In [1]:
```
#Data Preparation

#Code goes after this line by adding cells

#Import packages
import pandas as pd
import numpy as np
```

In [2]:
```
#Create variable to store CSV file
starwars_p = 'StarWars.csv'
```

In [3]:
```
#The imported csv file has two Columns, hence I have combined the 2 columns to
create 1 heading
#The combined columns are as follows
#RespondentID
#Q1
#Q2
#Q3-Star Wars: Episode I  The Phantom Menace
#Q3-Star Wars: Episode II  Attack of the Clones
#Q3-Star Wars: Episode III  Revenge of the Sith
#Q3-Star Wars: Episode IV  A New Hope
#Q3-Star Wars: Episode V The Empire Strikes Back
#Q3-Star Wars: Episode VI Return of the Jedi
#Q4-Star Wars: Episode I  The Phantom Menace
#Q4-Star Wars: Episode II  Attack of the Clones
#Q4-Star Wars: Episode III  Revenge of the Sith
#Q4-Star Wars: Episode IV  A New Hope
#Q4-Star Wars: Episode V The Empire Strikes Back
#Q4-Star Wars: Episode VI Return of the Jedi
#Q5-Han Solo
#Q5-Luke Skywalker
#Q5-Princess Leia Organa
#Q5-Anakin Skywalker
#Q5-Obi Wan Kenobi
#Q5-Emperor Palpatine
#Q5-Darth Vader
#Q5-Lando Calrissian
#Q5-Boba Fett
#Q5-C-3P0
#Q5-R2 D2
#Q5-Jar Jar Binks
#Q5-Padme Amidala
#Q5-Yoda
#Q6
#Q7
#Q8
#Q9
#Gender
#Age
#Household Income
#Education
#Location (Census Region)

#Q1     Have you seen any of the 6 films in the Star Wars franchise?
#Q2     Do you consider yourself to be a fan of the Star Wars film franchise?
#Q3     Which of the following Star Wars films have you seen? Please select al
l that apply.
#Q4     Please rank the Star Wars films in order of preference with 1 being yo
ur favorite film in the franchise and 6 being your least favorite film.
#Q5     Please state whether you view the following characters favorably, unfa
vorably, or are unfamiliar with him/her.
#Q6     Which character shot first?
#Q7     Are you familiar with the Expanded Universe?
#Q8     Do you consider yourself to be a fan of the Expanded Universe?Œæ
#Q9     Do you consider yourself to be a fan of the Star Trek franchise?

starwars = pd.read_csv(starwars_p, sep=',', decimal='.', header = None, skipro
```

```
ws = 2, names=['RespondentID','Q1','Q2','Q3-Star Wars: Episode I  The Phantom
 Menace','Q3-Star Wars: Episode II  Attack of the Clones','Q3-Star Wars: Episo
de III  Revenge of the Sith','Q3-Star Wars: Episode IV  A New Hope','Q3-Star W
ars: Episode V The Empire Strikes Back','Q3-Star Wars: Episode VI Return of th
e Jedi','Q4-Star Wars: Episode I  The Phantom Menace','Q4-Star Wars: Episode I
I  Attack of the Clones','Q4-Star Wars: Episode III  Revenge of the Sith','Q4-
Star Wars: Episode IV  A New Hope','Q4-Star Wars: Episode V The Empire Strikes
Back','Q4-Star Wars: Episode VI Return of the Jedi','Q5-Han Solo','Q5-Luke Sky
walker','Q5-Princess Leia Organa','Q5-Anakin Skywalker','Q5-Obi Wan Kenobi','Q
5-Emperor Palpatine','Q5-Darth Vader','Q5-Lando Calrissian','Q5-Boba Fett','Q5
-C-3P0','Q5-R2 D2','Q5-Jar Jar Binks','Q5-Padme Amidala','Q5-Yoda','Q6','Q7',
'Q8','Q9','Gender','Age','Household Income','Education','Location (Census Regi
on)'])
```

In [4]:
```
#The dataset imported from the csv file
starwars.head()
```

Out[4]:

	RespondentID	Q1	Q2	Q3-Star Wars: Episode I The Phantom Menace	Q3-Star Wars: Episode II Attack of the Clones	Q3-Star Wars: Episode III Revenge of the Sith	Q3-Star Wars: Episode IV A New Hope	Q3-Star Wars: Episode V The Empire Strikes Back	Q3-Star Wars: Episode VI Return of the Jedi	Q4-Star Wars: Episode I The Phantom Menace
0	3292879998	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	3.0
1	3292879538	No	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	3292765271	Yes	No	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	NaN	NaN	NaN	1.0
3	3292763116	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	5.0
4	3292731220	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	5.0

5 rows × 38 columns

In [5]:
```
#Checked and corrected the Typos in the all the columns
starwars['Q2'].value_counts()
```

Out[5]:
```
Yes     551
No      283
Noo       1
Yess      1
Name: Q2, dtype: int64
```

In [6]:
```
starwars['Q2'].replace({"Yess": "Yes", "Noo": "No"}, inplace=True)
```

In [7]:
```
starwars['Q2'].value_counts()
```

Out[7]:
```
Yes    552
No     284
Name: Q2, dtype: int64
```

In [8]:
```
starwars['Q1'].value_counts()
```

Out[8]:
```
Yes     935
No      250
Yes       1
Name: Q1, dtype: int64
```

In [9]:
```
starwars['Q1'] = starwars['Q1'].str.strip()
```

In [10]:
```
starwars['Q1'].value_counts()
```

Out[10]:
```
Yes    936
No     250
Name: Q1, dtype: int64
```

In [11]:
```
starwars['Q5-Boba Fett'].value_counts()
```

Out[11]:
```
Neither favorably nor unfavorably (neutral)    248
Somewhat favorably                             153
Very favorably                                 138
Unfamiliar (N/A)                               132
Somewhat unfavorably                            96
Very unfavorably                                45
Name: Q5-Boba Fett, dtype: int64
```

In [12]:
```
starwars['Q8'].value_counts()
```

Out[12]:
```
No      114
Yes      98
Yess      1
Name: Q8, dtype: int64
```

In [13]:
```
starwars['Q8'].replace({"Yess": "Yes"}, inplace=True)
```

In [14]:
```
starwars['Q8'].value_counts()
```

Out[14]:
```
No     114
Yes     99
Name: Q8, dtype: int64
```

In [15]:
```
starwars['Q9'].value_counts()
```

Out[15]:
```
No     639
Yes    426
Noo      1
no       1
yes      1
Name: Q9, dtype: int64
```

In [16]:
```
starwars['Q9'].replace({"yes": "Yes","Noo":"No","no":"No"}, inplace=True)
```

In [17]:
```
starwars['Q9'].value_counts()
```

Out[17]:
```
No     640
Yes    427
no       1
Name: Q9, dtype: int64
```

In [18]:
```
#Removed the whitspaces, new line characters, etc. from the columns using str.
strip()
starwars['Q9'] = starwars['Q9'].str.strip()
```

In [19]:
```
starwars['Q9'].replace({"yes": "Yes","Noo":"No","no":"No"}, inplace=True)
```

In [20]:
```
starwars['Q9'].value_counts()
```

Out[20]:
```
No     641
Yes    427
Name: Q9, dtype: int64
```

In [21]:
```
starwars['Gender'].value_counts()
```

Out[21]:
```
Female    546
Male      496
female      2
male        1
F           1
Name: Gender, dtype: int64
```

In [22]:
```
starwars['Gender'] = starwars['Gender'].str.strip()
```

In [23]:
```
#Case of the text data was changed wherever necessary
starwars['Gender'].replace({"female": "Female","male":"Male","F":"Female"}, in
place=True)
```

In [24]:
```
starwars['Gender'].value_counts()
```

Out[24]:
```
Female    549
Male      497
Name: Gender, dtype: int64
```

In [25]:
```
starwars['Age'].value_counts()
```

Out[25]:
```
45-60    291
> 60     269
30-44    268
18-29    217
500        1
Name: Age, dtype: int64
```

In [26]:
```
#Age column had value 500, which is an impossible value, which was removed by
 replacing it with the Mode value as finding out the mean and the median is im
possible for a range
starwars['Age'].replace({"500": "45-60"}, inplace=True)
```

In [27]:
```
starwars['Age'].value_counts()
```

Out[27]:
```
45-60    292
> 60     269
30-44    268
18-29    217
Name: Age, dtype: int64
```

In [28]:
```
starwars['Household Income'].value_counts()
```

Out[28]:
```
$50,000 - $99,999      298
$25,000 - $49,999      186
$100,000 - $149,999    141
$0 - $24,999           138
$150,000+               95
Name: Household Income, dtype: int64
```

In [29]:
```
#NaN values were filled with "NotAnswered"
starwars = starwars.fillna('NotAnswered')
```

In [30]:
```
starwars.head()
```

Out[30]:

	RespondentID	Q1	Q2	Q3-Star Wars: Episode I The Phantom Menace	Q3-Star Wars: Episode II Attack of the Clones	Q3-Star Wars: Episode III Revenge of the Sith	Q3-Star Wars: Episode IV A New Hope	Ep The Strik
0	3292879998	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	St Ep The Strik
1	3292879538	No	NotAnswered	NotAnswered	NotAnswered	NotAnswered	NotAnswered	NotA
2	3292765271	Yes	No	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	NotAnswered	NotA
3	3292763116	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	St Ep The Strik
4	3292731220	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	St Ep The Strik

5 rows × 38 columns

In [31]:
```
starwars.drop(starwars[(starwars.Q1 == 'Yes') & (starwars.Q2 == 'NotAnswered')
& (starwars.Age == 'NotAnswered')].index, inplace=False)
```

Out[31]:

	RespondentID	Q1	Q2	Q3-Star Wars: Episode I The Phantom Menace	Q3-Star Wars: Episode II Attack of the Clones	Q3-Star Wars: Episode III Revenge of the Sith	Q3-Star Wars: Episode IV A New Hope	S
0	3292879998	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	S
1	3292879538	No	NotAnswered	NotAnswered	NotAnswered	NotAnswered	NotAnswered	N
2	3292765271	Yes	No	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	NotAnswered	N
3	3292763116	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	S
4	3292731220	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	S
...	...	...	...	...	...	...	...	
1181	3288388730	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	S
1182	3288378779	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	S
1183	3288375286	No	NotAnswered	NotAnswered	NotAnswered	NotAnswered	NotAnswered	N
1184	3288373068	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	S
1185	3288372923	Yes	No	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	NotAnswered	NotAnswered	S

1086 rows × 38 columns

In [32]:
```
type('Location (Census Region)')
```

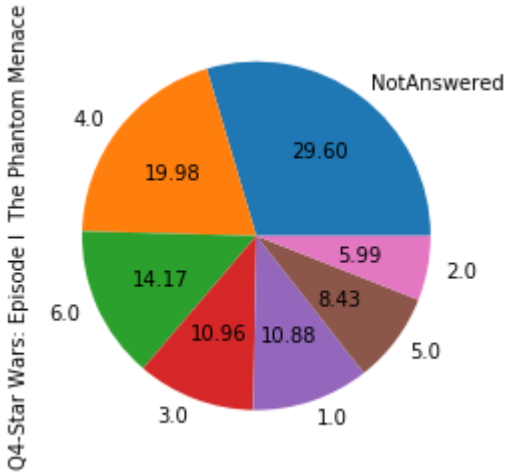
Out[32]:
```
str
```

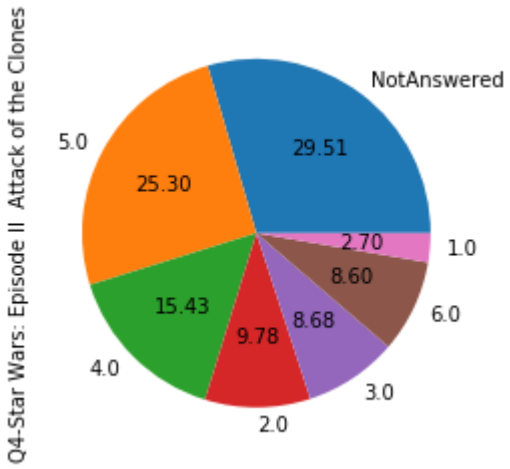
In [33]:
```
starwars['Q4-Star Wars: Episode I  The Phantom Menace'].value_counts()
```

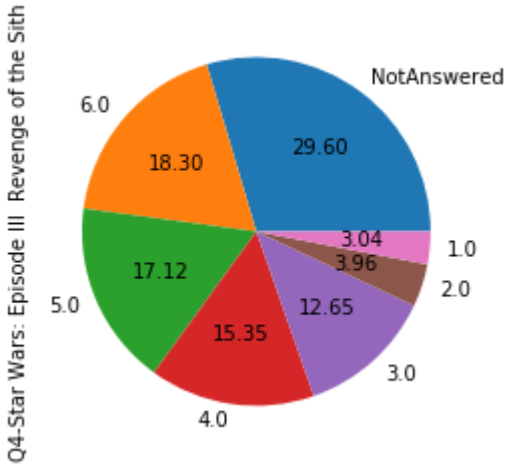
Out[33]:
```
NotAnswered    351
4.0            237
6.0            168
3.0            130
1.0            129
5.0            100
2.0             71
Name: Q4-Star Wars: Episode I  The Phantom Menace, dtype: int64
```

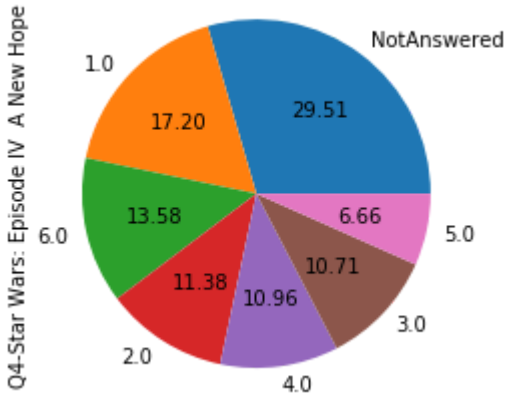
In [34]:
```
# (Data Exploration) The question is analyzed and the visualization used for t
he comparison is the pie chart.
import matplotlib.pyplot as plt
```

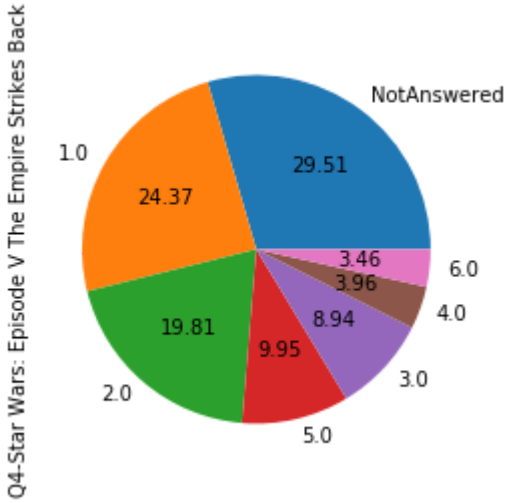
In [35]:
```
starwars['Q4-Star Wars: Episode I  The Phantom Menace'].value_counts(normalize
=True).plot(kind='pie',autopct='%.2f')
plt.show()
starwars['Q4-Star Wars: Episode II  Attack of the Clones'].value_counts(normal
ize=True).plot(kind='pie',autopct='%.2f')
plt.show()
starwars['Q4-Star Wars: Episode III  Revenge of the Sith'].value_counts(normal
ize=True).plot(kind='pie',autopct='%.2f')
plt.show()
starwars['Q4-Star Wars: Episode IV  A New Hope'].value_counts(normalize=True).
plot(kind='pie',autopct='%.2f')
plt.show()
starwars['Q4-Star Wars: Episode V The Empire Strikes Back'].value_counts(norma
lize=True).plot(kind='pie',autopct='%.2f')
plt.show()
starwars['Q4-Star Wars: Episode VI Return of the Jedi'].value_counts(normalize
=True).plot(kind='pie',autopct='%.2f')
plt.show()
```

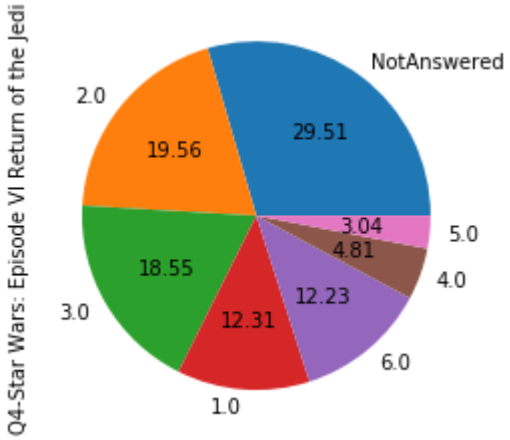

Q4-Star Wars: Episode I The Phantom Menace
NotAnswered
29.60
4.0
19.98
5.99
2.0
14.17
8.43
6.0
10.96
10.88
5.0
3.0
1.0


Q4-Star Wars: Episode II Attack of the Clones
NotAnswered
29.51
5.0
25.30
2.70
1.0
8.60
15.43
6.0
9.78
8.68
4.0
3.0
2.0

