# IMDB Movie Ratings Sentiment Analysis

# Description:

The dataset is comprised of tab-separated files with phrases from the Rotten Tomatoes dataset.

**The sentiment labels are:**

0 - negative

1 - somewhat negative

2 - neutral

3 - somewhat positive

4 – positive

**Tasks**

1. **Read bank.csv.**
2. **Check if data is clean/clean**
3. **Convert the sentiments from the numerical numbers to their actual lable.**
4. **Find the count of positive, negative and neutral sentiments etc**
5. **Plot a countplot to show the count of sentiments.**
6. **Plot a piechart to show the distribution of the sentiments on the movies**
7. **Perform Sentiment Analysis. Which algorithm will you use? How accurate is it?**
8. **Create a csv for each Sentiment. How many positive and somewhat positive sentiments are there?**

**BBC NEWS NLP ANALYSIS**

**Find out what this article is about, and generate a wordcloud from the frequent words**

**BBC News:**

* **URL:** [**https://www.bbc.com/news/world-57026936**](https://www.bbc.com/news/world-57026936)
* **Class Name: "ssrcss-uf6wea-RichTextComponentWrapper"**

# California Independent Medical Review Dataset

This data is from the California Department of Managed Health Care (DMHC). It contains all decisions from Independent Medical Reviews (IMR) administered by the DMHC since January 1, 2001. An IMR is an independent review of a denied, delayed, or modified health care service that the health plan has determined to be not medically necessary, experimental/investigational or non-emergent/urgent. If the IMR is decided in an enrollee's favor, the health plan must authorize the service or treatment requested.

Read Medical\_Reviews.cv

Clean Data

Drop all duplicates

Group data by Diagnosis Category, Diagnosis SubCategory and Patients Gender(Show the count).

Plot a barplot to show the relationship between Report year and Patients Gender.

**Bar plots** for categorical data distributions ( diagnosis categories, treatment categories).

**What are the most common diagnosis categories and sub-categories?**

* Which diagnosis category appears most frequently?
* What are the top 10 most frequent diagnosis sub-categories?

**How are determinations distributed?**

* What are the unique determination values?
* How many cases fall under each determination category?

**What is the age range distribution of patients?**

* How many patients fall into each age range?
* What are the diagnosis and treatment distributions by age range?

## CHATGPT(Sentiment Analysis)

## About Dataset

ChatGPT has been a major talk in the tech world. The tweets about chatgpt were gathered for a month and then the sentiment analysis was made using Natural Language Processing.

1. Perform Sentimental Classification on this chatgpt dataset
2. Read chatgpt.csvs