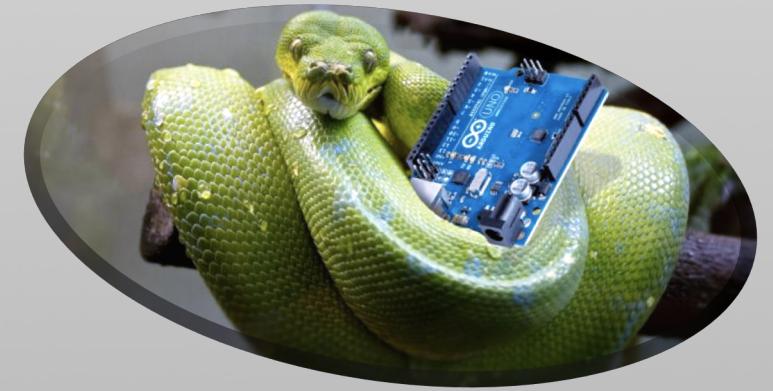
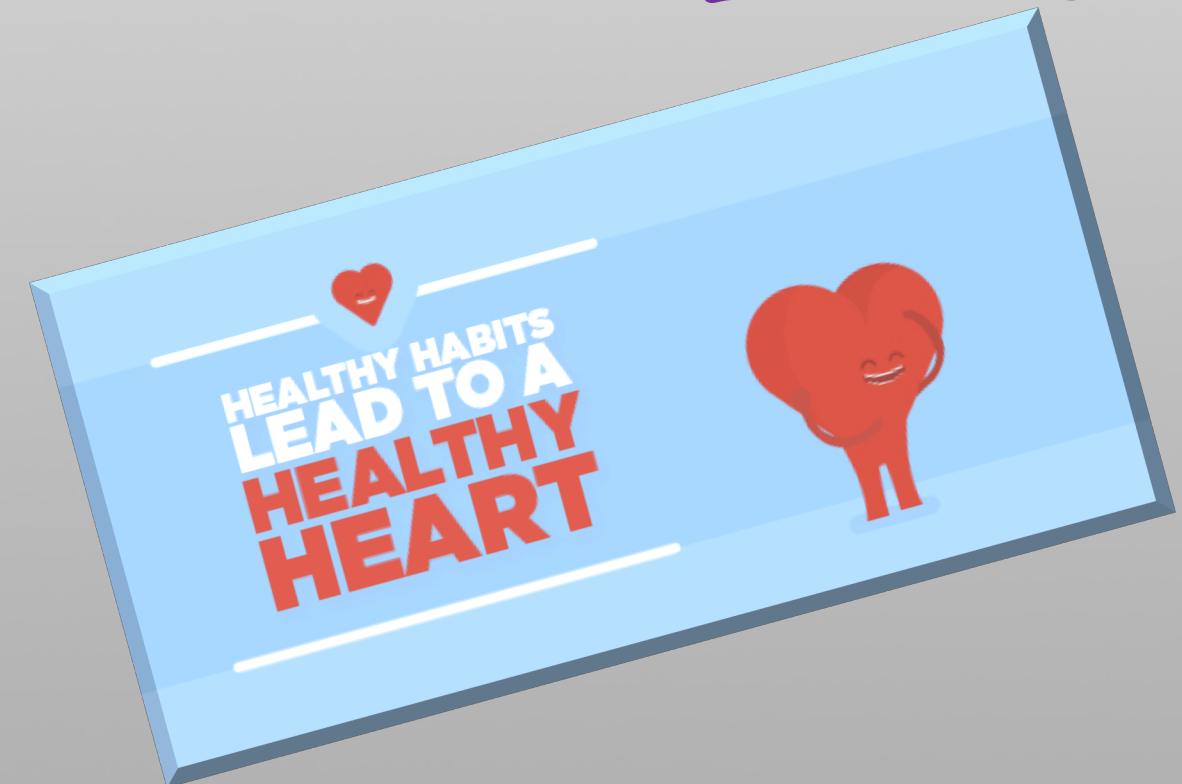




# BRFSS-Healthcare Data Analytics using Python



Presented By,  
**Harsh Raizada**



# Presentation Path.....

- ✓ Rationale for selection and description of dataset.
- ✓ Description of analysis procedure and objectives.
- ✓ Objectives
- ✓ Python coding and output for the objective analysis.





# Rationale for selection and description of dataset.

My interest in Healthcare forced me to select the Behavioural Risk Factor Surveillance System (BRFSS) open data set, which is the United states premier system of health-related telephone surveys that collect data about U.S. residents (18 year or older) regarding their health-related risk behaviours, chronic health conditions, and use of preventive services. Established in 1984 with 15 states, BRFSS now collects data in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world.

The BRFSS Questionnaire: Each year, the states—represented by their BRFSS coordinators and CDC—agree on the content of the questionnaire. The BRFSS questionnaire consists of three sections

1. Core component
2. Optional modules, and
3. State-added questions.





# What is risk factor surveillance?

Keeping track of the rates of risk factors which are the things or states in our daily lives that confers risk to our health is defined as a Risk Factor Surveillance.

By collecting behavioural health risk data at the federal and state level, BRFSS has become a powerful tool for targeting and designing health promotion activities for the US population.





# Description of analysis procedure and objectives:

I have tried to done Descriptive analytics which usually aims at developing population based rates (or percentage) with this data set.

Originally BRFSS data set were having sample of 486303 US residents (observations) and 275 variables in the data set but I have taken only those observations into consideration who have completed the interview and done slicing after which I have left with 421192 observation with 275 variables (which were required for my analysis).

## **A. IMMUNIZATION COVERAGE STATUS**

- Objective-1 To find out the existing coverage rate of Tetanus shot among US residents.
- Objective-2 To find out the existing coverage rate of Pneumococcal vaccine among US residents.

## **B.CHRONIC HEALTH CONDITION STATUS-ASTHMA**

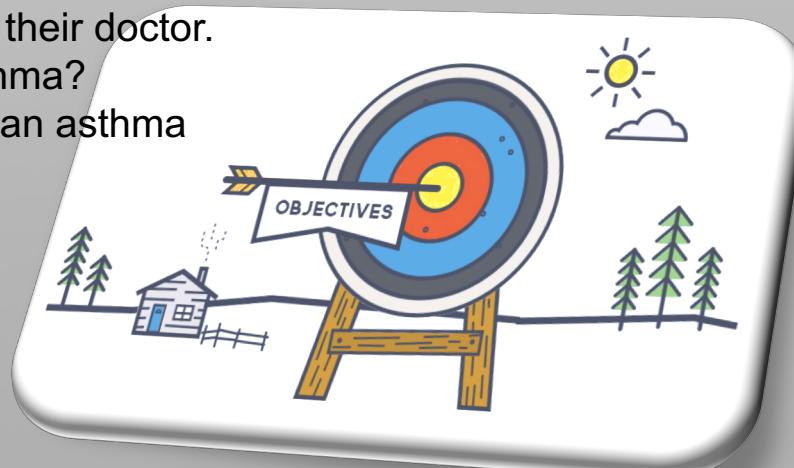
- Objective-3 To find out how much percentage of US residents ever told about asthma by their doctor.
- Objective-4 Among those who ever told about asthma by doctor how many still have asthma?
- Objective-5 Among those who still have asthma how many had an episode of asthma or an asthma attack in past 12 months?

## **C. BREAST & CERVICAL CANCER SCREENING STATUS**

- Objective -6 To find out the existing Breasts cancer screening rate among US residents.
- Objective-7 To find out the existing Cervical cancer screening rate among US residents.