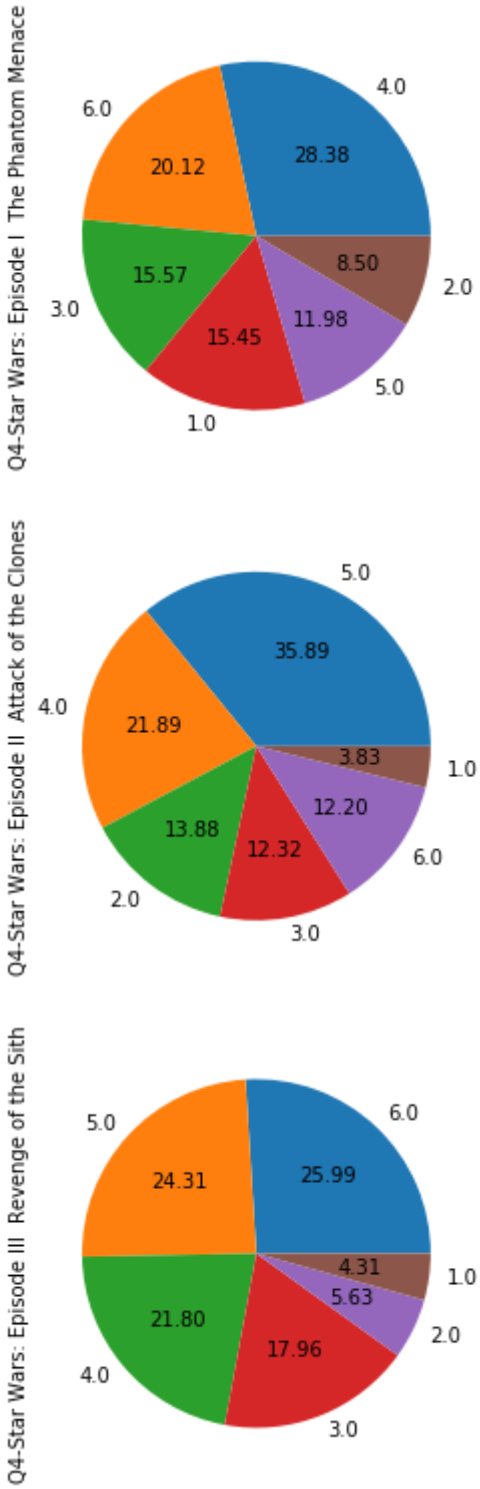

Q4-Star Wars: Episode III Revenge of the Sith
NotAnswered
6.0
29.60
18.30
3.04
1.0
3.96
17.12
2.0
5.0
12.65
15.35
3.0
4.0

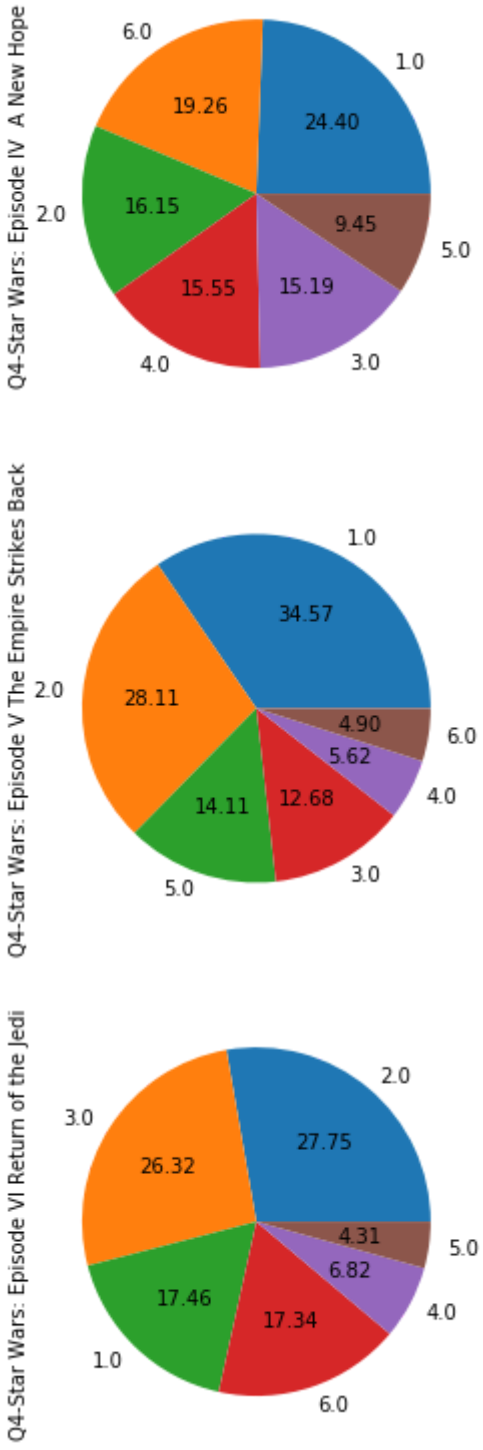

Q4-Star Wars: Episode IV A New Hope
NotAnswered
29.51
1.0
17.20
6.0
13.58
5.0
6.66
10.71
3.0
11.38
10.96
2.0
4.0


Q4-Star Wars: Episode V The Empire Strikes Back
NotAnswered
29.51
1.0
24.37
3.46
6.0
3.96
4.0
8.94
19.81
9.95
3.0
2.0
5.0


Q4-Star Wars: Episode VI Return of the Jedi
NotAnswered
29.51
2.0
19.56
3.04
5.0
4.81
18.55
4.0
12.23
3.0
12.31
6.0
1.0

In [36]:
```
#The pie charts were plotted after ignoring the NotAnswered values.

plt.figure(0)
starwars['Q4-Star Wars: Episode I  The Phantom Menace'].replace({"NotAnswered"
:np.nan}).value_counts(normalize=True).plot(kind='pie',autopct='%.2f')
plt.show()
plt.figure(1)
starwars['Q4-Star Wars: Episode II  Attack of the Clones'].replace({"NotAnswer
ed":np.nan}).value_counts(normalize=True).plot(kind='pie',autopct='%.2f')
plt.show()
plt.figure(2)
starwars['Q4-Star Wars: Episode III  Revenge of the Sith'].replace({"NotAnswer
ed":np.nan}).value_counts(normalize=True).plot(kind='pie',autopct='%.2f')
plt.show()
plt.figure(3)
starwars['Q4-Star Wars: Episode IV  A New Hope'].replace({"NotAnswered":np.nan
}).value_counts(normalize=True).plot(kind='pie',autopct='%.2f')
plt.show()
plt.figure(4)
starwars['Q4-Star Wars: Episode V The Empire Strikes Back'].replace({"NotAnswe
red":np.nan}).value_counts(normalize=True).plot(kind='pie',autopct='%.2f')
plt.show()
plt.figure(5)
starwars['Q4-Star Wars: Episode VI Return of the Jedi'].replace({"NotAnswered"
:np.nan}).value_counts(normalize=True).plot(kind='pie',autopct='%.2f')
plt.show()
```


Q4-Star Wars: Episode I The Phantom Menace
4.0
28.38
6.0
20.12
15.57
3.0
8.50
2.0
11.98
15.45
5.0
1.0
Q4-Star Wars: Episode II Attack of the Clones
5.0
35.89
4.0
21.89
3.83
1.0
12.20
13.88
12.32
6.0
2.0
3.0
Q4-Star Wars: Episode III Revenge of the Sith
6.0
5.0
24.31
25.99
4.31
1.0
5.63
21.80
17.96
2.0
4.0
3.0

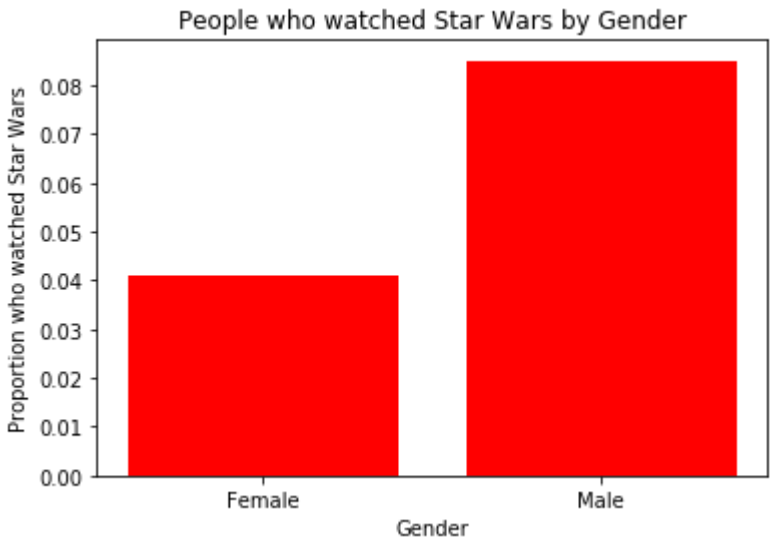


In [37]:
```
starwars['Age'].value_counts()
```

Out[37]:
```
45-60          292
> 60           269
30-44          268
18-29          217
NotAnswered    140
Name: Age, dtype: int64
```

