**Data Visualization Mini-Project**

**Section 1: Select topic and data**

When I was in childhood, I was very interested in seeing a film because movies represent dreams. The film industry is now more than a century old, but it still contributes to people. IMDb is the most popular database for movies. Due to more than 6 million movies being stored at IMDb, Amazon has owned this database since 1998. In this data visualization mini project, we utilize the dataset of IMDb to explore and discover some creative insight into the film industry. From 2004 to 2010, the film industry is the best period time in the history. We found that global movie gross income firstly proceeds from the US gross income. In other words, US movie income mainly came from the outside of the US market. Thus, we ask two questions which year and which month should have the highest average of US gross income between 2004 and 2010?

**Section 2: Target audience**

Movie production companies and executives are target audiences of my data visualization mini project.

**Section 3: Formulate a question and exploratory visual analysis**

1. Which year has the highest average of US gross income from 2004 to 2010?

A picture containing text, sky

Description automatically generated

From the control chart above, we see the highest average of US gross from 2004 to 2010 is 2009, which is $711,291,181.5.

2. Which month is the maximum upper limit in 2009?

Maximum upper control limit: $1,120,024,831

3. Which months had a higher average US gross from January 1, 1929, to December 19, 2069?

A picture containing boat, sky, outdoor, several

Description automatically generated

From the cycle plot above, we see five higher average months of US gross from January 1, 1929, to December 19, 2069 July, June, May, December, and November.

4.

Calendar

Description automatically generated

On average,

Action: May June July

Adventure: May July November June

Comedy: June November July December

Drama: December June November July

2009 December Action $760,167,650

2008 July Action $858,260,613

2007 May Adventure $968,670,672

2007 June December Comedy

**Section 4: Interpretation of discover**

Most people think that July and December have the highest gross because July 4 Independence Day and December 25 Christmas holidays might bring more customers. Our visualization displays May, June, and November are the hottest. Thus, if movies were released earlier one month before July and December, they would be better.

**References**

Boulenger, T. (2016). An analysis of movies by ratings.

<https://rpubs.com/ashtom/movierating>

Kaggle. IMDb movies extensive dataset. [Dataset].

<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset?select=IMDb+movies.csv>

Mccoy, S. (2021). Murach's Python for Data Analysis. Tableau & Python integration.