DRNKWK2'],
dtype='object')

```

```

df2 = df.rename(columns={'_STATE': 'state', 'FMONTH': 'file_month',
'PVTRESID1': 'private_residence', 'COLGHOUS': 'college_housing',
'STATERE1': 'resident_of_state', 'SEXVAR': 'sex_respondant',
'GENHLTH': 'general_health', 'PHYSHLTH': 'physical_health',
'MENTHLTH': 'mental_health', 'PERSDOC3': 'personal_hc_provider',
'MEDCOST1': 'not_afford_doc', 'CHECKUP1': 'last_checkup', 'EXERANY2':
'exercise_last_30_days', 'SLEPTIM1': 'sleep_time', 'LASTDEN4':
'last_dentist_checkup', 'RMVTETH4': 'num_permanent_teeth_removed',
'CVDINFR4': 'diagnosed_heart_attack', 'CVDCRHD4':
'diagnosed_angina_chd', 'CVDSTRK3': 'diagnosed_stroke', 'ASTHMA3':
'diagnosed_asthma', 'ASTHNOW': 'still_have_asthma', 'CHCSCNC1':
'told_had_skin_cancer', 'CHCOCNC1': 'told_had_melanoma_or_cancer',
'CHCCOPD3': 'told_had_bronchitis_emphysema_copd', 'ADDEPEV3':
'told_had_depressive_disorder', 'CHCKDNY2': 'told_had_kidney_disease',
'HAVARTH4': 'told_had_arthritis', 'DIABETE4': 'told_had_diabetes',
'MARITAL': 'marital_status', 'EDUCA': 'educational_level', 'VETERAN3':
'veteran', 'EMPLOY1': 'employment_status', 'CHILDREN': 'num_children',
'INCOME3': 'income_level', 'PREGNANT': 'pregnant', 'DEAF': 'deaf',
'BLIND': 'blind', 'DIFFWALK': 'difficulty_walking', 'SMOKDAY2':
'smoking_days', 'USENOW3': 'snus_chemma', 'ECIGNOW2': 'e_cigs_vaping',
'LCSNUMCG': 'cigarettes_a_day', 'ALCDAY4': 'alcohol_last_30',
'DIABTYPE': 'diabetes_type', 'INSULIN1': 'taking_insulin', 'COPDSMOK':
'years_smoking_tobacco', 'CNCRAGE': 'cancer_age', 'CNCRTYP2':
'cancer_type', 'CIMEMLoS': 'memory_loss', 'ACEDEPRS':
'living_with_depressed', 'ACEDRINK': 'living_with_alcoholic',
'ACEDRUGS': 'living_with_drogué', 'ACEHURT1':
'parent_hirt_physically', 'LSATISFY': 'life_satisfaction', 'EMTSUPRT':
'got_emotional_support', 'SDHISOLT': 'feel_socially_isolated',
'SDHFOOD1': 'food_didnt_last', 'SDHBILLS': 'not_able_pay_bills',
'SDHSTRE1': 'stress', 'MARIJAN1': 'used_marijuana', 'TYPCNTR9':

```

```
'contraception_method', 'HTM4': 'height', 'WTKG3': 'weight', '_BMI5':
'BMI', '_YRSSMOK': 'num_smoking_years', '_PACKDAY': 'packs_cigs_day',
'_PACKYRS': 'packs_cigs_years', '_DRNKWK2': 'alcohol_per_week'})
```

```
print(df['DIABAGE4']) #NaN values
print(df['NUMPHON4']) #NaN values except the first 2
print(df['CPDEM01C']) #not in the code book
print(df['USEMRJN4']) #NaN values
```

```
0      80.0
1      NaN
2      NaN
3      NaN
4      NaN
```

```
...
445127    NaN
445128    NaN
445129    NaN
445130    NaN
445131    NaN
```

Name: DIABAGE4, Length: 445132, dtype: float64

```
0      1.0
1      2.0
2      NaN
3      NaN
4      NaN
```

```
...
445127    NaN
445128    NaN
445129    NaN
445130    NaN
445131    NaN
```

Name: NUMPHON4, Length: 445132, dtype: float64

```
0      2.0
1      1.0
2      1.0
3      1.0
4      2.0
```

```
...
445127    2.0
445128    1.0
445129    1.0
445130    1.0
445131    1.0
```

Name: CPDEM01C, Length: 445132, dtype: float64

