**Code for**

**Import data set**

**Fill the missing values with mean of column in which null values is present**

**Removing any kind of spaces which cause any problem in getting data**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import csv

data = pd.read\_csv("Project Data.csv")

df= pd.DataFrame(data)

print(df)

df=df.drop(columns=['Timestamp'])

df=df.drop(columns=['Is there is any gum and teeth related diseases have been there in your family trade '])

columns\_with\_missing\_values = df.columns[df.isnull().any()].tolist()

for column in columns\_with\_missing\_values:

if df[column].dtype == 'float64':

df[column].fillna(df[column].mean(), inplace=True)

for column in columns\_with\_missing\_values:

if df[column].dtype == 'object':

df[column].fillna(df[column].mode().iloc[0], inplace=True)

for column in df.columns:

if df[column].dtype == 'object': # Check if the column contains object/string data

df[column] = df[column].str.upper()

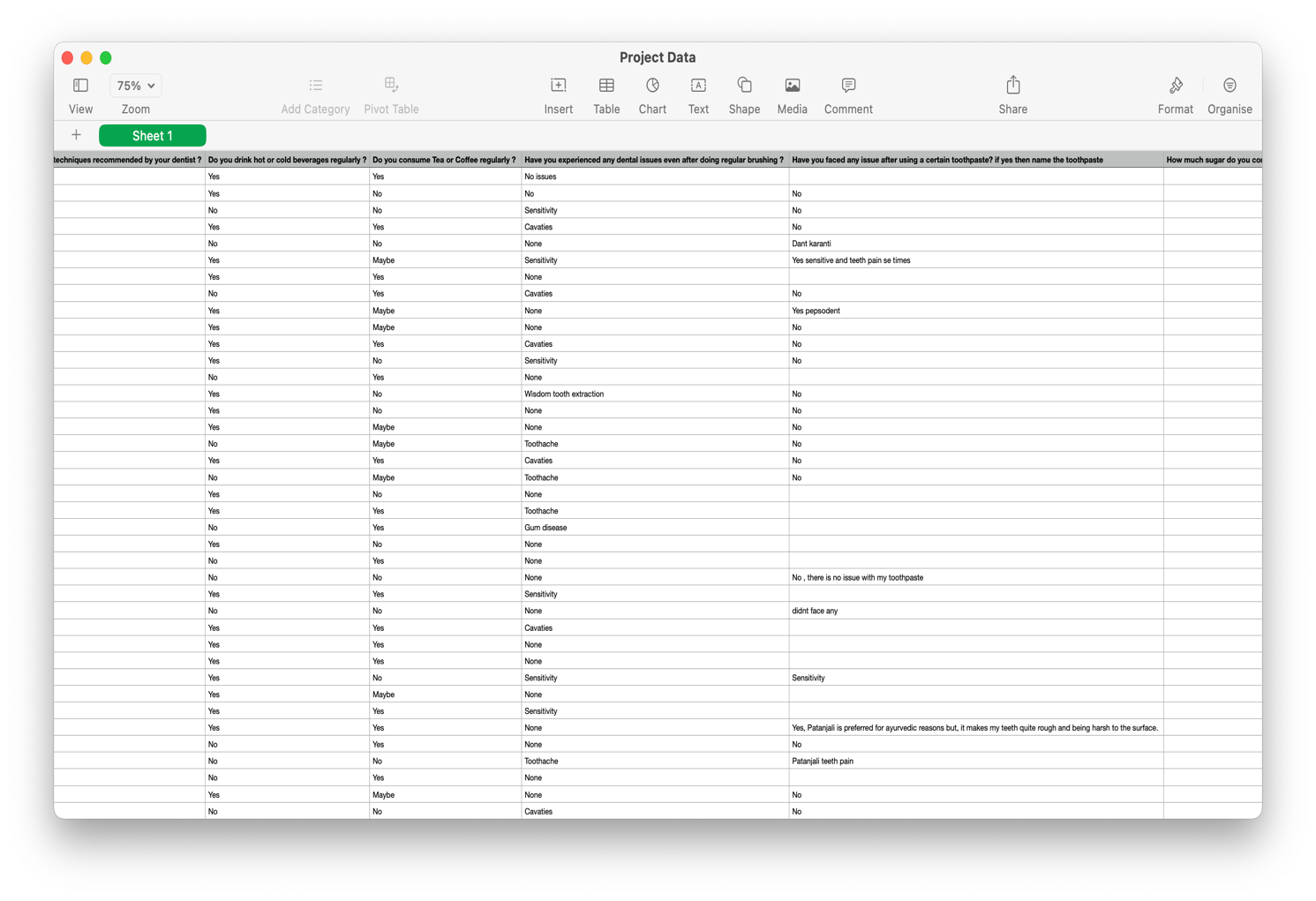
for column in df.columns:

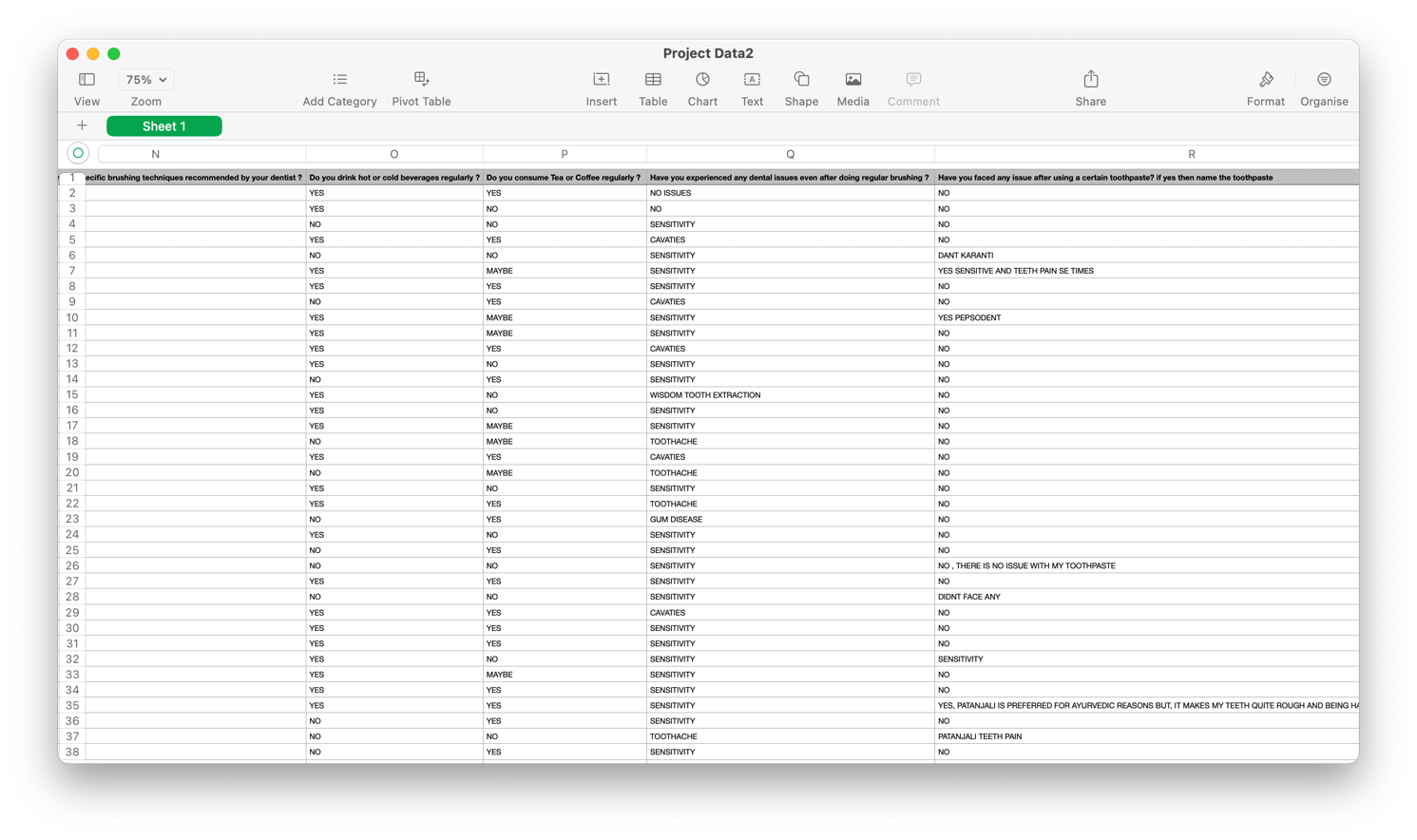
if df[column].dtype == 'object': # Check if the column contains object/string data

df[column] = df[column].str.strip()

df.to\_csv('Project Data2.csv', index=False)

