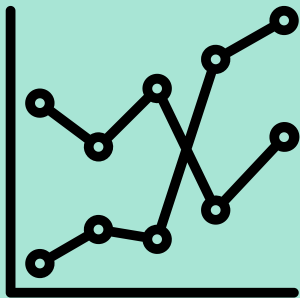
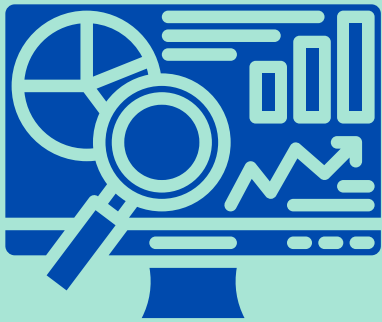


DATA ANALYST PORTFOLIO

Submitted by: Naveen A



Professional Background



Currently I have completed my B.E in Mechanical Engineering department, I have secured 7.9 CGPA and have several skills including Data Analysis, MySQL, Cloud Computing (AWS).

I have worked as an intern in trainity as a Data Analyst trainee. I have also completed my Cloud Computing (AWS) from Ethnus Codemitra. I have worked on Data Analytics course from scratch and managed different projects.

As I am a fresher it would be great to experience the real challenges of the corporate world and understand how things work. Being a fresher, I think I am very flexible and adaptive to learn new things. I have theoretical knowledge. But I am waiting to use my theoretical knowledge in a practical way. And I believe by putting significant efforts I will learn.

TABLE OF CONTENTS



PROFESSIONAL BACKGROUND	1
TABLE OF CONTENTS	2
TRAINITY PROJECT DESCRIPTION	3
THE PROBLEM	4
DESIGN	5
PROJECT 1	6-9
PROJECT 2	10-13
CAPSTONE PROJECT DESCRIPTION	14
DESIGN	15
FINDINGS	16-19
DATA ANALYSIS	20
CONCLUSION	21
REFERENCES	22

Trainity Project Description



The dataset used for the analysis is this Course, Data of various courses containing various columns and tuples. This report shows the findings and the insights derived after cleaning the data and plotting various plots to find the relationship among various items. The main aim of this project was to find insights among the data provided.





The Problem

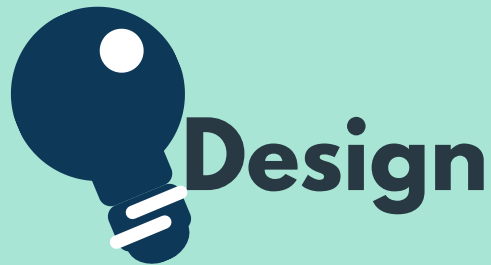
You're a Data Analyst working for the education tech company. You have been asked by your manager, Head of Curriculum at the company, to present the data on course revenue, and you have been provided with data on courses from different topics to understand where opportunities to increase revenue may lie, and track the performance of courses.

Your manager has suggested encouraging Data Analyst courses to charge more because she believes that these are the most popular courses. She needs to send a report to the CEO in the next three weeks on how they will increase their next quarterly earnings.

Goals/Objectives

This part of the project plan helps me move toward project execution smoothly. Make sure to include:

- **This is an individual project**
- **How much time each task would take to complete**
- **Resources assigned to each task**
- **Timeline for me to approve of project roles and responsibilities**



Steps taken to clean the data:

- **First importing the datasets provided and merging them.**
- **Then removing the duplicates and the blank cells (if required).**
- **Improving the headers of each column with proper values.**

Finding and replacing the data with correct values.

Tools used for Query and Visualization:

- **MySQL**
- **PowerBI**
- **Microsoft Excel**



A. Instagram User Analytics

Finding -1

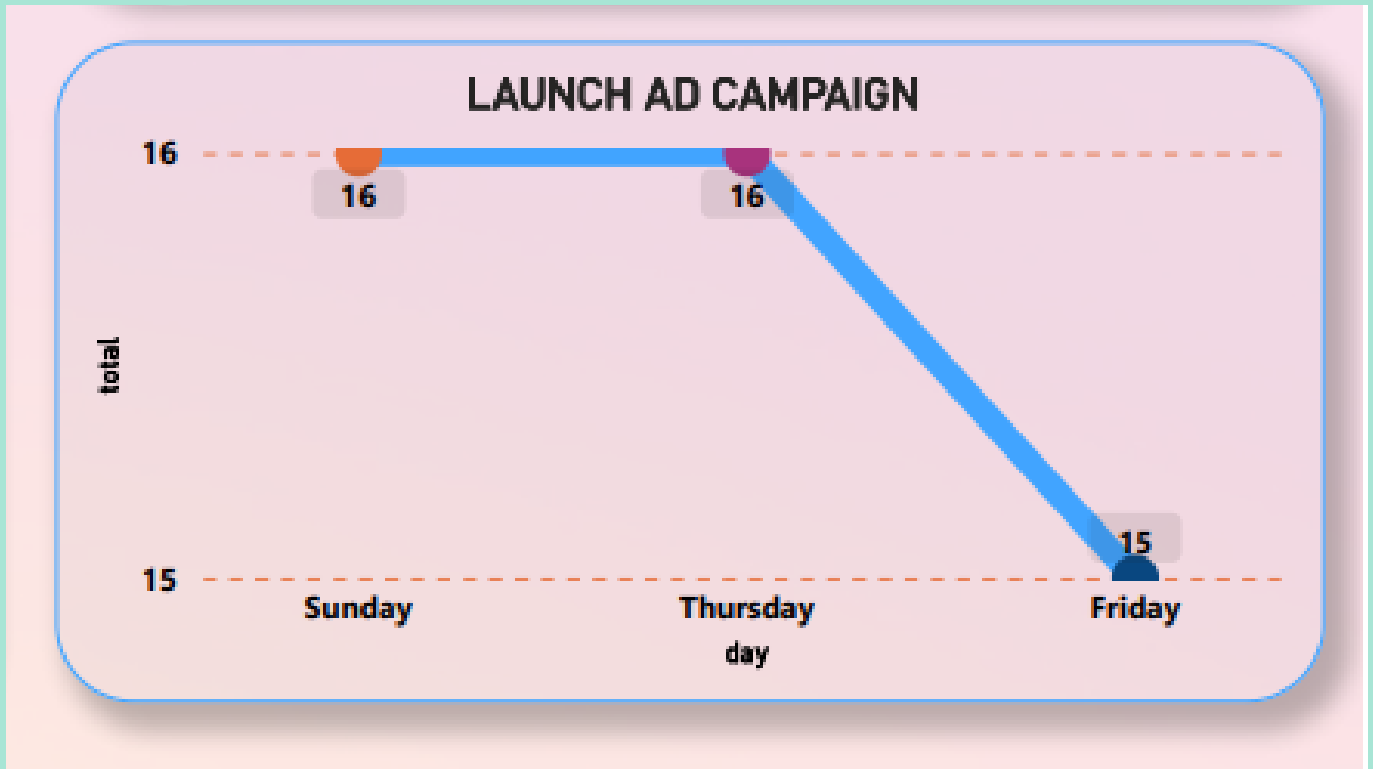


fig.1

When plotting the AD campaign for the companies to post it on Instagram considering each day for analysis (PowerBI) as in Fig. 1 , I found out that “Sunday and Thursday” are the best days to post.



