DATA ANALYSIS PORTFOLIO

By
VISHAL SONI

TABLE OF CONTENTS

PROFESSIONAL BACKGROUND

Proficient Data Analyst with 2 years of hands on experience in Microsoft SQL server, Excel and data visualization tool Kibana. I have been a part of couple of startups. Working as a Data Analyst for Voziq AI was a great experience in terms of learning and developing my technical skills. My expertise lies in exploratory data analysis , visualizations , reporting , dashboarding and writing SQL queries. I am currently honing my skills and looking for better opportunities to grow as a Data Analyst.

PROJECT 1:INSTAGRAM USER ANALYTICS

Description : This project aims to carry out the in-depth analysis of user engagement process with the Instagram platform which will help the product team to launch better features for the platform.

User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) in an attempt to derive business insights for marketing, product & development teams. These insights are then used by teams across the business to launch a new marketing campaign, decide on features to build for an app, track the success of the app by measuring user engagement and improve the experience altogether while helping the business grow. Work with the product team of Instagram and the product manager has asked you to provide insights on the questions asked by the management team.

This project focuses mainly on two key aspects Marketing and Investors Metrics.
Based on the user engagement and the data collected the insights needs to be carried out and presented to the product team. The project will answer to the important questions like :

> Rewarding Loyal Users
> Remind Inactive Users to Start Posting
>Declaring Contest Winner
>Hashtag Researching
>Launch AD Campaign
>User Engagement

Top Five Instagram Users

User Name	ID	Start Date	End Date	Tenure
Aniya Hackett	5	2016-12-07	2022-12-17	6 Yrs.
Arel Bogan63	4	2016-08-13	2022-12-17	6 Yrs.
Kasandra Homenick	7	2016-12-12	2022-12-17	6 Yrs.
Tabitha_Schamberger11	8	2016-08-20	2022-12-17	6 Yrs.
Gus93	9	2016-06-24	2022-12-17	6 Yrs.

Based on tenure the top five users has been selected. Whereas tenure being the difference in the created date/sign update and present date.

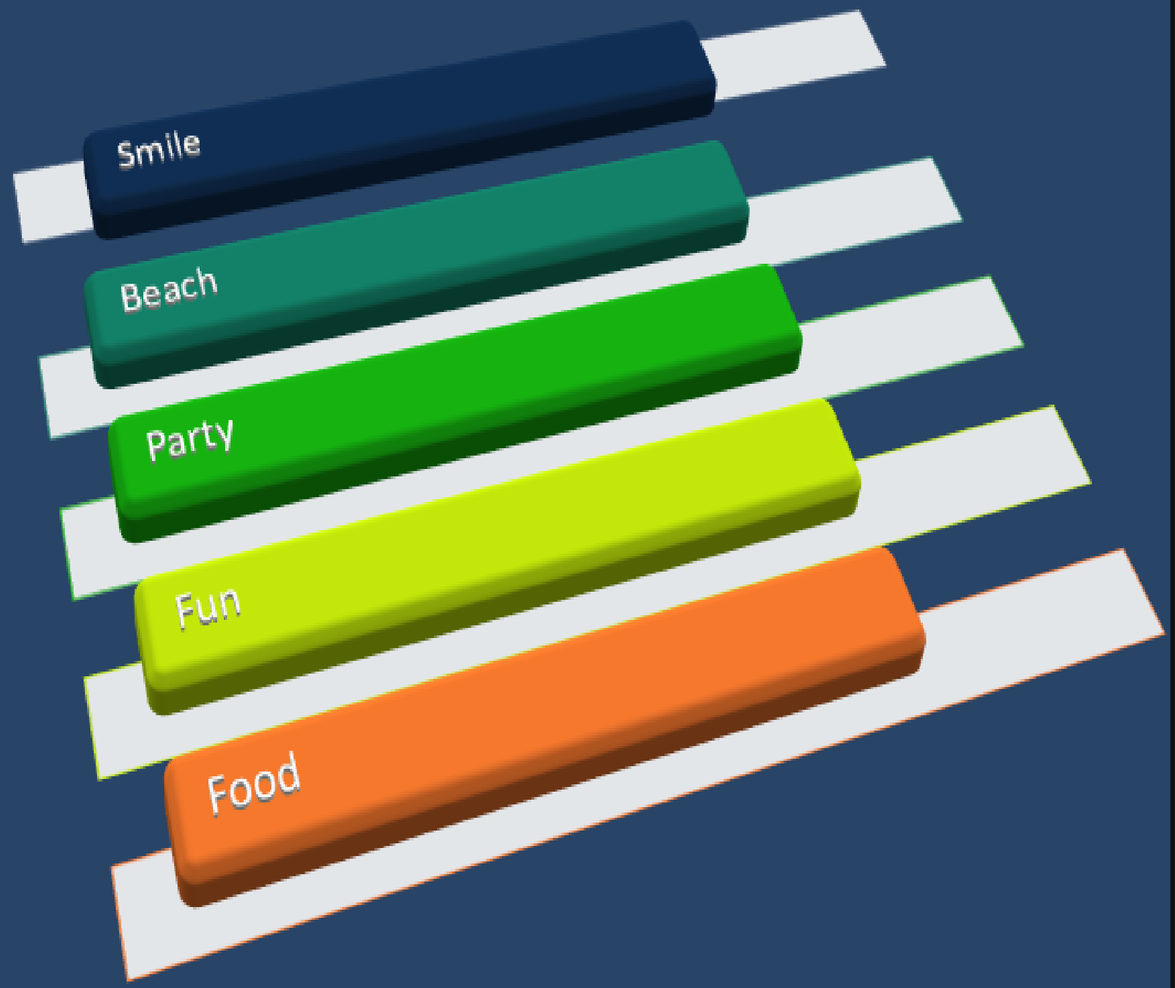
Inactive Users Never Posted a Photo

User ID	User Name
5	Aniya Hackett
7	Kasandra Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell Halvorson
25	Tierra Trantow
34	Pearl7
36	Ollie Ledner37
41	Mckenna17
45	David Osinski47
49	Morgan Kassulke
53	Linnea59
54	Duane60
57	Julien Schmidt
66	Mike Auer39
68	Franco Keebler64
71	Nia Haag
74	Hulda Macejkovic
75	Leslie67
76	Janelle Nikolaus81
80	Darby Herzog
81	Esther Zulauf61
83	Bartholome Bernhard
89	Jessyca West
90	Esmeralda Mraz57
91	Bethany20

To identify inactive users we need to identify users which are not present in the photos table. After querying inactive users from the database the list of inactive users was prepared.

To identify the winner of the contest we need to check which user has received most likes. After querying tables likes , users and photos table Zack has been announced as the winner who has received 48 likes on the photo which he posted.

Hashtag Researching



After querying the table phototags following are the recommended hashtags to use in the posts to reach most people on platform.

AD Campaign

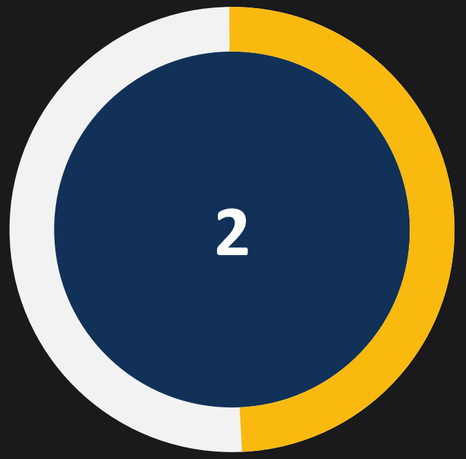
- According to the data users have registered most on Sundays and Thursdays.

- The Recommended days to launch ADs are Thursday and Sunday.

User Engagement

Total Users

Total Photos



Average Posts Per User

CONCLUSION

This task helps understand the basics fundamentals of structured query language (SQL). The aim was to derive insights by leveraging data. The emphasis of the project was more on understating the structure of the database rather than deriving insights only. The project also helps in understanding the relevance of user engagement process with a digital product which can actually help a business grow.

PROJECT 2 :OPERATION AND METRIC SPIKE ANALYTICS

Description : Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. You work closely with the ops team, support team, marketing team, etc. and help them derive insights out of the data they collect. Being one of the most important parts of a company, this kind of analysis is further used to predict the overall growth or decline of a company's fortune. It means better automation, better understanding between cross-functional teams, and more effective workflows. Investigating metric spike is also an important part of operation analytics as being a Data Analyst you must be able to understand or make other teams understand questions like- Why is there a dip in daily engagement? Why have sales taken a dip? Etc. Questions like these must be answered daily and for that its very important to investigate metric spike.

Task : To provide a detailed report for the below two operations mentioning the answers for the related questions:

Case Study 1 (Job Data)

- **Number of jobs reviewed:** Amount of jobs reviewed over time.
  Task: Calculate the number of jobs reviewed per hour per day for November 2020?
- **Throughput:** It is the no. of events happening per second.
  Task: Let's say the above metric is called throughput. Calculate 7 day rolling average of throughput.
- **Percentage share of each language:** Share of each language for different contents.
  Task: Calculate the percentage share of each language in the last 30 days?
- **Duplicate rows:** Rows that have the same value present in them.
  Task: Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

Case Study 2 (Investigating metric spike)

- **User Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service.
  Task: Calculate the weekly user engagement?
- **User Growth:** Amount of users growing over time for a product.
  Task: Calculate the user growth for product?
- **Weekly Retention:** Users getting retained weekly after signing-up for a product.
  Task: Calculate the weekly retention of users-sign up cohort?
- **Weekly Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.
  Task: Calculate the weekly engagement per device?
- **Email Engagement:** Users engaging with the email service.
  Task: Calculate the email engagement metrics?

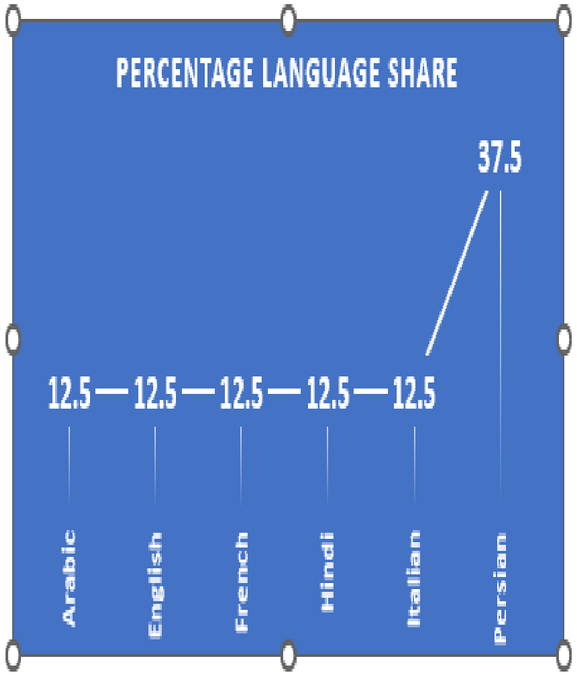
JOBS REVIEWED

DATE	JOBS PER DAY	HOURS SPENT
25/11/2020	1	0.01
26/11/2020	1	0.02
27/11/2020	1	0.03
28/11/2020	2	0.01
29/11/2020	1	0.01
30/11/2020	2	0.01

THROUGHPUT 7DAY ROLLING AVERAGE

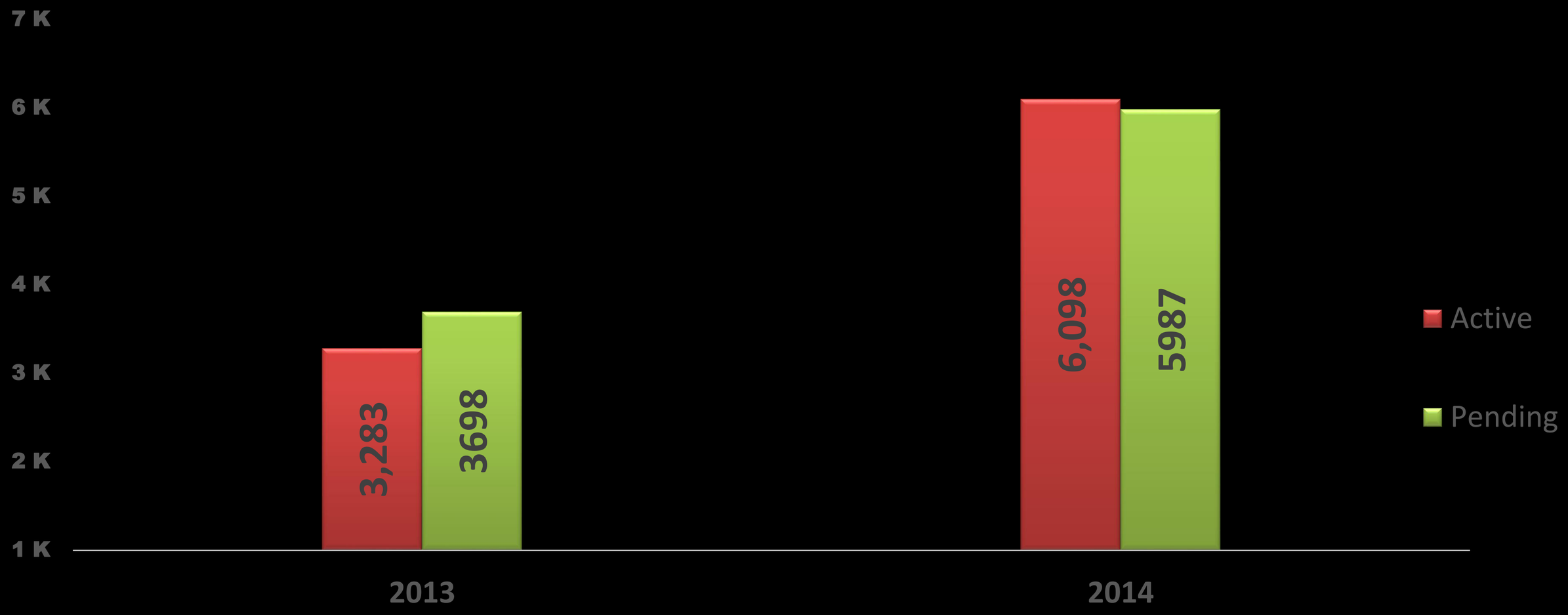
THROUGHPUT	JOB ID	DATE	AVERAGE
0.02	20	25/11/2020	0.02
0.02	23	26/11/2020	0.02
0.01	11	27/11/2020	0.01
0.09	25	28/11/2020	0.09
0.05	23	28/11/2020	0.03
0.05	23	29/11/2020	0.04
0.07	21	30/11/2020	0.07
0.04	22	30/11/2020	0.04

LANGUAGE SHARE

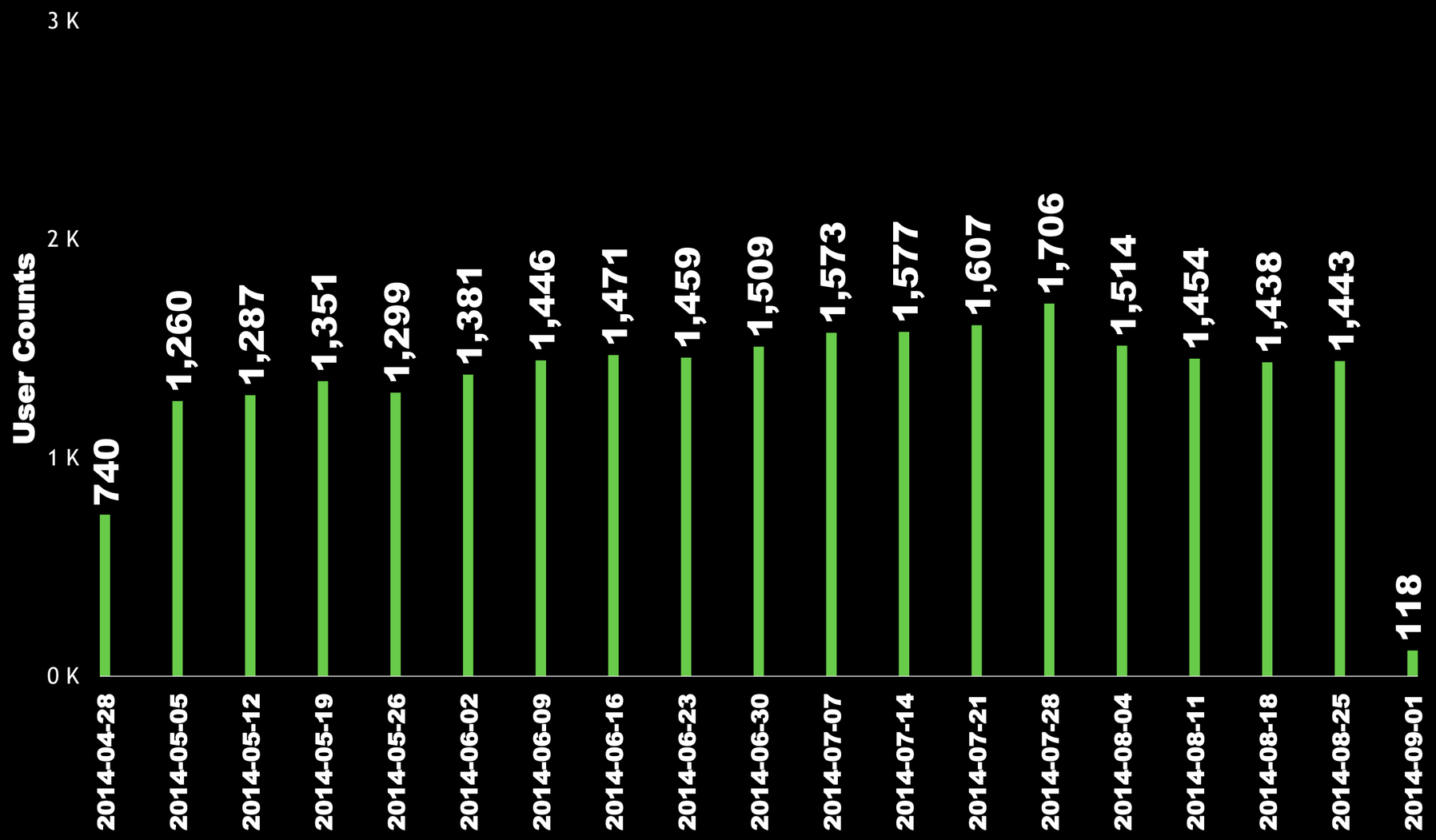


The dashboard showcase key metrics like jobs reviewed per day per hour ,throughput and language percentage share. These metrics are key in identifying the operations.