```
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
```

```

..
445127    NaN
445128    NaN
445129    NaN
445130    NaN
445131    NaN
Name: USEMRJN4, Length: 445132, dtype: float64

```

```
print(df2)
```

	state	file_month	private_residence	college_housing	\
0	1.0	1.0	1.0	NaN	
1	1.0	1.0	1.0	NaN	
2	1.0	1.0	1.0	NaN	
3	1.0	1.0	1.0	NaN	
4	1.0	1.0	1.0	NaN	
...
445127	78.0	11.0	NaN	NaN	
445128	78.0	11.0	NaN	NaN	
445129	78.0	11.0	NaN	NaN	
445130	78.0	11.0	NaN	NaN	
445131	78.0	11.0	NaN	NaN	

	resident_of_state	sex_respondant	general_health
physical_health \			
0	1.0	2.0	2.0
88.0			
1	1.0	2.0	1.0
88.0			
2	1.0	2.0	2.0
2.0			
3	1.0	2.0	1.0
88.0			
4	1.0	2.0	4.0
2.0			
...
...			
445127	NaN	2.0	3.0
88.0			
445128	NaN	2.0	1.0
2.0			
445129	NaN	2.0	5.0
30.0			
445130	NaN	1.0	2.0
88.0			
445131	NaN	1.0	2.0
88.0			

	mental_health	personal_hc_provider	...	USEMRJN4	\
0	88.0	1.0	...	NaN	

1	88.0	2.0	...	NaN
2	3.0	1.0	...	NaN
3	88.0	1.0	...	NaN
4	88.0	2.0	...	NaN
...
445127	3.0	3.0	...	NaN
445128	2.0	2.0	...	NaN
445129	30.0	3.0	...	NaN
445130	88.0	2.0	...	NaN
445131	1.0	3.0	...	NaN

	contraception_method	_AGE80	height	weight	BMI	\
0	NaN	80.0	NaN	NaN	NaN	
1	NaN	80.0	160.0	6804.0	2657.0	
2	NaN	56.0	157.0	6350.0	2561.0	
3	NaN	73.0	165.0	6350.0	2330.0	
4	NaN	43.0	157.0	5398.0	2177.0	
...	
445127	NaN	19.0	165.0	6985.0	2563.0	
445128	NaN	51.0	170.0	8301.0	2866.0	
445129	NaN	65.0	170.0	4990.0	1723.0	
445130	NaN	73.0	183.0	10886.0	3255.0	
445131	NaN	42.0	168.0	6350.0	2260.0	

	num_smoking_years	packs_cigs_day	packs_cigs_years
alcohol_per_week			
0	NaN	NaN	NaN
5.397605e-79			
1	NaN	NaN	NaN
5.397605e-79			
2	NaN	NaN	NaN
5.397605e-79			
3	56.0	0.1	6.0
5.397605e-79			
4	NaN	NaN	NaN
1.400000e+02			
...
...			
445127	NaN	NaN	NaN
9.990000e+04			
445128	NaN	NaN	NaN
5.397605e-79			
445129	44.0	1.0	44.0
9.990000e+04			
445130	NaN	NaN	NaN
5.397605e-79			
445131	NaN	NaN	NaN
9.990000e+04			

[445132 rows x 73 columns]


```
df2.to_csv("second_clean.csv", sep=',', index=False)
```

```
df3 = pd.read_csv("C:\\Users\\shera\\Downloads\\second_clean00.csv")
print(df3)
```

	state	file_month	private_residence	college_housing	\
0	1.0	1.0	1.0	NaN	
1	1.0	1.0	1.0	NaN	
2	1.0	1.0	1.0	NaN	
3	1.0	1.0	1.0	NaN	
4	1.0	1.0	1.0	NaN	
...	
445127	78.0	11.0	NaN	NaN	
445128	78.0	11.0	NaN	NaN	
445129	78.0	11.0	NaN	NaN	
445130	78.0	11.0	NaN	NaN	
445131	78.0	11.0	NaN	NaN	

	resident_of_state	sex_respondant	general_health
0	1.0	2.0	2.0
1	1.0	2.0	1.0
2	1.0	2.0	2.0
3	1.0	2.0	1.0
4	1.0	2.0	4.0
...
...
445127	NaN	2.0	3.0
445128	NaN	2.0	1.0
445129	NaN	2.0	5.0
445130	NaN	1.0	2.0
445131	NaN	1.0	2.0

	mental_health	personal_hc_provider	...	USEMRJN4	\
0	88.0	1.0	...	NaN	
1	88.0	2.0	...	NaN	
2	3.0	1.0	...	NaN	
3	88.0	1.0	...	NaN	
4	88.0	2.0	...	NaN	

