ONE PIECE EDA

1. Import Libraries
2. Load the Dataset
3. EDA
4. Visualization

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

```
data = pd.read_csv('/content/OnePieceArcs.csv')
```

```
data.head()
```

	Arc	Start onChapter	TotalChapters	TotalPages	Manga%	Start onEpisode	TotalEpisodes	TotalMi
0	Romance Dawn Arc	1	7	178	0.9%	1	3	
1	Orange Town Arc	8	14	273	1.4%	4	5	
2	Syrup Village	22	20	396	2.0%	9	10	

```
data.tail()
```

	Arc	Start onChapter	TotalChapters	TotalPages	Manga%	Start onEpisode	TotalEpisodes	TotalMi
46	Levely Arc	903	6	100	0.5%	878	12	
47	Wano Country Arc: Act 1	909	16	278	1.4%	890	26	

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 9 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   Arc                   51 non-null     object
 1   Start onChapter       51 non-null     int64 
 2   TotalChapters         51 non-null     int64 
 3   TotalPages            51 non-null     int64 
 4   Manga%                51 non-null     object
 5   Start onEpisode       51 non-null     int64 
 6   TotalEpisodes         51 non-null     int64 
 7   TotalMinutes(avg 24)  51 non-null     int64 
 8   Anime%                51 non-null     object
dtypes: int64(6), object(3)
memory usage: 3.7+ KB
```

```
max_chap = data[data['TotalEpisodes'] == data['TotalEpisodes'].max()]
max_chap
```

	Arc	Start onChapter	TotalChapters	TotalPages	Manga%	Start onEpisode	TotalEpisodes	TotalM
	Dressrosa							