YEARLY ACTIVE AND PENDING USER REGISTRRATIONS



WEEKLY USER ENGAGEMENT



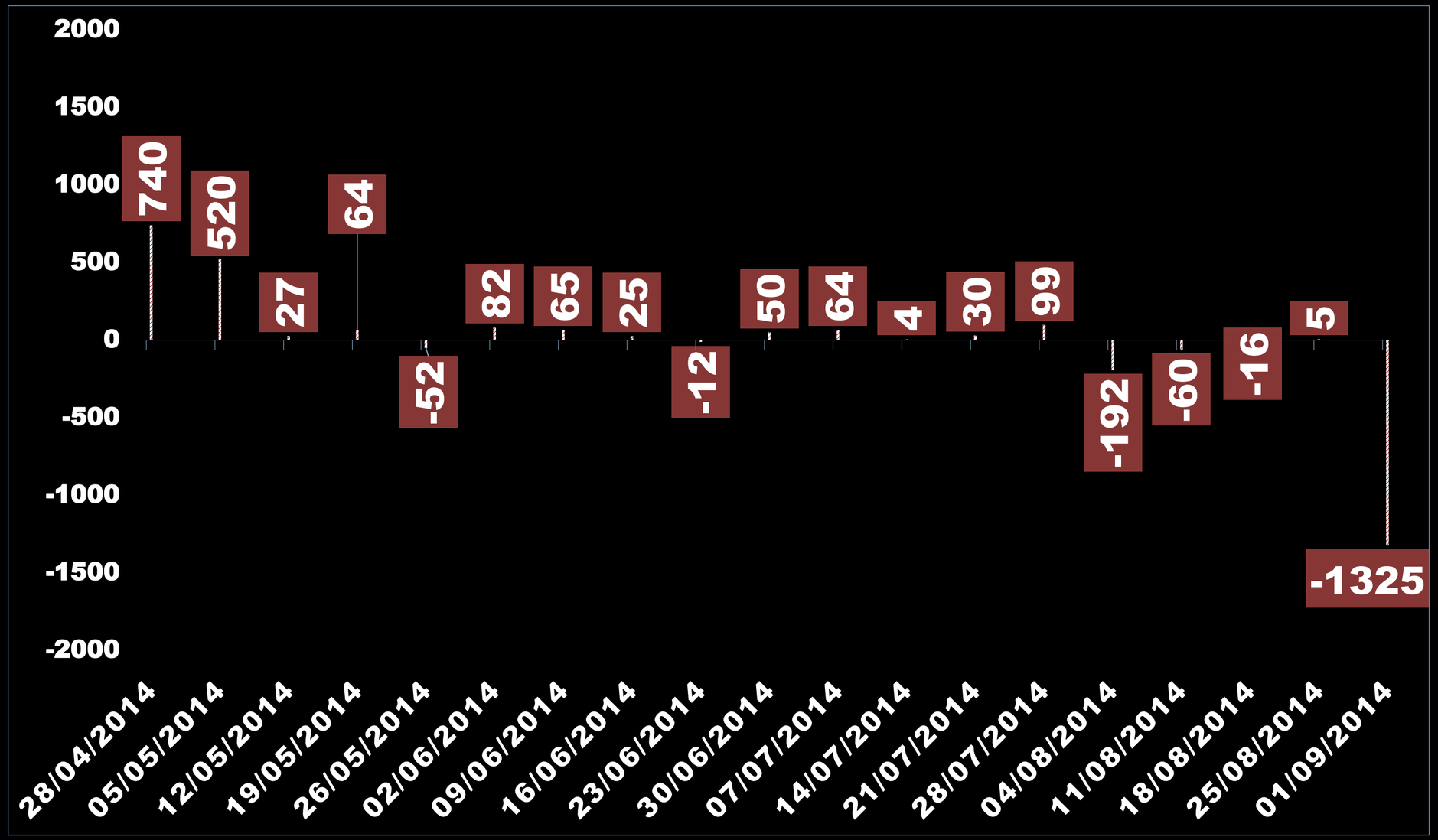
- Regular user engagements are being observed on a weekly basis for the year of 2014
- No user engagements has been observed for the year of 2013

WEEKLY RETENTION USER-SIGNUP COHORT

Signed-up Week	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2014-04-28	740	64%	44%	34%	28%	25%	23%	20%	20%	20%	18%	18%	18%	19%	16%	12%	11%	10%	1%
2014-05-05	788	46%	33%	26%	21%	19%	18%	16%	14%	15%	13%	15%	16%	14%	12%	11%	9%	1%	0%
2014-05-12	601	47%	29%	25%	19%	16%	15%	13%	16%	14%	11%	11%	10%	7%	8%	8%	0%	0%	0%
2014-05-19	555	40%	30%	22%	16%	13%	11%	12%	11%	12%	12%	7%	7%	6%	7%	0%	0%	0%	0%
2014-05-26	495	38%	26%	18%	15%	13%	15%	15%	12%	10%	9%	8%	7%	6%	0%	0%	0%	0%	0%
2014-06-02	521	43%	29%	21%	17%	14%	12%	12%	11%	9%	8%	7%	6%	0%	0%	0%	0%	0%	0%
2014-06-09	542	40%	25%	19%	17%	15%	13%	11%	10%	9%	6%	6%	0%	0%	0%	0%	0%	0%	0%
2014-06-16	535	38%	27%	19%	15%	12%	12%	11%	7%	7%	5%	0%	0%	0%	0%	0%	0%	0%	0%
2014-06-23	500	44%	28%	20%	15%	13%	10%	9%	8%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-06-30	495	37%	23%	17%	15%	11%	9%	9%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-07-07	493	40%	25%	22%	14%	11%	8%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-07-14	486	40%	23%	14%	9%	6%	6%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-07-21	501	37%	20%	13%	9%	8%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-07-28	533	38%	23%	15%	10%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-08-04	430	34%	18%	13%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-08-11	496	38%	19%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-08-18	499	40%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-08-25	518	8%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2014-09-01	32	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

The above slide showcase data as a heat map which is actually a user-signup weekly retention cohort. Cohort here means dividing data into groups which share a common experience within a defined time-span. Cohort analysis is a tool to measure user engagement over time. Customers retained over time can give insight on how actively customers are engaging with a product or how interested the customers are in a product. The above heat map showcases on the very first week 740 users signed up and the next week only 69.9% users came back/logged inn , the 10^{th} week only 18% and on the 18^{th} week only 1% users came back which means only 5 in 740 users are still active.

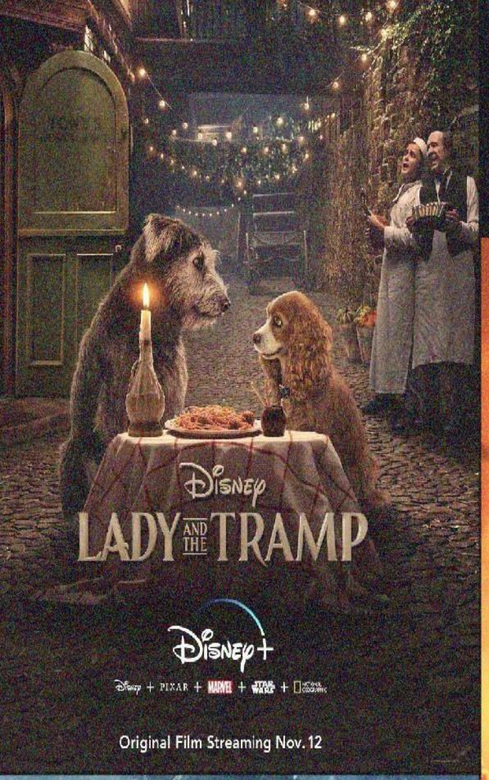
USER GROWTH OVER TIME

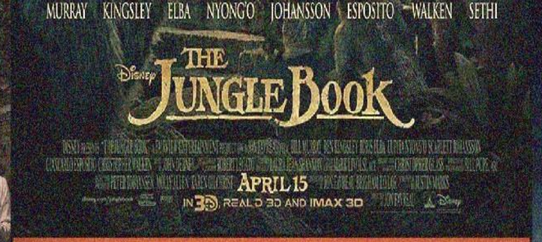


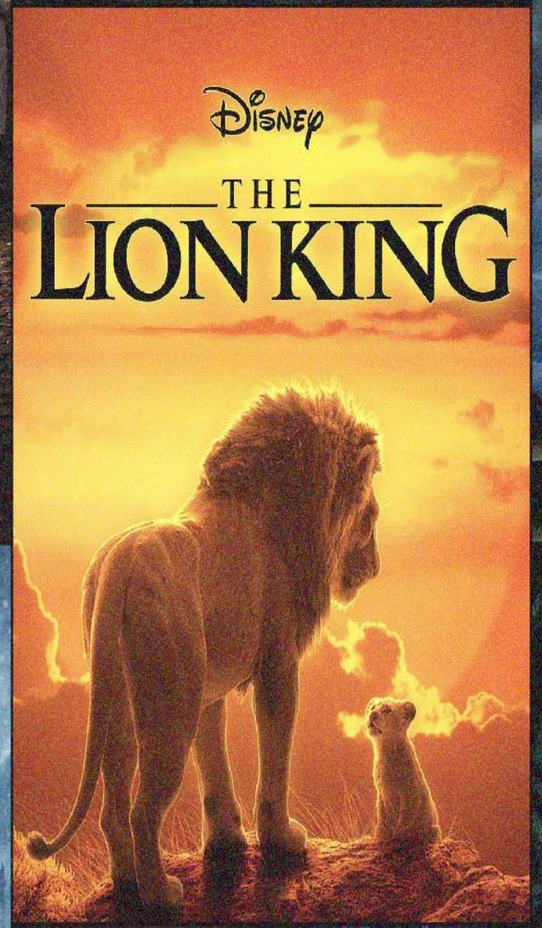
- A major dip of 192 has been observed in the week 33

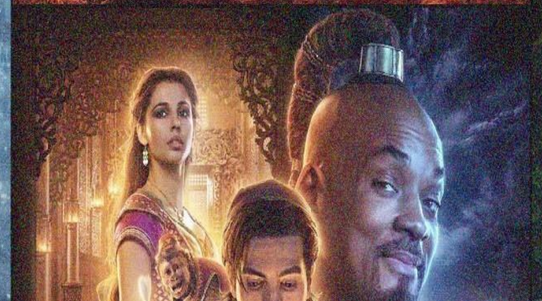
CONCLUSION

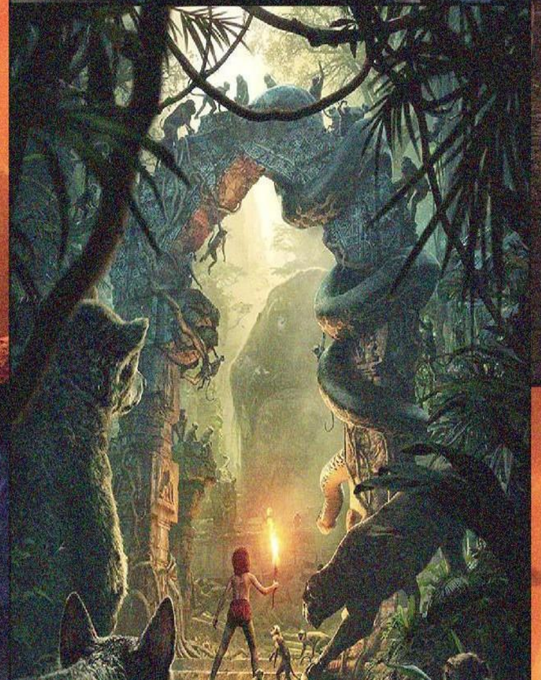
This task helps understand the advance fundamentals of structured query language (SQL).The problem statements help to dig deeper in order to draw insights. Operations metric analytics help understand key metrics like throughput and jobs reviewed where as investigating spike metrics helps detect the unusual behavior. Cohort retention analysis helps track user growth over time. Weekly user engagement and user growth over time helps understand the user behavior.

