**Your Task - 8:** Your task is to understand and apply the above graphs to any of the data set you want and send me the file along with a report explaining what have you analyzed in these graphs and why you applied that particular graphs to the data.

To

Rishabh Goyal,

From

Pampana Jai Kiran.

Sir, I have successfully completed the task 8.

**Data Analysis Report**

In this Analysis we have gone through the data and understand the nature of the data and I have made the following charts and plots in the colab.

They are

* Histogram
* Scatter plot
* Bar chart(both horizontal and vertical)
* Box plot
* Line chart

**Histogram:**

I have made a histogram chart which show the gender category of the dataset in a histogram chart. Which helps us to understand the gender category and its population an frequency.

**Code:**

plt.hist(df.gender, bins=3, edgecolor='k')

plt.xlabel('Gender')

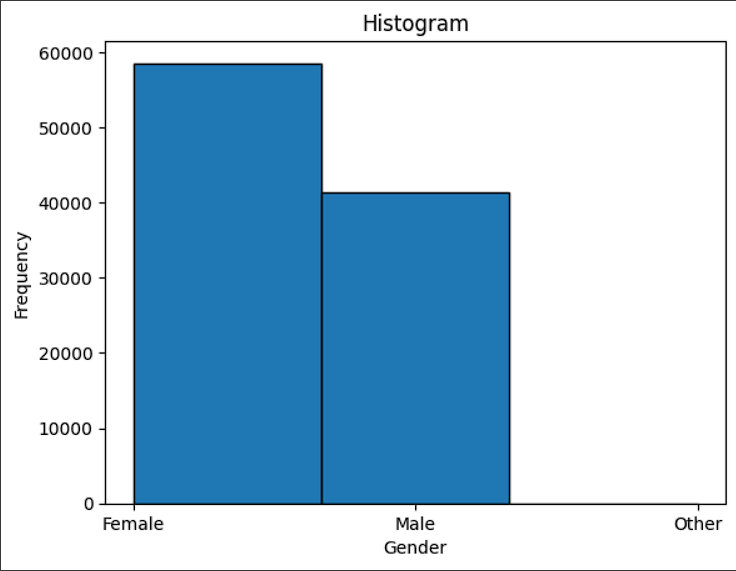
plt.ylabel('Frequency')

plt.title('Histogram')

plt.show()

**Observation:**

* We can observe that the no. of females is more than the no. of males
* We can observe that the no. of females is nearly at 60000.
* We can observe that the no. of male is over 40000.
* The Other is the least popularity.



**Scatter plot:**

I have made a Scatter plot which show the gender category of the dataset which have a Smoking history in a Scatter plot. Which helps us to understand the gender category and its Smoking history which show the current, never, ever, former, etc.

**Code:**

plt.scatter(x=df['gender'], y=df['smoking\_history'], c='g', marker='x', label='Data')

plt.xlabel('Gender')

plt.ylabel('smoking\_history')

plt.title('Scatter Plot')

plt.legend()

plt.show()

**Observation:**

* Overall, more males than females have ever smoked.
* Current smokers are more likely to be male than female.
* There are more former smokers than current smokers, and the majority of former smokers are male.
* The largest group of people are those who have never smoked, and this group is more likely to be female than male.
* There are no former smokers in other category in gender.



**Bar chart:**

I have made a bar chart which show the gender and Age category of the dataset and which is helps us to understand the life span of the gender category of male, female and other.

**Code:**

plt.bar(df.gender,df.age, color='skyblue')

plt.xlabel('Gender')

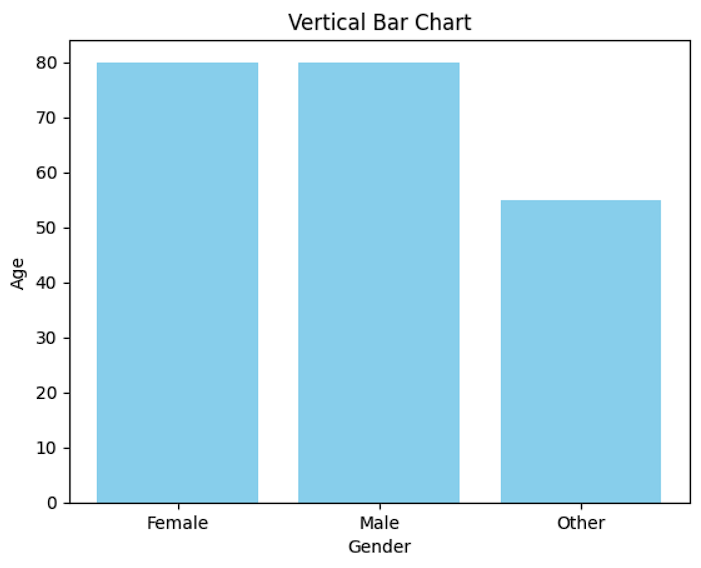
plt.ylabel('Age')