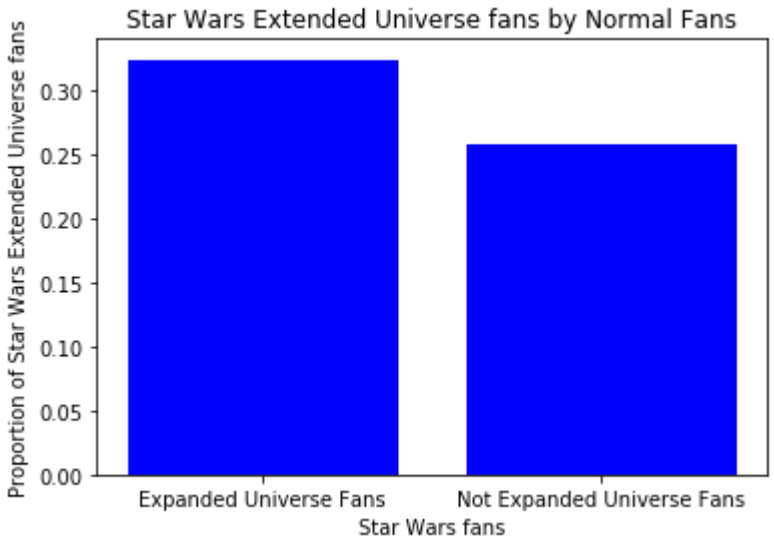
In [38]:
```
# Have you seen any of the 6 films in the Star Wars franchise? and Gender
starwars_gender = starwars['Gender'].replace({"NotAnswered":np.nan}).dropna().
value_counts()

gender_f_mask = starwars['Gender'].replace({"NotAnswered":np.nan}).dropna() ==
'Female'
gender_m_mask = starwars['Gender'].replace({"NotAnswered":np.nan}).dropna() ==
'Male'

gender_Q1_f = starwars.replace({"NotAnswered":np.nan}).dropna().loc[gender_f_m
ask,'Q1'].value_counts()
gender_Q1_m = starwars.replace({"NotAnswered":np.nan}).dropna().loc[gender_m_m
ask,'Q1'].value_counts()

rate = [gender_Q1_f['Yes']/float(starwars_gender['Female']+starwars_gender['Ma
le']),
        gender_Q1_m['Yes']/float(starwars_gender['Female']+starwars_gender['Ma
le'])]

plt.bar(list(range(2)), rate, color='r', align='center')
plt.xticks(list(range(2)),['Female','Male'])
plt.xlabel('Gender')
plt.ylabel('Proportion who watched Star Wars')
plt.title('People who watched Star Wars by Gender')
```

Out[38]: Text(0.5, 1.0, 'People who watched Star Wars by Gender')



In [39]:
```
# Do you consider yourself to be a fan of the Star Wars film franchise? and Do you consider yourself to be a fan of the Expanded Universe?
starwars_Q8 = starwars['Q8'].replace({"NotAnswered":np.nan}).dropna().value_counts()

Q8_y_mask = starwars['Q8'].replace({"NotAnswered":np.nan}).dropna() == 'Yes'
Q8_n_mask = starwars['Q8'].replace({"NotAnswered":np.nan}).dropna() == 'No'

Q8_y_Q2 = starwars.replace({"NotAnswered":np.nan}).dropna().loc[Q8_y_mask,'Q2'].value_counts()
Q8_n_Q2 = starwars.replace({"NotAnswered":np.nan}).dropna().loc[Q8_n_mask,'Q2'].value_counts()

rate = [Q8_y_Q2['Yes']/float(starwars_Q8['Yes']+starwars_Q8['No']),
        Q8_n_Q2['Yes']/float(starwars_Q8['Yes']+starwars_Q8['No'])]

plt.bar(list(range(2)), rate, color='b', align='center')
plt.xticks(list(range(2)),['Expanded Universe Fans','Not Expanded Universe Fans'])
plt.xlabel('Star Wars fans')
plt.ylabel('Proportion of Star Wars Extended Universe fans')
plt.title('Star Wars Extended Universe fans by Normal Fans')
```

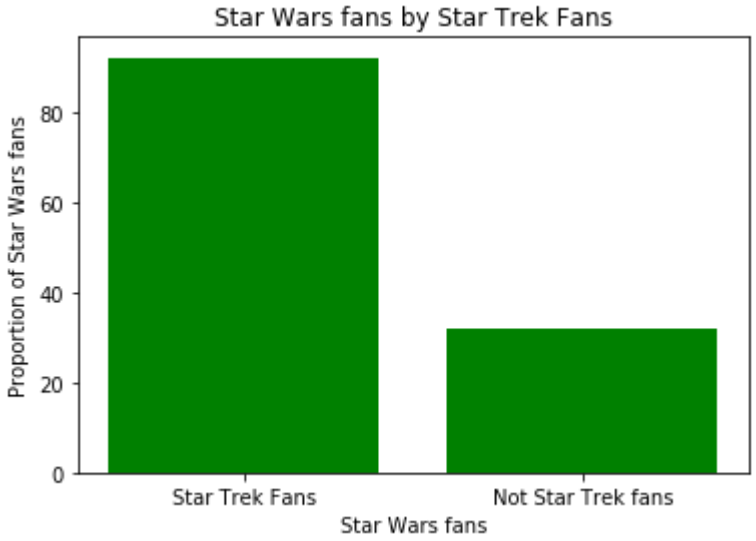
Out[39]: Text(0.5, 1.0, 'Star Wars Extended Universe fans by Normal Fans')



In [40]:
```
# Do you consider yourself to be a fan of the Star Wars film franchise? and Do you consider yourself to be a fan of the Star Trek franchise?
starwars_Q9 = starwars['Q9'].replace({"NotAnswered":np.nan}).dropna().value_counts()

Q9_y_mask = starwars['Q9'].replace({"NotAnswered":np.nan}).dropna() == 'Yes'
Q9_n_mask = starwars['Q9'].replace({"NotAnswered":np.nan}).dropna() == 'No'

Q9_y_Q2 = starwars.replace({"NotAnswered":np.nan}).dropna().loc[Q9_y_mask,'Q2'].value_counts()
Q9_n_Q2 = starwars.replace({"NotAnswered":np.nan}).dropna().loc[Q9_n_mask,'Q2'].value_counts()

print(Q9_y_Q2['Yes'])
print(Q9_n_Q2['Yes'])

#rate = [Q9_y_Q2['Yes']/float(starwars_Q9['Yes']+starwars_Q9['No']), Q9_n_Q2['Yes']/float(starwars_Q9['Yes']+starwars_Q9['No'])]

rate = [Q9_y_Q2['Yes'], Q9_n_Q2['Yes']]

plt.bar(list(range(2)), rate, color='g', align='center')
plt.xticks(list(range(2)),['Star Trek Fans','Not Star Trek fans'])
plt.xlabel('Star Wars fans')
plt.ylabel('Proportion of Star Wars fans')
plt.title('Star Wars fans by Star Trek Fans')
```

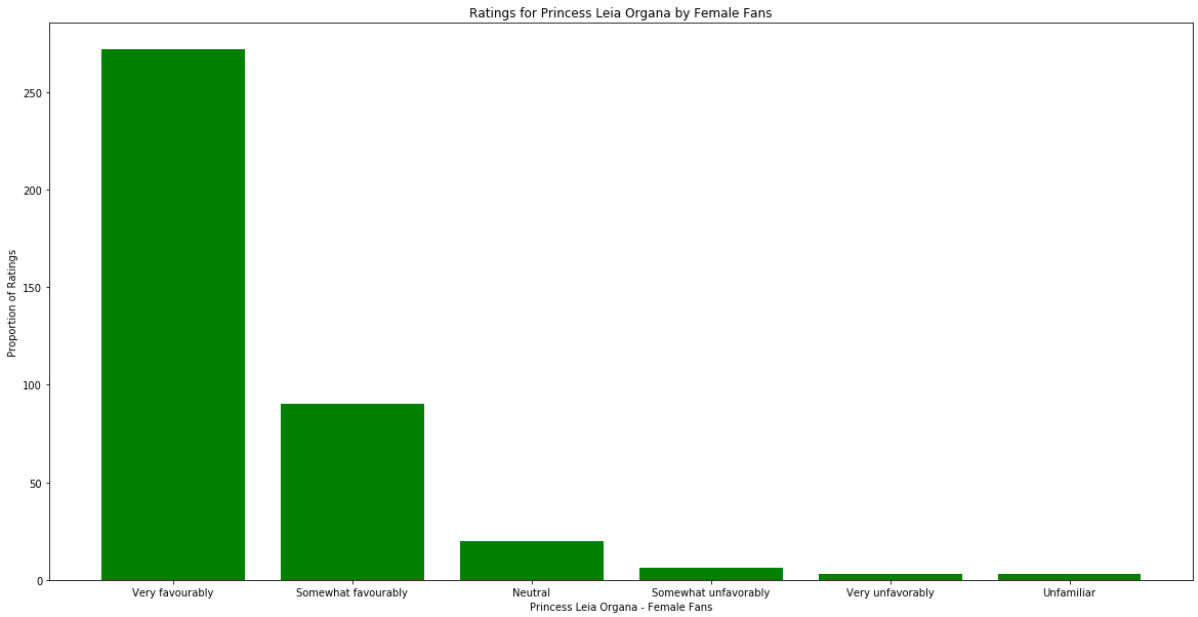
```
92
32
```

Out[40]: Text(0.5, 1.0, 'Star Wars fans by Star Trek Fans')