```

...
445127      3.0      3.0  ...      NaN
445128      2.0      2.0  ...      NaN
445129     30.0      3.0  ...      NaN
445130     88.0      2.0  ...      NaN
445131      1.0      3.0  ...      NaN

contraception_method  _AGE80  height  weight  BMI  \
0      NaN      80.0      NaN      NaN      NaN
1      NaN      80.0     160.0     6804.0    2657.0
2      NaN      56.0     157.0     6350.0    2561.0
3      NaN      73.0     165.0     6350.0    2330.0
4      NaN      43.0     157.0     5398.0    2177.0
...
445127      NaN     19.0     165.0     6985.0    2563.0
445128      NaN     51.0     170.0     8301.0    2866.0
445129      NaN     65.0     170.0     4990.0    1723.0
445130      NaN     73.0     183.0    10886.0    3255.0
445131      NaN     42.0     168.0     6350.0    2260.0

num_smoking_years  packs_cigs_day  packs_cigs_years
alcohol_per_week
0      NaN      NaN      NaN
5.397605e-79
1      NaN      NaN      NaN
5.397605e-79
2      NaN      NaN      NaN
5.397605e-79
3      56.0      0.1      6.0
5.397605e-79
4      NaN      NaN      NaN
1.400000e+02
...      ...      ...      ...
...
445127      NaN      NaN      NaN
9.990000e+04
445128      NaN      NaN      NaN
5.397605e-79
445129      44.0      1.0      44.0
9.990000e+04
445130      NaN      NaN      NaN
5.397605e-79
445131      NaN      NaN      NaN
9.990000e+04

[445132 rows x 73 columns]
print(df3.columns)

```

```

Index(['state', 'file_month', 'private_residence', 'college_housing',
      'resident_of_state', 'sex_respondant', 'general_health',
      'physical_health', 'mental_health', 'personal_hc_provider',
      'not_afford_doc', 'last_checkup', 'exercise_last_30_days',
      'sleep_time',
      'last_dentist_checkup', 'num_permanent_teeth_removed',
      'diagnosed_heart_attack', 'diagnosed_angina_chd',
      'diagnosed_stroke',
      'diagnosed_asthma', 'still_have_asthma',
      'told_had_skin_cancer',
      'told_had_melanoma_or_cancer',
      'told_had_bronchitis_emphysema_copd',
      'told_had_depressive_disorder', 'told_had_kidney_disease',
      'told_had_arthritis', 'told_had_diabetes', 'DIABAGE4',
      'marital_status',
      'educational_level', 'NUMPHON4', 'CPDEM01C', 'veteran',
      'employment_status', 'num_children', 'income_level',
      'pregnant', 'deaf',
      'blind', 'difficuly_walking', 'smoking_days', 'snus_chemma',
      'e_cigs_vaping', 'cigarettes_a_day', 'alcohol_last_30',
      'diabetes_type',
      'taking_insulin', 'years_smoking_tobacco', 'cancer_age',
      'cancer_type',
      'memory_loss', 'living_with_depressed',
      'living_with_alcoholic',
      'living_with_drogué', 'parent_hirt_physically',
      'life_satisfaction',
      'got_emotional_support', 'feel_socially_isolated',
      'food_didnt_last',
      'not_able_pay_bills', 'stress', 'used_marijuana', 'USEMRJN4',
      'contraception_method', '_AGE80', 'height', 'weight', 'BMI',
      'num_smoking_years', 'packs_cigs_day', 'packs_cigs_years',
      'alcohol_per_week'],
      dtype='object')

```