```
data['Manga Percentage'] = data['Manga%'].apply(lambda x: float(x[:-1]))
data['Anime Percentage'] = data['Anime%'].apply(lambda x: float(x[:-1]))
```

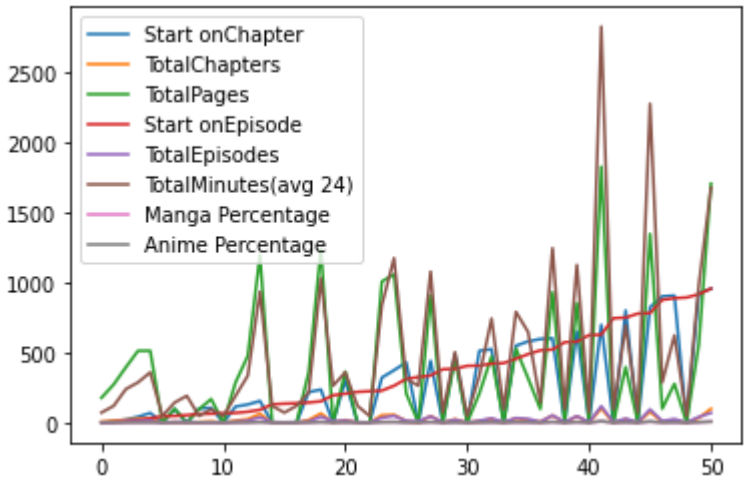
```
data.drop(['Manga%', 'Anime%'], axis = 1, inplace = True)
data.head()
```

	Arc	Start onChapter	TotalChapters	TotalPages	Start onEpisode	TotalEpisodes	TotalMinutes(av 24
0	Romance Dawn Arc	1	7	178	1	3	7
1	Orange Town Arc	8	14	273	4	5	12
2	Syrup Village	22	20	396	9	10	24

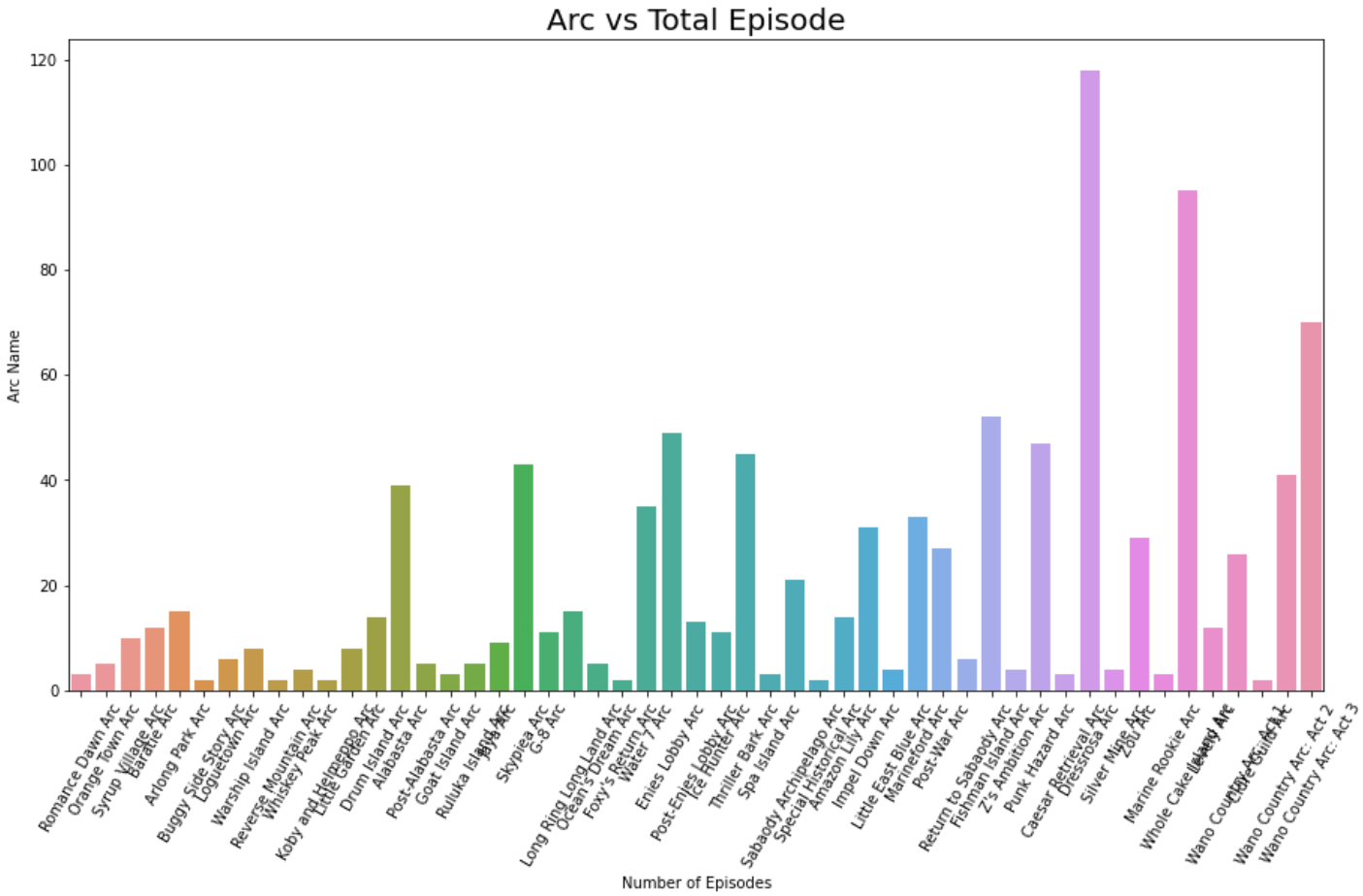
```
filler = data[data['Manga Percentage'] == 0]
filler
```

	Arc	Start onChapter	TotalChapters	TotalPages	Start onEpisode	TotalEpisodes	TotalMinutes(a
5	Buggy Side Story Arc	0	0	0	46	2	
7	Warship Island Arc	0	0	0	54	8	
10	Koby and Helmeppo Arc	0	0	0	68	2	