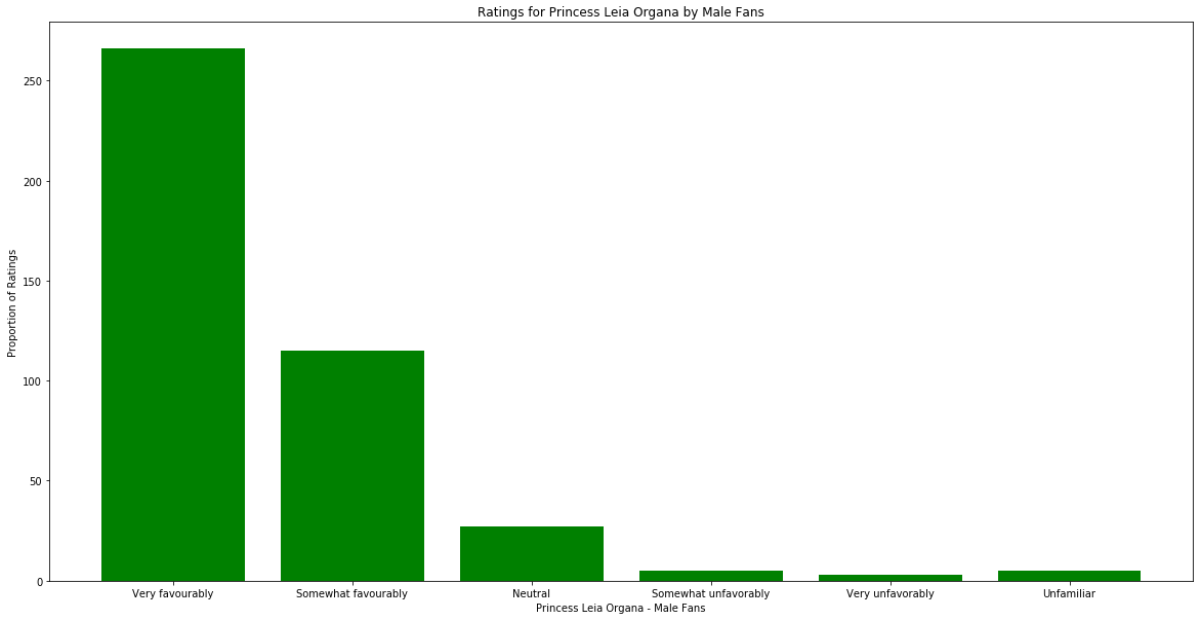
In [41]:
```
# Data Exploration
# Explore whether there are relationship between people's demographics (Gende
r, Age, Household Income, Education, Location) and their attitude to Start War
characters.


#Code goes after this line by adding cells

Leia_vf_mask = starwars["Q5-Princess Leia Organa"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Princess Leia Organa"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Princess Leia Organa"]== 'Neither favorably nor un
favorably (neutral)'
Leia_su_mask = starwars["Q5-Princess Leia Organa"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Princess Leia Organa"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Princess Leia Organa"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Female'], Leia_sf['Female'], Leia_nf['Female'], Leia_su['Fem
ale'], Leia_vu['Female'], Leia_uf['Female']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='g', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Princess Leia Organa - Female Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Princess Leia Organa by Female Fans')
```

Out[41]: Text(0.5, 1.0, 'Ratings for Princess Leia Organa by Female Fans')



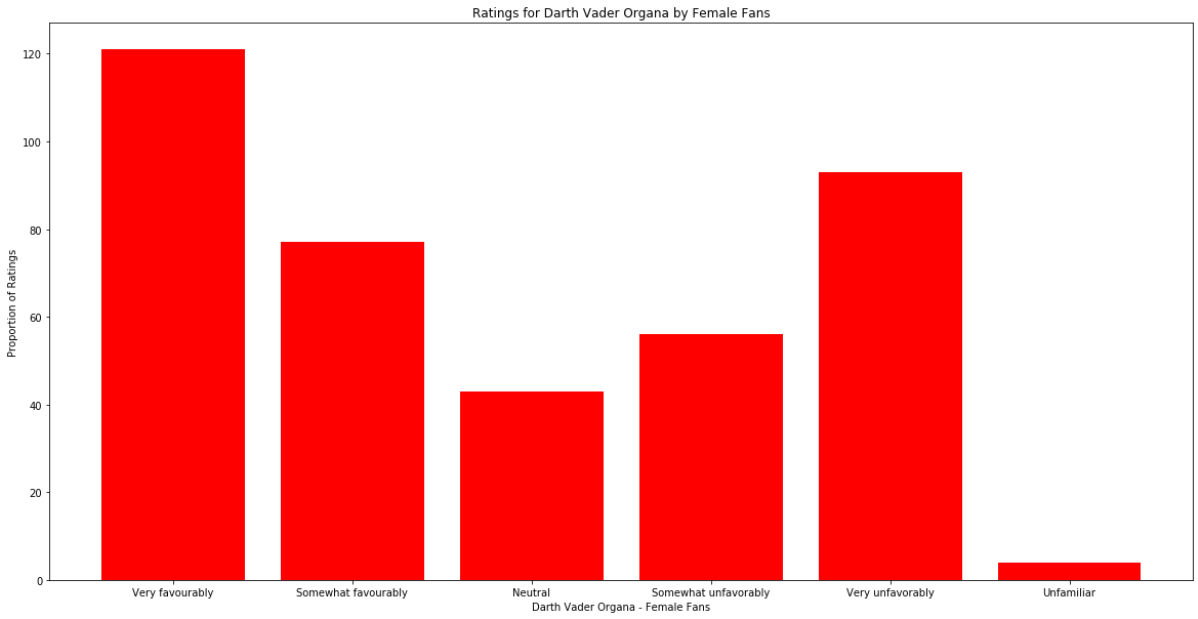
In [42]:
```
Leia_vf_mask = starwars["Q5-Princess Leia Organa"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Princess Leia Organa"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Princess Leia Organa"]== 'Neither favorably nor un
favorably (neutral)'
Leia_su_mask = starwars["Q5-Princess Leia Organa"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Princess Leia Organa"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Princess Leia Organa"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Male'], Leia_sf['Male'], Leia_nf['Male'], Leia_su['Male'], L
eia_vu['Male'], Leia_uf['Male']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='g', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Princess Leia Organa - Male Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Princess Leia Organa by Male Fans')
```

Out[42]: Text(0.5, 1.0, 'Ratings for Princess Leia Organa by Male Fans')



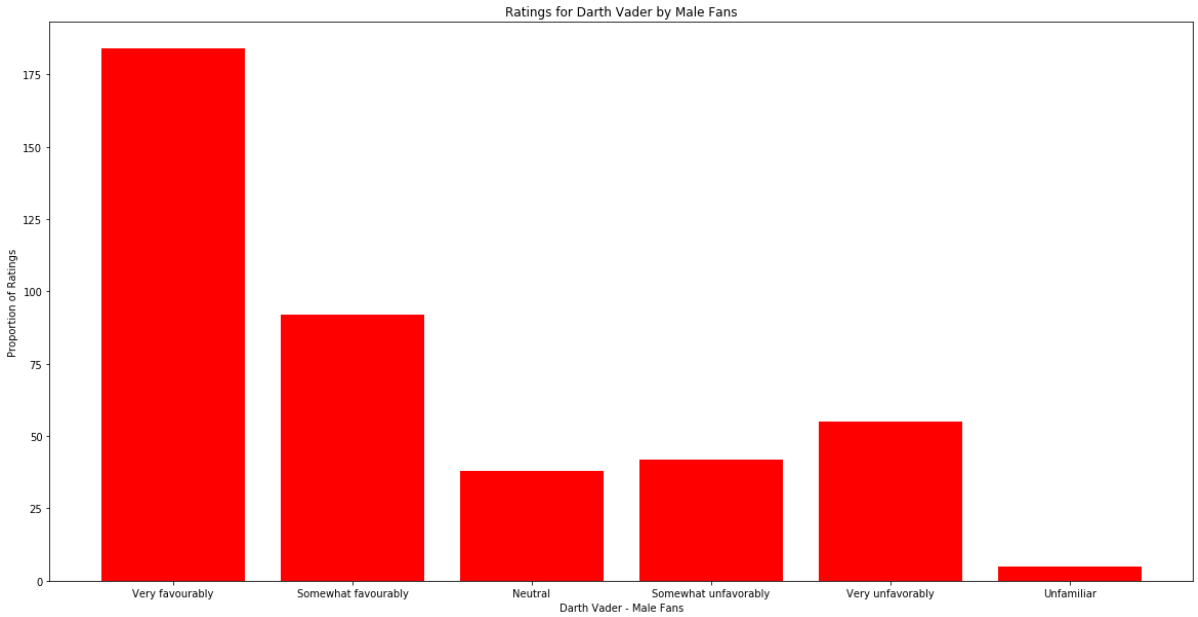
In [43]:
```
Leia_vf_mask = starwars["Q5-Darth Vader"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Darth Vader"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Darth Vader"]== 'Neither favorably nor unfavorably
(neutral)'
Leia_su_mask = starwars["Q5-Darth Vader"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Darth Vader"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Darth Vader"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Female'], Leia_sf['Female'], Leia_nf['Female'], Leia_su['Fem
ale'], Leia_vu['Female'], Leia_uf['Female']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='r', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Darth Vader Organa - Female Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Darth Vader Organa by Female Fans')
```

