DHP Survey Analysis

May 16, 2021

1 Introduction

In our study on network based algorithms, specifically knowledge graph based methods, we studied techniques that aimed at removing algorithmic bias.

One such approach which we analyzed in detail was by Bose, et al.: Compositional Fairness Constraints for Graph Embeddings. It discusses an algorithm for removing bias from recommendation systems by learning graph embeddings that are invariant to chosen sensitive attributes (could be age, gender, occupation etc.). This would allow users themselves decide what attributes/factors are to be used/removed while giving them recommendations, in a very flexible way. It makes the recommendation algorithm "personalized" as opposed to what has been a standard algorithm for everyone.

An excerpt from the paper : *For instance, in the context of social recommendations, our framework would allow one user to request that their recommendations are invariant to both their age and gender, while also allowing another user to request invariance to just their age.*

Click here to open this notebook in google colab

2 Social Survey

2.1 Few interesting questions were raised:

- Are users are informed enough to decide which attributes to choose, to get good recommendations ?
- Are people interested in having the option to decide which attributes influence their recommendations ?

- Are there differences of opinion between demographics about the same?

2.1.1 We conducted an online survey to answer these questions. We surveyed around 320 people across various age groups and asked them the following questions:

- What is your Gender?
  - Female
  - Male
  - Non-Binary
  - Prefer not to say
- What is your age?
- Which of the following platforms do you use regularly?

- Netflix
- Amazon Prime
- YouTube
- Spotify
- JioSaavn
- Do you know what information about you do the above platforms use in order to provide recommendations?
  - Yes
  - No
- Would you like to have an option to choose which attributes to use while giving you recommendations?
  The attributes could be your gender, region you live in, previously watched movie genres etc.
  - Yes
  - No (If they answer yes for the previous question, one further question)
- Which of these attributes would you like to NOT influence your recommendations?
  - Gender
  - Age
  - Region you stay in
  - Previous usage history
  - People with similar watch preferences
  - The device you use
  - Time of the day you usually use the platform
  - Duration for which you use the platform

2.2 Data collected

Link to analytics page for responses

Link to raw response data

2.3 Analysis

Link to all plots and diagrams : Generated and discussed below.

3 Analysis Setup

3.1 Imports

```
[2]: import pandas as pd
     import matplotlib.pyplot as plt
     import numpy as np
     from google.colab import drive
     drive.mount('/content/drive')
     from IPython.display import Image
```

Mounted at /content/drive

3.2 Reading in the data

```
[5]:  df = pd.read_csv('./responses.csv')
      # Renaming the Columns with appropriate names
      df.columns = ['Timestamp', 'Gender', 'Age',  'PlatformsUsed', 'IsInformed',␣
      ↪'Option','Attributes']
```

4 Overall Statistics

Gathered from google forms data.

4.0.1 A whopping 87% of people wanted to have the option to choose which attributes are taken into account while getting recommendations!

```
[6]:  Image("/content/drive/MyDrive/SocialSurvey/choose.png", width = 800)
```

[6]:

Would you like to have an option to choose which attributes to use while giving you
recommendations? The attributes could be your gen...ou live in, previously watched movie genres etc.
318 responses
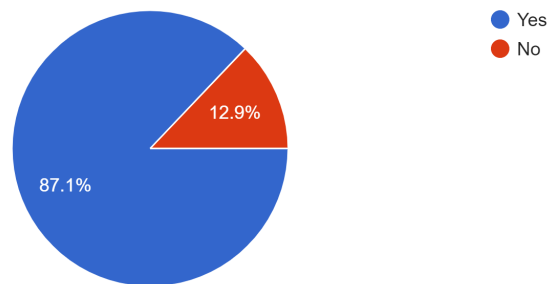
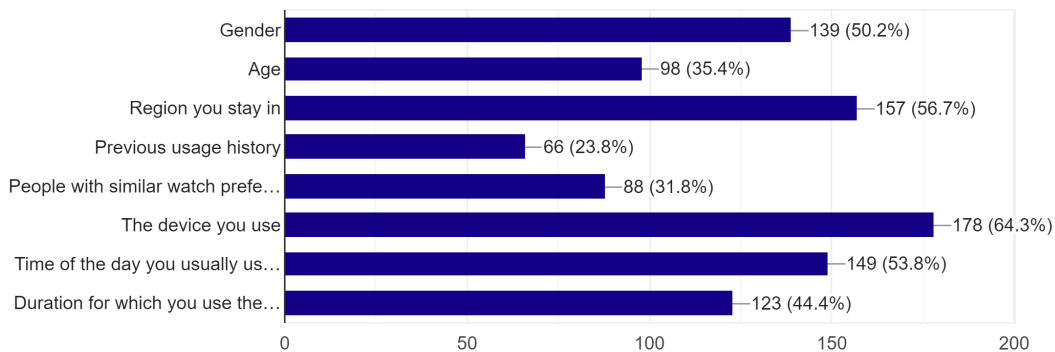### 4.0.2 Attributes that people do not want to influence their recommendations

- 50% of the people didn't want their gender to be considered.
- An interesting factor to be noticed is that around 24% of people did not want previous watch history to be considered, while it is one of the prime factors used by recommendation engines.
- Most people (~65%) were okay with their age being used to give them recommendations.

```
[7]:  Image("/content/drive/MyDrive/SocialSurvey/attributes.png", width = 800)
```

[7]:

Which of these attributes would you like to NOT influence your recommendations?
277 responses

|	
Gender	139 (50.2%)
Age	98 (35.4%)
Region you stay in	157 (56.7%)
Previous usage history	66 (23.8%)
People with similar watch prefe…	88 (31.8%)
The device you use	178 (64.3%)
Time of the day you usually us…	149 (53.8%)
Duration for which you use the…	123 (44.4%)

4.0.3 46% people said they were not aware about which attributes are used by platforms to give them recommendations

```
[8]: Image("/content/drive/MyDrive/SocialSurvey/information.png", width=800)
```

[8]:

Do you know what information about you do the above platforms use in order to provide recommendations?
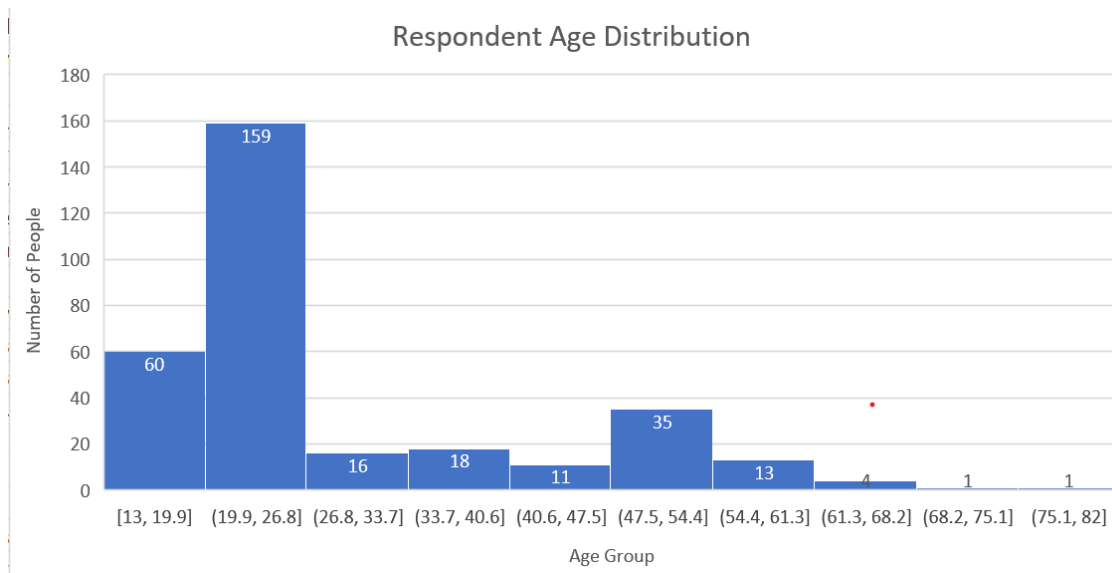318 responses

- Yes
- No

45.9%

54.1%
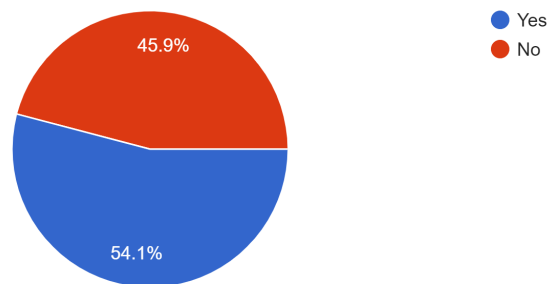
4.0.4 Age distribution of respondents

We tried to get responses from various age groups, to get diverse opinions.

```
[9]: Image("/content/drive/MyDrive/SocialSurvey/age.png", width=800)
```
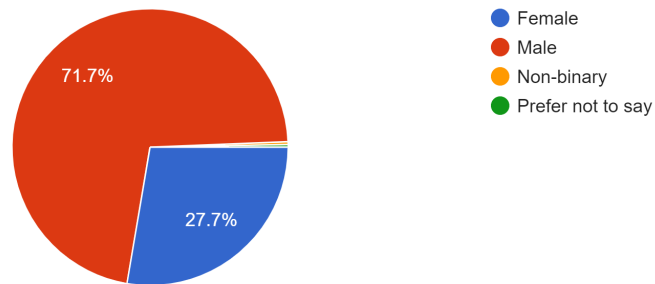
[9]:

Respondent Age Distribution

4.0.5 Gender distribution

```
[10]: Image("/content/drive/MyDrive/SocialSurvey/gender.png", width=800)
```

[10]:

What is your Gender?
318 responses

- Female
- Male
- Non-binary
- Prefer not to say
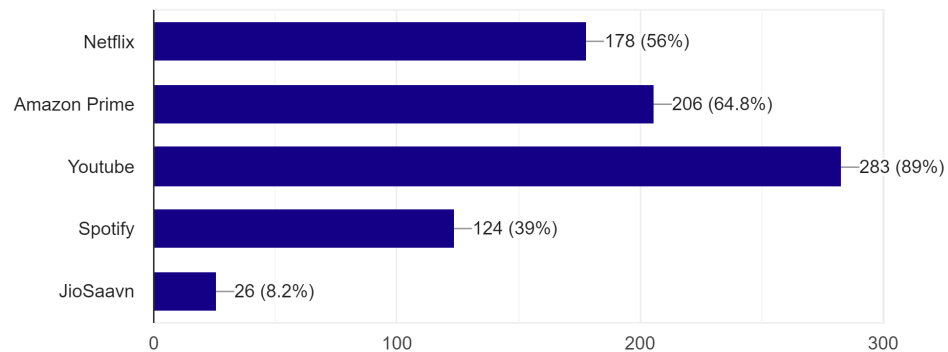
71.7%
27.7%

4.0.6 Platforms used regularly

```
[11]: Image("/content/drive/MyDrive/SocialSurvey/medium.png", width=800)
```

[11]:

Which of the following platforms do you use regularly?
318 responses

5 Interesting Observations

- 57% of people below 40 knew about what influences their recommendations compared to 47% of people above 40. The younger population is more aware.

- 48% of people below 40 didn't want their gender to be considered while only 68% people above 40 were okay with their gender being while getting recommendations.

- 75% of people below 40 wanted to watch what their peers were watching (people with similar watch preferences). People above 40 tended to be more individualistic and more than 40% did not want peers to influence their recommendations.

- People of all age groups were okay with their age being used to give them recommendations, with over 70% being okay with it.

- Women are more sensitive about their gender being used as compared to men. 39% of men did not want their gender to be considered compared to 50% of women.

- Over 70% of both men and women were okay with their age being used.

- Men were a bit more aware of attributes used by recommendation engines, around 60% responded they knew, compared to 49% of women.

- Almost all women, ~93% wanted the option to choose the attributes being used to give recommendation, ~85% men wanted to have the option.

- Both men and women wanted peer influence on their recommendations (>70% for both).

