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

Social Survey

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?

We conducted an online survey to answer these questions. We surveyed around `384` 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
```

Results

Link to analytics page for responses

Link to raw response data

Analysis Setup

Imports

In [1]:
```
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

Reading in the data

In [2]:
```
df = pd.read_csv('./responses.csv')
# Renaming the Columns with appropriate names
df.columns = ['Timestamp', 'Gender', 'Age',  'PlatformsUsed', 'IsInformed', 'Option','Attributes']
```

Overall Statistics

Gathered from google forms data.

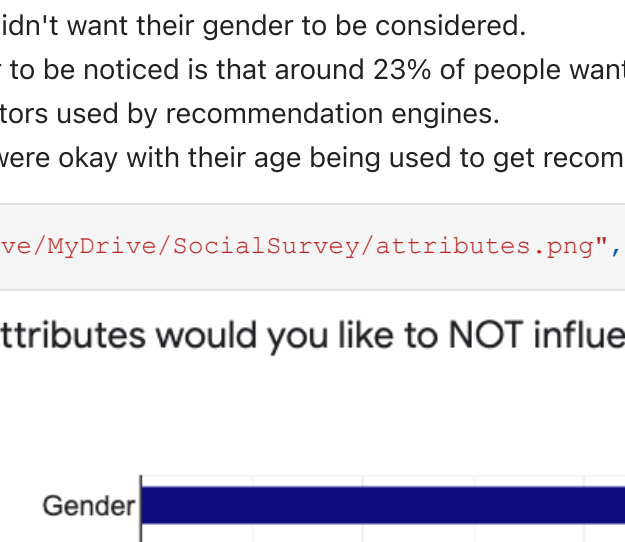
A whopping 87% of people wanted to have the option to choose which attributes are taken into account while getting recommendations!

In [3]:
```
Image("/content/drive/MyDrive/SocialSurvey/choose.png", width = 800)
```

Out[3]:

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.

384 responses



Attributes that people do not want to influence their recommendations

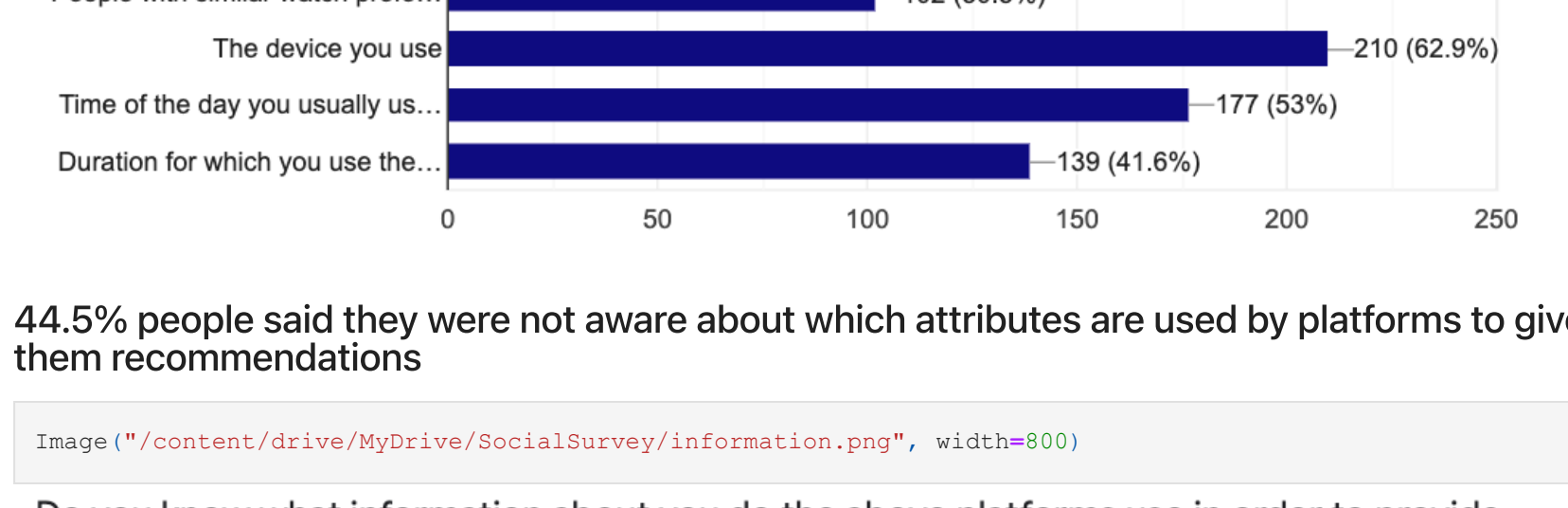
- 50% of the people didn't want their gender to be considered.
- An interesting factor to be noticed is that around 23% of people wanted previous watch history to not be considered, while it is one of the prime factors used by recommendation engines.
- Only around 636% were okay with their age being used to get recommendations.