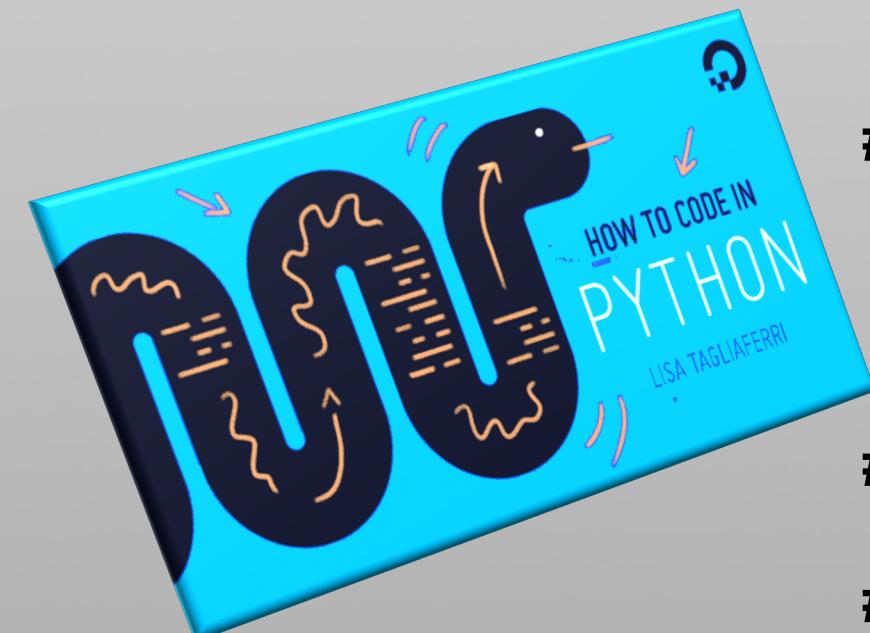
## **D. HIV/AIDS TEST STATUS**

- Object-8 To find out the existing HIV testing rate among US residents.





# Setting up python environment for analysis



## # changing of working directory

- import os
- os.getcwd()
- os.chdir(r"/Users/harsh/Desktop/Python programming course /pproject")

## # importing of important packages for analysis

- import numpy as np
- import pandas as pd
- import matplotlib.pyplot as plt
- import seaborn as sns

## #reading of data CSV file.

- brfss=pd.read\_csv("brfssp.csv")

## #copy of main data set i.e brfss into another dataset.

- cbrfss=brfss

## #slicing & subsetting of data frame with selected variables.

- cbrfssci=cbrfss[(cbrfss.DISPCODE==1100)]



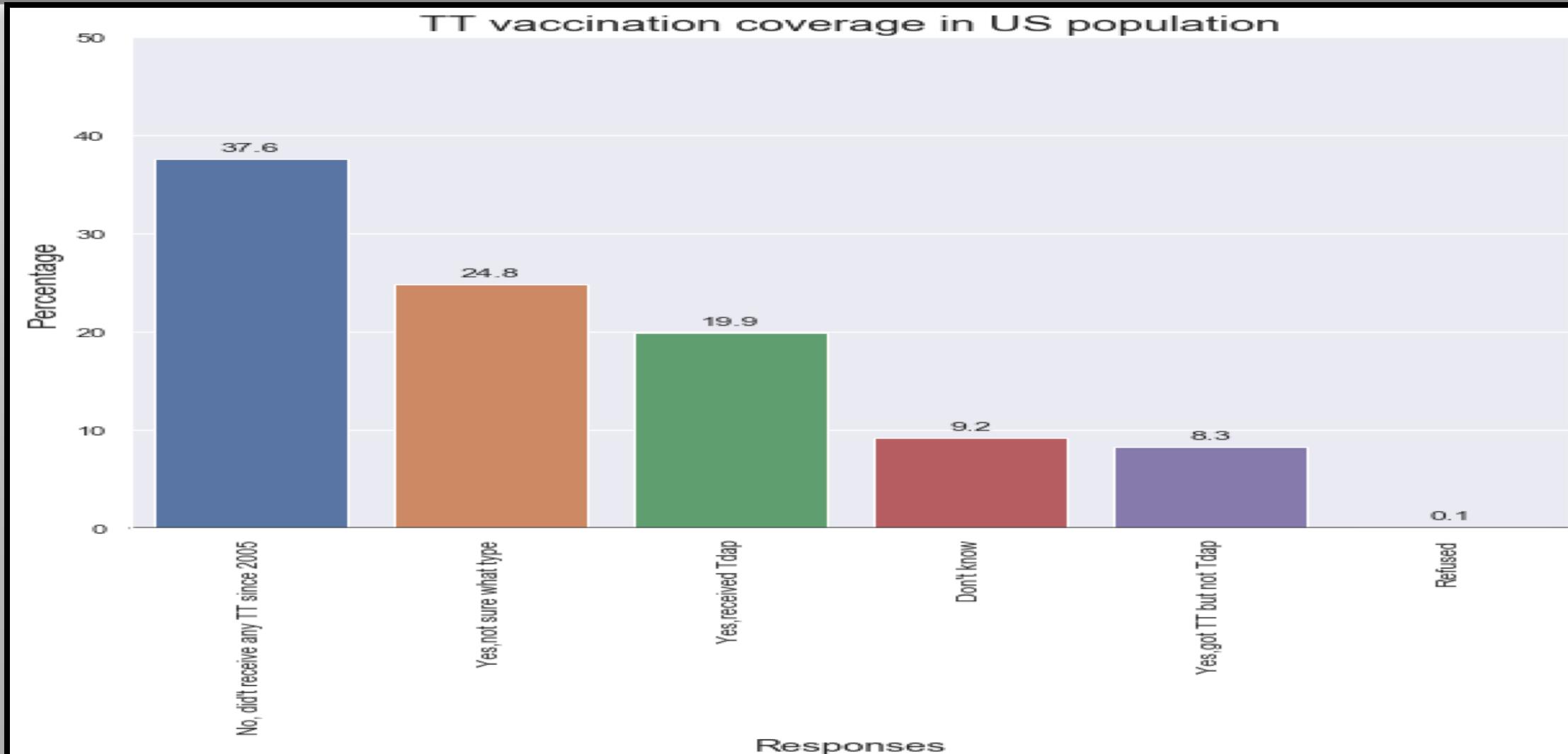
# Objective-1 To find out the existing coverage rate of Tetanus shot among US residents.