```
data.plot()
plt.rcParams['figure.figsize']=(15,8)
```

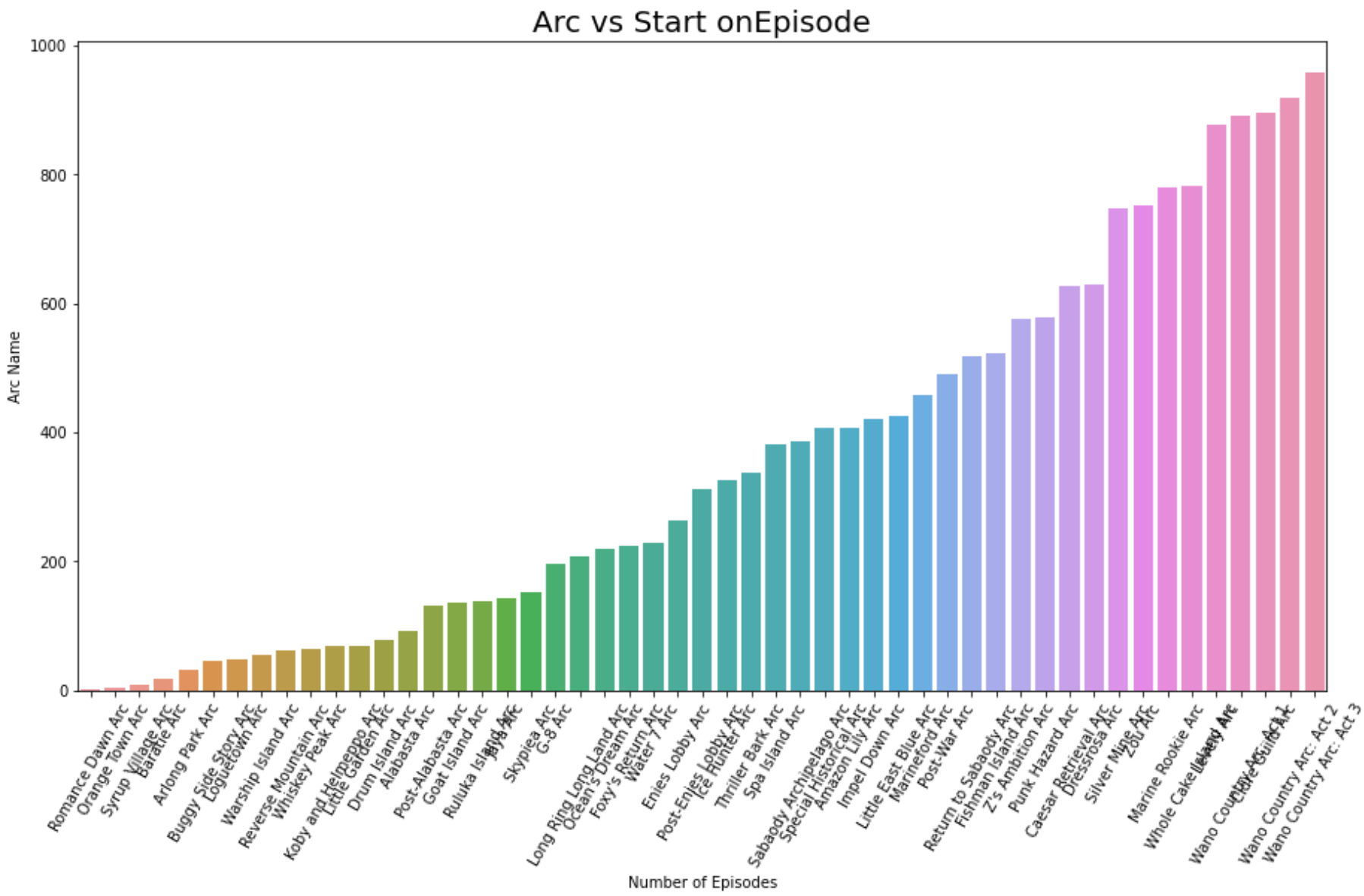



