Introduction:

This is a report for analyzing (No-Show Appointment) data . This dataset collects information from 100k medical appointments in Brazil and is focused on the question of whether patients show up for their appointment. A number of characteristics about the patient are included in each row.

I used some libraries in analysis :

Pandas

Numpy

Seaborn

Matplotlib

This report is an answer for some questions :

(1)What is the age range of the patients?

(2) What are the top 10 neighborhoods where patients come from?

(3)what is the most common day which patients have an appointment?

(4) who have more scholarships females or males ?

(5) who came to the appointment more ?(males or females)

(6) Did the patients with scholarship come to the appointment more than the patients without it?

(7) who came more to the appointment the patients with diabetes or the patients without it?

(8) who came more to the appointment the patients who have a handicap or did not have?

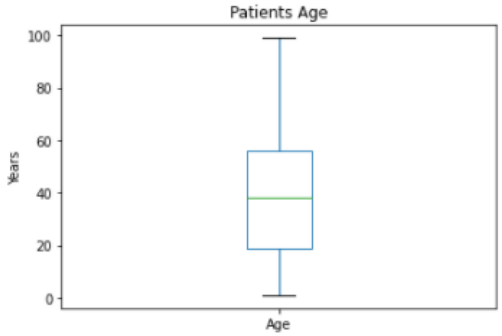
(9) who came more to the appointment the patients who are alcoholism or not?

(10) who came more to the appointment the patients who have hypertension or not?

(11) what is the age group for most patient?

(1)What is the age range of the patients?

```
In [86]: df_clean.Age.plot(kind='box');
         plt.title('Patients Age');
         plt.ylabel("Years");
```

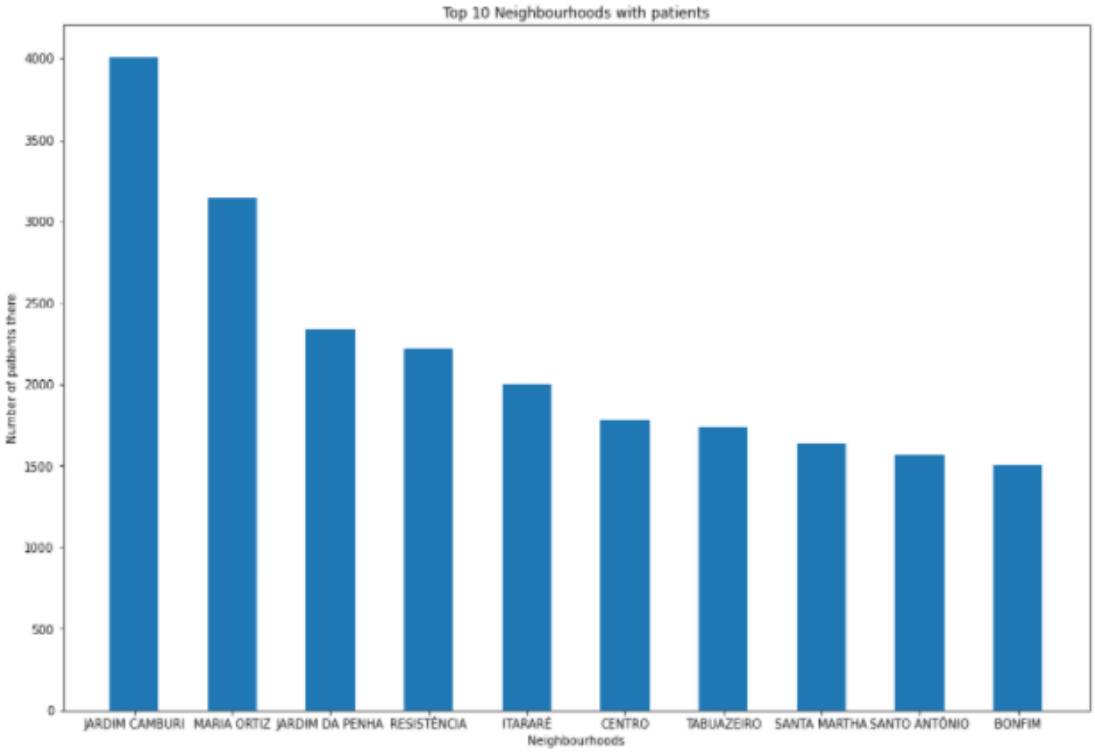


Answer: the age range is between 1 and 99.