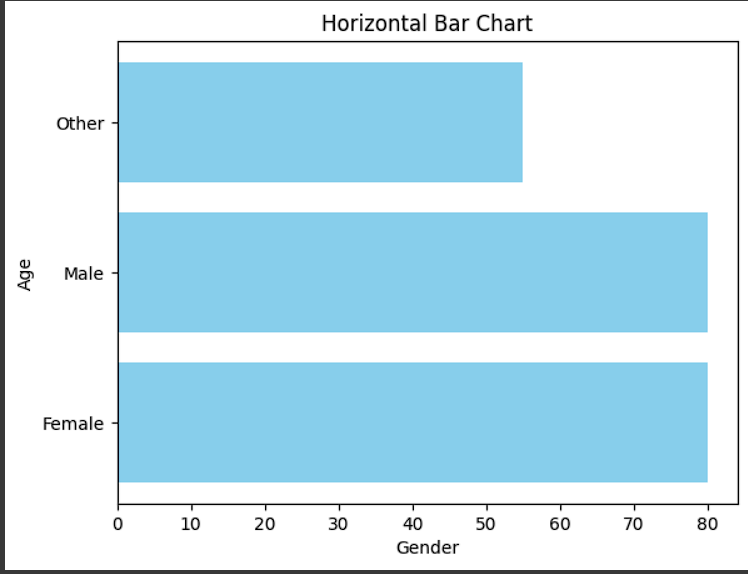
plt.title('Vertical Bar Chart')

plt.show()

**Observation:**

* The chart shows the number of people of each gender in different age groups.
* The difference in the number of males and females is greatest in the oldest age group (80+).
* More people are born female than male.
* Females have a longer life expectancy than males.
* Males are more likely to die from accidents and other causes at a younger age.
* It is important to note that this chart is a general representation of the age and gender distribution of the population.





**Box Plot:**

I have made a bar chart which show the gender and Age categories of the dataset. We can see that the Minimum maximum and average Age of the gender category.

**Code:**

plt.boxplot(df.age)

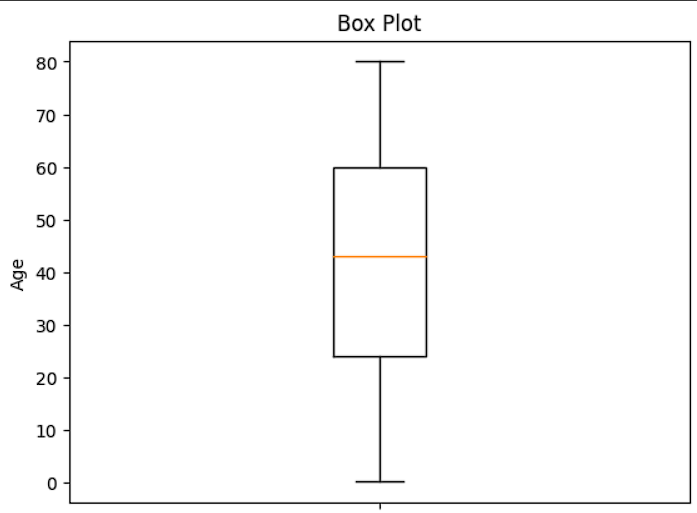
plt.ylabel('Age')

plt.title('Box Plot')

plt.show()

**Observation:**

* we can observe that the minimum age is 1
* we can the Q2 at the more than 20+ years of age.
* We can see the the median or the average age is 40+ years of age.
* We can see the the Q3 at the 60 years of age which the 75% of the populations lifespan.
* We can see that the maximum age is 80 to 80+.
* The yellow line represents the median of Age.



**Line chart:**

I have made a line on the rating of the costomers on the cities in which the line chart shows the number of customers over time. The x-axis represents time, and the y-axis represents the number of customers. The chart shows a steady increase in the number of customers over time, with a few small dips.

**Observations:**

* The number of customers is increasing over time.
* There are a few small dips in the number of customers, but the overall trend is upwards.
* The largest dip in the number of customers occurs at the 400 data point.
* The average rating is a 7star rating
* We can observe that the customers are increasingly by the increase in the rating.