```
df3['_AGE80'].astype(int)
```

```

0      80
1      80
2      56
3      73
4      43
..
445127  19
445128  51
445129  65
445130  73
445131  42

```

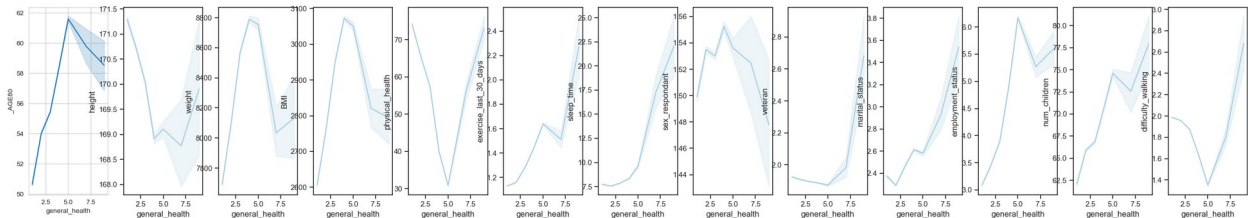
```
Name: _AGE80, Length: 445132, dtype: int32
```

```
df3.describe()[1:]
[list(df3)].T.style.background_gradient(cmap='Blues')
```

```
<pandas.io.formats.style.Styler at 0x22000125cd0>
```

```
def numeric_features_func(f):
    plt.figure(figsize=(35, 5))
    i = 1
    new = df3.filter(items=['_AGE80', 'height', 'weight', 'BMI',
'physical_health', 'exercise_last_30_days', 'sleep_time',
'sex_respondant', 'veteran', 'marital_status', 'education_level',
'employment_status', 'num_children', 'difficulty_walking'])
    for feature in new.columns:
        plt.subplot(1, 14, i)
        sns.set(palette='Paired')
        sns.set_style("ticks")
        sns.lineplot(y=new[feature], x=df3[f])
        i += 1
```