In [4]:
```
Image("/content/drive/MyDrive/SocialSurvey/attributes.png", width = 800)
```

Out[4]:

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

334 responses



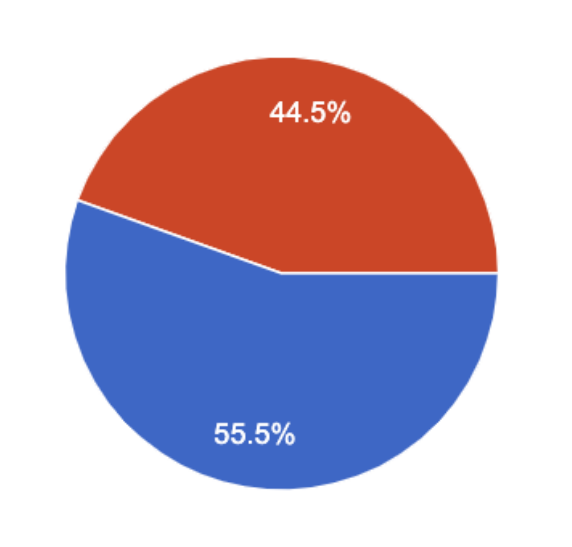
44.5% people said they were not aware about which attributes are used by platforms to give them recommendations

In [5]:
```
Image("/content/drive/MyDrive/SocialSurvey/information.png", width=800)
```

Out[5]:

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

384 responses

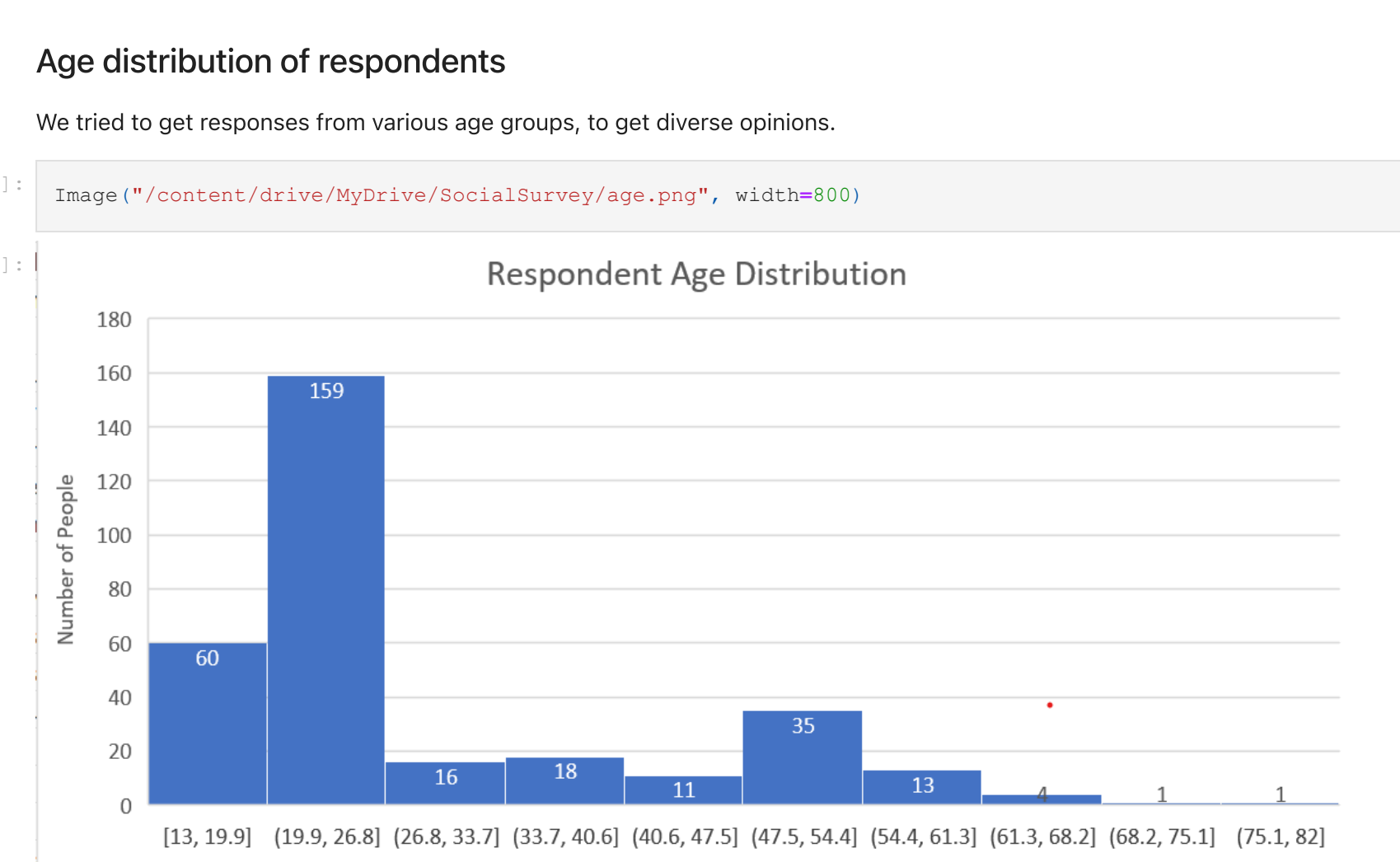


Age distribution of respondents

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

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

Out[6]:



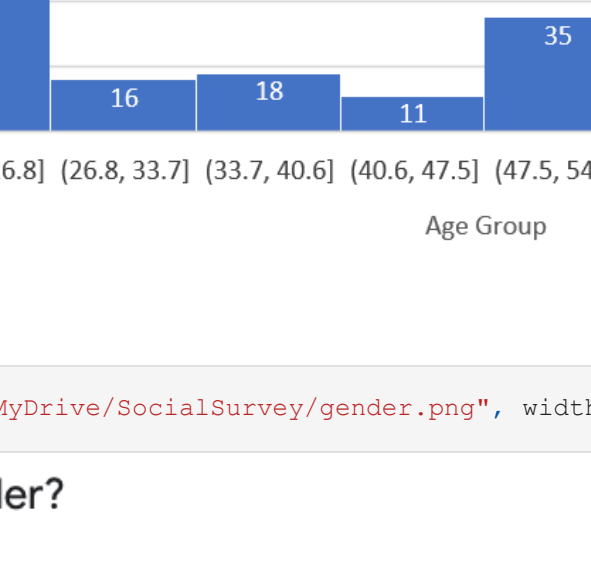
Gender distribution

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

Out[7]:

What is your Gender?

384 responses



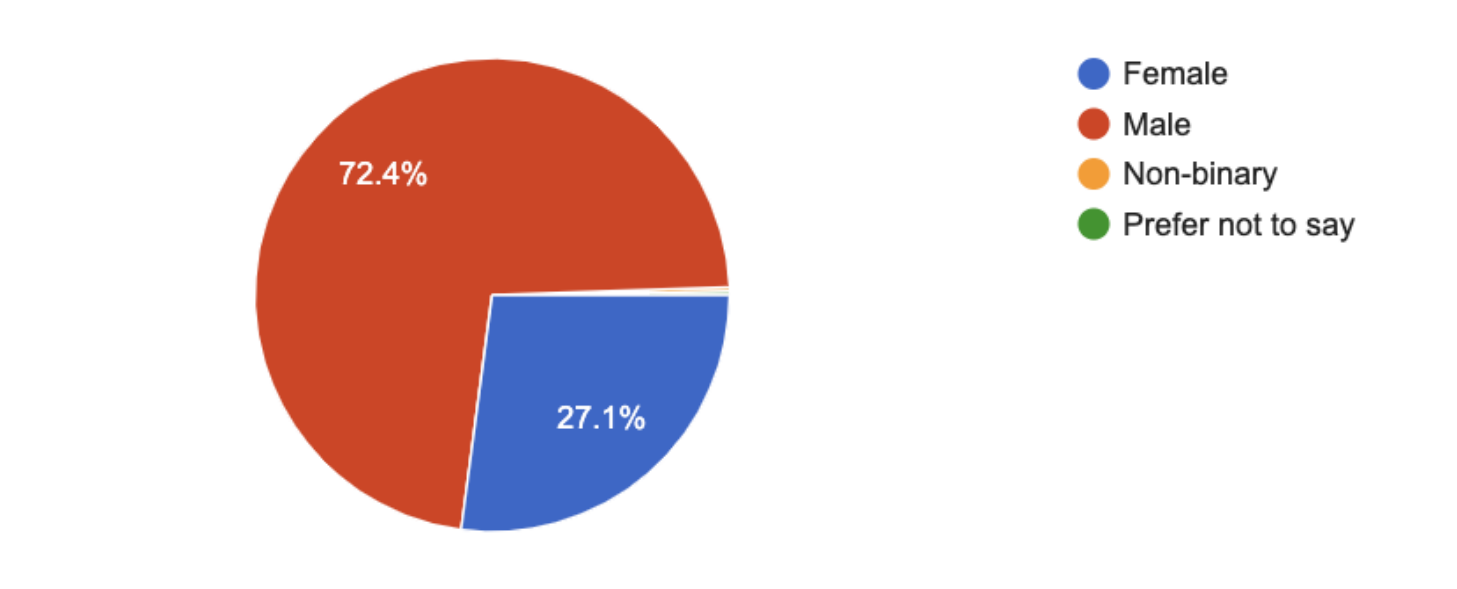
Platforms used regularly

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

Out[8]:

Which of the following platforms do you use regularly?

384 responses



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.

