Business Intelligence and Data Warehousing (ANL408)

- By Sabarish Nair

About me...

9 years of work experience

Bachelors in Information Technology

Masters in Business Analytics

Worked as a software developer, senior analyst and product manager

Data and Data Explosion

- Data getting generated at a rapid pace.
- Internet Usage, Digital Transformation, IoT Devices, Big Data, Social Media, etc, contributing to the data explosion.

SUCCESS = DATA-DRIVEN INSIGHTS!!

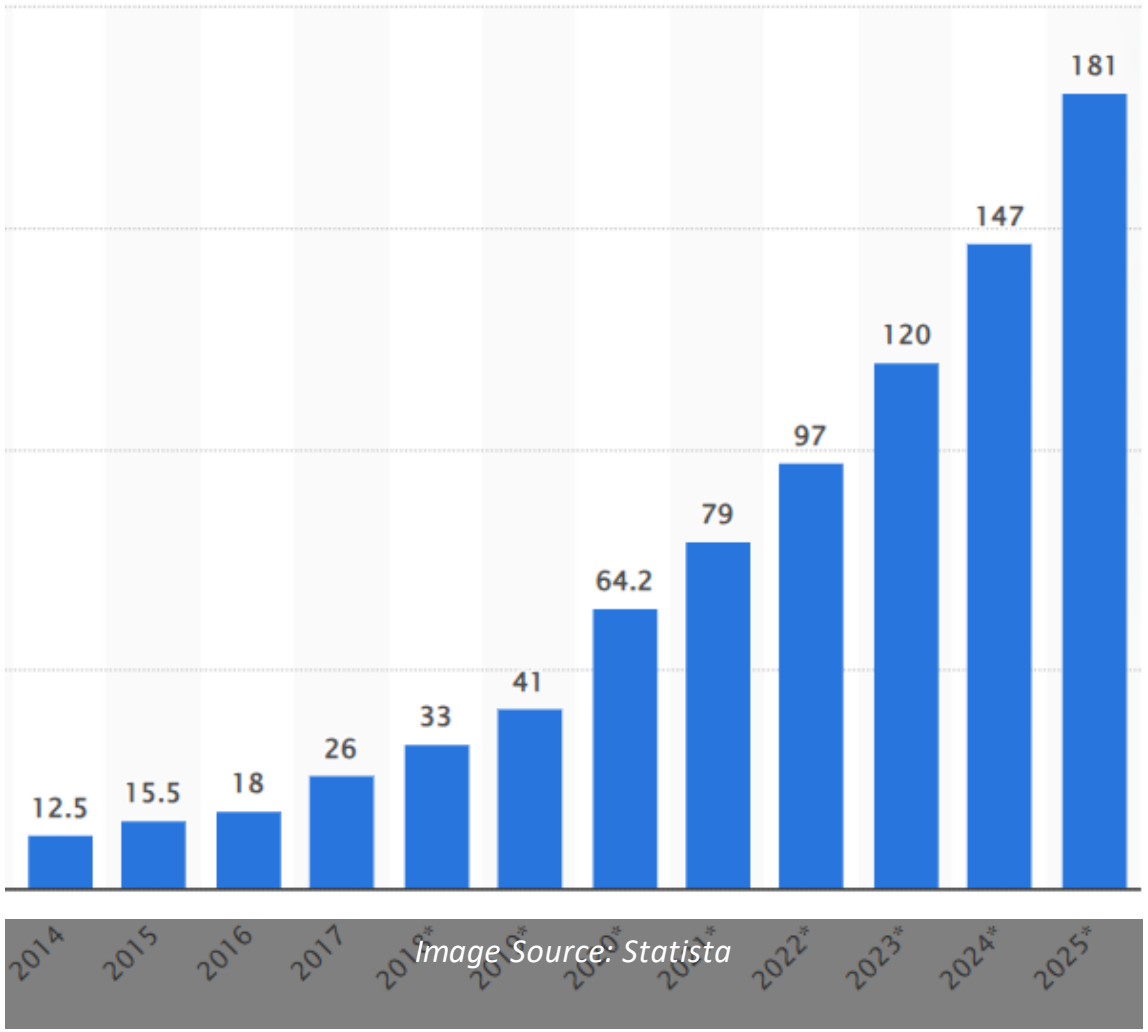


Image Source: Statista

Types of Data

Structured Data: Highly Organized and formatted in a predefined manner. (E.g., Relational database, spreadsheet and CSV Files)

Unstructured Data: Lacks a predefined model or structure. (E.g., Images, Video and Audio)

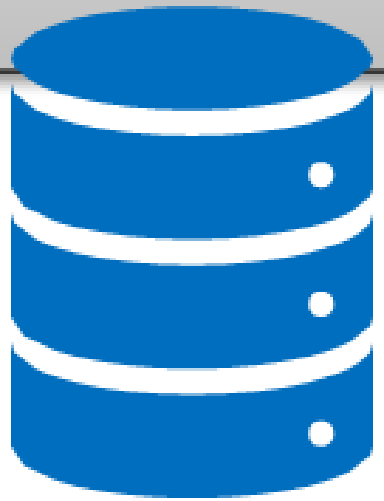
Semi-Structured Data: Has some organizational properties making it easier to analyze. (E.g., JSON, XML)

Database

"*An organized collection of data designed to efficiently manage, store, retrieve, and update large amounts of data.*"

- **Structured Data:** Relational Databases (MySQL, PostgreSQL, etc.)
- **Unstructured Data:** Object Stores (Amazon S3, Google Cloud Storage, etc.)
- **Semi-Structured Data:** NoSQL Databases (MongoDB, Cassandra, etc.)