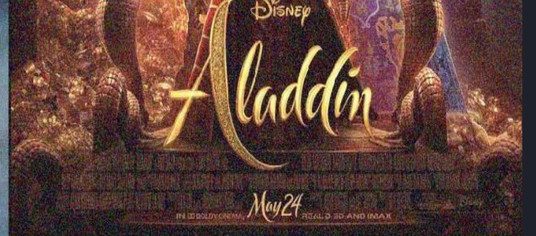


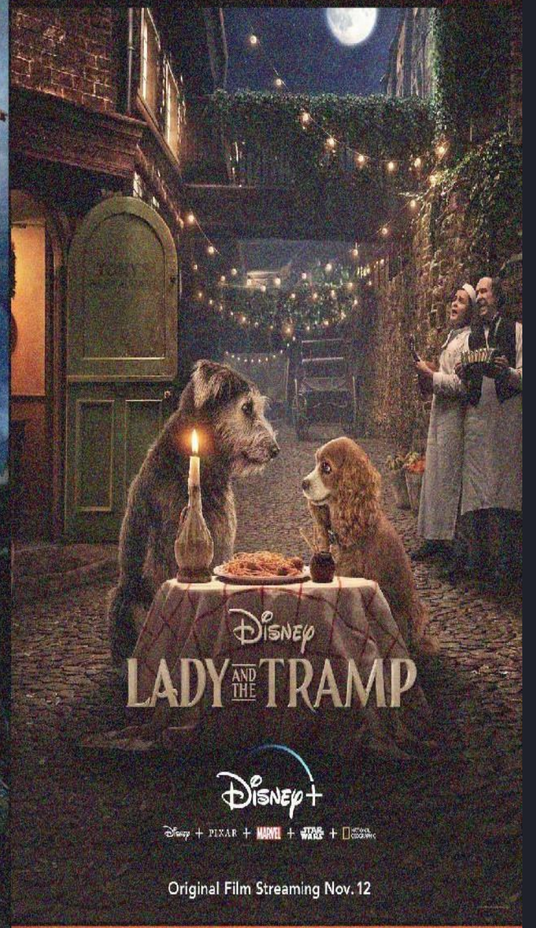


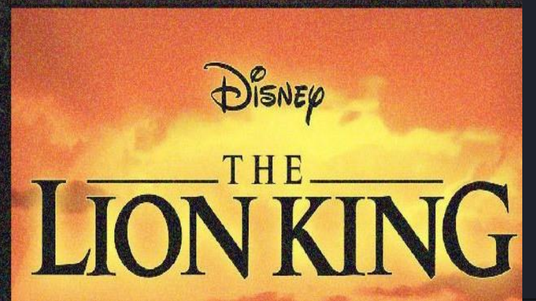










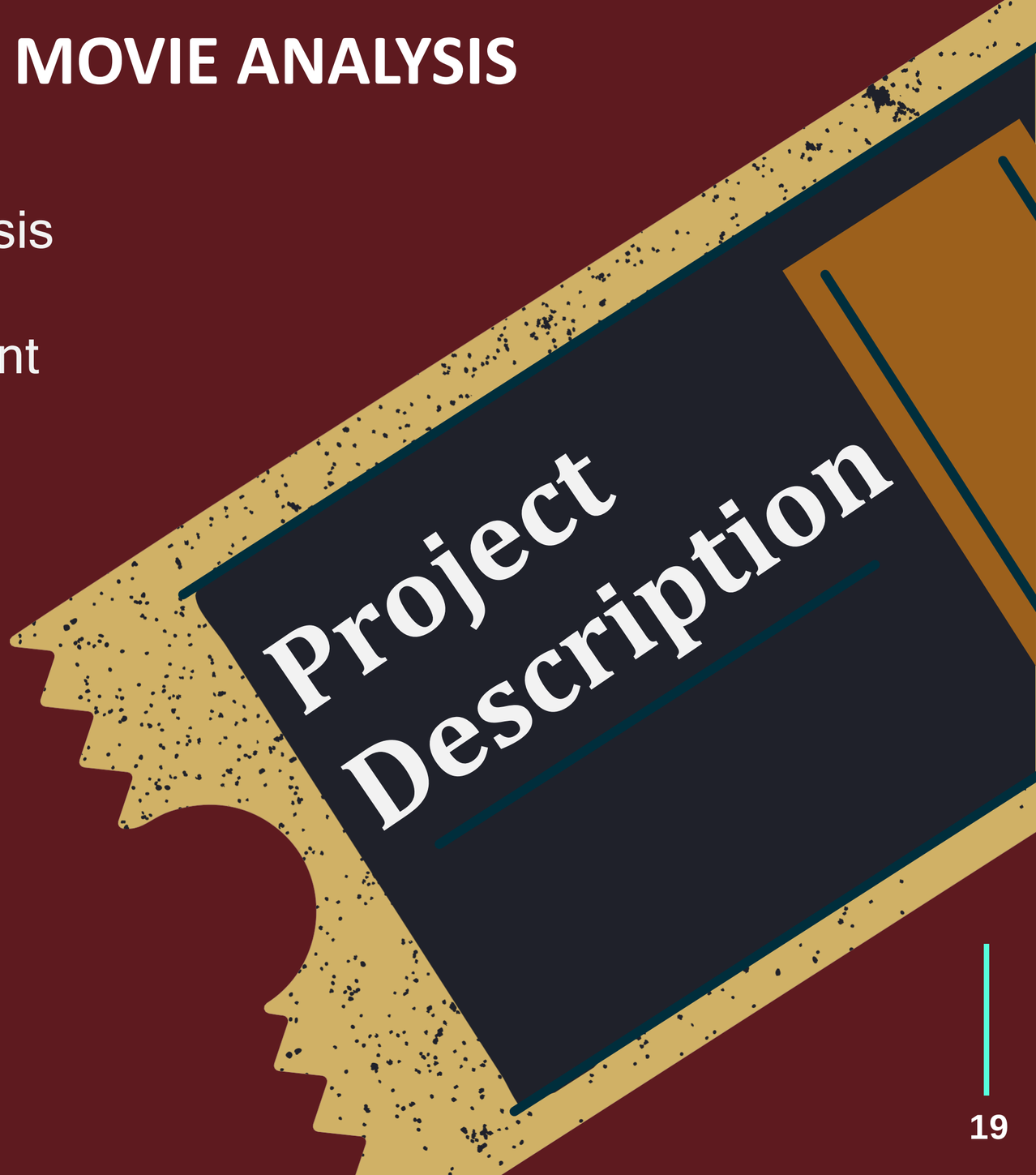


IMDB MOVIE ANALYSIS

PROJECT 4: IMDB MOVIE ANALYSIS

This project is focused on carrying out analysis an which will help bring insights relevant to IMDB. The project will answer to the important questions like :

- ❑ Cleaning Data
- ❑ Movies With Highest Profit
- ❑ Top Movies With Best IMDB Rating Score
- ❑ Best Directors
- ❑ Movie Releases Over Decades
- ❑ User Rating Over Decades
- ❑ Critic Favorite Actors
- ❑ Audience Favorite Actors



APPROACH

First I calculated the number of null values present in each column using COUNTA function. As the percentage were less than 30% ,instead of removing those null values I performed missing value treatment using mean ,median and mode. After that I removed the duplicate values. After cleaning the data, analysis was carried out to in order to generate insights. I have also attached the excel sheet for the reference.

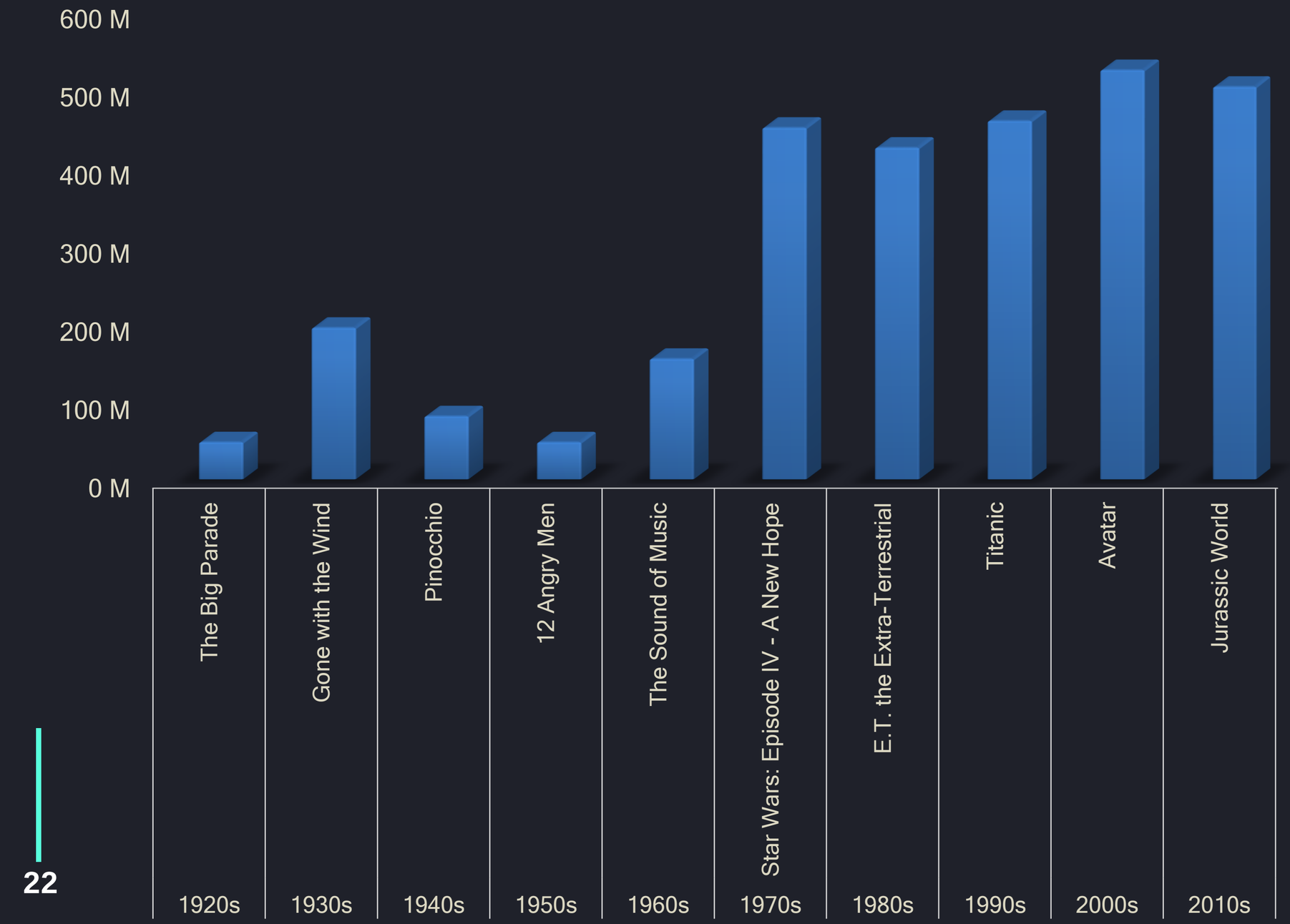
Avatar has the highest profit of 523 Million.

Jurassic World made a profit of 502 Million with the budget of 150 Million.

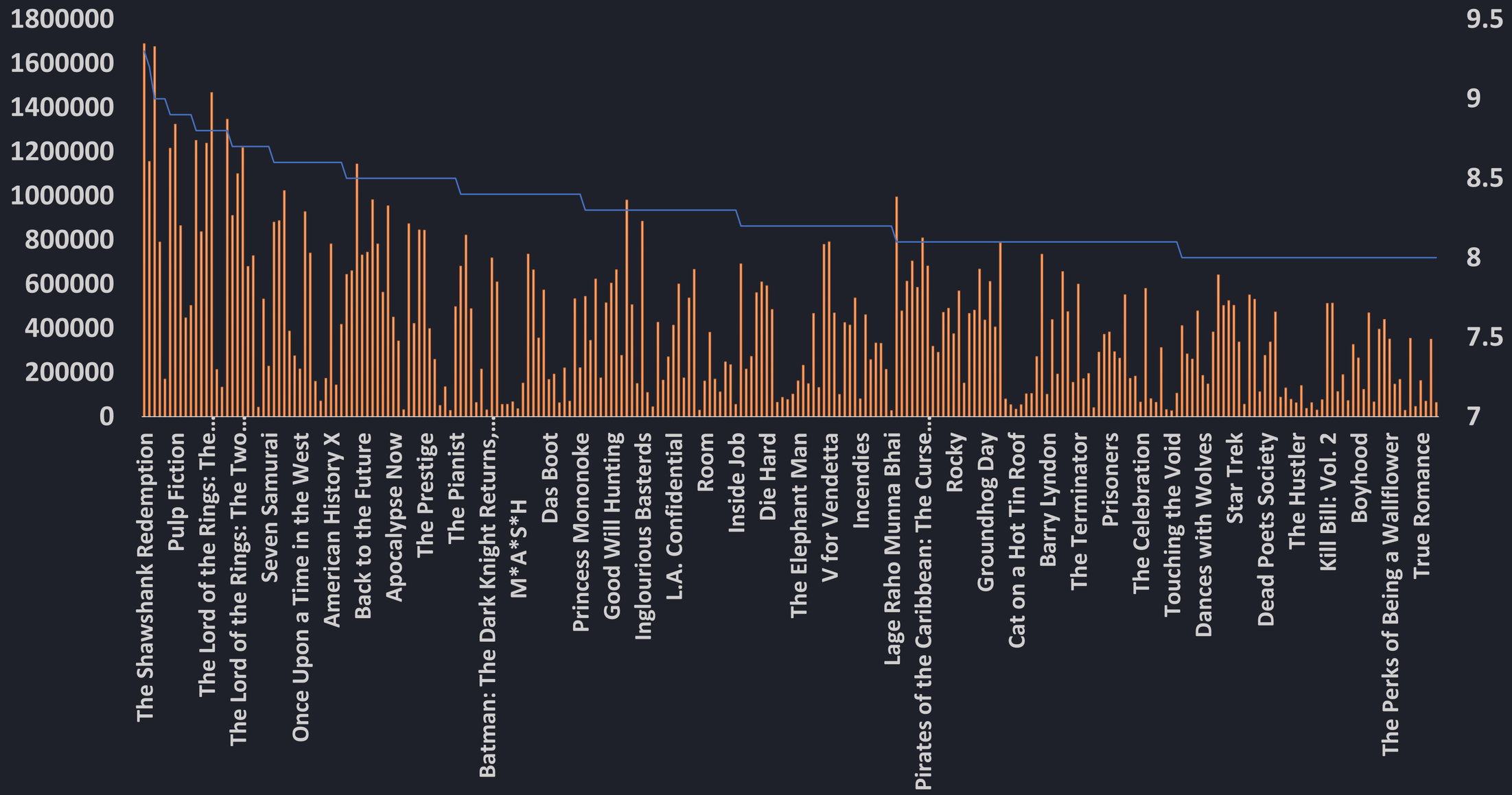
Titanic being the most profitable movie of the 1990s.

Star Wars: Episode IV - A New Hope made a profit of 450 Million despite having the least budget amongst the top three profitable movies.

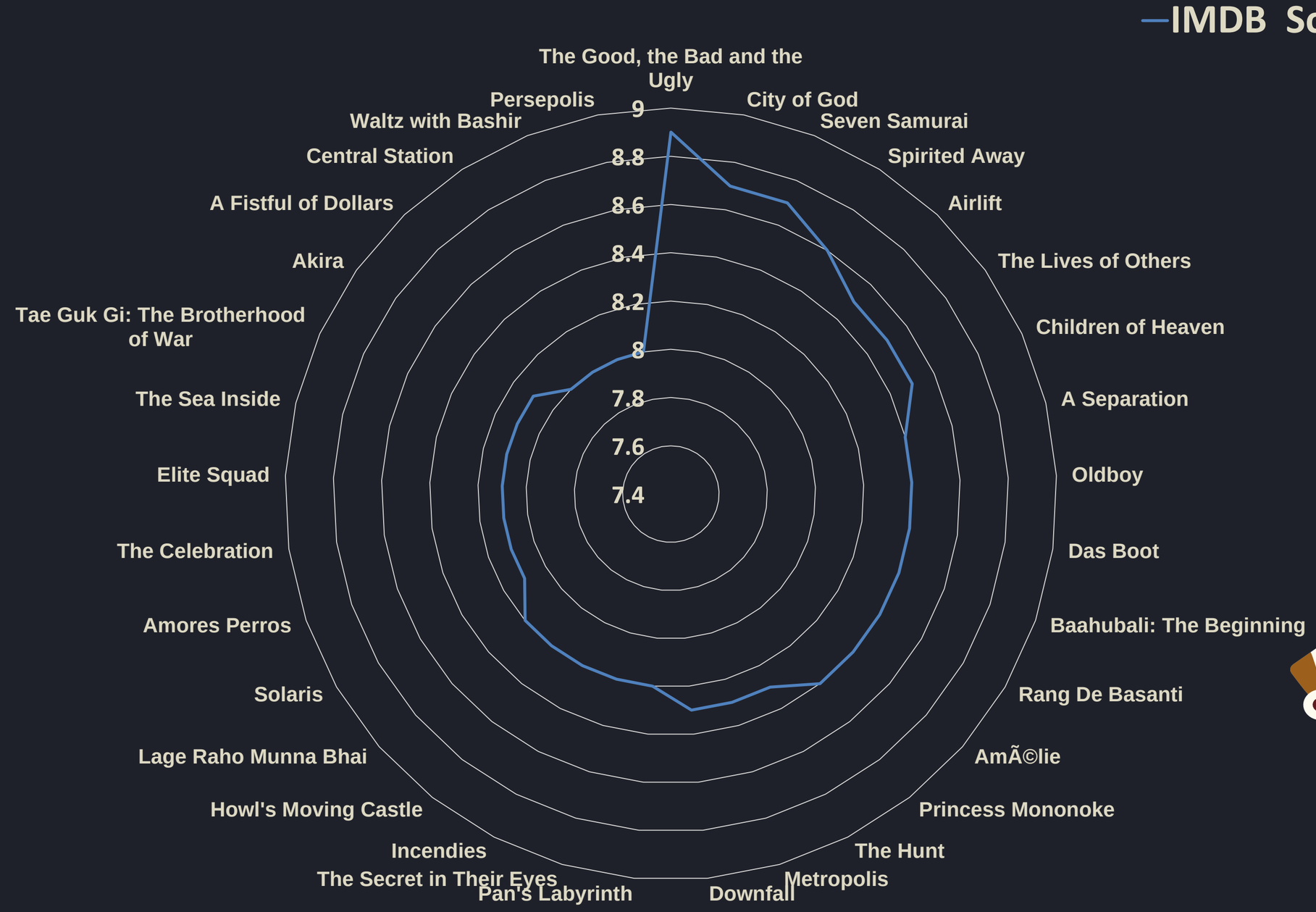
Most Profitable Movies Over Each Decades



