

MICROSOFT MOVIES ANALYSIS

- Student name: Frida Oyucho
- Student pace: part time
- Instructor name: William Okombo/Noah Kandie

Business Understanding

Microsoft is exploring top-performing genres in the current box office landscape for their new movie studio. This analysis aims to provide actionable insights on successful film trends to guide strategic decisions and ensure the creation of content that resonates with audiences, maximizing the studio's chances of success in the competitive entertainment industry.

Research Questions:

- What is the average duration of a movie?
- Which are the top 10 movie genres?
- Which movie genres have the highest revenue?
- Which is the highest rated movie genre?
- What is the correlation between movies and revenue variables collected?

Data Understanding

The datasets used in this analysis were from:

- tmdb database
- tn.movie_budgets database
- title.basics database

The selected datasets from the listed databases provide information needed about movies and the different genres i.e runtime, Production Budget, movie popularity, worldwide gross e.t.c

```
<!-- Importing necessary libraries -->
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

```
<!-- Importing csv files and converting them to dataframes -->
df1 = pd.read_csv('ZippedData/tmdb.movies.csv',index_col=0)
df2 = pd.read_csv('ZippedData/title.basics.csv')
df3 = pd.read_csv('ZippedData/tn.movie_budgets.csv')
```

```
<!-- Preview of dataset1 -->  
df1.head()
```

```
<!-- Preview of dataset2 -->  
df2.head()
```

```
<!-- Preview of dataset3 -->  
df3.head()
```

```
<!-- Joining DataSets( df1, df2 & df3) -->  
merged_df = pd.merge(df1,df2)  
  
joined_df = merged_df.merge(df3,left_index=True, right_index=True,how =  
'outer')  
joined_df.head()
```

```
<!-- Preview of the joined Dataset -->  
joined_df.shape
```

```
<!-- Exporting Data Set to CSV -->  
joined_df.to_csv('FinalMergedData.csv')
```