Why do we need a data warehouse?



Problem Statements

"Yes, we have a lot of data but we don't use it"

"Our data is very complicated and difficult to analyze"

"It's spread all over the different systems and difficult to access"

"I just want to see what is relevant!"

"We need to access data quick and easily"

"We want to make fact-based decisions!"

Two Purposes!

- Thousands of records at a time
- Fast query performance
- Historical context
- Usablity

Analytical
Decision
Making

Operational
Data Keeping

- One record at a time
- Data input
- No long history

Why do we need data warehouse?

Consolidation of information sources.

Improved query performance

Separate research and decision support systems from operational systems.

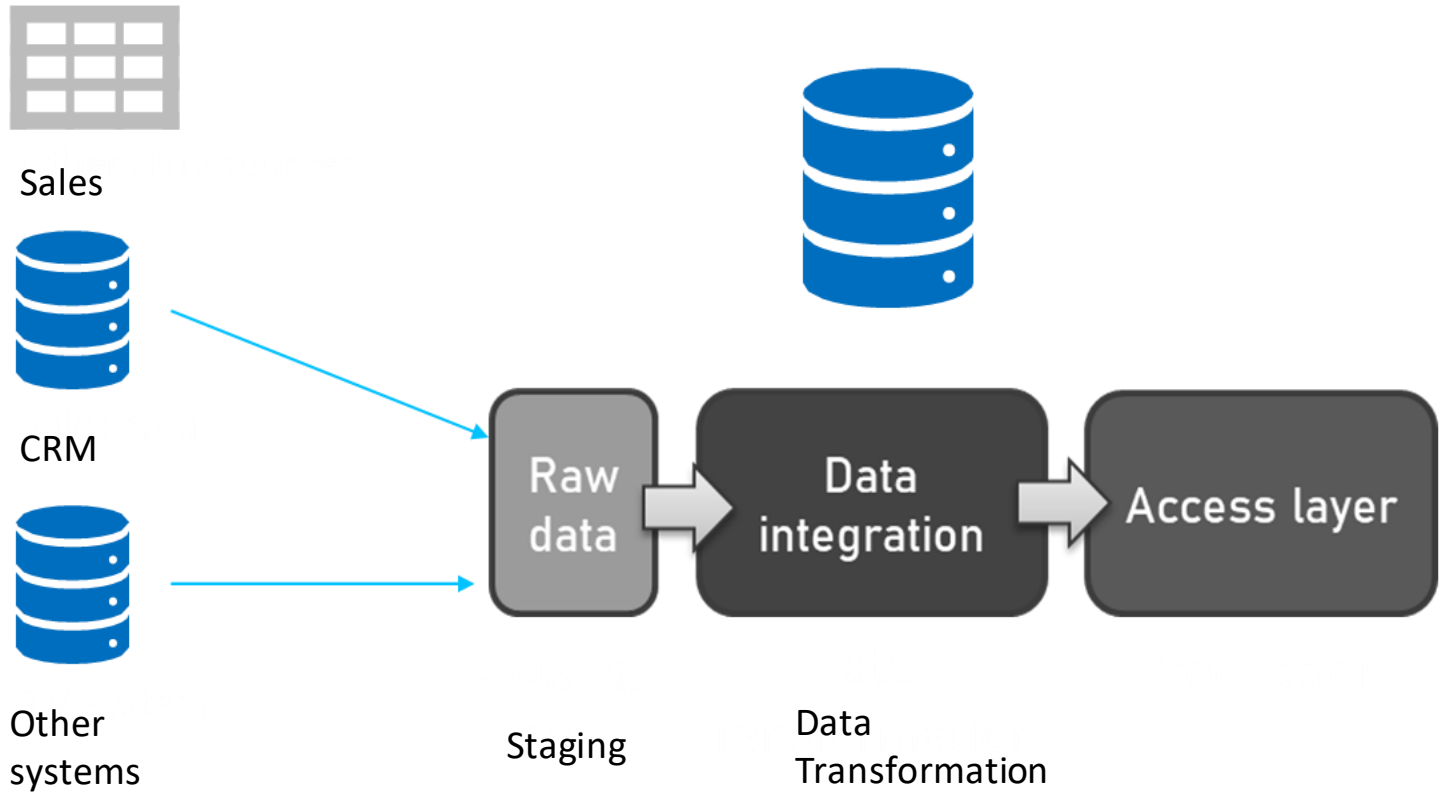
Foundation for data mining, data visualization, advanced reporting, and OLAP tools.

What is data warehouse?

Database used and optimized for analytical purposes.