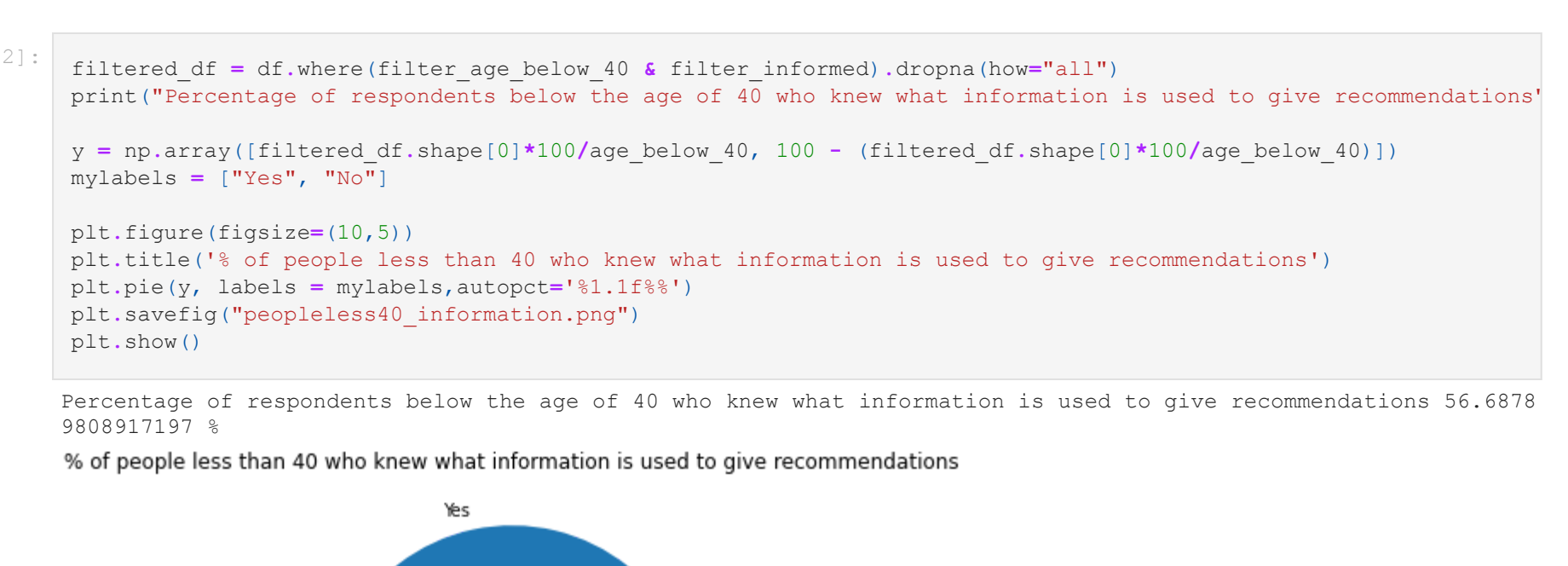
- Almost all women, ~93% wanted the option to choose the attributes being used to give recommendation, ~85% men wanted to have the option.

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

Demographic 1.a: Age below 40

In [21]:
```
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 below age of 40:", age_below_40)

filtered_df = df.where(filter_age_below_40 & filter_attributes_gender).dropna(how="all")
print("Percentage of respondents below the age of 40 who didn't want their gender to be considered while gettin
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 below  the age of 40 who didn't want their age to be considered while getting
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 below  the age of 40 who didn't want their region of stay to be considered whi
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 below  the age of 40 who didn't want their watch history to be considered whil
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 below  the age of 40 who didn't want the people with similar watch preferences
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 below  the age of 40 who didn't want their device of use to be considered whil
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 below  the age of 40 who didn't want their time of watch to be considered whil
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 below  the age of 40 who didn't want their watch duration to be considered whi
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', 'D
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()
```

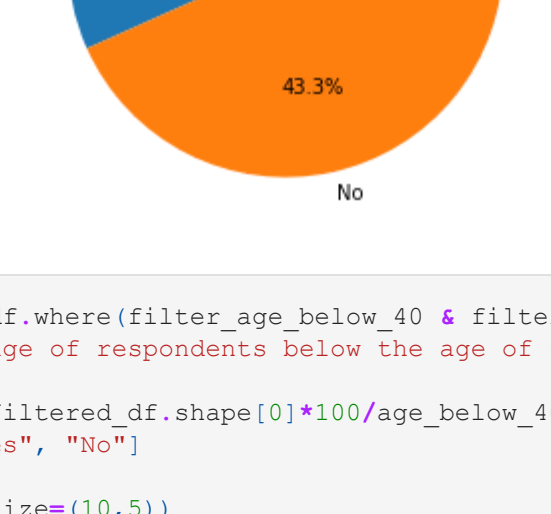
```
Respondents below age of 40: 314
Percentage of respondents below  the age of 40 who didn't want their gender to be considered while getting recom
mendations 45.22292993630573 %
Percentage of respondents below  the age of 40 who didn't want their age to be considered while getting recomme
ndations 31.528662420382165 %
Percentage of respondents below  the age of 40 who didn't want their region of stay to be considered while gett
ing recommendations 48.40764331210191 %
Percentage of respondents below  the age of 40 who didn't want their watch history to be considered while getti
ng recommendations 18.15286624203822 %
Percentage of respondents below  the age of 40 who didn't want the people with similar watch preferences to be
considered while getting recommendations 24.203821656050955 %
Percentage of respondents below  the age of 40 who didn't want their device of use to be considered while getti
ng recommendations 56.68789808917197 %
Percentage of respondents below  the age of 40 who didn't want their time of watch to be considered while getti
ng recommendations 45.22292993630573 %
Percentage of respondents below  the age of 40 who didn't want their watch duration to be considered while gett
ing recommendations 33.75796178343949 %
```



In [22]:
```
filtered_df = df.where(filter_age_below_40 & filter_informed).dropna(how="all")
print("Percentage of respondents below the age of 40 who knew what information is used to give recommendations"

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=(10,5))
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 below the age of 40 who knew what information is used to give recommendations 56.6878
9808917197 %
```