**Data Set before and after cleaning**

**Before**

**After**

**Analysis and Visualization of data set using pandas**

**Data Set 1: Age**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import csv

df = pd.read\_csv('Project Data2.csv')

###################'Age'########################

value\_counts = df['Age '].value\_counts()

plt.figure(figsize=(10, 6))

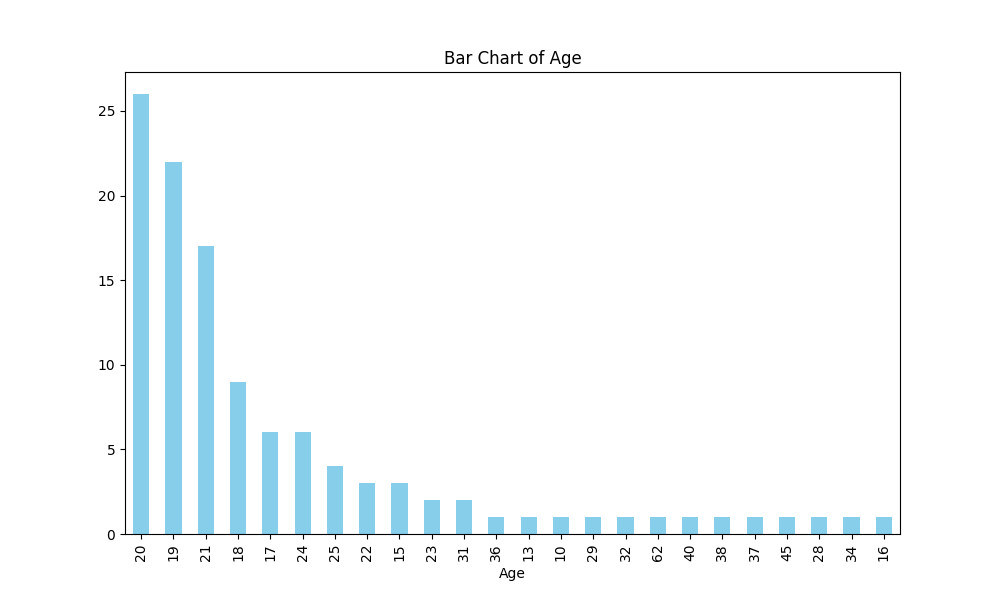
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Bar Chart of Age')

plt.xlabel('Age')

plt.ylabel(' ')

plt.show()

****

**Data Set 2: Gender**

df = pd.read\_csv('Project Data2.csv')

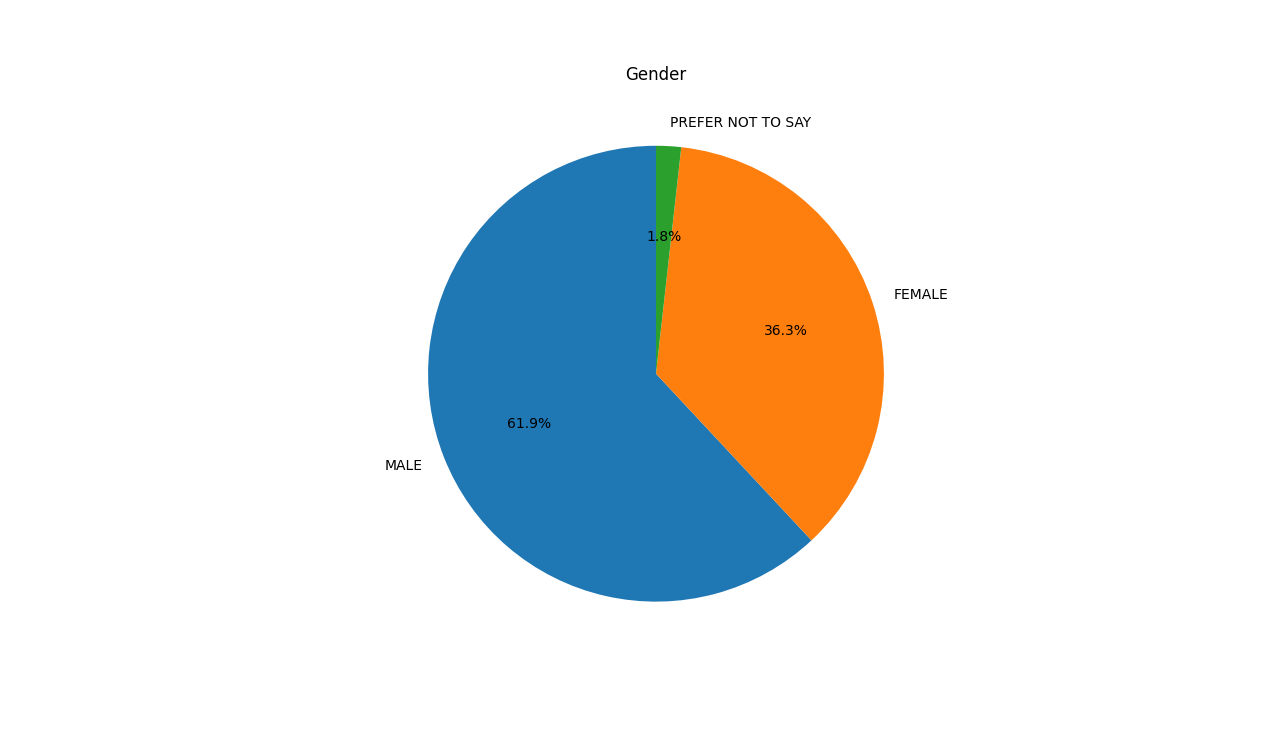
value\_counts = df['Gender '].value\_counts()

plt.figure(figsize=(8, 8))

plt.pie(value\_counts, labels=value\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title(f'Gender')

plt.show()

****

**Data Set 3:** **Which brand of toothpaste do you commonly use ?**

value\_counts = df['Which brand of toothpaste do you commonly use ?'].value\_counts()

plt.figure(figsize=(10, 6))

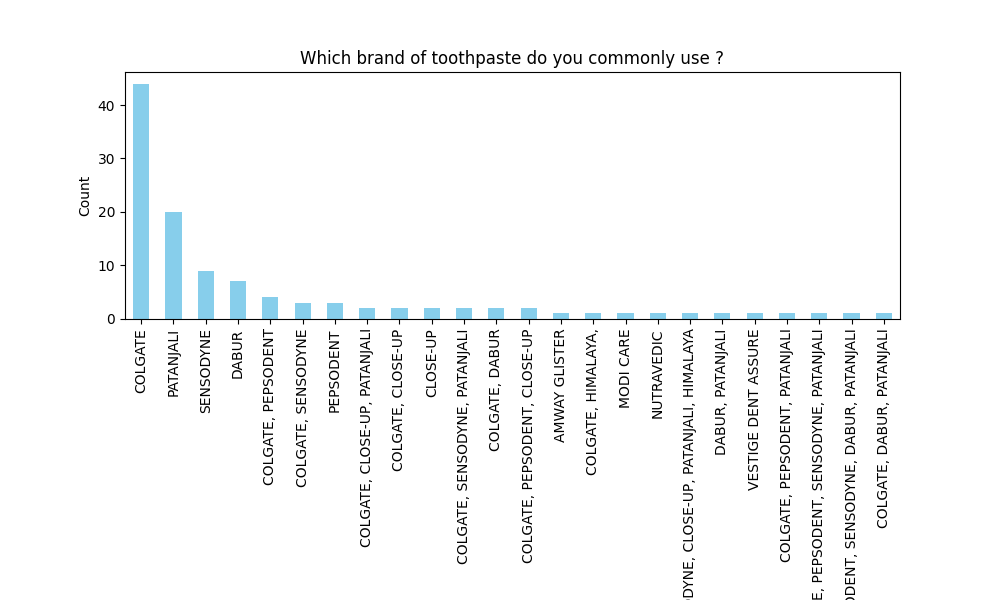
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Which brand of toothpaste do you commonly use ?')

plt.xlabel('Brand')

plt.ylabel('Count')

plt.show()