```
numeric_features_func('general_health')
```



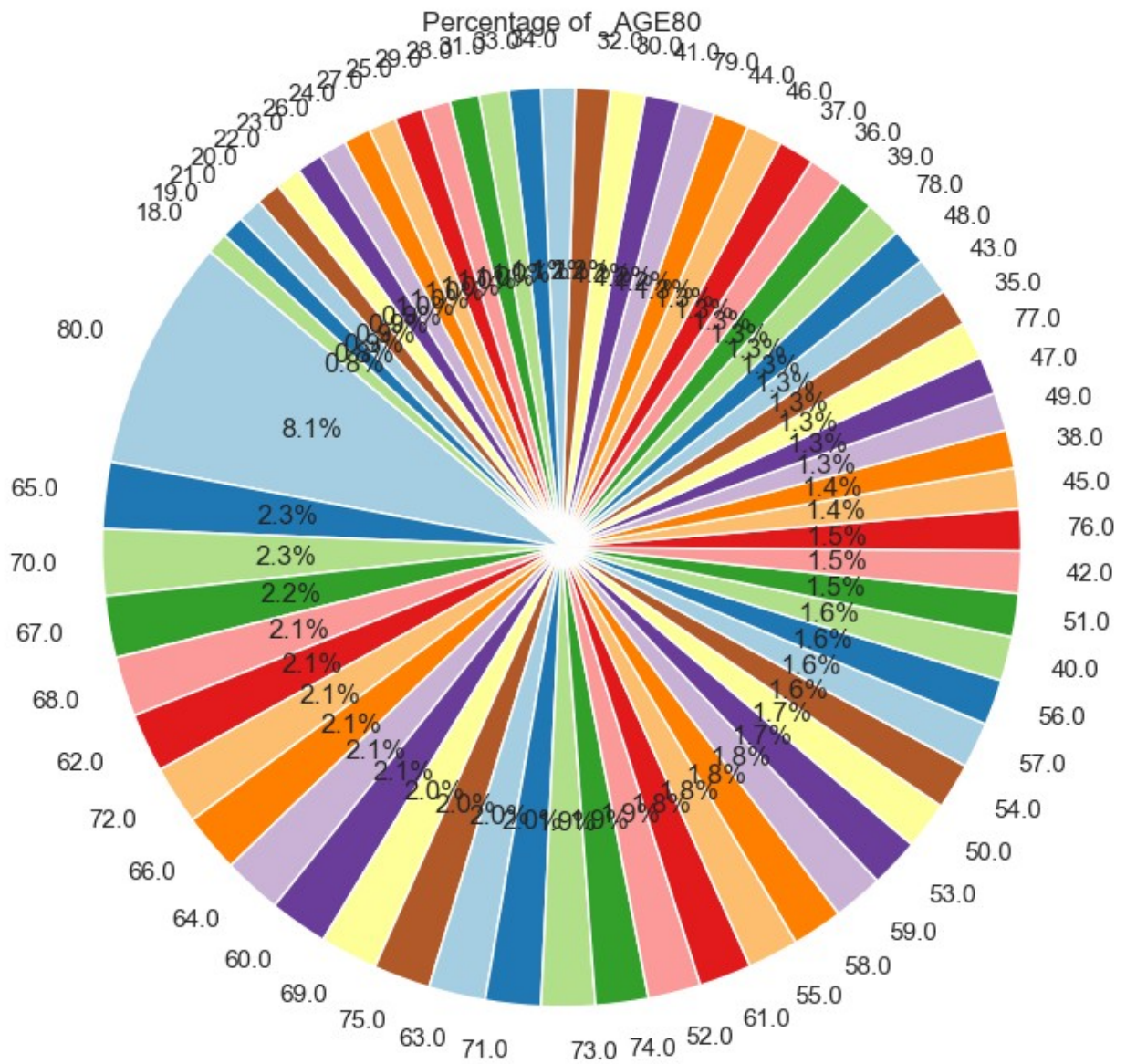
```
df3['physical_health']
```

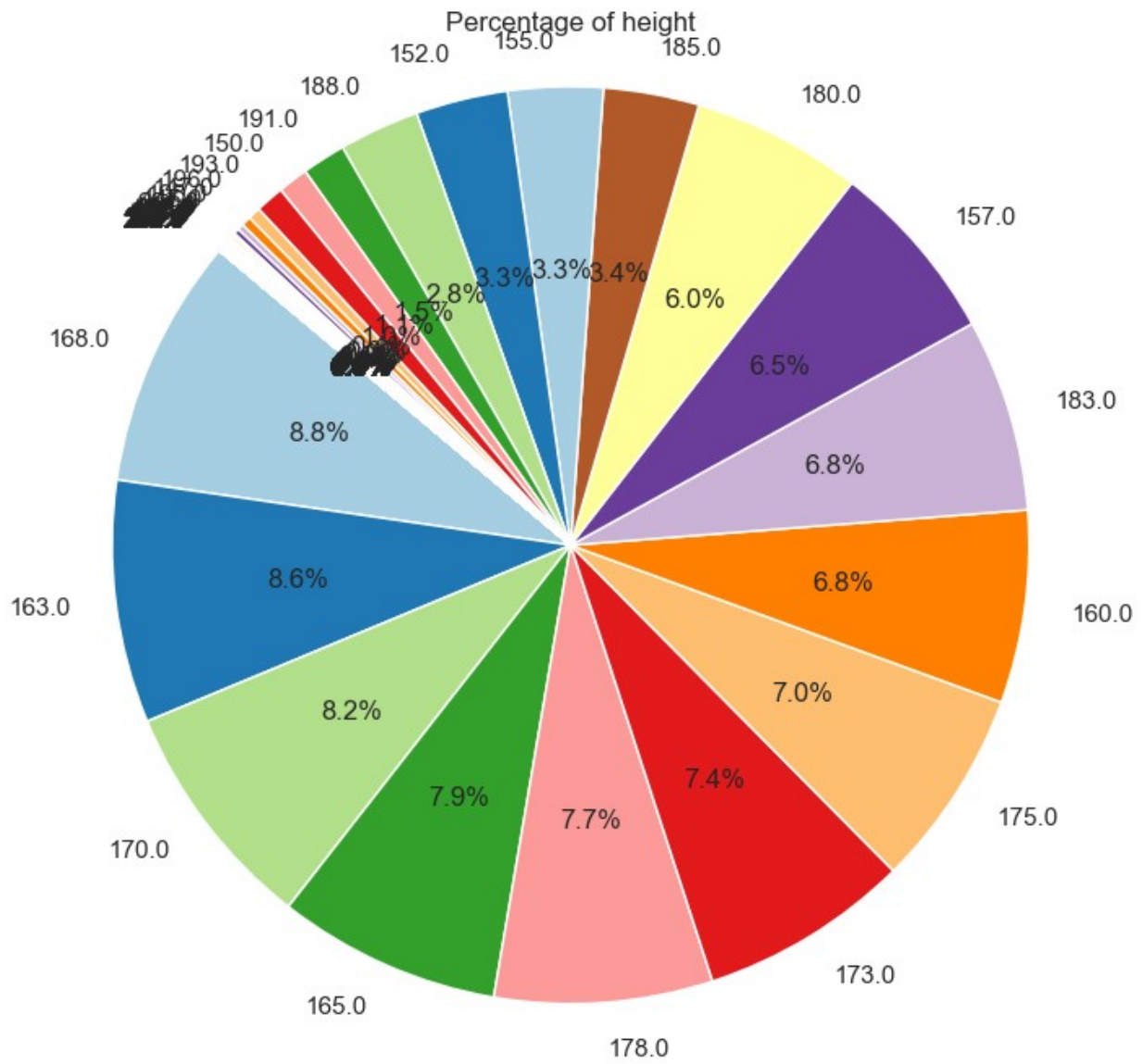
```
0      88.0
1      88.0
2       2.0