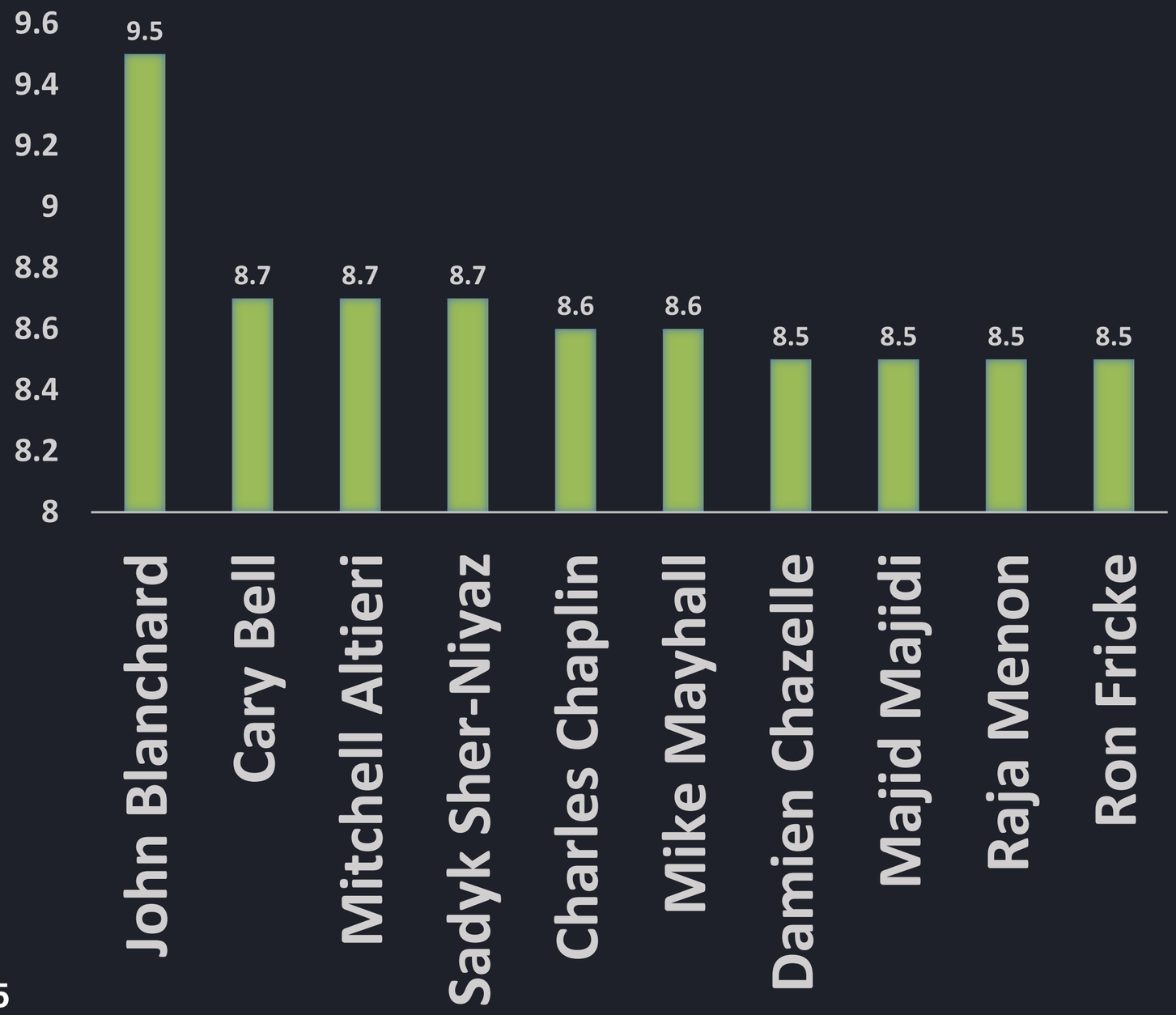
Top 250 IMDB Movie Based On IMDB Score



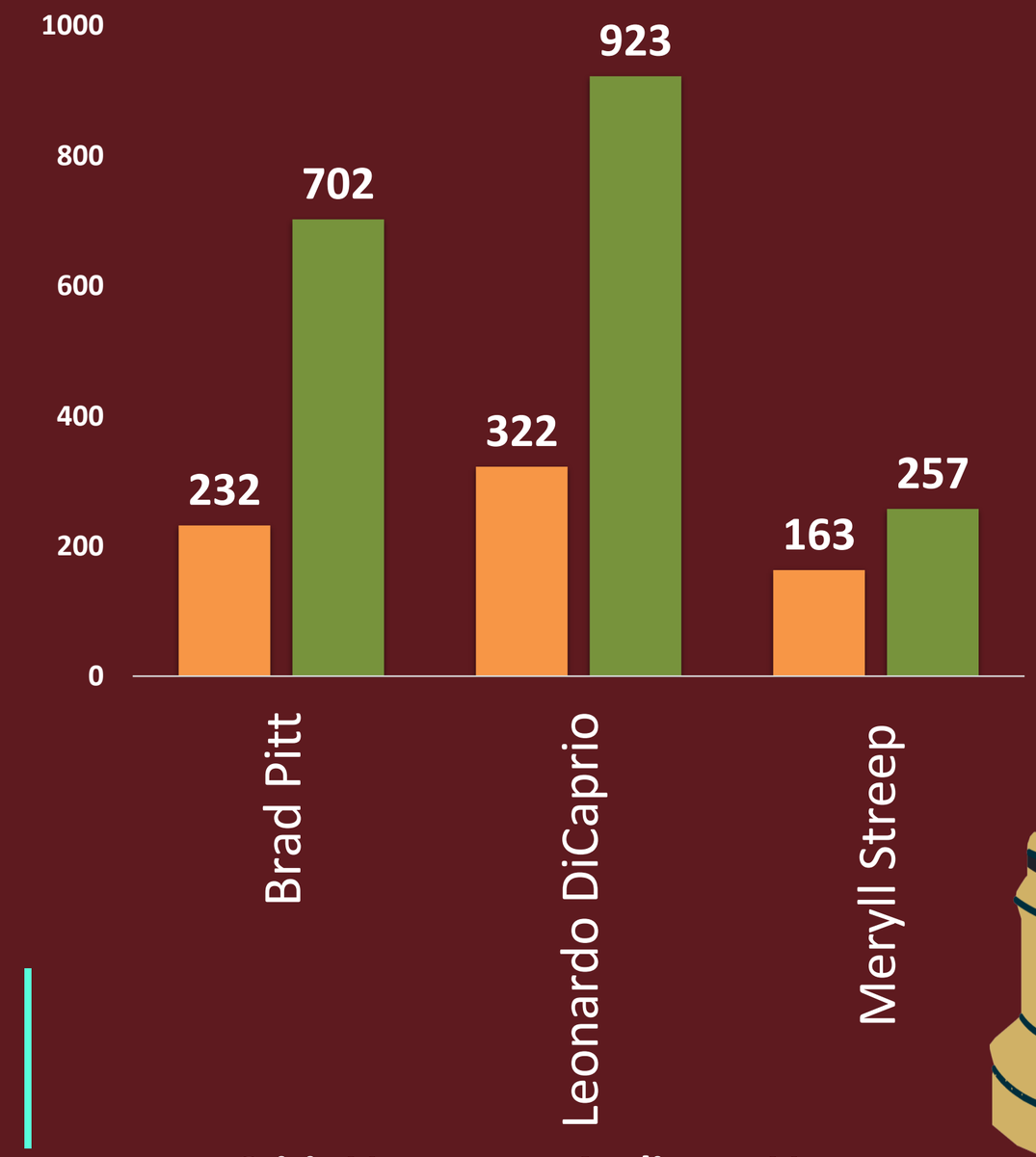
Top Foreign Language Movies



TOP 10 DIRECTORS WITH HIGHEST MEAN IMDB SCORE

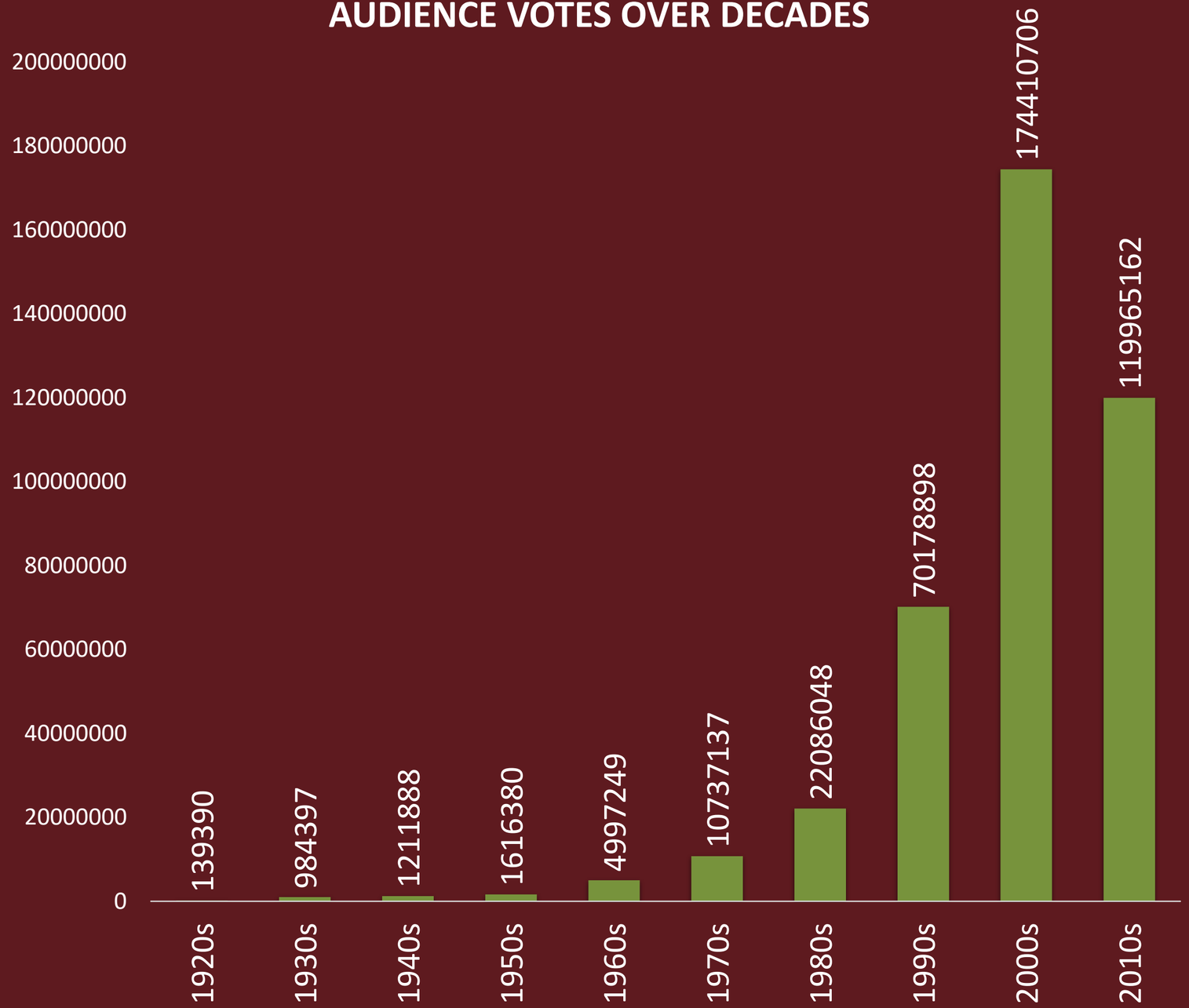


CRITIC VS AUDIENCE FAVOURITE ACTOR



Leonardo DiCaprio being the both audience and critic favorite as per the score calculated.

AUDIENCE VOTES OVER DECADES



Audience votes are seen to be exponentially increasing over decades indicating their keen interest towards movies.

CONCLUSION

The project successfully brings out the required insights. The insights drawn were important and answers the required questions. Attached are the links of the SQL queries and excel analysis files.

PROJECT 5 : HIRING PROCESS ANALYTICS

This project aims to carry out the in-depth analysis of major underlying trends about the hiring process and performing exploratory data analysis on the dataset as well.

The project will answer to the important questions like :

Hiring
Average Salary
Class Intervals
Drawing Charts and Plots
Post Tiers

For EDA Perform below steps

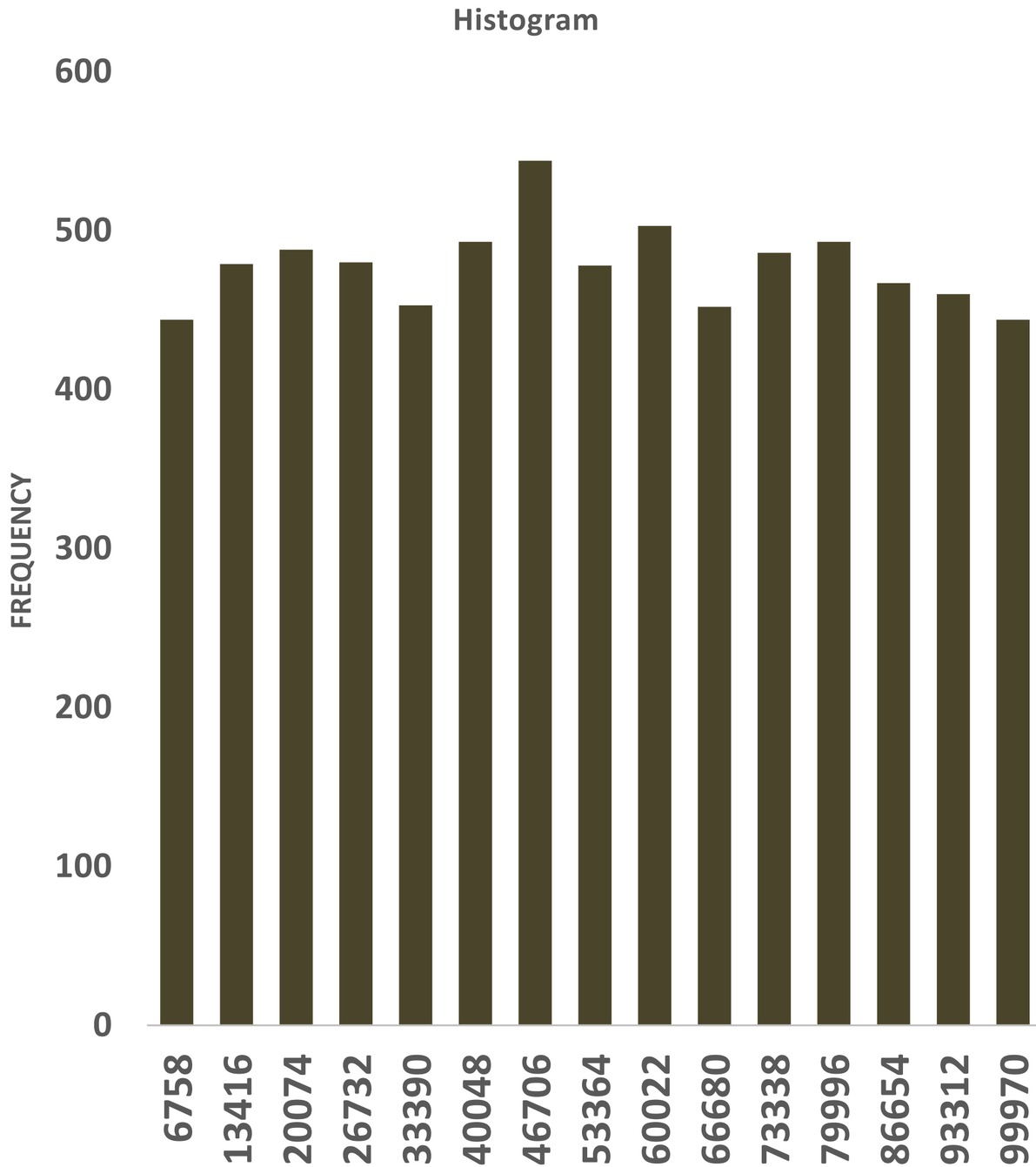
Understanding data columns and data
Checking for missing data
Clubbing columns with multiple categories
Checking for outliers
Removing outliers
Drawing Data Summary

APPROACH

First, I carried out the exploratory data analysis using excel data analysis tool pack. Calculated descriptive statistics on salary offered column. Find out missing values and outliers using quartile function in excel. After removing the outliers further insights were carried out using excel formulas and graphs.