Out[43]: Text(0.5, 1.0, 'Ratings for Darth Vader Organa by Female Fans')



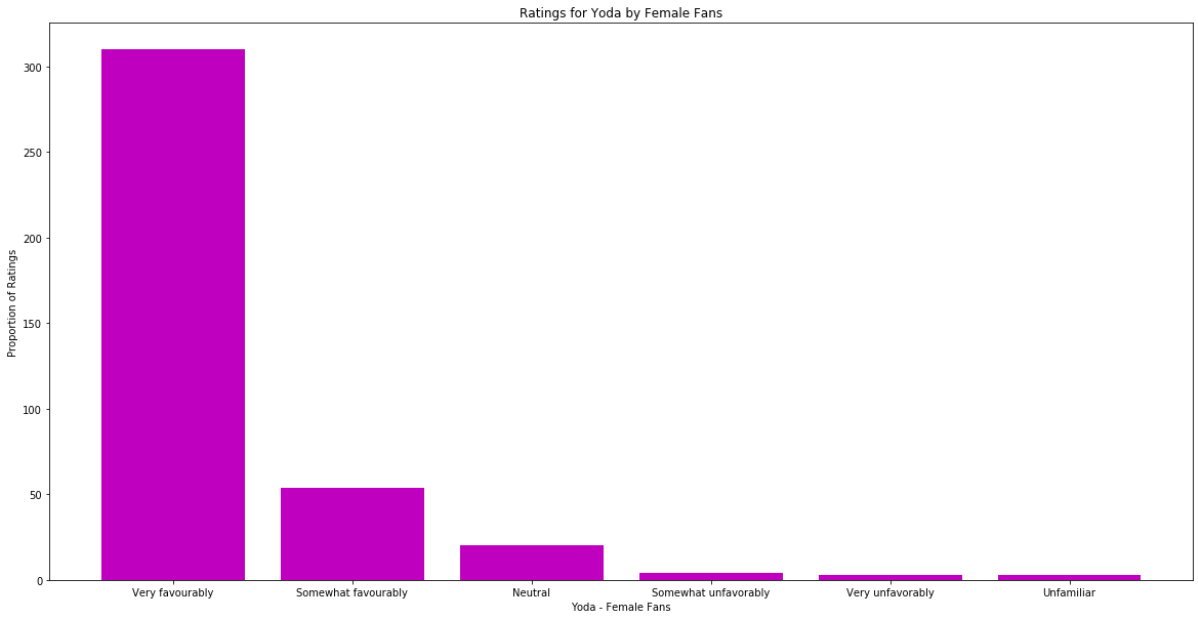
In [44]:
```
Leia_vf_mask = starwars["Q5-Darth Vader"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Darth Vader"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Darth Vader"]== 'Neither favorably nor unfavorably
(neutral)'
Leia_su_mask = starwars["Q5-Darth Vader"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Darth Vader"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Darth Vader"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Male'], Leia_sf['Male'], Leia_nf['Male'], Leia_su['Male'], L
eia_vu['Male'], Leia_uf['Male']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='r', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Darth Vader - Male Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Darth Vader by Male Fans')
```

Out[44]: Text(0.5, 1.0, 'Ratings for Darth Vader by Male Fans')



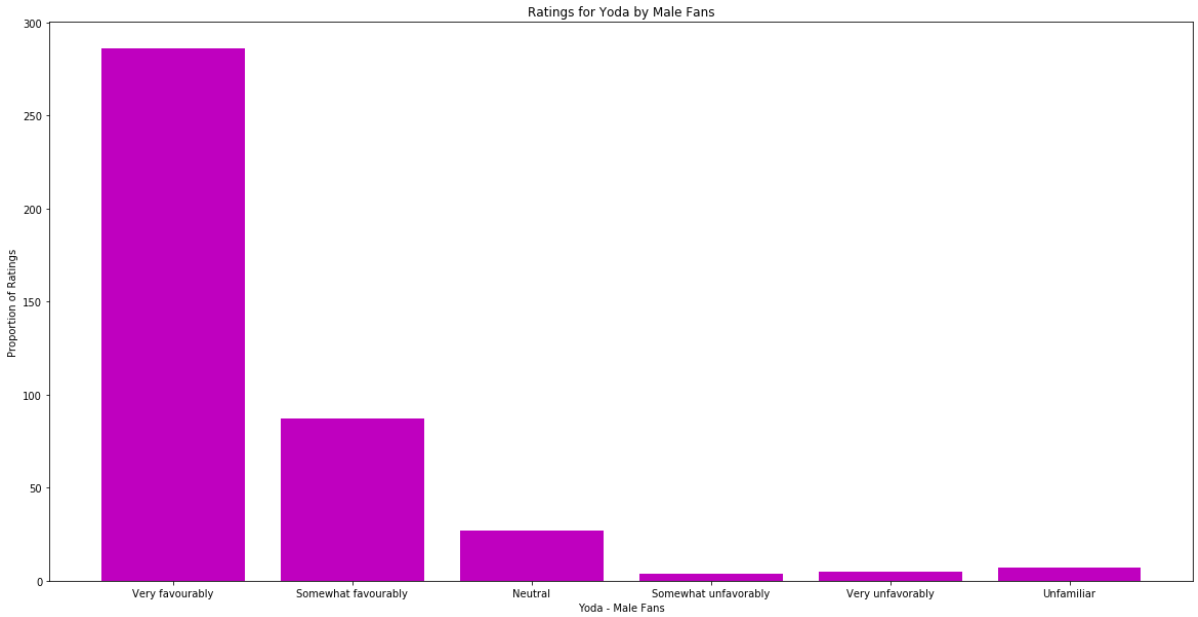
In [45]:
```
Leia_vf_mask = starwars["Q5-Yoda"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Yoda"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Yoda"]== 'Neither favorably nor unfavorably (neutr
al)'
Leia_su_mask = starwars["Q5-Yoda"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Yoda"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Yoda"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Female'], Leia_sf['Female'], Leia_nf['Female'], Leia_su['Fem
ale'], Leia_vu['Female'], Leia_uf['Female']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='m', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Yoda - Female Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Yoda by Female Fans')
```

Out[45]: Text(0.5, 1.0, 'Ratings for Yoda by Female Fans')



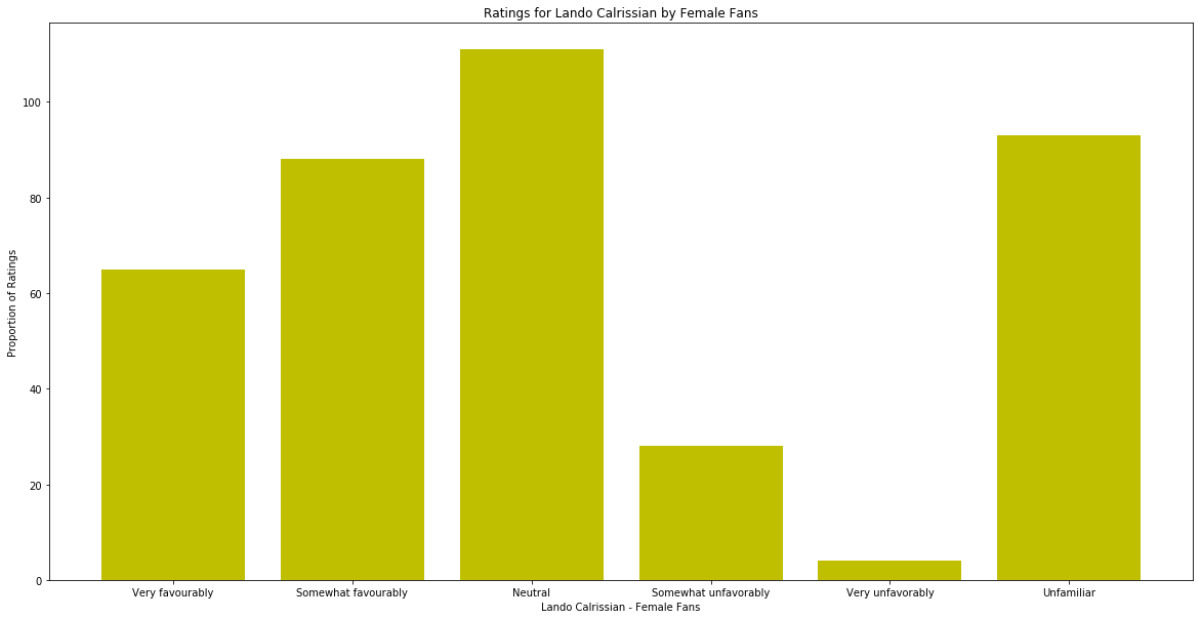
In [46]:
```
Leia_vf_mask = starwars["Q5-Yoda"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Yoda"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Yoda"]== 'Neither favorably nor unfavorably (neutr
al)'
Leia_su_mask = starwars["Q5-Yoda"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Yoda"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Yoda"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Male'], Leia_sf['Male'], Leia_nf['Male'], Leia_su['Male'], L
eia_vu['Male'], Leia_uf['Male']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='m', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Yoda - Male Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Yoda by Male Fans')
```