## Python coding-

```
tt=cbrrfssci.TETANUS.value_counts()  
tt1=tt.to_frame()  
tt1["Percentage"]=round((tt1/tt1.sum())*100,2)  
tt1["Category"]={"No, did't receive any TT since 2005":4.0,"Yes,not sure what type":3.0,"Yes,received Tdap":1.0,"Don't know":7.0,"Yes,got TT but not Tdap":2.0,"Refused":9.0}  
ttbar=sns.barplot(data=tt1,x="Category",y="Percentage"\n                  ,palette="deep")  
plt.axhline(0, color="k", clip_on=False)  
plt.title("TT vaccination coverage in US population")  
plt.xlabel("Responses")  
plt.ylim(0,50)  
plt.xticks(rotation='vertical')  
rects = ttbar.patches  
for rect in rects:  
    y_value = rect.get_height()  
    x_value = rect.get_x() + rect.get_width() / 2  
    space = 5  
    va = 'bottom'  
    label = "{:.1f}".format(y_value)  
    plt.annotate(label,(x_value, y_value),xytext=(0, space),\n                textcoords="offset points",ha='center',va=va)  
plt.figure(figsize=(10,7))
```



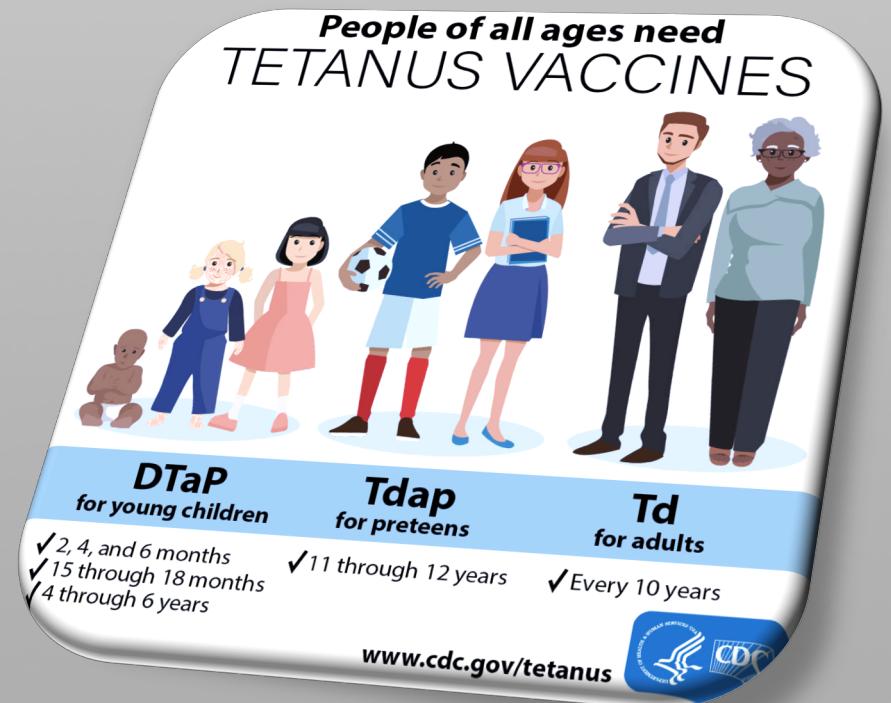
# Output-





## Conclusion-

- Around 37.6% of the respondents have not any TT shot in past 10 years which makes them more prone to Tetanus infection.





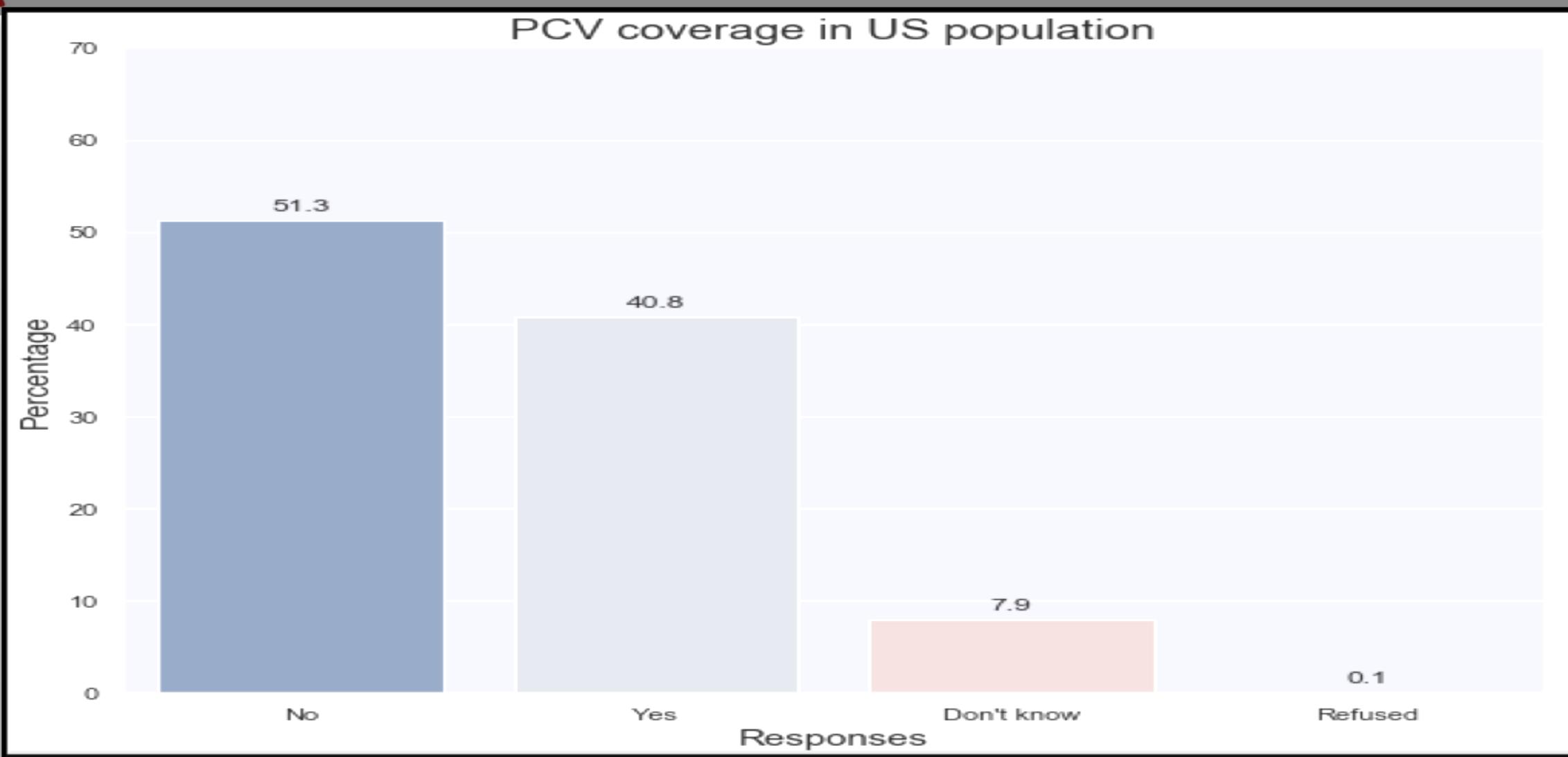
## Objective-2 To find out the existing coverage rate of Pneumococcal vaccine (PCV) among US residents.

Python coding-

```
pcv=cbrfssci.PNEUVAC3.value_counts()  
pcv1=pcv.to_frame()  
pcv1["Percentage"]=round((pcv1/pcv1.sum())*100,2)  
pcv1["Category"]={"No":2.0,"Yes":1.0,"Don't know":7.0,"Refused":9.0}  
pcvbar=sns.barplot(data=pcv1,x="Category",y="Percentage",  
                    palette="vlag",)  
plt.title("PCV coverage in US population")  
plt.xlabel("Responses")  
plt.ylim(0,70)  
rects = pcvbar.patches  
for rect in rects:  
    y_value = rect.get_height()  
    x_value = rect.get_x() + rect.get_width() / 2  
    space = 5  
    va = 'bottom'  
    label = "{:.1f}".format(y_value)  
    plt.annotate(label,(x_value, y_value),xytext=(0, space),  
                textcoords="offset points",ha='center',va=va)
```