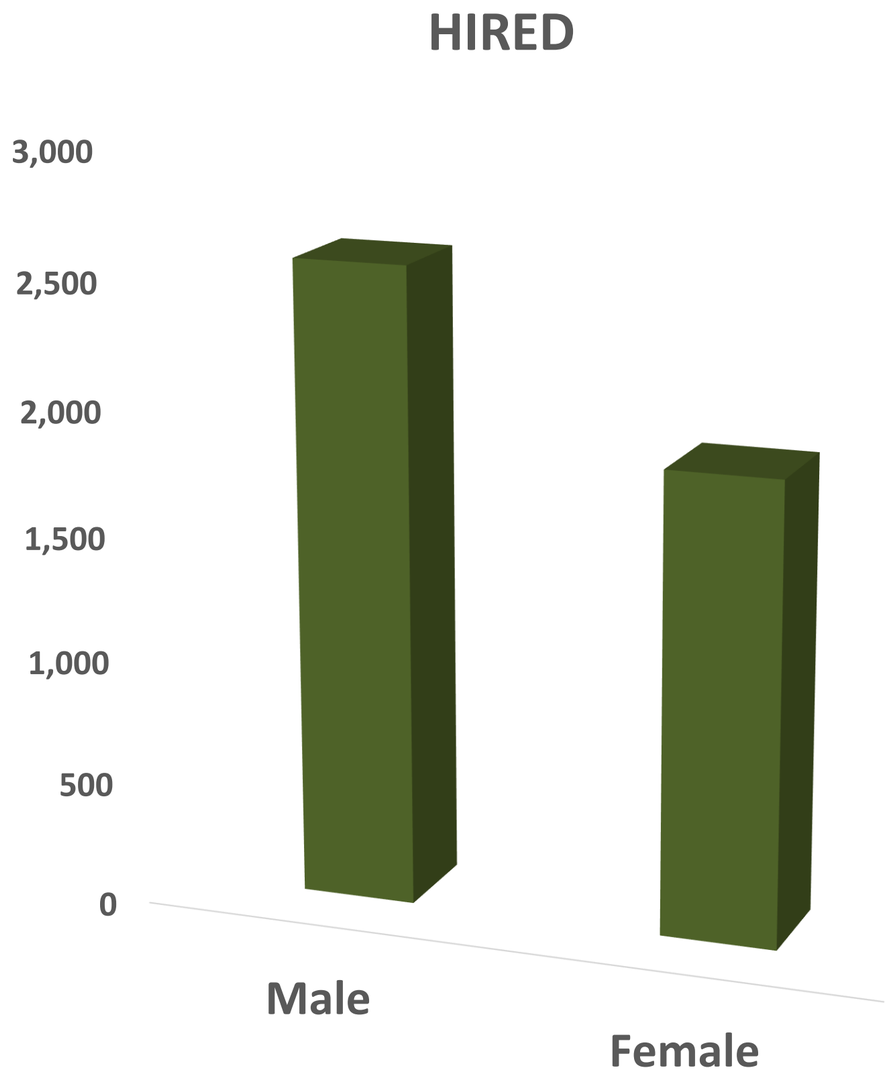
SALARY CLASS INTERVAL

- Histogram shows the salary class intervals.
- Majority of applicants were offered salary between 40K – 47K.

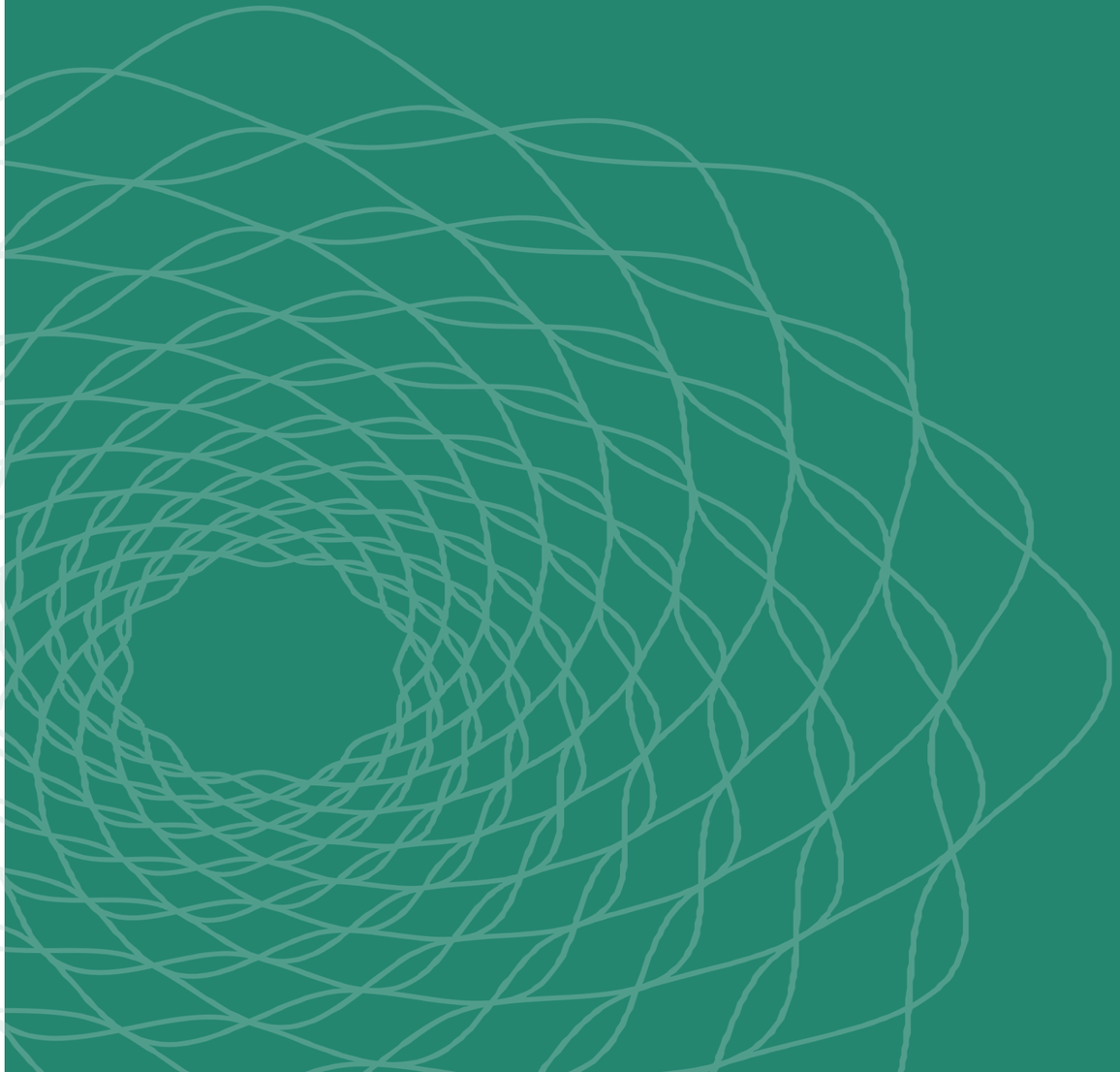


HIRING

- 2562 Males are hired
- 1854 Females are hired



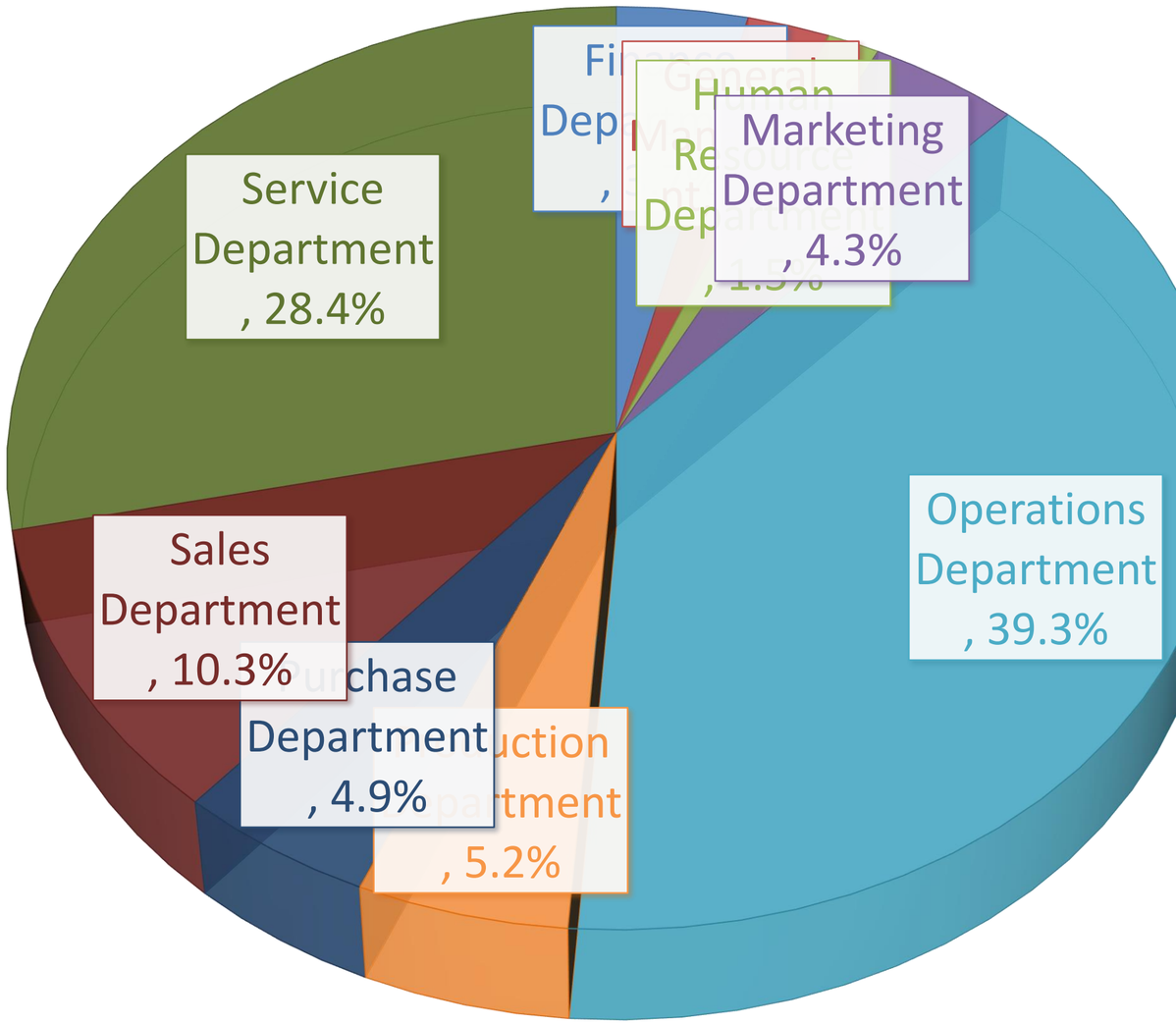
AVERAGE SALARY



Average salary offered in this company is 49,878

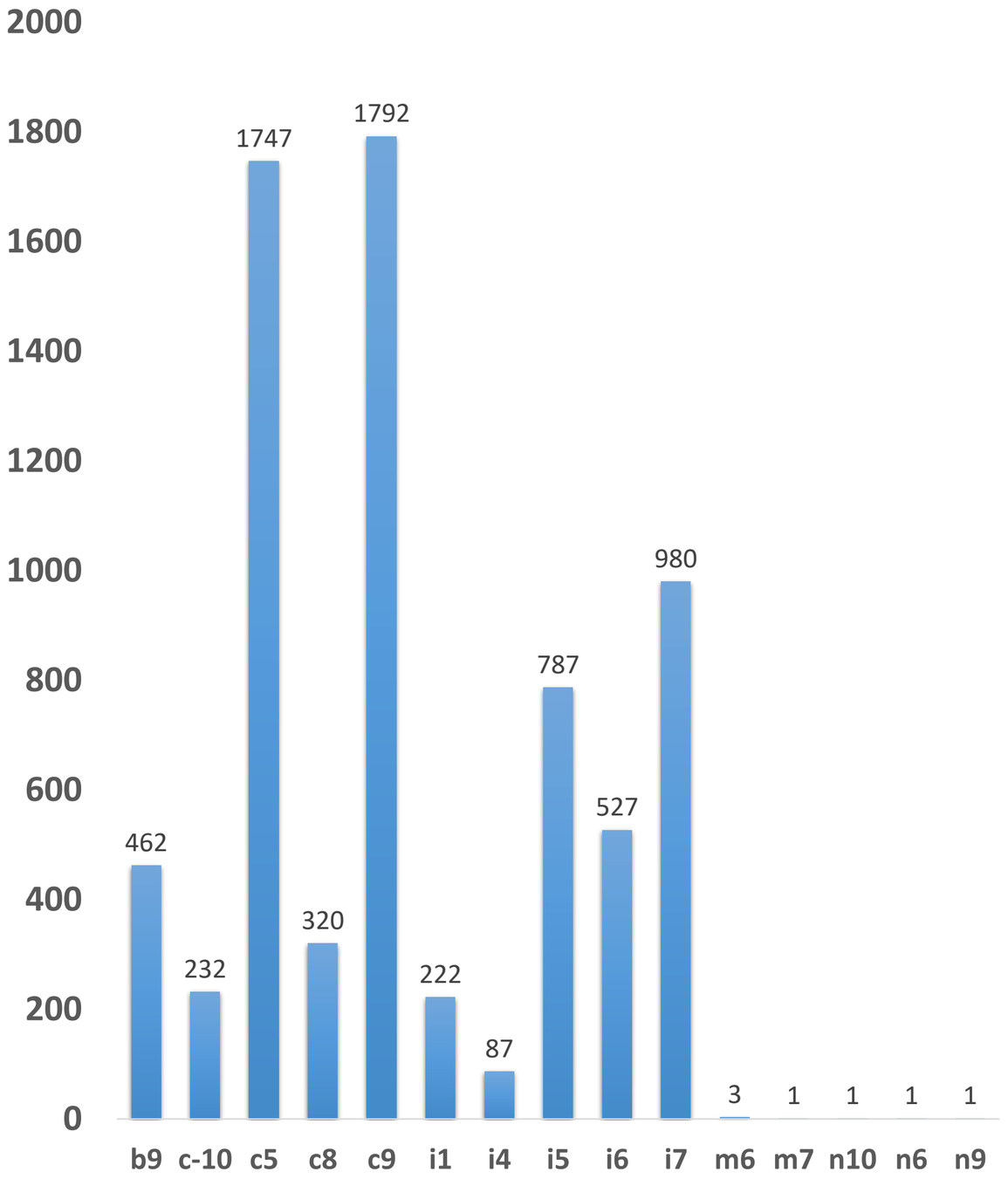
DEPARTMENT PROPORTION

- 4694 applicants were hired out of 7164 application ids.
- Majority of applicants were hired in operations department.
- The pie chart represents the proportion of applicants working in different departments.



TIERS

Total of 15 different tiers are present in this company.



CONCLUSION

This project has helped me to understand the concept of exploratory data analysis.

This project helps understand the process of data cleaning and how important it is to handle outliers in a given dataset.