****

**Data Set 4:** **Do you drink hot or cold beverages regularly ?**

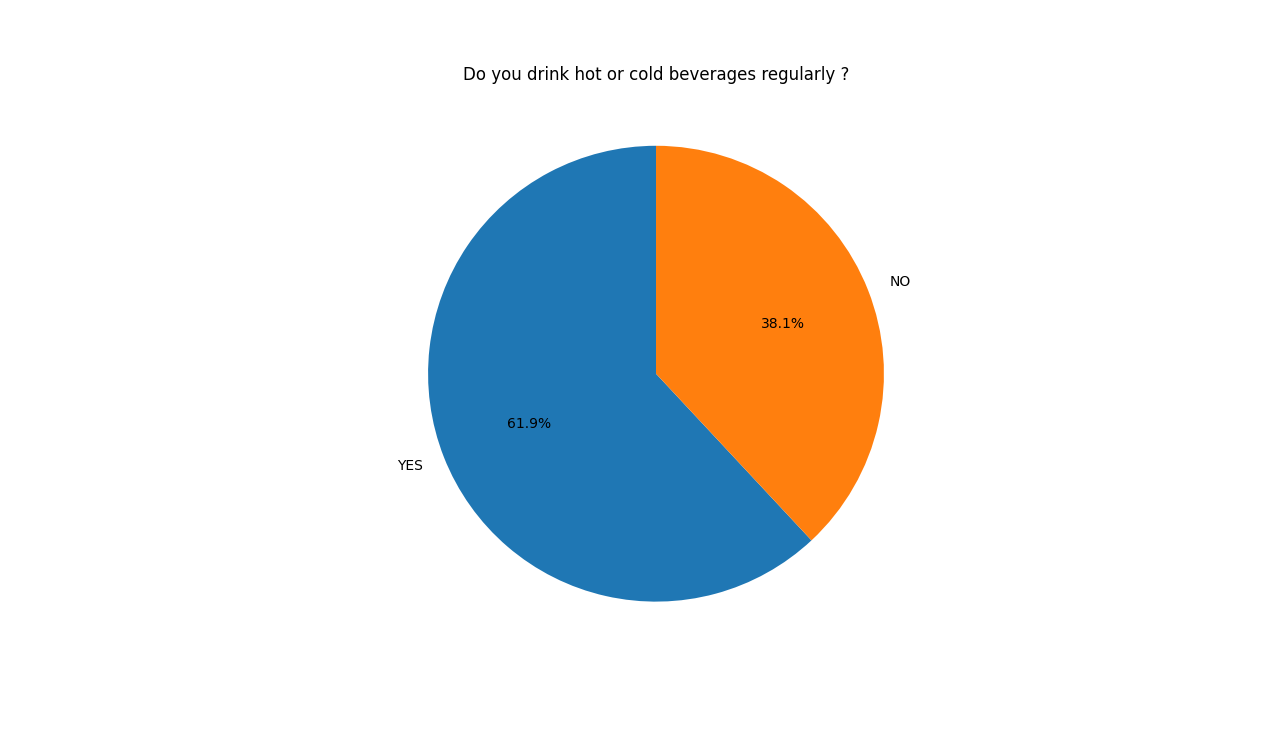
value\_counts = df['Do you drink hot or cold beverages regularly ?'].value\_counts()

plt.figure(figsize=(8, 8))

plt.pie(value\_counts, labels=value\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title(f'Do you drink hot or cold beverages regularly ?')

plt.show()

****

**Data Set 5: 'Are you satisfied with the toothpaste you use ?**

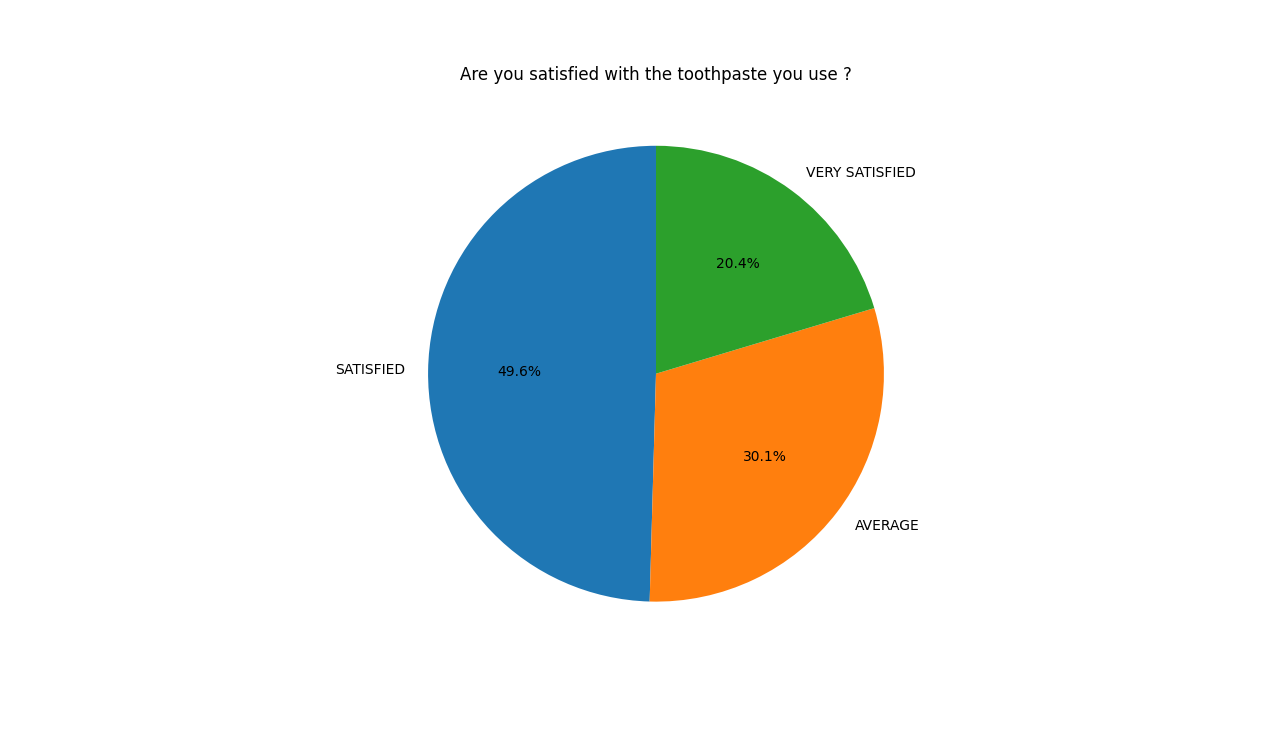
value\_counts = df['Are you satisfied with the toothpaste you use ?'].value\_counts()

plt.figure(figsize=(8, 8))

plt.pie(value\_counts, labels=value\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title(f'Are you satisfied with the toothpaste you use ?')

plt.show()

****

**Data Set 6:** **Have you switched toothpaste brands in the past year? if yes then why?**

value\_counts = df['Have you switched toothpaste brands in the past year? if yes then why?'].value\_counts()

plt.figure(figsize=(10, 6))

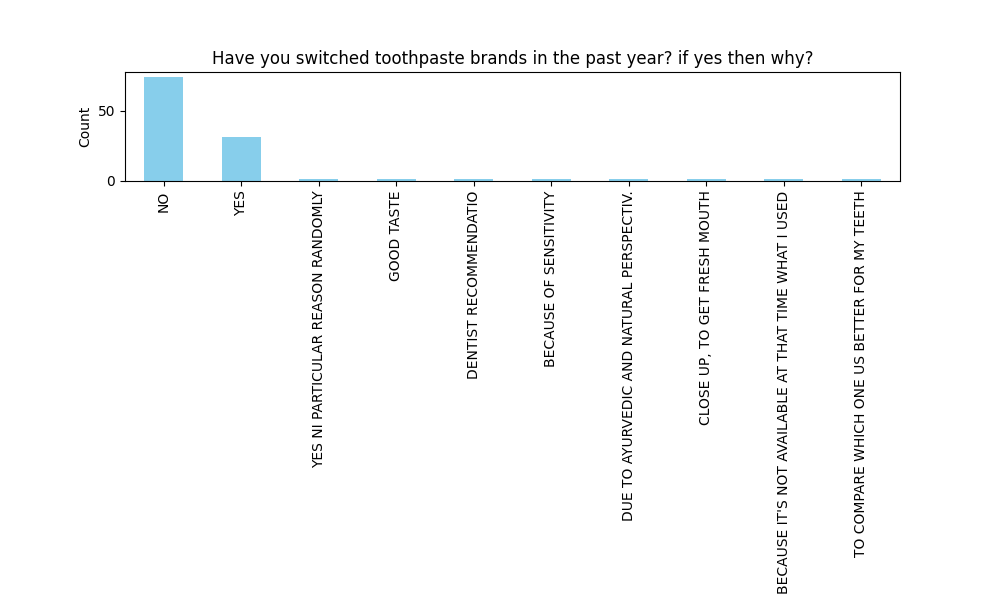
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Have you switched toothpaste brands in the past year? if yes then why?')

plt.xlabel('')

plt.ylabel('Count')

plt.show()

****

**Data Set 7: For how long you brush your teeth ?**

value\_counts = df['For how long you brush your teeth ?'].value\_counts()

plt.figure(figsize=(10, 6))