---

(2) What are the top 10 neighborhoods where patients come from?

```
In [67]: Neighbourhood=['JARDIM CAMBURI','MARIA ORTIZ','JARDIM DA PENHA','RESISTÊNCIA','ITARARÉ','CENTRO','TABUAZEIRO','SANTA MARTHA',
                        'SANTO ANTÔNIO','BONFIM']
         fig=plt.figure()
         ax = fig.add_axes([1,1,2,2])
         ax.bar(Neighbourhood,Top10_Neighbourhood,width = 0.5)
         ax.set_title('Top 10 Neighbourhoods with patients')
         ax.set_ylabel('Number of patients there')
         ax.set_xlabel('Neighbourhoods')
         plt.show()
```



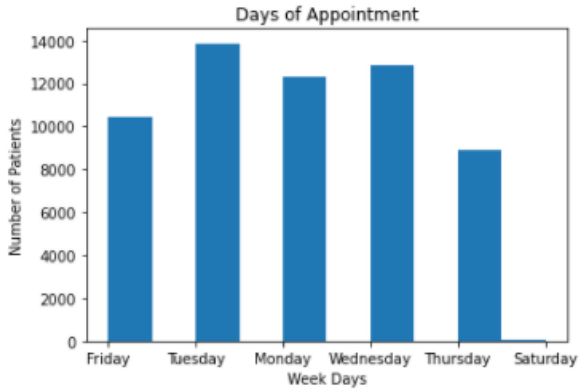
Answer: The top 10 are

JARDIM CAMBURI	4006
MARIA ORTIZ	3146
JARDIM DA PENHA	2335
RESISTÊNCIA	2215
ITARARÉ	1996
CENTRO	1779
TABUAZEIRO	1736
SANTA MARTHA	1633
SANTO ANTÔNIO	1567
BONFIM	1502

---

(3)what is the most common day which patients have an appointment?

```
In [90]: plt.hist(df_clean.Day_Of_Appointment);
         plt.title('Days of Appointment');
         plt.ylabel("Number of Patients");
         plt.xlabel("Week Days");
```

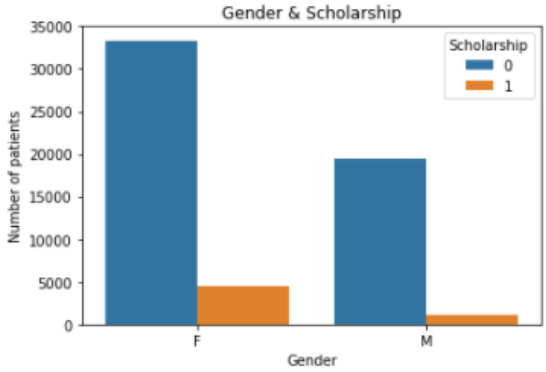


Answer: Tuesday is the most common day.

---

(4) who have more scholarships females or males ?

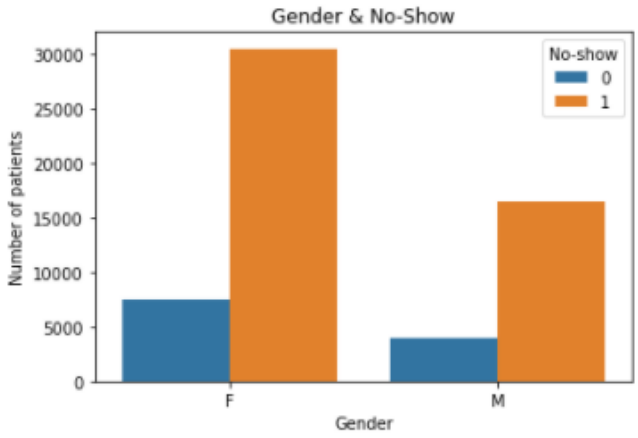
```
In [71]: sns.countplot(x=df_clean.Gender,hue=df_clean.Scholarship);
         plt.ylabel("Number of patients");
         plt.title("Gender & Scholarship");
```



Answer: Females have scholarship more than males.

