Analysis of Search Terms done on Google

Student Name:-

1) Rohit Shidid MIS-NO: 142103012

2) Vineet V. Shinde MIS -NO:112003131

3) Siddharth Chaudhari MIS-NO:112003136

INTRODUCTION:-

- Sports are a major source of entertainment and excitement for people all over the world. From amateur players to professional athletes, millions of individuals are passionate about different types of sports. In this mini project, we will focus on three of the most popular sports: football, cricket, and basketball. The aim of this project is to analyse the search terms related to these sports and understand the global interest and attention towards these games.
- To achieve this goal, we will use data analysis techniques to examine the frequency and volume of search terms related to football, cricket, and basketball on various platforms, including YouTube, web, image, Google Shopping, news searches, and others. This will include analysing search trends on search engines, social media, and other relevant sources. Our analysis will cover a period of time to understand how the popularity of these sports has evolved over time.

- This project will provide valuable insights for sports enthusiasts, marketers, and industry professionals who are interested in the growth and popularity of these sports. It will also help us understand the impact of events such as world championships, league matches, and other major competitions on the popularity of these sports.
- In conclusion, this mini project will provide a comprehensive analysis of the search terms related to football, cricket, and basketball, and help us understand the global interest and attention towards these sports. Whether you are a fan, marketer, or industry professional, this project will provide you with valuable insights and data to help you make informed decisions.

Mini-Project problem statement

1) Requirement Analysis:

Technical Requirements:

- Data collection from various platforms (YouTube, web, image, Google Shopping, news searches, etc.)
- Data cleaning and analysis capabilities
- Visualisation of data analysis results
- Implementation using appropriate tools and technologies (data analysis software, visualisation software, report writing software)

Functional Requirements:

- Data export in various formats
- Data filtering and comparison capabilities
- Report generation and customization

2) Project Objectives:

Vineet(112003131):

Proj_obj1 : To measure the popularity and trend of the three sports based on multiple types of data.

Proj_obj2 : To measure the domination of each sport in high populated countries .

Siddharth(112003136):

Proj_obj1 : Overall conclusion of data over a period of 5 years

Proj_obj2 : To measure the popularity increase of the sports in low interest level regions during significant events.

Rohit (142103012):

Proj_obj1 : To finalise region wise conclusion of data over a period of 5 years

Proj_obj2 : To measure the off season regional Analysis

3) <u>The Data Required</u>: To measure the popularity and trend of the three sports based on online search data would include:

- Volume and frequency of searches made on various platforms, such as YouTube, web, image, Google Shopping, news searches, and others.
- Date and time of the searches, to determine the popularity and trend of the sports over time.
- Keywords used in the searches, to identify the most common terms and phrases used in searches related to the sports.
- Historical data about the popularity and trend of the sports over time, to provide context and insights into current trends.
- Event data, such as information about major events and competitions related to the three sports, to analyse the impact of events on the popularity of the sports.

4) <u>Input Required:</u> Information about the type of content that users are searching for, such as video, image, news, etc.

- Keywords used in the searches: To identify the most common terms and phrases used in searches related to the sports.
- Date and time of the searches: To determine the popularity and trend of the sports over time.

- Search data: The volume and frequency of searches made on various platforms, such as YouTube, web, image, Google Shopping, news searches, and others.
- Geographical location data: Information about the geographical location of the users making the searches.

Data cleaning:

Loading data in data frame needs uniform data which has same number of columns throughout out the entire csv file.

```
[ ]  import numpy as np
     import pandas as pd
```

```
# Loop the data lines
with open("data.csv", 'r') as temp_f:
    # get No of columns in each line
    col_count = [ len(l.split(",")) for l in temp_f.readlines() ]

# Generate column names  (names will be 0, 1, 2, ..., maximum columns - 1)
column_names = [i for i in range(0, max(col_count))]

#Read csv
df = pd.read_csv("data.csv", header=None, delimiter=",", names=column_names)
```

