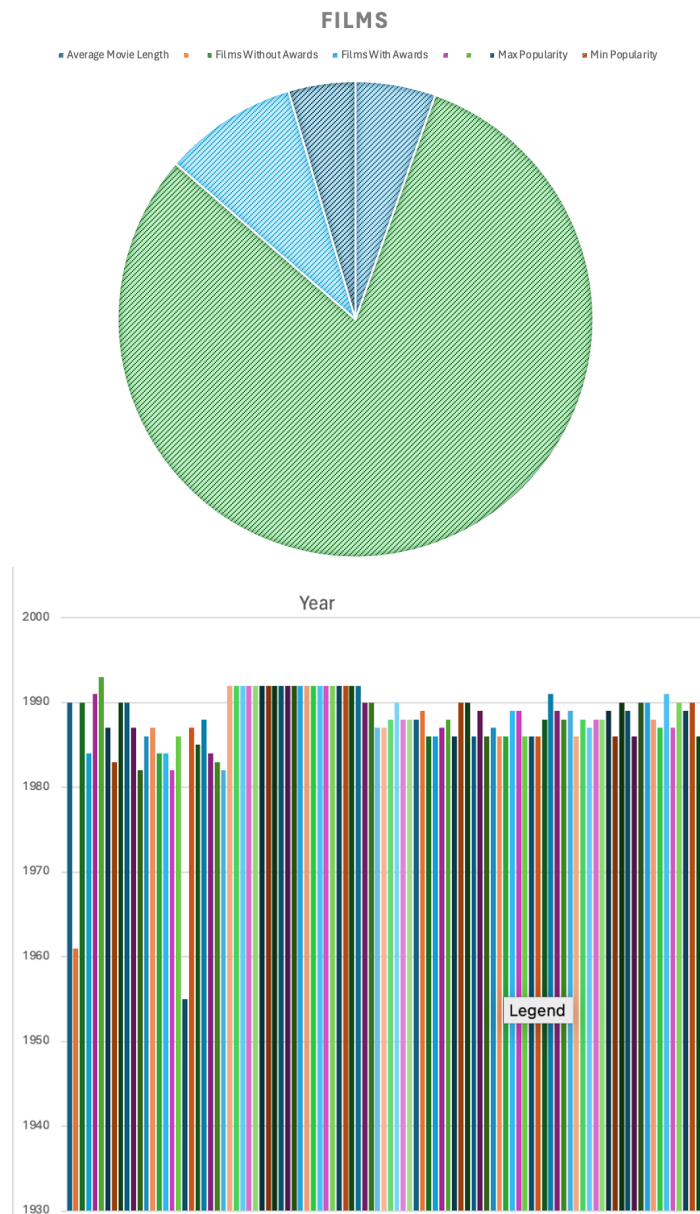


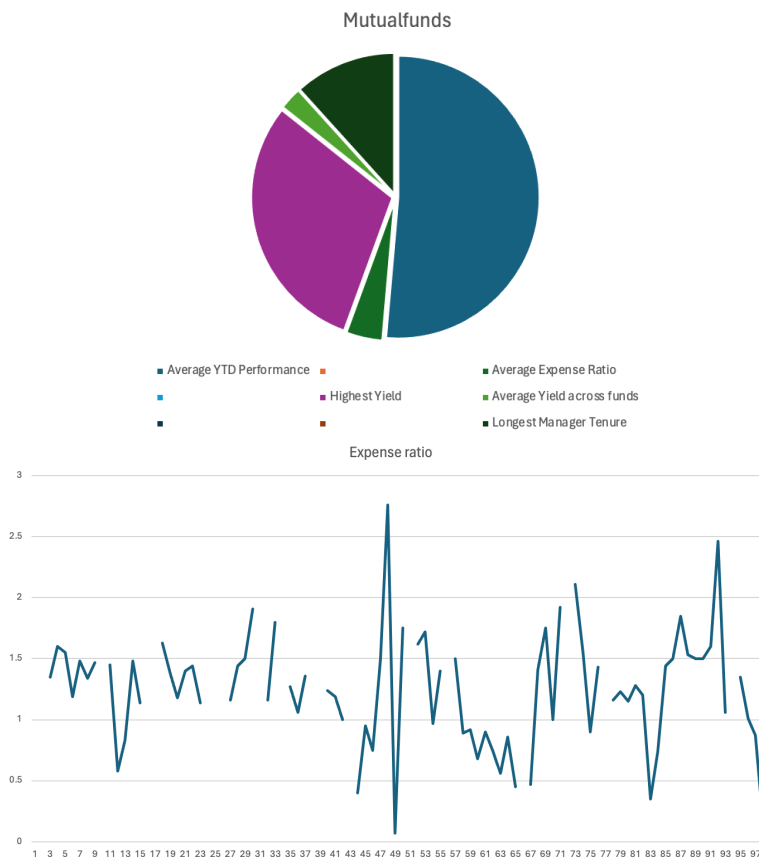
Project 1: Traditional Data analysis

For the assignment on traditional data analysis, two datasets were selected and explored using Excel: one related to films and the other to mutual funds. The analysis focused on uncovering non-obvious insights without the use of advanced visualization systems, although Excel's basic graph facilities were utilized for exploration. Here's a detailed write-up of the findings and the challenges encountered during the analysis:

Films Dataset



Mutual funds



Films Dataset Insights:

1. **Unique Genres:** There are 9 unique genres within this dataset. This was discovered by using Excel's data filtering feature on the 'Subject' column to count distinct genre values.
2. **Average Movie Length:** The average length of movies is approximately 105.4 minutes. This insight was found by applying Excel's AVERAGE function to the 'Length' column.
3. **Most Common Director:** A significant number of entries lacked director information, indicated by 'Missing'. This was identified through sorting the 'Director' column and observing the most frequent entry.
4. **Awards Distribution:** A vast majority of films (1564) have not received awards, while only 176 have. This distribution was uncovered by applying a COUNTIF formula to the 'Awards' column.
5. **Popularity Range:** The popularity scores range from 0 to 88. This was determined by using Excel's MIN and MAX functions on the 'Popularity' column.

Mutual Funds Dataset Insights:

1. **Average Year-to-Date (YTD) Performance:** The average YTD performance is around 14.67%. This calculation was performed by using the AVERAGE function on the 'YTD' column.

2. **Average Yield Across Funds:** The analysis revealed that the average yield across all mutual funds in the dataset is approximately 0.75%. This insight was obtained by applying the **AVERAGE** function on the 'Yield' column in Excel.
3. **Average Expense Ratio:** The average expense ratio is approximately 1.19%. This was calculated by applying the AVERAGE function to the 'Expense ratio' column.
4. **Highest Yield:** The highest yield recorded is 8.61%, identified by applying the MAX function to the 'Yield' column.