6 Demographic 1.a: Age below 40

```python
[12]: filter_age_below_40 = df['Age'] < 40
      filter_attributes_gender = df['Attributes'].str.contains('Gender',na=False)
      filter_attributes_age = df['Attributes'].str.contains('Age',na=False)
      filter_attributes_region = df['Attributes'].str.contains('Region',na=False)
      filter_attributes_history = df['Attributes'].str.contains('history',na=False)
      filter_attributes_preferences = df['Attributes'].str.
       ↪contains('preferences',na=False)
      filter_attributes_device = df['Attributes'].str.contains('device',na=False)
      filter_attributes_time = df['Attributes'].str.contains('Time',na=False)
      filter_attributes_duration = df['Attributes'].str.contains('Duration',na=False)
      filter_informed = df['IsInformed'] == "Yes"
      filter_choose = df['Option'] == "Yes"
      bar_graph_ageless40=[]

      age_below_40 = df.where(filter_age_below_40).dropna(how="all").shape[0]
      print("Respondents above age of 40:", age_below_40)

      filtered_df = df.where(filter_age_below_40 & filter_attributes_gender).
       ↪dropna(how="all")
      print("Percentage of respondents above the age of 40 who didn't want their␣
       ↪gender to be considered while getting recommendations", filtered_df.
       ↪shape[0]*100/age_below_40, "%")
      bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)

      filtered_df = df.where(filter_age_below_40 & filter_attributes_age).
       ↪dropna(how="all")
      print("Percentage of respondents above the age of 40 who didn't want their age␣
       ↪to be considered while getting recommendations", filtered_df.shape[0]*100/
       ↪age_below_40, "%")
      bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)

      filtered_df = df.where(filter_age_below_40 & filter_attributes_region).
       ↪dropna(how="all")
      print("Percentage of respondents above the age of 40 who didn't want their␣
       ↪region of stay to be considered while getting recommendations", filtered_df.
       ↪shape[0]*100/age_below_40, "%")
      bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)

      filtered_df = df.where(filter_age_below_40 & filter_attributes_history).
       ↪dropna(how="all")
      print("Percentage of respondents above the age of 40 who didn't want their␣
       ↪watch history to be considered while getting recommendations", filtered_df.
       ↪shape[0]*100/age_below_40, "%")
      bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)
```

```
filtered_df = df.where(filter_age_below_40 & filter_attributes_preferences).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want the people␣
 ↪with similar watch preferences to be considered while getting␣
 ↪recommendations", filtered_df.shape[0]*100/age_below_40, "%")
bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)

filtered_df = df.where(filter_age_below_40 & filter_attributes_device).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their␣
 ↪device of use to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_below_40, "%")
bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)

filtered_df = df.where(filter_age_below_40 & filter_attributes_time).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their time␣
 ↪of watch to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_below_40, "%")
bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)

filtered_df = df.where(filter_age_below_40 & filter_attributes_duration).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their␣
 ↪watch duration to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_below_40, "%")
bar_graph_ageless40.append(filtered_df.shape[0]*100/age_below_40)

labels=['Gender','Age','Region of Stay', 'Watch History', 'Peer Preferences',␣
 ↪'Device of Use','Time of Day', 'Duration']
plt.figure(figsize=(15,5))
plt.bar(labels, bar_graph_ageless40)
for index, value in enumerate(bar_graph_ageless40):
    plt.text(index,value+0.5, s=round(value, 2), weight='bold')
plt.title('Attributes chosen by people less than 40 years to not influence␣
 ↪their recommendations')
plt.xlabel('Attributes')
plt.ylabel('Percentage of people')
plt.savefig("peopleless40_attributes.png")
plt.show()
```
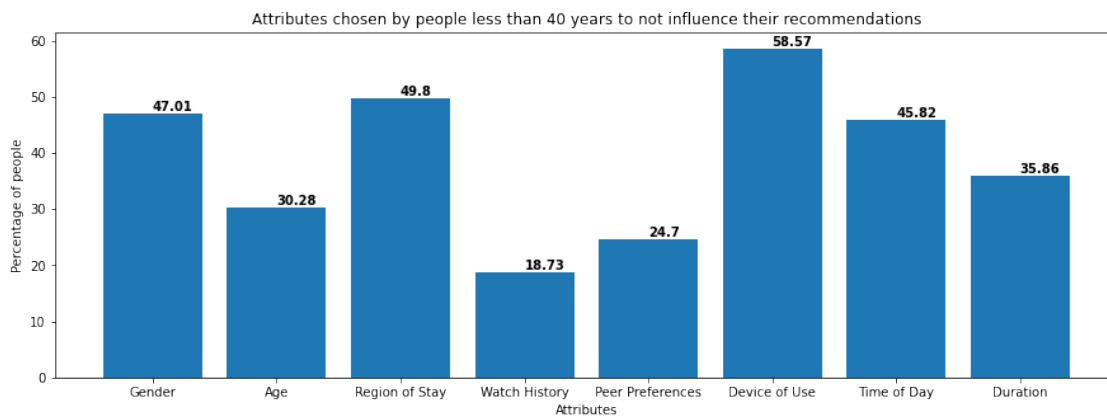
Respondents above age of 40: 251
Percentage of respondents above the age of 40 who didn't want their gender to be
considered while getting recommendations 47.01195219123506 %
Percentage of respondents above the age of 40 who didn't want their age to be
considered while getting recommendations 30.278884462151396 %

Percentage of respondents above the age of 40 who didn't want their region of stay to be considered while getting recommendations 49.800796812749006 %
Percentage of respondents above the age of 40 who didn't want their watch history to be considered while getting recommendations 18.725099601593627 %
Percentage of respondents above the age of 40 who didn't want the people with similar watch preferences to be considered while getting recommendations 24.701195219123505 %
Percentage of respondents above the age of 40 who didn't want their device of use to be considered while getting recommendations 58.56573705179283 %
Percentage of respondents above the age of 40 who didn't want their time of watch to be considered while getting recommendations 45.81673306772908 %
Percentage of respondents above the age of 40 who didn't want their watch duration to be considered while getting recommendations 35.85657370517928 %
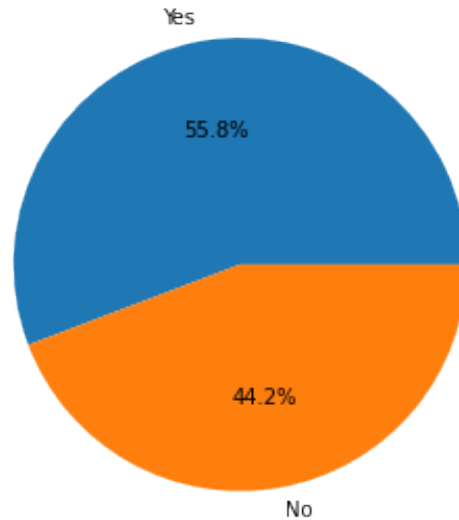


```
[13]: filtered_df = df.where(filter_age_below_40 & filter_informed).dropna(how="all")
      print("Percentage of respondents above the age of 40 who knew what information␣
       ↪is used to give recommendations", filtered_df.shape[0]*100/age_below_40, "%")

      y = np.array([filtered_df.shape[0]*100/age_below_40, 100 - (filtered_df.
       ↪shape[0]*100/age_below_40)])
      mylabels = ["Yes", "No"]

      plt.figure(figsize=(5,5))
      plt.title('% of people less than 40 who knew what information is used to give␣
       ↪recommendations')
      plt.pie(y, labels = mylabels,autopct='%1.1f%%')
      plt.savefig("peopleless40_information.png")
      plt.show()
```