Subject-Oriented, Integrated, Time-Variant Non-volatile collection of data in support of decision making

Data warehouse landscape



Components of Data warehouse

ETL Solutions

Reporting

Data Mining Capabilities

Other in-house applications

Characteristics of Data Warehouse

Centralized and consistent location of data

Fast query performance

User friendly

Consistent and repeated data loading (ETL)

Reporting and data visualization (built on top)

Data warehouse software

- Amazon Redshift
- Azure Synapse Analytics and Azure SQL database
- Google Big Query
- Snowflake data warehouse
- Teradata
- And many more....

Database vs Data Warehouse

Purpose	Data warehouse	Database
Purpose	Analysis	Reporting
Database	OLAP	OLTP
Type of collection	Subject-oriented	Application-oriented
Query	Complex analytical queries	Simple transaction queries

We create a data warehouse for Business Intelligence...

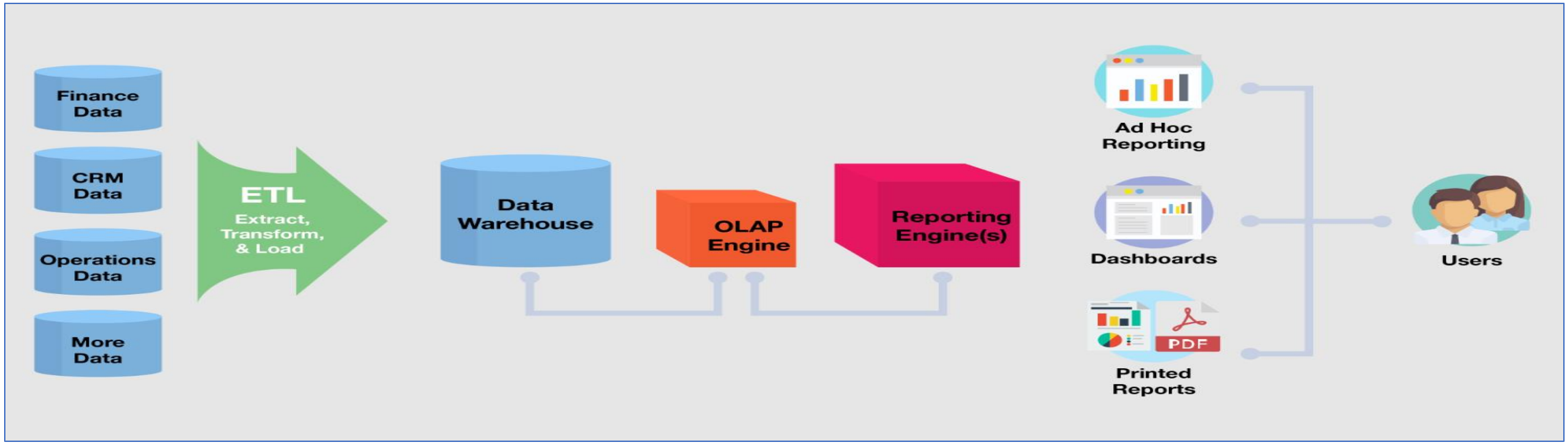
What is Business Intelligence?

Business Intelligence (BI)

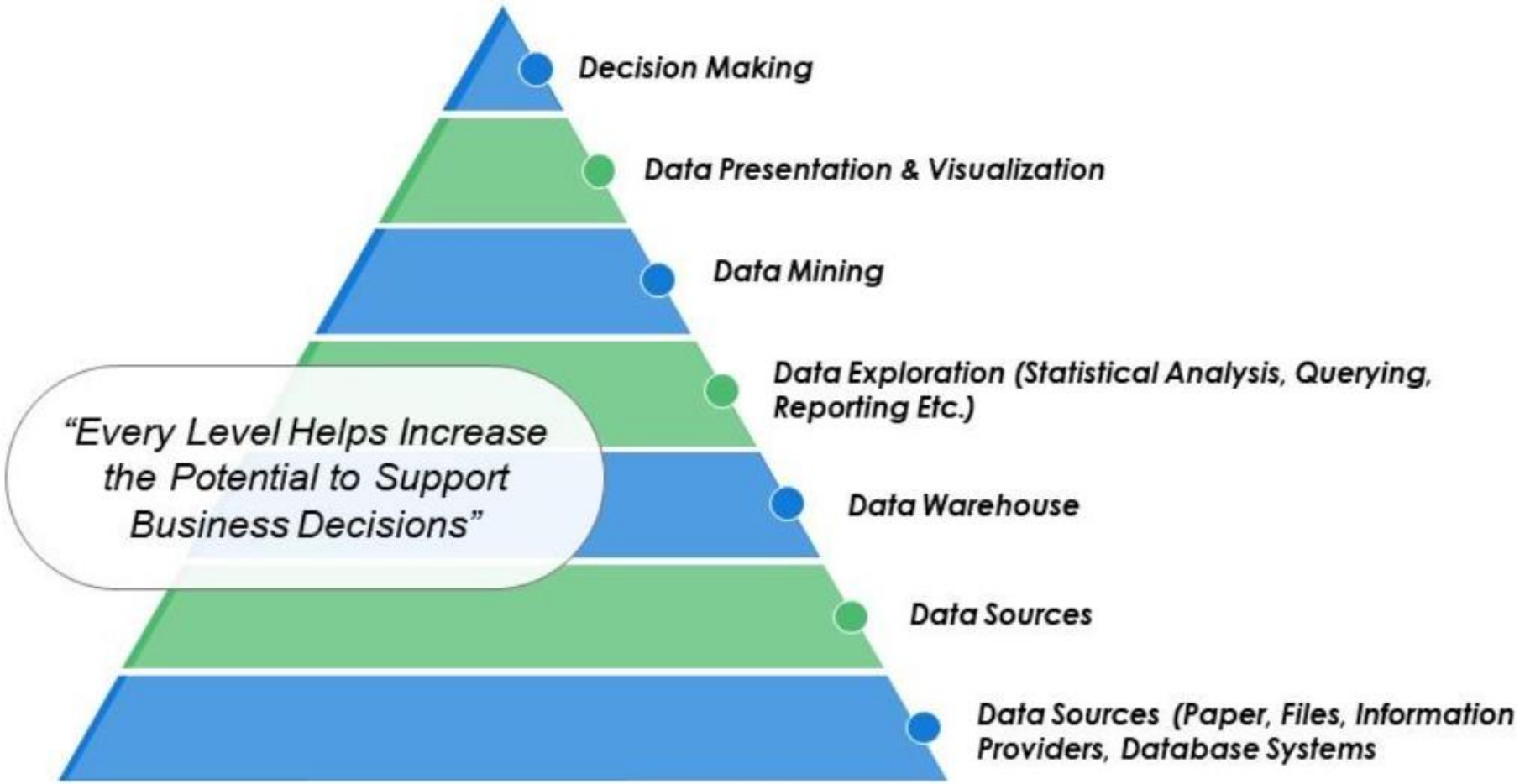
The process of drawing critical insights from data to inform decisions.