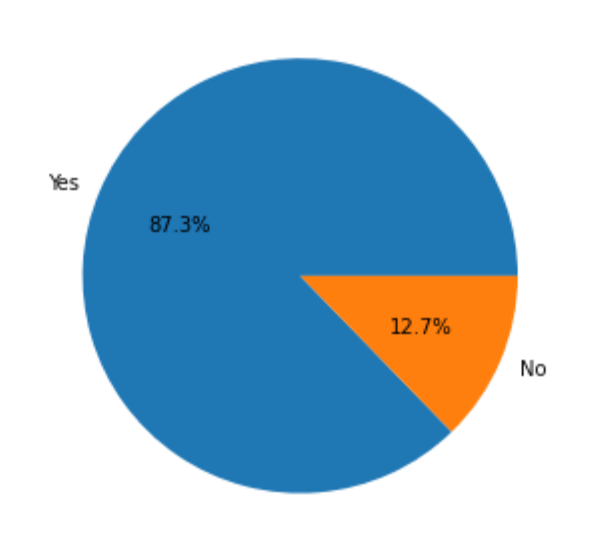
% of people less than 40 who knew what information is used to give recommendations



In [23]:
```
filtered_df = df.where(filter_age_below_40 & filter_choose).dropna(how="all")
print("Percentage of respondents below the age of 40 who wanted an option to choose which attributes to use whi

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=(10,5))
plt.title('% of people less than 40 wanted an option to choose which attributes to use while giving recommendat
plt.pie(y, labels = mylabels,autopct='%1.1f%%')
plt.savefig("peopleless40_choose.png")
plt.show()
```

```
Percentage of respondents below the age of 40 who wanted an option to choose which attributes to use while givi
ng recommendations 87.26114649681529 %
```

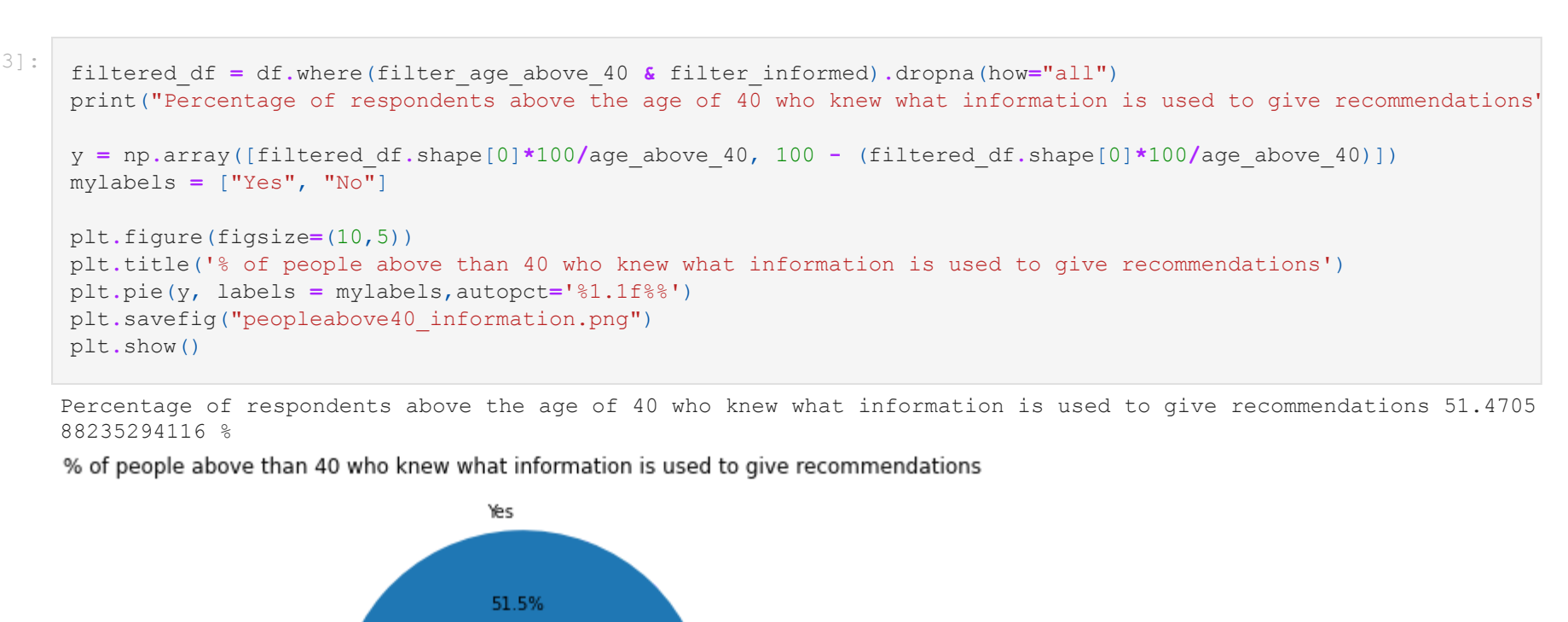
% of people less than 40 wanted an option to choose which attributes to use while giving recommendations