Percentage of respondents above the age of 40 who knew what information is used to give recommendations 55.776892430278885 %

% of people less than 40 who knew what information is used to give recommendations



```
[14]: filtered_df = df.where(filter_age_below_40 & filter_choose).dropna(how="all")
      print("Percentage of respondents above the age of 40 who wanted an option to␣
       ↪choose which attributes to use while giving recommendations", filtered_df.
       ↪shape[0]*100/age_below_40, "%")

      y = np.array([filtered_df.shape[0]*100/age_below_40, 100 - (filtered_df.
       ↪shape[0]*100/age_below_40)])
      mylabels = ["Yes", "No"]

      plt.figure(figsize=(5,5))
      plt.title('% of people less than 40 wanted an option to choose which attributes␣
       ↪to use while giving recommendations')
      plt.pie(y, labels = mylabels,autopct='%1.1f%%')
      plt.savefig("peopleless40_choose.png")
      plt.show()
```

Percentage of respondents above the age of 40 who wanted an option to choose
which attributes to use while giving recommendations 87.64940239043824 %

% of people less than 40 wanted an option to choose which attributes to use while giving recommendations
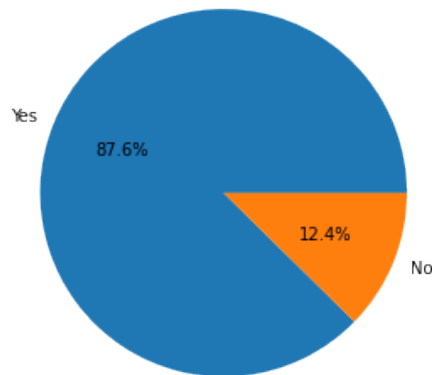
7 Demographic 1.b: Age above 40

```
[15]: filter_age_above_40 = df['Age'] > 40
      filter_attributes_gender = df['Attributes'].str.contains('Gender',na=False)
      filter_attributes_age = df['Attributes'].str.contains('Age',na=False)
      filter_attributes_region = df['Attributes'].str.contains('Region',na=False)
      filter_attributes_history = df['Attributes'].str.contains('history',na=False)
      filter_attributes_preferences = df['Attributes'].str.
       ↪contains('preferences',na=False)
      filter_attributes_device = df['Attributes'].str.contains('device',na=False)
      filter_attributes_time = df['Attributes'].str.contains('Time',na=False)
      filter_attributes_duration = df['Attributes'].str.contains('Duration',na=False)
      filter_informed = df['IsInformed'] == "Yes"
      filter_choose = df['Option'] == "Yes"
      bar_graph_ageabove40=[]

      age_above_40 = df.where(filter_age_above_40).dropna(how="all").shape[0]
      print("Respondents above age of 40:", age_above_40)

      filtered_df = df.where(filter_age_above_40 & filter_attributes_gender).
       ↪dropna(how="all")
      print("Percentage of respondents above the age of 40 who didn't want their␣
       ↪gender to be considered while getting recommendations", filtered_df.
       ↪shape[0]*100/age_above_40, "%")
      bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)

      filtered_df = df.where(filter_age_above_40 & filter_attributes_age).
       ↪dropna(how="all")
```

```python
print("Percentage of respondents above the age of 40 who didn't want their age␣
 ↪to be considered while getting recommendations", filtered_df.shape[0]*100/
 ↪age_above_40, "%")
bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)

filtered_df = df.where(filter_age_above_40 & filter_attributes_region).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their␣
 ↪region of stay to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_above_40, "%")
bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)

filtered_df = df.where(filter_age_above_40 & filter_attributes_history).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their␣
 ↪watch history to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_above_40, "%")
bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)

filtered_df = df.where(filter_age_above_40 & filter_attributes_preferences).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want the people␣
 ↪with similar watch preferences to be considered while getting␣
 ↪recommendations", filtered_df.shape[0]*100/age_above_40, "%")
bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)

filtered_df = df.where(filter_age_above_40 & filter_attributes_device).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their␣
 ↪device of use to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_above_40, "%")
bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)

filtered_df = df.where(filter_age_above_40 & filter_attributes_time).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their time␣
 ↪of watch to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_above_40, "%")
bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)

filtered_df = df.where(filter_age_above_40 & filter_attributes_duration).
 ↪dropna(how="all")
print("Percentage of respondents above the age of 40 who didn't want their␣
 ↪watch duration to be considered while getting recommendations", filtered_df.
 ↪shape[0]*100/age_above_40, "%")
bar_graph_ageabove40.append(filtered_df.shape[0]*100/age_above_40)
```

```
labels=['Gender','Age','Region of Stay', 'Watch History', 'Peer Preferences',␣
 ↪'Device of Use','Time of Day', 'Duration']
plt.figure(figsize=(15,5))
plt.bar(labels, bar_graph_ageabove40)
for index, value in enumerate(bar_graph_ageabove40):
    plt.text(index,value+0.5, s=round(value, 2), weight='bold')
plt.title('Attributes chosen by people above than 40 years to not influence␣
 ↪their recommendations')
plt.xlabel('Attributes')
plt.ylabel('Percentage of people')
plt.savefig("peopleabove40_attributes.png")
plt.show()
```