Removing unnecessary lines and empty lines from the data

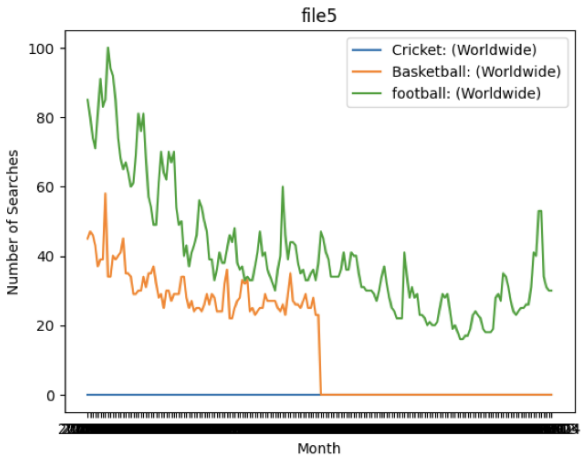
```python
# Read the file into a list
with open('data.csv', 'r') as f:
    lines = f.readlines()

# Filter out any elements that match the line
lines = [line for line in lines if not line.startswith('Category: All categories')]

# Filter out any elements that start with the word "Country"
lines = [line for line in lines if not line.startswith('Country')]

# Modify all lines starting with the word "Month"
for i in range(len(lines)):
    words = lines[i].strip().split(',')
    if words[0] == 'Month':
        lines[i] = 'Month\n'

# Filter out any empty elements
lines = [line for line in lines if line.strip()]

# Write the modified list back to the file
with open('data2.csv', 'w') as f:
    f.writelines(lines)
```

5) ***Output Expected***: Trend analysis: A detailed analysis of the trends in the popularity of the three sports over time, based on the volume and frequency of searches.

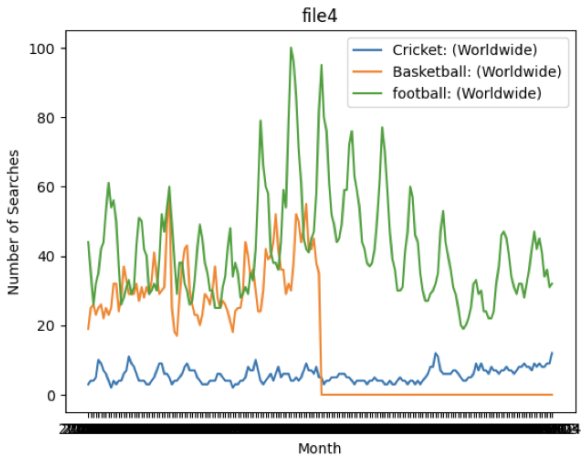
- Keyword analysis: An analysis of the most common terms and phrases used in searches related to the three sports.
- Event impact analysis: An analysis of the impact of major events and competitions on the popularity of the three sports.
- Geographical analysis: An analysis of the geographical location of users who are searching for the sports.
- Type of content analysis: An analysis of the type of content that users are searching for, such as video, image, news, etc.

EDA(https://colab.research.google.com/drive/16iTwsq9EzSx_YfF6bbjRkfZnnzZAO2Zr):

YOUTUBE SEARCHES



GOOGLE SHOPPING



NEWS SEARCHES

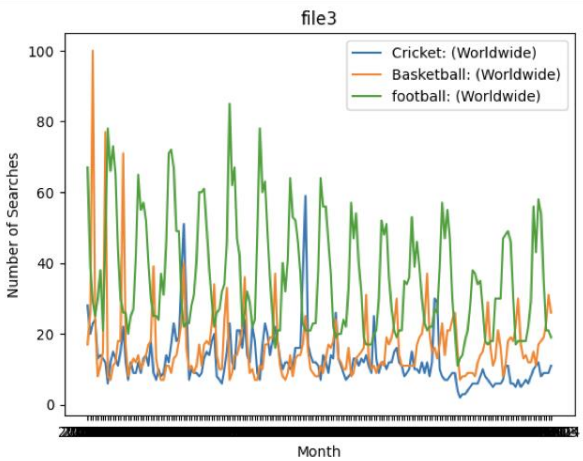
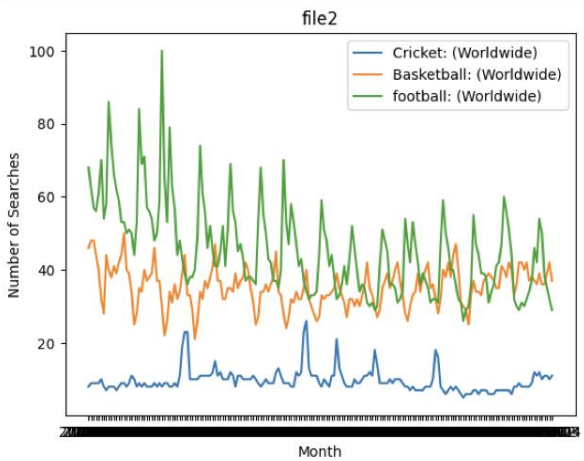
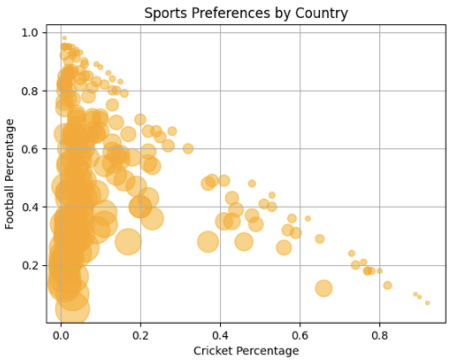
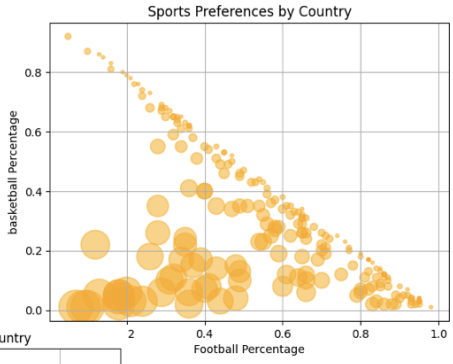
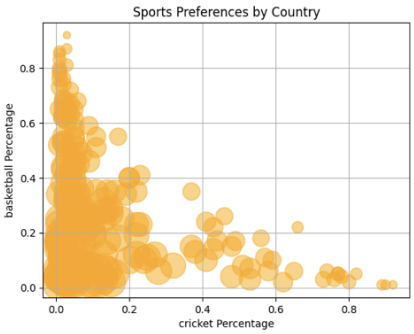


