

PROGRAM-11: DATA VISUALIZATION

REQUIREMENT:

With the help of suitable data and plots of your choice discuss how the data visualization can lead to misleading information.

IMPORTING LIBRARIES

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

ABOUT DATASET

This dataset consists of data if an advertisement will be success or not.

IMPORTING DATASET AND DISPLAYING IT

```
In [3]: media = pd.read_csv ('Media.csv')
actual = pd.read_csv ('Actual.csv')
```

```
In [4]: media.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6513 entries, 0 to 6512
Data columns (total 11 columns):
id                6513 non-null int64
realionship_status  6513 non-null object
Industry          6513 non-null object
genre             6513 non-null object
Targeted Sex      6513 non-null object
average_min_perweek  6513 non-null int64
Airtime           6513 non-null object
airlocation       6513 non-null object
ratings           6513 non-null float64
expensive         6513 non-null object
money_back_guarantee 6513 non-null object
dtypes: float64(1), int64(2), object(8)
memory usage: 559.8+ KB
```

```
In [5]: actual.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26048 entries, 0 to 26047
Data columns (total 12 columns):
id                26048 non-null int64
realionship_status  26048 non-null object
industry          26048 non-null object
genre             26048 non-null object
Targeted Sex      26048 non-null object
average_min_perweek  26048 non-null int64
Airtime           26048 non-null object
airlocation       26048 non-null object
ratings           26048 non-null float64
expensive         26048 non-null object
money_back_guarantee 26048 non-null object
netgain           26048 non-null bool
dtypes: bool(1), float64(1), int64(2), object(8)
memory usage: 2.2+ MB
```

```
In [6]: media=media.drop(['money_back_guarantee'],axis = 1)
media.head(7)
```

Out[6]:

	id	realtionship_status	Industry	genre	Targeted Sex	average_min_perweek	Airtime	airlocation	ratings	expensive
0	1	Widowed	Auto	Comedy	Male	10	Daytime	United-States	0.037465	High
1	4	Married-civ-spouse	Pharma	Comedy	Female	40	Morning	United-States	0.056262	Low
2	5	Divorced	Entertainment	Comedy	Male	50	Morning	United-States	0.037465	High
3	9	Married-civ-spouse	Pharma	Infomercial	Female	40	Primetime	United-States	0.037465	High
4	10	Married-civ-spouse	Pharma	Comedy	Female	40	Primetime	United-States	0.037465	High
5	20	Never-married	Entertainment	Comedy	Male	40	Primetime	United-States	0.037465	High
6	28	Divorced	Auto	Comedy	Female	40	Primetime	United-States	0.063196	Low

```
In [7]: actual=actual.drop(['money_back_guarantee', 'netgain'],axis = 1)
actual.head(7)
```

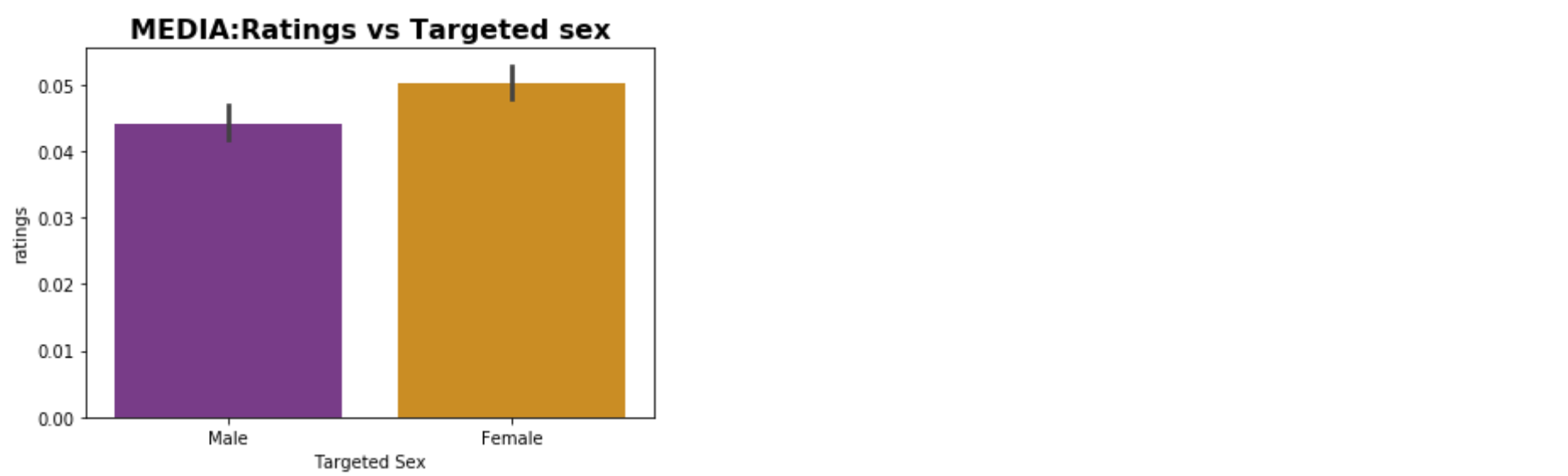
Out[7]:

	id	realtionship_status	industry	genre	Targeted Sex	average_min_perweek	Airtime	airlocation	ratings	expensive
0	19717	Married-spouse-absent	Auto	Comedy	Male	45	Primetime	United-States	0.027465	High
1	31593	Married-civ-spouse	Pharma	Comedy	Male	45	Primetime	United-States	0.027465	Low
2	5681	Divorced	Entertainment	Comedy	Female	45	Primetime	United-States	0.027465	High
3	15491	Separated	Political	Infomercial	Female	40	Primetime	United-States	0.027465	Low
4	23587	Married-civ-spouse	Pharma	Comedy	Male	48	Primetime	United-States	0.027465	High
5	28523	Divorced	Auto	Comedy	Female	40	Primetime	United-States	0.027465	Low
6	12290	Married-civ-spouse	Pharma	Infomercial	Male	50	Morning	Outlying-US(Guam-USVI-etc)	0.027465	High

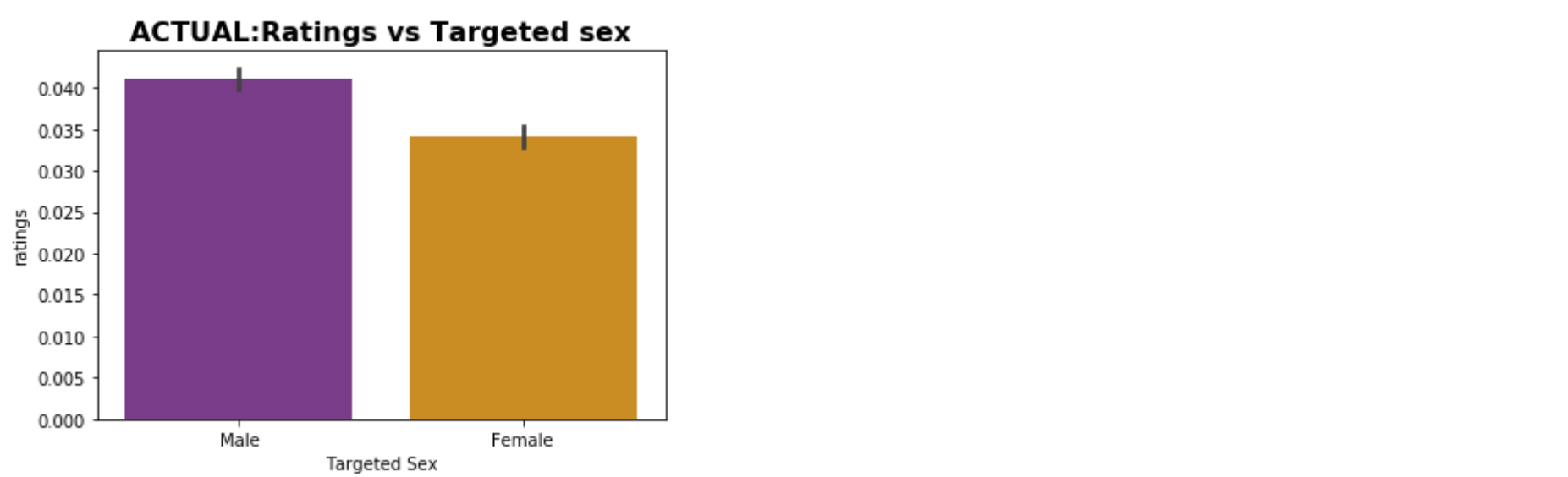
VISUALIZATION

1. RATINGS VS TARGET SEX

```
In [8]: ax = sns.barplot(x="Targeted Sex", y="ratings", palette="CMRmap",data=media).set_title('MEDIA:Ratings vs Targeted sex',w
```



```
In [9]: ax = sns.barplot(x="Targeted Sex", y="ratings", palette="CMRmap",data=actual).set_title('ACTUAL:Ratings vs Targeted sex'
```