(5) who came to the appointment more? (males or females)

```
In [72]:
#1 --> show up their appointment
#0---> don't show up their appointment
sns.countplot(x=df_clean.Gender,hue=df_clean['No-show']);
plt.ylabel("Number of patients");
plt.title("Gender & No-Show");
```

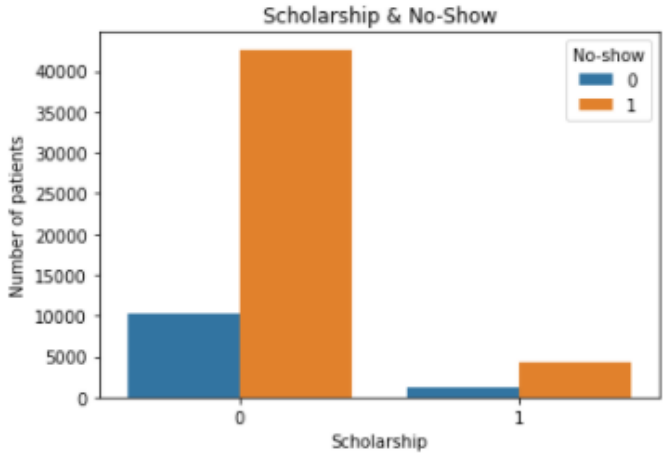


Answer: Females came more than males.

---

(6) Did the patients with scholarship come to the appointment more than the patients without it?

```
In [73]:
sns.countplot(x=df_clean.Scholarship,hue=df_clean['No-show']);
plt.ylabel("Number of patients");
plt.title("Scholarship & No-Show");
```

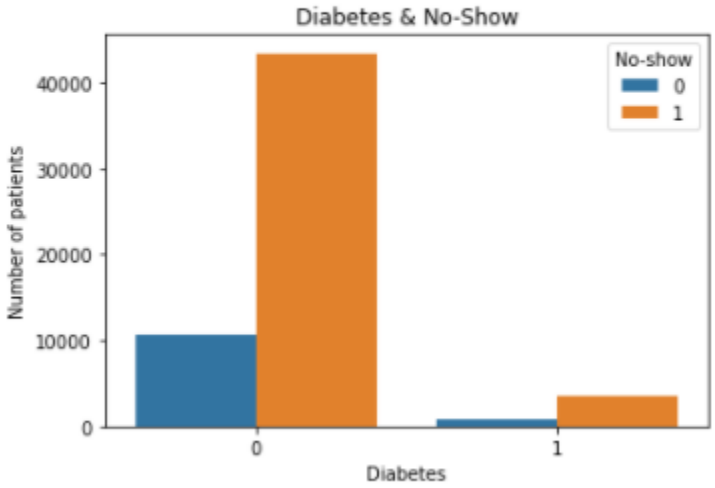


Answer: The patients without scholarship came more than with scholarship.

---

(7) who came more to the appointment the patients with diabetes or the patients without it?

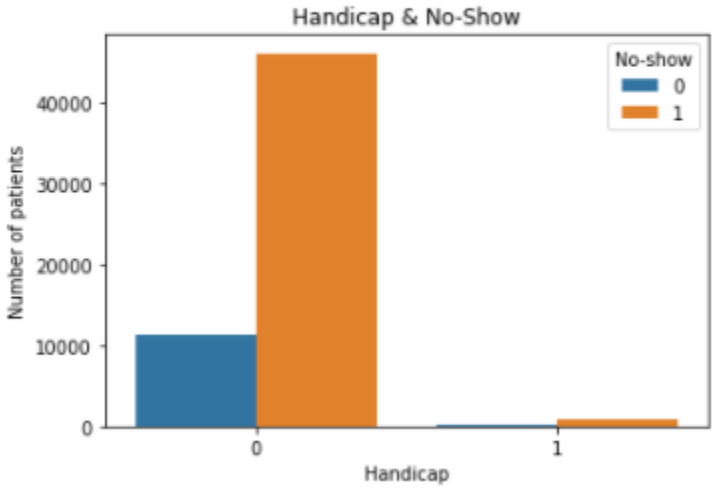
In [74]:
```
sns.countplot(x=df_clean['Diabetes'],hue=df_clean['No-show']);
plt.title("Diabetes & No-Show");
plt.ylabel("Number of patients");
```



Answer: The patients which did not have diabetes come more than the patients which have.

(8) who came more to the appointment the patients who have a handicap or did not have?

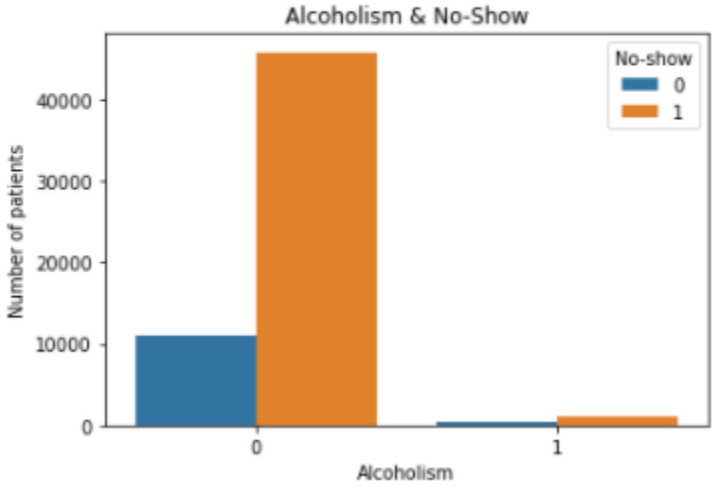
```
sns.countplot(x=df_clean['Handicap'],hue=df_clean['No-show']);
plt.ylabel("Number of patients");
plt.title("Handicap & No-Show");
```