Demographic 1.b: Age above 40

In [12]:
```
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 gettin
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")
print("Percentage of respondents above the age of 40 who didn't want their age to be considered while getting r
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 whil
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
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
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
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
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 whil
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', 'D
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()
```

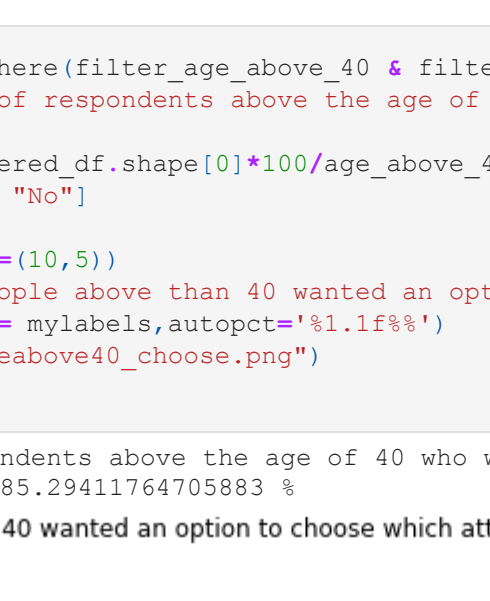
```
Respondents above age of 40: 68
Percentage of respondents above the age of 40 who didn't want their gender to be considered while getting recom
mendations 32.35294117647059 %
Percentage of respondents above the age of 40 who didn't want their age to be considered while getting recommen
dations 32.35294117647059 %
Percentage of respondents above the age of 40 who didn't want their region of stay to be considered while getti
ng recommendations 45.588235294117645 %
Percentage of respondents above the age of 40 who didn't want their watch history to be considered while gettin
g recommendations 27.941176470588236 %
Percentage of respondents above the age of 40 who didn't want the people with similar watch preferences to be c
onsidered while getting recommendations 38.23529411764706 %
Percentage of respondents above the age of 40 who didn't want their device of use to be considered while gettin
g recommendations 45.588235294117645 %
Percentage of respondents above the age of 40 who didn't want their time of watch to be considered while gettin
g recommendations 48.529411764705884 %
Percentage of respondents above the age of 40 who didn't want their watch duration to be considered while getti
ng recommendations 47.05882352941177 %
```



In [13]:
```
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"

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=(10,5))
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 51.4705
88235294116 %
```

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



In [14]:
```
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 whi

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=(10,5))
plt.title('% of people above than 40 wanted an option to choose which attributes to use while giving recommenda
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 givi
ng recommendations 85.29411764705883 %
```

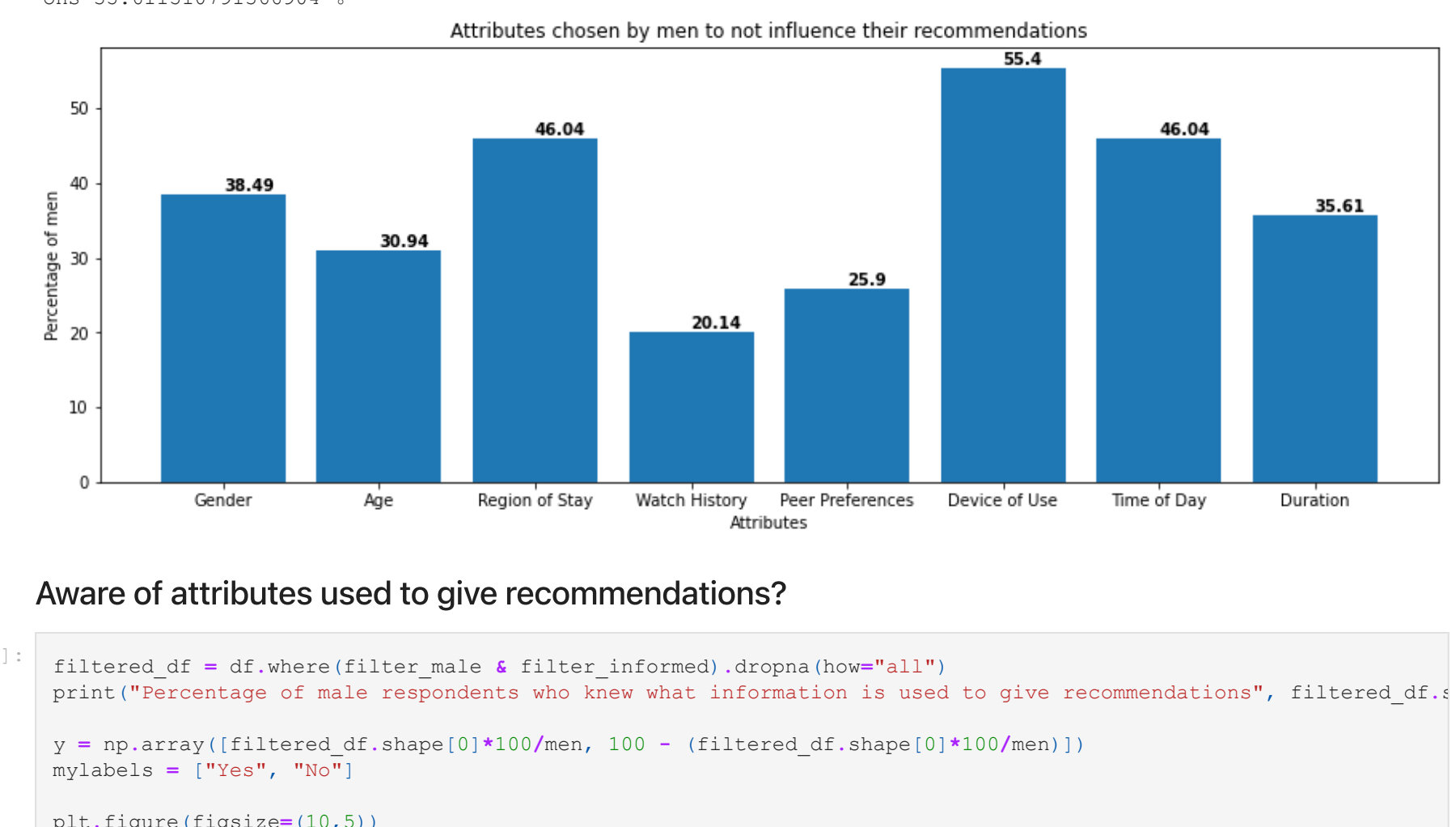
% of people above than 40 wanted an option to choose which attributes to use while giving recommendations