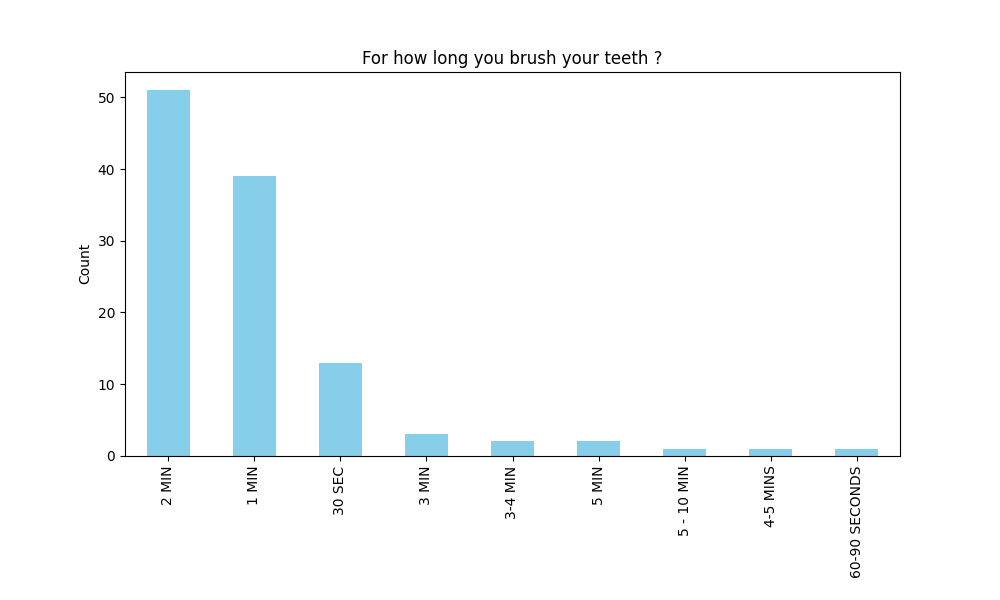
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'For how long you brush your teeth ?')

plt.xlabel('')

plt.ylabel('Count')

plt.show()

****

**Data Set 8: 'Which type of brush are you using ?**

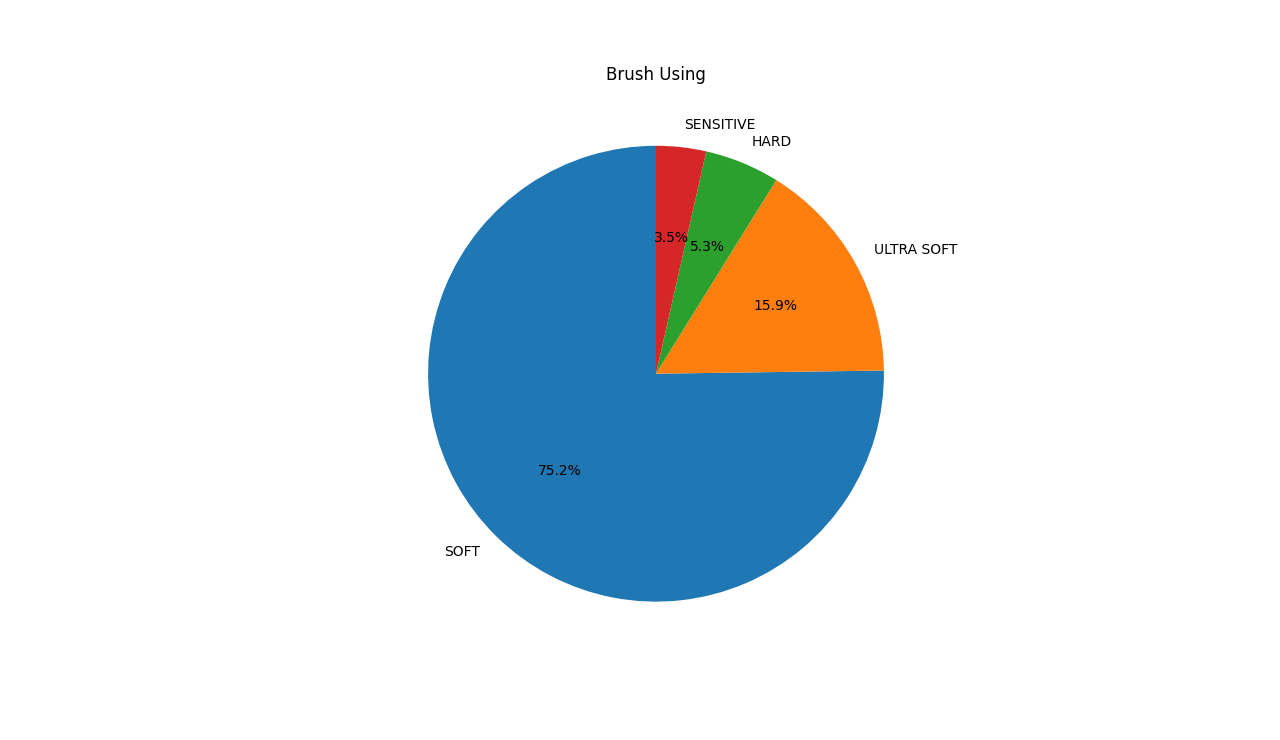
value\_counts = df['Which type of brush are you using ?'].value\_counts()

plt.figure(figsize=(8, 8))

plt.pie(value\_counts, labels=value\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title(f'Brush Using')

plt.show()

****

**Data Set 9: How often do you brush your teeth in a typical day ?**

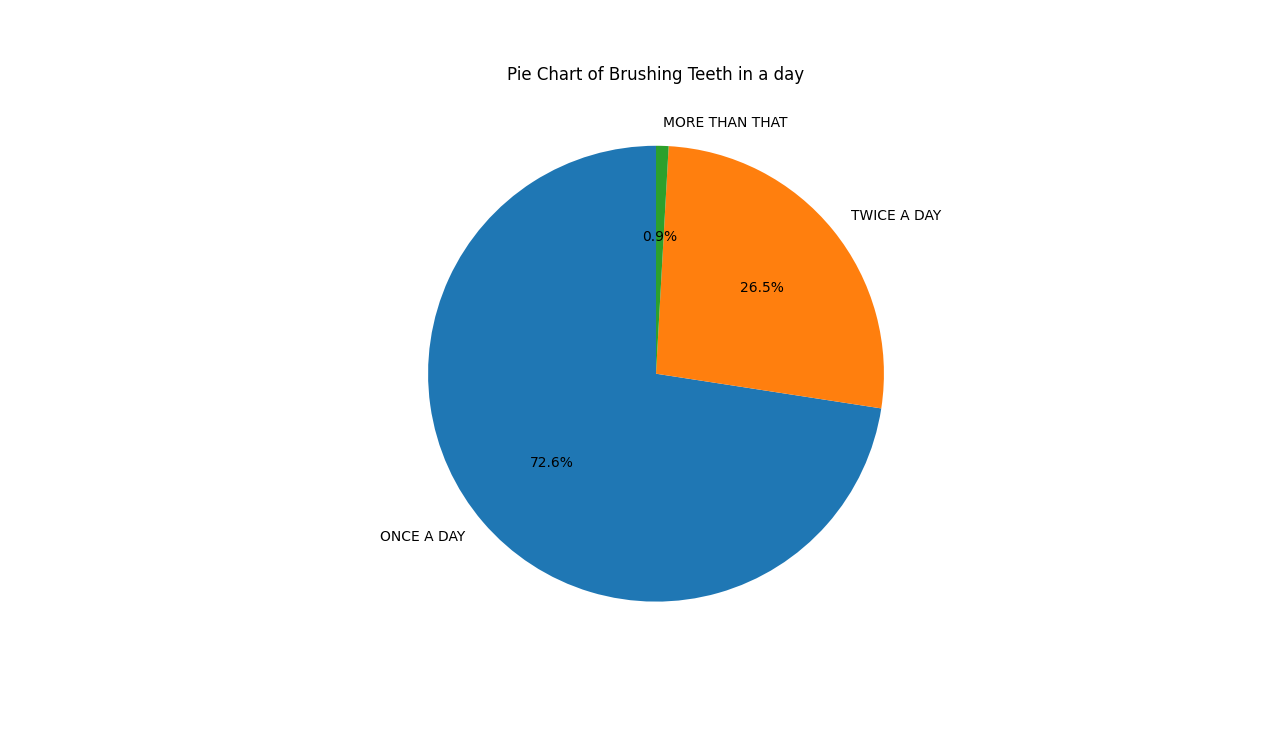
value\_counts = df['How often do you brush your teeth in a typical day ?'].value\_counts()

plt.figure(figsize=(8, 8))

plt.pie(value\_counts, labels=value\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title(f'Pie Chart of Brushing Teeth in a day')

plt.show()