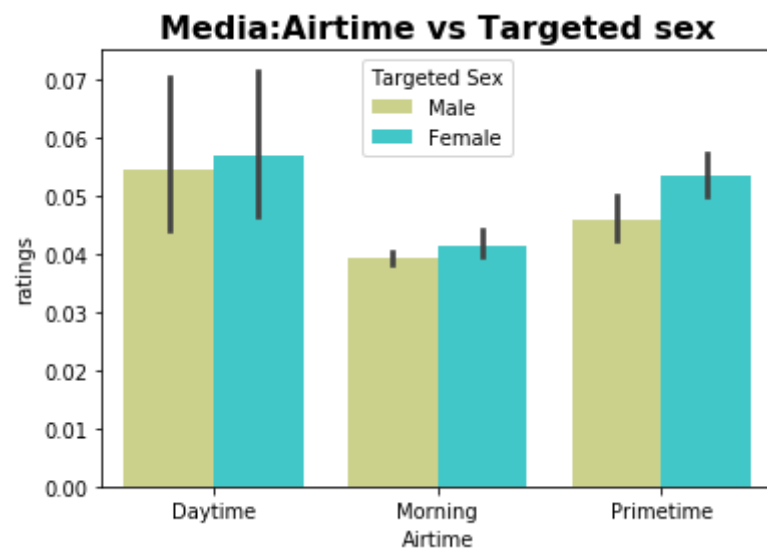


INFERENCE:

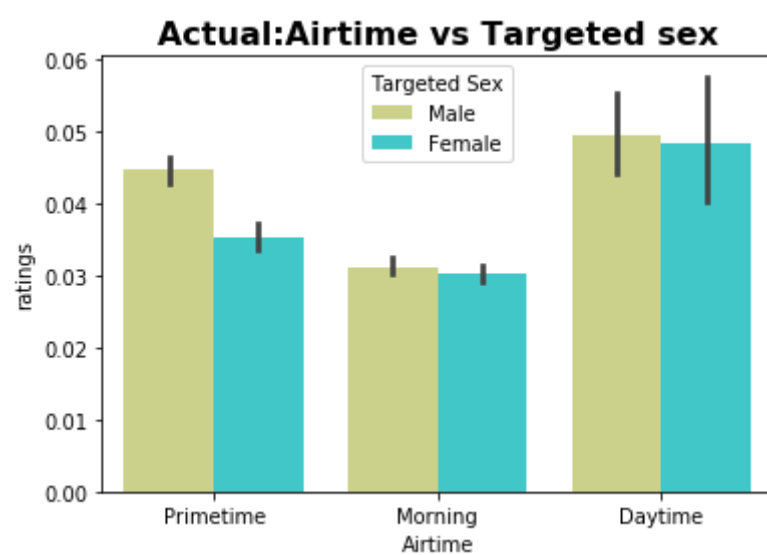
Comparing Media and Actual Dataset its clear that media has manipulated the dataset. In Actual, the target sex is Male but when we compare with Media its clear that target sex is female.

2. RATINGS vs AIRTIME

```
In [10]: ax = sns.barplot(x="Airtime", y="ratings", hue="Targeted Sex",palette="rainbow_r", data=media).set_title('Media:Airtime
```



```
In [11]: ax = sns.barplot(x="Airtime", y="ratings", hue="Targeted Sex", palette="rainbow_r",data=actual).set_title('Actual:Airtime
```