Finding -2

Instagram User Analytics

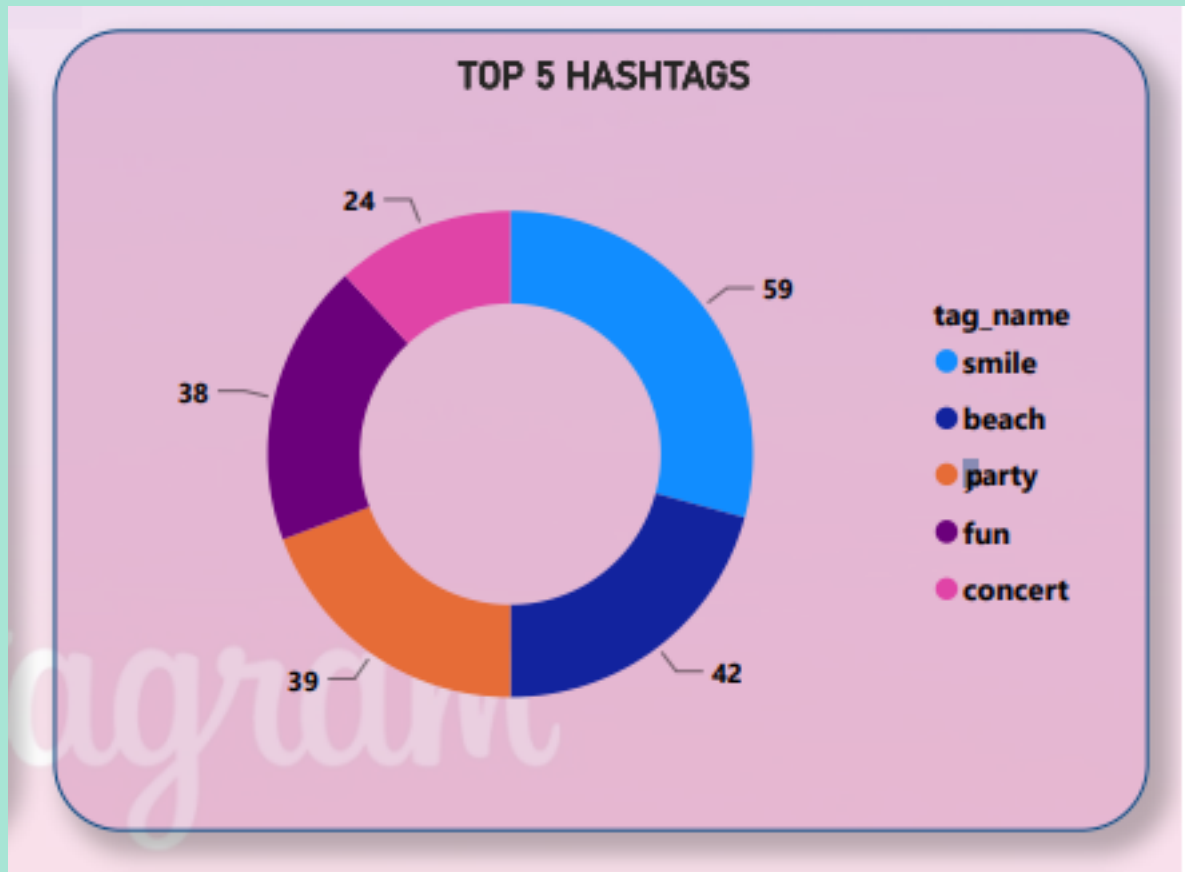


fig.2

When plotting the hashtags used by the user on Instagram, I found out that the top 5 most used hashtags in the given data set. Using this vision user can reach out to the audience faster and get more number of views.