This project has lead me to understand the analysis required in the company's hiring process.

PROJECT 6 : BANK LOAN CASE STUDY

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study we will develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers. When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Task : You are required to provide a detailed report for the below data record mentioning the answer to the questions that follows:

- Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value).
- Identify if there are outliers in the dataset. Also, mention why do you think it is an outlier. Again, remember that for this exercise, it is not necessary to remove any data points.
- Identify if there is data imbalance in the data. Find the ratio of data imbalance.
- Find the top 10 correlation for the Client with payment difficulties and all other cases (Target variable). Note that you have to find the top correlation by segmenting the data frame w.r.t to the target variable and then find the top correlation for each of the segmented data and find if any insight is there. Say, there are 5+1(target) variables in a dataset: Var1, Var2, Var3, Var4, Var5, Target. And if you have to find top 3 correlation, it can be: Var1 & Var2, Var2 & Var3, Var1 & Var3. Target variable will not feature in this correlation as it is a categorical variable and not a continuous variable which is increasing or decreasing.

APPROACH

I have used COUNTA function to count the total rows in each column. After that I have found the percentage of null values in each column using the formula 1- (Total Row Counts for each columns /Total Row Counts). After that I have removed all the columns having null value percentages more than 30%. For column having less than 30% null value percentages I have done mean, median and mode imputations for the missing values for columns having null value percentages less than 30%. I have also found the outliers using interquartile range method considering relevant columns. After going through each column description, I have kept only relevant columns to bring out the insights. The columns having days are converted in to years by simply dividing the days by 365.Click on the below link to open the excel file.

CONCLUSION

This project helps in handling the large datasets. How exploratory data analysis can be applied to large datasets. When dealing with the large datasets it is also important to select only those columns which are very useful to our analysis. Finding correlations between columns can become very convenient while dealing with large datasets as it saves time selecting which columns should be considered for analysis. The project also help in understanding the various terminologies used in the banking domain. The insight drawn from the project are as follows:

- Applicants drawing higher income were offered higher loan amount by the bank.
- Majority of applicants drawn an income range between 1.25 Lacs - 1.5 Lacs , also the defaults drawn the income between the same range .
- Majority of applicants were offered in the credit range of 9 Lacs and above.

PROJECT 7 : ADS AIRING REPORT

PROJECT DESCRIPTION

This project aims at drawing insights from a given data set of Ads Airing. The project is a detailed study of the TV ads airing brand and their ads placements. The detailed analysis will lead to the answers of the following questions:

PROBLEM 1

What is Pod Position? Does the Pod position number affect the amount spent on Ads for a specific period of time by a company?

PROBLEM 2

What is the share of various brands in TV airings and how has it changed from Q1 to Q4 in 2021?

PROBLEM 3

Conduct a competitive analysis for the brands and define advertisement strategy of different brands and how it differs across the brands.

PROBLEM 4

Mahindra and Mahindra wants to run a digital ad campaign to complement its existing TV ads in Q1 of 2022. Based on the data from 2021, suggest a media plan to the CMO of Mahindra and Mahindra. Which audience should they target? *Assume XYZ Ads has the ad viewership data and TV viewership for the people in India.

My approach to this project involves understanding the dataset. After understanding the dataset I have used SQL and Excel to draw insights. I have imported the excel file in SQL and extracted data by running SQL queries and I have used excel to create charts and graphs for the better understanding of the result.

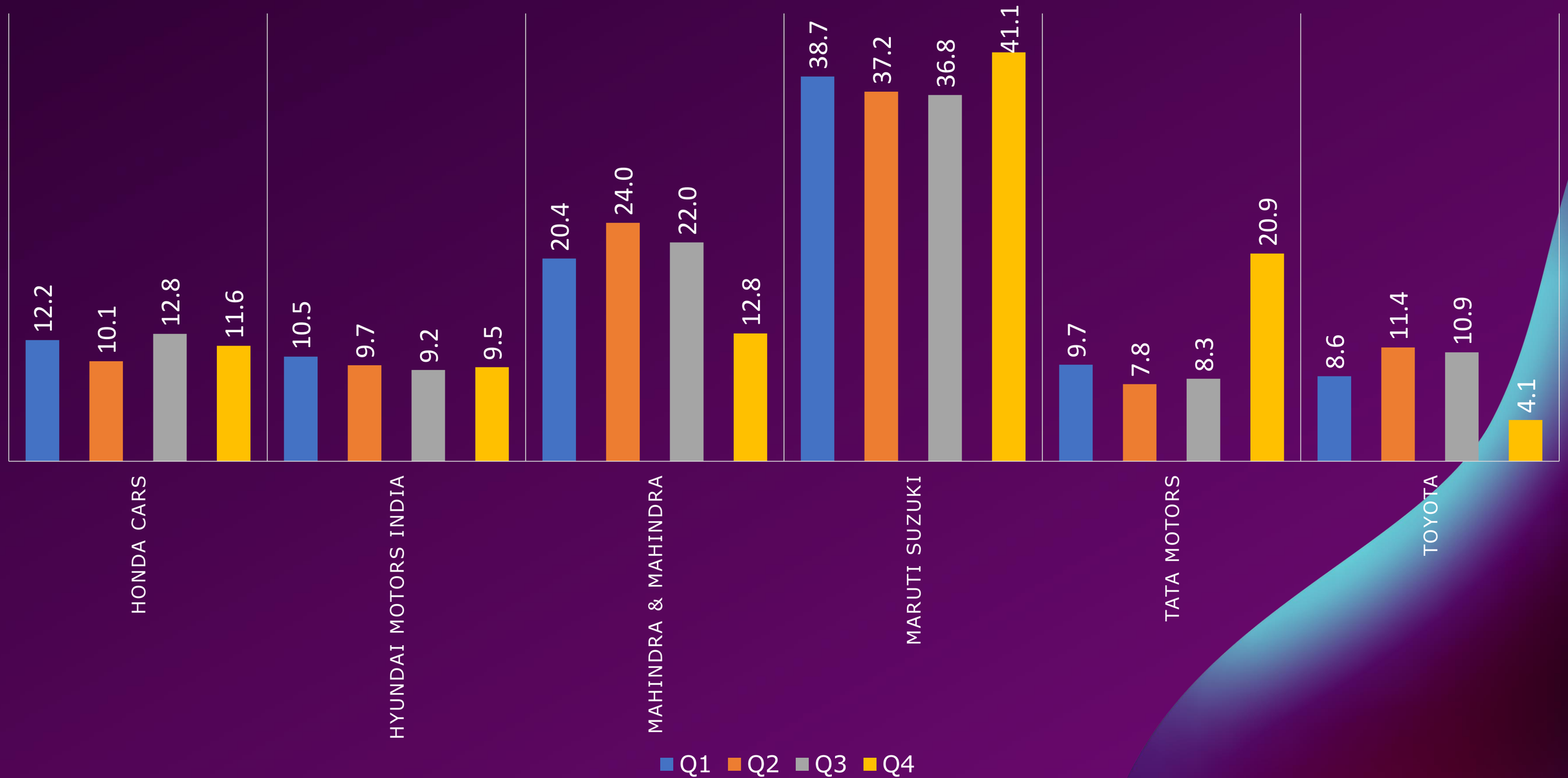
APPROACH

TOTAL SPEND PER POD POISTION BY COMPANIES IN 2021

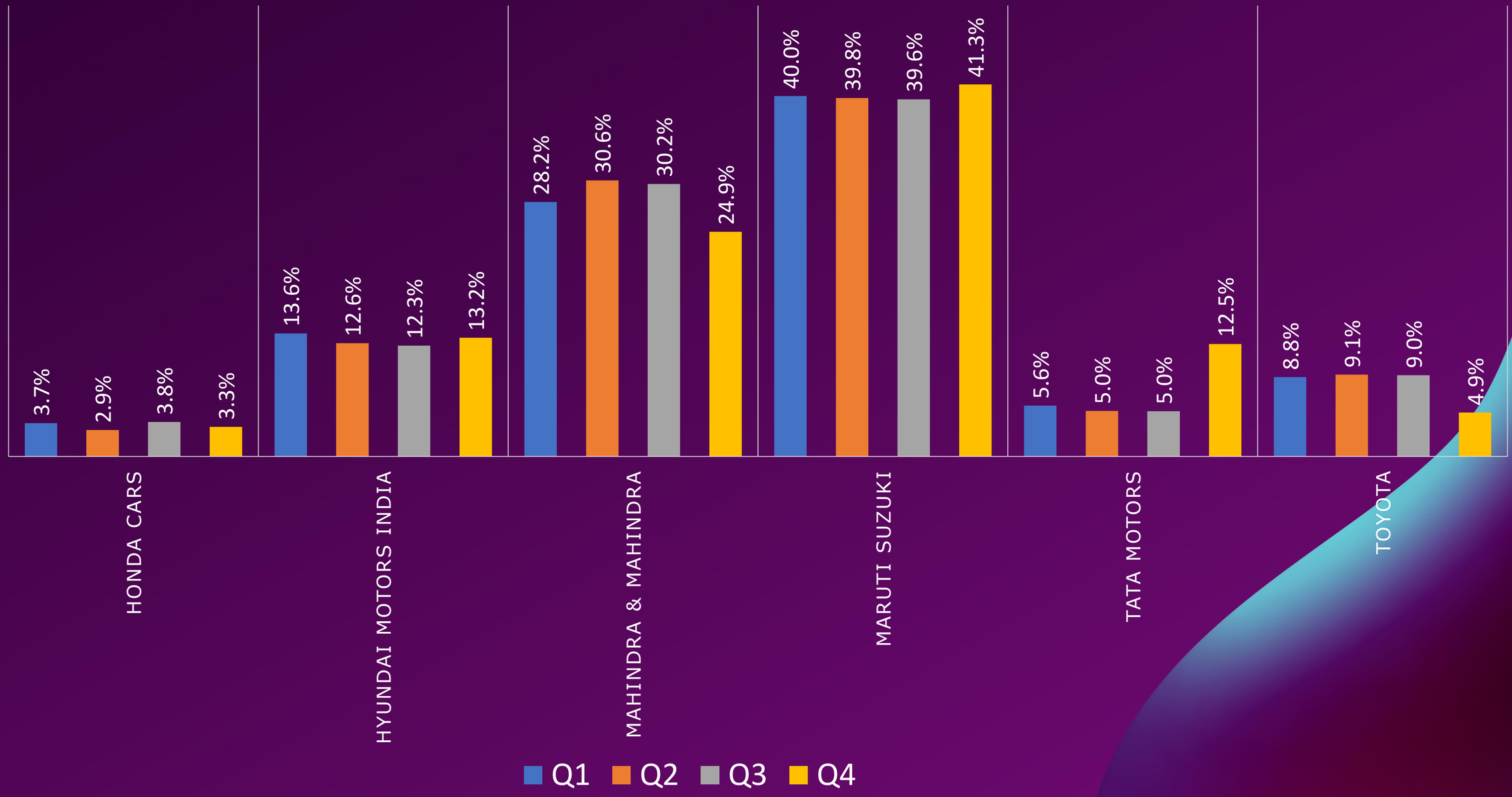
BRAND	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Maruti Suzuki	$131.2 M	$75.4 M	$78.7 M	$68.5 M	$52.0 M	$42.0 M	$30.8 M	$22.7 M	$17.0 M	$12.5 M	$8.8 M	$6.4 M	$4.1 M	$2.9 M	$1.9 M	$1.4 M	$.8 M	$.5 M	$.5 M	$.2 M	$.1 M	$.1 M	$.1 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	-	$.0 M
Mahindra and Mahindra	$97.2 M	$51.2 M	$53.6 M	$49.9 M	$40.0 M	$31.3 M	$23.0 M	$17.2 M	$11.7 M	$7.6 M	$4.9 M	$3.1 M	$2.1 M	$1.8 M	$1.0 M	$.5 M	$.4 M	$.4 M	$.2 M	$.1 M	$.1 M	$.0 M	$.0 M	$.0 M	$.0 M	$.1 M	-	$.0 M	$.0 M	-	$.0 M
Hyundai Motors India	$42.0 M	$22.8 M	$26.2 M	$27.8 M	$15.6 M	$13.0 M	$8.8 M	$7.8 M	$5.3 M	$3.2 M	$2.4 M	$1.9 M	$1.5 M	$.9 M	$.5 M	$.3 M	$.2 M	$.1 M	$.1 M	$.1 M	$.1 M	$.1 M	$.0 M	$.0 M	$.0 M	-	-	-	$.0 M	-	-
Toyota	$26.5 M	$17.6 M	$17.4 M	$14.7 M	$10.7 M	$8.0 M	$5.5 M	$3.8 M	$2.9 M	$1.9 M	$1.2 M	$.8 M	$.4 M	$.4 M	$.2 M	$.1 M	$.1 M	$.1 M	$.1 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	-	-	-	-
Tata Motors	$19.7 M	$13.6 M	$13.9 M	$11.6 M	$9.3 M	$7.8 M	$5.9 M	$4.1 M	$3.0 M	$1.9 M	$1.4 M	$1.0 M	$.6 M	$.4 M	$.2 M	$.1 M	$.1 M	$.1 M	$.0 M	$.1 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	-	$.0 M	-
Honda Cars	$7.5 M	$7.2 M	$6.7 M	$6.1 M	$4.5 M	$3.9 M	$3.2 M	$2.5 M	$2.1 M	$1.5 M	$1.0 M	$.7 M	$.5 M	$.3 M	$.2 M	$.1 M	$.1 M	$.1 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M	$.0 M

Pod Position can be defined as the placement/position of an individual advertisement within a certain commercial duration in a show. Yes the pod position does affect the total amount spent on ads over time (1 Year) by a company. The heat map in the above slide clearly depicts that companies has spent most on the pod position 1 airing on multiple shows over the span of 1 year. As evident from the heat map the maximum amount has been spent on pod position 1 by each companies. With the increasing pod position the total amount spent is clearly reducing. Thus proving ads which are placed first are paid more attention.