Out[46]: Text(0.5, 1.0, 'Ratings for Yoda by Male Fans')



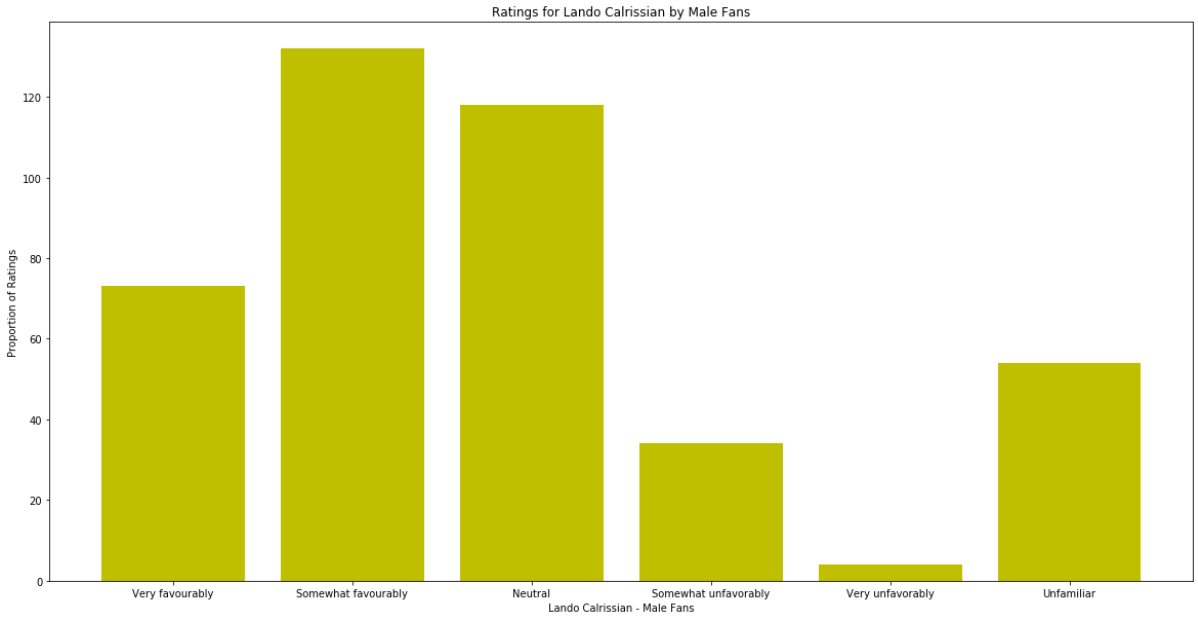
In [47]:
```
Leia_vf_mask = starwars["Q5-Lando Calrissian"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Lando Calrissian"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Lando Calrissian"]== 'Neither favorably nor unfavo
rably (neutral)'
Leia_su_mask = starwars["Q5-Lando Calrissian"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Lando Calrissian"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Lando Calrissian"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Female'], Leia_sf['Female'], Leia_nf['Female'], Leia_su['Fem
ale'], Leia_vu['Female'], Leia_uf['Female']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='y', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Lando Calrissian - Female Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Lando Calrissian by Female Fans')
```

Out[47]: Text(0.5, 1.0, 'Ratings for Lando Calrissian by Female Fans')



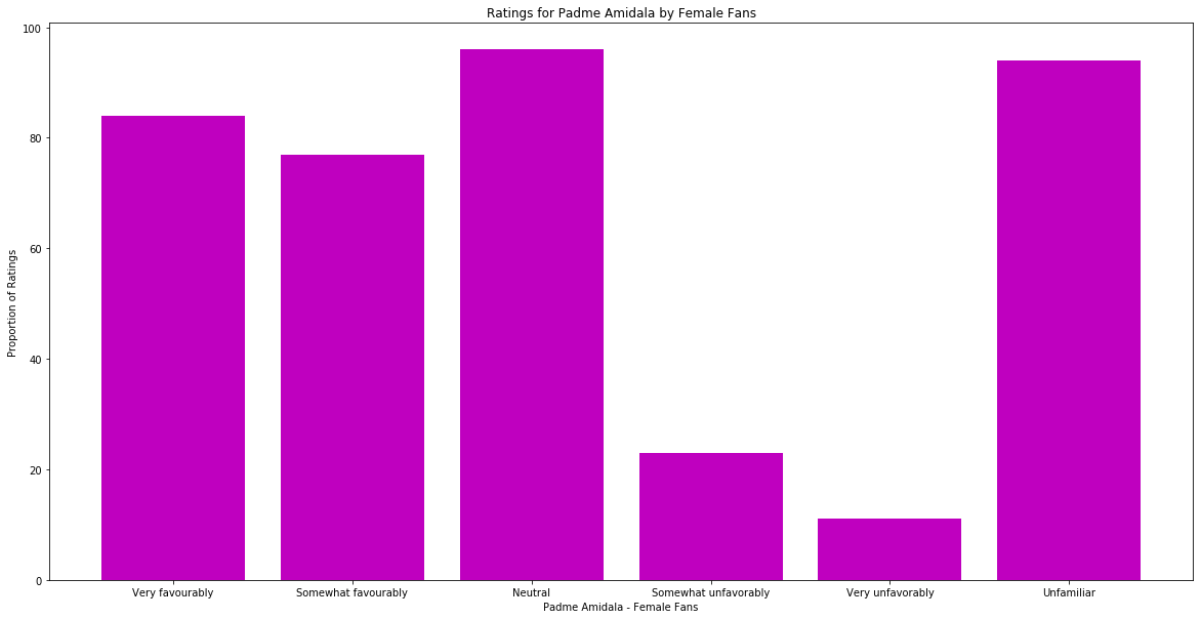
In [48]:
```
Leia_vf_mask = starwars["Q5-Lando Calrissian"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Lando Calrissian"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Lando Calrissian"]== 'Neither favorably nor unfavo
rably (neutral)'
Leia_su_mask = starwars["Q5-Lando Calrissian"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Lando Calrissian"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Lando Calrissian"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Male'], Leia_sf['Male'], Leia_nf['Male'], Leia_su['Male'], L
eia_vu['Male'], Leia_uf['Male']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='y', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Lando Calrissian - Male Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Lando Calrissian by Male Fans')
```

Out[48]: Text(0.5, 1.0, 'Ratings for Lando Calrissian by Male Fans')



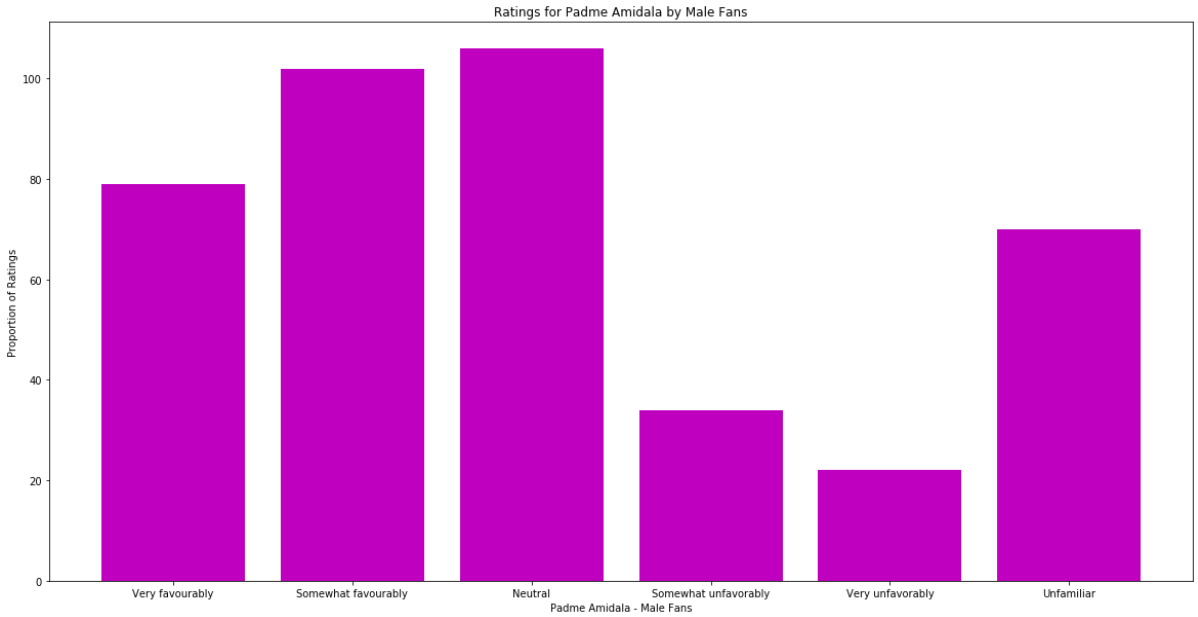
In [49]:
```
Leia_vf_mask = starwars["Q5-Padme Amidala"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Padme Amidala"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Padme Amidala"]== 'Neither favorably nor unfavorab
ly (neutral)'
Leia_su_mask = starwars["Q5-Padme Amidala"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Padme Amidala"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Padme Amidala"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Female'], Leia_sf['Female'], Leia_nf['Female'], Leia_su['Fem
ale'], Leia_vu['Female'], Leia_uf['Female']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='m', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Padme Amidala - Female Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Padme Amidala by Female Fans')
```