Answer: The patients did not have come more.

(9) who came more to the appointment the patients who are alcoholism or not?

```
sns.countplot(x=df_clean['Alcoholism'],hue=df_clean['No-show']);
plt.ylabel("Number of patients");
plt.title("Alcoholism & No-Show");
```

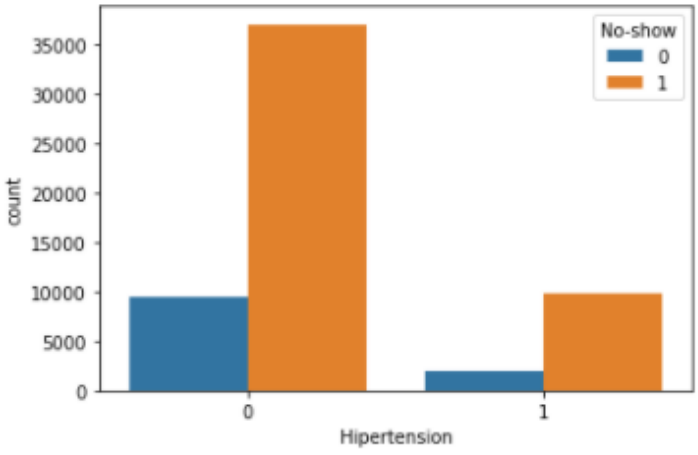


Answer: The patients aren't have .

---

(10) who came more to the appointment the patients who have hypertension or not?

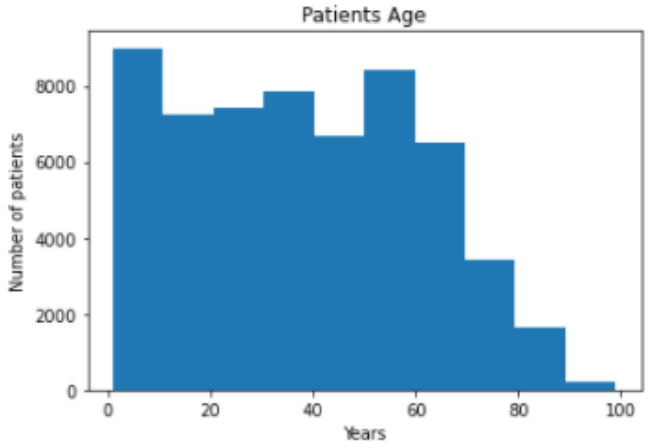
```
In [76]: sns.countplot(x=df_clean['Hipertension'],hue=df_clean['No-show']);
```



Answer: The patients did not have hypertension.

(11) what is the age group for most patient?

```
In [88]: plt.hist(df_clean.Age);
         plt.title('Patients Age');
         plt.ylabel("Number of patients");
         plt.xlabel("Years");
```



Answer: Young people(from 1 to 10 years old)

Conclusion:

We recognized some important information from this visualization:

(1) young people (from 1 to 10 years old) are the most patients and more susceptible to disease.

(2) Most patients came from 10 neighborhoods ('JARDIM CAMBURI','MARIA ORTIZ','JARDIM DA PENHA','RESISTÊNCIA','ITARARÉ','CENTRO','TABUAZEIRO','SANTA MARTHA', 'SANTO ANTÔNIO','BONFIM')

(3) Tuesday is the most common day which patients came to their appointments.

(4)Females are more committed than males and came to the appointments more than males.

(5) Females have scholarship more than males.

(7) The most patients came to their appointments did not have (handicap, diabetes, hypertension and alcoholism).

Limitations:

(1) As we recognized that (handicap, diabetes, hypertension and alcoholism) did not be the most reasons that would make patient go to the appointment….. so what are the most diseases that will make patients come to their appointment?

(2) Some important data was missing like :
(3) What are medicines that patients taking?
(4) How much time patient go to the hospital ?
(5) Why patients did not come to their appointments?