```
sns.barplot(x='Arc', y='TotalEpisodes', data=data)
plt.rcParams['figure.figsize']=(15,8)
plt.xticks(rotation=60)
plt.xlabel("Number of Episodes")
plt.ylabel("Arc Name")
plt.title("Arc vs Total Episode",fontsize=20)
plt.show()
```



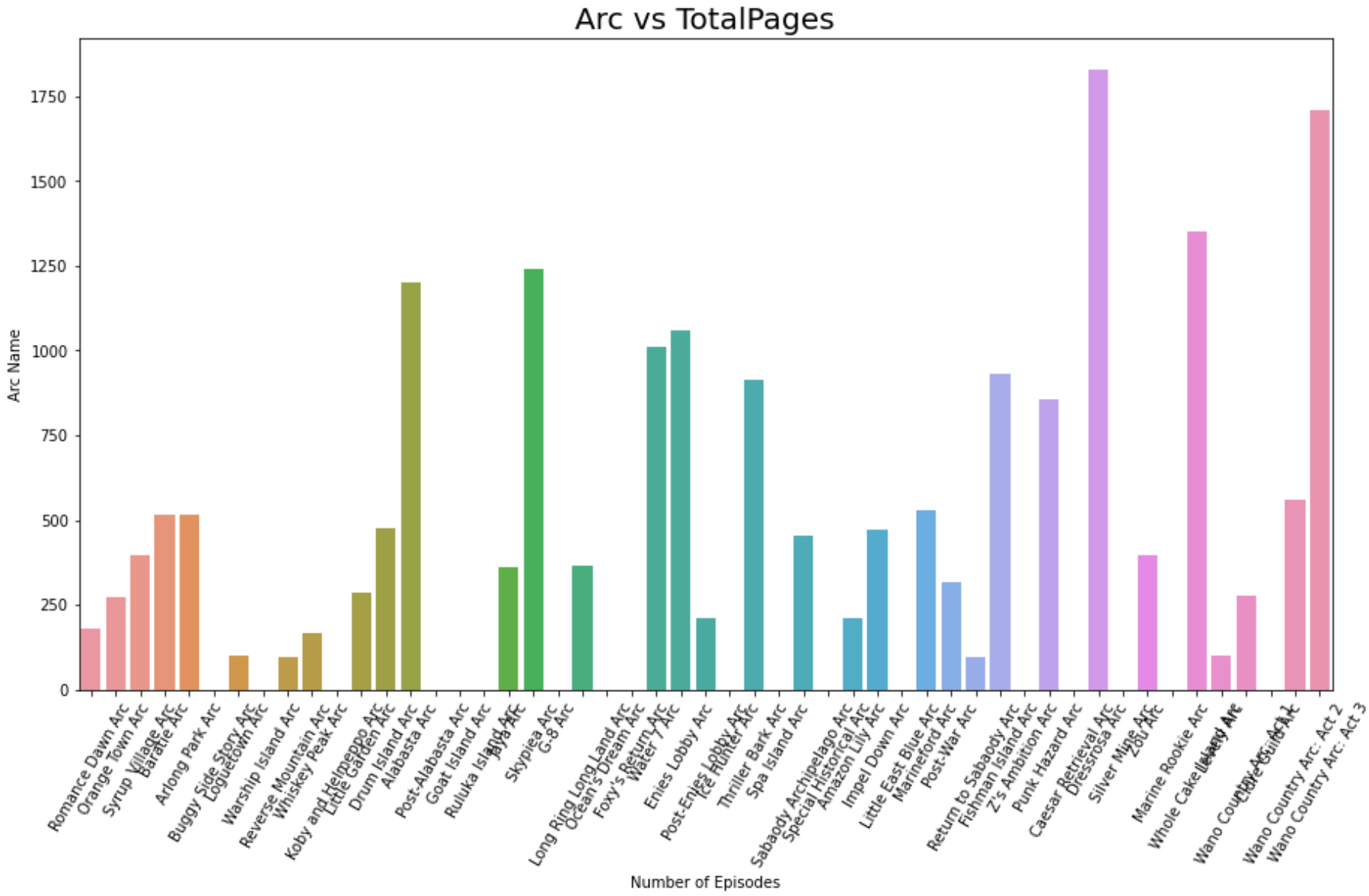

```
sns.barplot(x='Arc', y='Start onEpisode', data=data)
plt.rcParams['figure.figsize']=(15,8)
plt.xticks(rotation=60)
plt.xlabel("Number of Episodes")
plt.ylabel("Arc Name")
plt.title("Arc vs Start onEpisode",fontsize=20)
```