Finding -3

Instagram User Analytics

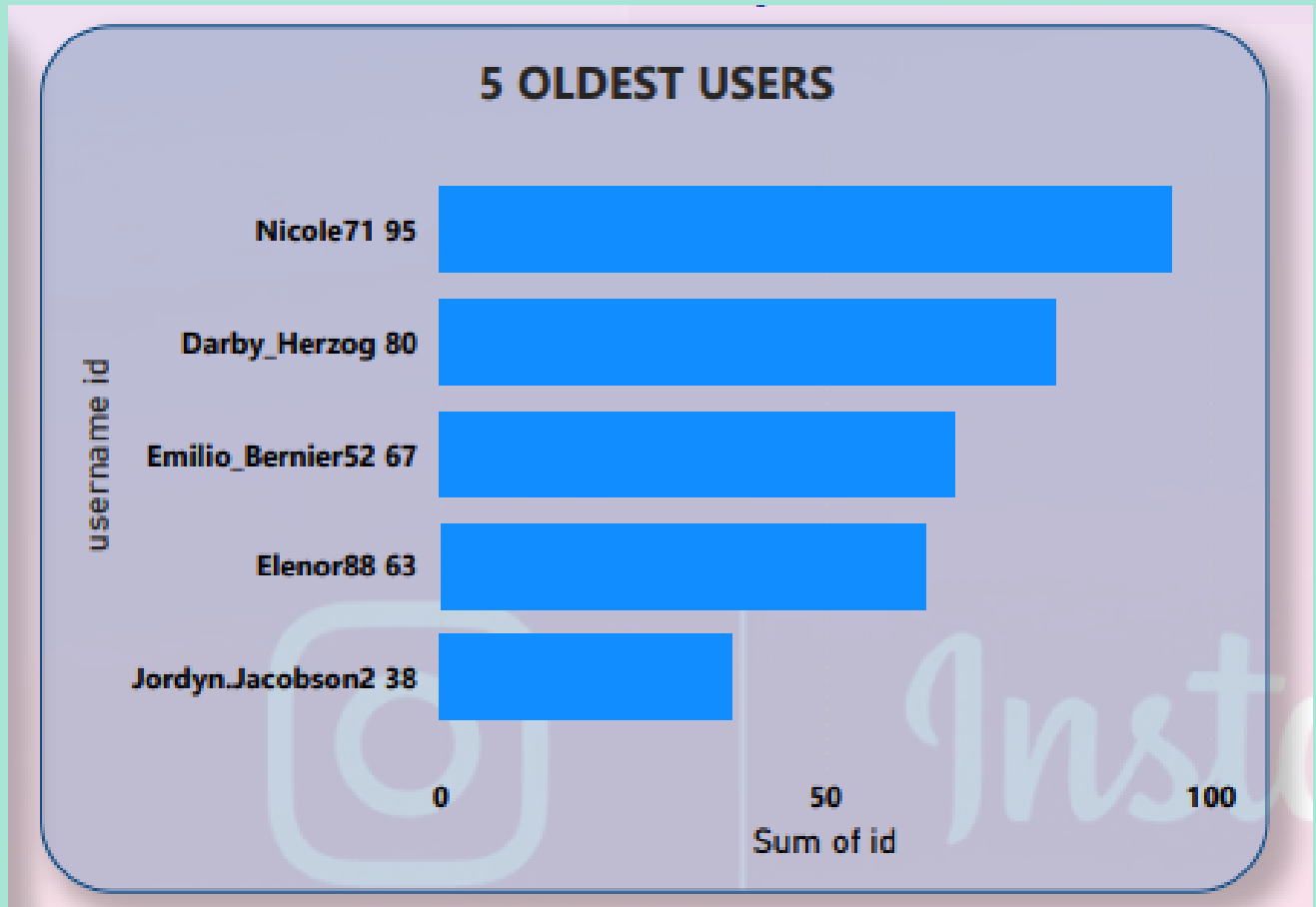


fig.3

Considering the username_id and the date they created their account I was able to find out the top 5 oldest users.

Condition: Only active users (the active users are considered on the basis of likes given to a post).

Analysis

Instagram User Analytics



After the analysis, I found out the results for various questions asked, some of them were shown in the above pages.