****

**Data set 10:** '**Do you use any additional oral care product such as mouthwash or dental floss ?**

value\_counts = df['Do you use any additional oral care product such as mouthwash or dental floss ?'].value\_counts()

plt.figure(figsize=(10, 6))

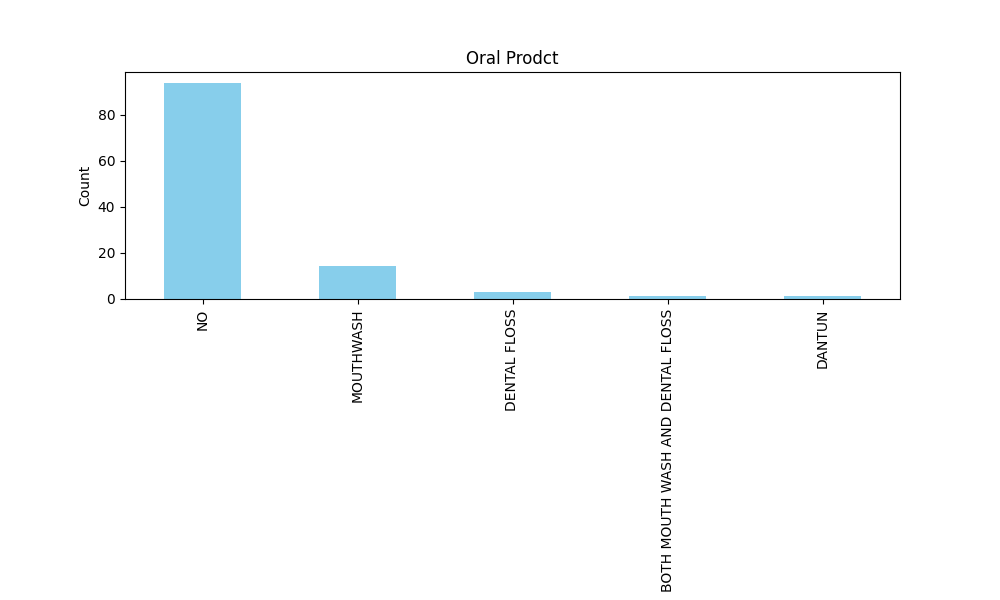
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Oral Prodct')

plt.xlabel('')

plt.ylabel('Count')

plt.show()

****

**Data Set 11: Do you follow any specific brushing techniques recommended by your dentist ?**

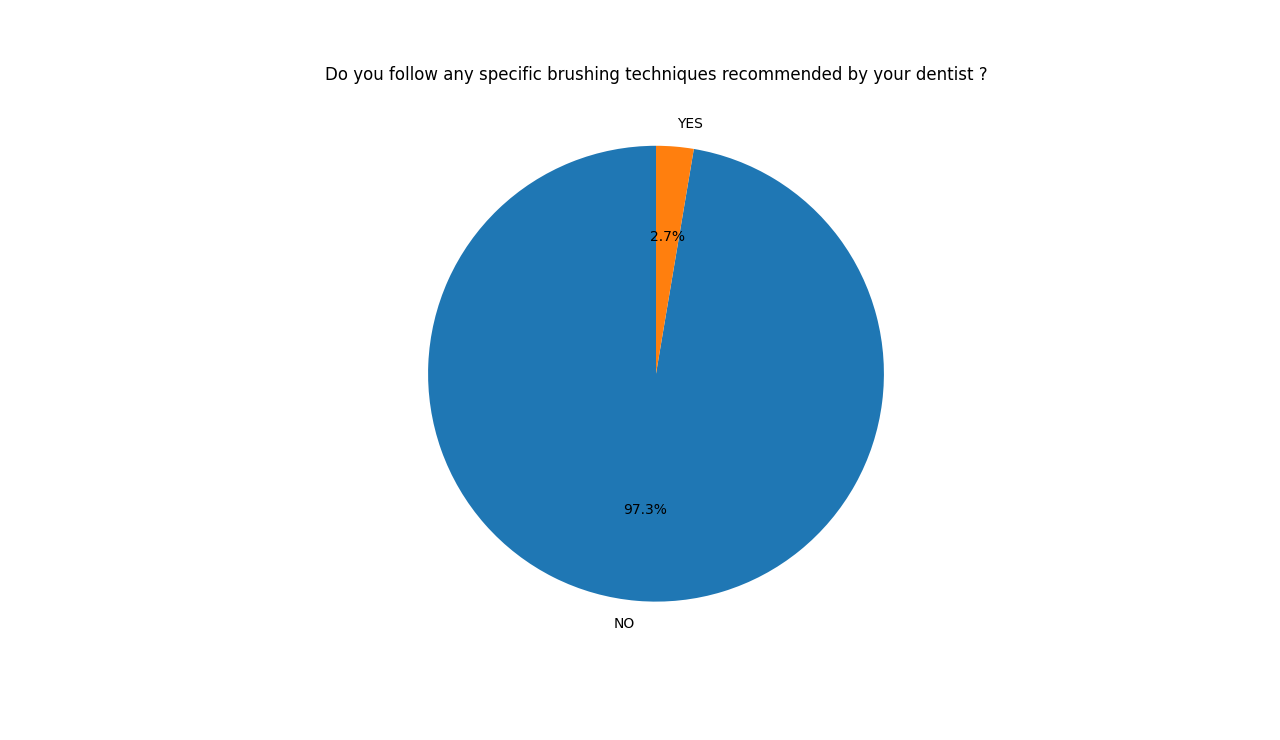
value\_counts = df['Do you follow any specific brushing techniques recommended by your dentist ?'].value\_counts()

plt.figure(figsize=(8, 8))

plt.pie(value\_counts, labels=value\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title(f'Do you follow any specific brushing techniques recommended by your dentist ?')

plt.show()

****

**Data set 12: Do you consume Tea or Coffee regularly ?**

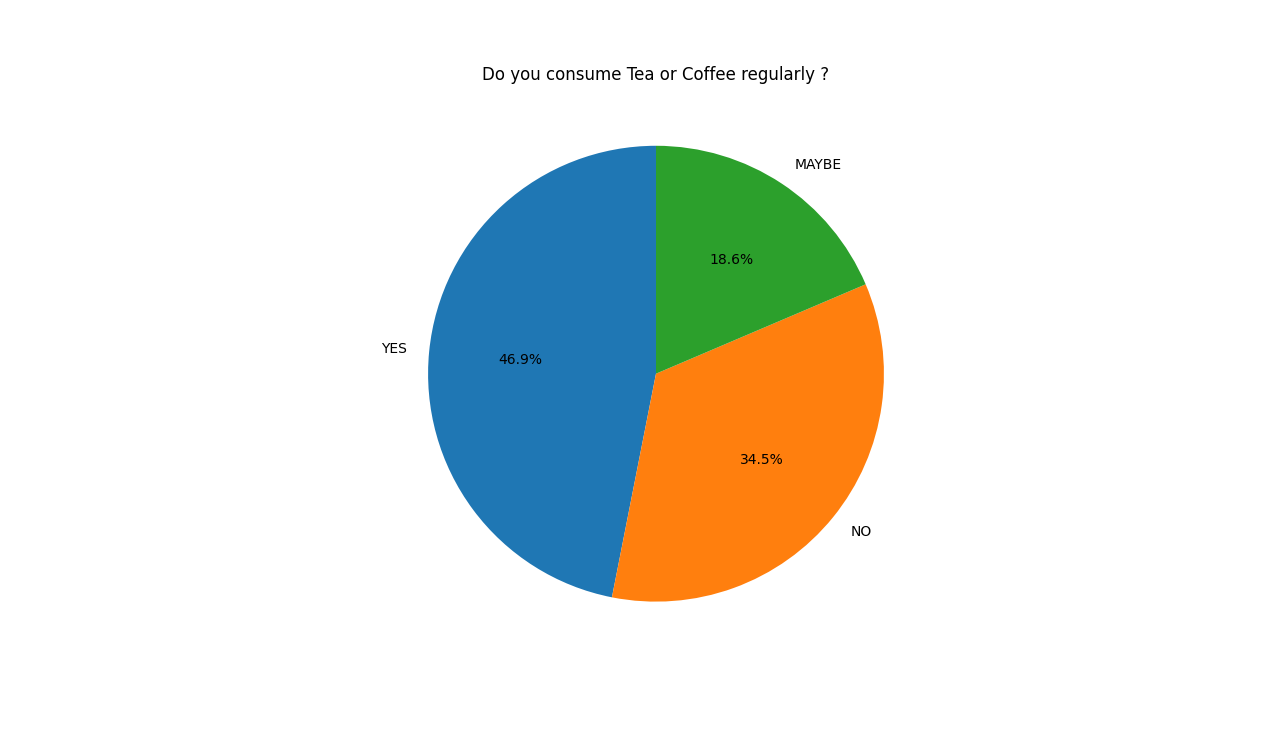
value\_counts = df['Do you consume Tea or Coffee regularly ? '].value\_counts()

plt.figure(figsize=(8, 8))

plt.pie(value\_counts, labels=value\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title(f'Do you consume Tea or Coffee regularly ?')

