markdown

May 6, 2025

1 The Office

The Office is an American **mockumentary sitcom television** series based on the 2001–2003 BBC series The Office created by *Ricky Gervais* and *Stephen Merchant* and starring the former. Adapted for NBC by *Greg Daniels*, a veteran writer for **Saturday Night Live**, **King of the Hill**, and **The Simpsons**, the show depicts the everyday work lives of office employees at the Scranton, Pennsylvania, branch of the fictional Dunder Mifflin Paper Company, and aired from March 24, 2005, to May 16, 2013, with a total of nine seasons consisting of 201 episodes.[1] The show was co-produced by **Daniels' Deedle-Dee Productions**, **Reveille Productions** (later Shine America) and 3 Arts Entertainment (although uncredited) in association with Universal Television. The original executive producers were *Daniels*, *Gervais*, *Merchant*, *Howard Klein* and *Ben Silverman*, with numerous others being promoted in later seasons. V Like its British counterpart, the series was filmed in a single-camera setup without a studio audience or a laugh track, to mirror the look of an actual documentary. It debuted on NBC as a mid-season replacement and ended its nine-season run on May 16, 2013, with a two-part series finale. Its original main cast was *Steve Carell*, *Rainn Wilson*, *John Krasinski*, *Jenna Fischer*, and *B. J. Novak*. It experienced numerous changes to its ensemble cast during its run. Stars outside the original main cast include *Ed Helms*, *Rashida Jones*, *Amy Ryan*, *Mindy Kaling*, *Craig Robinson*, *James Spader*, *Ellie Kemper*, and *Catherine Tate*.

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
from IPython.display import Markdown, display

data = {
    "season": [1, 2, 3, 4, 5, 6, 7, 8, 9],
    "episodes": [6, 22, 25, 19, 28, 26, 26, 24, 25],
    "rank": [102, 67, 68, 77, 52, 52, 53, 87, 94],
    "viewership": [5.4, 8.0, 8.3, 8.9, 9.0, 7.8, 7.7, 6.5, 5.1]
}

overall_df = pd.DataFrame(data)

data3 = {
    "overall_episode": [
        29, 30, 31, 32, 33, 34, 35, 36, 37, 38,
        39, 40, 41, 42, 43, 44, 45, 46, 47, 48,
        49, 50, 51, 52, 53
    ],
    "episode": [
        1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
        11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
        21, 22, 23, 24, 25
    ],
    "viewership": [
        9.11, 7.78, 8.89, 8.83, 8.46, 8.81, 8.05, 8.43, 9.07, 8.44,
        9.0, 8.80, 10.15, 9.32, 10.01, 8.84, 8.84, 8.25, 6.74, 7.71,
        7.56, 6.99, 7.17, 7.88, 7.98
    ]
}

df_season3 = pd.DataFrame(data3)
df_season3.loc[:, 'season'] = 3


data4 = {
    "overall_episode": [
        54, 55, 56, 57, 58, 59, 60, 61, 62, 63,
        64, 65, 66, 67, 68, 69, 70, 71, 72
    ],
    "episode": [
        1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
        11, 12, 13, 14, 15, 16, 17, 18, 19
    ],
    "viewership": [
        9.65, 9.0, 8.57, 9.1, 8.87, 8.9, 8.61, 8.5, 8.96, 8.36,
        8.27, 8.80, 9.33, 9.86, 7.69, 7.75, 7.16, 8.21, 8.3
```