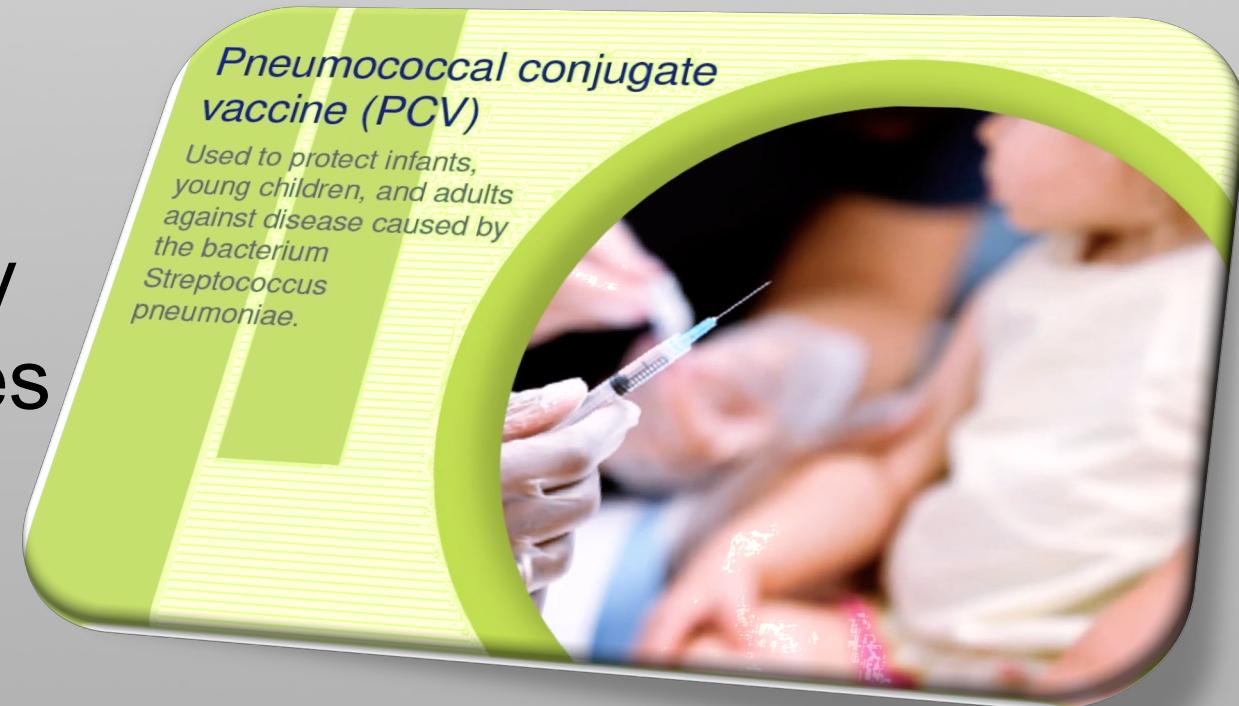
# Output:





## Conclusion-

PCV coverage in population is only 40.8 % and still 51.3 % of population have not received any shot of PCV vaccine which makes them more prone to lung infections



### *Pneumococcal conjugate vaccine (PCV)*

*Used to protect infants, young children, and adults against disease caused by the bacterium *Streptococcus pneumoniae*.*



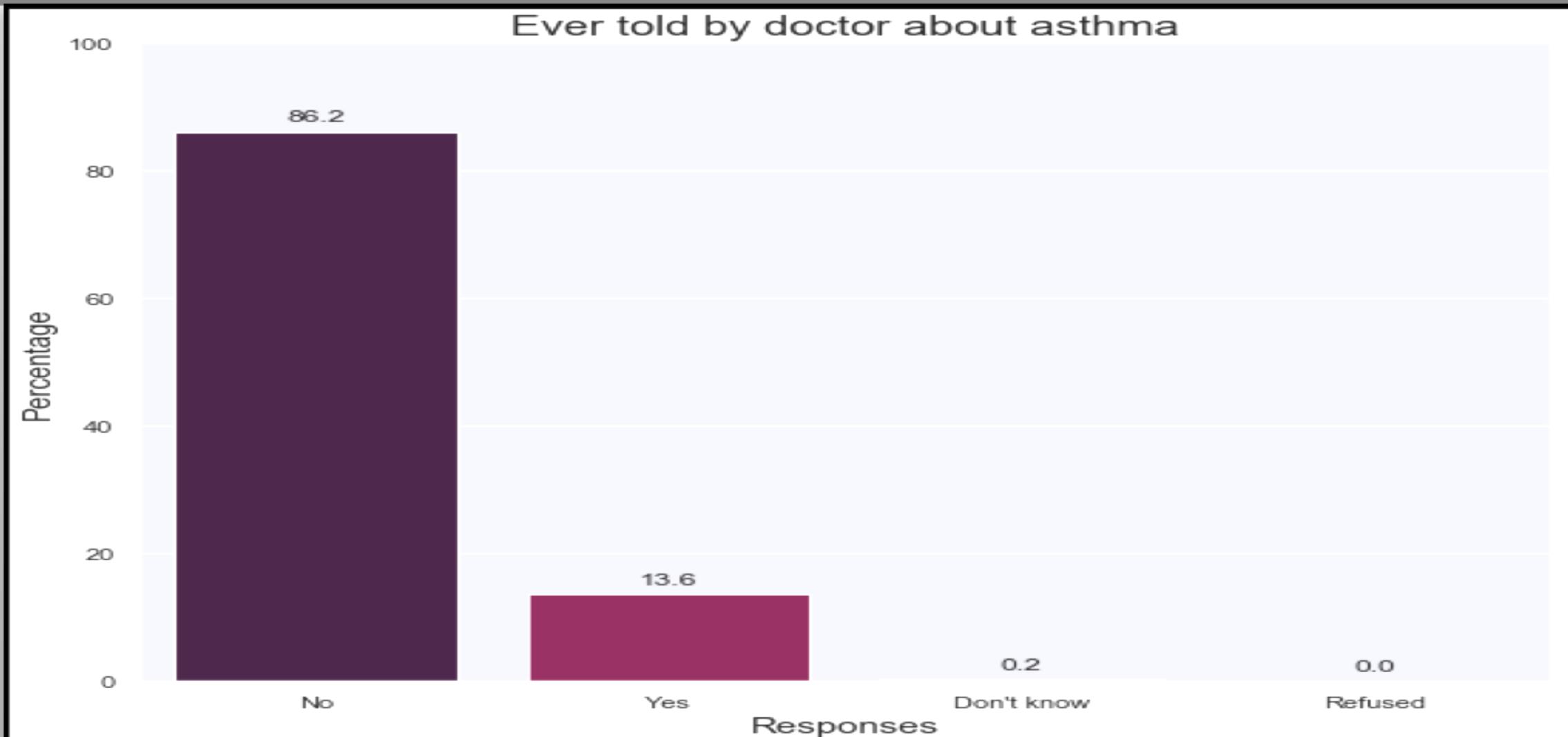
## Objective-3 To find out how much percentage of US residents ever told about asthma by their doctor.