Technology that allows gathering, storing, accessing, and analyzing data to help users make better decisions.

BI Example


Finance Data
CRM Data
Operations Data
More Data
ETL
Extract, Transform, & Load
Data Warehouse
OLAP Engine
Reporting Engine(s)
Ad Hoc Reporting
Dashboards
PDF
Printed Reports
Users

BI Process


Decision Making
Data Presentation & Visualization
Data Mining
Data Exploration (Statistical Analysis, Querying, Reporting Etc.)
Data Warehouse
Data Sources
Data Sources (Paper, Files, Information Providers, Database Systems
"Every Level Helps Increase the Potential to Support Business Decisions"

BI Components

Data Warehouse: A decision support database that is maintained separately from the organization's operational database.

Online Transactional Processing (OLTP) – A class of information systems that facilitate and manage transaction-oriented applications (ATMs, online booking applications, etc.)

Online Analytical Processing (OLAP) – Approach to answer multi-dimensional queries (drill down, slicing and dicing, marketing analysis, supply and demand forecasting)

BI Uses

- Business Operations Reporting
- Forecasting
- Dashboard
- Multidimensional analysis
- Finding correlation among different factors

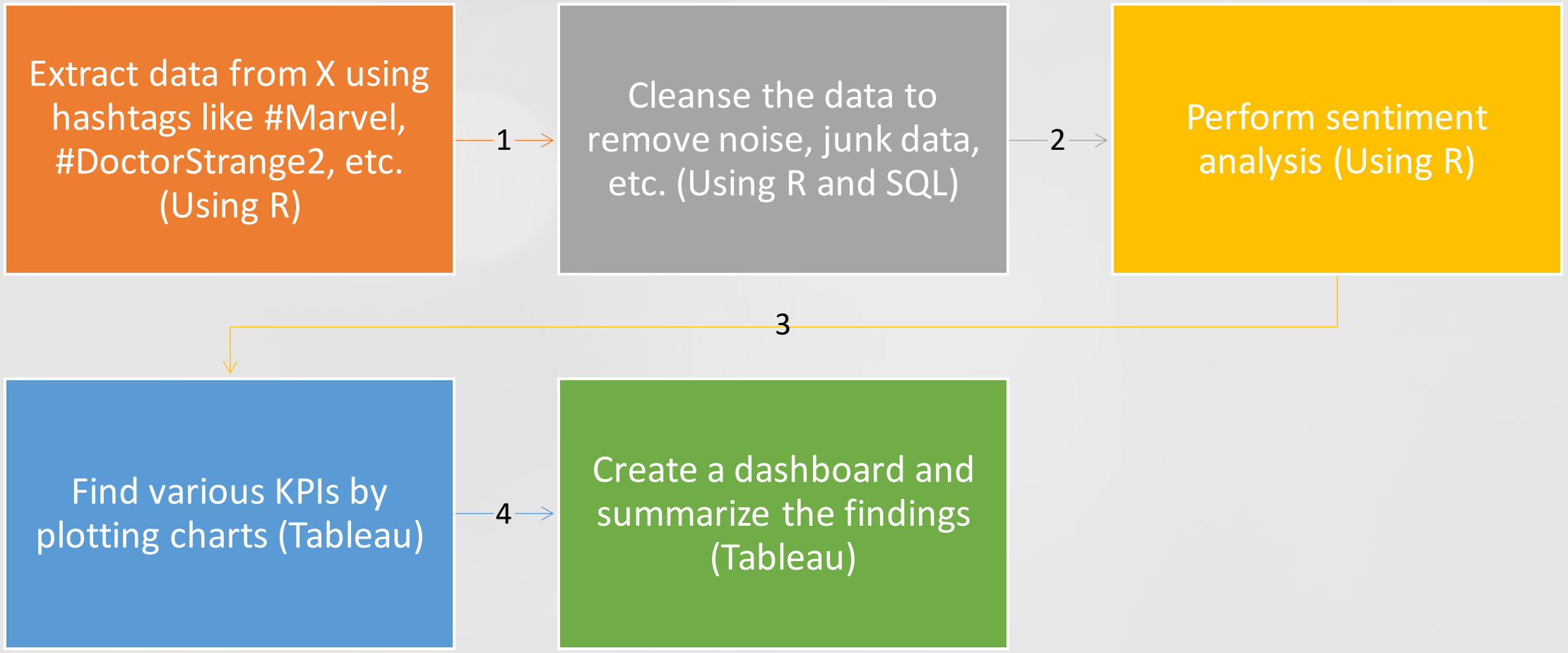
BI Tool

- Tableau
- Power BI
- QlikView
- Google Data Studio
- And many more.....

Practical Example: Doctor Strange in the Multiverse of Madness

- How can you help Marvel Studios identify trends and insights from social media (X, formerly Twitter)?
- Suggest recommendations to Marvel Studio on how to utilize the findings.
- Find distinct trends, analyses, and patterns in your data.
- Analysis to be performed on at least 800 tweets collected over a time period.