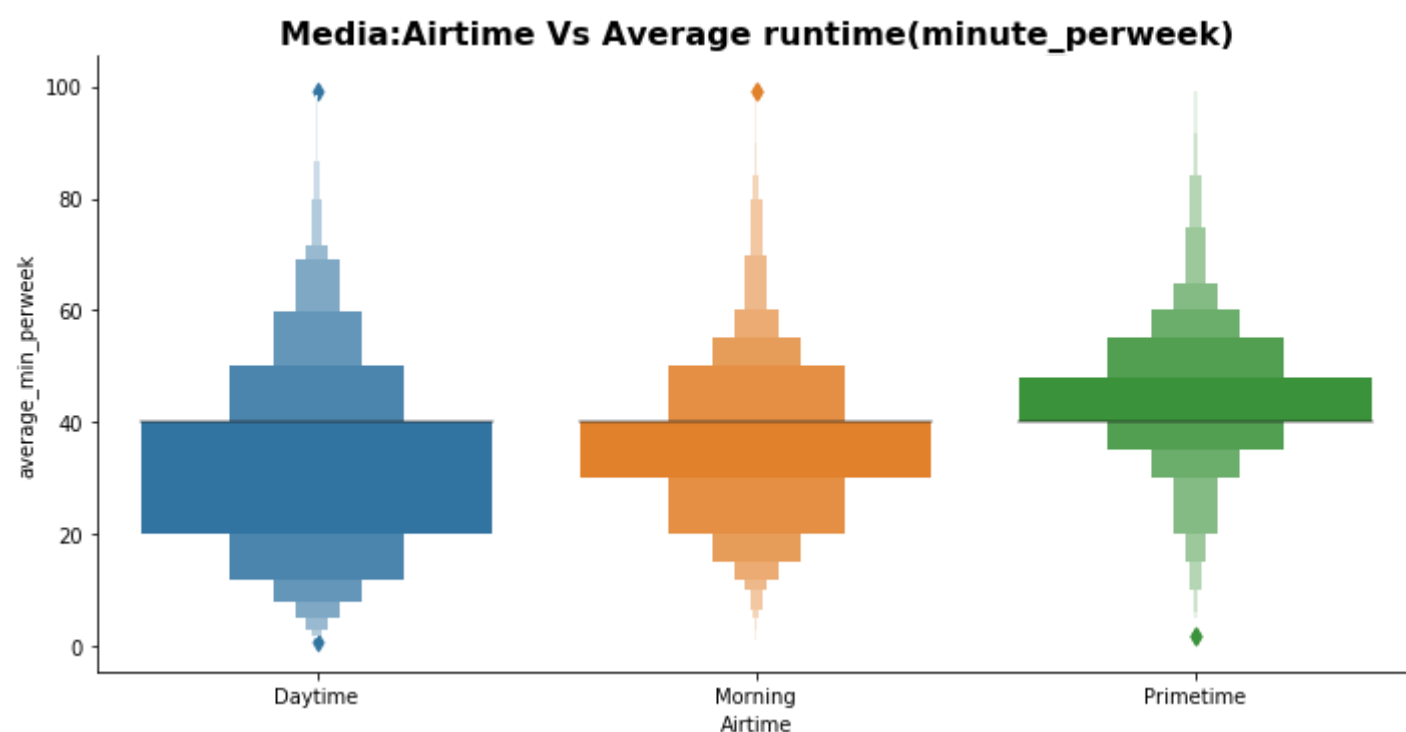


INFERENCE:

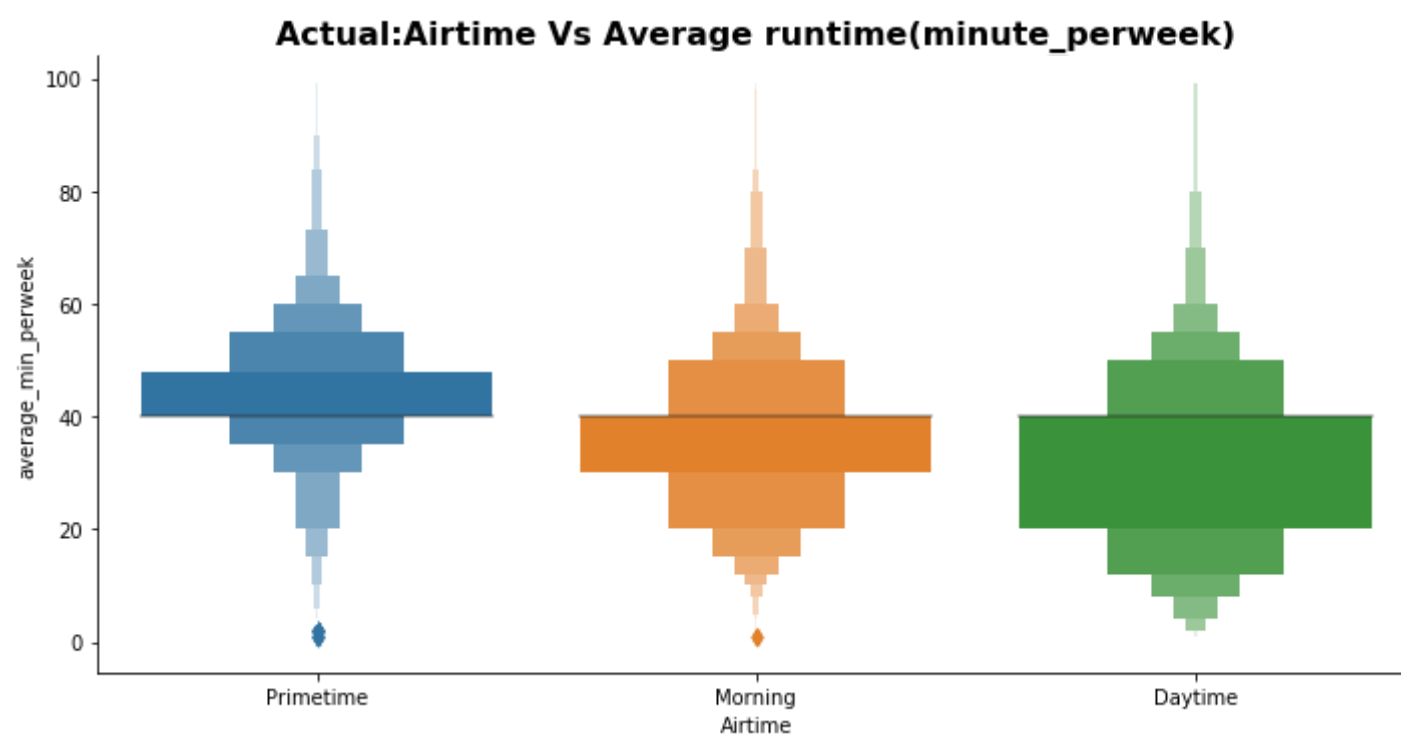
Comparing Media and Actual Dataset its clear that media has manipulated the dataset. In Actual, the target sex and airtime anlysis is direct opposite from Media.

3. AIRTIME VS AVERAGE RUNTIME(minute_perweek)

```
In [12]: sns.catplot(x='Airtime', y='average_min_perweek', data=media, kind='boxen', aspect=2)
plt.title('Media:Airtime Vs Average runtime(minute_perweek)', weight='bold', fontsize=16)
plt.show()
```



```
In [13]: sns.catplot(x='Airtime', y='average_min_perweek', data=actual, kind='boxen', aspect=2)
plt.title('Actual:Airtime Vs Average runtime(minute_perweek)', weight='bold', fontsize=16)
plt.show()
```

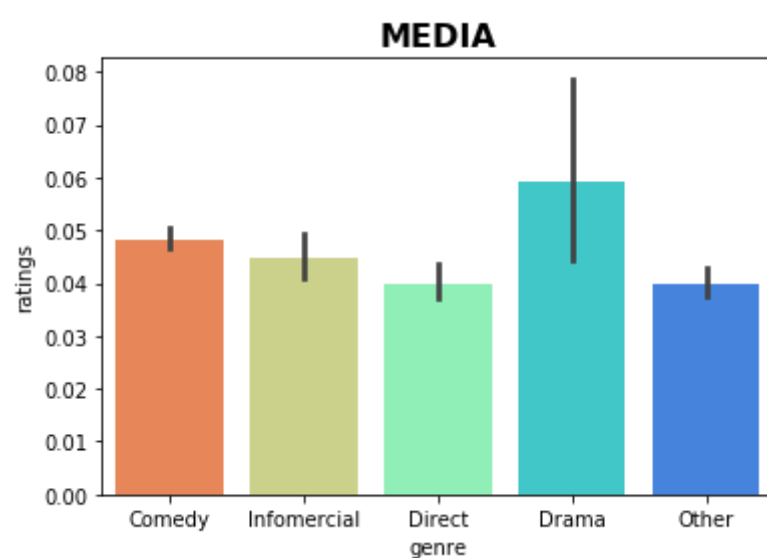


INFERENCE:

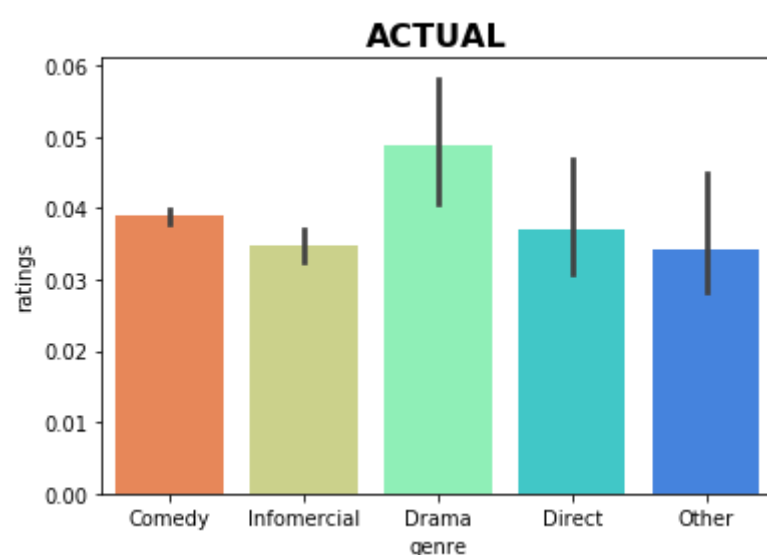
Compairing Media and Actual Dataset there is no manipulation done.

4. GENRE VS RATINGS

```
In [14]: ax = sns.barplot(x="genre", y="ratings", data=media,
                        palette="rainbow_r").set_title('MEDIA', weight='bold', fontsize=16)
```



```
In [15]: ax = sns.barplot(x="genre", y="ratings", data=actual,
                        palette="rainbow_r").set_title('ACTUAL', weight='bold', fontsize=16)
```

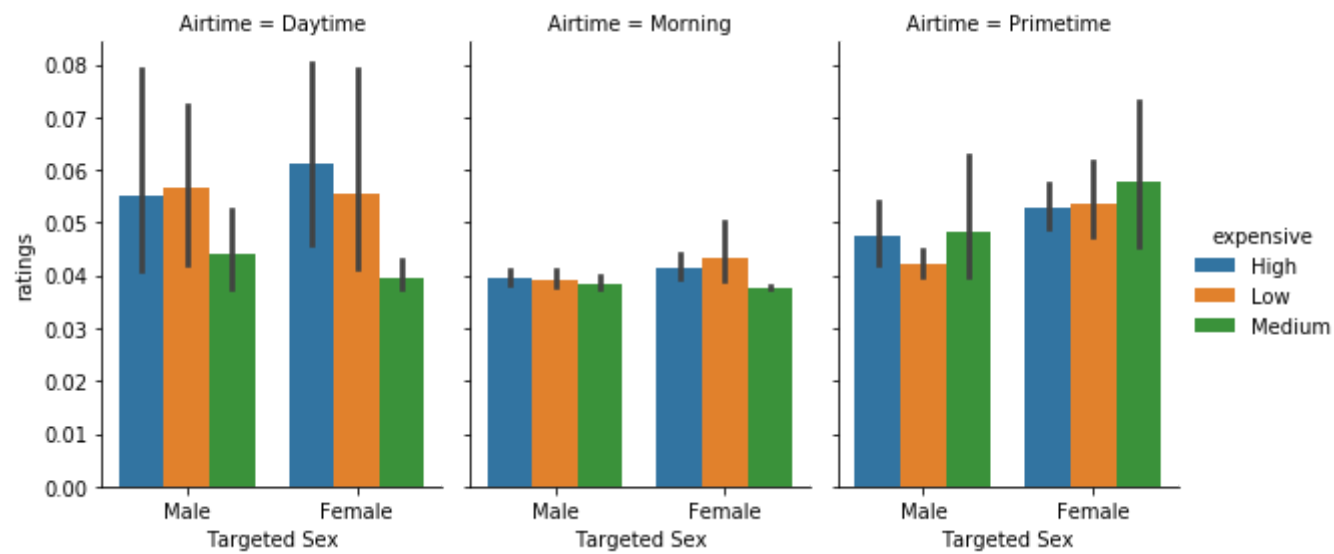


INFERENCE:

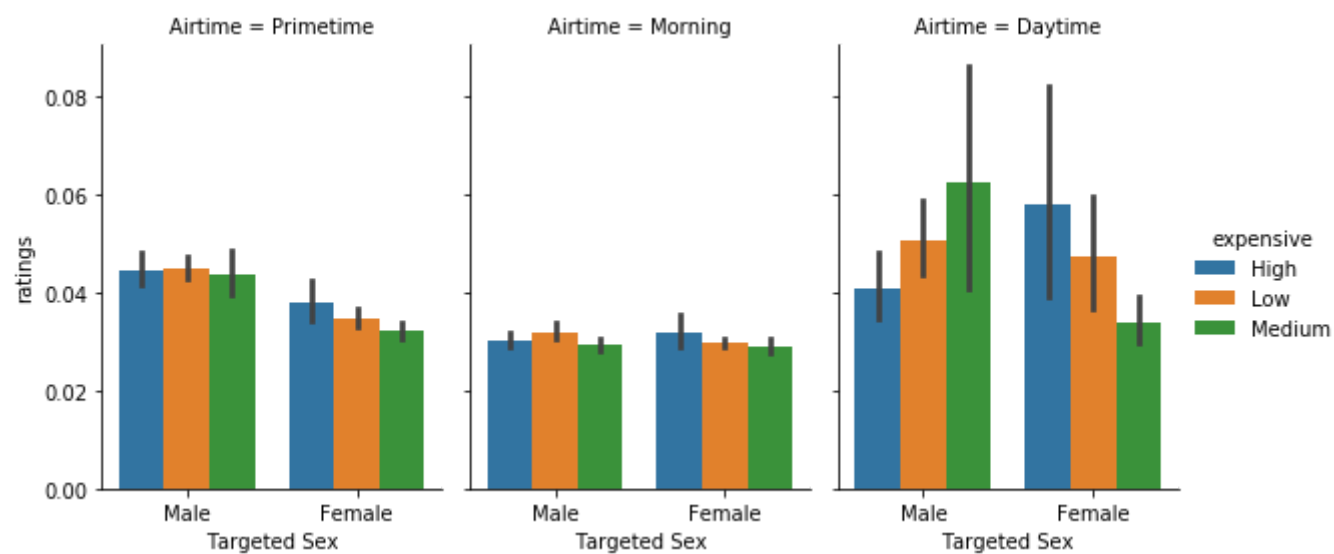
Compairing Media and Actual Dataset its clear that media has manipulated the dataset. The visualization shows slight diffence in genre between Actual and Media.

5. TARGETED SEX VS EXPENSE

```
In [16]: # Media
g=sns.catplot(x="Targeted Sex", y="ratings",
             hue="expensive", col="Airtime",
             data=media, kind="bar",
             height=4, aspect=.7);
```



```
In [17]: # Actual
g=sns.catplot(x="Targeted Sex", y="ratings",
             hue="expensive", col="Airtime",
             data=actual, kind="bar",
             height=4, aspect=.7);
```



INFERENCE:

Comparing Media and Actual Dataset its clear that media has manipulated the dataset. There is a drastic change in Expense vs targeted sex when compared between Actual and media.

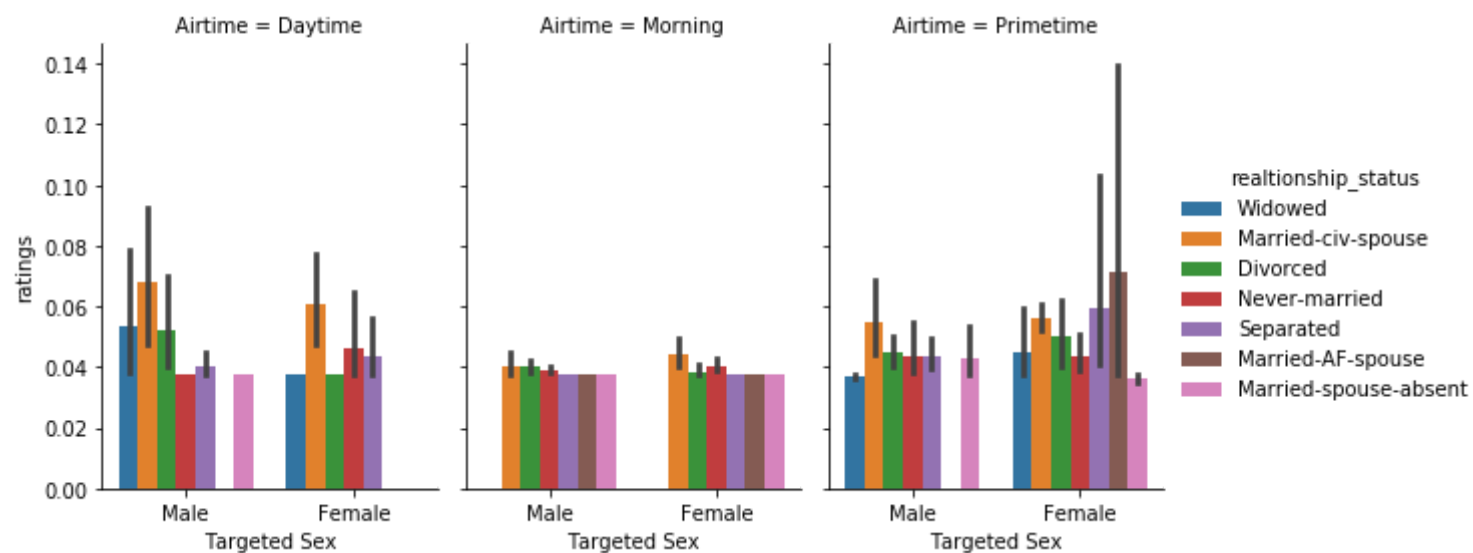
6. TARGETED SEX VS RELATIONSHIP STATUS

```
In [18]: g = sns.catplot(x="Targeted Sex", y="ratings",
                        hue="realtionship_status", col="Airtime",
                        data=media, kind="bar",
                        height=4, aspect=.7);

print("MEDIA")
g
```