### Python coding-

```
astd=cbrfssci.ASTHMA3.value_counts()
astd1=astd.to_frame()
astd1["Percentage"]=round((astd1/astd1.sum())*100,2)
astd1["Category"]={"No":2.0,"Yes":1.0,"Don't know":7.0,"Refused":9.0}
astdbar=sns.barplot(data=astd1,x="Category",y="Percentage",\
                     palette="rocket",)
plt.title("Ever told by doctor about asthma")
plt.xlabel("Responses")
plt.ylim(0,100)
rects = astdbar.patches
for rect in rects:
    y_value = rect.get_height()
    x_value = rect.get_x() + rect.get_width() / 2
    space = 5
    va = 'bottom'
    label = "{:.1f}".format(y_value)
    plt.annotate(label,(x_value, y_value),xytext=(0, space),\
                textcoords="offset points",ha='center',va=va)
```



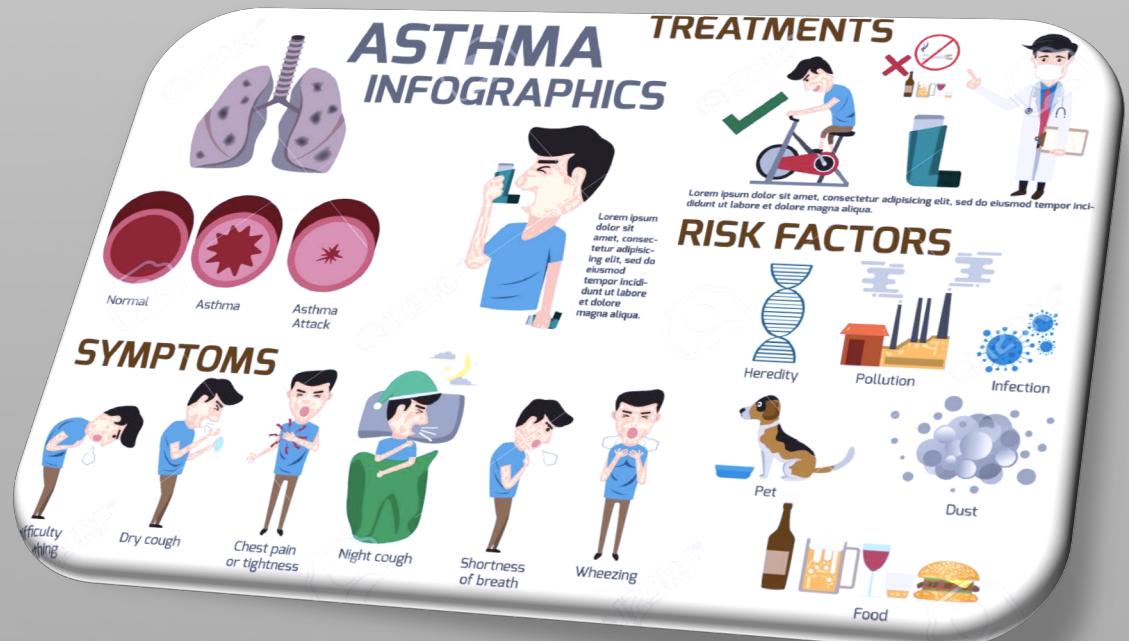
# Output:





## Conclusion-

- Among total respondents only 13.6% of respondents told by their doctor that they have asthma.

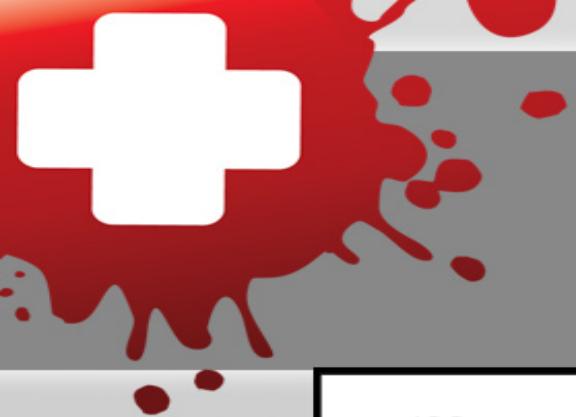




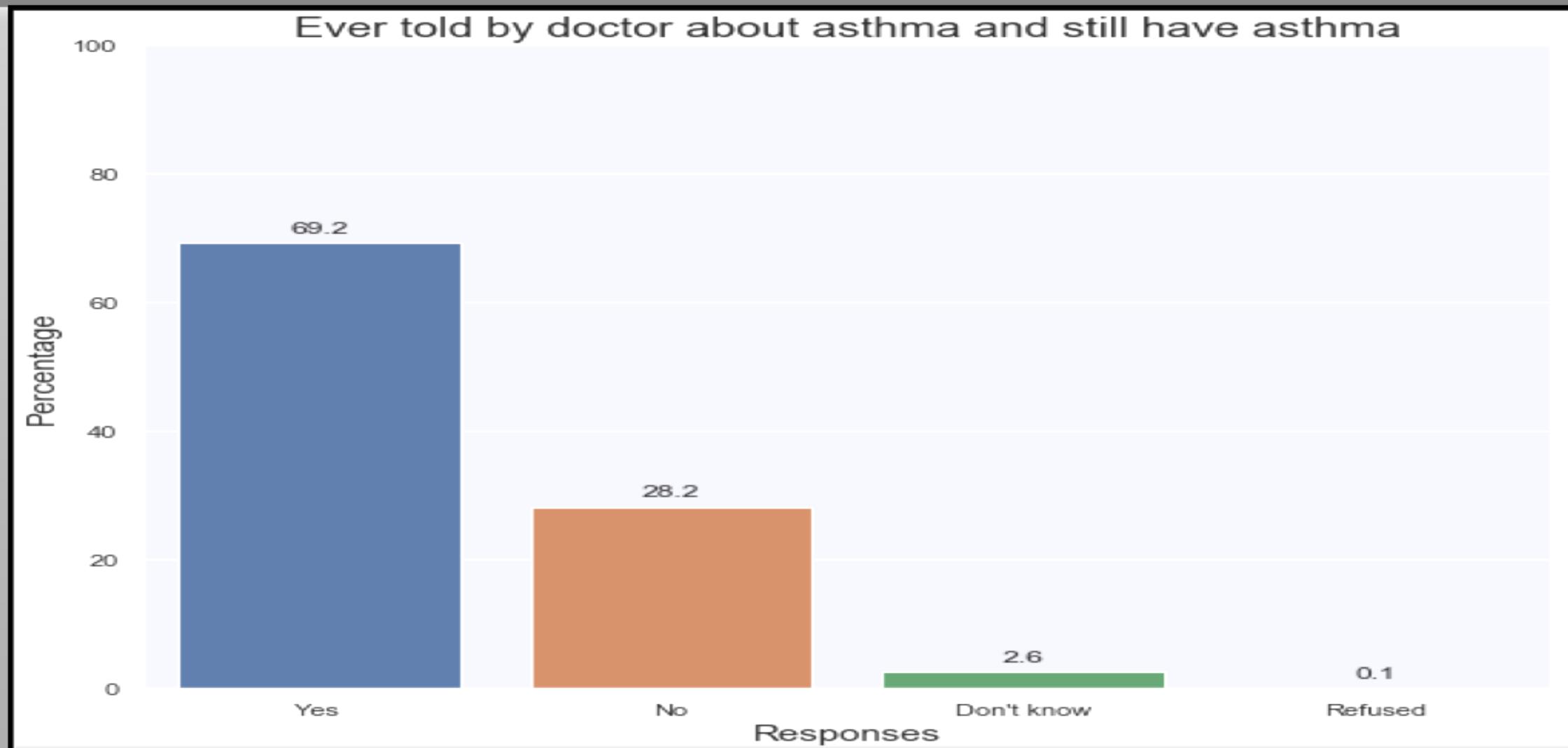
## Objective-4 Among those who ever told about asthma by doctor how many still have asthma?