3      88.0
4       2.0
...
445127  88.0
445128   2.0
445129  30.0
445130  88.0
445131  88.0
```

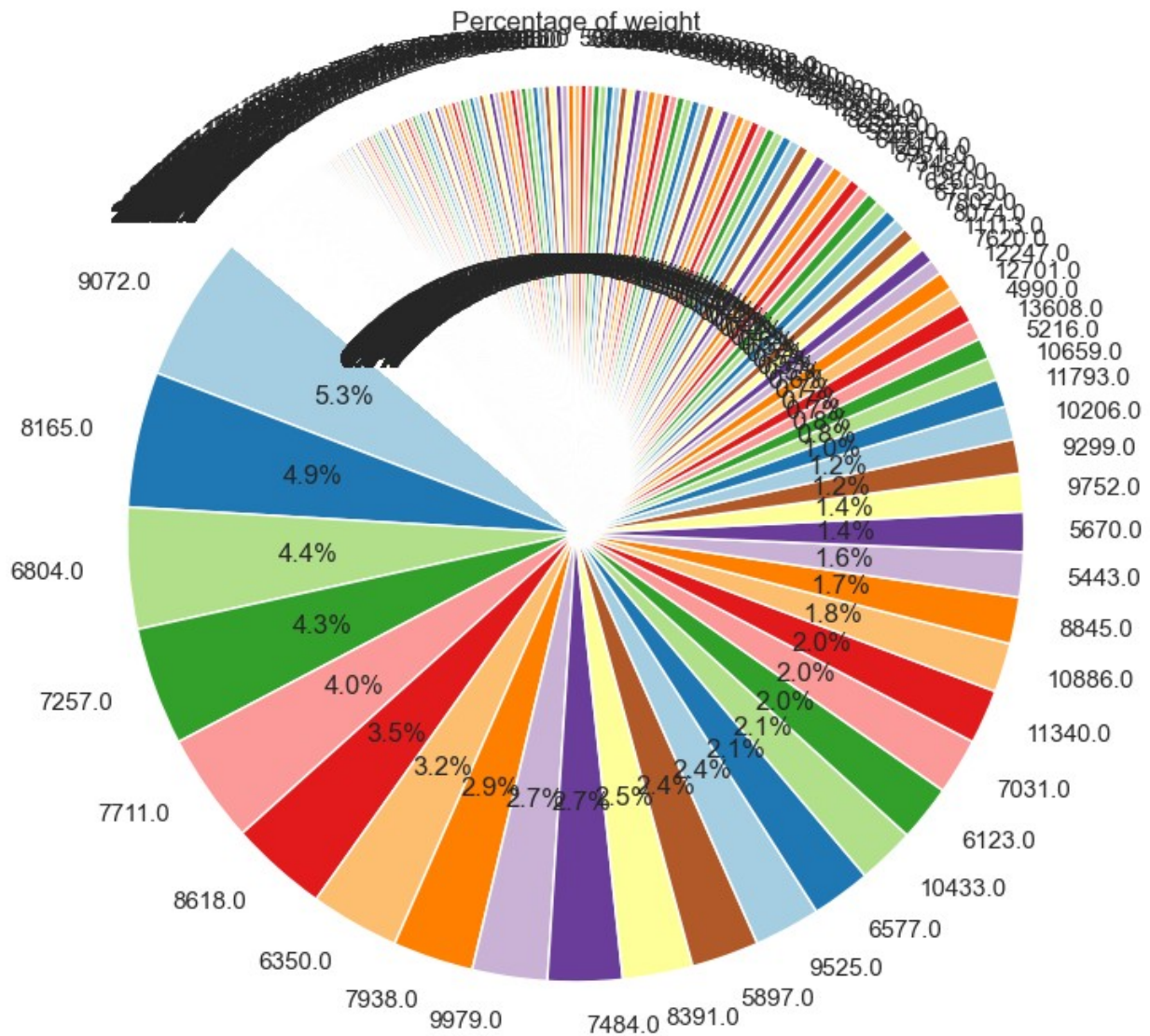
```
Name: physical_health, Length: 445132, dtype: float64
```