Demographic 2.a: Male respondents

Attributes to not be considered

In [15]:
```
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)

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 recommendatio
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"
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 recom
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 recomm
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 considere
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 recomm
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 recomm
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 recom
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', 'D
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: 278
Percentage of male respondents who didn't want their gender to be considered while getting recommendations 38.4
89208633093526 %
Percentage of male respondents who didn't want their age to be considered while getting recommendations 30.9352
5179856115 %
Percentage of male respondents who didn't want their region of stay to be considered while getting recommendati
ons 46.043165467625904 %
Percentage of male respondents who didn't want their watch history to be considered while getting recommendatio
ns 20.14388489208633 %
Percentage of male respondents who didn't want the people with similar watch preferences to be considered while
getting recommendations 25.899280575539567 %
Percentage of male respondents who didn't want their device of use to be considered while getting recommendatio
ns 55.39568345323741 %
Percentage of male respondents who didn't want their time of watch to be considered while getting recommendatio
ns 46.043165467625904 %
Percentage of male respondents who didn't want their watch duration to be considered while getting recommendati
ons 35.611510791366904 %
```

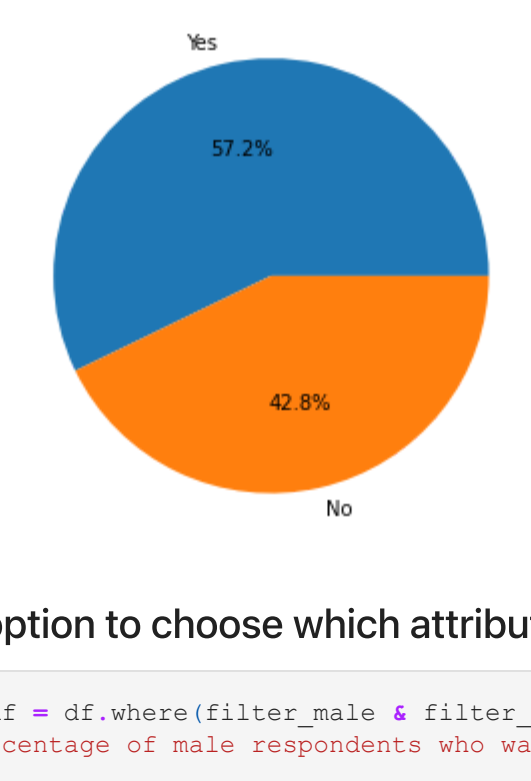


Aware of attributes used to give recommendations?

In [16]:
```
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.s

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

plt.figure(figsize=(10,5))
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 57.194244604316545 %
```

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

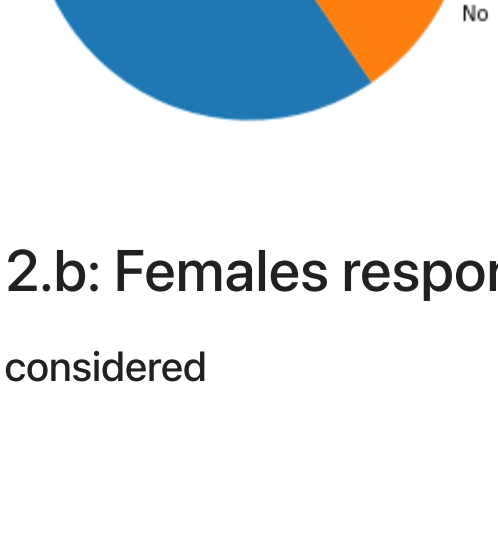


Want an option to choose which attributes are used to give recommendations?