Steps to be followed



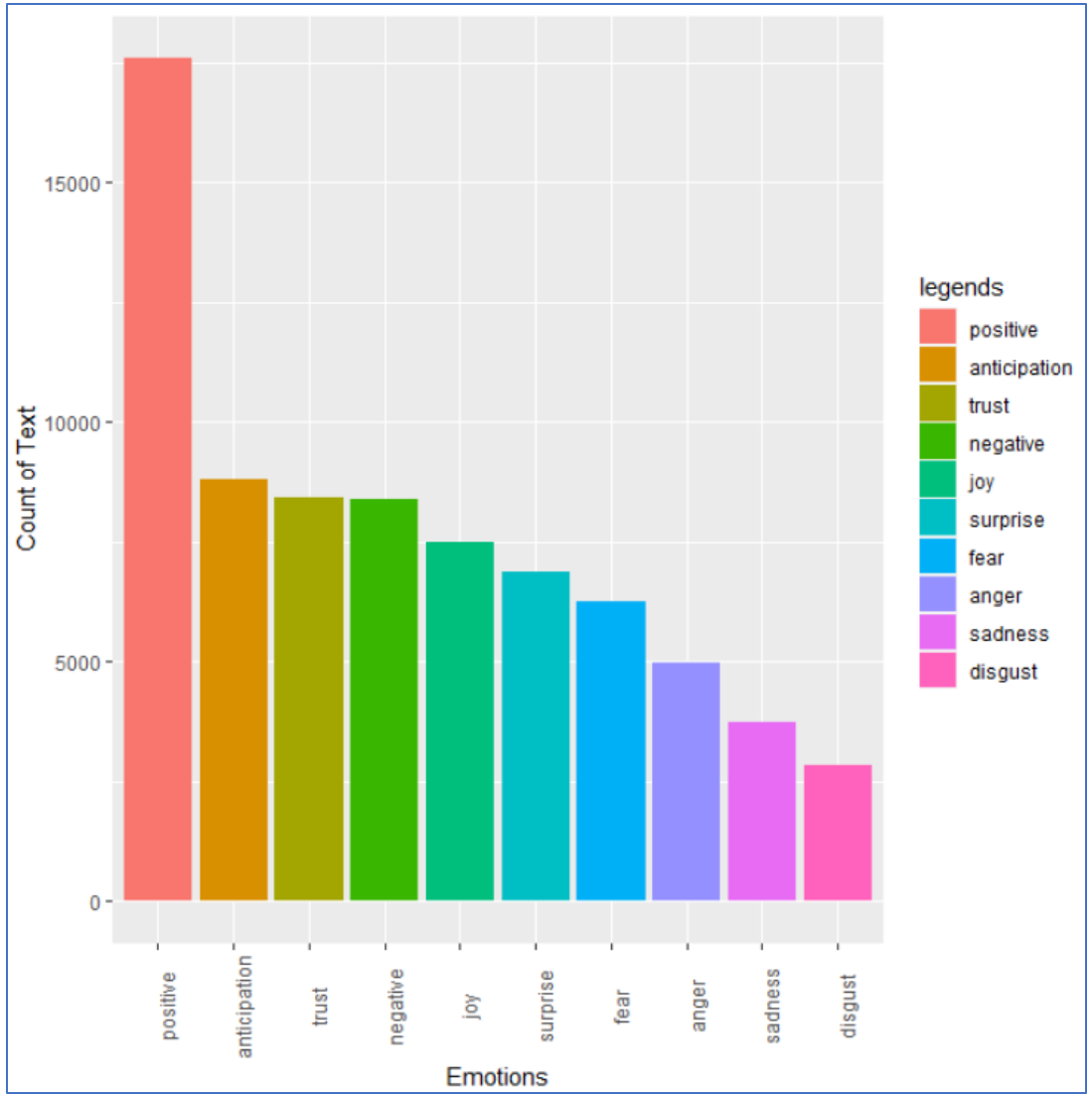
Code Snippets

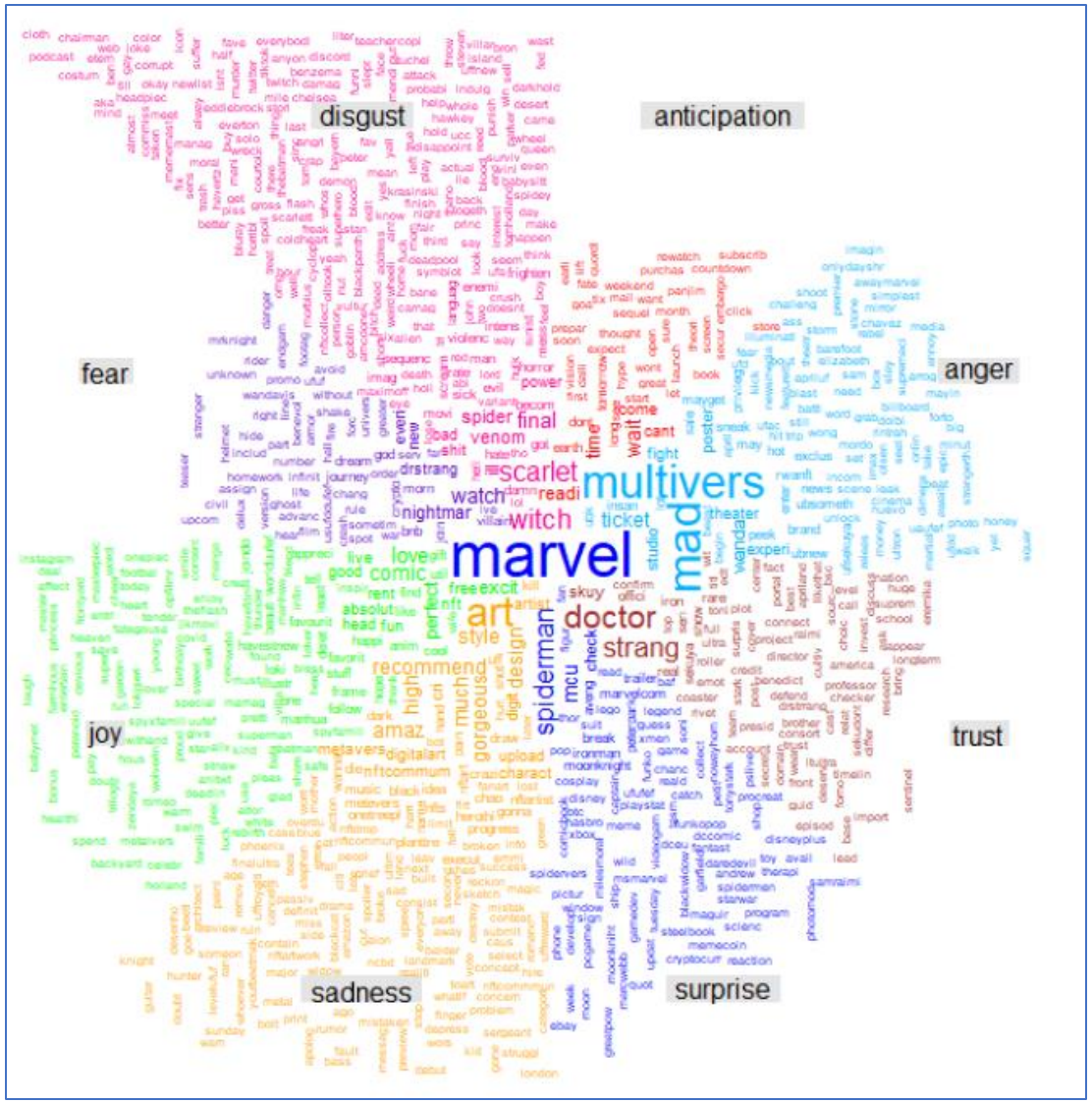
```r
install.packages("twitteR") #install package
library(twitteR) #load package

setup_twitter_oauth(api_key, api_secret_key, access_token, access_token_secret)

multiversedataset4 <- searchTwitter('#MultiverseofMadness + #DoctorStrange2',
                                    n = 800,
                                    retryOnRateLimit = 1e3)
```

```r
search_criteria_filterquery1 <- "#MultiverseofMadness
                                 OR (#MultiverseofMadness AND #DoctorStrange2)
                                 OR (#MultiverseofMadness AND #ScarletWitch)"

search_criteria_filterquery1_results <- search_tweets(search_criteria_filterquery1,
                                                      n = 54000,
                                                      retryonratelimit = TRUE,
                                                      lang = 'en',
                                                      since='2022-03-07',
                                                      until='2022-04-14')
```

```r
#Loading Library Tidyverse
library(tidyverse)

#Selecting only the required attributes from the dataset
filtered_dataset <- merged_datset %>%
  select(
    created_at,
    screen_name,
    text,
    source,
    is_retweet,
    favorite_count,
    retweet_count,
    hashtags,
    media_expanded_url,
    mentions_screen_name,
    quoted_text,
    retweet_text,
    retweet_source,
    retweet_favorite_count,
    retweet_screen_name,
    retweet_name,
    place_name,
    place_type,
    place_full_name,
    country,
    country_code,
    name,
    location,
    description,
    followers_count,
    friends_count,
    account_created_at,
    verified
  )
```