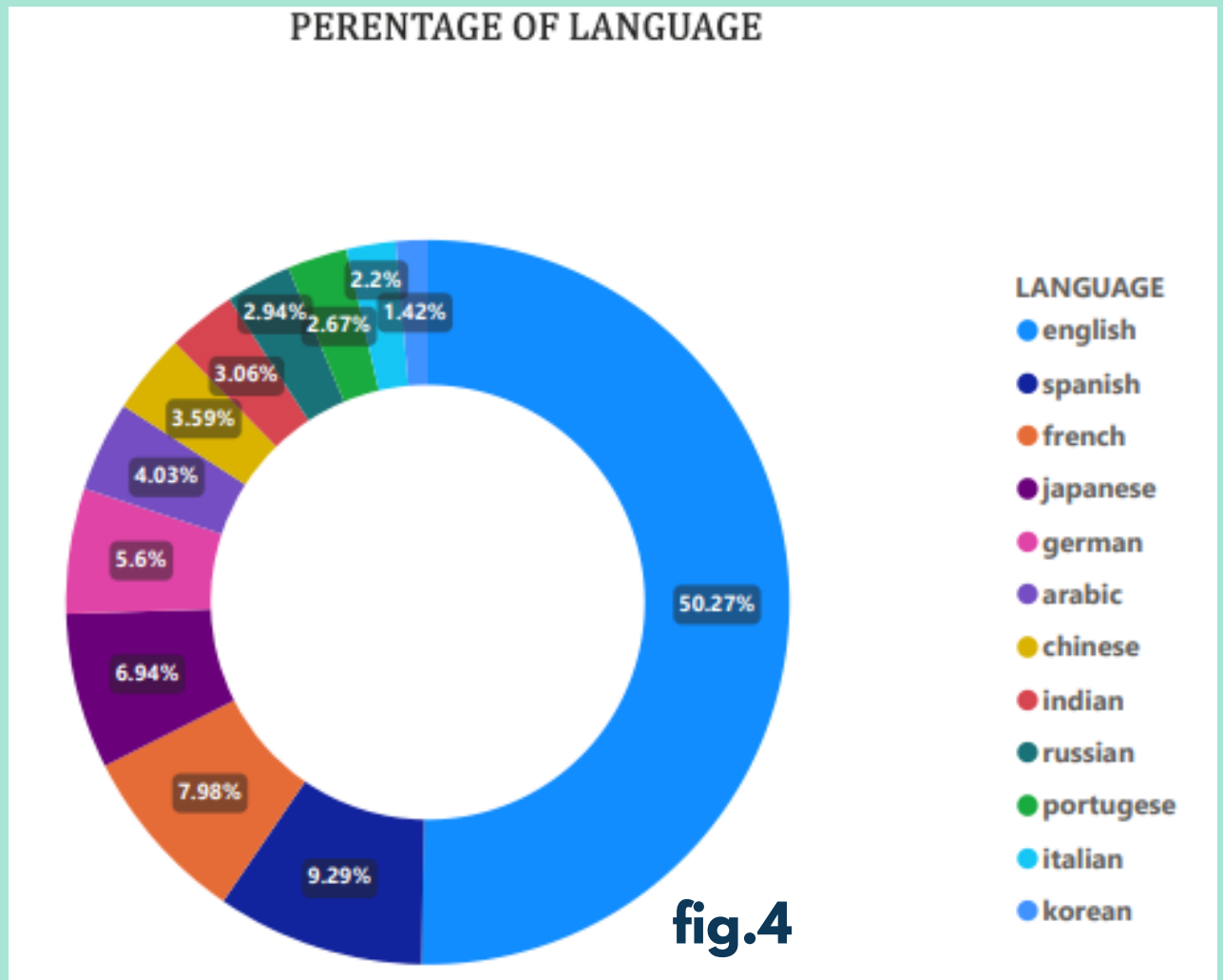
Using the 5 Whys approach I am finding the root cause:-

“Why is the AD Campaign?” Maybe the marketing done is not up to the mark.

“Why is the top 5 Hashtags?” -May be the reach is not up-to the mark.

B. Operation and Metric Analysis

Finding -1



From the above fig.4, I was able to conclude that in a company the percentage share of each language in last 30 days. In that English was in the upper hand with **50.27%**.

