

# Project: Investigate a Dataset - TMDb movie data

## Table of Contents

- [Introduction](#)
- [Data Wrangling](#)
- [Exploratory Data Analysis](#)
- [Conclusions](#)

## Introduction

### Dataset Description

Using the Movie Database(TMDb) with data from over 10,000 movies I will analyze the data to uncover trends among different movies and genres. This analysis will explore their popularity, runtimes, genres, and budget, as well as the correlations among these factors.

### Question(s) for Analysis

**Question 1:** What are the top 10 favorite movies?

**Question 2:** What year was the highest budget movie produce?

**Question 3:** What is the correlation between the average runtime of movies and the passage of time?

```
In [2]: # Import packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

## Data Wrangling

### General Properties

```
In [3]: #Load data
df = pd.read_csv('tmdb-movies.csv')

#show first 3 rows
```