plt.show()

****

**Data Set 13: Have you experienced any dental issues even after doing regular brushing ?**

value\_counts = df['Have you experienced any dental issues even after doing regular brushing ? '].value\_counts()

plt.figure(figsize=(10, 6))

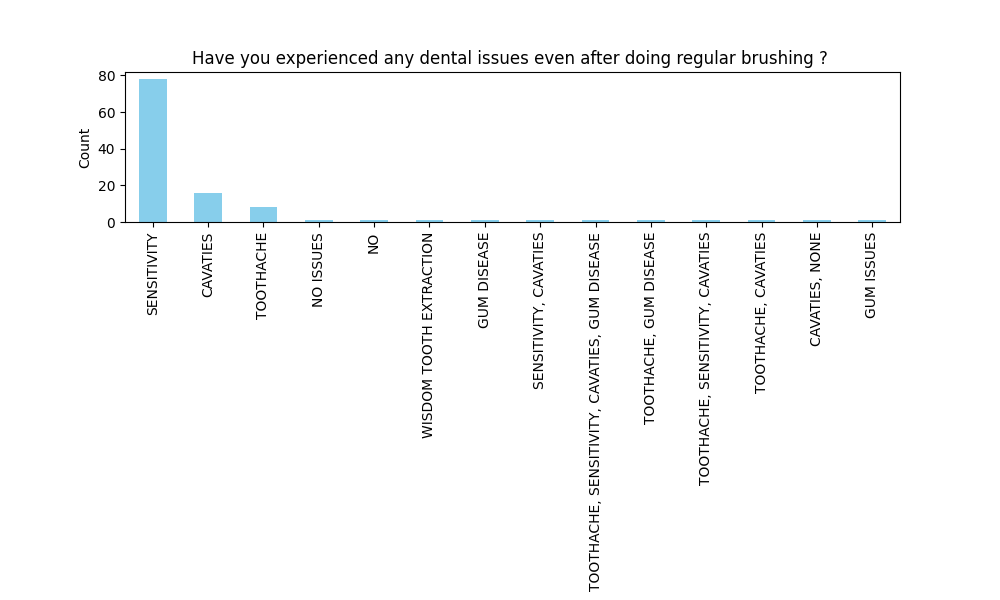
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Have you experienced any dental issues even after doing regular brushing ? ')

plt.xlabel('')

plt.ylabel('Count')

plt.show()

****

**Data Set 14:** **Have you faced any issue after using a certain toothpaste? if yes then name the toothpaste**

value\_counts = df['Have you faced any issue after using a certain toothpaste? if yes then name the toothpaste '].value\_counts()

plt.figure(figsize=(10, 6))

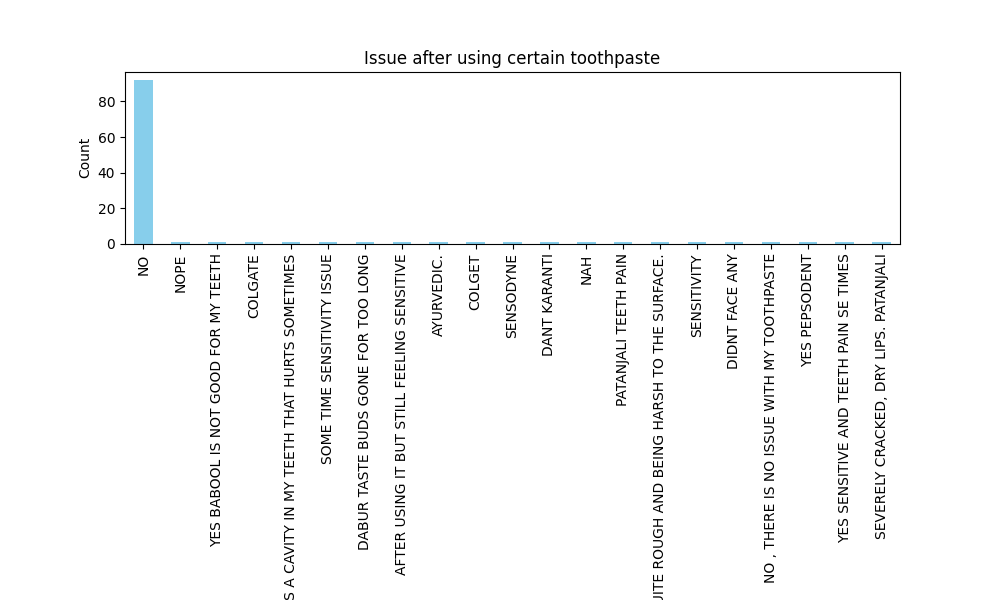
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Issue after using certain toothpaste')

plt.xlabel(' ')

plt.ylabel('Count')

plt.show()

****

**Data Set 15: How much sugar do you consume daily ?**

value\_counts = df['How much sugar do you consume daily ?'].value\_counts()

plt.figure(figsize=(10, 6))

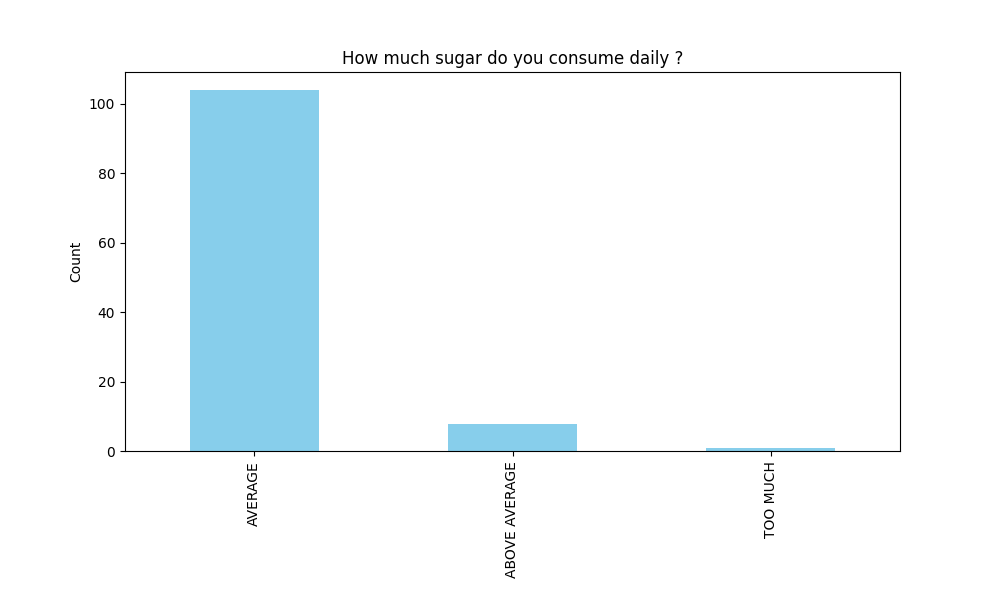
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'How much sugar do you consume daily ?')

plt.xlabel('')

plt.ylabel('Count')

plt.show()

****

**Data Set 16: Have you ever diagnose with any problem related to teeth ?**

value\_counts = df['Have you ever diagnose with any problem related to teeth ?'].value\_counts()

plt.figure(figsize=(10, 6))

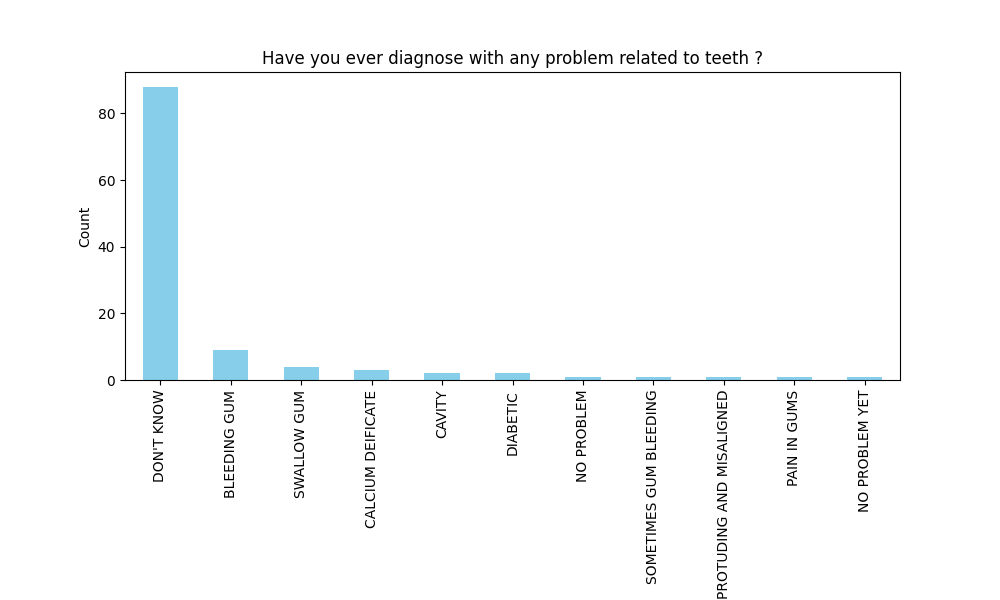
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Have you ever diagnose with any problem related to teeth ?')