Respondents above age of 40: 65
Percentage of respondents above the age of 40 who didn't want their gender to be
considered while getting recommendations 32.30769230769231 %
Percentage of respondents above the age of 40 who didn't want their age to be
considered while getting recommendations 33.84615384615385 %
Percentage of respondents above the age of 40 who didn't want their region of
stay to be considered while getting recommendations 46.15384615384615 %
Percentage of respondents above the age of 40 who didn't want their watch
history to be considered while getting recommendations 27.692307692307693 %
Percentage of respondents above the age of 40 who didn't want the people with
similar watch preferences to be considered while getting recommendations 40.0 %
Percentage of respondents above the age of 40 who didn't want their device of
use to be considered while getting recommendations 46.15384615384615 %
Percentage of respondents above the age of 40 who didn't want their time of
watch to be considered while getting recommendations 49.23076923076923 %
Percentage of respondents above the age of 40 who didn't want their watch
duration to be considered while getting recommendations 49.23076923076923 %
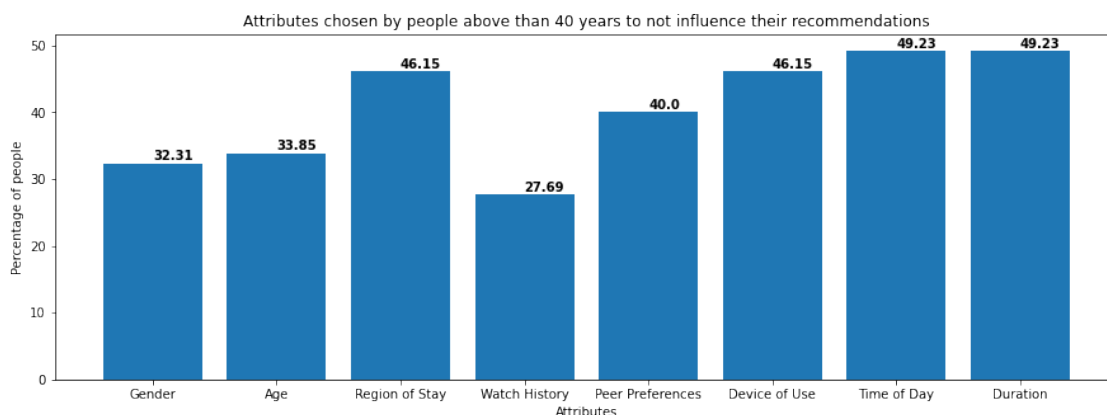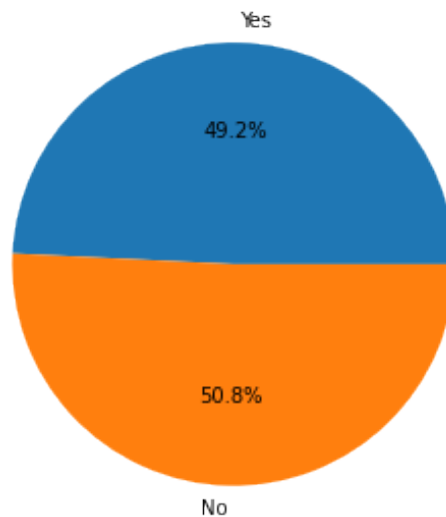
```
[16]:  filtered_df = df.where(filter_age_above_40 & filter_informed).dropna(how="all")
       print("Percentage of respondents above the age of 40 who knew what information␣
        ↪is used to give recommendations", filtered_df.shape[0]*100/age_above_40, "%")

       y = np.array([filtered_df.shape[0]*100/age_above_40, 100 - (filtered_df.
        ↪shape[0]*100/age_above_40)])
       mylabels = ["Yes", "No"]

       plt.figure(figsize=(5,5))
       plt.title('% of people above than 40 who knew what information is used to give␣
        ↪recommendations')
       plt.pie(y, labels = mylabels,autopct='%1.1f%%')
       plt.savefig("peopleabove40_information.png")
       plt.show()
```

Percentage of respondents above the age of 40 who knew what information is used
to give recommendations 49.23076923076923 %

% of people above than 40 who knew what information is used to give recommendations

```
[17]:  filtered_df = df.where(filter_age_above_40 & filter_choose).dropna(how="all")
       print("Percentage of respondents above the age of 40 who wanted an option to␣
        ↪choose which attributes to use while giving recommendations", filtered_df.
        ↪shape[0]*100/age_above_40, "%")

       y = np.array([filtered_df.shape[0]*100/age_above_40, 100 - (filtered_df.
        ↪shape[0]*100/age_above_40)])
       mylabels = ["Yes", "No"]
```

```
plt.figure(figsize=(5,5))
plt.title('% of people above than 40 wanted an option to choose which␣
 ↪attributes to use while giving recommendations')
plt.pie(y, labels = mylabels,autopct='%1.1f%%')
plt.savefig("peopleabove40_choose.png")
plt.show()
```

Percentage of respondents above the age of 40 who wanted an option to choose
which attributes to use while giving recommendations 84.61538461538461 %

% of people above than 40 wanted an option to choose which attributes to use while giving recommendations
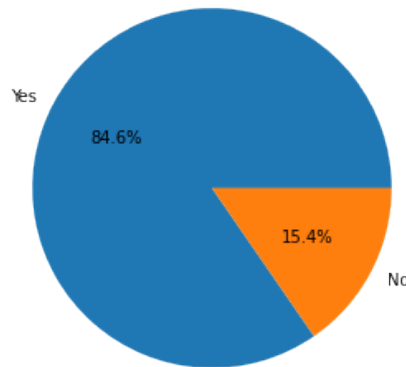
8 Demographic 2.a: Male respondents