```
import matplotlib.pyplot as plt
```

```
def draw_pie_plot(df, attribute):  
    attribute_counts = df3[attribute].value_counts()  
    percentages = attribute_counts / len(df3) * 100  
  
    plt.figure(figsize=(8, 8))  
    plt.pie(percentages, labels=attribute_counts.index, autopct='%1.1f  
%%', startangle=140)  
    plt.title(f"Percentage of {attribute}")  
    plt.axis('equal')  
    plt.show()  
  
attributes_to_plot = ['_AGE80', 'height', 'weight']  
for attribute in attributes_to_plot:  
    draw_pie_plot(df3, attribute)
```







```
import pandas as pd
import matplotlib.pyplot as plt

def plot_weight_distribution_pie(df):
    df['weight_intervals'] = pd.cut(df['weight'] / 500, bins=range(0,
int(df['weight'].max() / 500) + 1, 1))

    poor_data = df[df['general_health'] == 5]

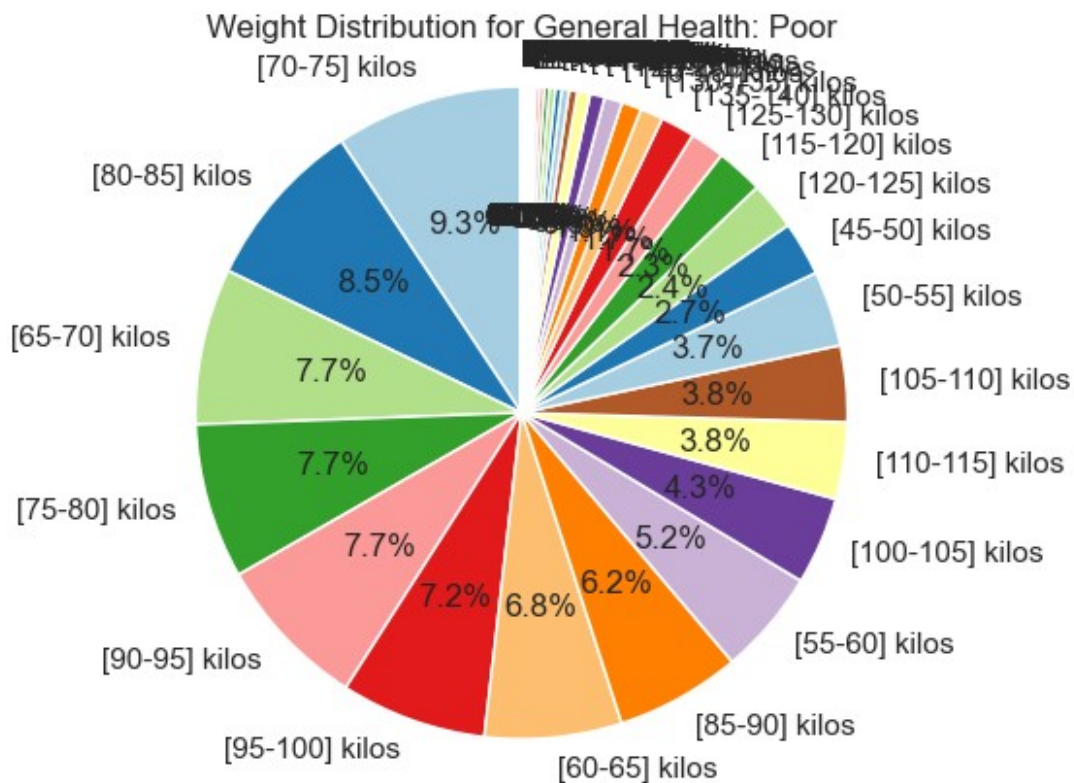
    weight_distribution =
poor_data['weight_intervals'].value_counts(normalize=True)

    labels = [f"[{int(interval.left * 5)}-{int(interval.right * 5)}]
kilos" for interval in weight_distribution.index]
```