BRANDS QUATERLY ADS SHARE IN 2021

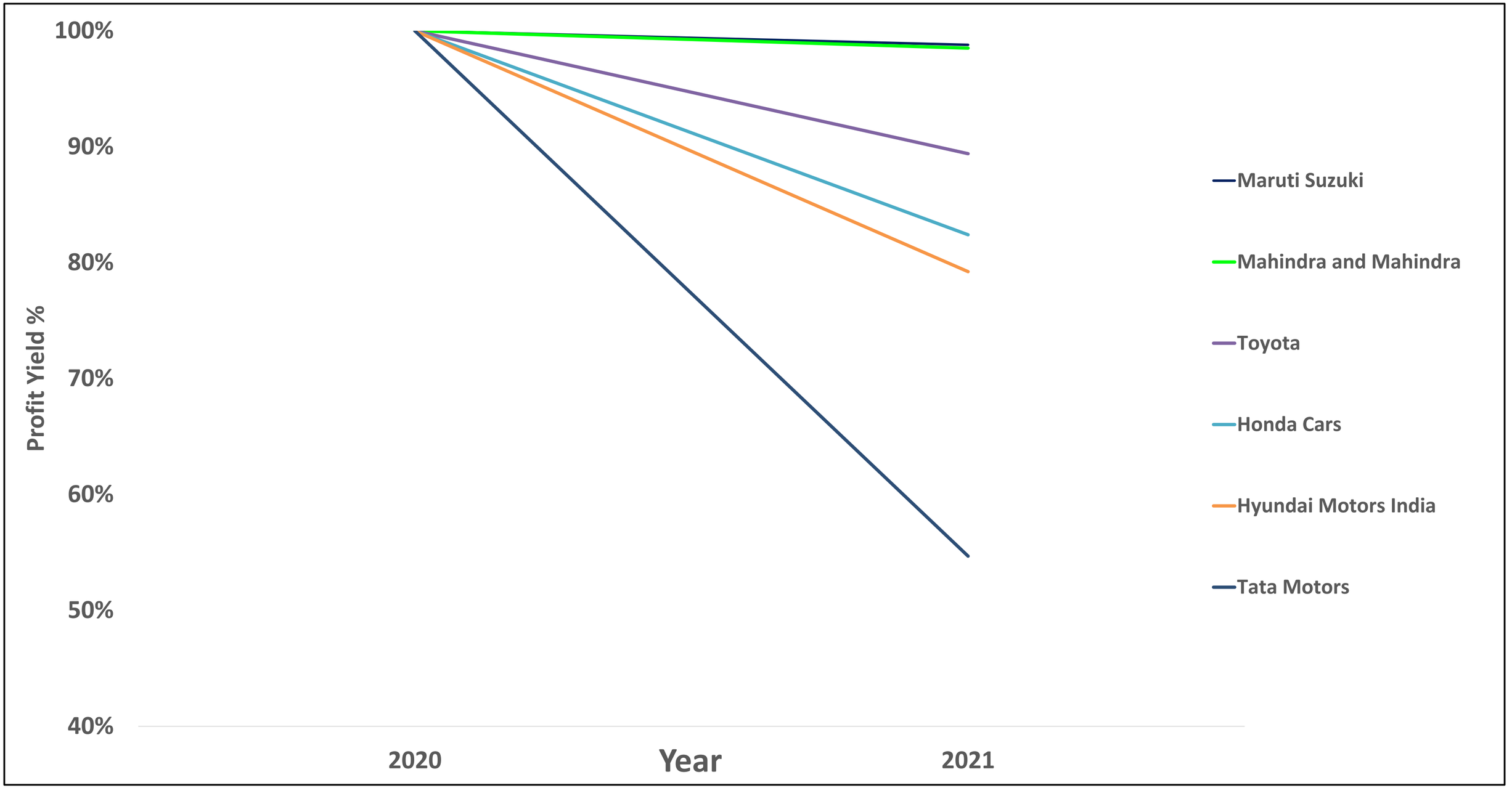


BRANDS QUATERLY SPEND SHARES IN 2021



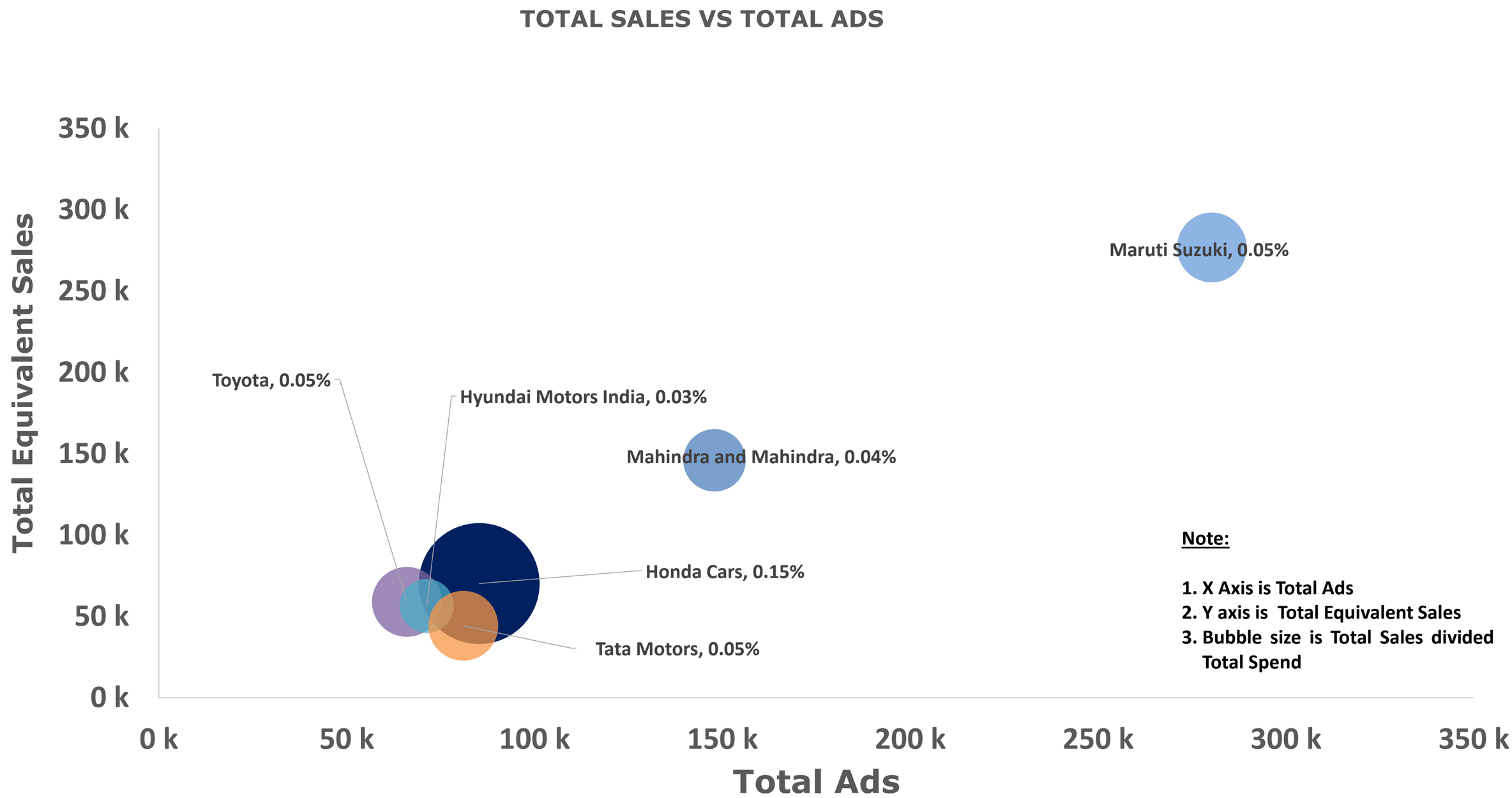
If we look at the above two slides we can bring out the analogy that Brand Maruti Suzuki has spend on the ads proportionally to the percentage of aids airing on multiple shows.

% PROFIT YIELD BRAND CURVES YEAR WISE

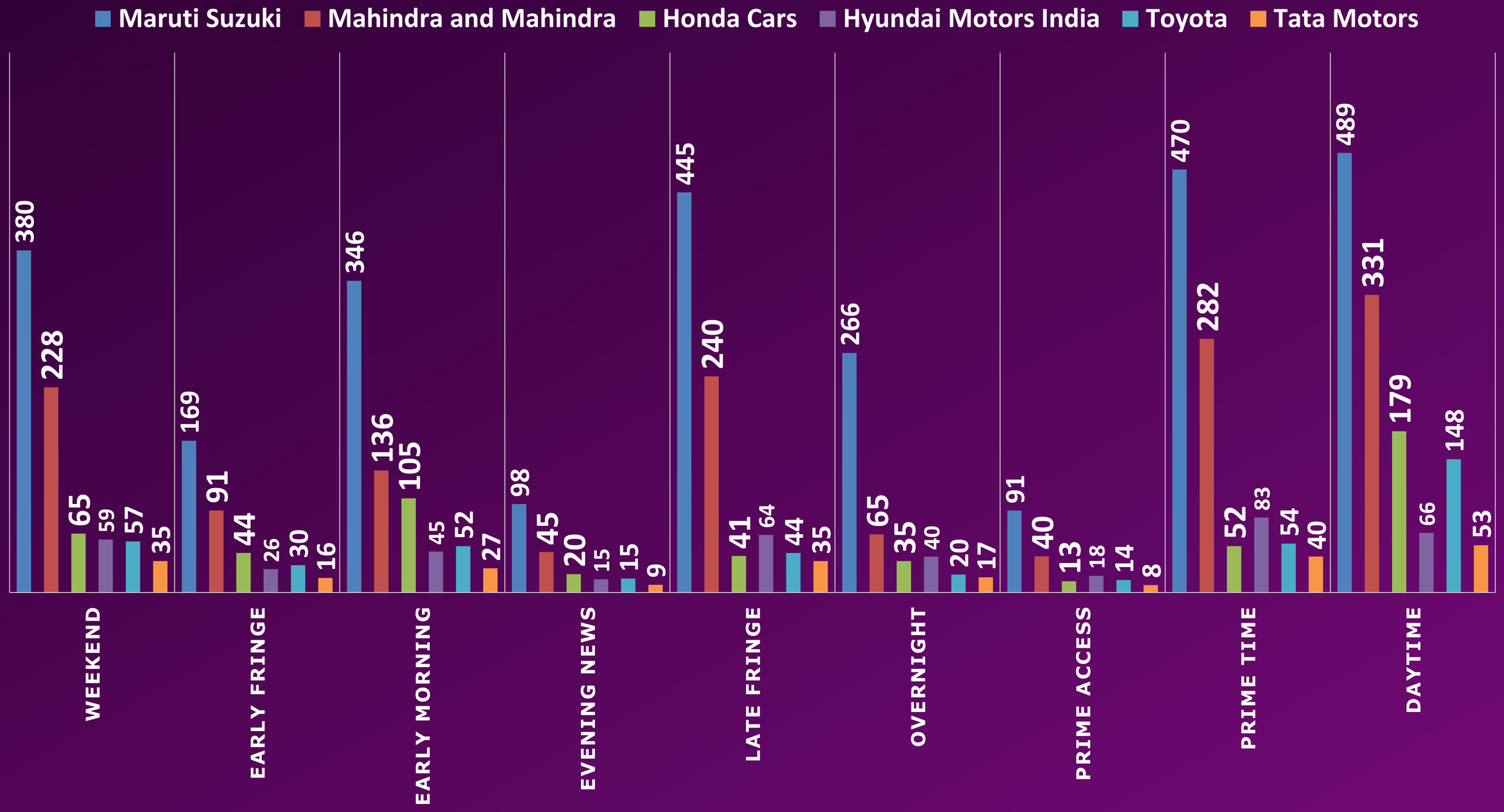


In the above slide % Profit Yield has been calculated as the Total Equivalent Sales Per Brand divided by the Total Ads Per Brand. Considering 2020s Profit Yield as 100 % the curves are being drawn for 2021. As evident from the graph from 2020 to 2021 the % profit yield has for Maruti Suzuki and Mahindra and Mahindra are slightly better compared to the other brands. Although the below slide showcase the total sales to total spend ratio is better for Honda Cars (0.015%). Slide 13 showcase that Brand Maruti Suzuki and Mahindra and Mahindra has aired ads for longer duration for every dayparts. The below slides showcase patterns which is evident that almost every brands are following the same strategy in terms of ads aired via different network types. In terms of duration of the ads aired every brand has aired ads in different dayparts but Maruti Suzuki and Mahindra and Mahindra being consistent in terms of airing ads on every dayparts for a longer duration.