plt.xlabel(' ')

plt.ylabel('Count')

plt.show()

****

**Data Set 17: Do you consume any of these Tobacco, alcohol and smoking ?**

value\_counts = df['Do you consume any of these Tobacco, alcohol and smoking ?'].value\_counts()

plt.figure(figsize=(10, 6))

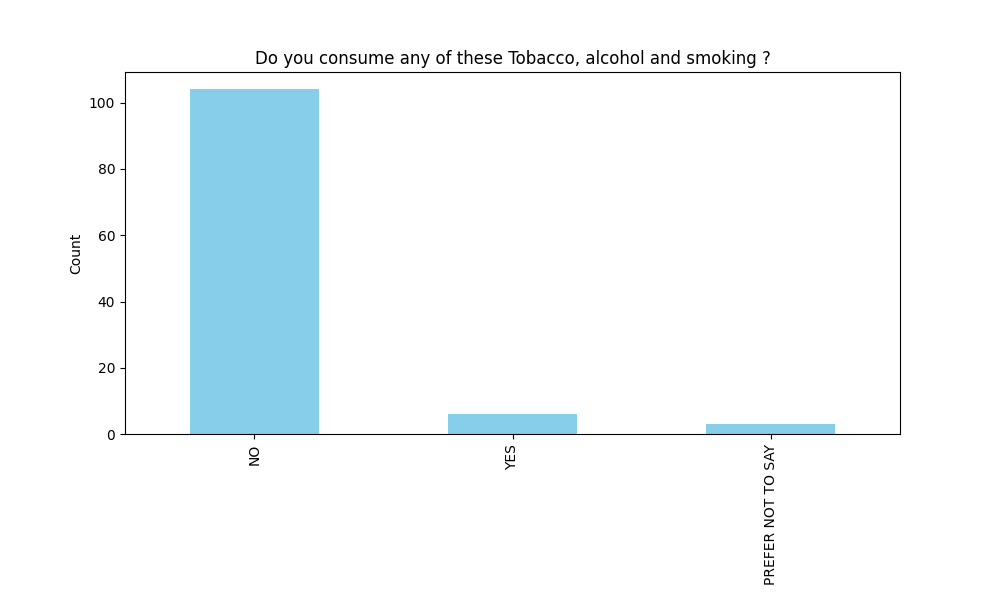
value\_counts.plot(kind='bar', color='skyblue')

plt.title(f'Do you consume any of these Tobacco, alcohol and smoking ?')

plt.xlabel(' ')

plt.ylabel('Count')

plt.show()

****

**Compression between toothpaste use and dental issue after using toothpaste**

This show if certain toothpaste doing any problem to people teeth

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import csv

df = pd.read\_csv('Project Data2.csv')

toothpaste\_brand = df['Which brand of toothpaste do you commonly use ?']

brushing\_frequency = df['How often do you brush your teeth in a typical day ?']

brushing\_duration = df['For how long you brush your teeth ?']

hot\_cold\_beverage = df['Do you drink hot or cold beverages regularly ?']

tea\_coffee\_consumption = df['Do you consume Tea or Coffee regularly ? ']

sugar\_consumption = df['How much sugar do you consume daily ?']

dental\_issues = df['Have you experienced any dental issues even after doing regular brushing ? ']

hot\_cold\_beverage = np.where(hot\_cold\_beverage == 'Yes', 1, 0)

tea\_coffee\_consumption = np.where(tea\_coffee\_consumption == 'Yes', 1, 0)

toothpaste\_dental\_issues = pd.crosstab(toothpaste\_brand, dental\_issues)

brushing\_dental\_issues = pd.crosstab(brushing\_frequency, dental\_issues)

duration\_dental\_issues = pd.crosstab(brushing\_duration, dental\_issues)

plt.figure(figsize=(12, 8))

plt.subplot(2, 2, 1)

toothpaste\_dental\_issues.plot(kind='bar', stacked=True, colormap='viridis')

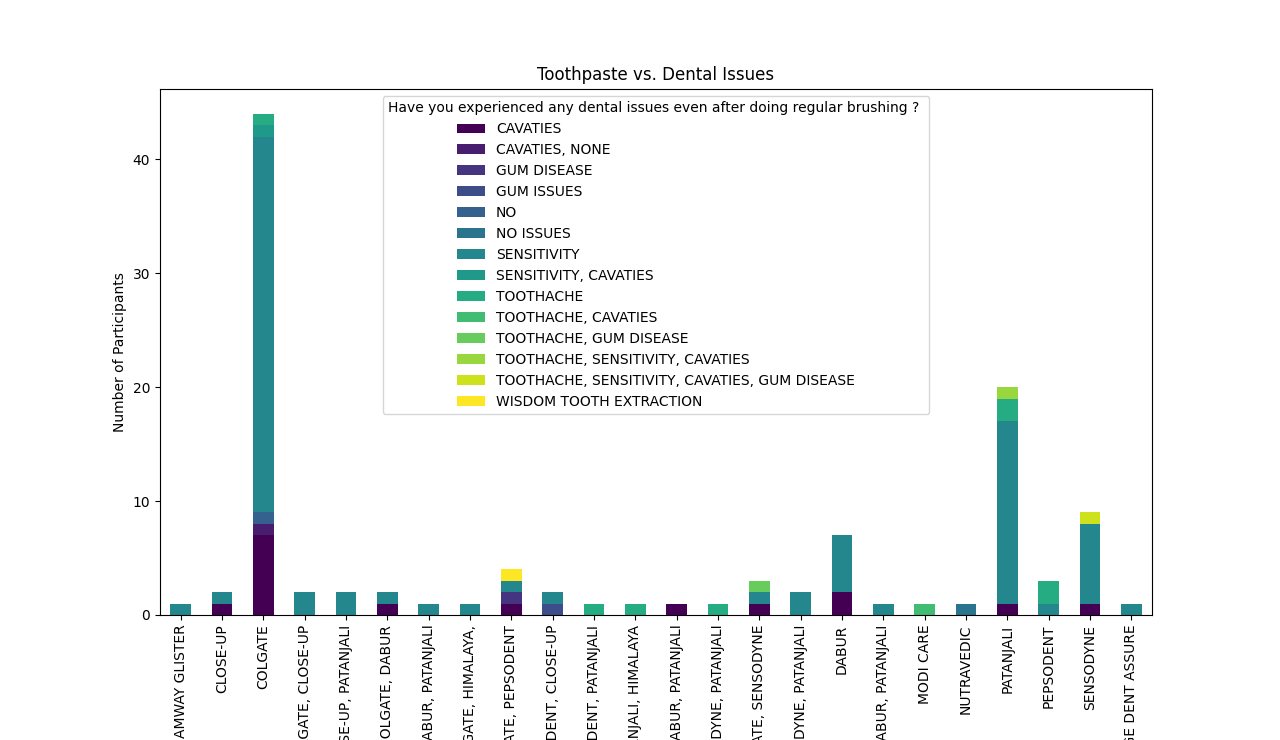
plt.title('Toothpaste vs. Dental Issues')

plt.xlabel('Toothpaste Brand')

plt.ylabel('Number of Participants')

plt.tight\_layout()

plt.show()

****

**Compression between frequency of brush vs dental issue**

This compression show if how many times you doing brush in a day but still getting dental issue

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

df = pd.read\_csv('Project Data2.csv')

toothpaste\_brand = df['Which brand of toothpaste do you commonly use ?']

brushing\_frequency = df['How often do you brush your teeth in a typical day ?']

brushing\_duration = df['For how long you brush your teeth ?']

hot\_cold\_beverage = df['Do you drink hot or cold beverages regularly ?']

tea\_coffee\_consumption = df['Do you consume Tea or Coffee regularly ? ']

sugar\_consumption = df['How much sugar do you consume daily ?']