8.0.1 Attributes to not be considered

```
[18]: filter_male = df['Gender'] == "Male"
      filter_attributes_gender = df['Attributes'].str.contains('Gender',na=False)
      filter_attributes_age = df['Attributes'].str.contains('Age',na=False)
      filter_attributes_region = df['Attributes'].str.contains('Region',na=False)
      filter_attributes_history = df['Attributes'].str.contains('history',na=False)
      filter_attributes_preferences = df['Attributes'].str.
       ↪contains('preferences',na=False)
      filter_attributes_device = df['Attributes'].str.contains('device',na=False)
      filter_attributes_time = df['Attributes'].str.contains('Time',na=False)
      filter_attributes_duration = df['Attributes'].str.contains('Duration',na=False)
      filter_informed = df['IsInformed'] == "Yes"
      filter_choose = df['Option'] == "Yes"
      bar_graph_men=[]

      men = df.where(filter_male).dropna(how="all").shape[0]
      print("Male respondents:", men)
```

```python
filtered_df = df.where(filter_male & filter_attributes_gender).dropna(how="all")
print("Percentage of male respondents who didn't want their gender to be␣
 ↪considered while getting recommendations", filtered_df.shape[0]*100/men, "%")
bar_graph_men.append(filtered_df.shape[0]*100/men)

filtered_df = df.where(filter_male & filter_attributes_age).dropna(how="all")
print("Percentage of male respondents who didn't want their age to be␣
 ↪considered while getting recommendations", filtered_df.shape[0]*100/men, "%")
bar_graph_men.append(filtered_df.shape[0]*100/men)

filtered_df = df.where(filter_male & filter_attributes_region).dropna(how="all")
print("Percentage of male respondents who didn't want their region of stay to␣
 ↪be considered while getting recommendations", filtered_df.shape[0]*100/men,␣
 ↪"%")
bar_graph_men.append(filtered_df.shape[0]*100/men)

filtered_df = df.where(filter_male & filter_attributes_history).
 ↪dropna(how="all")
print("Percentage of male respondents who didn't want their watch history to be␣
 ↪considered while getting recommendations", filtered_df.shape[0]*100/men, "%")
bar_graph_men.append(filtered_df.shape[0]*100/men)

filtered_df = df.where(filter_male & filter_attributes_preferences).
 ↪dropna(how="all")
print("Percentage of male respondents who didn't want the people with similar␣
 ↪watch preferences to be considered while getting recommendations",␣
 ↪filtered_df.shape[0]*100/men, "%")
bar_graph_men.append(filtered_df.shape[0]*100/men)

filtered_df = df.where(filter_male & filter_attributes_device).dropna(how="all")
print("Percentage of male respondents who didn't want their device of use to be␣
 ↪considered while getting recommendations", filtered_df.shape[0]*100/men, "%")
bar_graph_men.append(filtered_df.shape[0]*100/men)

filtered_df = df.where(filter_male & filter_attributes_time).dropna(how="all")
print("Percentage of male respondents who didn't want their time of watch to be␣
 ↪considered while getting recommendations", filtered_df.shape[0]*100/men, "%")
bar_graph_men.append(filtered_df.shape[0]*100/men)

filtered_df = df.where(filter_male & filter_attributes_duration).
 ↪dropna(how="all")
print("Percentage of male respondents who didn't want their watch duration to␣
 ↪be considered while getting recommendations", filtered_df.shape[0]*100/men,␣
 ↪"%")
bar_graph_men.append(filtered_df.shape[0]*100/men)
```

```
labels=['Gender','Age','Region of Stay', 'Watch History', 'Peer Preferences',␣
 ↪'Device of Use','Time of Day', 'Duration']
plt.figure(figsize=(15,5))
plt.bar(labels, bar_graph_men)
for index, value in enumerate(bar_graph_men):
    plt.text(index,value+0.5, s=round(value, 2), weight='bold')
plt.title('Attributes chosen by men to not influence their recommendations')
plt.xlabel('Attributes')
plt.ylabel('Percentage of men')
plt.savefig("men_attributes.png")
plt.show()
```

Male respondents: 228
Percentage of male respondents who didn't want their gender to be considered
while getting recommendations 41.228070175438596 %
Percentage of male respondents who didn't want their age to be considered while
getting recommendations 31.140350877192983 %
Percentage of male respondents who didn't want their region of stay to be
considered while getting recommendations 46.05263157894737 %
Percentage of male respondents who didn't want their watch history to be
considered while getting recommendations 21.49122807017544 %
Percentage of male respondents who didn't want the people with similar watch
preferences to be considered while getting recommendations 28.07017543859649 %
Percentage of male respondents who didn't want their device of use to be
considered while getting recommendations 57.01754385964912 %
Percentage of male respondents who didn't want their time of watch to be
considered while getting recommendations 47.36842105263158 %
Percentage of male respondents who didn't want their watch duration to be
considered while getting recommendations 38.1578947368421 %
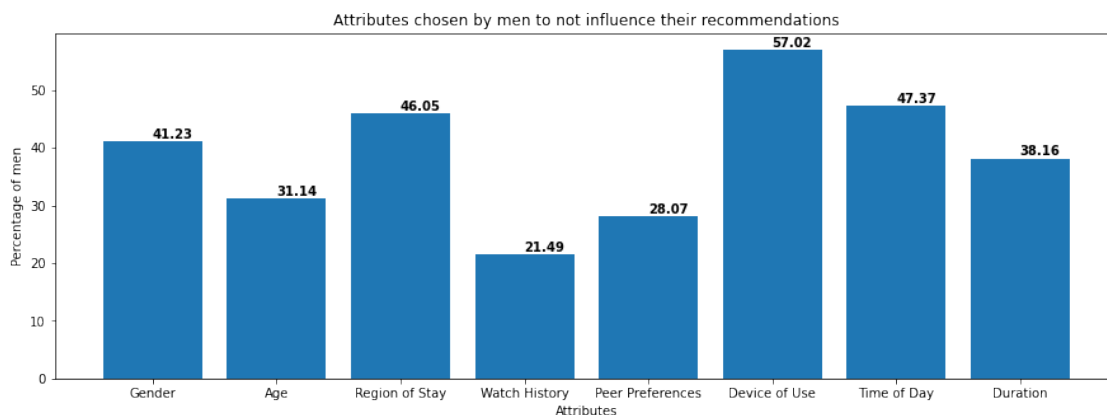
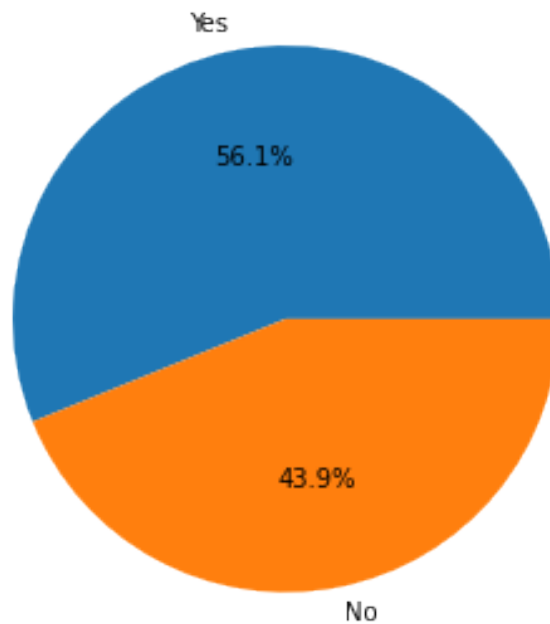### 8.0.2 Aware of attributes used to give recommendations?

```
[19]: filtered_df = df.where(filter_male & filter_informed).dropna(how="all")
      print("Percentage of male respondents who knew what information is used to give␣
       ↪recommendations", filtered_df.shape[0]*100/men, "%")

      y = np.array([filtered_df.shape[0]*100/men, 100 - (filtered_df.shape[0]*100/
       ↪men)])
      mylabels = ["Yes", "No"]

      plt.figure(figsize=(5,5))
      plt.title('% of men who knew what information is used to give recommendations')
      plt.pie(y, labels = mylabels,autopct='%1.1f%%')
      plt.savefig("men_information.png")
      plt.show()
```

```
Percentage of male respondents who knew what information is used to give
recommendations 56.14035087719298 %
```


