**TMDB\_Analysis\_Python\_Project**

**Overview**

This project analyzes the [Kaggle Dataset Full TMDB Movies Dataset 2024 (1M Movies),](https://www.kaggle.com/datasets/asaniczka/tmdb-movies-dataset-2023-930k-movies) which contains movie data pulled from the TMDB website API. The analysis focuses on what makes a successful film through various features.

**Introduction**

My name is Jorge Martinez and I am studying Data Science with Machine Learning at New York City Data Science Academy. This project demonstrates my skills in:

* Working with datasets
* Creating Python visualizations
* Using GitHub
* Communicating with data

**Project Focus**

I examined what makes a successful film by analyzing:

* Financial features
* Geographic features
* Runtime
* Genre

**Background**

While IMDB (Internet Movie Database) is widely discussed, I chose to analyze the less-examined TMDB (The Movie Database) website. TMDB operates more like a non-profit with:

* Individual users uploading and updating movie entries
* Moderators overseeing the platform

This crowdsourced approach creates some data limitations, particularly regarding budget and revenue information. Users who uploaded movie entries may have lacked this data or weren't required to provide it due to minimal upload requirements.

**Resources**

For my visualizations, I referenced the [Python Glossary](https://python-graph-gallery.com/) extensively.