5. **Fund with the Longest Manager Tenure:** The longest manager tenure among the mutual funds in the dataset was found to be approximately 3.35 years. This was discovered by utilizing the **MAX** function on the 'Mgr tenure' column, which identifies the maximum value within a range of numbers.

Challenges and Further Exploration:

For the **Films Dataset:**

1. **Predictive Analysis:** Forecasting future popularity based on factors like genre, director, or cast would be challenging due to Excel's limited statistical modeling capabilities.
2. **Text Analysis:** Analyzing the text data for patterns, such as commonalities in award-winning movie titles or actor name trends, is difficult without natural language processing tools.
3. **Complex Relationships:** Understanding the intricate relationships between movie length, genre, and popularity would require more sophisticated data visualization and statistical tools.
4. **Time Series Analysis:** Analyzing trends over years to identify the evolution of film genres and lengths is cumbersome and not straightforward in Excel.
5. **Sentiment Analysis:** Determining the sentiment or critical reception of films from the title or awards data would be beyond Excel's capabilities.

For the **Mutual Funds Dataset:**

1. **Detailed Performance Analysis:** Deeply analyzing the factors influencing fund performance over different time frames would require complex statistical or machine learning models.
2. **Correlation Analysis:** Examining the correlation between expense ratios, yields, and performance ratings involves statistical tests that are not directly available in Excel.
3. **Category Deep Dive:** Further breaking down categories by subcategories or investment focus areas would be challenging without additional data and more advanced analysis tools.
4. **Risk Assessment:** Assessing the risk levels associated with different mutual funds based on their performance and yield data is not straightforward in Excel.

5. **Trend Forecasting:** Predicting future trends in mutual fund performance or rating changes would require predictive modeling capabilities not present in Excel.

This assignment underscored the power of Excel for initial data exploration and basic analysis but also highlighted its limitations for more sophisticated data analysis tasks. These insights and challenges pave the way for a deeper dive into data science and analytics tools that can handle complex analyses and provide more in-depth insights.