Out[49]: Text(0.5, 1.0, 'Ratings for Padme Amidala by Female Fans')



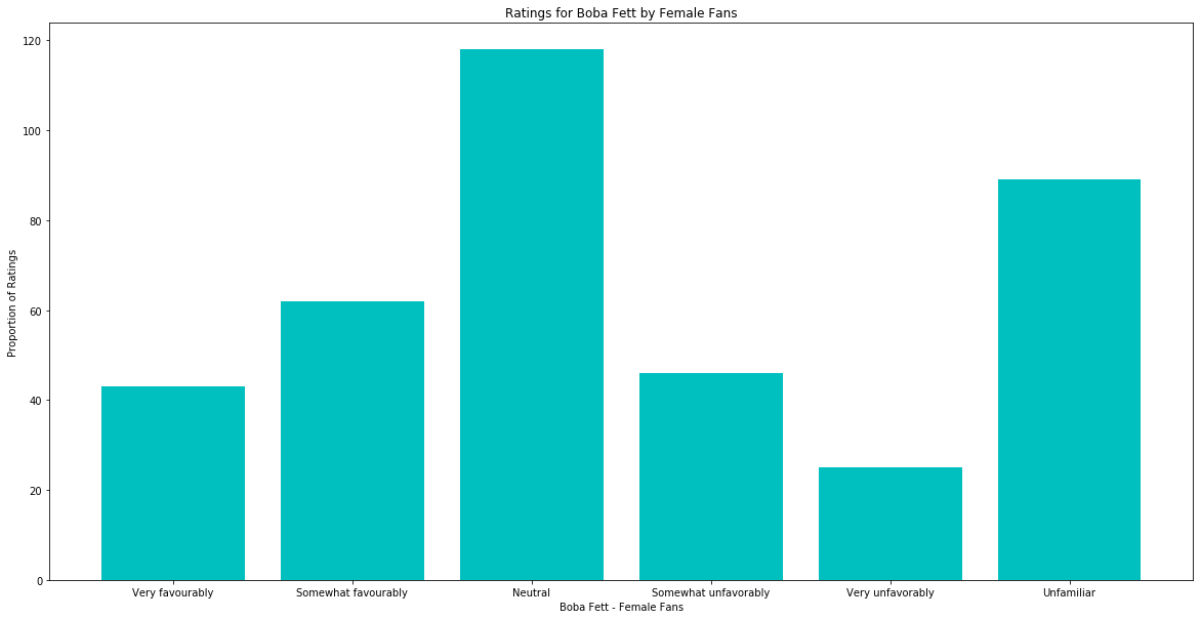
In [50]:
```
Leia_vf_mask = starwars["Q5-Padme Amidala"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Padme Amidala"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Padme Amidala"]== 'Neither favorably nor unfavorab
ly (neutral)'
Leia_su_mask = starwars["Q5-Padme Amidala"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Padme Amidala"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Padme Amidala"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Male'], Leia_sf['Male'], Leia_nf['Male'], Leia_su['Male'], L
eia_vu['Male'], Leia_uf['Male']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='m', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Padme Amidala - Male Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Padme Amidala by Male Fans')
```

Out[50]: Text(0.5, 1.0, 'Ratings for Padme Amidala by Male Fans')



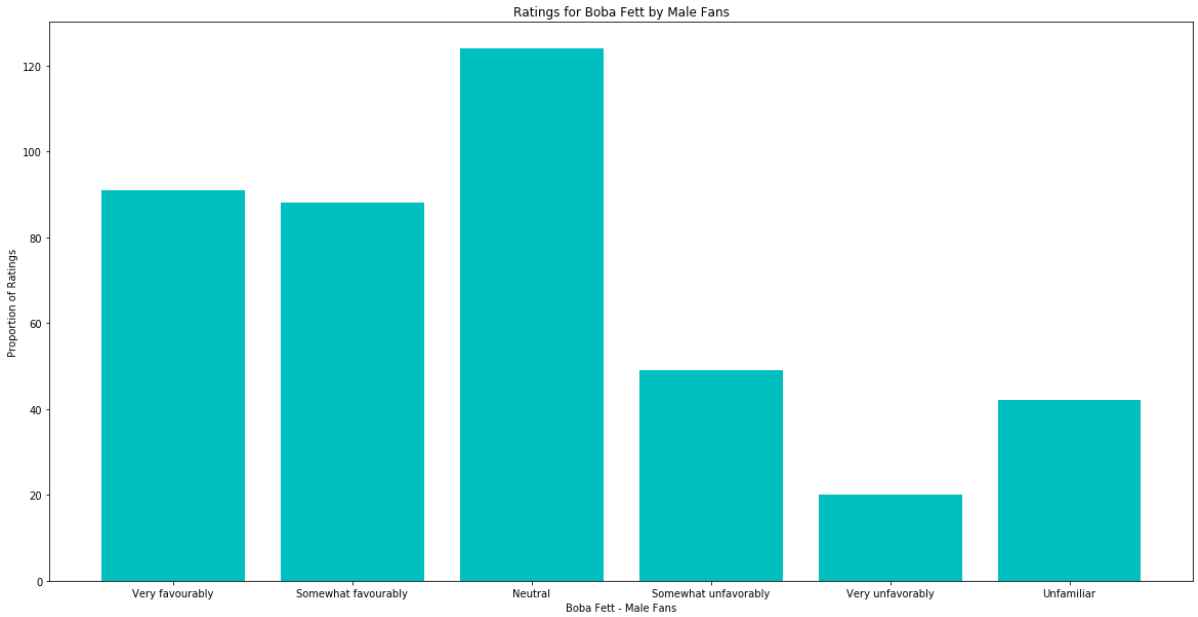
In [51]:
```
Leia_vf_mask = starwars["Q5-Boba Fett"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Boba Fett"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Boba Fett"]== 'Neither favorably nor unfavorably
 (neutral)'
Leia_su_mask = starwars["Q5-Boba Fett"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Boba Fett"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Boba Fett"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Female'], Leia_sf['Female'], Leia_nf['Female'], Leia_su['Fem
ale'], Leia_vu['Female'], Leia_uf['Female']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='c', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Boba Fett - Female Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Boba Fett by Female Fans')
```

Out[51]: Text(0.5, 1.0, 'Ratings for Boba Fett by Female Fans')



In [52]:
```
Leia_vf_mask = starwars["Q5-Boba Fett"]== 'Very favorably'
Leia_sf_mask = starwars["Q5-Boba Fett"]== 'Somewhat favorably'
Leia_nf_mask = starwars["Q5-Boba Fett"]== 'Neither favorably nor unfavorably
 (neutral)'
Leia_su_mask = starwars["Q5-Boba Fett"]== 'Somewhat unfavorably'
Leia_vu_mask = starwars["Q5-Boba Fett"]== 'Very unfavorably'
Leia_uf_mask = starwars["Q5-Boba Fett"]== 'Unfamiliar (N/A)'

Leia_vf = starwars.loc[Leia_vf_mask,'Gender'].value_counts()
Leia_sf = starwars.loc[Leia_sf_mask,'Gender'].value_counts()
Leia_nf = starwars.loc[Leia_nf_mask,'Gender'].value_counts()
Leia_su = starwars.loc[Leia_su_mask,'Gender'].value_counts()
Leia_vu = starwars.loc[Leia_vu_mask,'Gender'].value_counts()
Leia_uf = starwars.loc[Leia_uf_mask,'Gender'].value_counts()

rate2 = [Leia_vf['Male'], Leia_sf['Male'], Leia_nf['Male'], Leia_su['Male'], L
eia_vu['Male'], Leia_uf['Male']]

plt.figure(figsize=(20,10))
plt.bar(list(range(6)), rate2, color='c', align='center')
plt.xticks(list(range(6)),['Very favourably','Somewhat favourably','Neutral',
'Somewhat unfavorably','Very unfavorably','Unfamiliar'])
plt.xlabel('Boba Fett - Male Fans')
plt.ylabel('Proportion of Ratings')
plt.title('Ratings for Boba Fett by Male Fans')
```

Out[52]: Text(0.5, 1.0, 'Ratings for Boba Fett by Male Fans')



In [ ]:
```

```