```
plt.show()
```




```
sns.barplot(x='Arc', y='TotalPages', data=data)
plt.rcParams['figure.figsize']=(15,8)
plt.xticks(rotation=60)
```

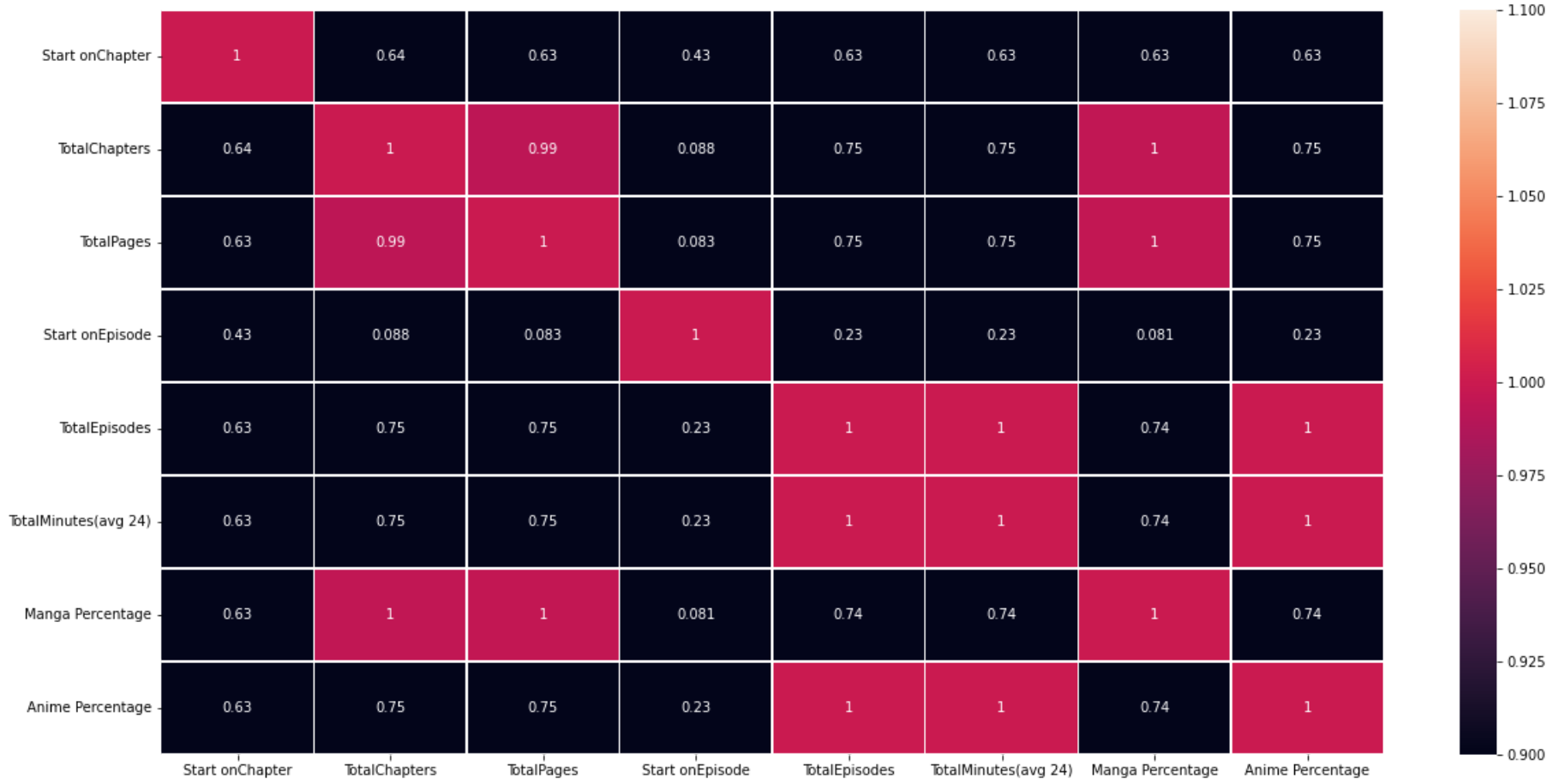
```
plt.xlabel("Number of Episodes")
plt.ylabel("Arc Name")
plt.title("Arc vs TotalPages",fontsize=20)
plt.show()
```




```
corr = data.corr('kendall')
```

```
plt.figure(figsize=(20,10))
sns.heatmap(corr, vmin=1, vmax=1, linewidth= 0.5, annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f80933b1b10>
```



✓ 0s completed at 14:02