MEDIA

Out[18]: <seaborn.axisgrid.FacetGrid at 0x21946c43c08>

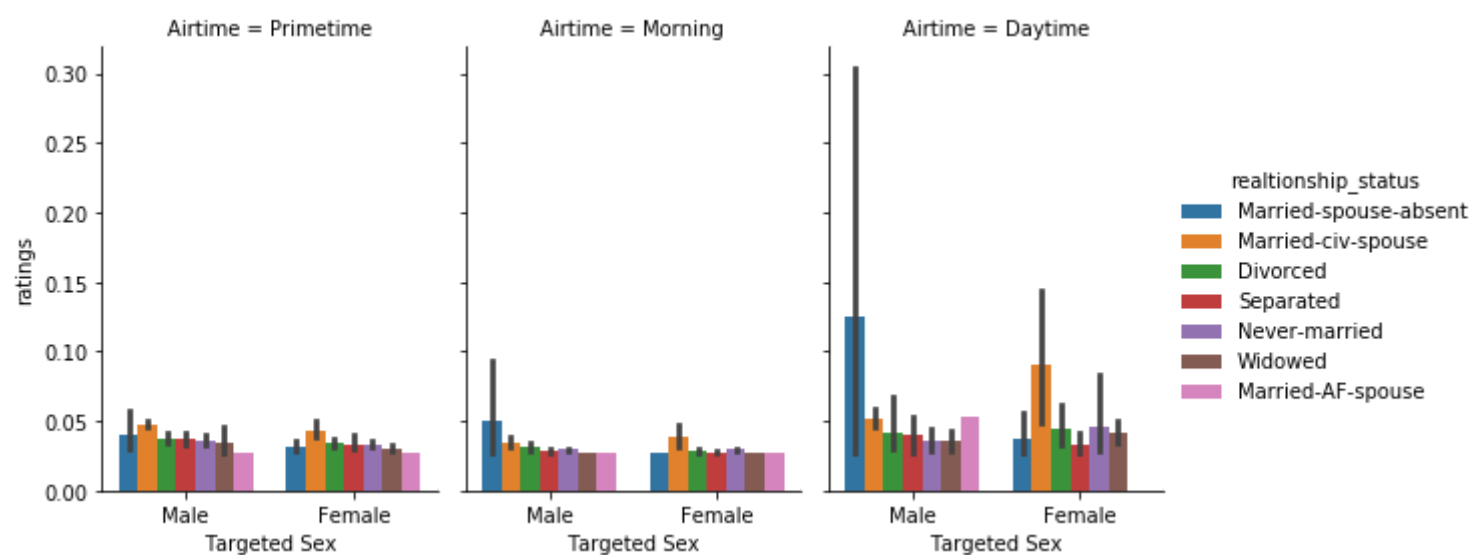


```
In [19]: g = sns.catplot(x="Targeted Sex", y="ratings",
                        hue="realtionship_status", col="Airtime",
                        data=actual, kind="bar",
                        height=4, aspect=.7);

print("ACTUAL")
g
```

ACTUAL

Out[19]: <seaborn.axisgrid.FacetGrid at 0x21947326288>



INFERENCE:

Compairing Media and Actual Dataset its clear that media has manipulated the dataset. There is a drastic change in Relationship status vs targeted sex when compared between Actual and media.