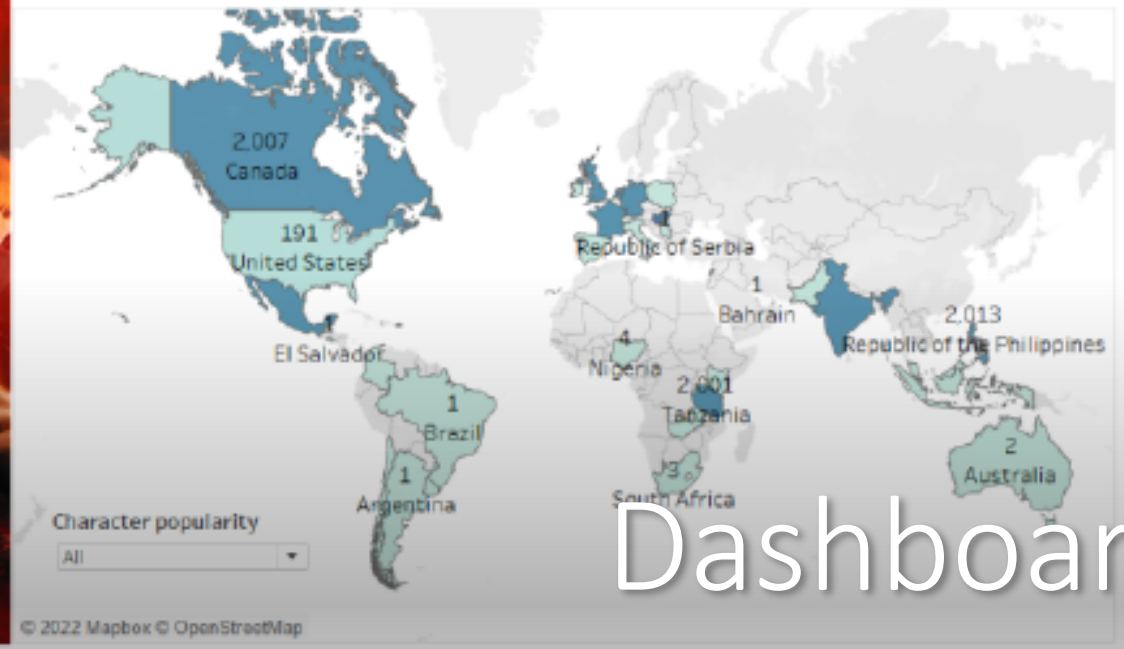
Sentiment Analysis



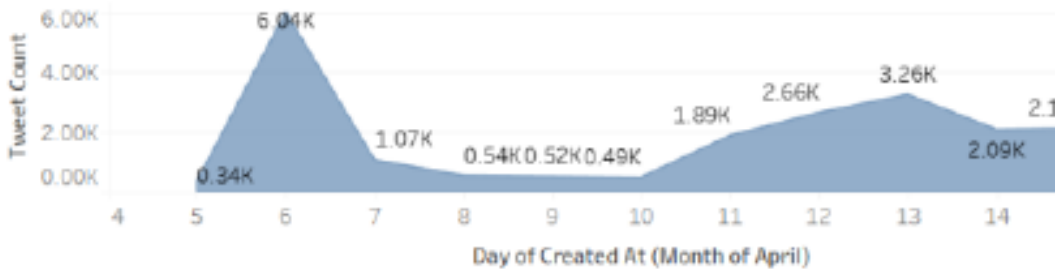


Movie Analytics: Doctor Strange in the Multiverse of Madness

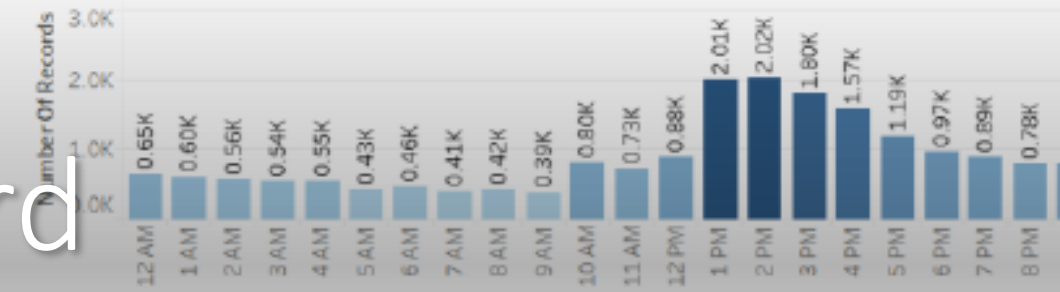
Geographic Customer Segmentation



Daily Tweet Engagement Rate: 21,029 tweets

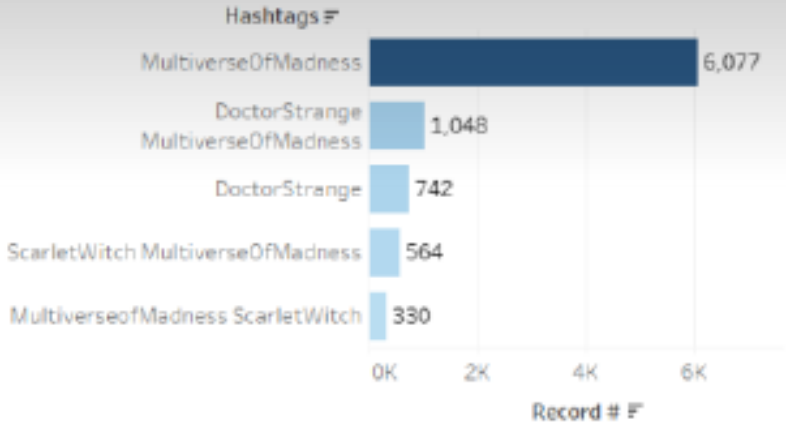


Hourly Engagement Rate

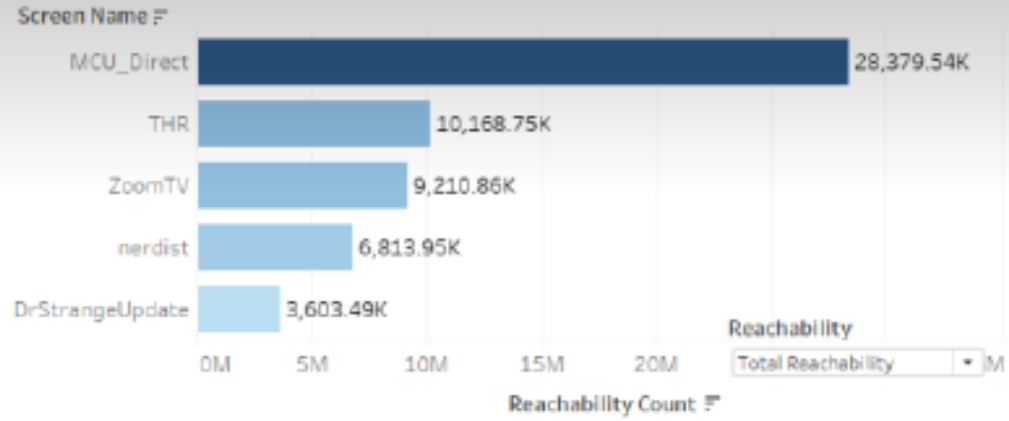


Dashboard

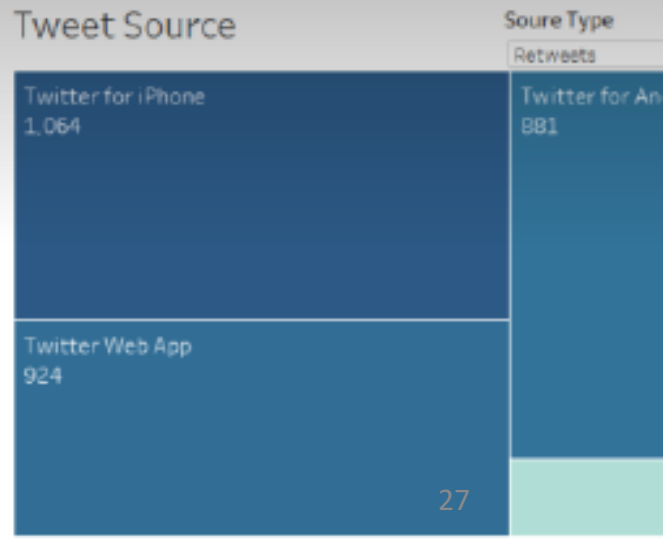
Hashtag Performance



Reachability (Top 5 Users)



Tweet Source



Recommendations to Marvel Studios

Sr. No.	KPI	Recommendations
1	Geographic customer segmentation and character popularity	1. Target countries with Marvel comic fans like the US, India, UK, etc. for movie promotion, 2. Revise marketing strategies in other countries, 3. Send the lead actors to regions with a huge fan Bases
2	Customer Engagement	1. Improve daily engagement rate threshold 2. Create social media buzz by lead actors 3. Utilize timeslot between 1 PM and 3 PM for promotional activities
3	Hashtag performance	1. Utilize top 10 hashtags for future posts 2. Analyze user sentiments from hashtags
4	Reachability	1. Use influencer marketing 2. Hire Top Influencers to promote films since it is cost-effective
5	Tweet Source	1. Devise marketing strategies for iPhone Users 2. Develop custom content and advertisements tailored to iPhone users based on their Demographics


Thank you