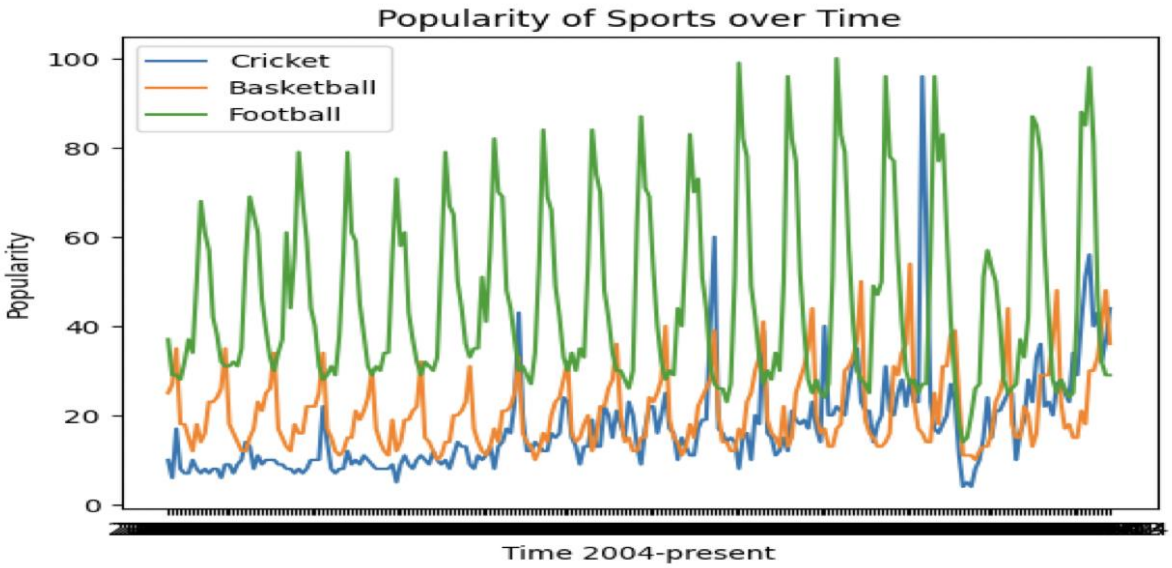
IMAGE SEARCH





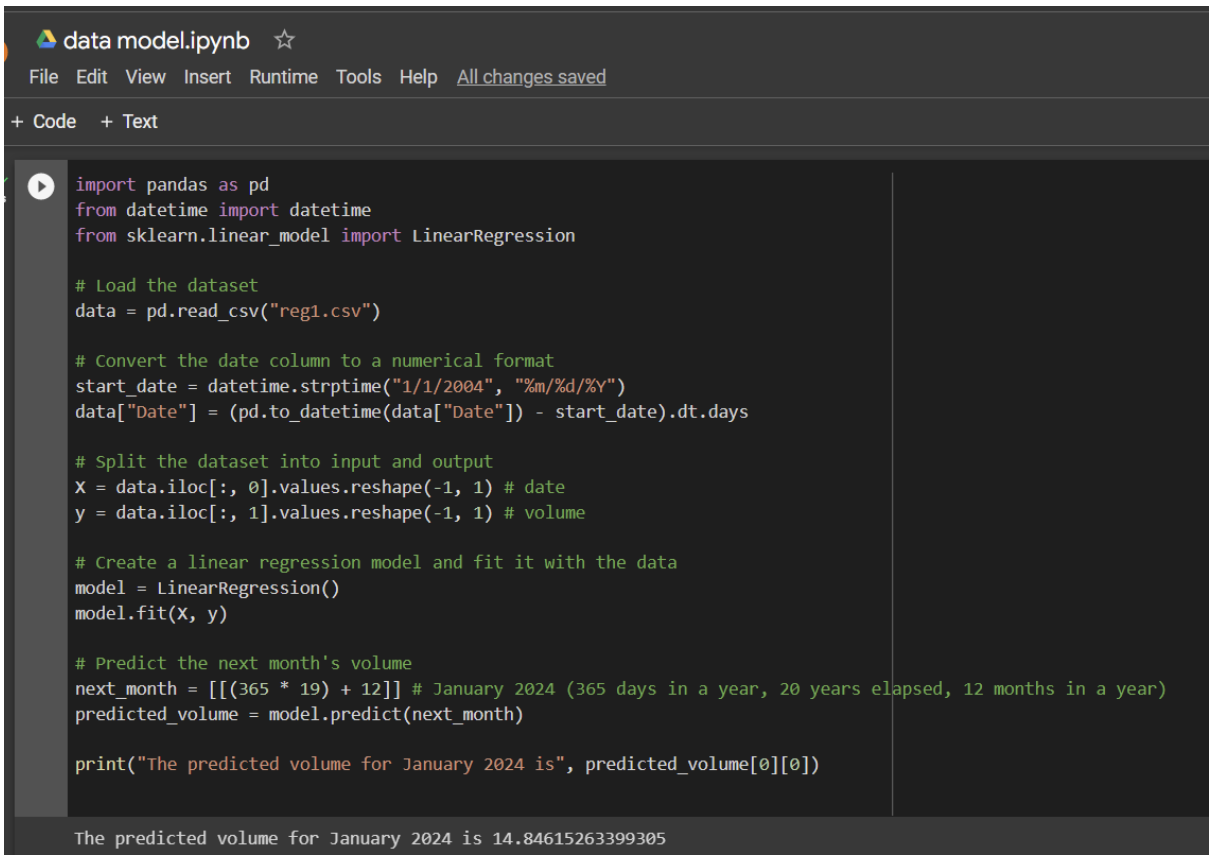






6) Design Model:

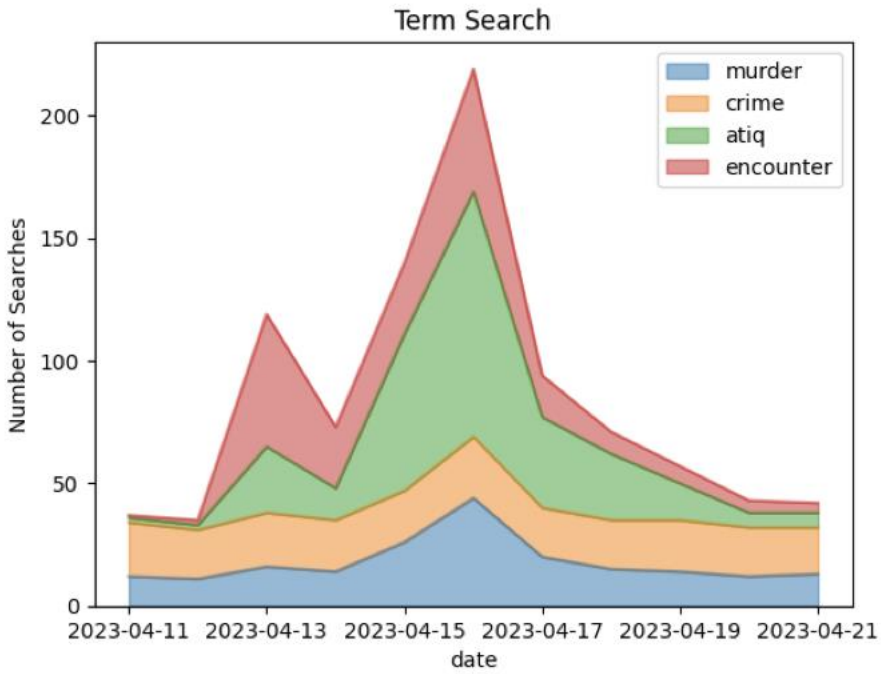
- For continuous data/categorical data a linear regression model works best in approximating/predicting the outcomes for the future.
- By using the linear regression data model on the data we can find the predictions for a sport for next year as to how much popularity it might achieve by next year this is obviously not taking into considerations unforeseen circumstances for example covid-19.