% of men who knew what information is used to give recommendations

8.0.3 Want an option to choose which attributes are used to give recommendations?

```
[20]: filtered_df = df.where(filter_male & filter_choose).dropna(how="all")
      print("Percentage of male respondents who wanted an option to choose which
       ↪attributes to use while giving recommendations", filtered_df.shape[0]*100/
       ↪men, "%")

      y = np.array([filtered_df.shape[0]*100/men, 100 - (filtered_df.shape[0]*100/
       ↪men)])
      mylabels = ["Yes", "No"]

      plt.figure(figsize=(5,5))
      plt.title('% of men who wanted an option to choose which attributes to use
       ↪while giving recommendations')
      plt.pie(y, labels = mylabels,autopct='%1.1f%%')
      plt.savefig("men_choose.png")
      plt.show()
```

Percentage of male respondents who wanted an option to choose which attributes
to use while giving recommendations 84.64912280701755 %

% of men who wanted an option to choose which attributes to use while giving recommendations
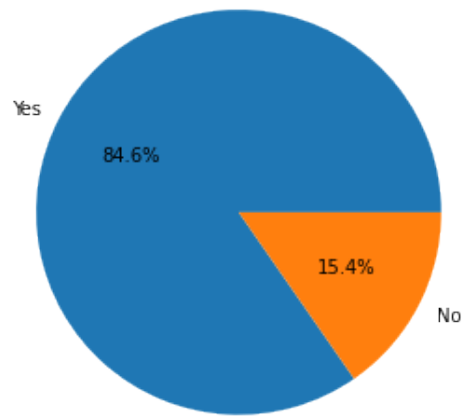
9 Demographic 2.b: Females respondents