dental\_issues = df['Have you experienced any dental issues even after doing regular brushing ? ']

hot\_cold\_beverage = np.where(hot\_cold\_beverage == 'Yes', 1, 0)

tea\_coffee\_consumption = np.where(tea\_coffee\_consumption == 'Yes', 1, 0)

toothpaste\_dental\_issues = pd.crosstab(toothpaste\_brand, dental\_issues)

brushing\_dental\_issues = pd.crosstab(brushing\_frequency, dental\_issues)

duration\_dental\_issues = pd.crosstab(brushing\_duration, dental\_issues)

plt.figure(figsize=(12, 8))

plt.subplot(2, 2, 2)

brushing\_dental\_issues.plot(kind='bar', stacked=True, colormap='viridis')

plt.title('Brushing Frequency vs. Dental Issues')

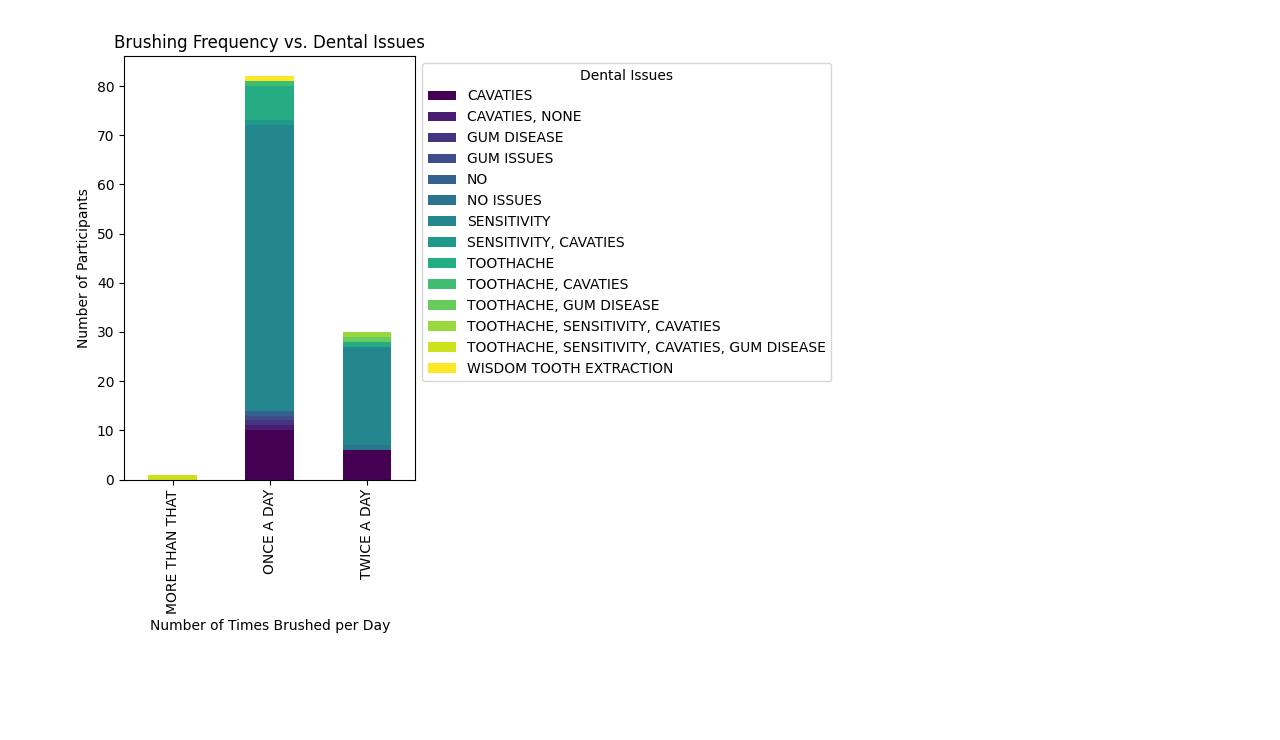
plt.xlabel('Number of Times Brushed per Day')

plt.ylabel('Number of Participants')

plt.legend(title='Dental Issues', bbox\_to\_anchor=(1, 1))

plt.tight\_layout()

plt.show()

****

**Compression between Brushing Duration vs Dental issue**

This compression show people doing regular and sufficient time for brushing but still get dental issue

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import csv

df = pd.read\_csv('Project Data2.csv')

toothpaste\_brand = df['Which brand of toothpaste do you commonly use ?']

brushing\_frequency = df['How often do you brush your teeth in a typical day ?']

brushing\_duration = df['For how long you brush your teeth ?']

hot\_cold\_beverage = df['Do you drink hot or cold beverages regularly ?']

tea\_coffee\_consumption = df['Do you consume Tea or Coffee regularly ? ']

sugar\_consumption = df['How much sugar do you consume daily ?']

dental\_issues = df['Have you experienced any dental issues even after doing regular brushing ? ']

hot\_cold\_beverage = np.where(hot\_cold\_beverage == 'Yes', 1, 0)

tea\_coffee\_consumption = np.where(tea\_coffee\_consumption == 'Yes', 1, 0)

toothpaste\_dental\_issues = pd.crosstab(toothpaste\_brand, dental\_issues)

brushing\_dental\_issues = pd.crosstab(brushing\_frequency, dental\_issues)

duration\_dental\_issues = pd.crosstab(brushing\_duration, dental\_issues)

plt.figure(figsize=(12, 8))

plt.subplot(2, 2, 3)

duration\_dental\_issues.plot(kind='bar', stacked=True, colormap='viridis')

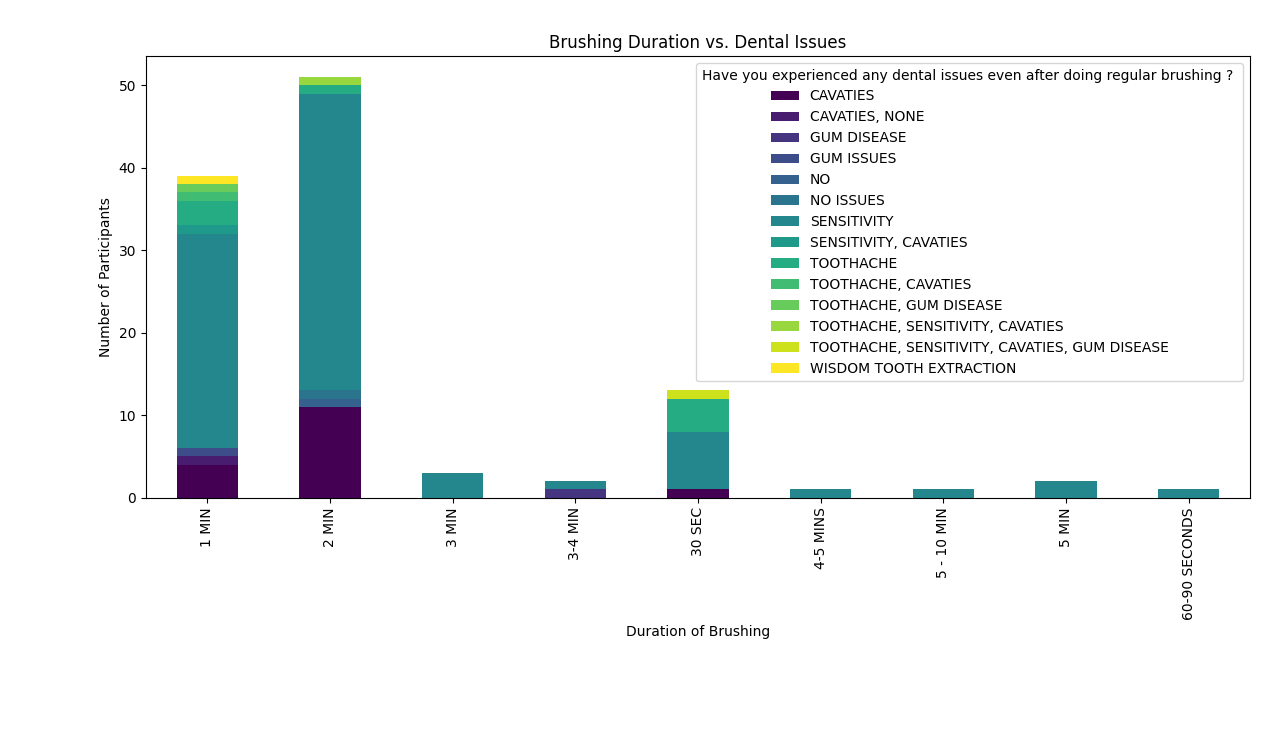
plt.title('Brushing Duration vs. Dental Issues')

plt.xlabel('Duration of Brushing')

plt.ylabel('Number of Participants')

plt.tight\_layout()

plt.show()

****