```
plt.pie(weight_distribution, labels=labels, autopct='%1.1f%%',
startangle=90)
plt.axis('equal')
plt.title('Weight Distribution for General Health: Poor')
plt.show()
```

```
plot_weight_distribution_pie(df3)
```



```
len(df3[(df3['weight'] >=12000) & (df3['weight'] <= 12500)])
```

```
6530
```

```
len(df3[(df3['weight'] >= 8000) & (df3['weight'] <= 8500)])
```

```
39934
```

```
len(df3)
```

```
445132
```

```
women=df3[df3['sex_respondant'] == 2]
```

```
print(women)
```

	state	file_month	private_residence	college_housing	\
0	1.0	1.0	1.0	NaN	

1	1.0	1.0	1.0	NaN
2	1.0	1.0	1.0	NaN
3	1.0	1.0	1.0	NaN
4	1.0	1.0	1.0	NaN
...
445119	78.0	11.0	NaN	NaN
445123	78.0	11.0	NaN	NaN
445127	78.0	11.0	NaN	NaN
445128	78.0	11.0	NaN	NaN
445129	78.0	11.0	NaN	NaN

	resident_of_state	sex_respondant	general_health
physical_health \			
0	1.0	2.0	2.0
88.0			
1	1.0	2.0	1.0
88.0			
2	1.0	2.0	2.0
2.0			
3	1.0	2.0	1.0
88.0			
4	1.0	2.0	4.0
2.0			
...
...			
445119	NaN	2.0	2.0
88.0			
445123	NaN	2.0	4.0
88.0			
445127	NaN	2.0	3.0
88.0			
445128	NaN	2.0	1.0
2.0			
445129	NaN	2.0	5.0
30.0			

	mental_health	personal_hc_provider	...	contraception_method
\				
0	88.0	1.0	...	NaN
1	88.0	2.0	...	NaN
2	3.0	1.0	...	NaN
3	88.0	1.0	...	NaN
4	88.0	2.0	...	NaN
...

445119	88.0	1.0	...	NaN
445123	7.0	3.0	...	NaN
445127	3.0	3.0	...	NaN
445128	2.0	2.0	...	NaN
445129	30.0	3.0	...	NaN
	_AGE80	height	weight	BMI
packs_cigs_day	\			num_smoking_years
0	80.0	NaN	NaN	NaN
NaN				
1	80.0	160.0	6804.0	2657.0
NaN				NaN
2	56.0	157.0	6350.0	2561.0
NaN				NaN
3	73.0	165.0	6350.0	2330.0
0.1				56.0
4	43.0	157.0	5398.0	2177.0
NaN				NaN
...
...				
445119	75.0	147.0	3765.0	1735.0
1.0				30.0
445123	29.0	193.0	9072.0	2434.0
NaN				NaN
445127	19.0	165.0	6985.0	2563.0
NaN				NaN
445128	51.0	170.0	8301.0	2866.0
NaN				NaN
445129	65.0	170.0	4990.0	1723.0
1.0				44.0
	packs_cigs_years	alcohol_per_week	weight_intervals	
0	NaN	5.397605e-79	NaN	
1	NaN	5.397605e-79	(13.0, 14.0]	
2	NaN	5.397605e-79	(12.0, 13.0]	
3	6.0	5.397605e-79	(12.0, 13.0]	
4	NaN	1.400000e+02	(10.0, 11.0]	
...	
445119	30.0	4.700000e+01	(7.0, 8.0]	
445123	NaN	5.397605e-79	(18.0, 19.0]	
445127	NaN	9.990000e+04	(13.0, 14.0]	
445128	NaN	5.397605e-79	(16.0, 17.0]	
445129	44.0	9.990000e+04	(9.0, 10.0]	
[235893 rows x 74 columns]				

```
len(women)
```

```
235893
```

```
women_filtrees= women[(women['general_health'] != 7) &  
(women['general_health'] != 9)]  
health_counts = women_filtrees['general_health'].value_counts()  
fig, ax = plt.subplots()  
ax.pie(health_counts, labels=health_counts.index, autopct='%1.1f%%',  
startangle=90)  
ax.axis('equal')  
ax.set_title('How women described their general health (Excluding the  
answers not sure and refused)')  
plt.show()
```

How women described their general health (Excluding the answers not sure and refused)