```
    ]
}

df_season4 = pd.DataFrame(data4)
df_season4.loc[:, 'season'] = 4

df_merged = pd.concat([df_season3, df_season4], axis=1)
```

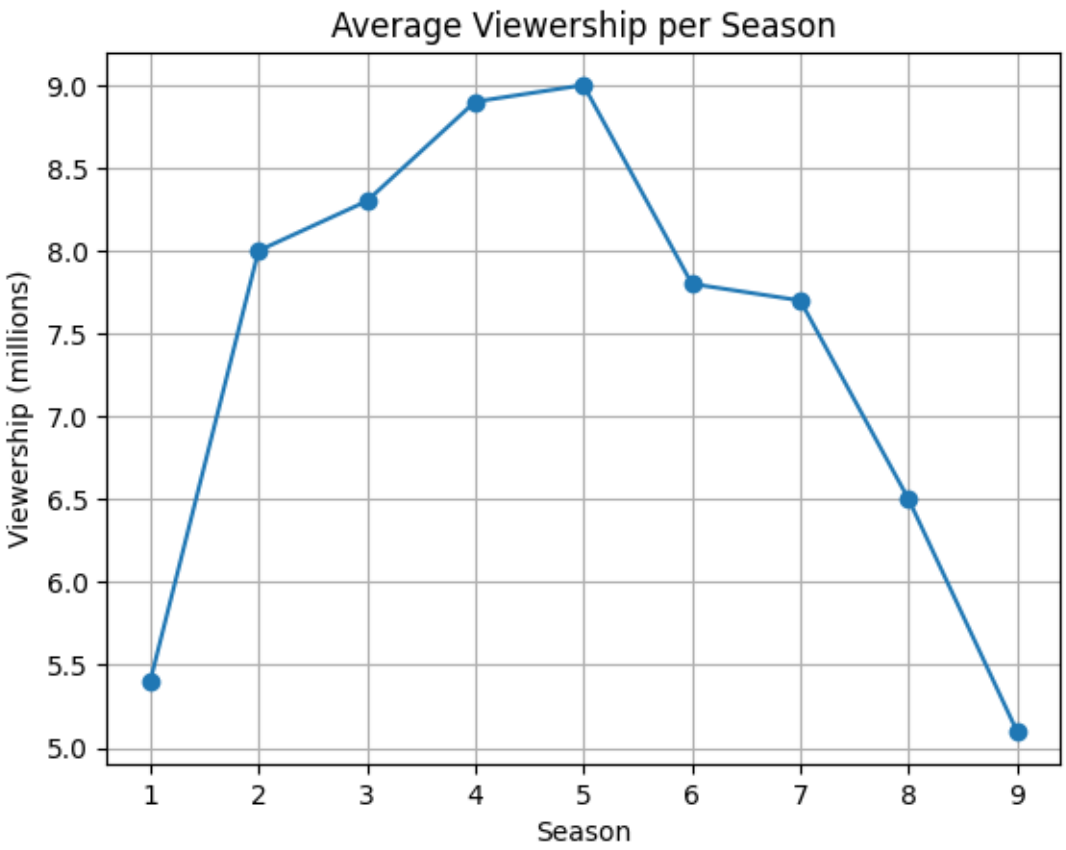
1.1 Summary over all seasons

[2]:
```
overall_df.drop(['episodes'], axis=1, inplace=False).describe()
```

[2]:

	season	rank	viewership
count	9.000000	9.000000	9.000000
mean	5.000000	72.444444	7.411111
std	2.738613	18.822268	1.428675
min	1.000000	52.000000	5.100000
25%	3.000000	53.000000	6.500000
50%	5.000000	68.000000	7.800000
75%	7.000000	87.000000	8.300000
max	9.000000	102.000000	9.000000

[3]:
```
plt.figure()
plt.plot(overall_df['season'], overall_df['viewership'], marker='o')
plt.title('Average Viewership per Season')
plt.xlabel('Season')
plt.ylabel('Viewership (millions)')
plt.grid(True)
```

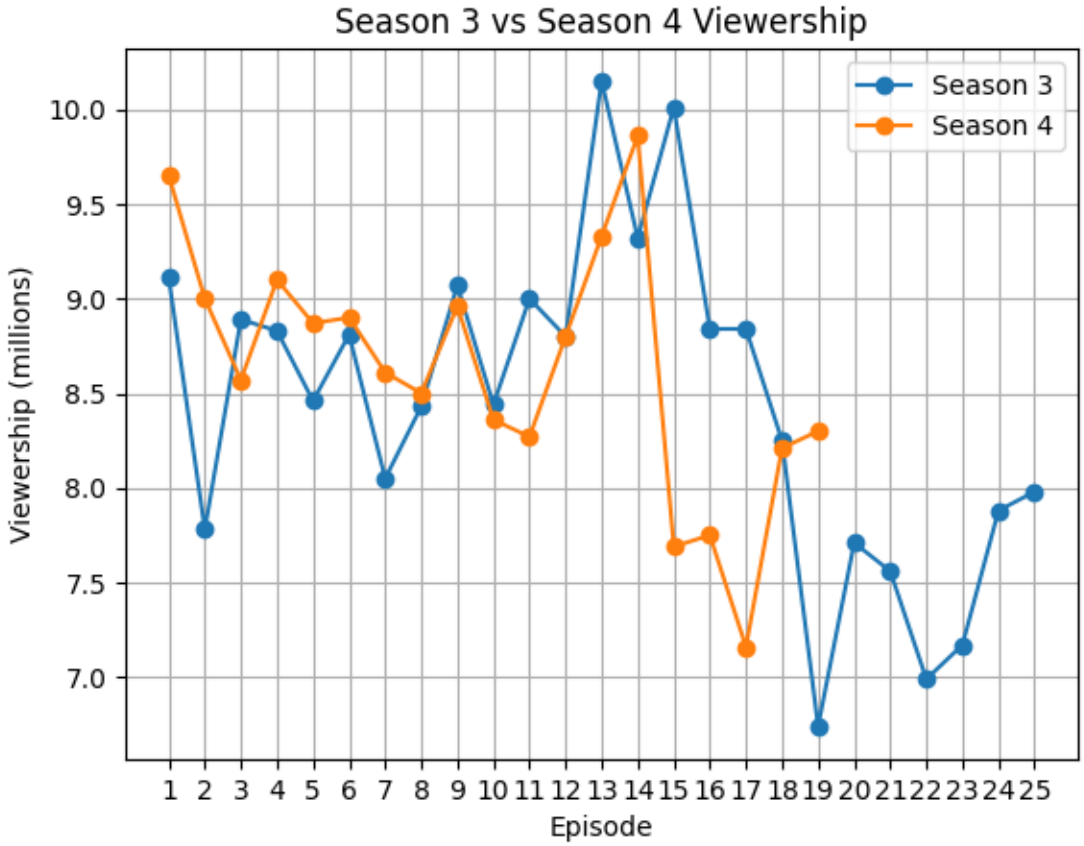


1.2 Comparison of Season 3 vs Season 4

[4]:
```
s3_mean = df_season3['viewership'].mean()
s4_mean = df_season4['viewership'].mean()

plt.figure()
plt.plot(df_season3['episode'],
         df_season3['viewership'],
         marker='o',
         label='Season 3')
plt.plot(df_season4['episode'],
         df_season4['viewership'],
         marker='o',
         label='Season 4')
plt.title('Season 3 vs Season 4 Viewership')
plt.xlabel('Episode')
plt.ylabel('Viewership (millions)')
plt.grid(True)
plt.xticks(range(1, max(df_season3['episode'].max(),
```

```
                        df_season4['episode'].max())+1))
plt.legend()
plt.show()
```



[5]:
```
display(Markdown(f"""
## Season 3 vs Season 4 Analysis
overall season 4 had marginally better viewership compared to season 3
- average viewership for season 3 = `{s3_mean}`
- average viewership for season 4 = `{s4_mean}`
"""))
```

1.3 Season 3 vs Season 4 Analysis

overall season 4 had marginally better viewership compared to season 3 - average viewership for season 3 = `8.4444` - average viewership for season 4 = `8.62578947368421`

[ ]:
```

```