### Python coding-

```
astdy=cbrfssci[(cbrfssci.ASTHMA3==1)]
asc=astdy.ASTHNOW.value_counts()
asc1=asc.to_frame()
asc1["Percentage"]=round((asc1/asc1.sum())*100,2)
asc1["Category"]={"Yes":1.0,"No":2.0,"Don't know":7.0,"Refused":9.0}
ascdbar=sns.barplot(data=asc1,x="Category",y="Percentage")
plt.title("Ever told by doctor about asthma and still have asthma")
plt.xlabel("Responses")
plt.ylim(0,100)
rects = ascdbar.patches
for rect in rects:
    y_value = rect.get_height()
    x_value = rect.get_x() + rect.get_width() / 2
    space = 5
    va = 'bottom'
    label = "{:.1f}".format(y_value)
    plt.annotate(label,(x_value, y_value),xytext=(0, space),\
                textcoords="offset points",ha='center',va=va)
```



# Output:





## Conclusion-

- Among 13.6 % told by doctor about asthma still 69.2 % of respondents are suffering from Asthma.





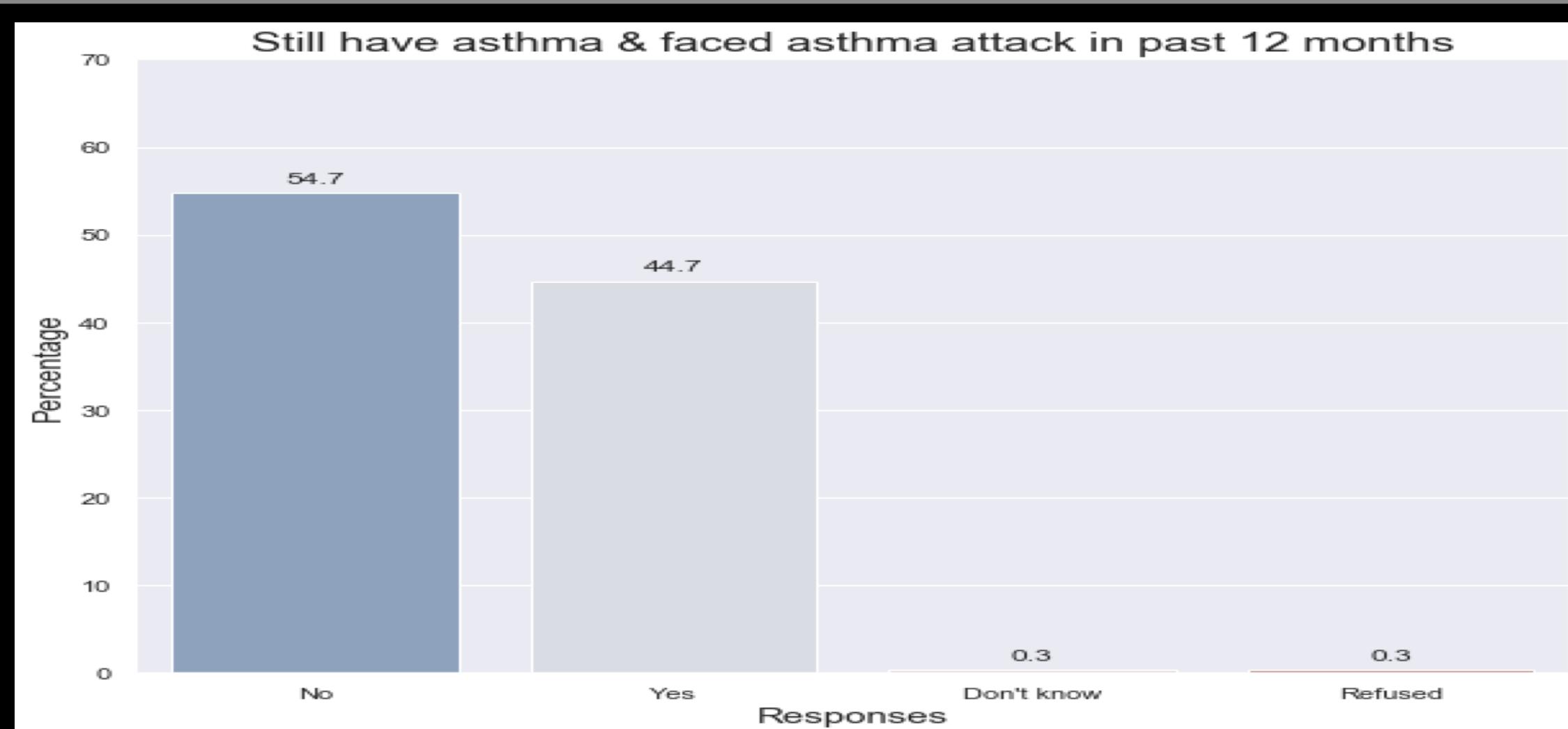
## Objective-5 Among those who still have asthma how many had an episode of asthma or an asthma attack in past 12 months?

### Python coding-

```
ascy=cbrfssci[(cbrfssci.ASTHNOW==1)]
asat=ascy.ASATTACK.value_counts()
asat1=asat.to_frame()
asat1["Percentage"]=round((asat1/asat1.sum())*100,2)
asat1["Category"]={"No":2.0,"Yes":1.0,"Don't know":7.0,"Refused":9.0}
asatdbar=sns.barplot(data=asat1,x="Category",y="Percentage",\
                      palette="vlag")
plt.title("Still have asthma & faced asthma attack in past 12 months")
plt.xlabel("Responses")
plt.ylim(0,70)
rects = asatdbar.patches
for rect in rects:
    y_value = rect.get_height()
    x_value = rect.get_x() + rect.get_width() / 2
    space = 5
    va = 'bottom'
    label = "{:.1f}".format(y_value)
    plt.annotate(label,(x_value, y_value),xytext=(0, space),\
                textcoords="offset points",ha='center',va=va)
```



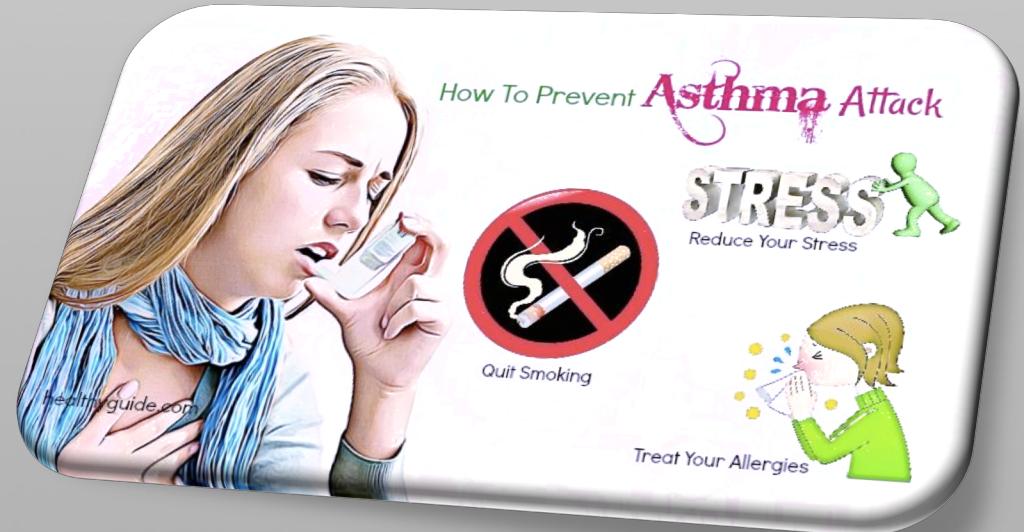
# Output:





## Conclusion-

- Among total existing asthmatic respondents 44.7% have faced asthma attack in past 12 months.