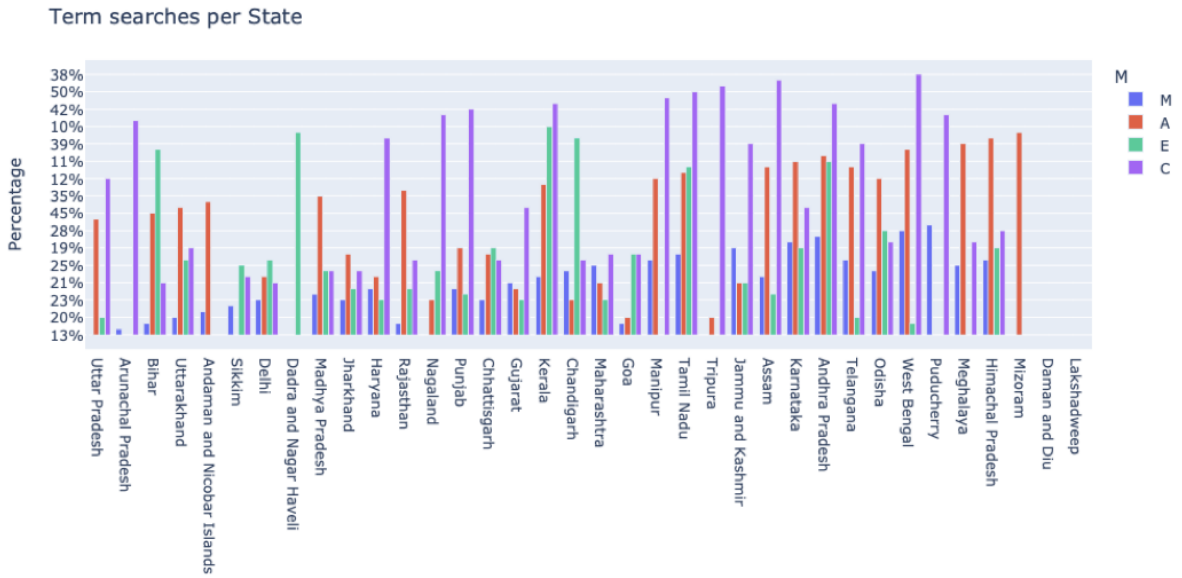
Case Study

(https://colab.research.google.com/drive/1ybtLJEChPdolMdMlAJsqHw8nLEBu3AVe):

In



addition to our primary analysis of sports-related search terms, we undertook a case study to examine how people search on Google using specific keywords. We specifically explored the search patterns surrounding keywords like "murder," "crime," "Atiq," and "encounter" in light of the unfortunate event of Atiq's death. This case study aimed to understand the public's interest, concerns, and information-seeking behavior related to these keywords during that specific incident. By analyzing the search trends and user queries, we aimed to uncover valuable insights into the public sentiment, perception, and curiosity surrounding such significant events. This case study provided a deeper understanding of how people utilize search engines to seek information and engage with current events and topics of interest.



Case Study Conclusion:

In conclusion, this case study aimed to analyze the changes in keyword searches related to murder, crime, encounter, and Atiq Ahmed immediately following his murder.

The study results revealed significant changes in search patterns across different states in India, with some states showing a marked increase in searches related to these keywords. The findings suggest that the public's interest and curiosity in high-profile criminal cases can significantly impact their search behaviour and by extension, their perceptions of the justice system.

Overall, this case study highlights the importance of monitoring and analyzing search trends related to crime and other high-profile events, as they can provide valuable insights into public sentiment and perceptions.

Project Conclusion:

In conclusion, this project aims to conduct a comprehensive analysis of keyword searches related to football, cricket, and basketball over the past two decades. By using a variety of data sources and analysis techniques, the project aims to identify trends and insights that can be used by various stakeholders in the sports industry.