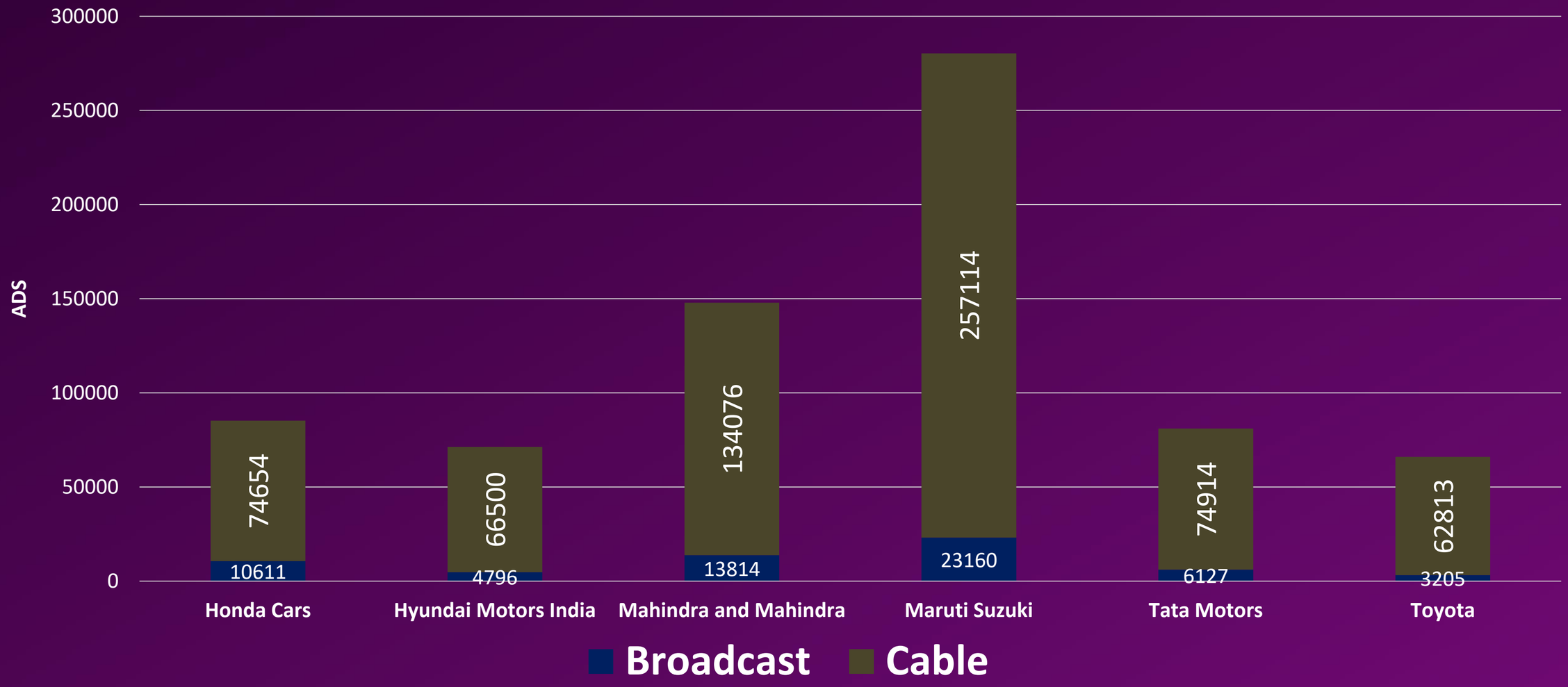
TOTAL EQUIVALENT SALES VS TOTAL ADS BY BRAND



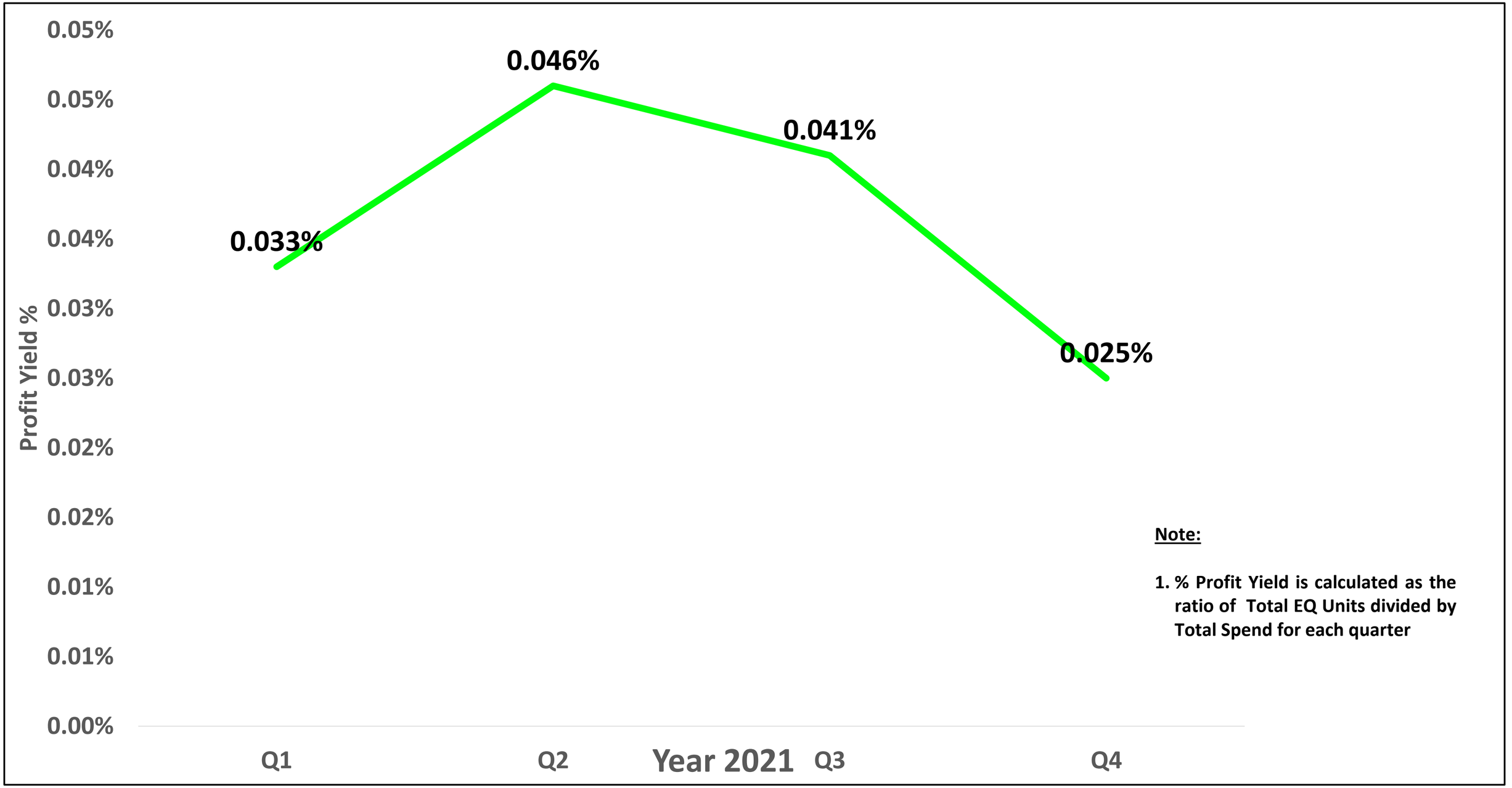
DURATION (HRS) ADS AIRED ON DAYPARTS PER BRAND



ADS BY NETWORK TYPE



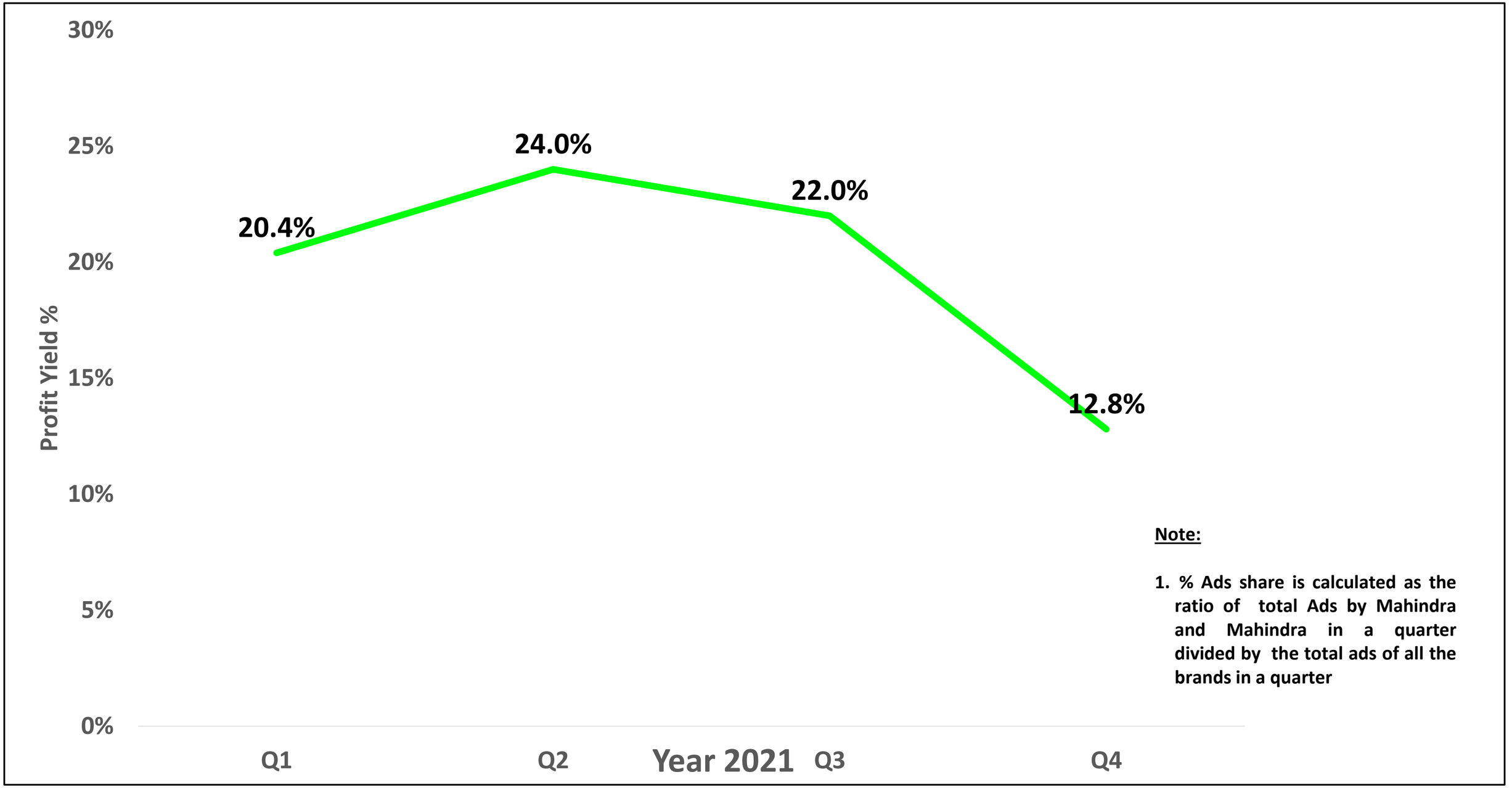
% PROFIT YIELD CURVE QUARTER WISE



Note:

1. **% Profit Yield is calculated as the ratio of Total EQ Units divided by Total Spend for each quarter**

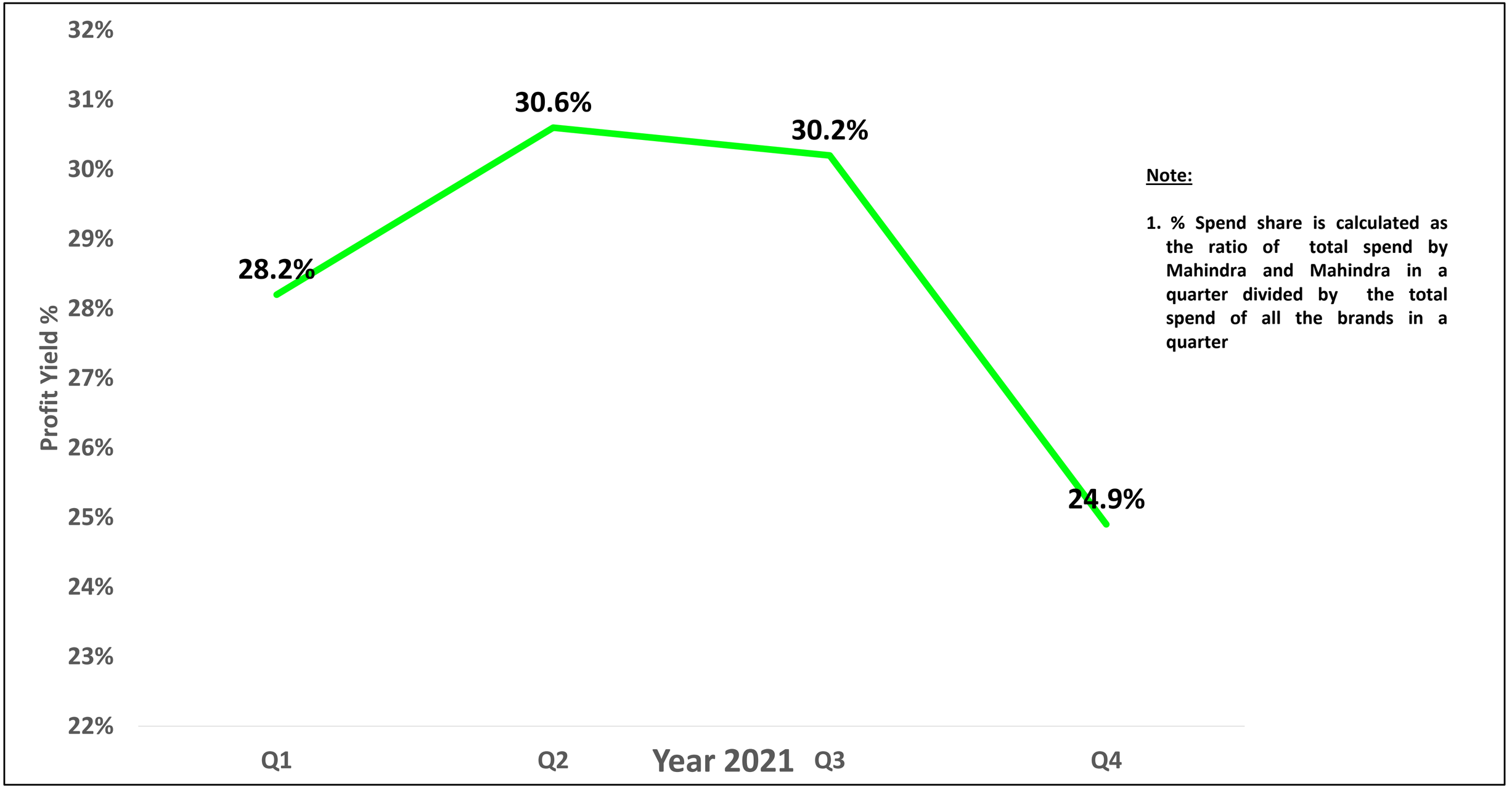
% ADS SHARE CURVE QUARTER WISE



Note:

1. **% Ads share is calculated as the ratio of total Ads by Mahindra and Mahindra in a quarter divided by the total ads of all the brands in a quarter**

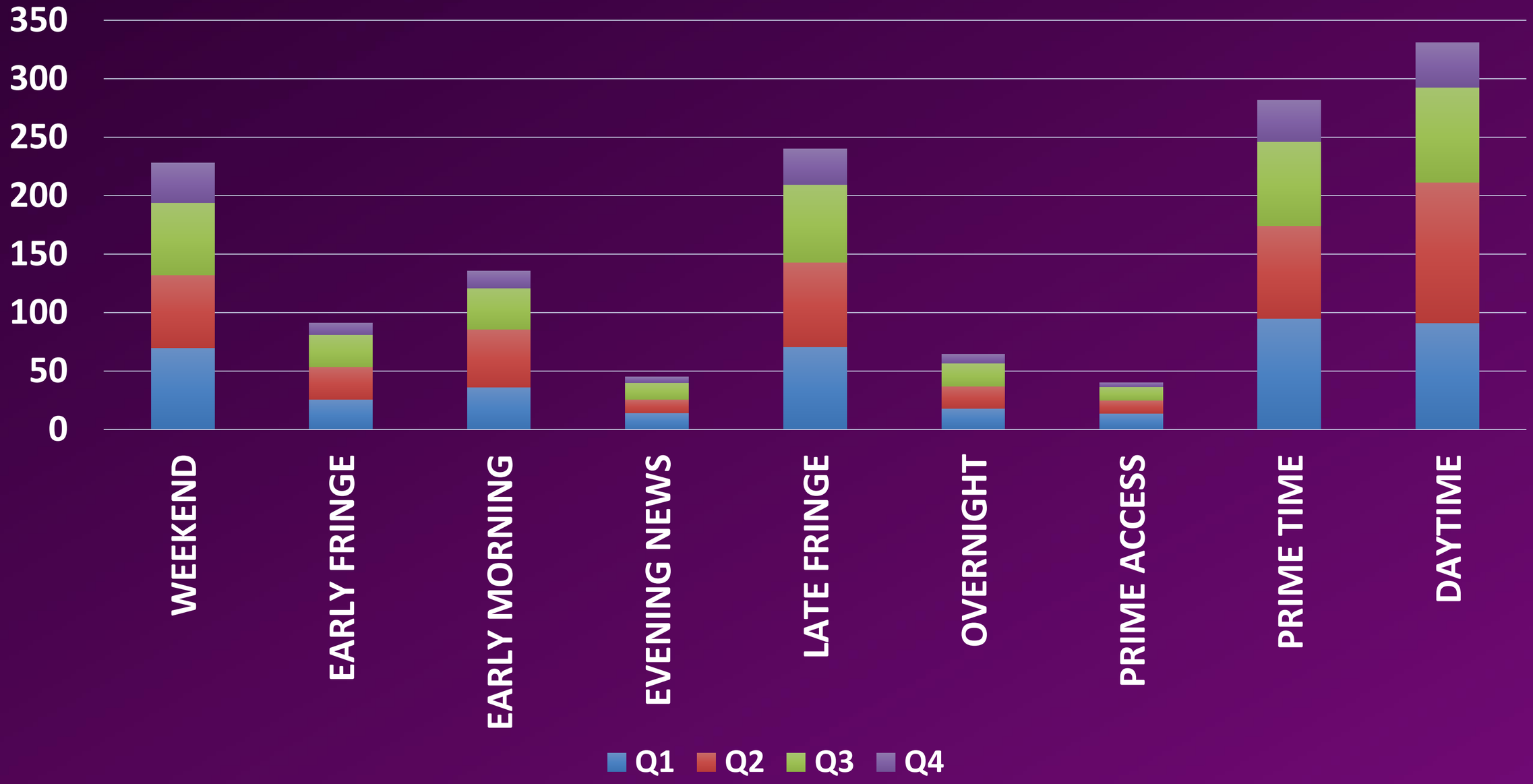
% SPEND SHARE CURVE QUARTER WISE



Note:

1. % Spend share is calculated as the ratio of total spend by Mahindra and Mahindra in a quarter divided by the total spend of all the brands in a quarter

DURATION (HRS) ADS AIRED ON DAYPARTS PER QUARTER



In the above slides if we track the growth of Mahindra and Mahindra the growth has been inconsistent per quarter. A better digital plan would be to increase the number of ads per quarter and maintain it throughout the rest of the quarters. Page 57 showcase the increase in % spend has resulted in better profit yield in Q2.Thus Mahindra and Mahindra should increase the amount spend on ads aired per quarter.

PROJECT 8 : CALL VOLUME TREND ANALYSIS

A customer experience (CX) team consists of professionals who analyze customer feedback and data, and share insights with the rest of the organization. Typically, these teams fulfil various roles and responsibilities such as: Customer experience programs (CX programs), Digital customer experience, Design and processes, Internal communications, Voice of the customer (VoC), User experiences, Customer experience management, Journey mapping, Nurturing customer interactions, Customer success, Customer support, Handling customer data, Learning about the customer journey.

Inbound customer support is defined as the call center which is responsible for handling inbound calls of customers. Inbound calls are the incoming voice calls of the existing customers or prospective customers for your business which are attended by customer care representatives. Inbound customer service is the methodology of attracting, engaging, and delighting your customers to turn them into your business' loyal advocates. By solving your customers' problems and helping them achieve success using your product or service, you can delight your customers and turn them into a growth engine for your business.

Case Study Objectives:

- Calculate the average call time duration for all incoming calls received by agents (in each Time Bucket).
- Show the total volume/ number of calls coming in via charts/ graphs [Number of calls v/s Time]. You can select time in a bucket form
- As you can see current abandon rate is approximately 30%. Propose a manpower plan required during each time bucket [between 9am to 9pm] to reduce the abandon rate to 10%. (i.e., You must calculate minimum number of agents required in each time bucket so that at least 90 calls should be answered out of 100.)

APPROACH

- My approach to this project involved understanding the data set.
- I have also gained the domain knowledge on how call centres leverage data to improve customer experience.
- After understanding the data set, I have checked there are null records on the columns agent name and agent id. Instead of removing those nulls, I have replaced those nulls with Executives 66 on the agent's name column and 1000066 on the agent ids column as removing the nulls will impact the data and in turn analysis as well. The null imputation has been done on the category columns as it will not impact the records of the other columns.
- The case study objectives were answered using both SQL and Excel.

CONCLUSION

All the case study objectives were successfully answered. This project helps understand the terminologies and metrics that are required to carry out analysis in a call center. The insights are as follows:

- The average call duration calculated for each time period helps track average handling time.
- The call volume percentage calculated for each call status helps track the total volume of calls.

THANK YOU !!

Let's work together

MOBILE

7906325424

MAILING ADDRESS

voraciousv555@gmail.com