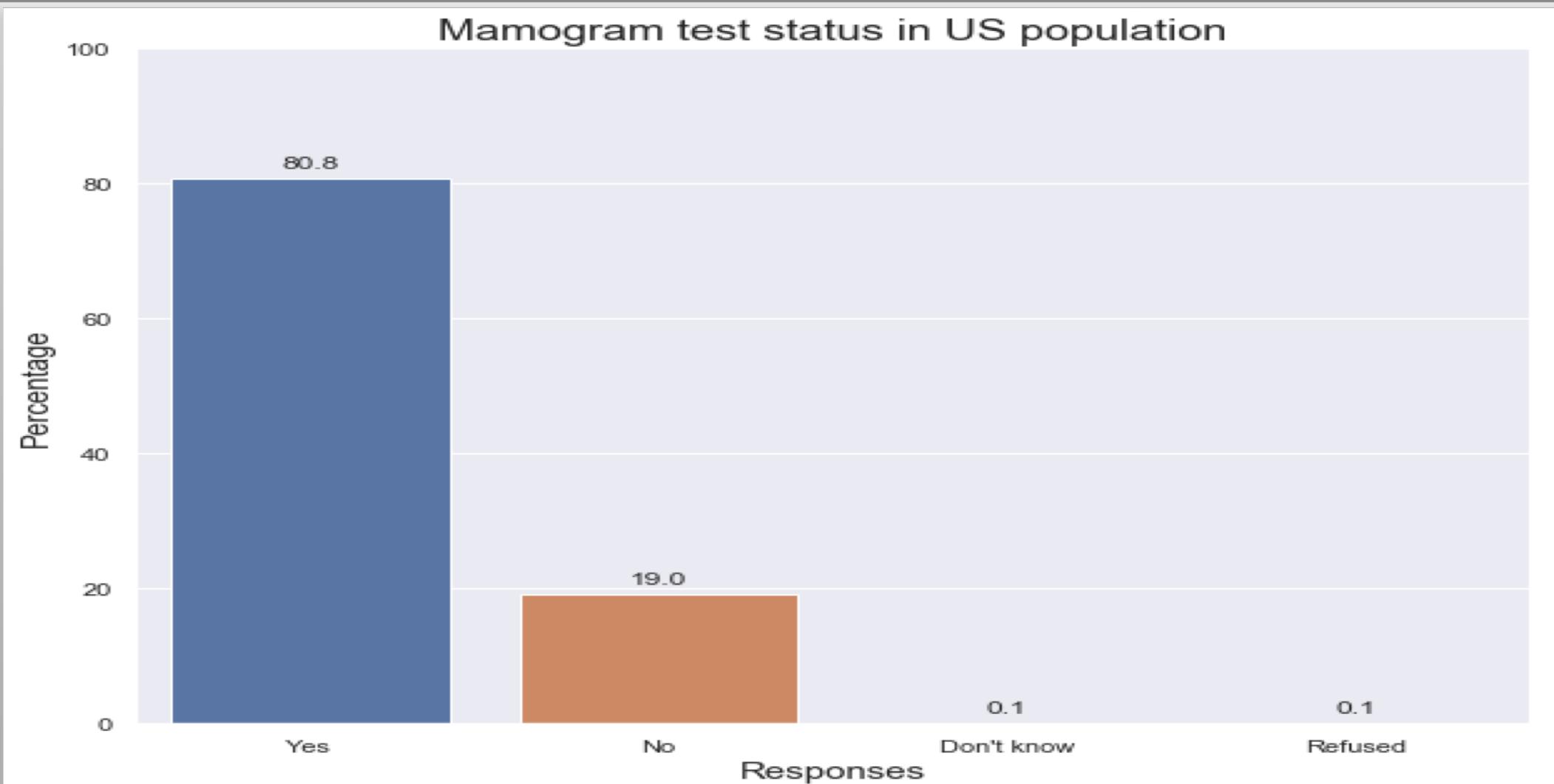
# Objective -6 To find out the existing Breasts cancer screening rate among US residents.

## Python coding-

```
bcf=cbrfssci[(cbrfssci.SEX==2)]
bcs=bcf.HADMAM.value_counts()
bcs1=bcs.to_frame()
bcs1["Percentage"]=round((bcs1/bcs1.sum())*100,2)
bcs1["Category"]={"Yes":1.0,"No":2.0,"Don't know":7.0,"Refused":9.0}
bcsbar=sns.barplot(data=bcs1,x="Category",y="Percentage")
plt.title("Mamogram test status in US population")
plt.xlabel("Responses")
plt.ylim(0,100)
rects = bcsbar.patches
for rect in rects:
    y_value = rect.get_height()
    x_value = rect.get_x() + rect.get_width() / 2
    space = 5
    va = 'bottom'
    label = "{:1f}".format(y_value)
    plt.annotate(label,(x_value, y_value),xytext=(0, space),\
                textcoords="offset points",ha='center',va=va)
```



# Output:





## Conclusion-

- Among total female respondents 80.2 % of them has undergone for mammogram which shows breast cancer awareness among females.





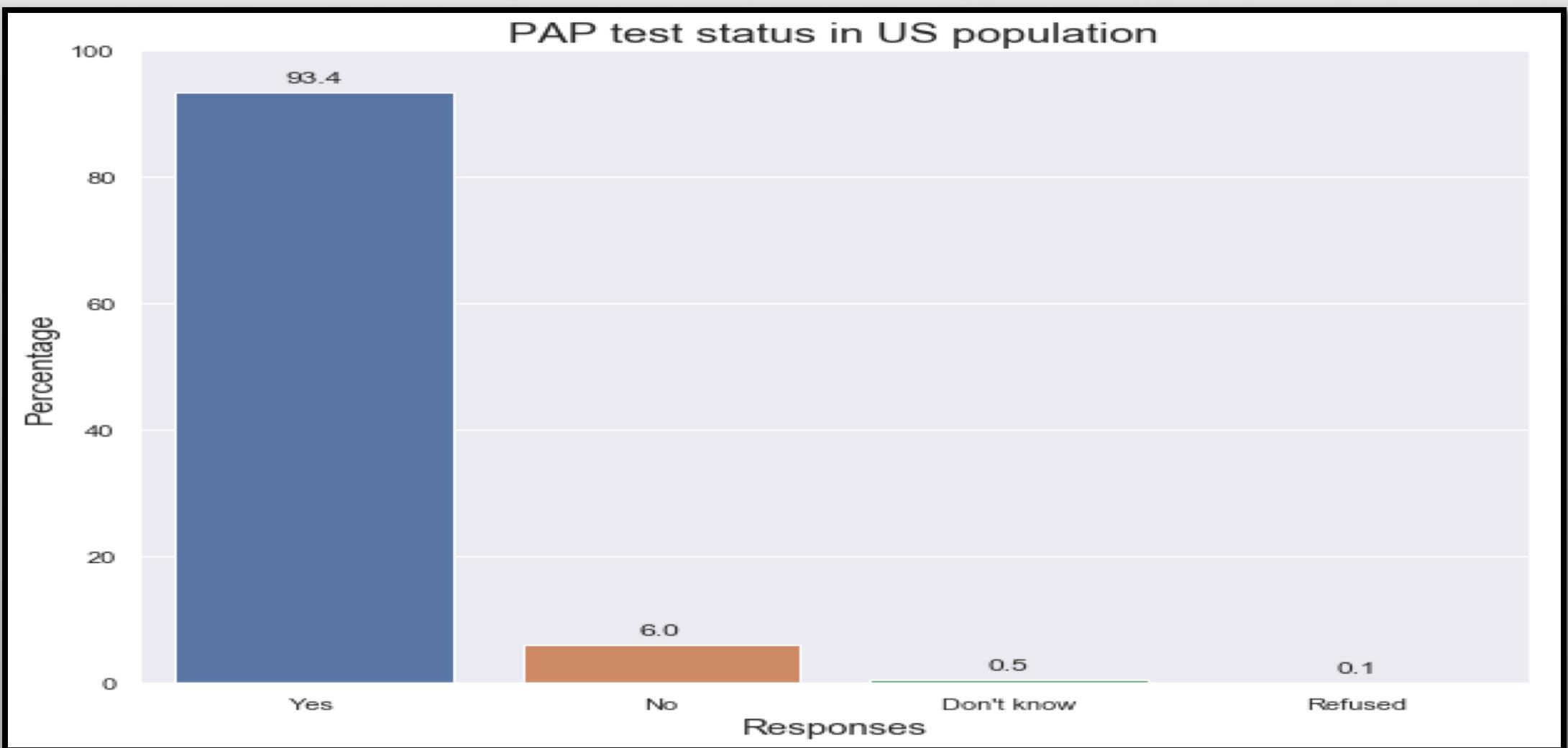
## Objective-7 To find out the existing Cervical cancer screening rate among US residents.