In [17]:
```
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 recom

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

plt.figure(figsize=(10,5))
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 recommendati
ons 84.53237410071942 %
```

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

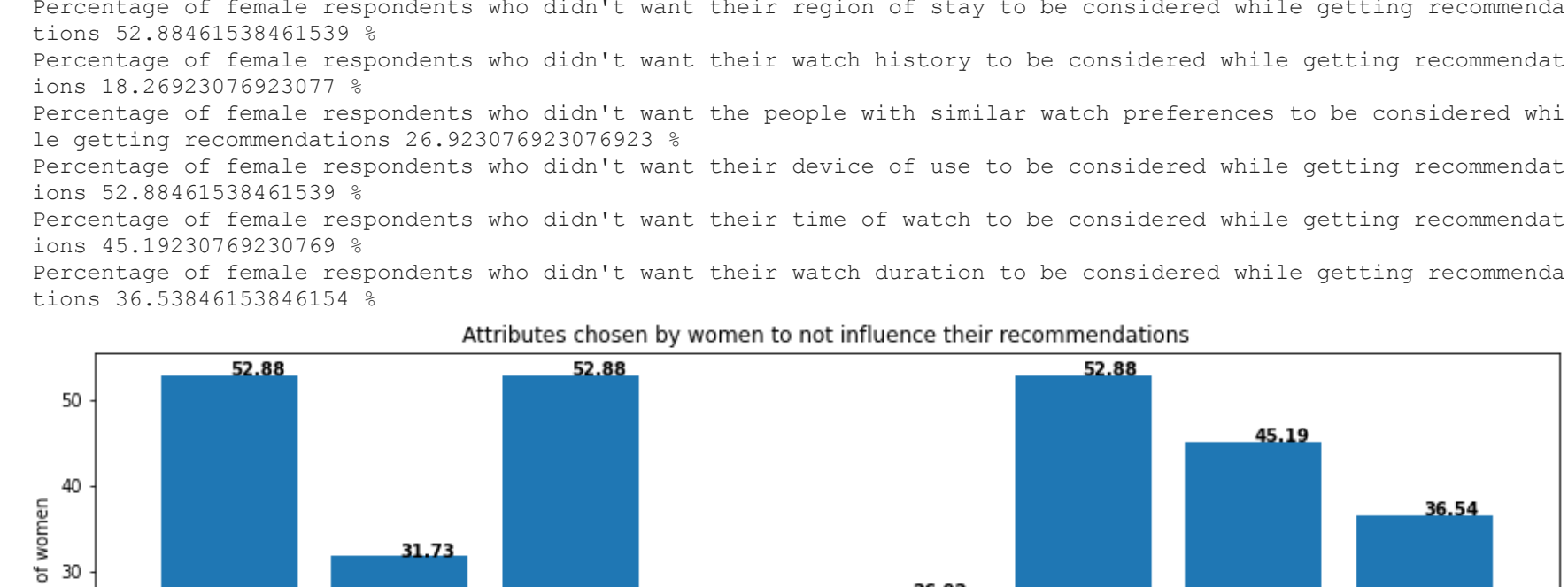


Demographic 2.b: Females respondents

Attributes to not be considered

In [18]:
```
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)
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 recommenda
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 recommendatio
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 rec
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 reco
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 conside
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 reco
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 reco
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 rec
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', 'D
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: 104
Percentage of female respondents who didn't want their gender to be considered while getting recommendations 5
2.884615384615
39 %
Percentage of female respondents who didn't want their age to be considered while getting recommendations 31.73
076923076923 %
Percentage of female respondents who didn't want their region of stay to be considered while getting recommenda
tions 52.88461538461539 %
Percentage of female respondents who didn't want their watch history to be considered while getting recommendat
ions 18.26923076923077 %
Percentage of female respondents who didn't want the people with similar watch preferences to be considered whi
le getting recommendations 26.923076923076923 %
Percentage of female respondents who didn't want their device of use to be considered while getting recommendat
ions 52.88461538461539 %
Percentage of female respondents who didn't want their time of watch to be considered while getting recommendat
ions 45.19230769230769 %
Percentage of female respondents who didn't want their watch duration to be considered while getting recommenda
tions 36.53846153846154 %
```

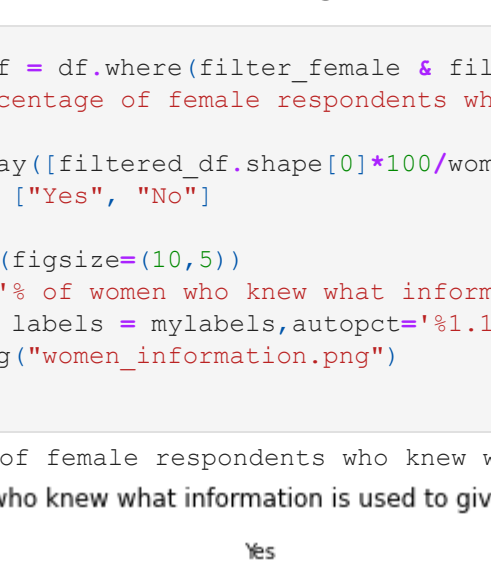


Aware of attributes used to give recommendations?

In [19]:
```
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

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

plt.figure(figsize=(10,5))
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 50.96153846153846 %
```

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



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

In [20]:
```
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 rec

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

plt.figure(figsize=(10,5))
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 recommenda
tions 93.26923076923077 %
```

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