9.0.1 Attributes to not be considered

```
[21]: filter_female = df['Gender'] == "Female"
      filter_attributes_gender = df['Attributes'].str.contains('Gender',na=False)
      filter_attributes_age = df['Attributes'].str.contains('Age',na=False)
```

```python
filter_attributes_region = df['Attributes'].str.contains('Region',na=False)
filter_attributes_history = df['Attributes'].str.contains('history',na=False)
filter_attributes_preferences = df['Attributes'].str.
 ↪contains('preferences',na=False)
filter_attributes_device = df['Attributes'].str.contains('device',na=False)
filter_attributes_time = df['Attributes'].str.contains('Time',na=False)
filter_attributes_duration = df['Attributes'].str.contains('Duration',na=False)
filter_informed = df['IsInformed'] == "Yes"
filter_choose = df['Option'] == "Yes"
bar_graph_women=[]

women = df.where(filter_female).dropna(how="all").shape[0]
print("Female respondents:", women)

filtered_df = df.where(filter_female & filter_attributes_gender).
 ↪dropna(how="all")
print("Percentage of female respondents who didn't want their gender to be␣
 ↪considered while getting recommendations", filtered_df.shape[0]*100/women,␣
 ↪"%")
bar_graph_women.append(filtered_df.shape[0]*100/women)

filtered_df = df.where(filter_female & filter_attributes_age).dropna(how="all")
print("Percentage of female respondents who didn't want their age to be␣
 ↪considered while getting recommendations", filtered_df.shape[0]*100/women,␣
 ↪"%")
bar_graph_women.append(filtered_df.shape[0]*100/women)

filtered_df = df.where(filter_female & filter_attributes_region).
 ↪dropna(how="all")
print("Percentage of female respondents who didn't want their region of stay to␣
 ↪be considered while getting recommendations", filtered_df.shape[0]*100/
 ↪women, "%")
bar_graph_women.append(filtered_df.shape[0]*100/women)

filtered_df = df.where(filter_female & filter_attributes_history).
 ↪dropna(how="all")
print("Percentage of female respondents who didn't want their watch history to␣
 ↪be considered while getting recommendations", filtered_df.shape[0]*100/
 ↪women, "%")
bar_graph_women.append(filtered_df.shape[0]*100/women)

filtered_df = df.where(filter_female & filter_attributes_preferences).
 ↪dropna(how="all")
print("Percentage of female respondents who didn't want the people with similar␣
 ↪watch preferences to be considered while getting recommendations",␣
 ↪filtered_df.shape[0]*100/women, "%")
```

```
bar_graph_women.append(filtered_df.shape[0]*100/women)

filtered_df = df.where(filter_female & filter_attributes_device).
 ↪dropna(how="all")
print("Percentage of female respondents who didn't want their device of use to␣
 ↪be considered while getting recommendations", filtered_df.shape[0]*100/
 ↪women, "%")
bar_graph_women.append(filtered_df.shape[0]*100/women)

filtered_df = df.where(filter_female & filter_attributes_time).dropna(how="all")
print("Percentage of female respondents who didn't want their time of watch to␣
 ↪be considered while getting recommendations", filtered_df.shape[0]*100/
 ↪women, "%")
bar_graph_women.append(filtered_df.shape[0]*100/women)

filtered_df = df.where(filter_female & filter_attributes_duration).
 ↪dropna(how="all")
print("Percentage of female respondents who didn't want their watch duration to␣
 ↪be considered while getting recommendations", filtered_df.shape[0]*100/
 ↪women, "%")
bar_graph_women.append(filtered_df.shape[0]*100/women)

labels=['Gender','Age','Region of Stay', 'Watch History', 'Peer Preferences',␣
 ↪'Device of Use','Time of Day', 'Duration']
plt.figure(figsize=(15,5))
plt.bar(labels, bar_graph_women)
for index, value in enumerate(bar_graph_women):
    plt.text(index,value, s=round(value, 2), weight='bold')
plt.title('Attributes chosen by women to not influence their recommendations')
plt.xlabel('Attributes')
plt.ylabel('Percentage of women')
plt.savefig("women_attributes.png")
plt.show()
```

```
Female respondents: 88
Percentage of female respondents who didn't want their gender to be considered
while getting recommendations 48.86363636363637 %
Percentage of female respondents who didn't want their age to be considered
while getting recommendations 28.40909090909091 %
Percentage of female respondents who didn't want their region of stay to be
considered while getting recommendations 56.81818181818182 %
Percentage of female respondents who didn't want their watch history to be
considered while getting recommendations 17.045454545454547 %
Percentage of female respondents who didn't want the people with similar watch
preferences to be considered while getting recommendations 25.0 %
Percentage of female respondents who didn't want their device of use to be
considered while getting recommendations 53.40909090909091 %
```
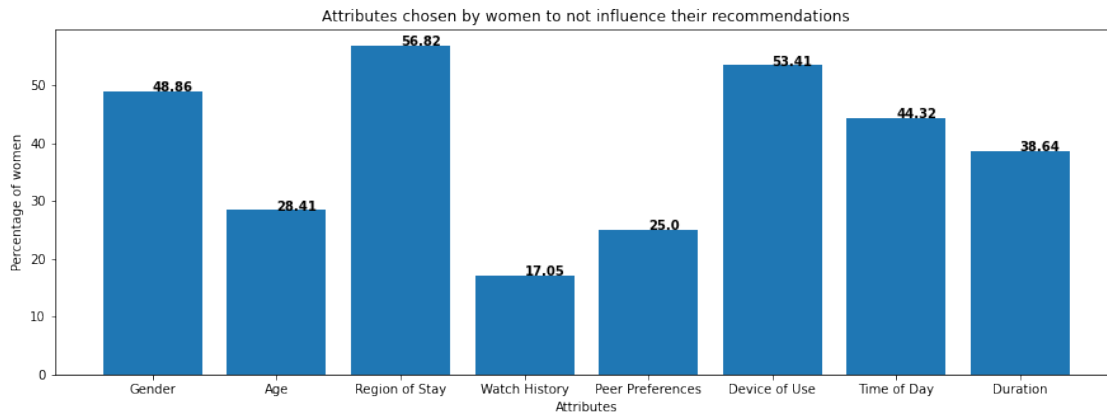
Percentage of female respondents who didn't want their time of watch to be considered while getting recommendations 44.31818181818182 %
Percentage of female respondents who didn't want their watch duration to be considered while getting recommendations 38.63636363636363 %

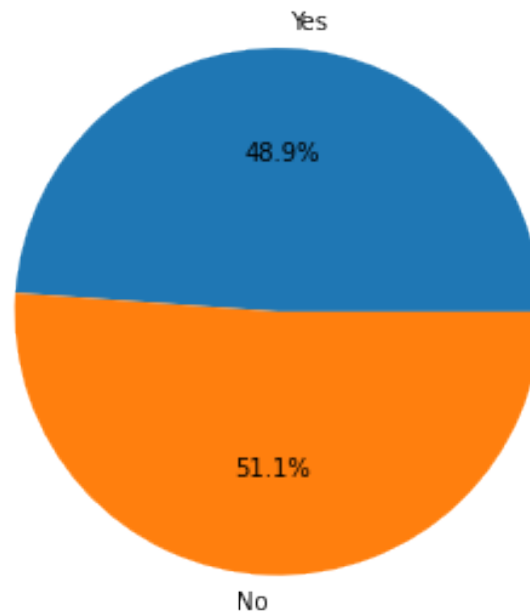### 9.0.2 Aware of attributes used to give recommendations?

```
[22]: filtered_df = df.where(filter_female & filter_informed).dropna(how="all")
      print("Percentage of female respondents who knew what information is used to␣
       ↪give recommendations", filtered_df.shape[0]*100/women, "%")


      y = np.array([filtered_df.shape[0]*100/women, 100 - (filtered_df.shape[0]*100/
       ↪women)])
      mylabels = ["Yes", "No"]

      plt.figure(figsize=(5,5))
      plt.title('% of women who knew what information is used to give␣
       ↪recommendations')
      plt.pie(y, labels = mylabels,autopct='%1.1f%%')
      plt.savefig("women_information.png")
      plt.show()
```

Percentage of female respondents who knew what information is used to give recommendations 48.86363636363637 %

% of women who knew what information is used to give recommendations

9.0.3 Want an option to choose which attributes are used to give recommendations?

```
[23]: filtered_df = df.where(filter_female & filter_choose).dropna(how="all")
      print("Percentage of female respondents who wanted an option to choose which␣
       ↪attributes to use while giving recommendations", filtered_df.shape[0]*100/
       ↪women, "%")

      y = np.array([filtered_df.shape[0]*100/women, 100 - (filtered_df.shape[0]*100/
       ↪women)])
      mylabels = ["Yes", "No"]

      plt.figure(figsize=(5,5))
      plt.title('% of women who wanted an option to choose which attributes to use␣
       ↪while giving recommendations')
      plt.pie(y, labels = mylabels,autopct='%1.1f%%')
      plt.savefig("women_choose.png")
      plt.show()
```

Percentage of female respondents who wanted an option to choose which attributes
to use while giving recommendations 93.18181818181819 %

% of women who wanted an option to choose which attributes to use while giving recommendations