### Pyhton coding-

```
ccf=cbrfssci[(cbrfssci.SEX==2)]
ccs=ccf.HADPAP2.value_counts()
ccs1=ccs.to_frame()
ccs1["Percentage"]=round((ccs1/ccs1.sum())*100,2)
ccs1["Category"]={"Yes":1.0,"No":2.0,"Don't know":7.0,"Refused":9.0}
ccsbar=sns.barplot(data=ccs1,x="Category",y="Percentage")
plt.title("PAP test status in US population")
plt.xlabel("Responses")
plt.ylim(0,100)
rects = ccsbar.patches
for rect in rects:
    y_value = rect.get_height()
    x_value = rect.get_x() + rect.get_width() / 2
    space = 5
    va = 'bottom'
    label = "{:.1f}".format(y_value)
    plt.annotate(label,(x_value, y_value),xytext=(0, space),\
                textcoords="offset points",ha='center',va=va)
```



# Output:





## Conclusion-

- Among total female respondents 93.4 % of them has undergone for PAP test which shows cervical cancer awareness among females.





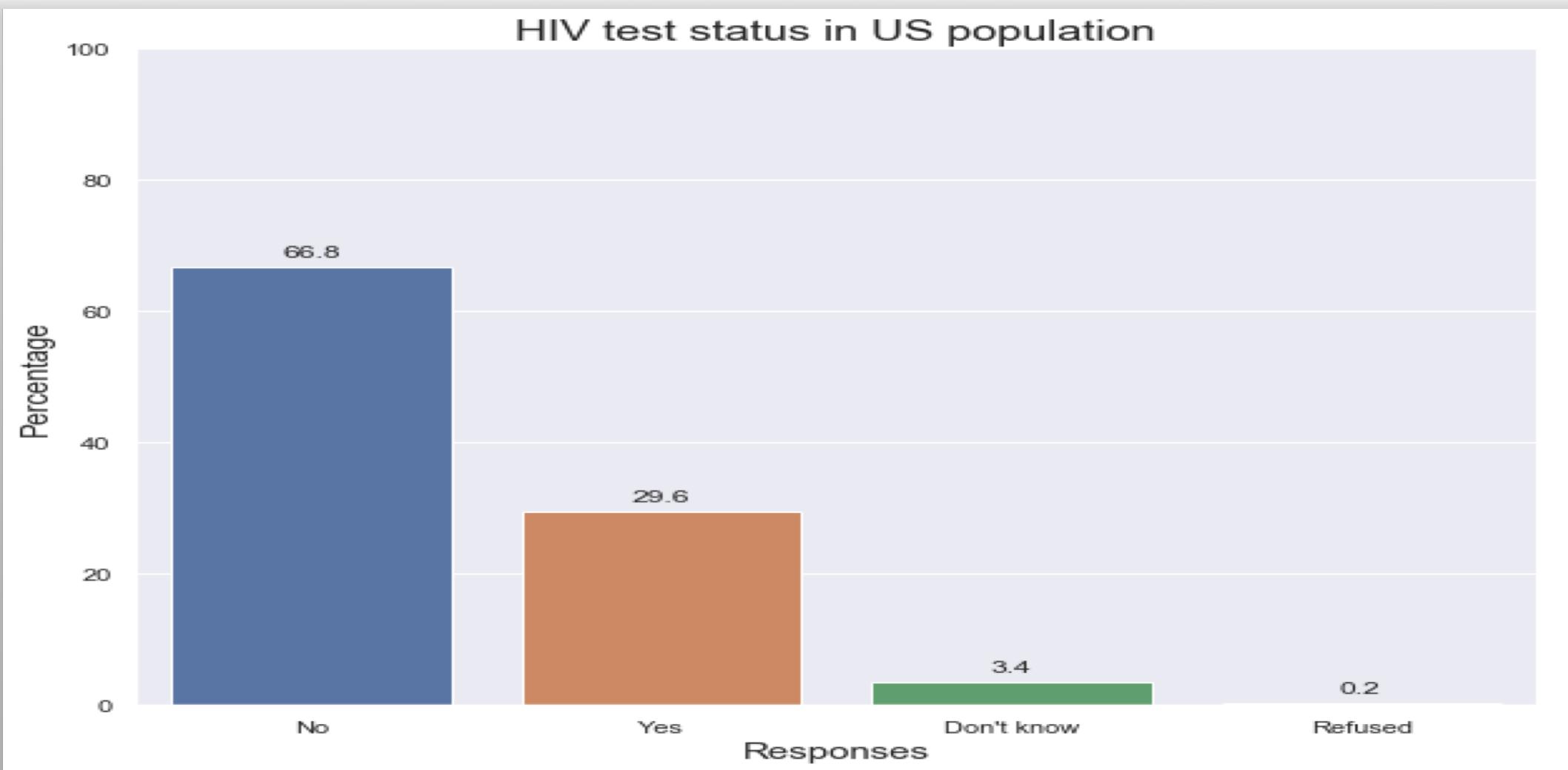
# Object-8 To find out the existing HIV testing rate among US residents.

- **Python coding-**

```
hivtst=cbrfssci.HIVTST6.value_counts()
hivtst1=hivtst.to_frame()
hivtst1["Percentage"]=round((hivtst1/hivtst1.sum())*100,2)
hivtst1["Category"]={"No":2.0,"Yes":1.0,"Don't know":7.0,"Refused":9.0}
hivtst1
hivbar=sns.barplot(data=hivtst1,x="Category",y="Percentage")
plt.title("HIV test status in US population")
plt.xlabel("Responses")
plt.ylim(0,100)
rects = hivbar.patches
for rect in rects:
    y_value = rect.get_height()
    x_value = rect.get_x() + rect.get_width() / 2
    space = 5
    va = 'bottom'
    label = "{:.1f}".format(y_value)
    plt.annotate(label,(x_value, y_value),xytext=(0, space),\
                textcoords="offset points",ha='center',va=va)
plt.rcParams['figure.figsize']=6,5
```



# Output:





## Conclusion-

- Among total respondents 29.6% of respondents has undergone for HIV test which shows that may be they think that they might get affected by HIV, but still percentage is less, but still awareness can be increased.





*Thank the People  
Who Care*