Finding -2

Operation and Metric Analysis

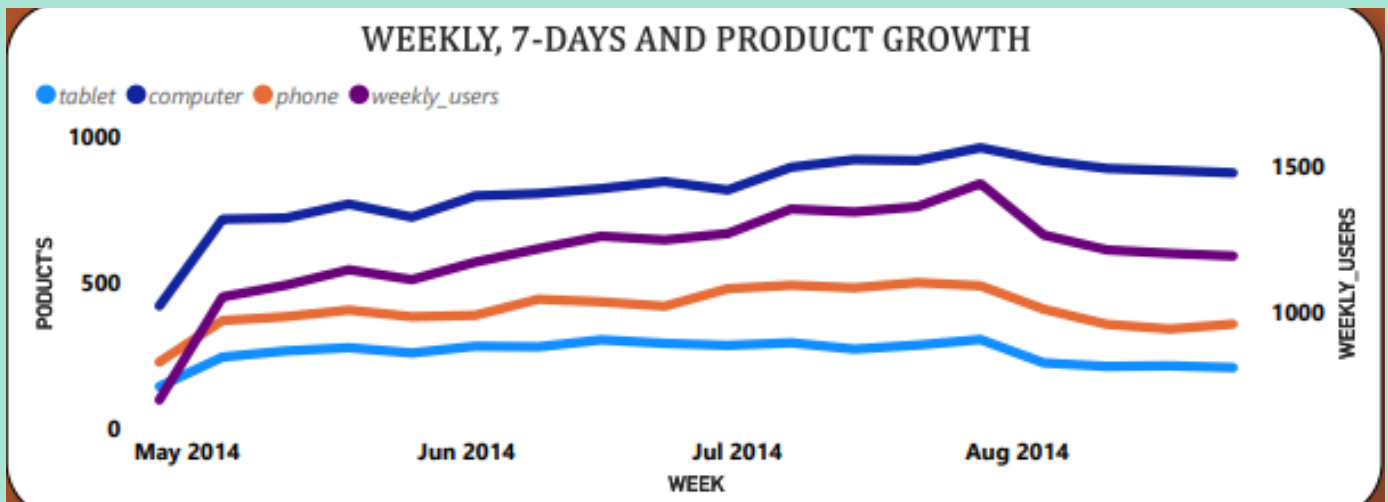


fig.5

From the above fig.5, I was able to conclude that in a company measuring the activeness of a user. Measuring if the user finds quality in a product/service. Considering a period of weeks I was able to say that the users gradually increased their interest in work and use of computers were more.

Finding -3

Operation and Metric Analysis

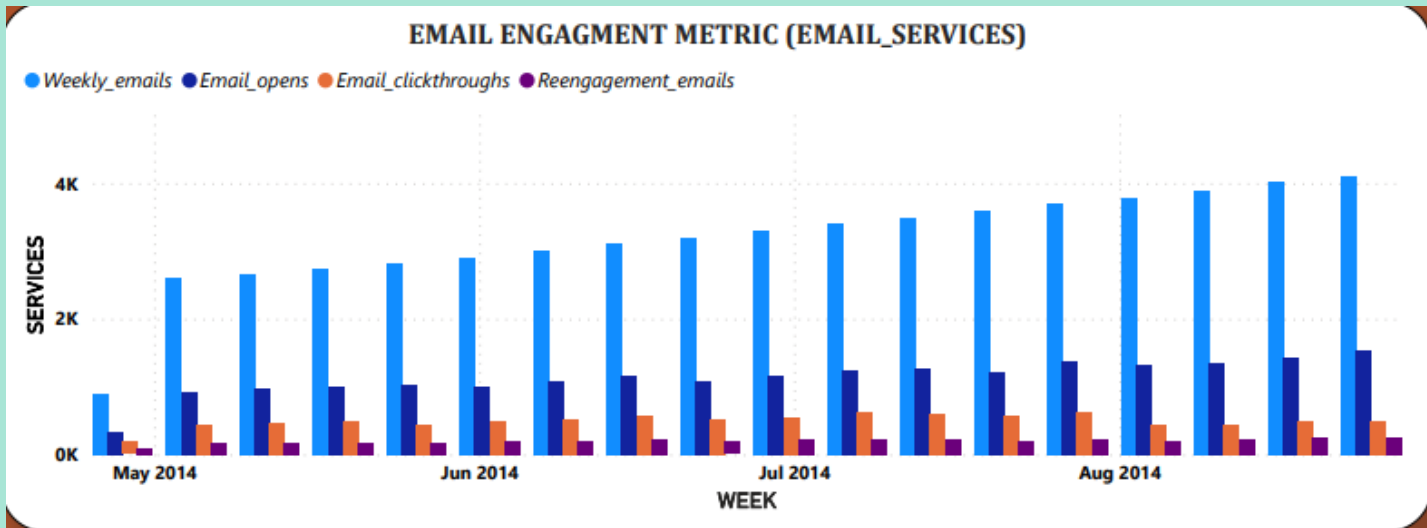


fig.6

From the above fig.6, I was able to conclude that in a company measuring the activeness of a user. Measuring if the user finds quality in a using the email service.

Analysis

Operation and Metric Analysis



After the analysis, I found out the results for various questions asked, some of them were shown in the above pages.

Using the 5 Whys approach I am finding the root cause:-

“Why is the percentage of language?” To know the employees preferences.

“Why is the top products growth?” -May be for budget propose.

"Why is the email metrics?" To know the users experience on that platform.



Capstone Project Description: **"XYZ Ads Airing Report Analysis"**

For your Final Project, we are providing with a dataset having different TV Airing Brands, their product, their category. Dataset includes the network through which Ads are airing, types of network like Cable/ Broadcast and the show name also on which Ads got aired. You can also see the data of Dayparts, Time zone and the time & date at which Ads got aired. IT also includes other data like Pod Position (the lesser the valuable), duration for which Ads aired on screen, Equivalent sales &, total amount spent on the Ads aired.

Design

XYZ Ads Airing Report Analysis



Steps taken to clean the data:

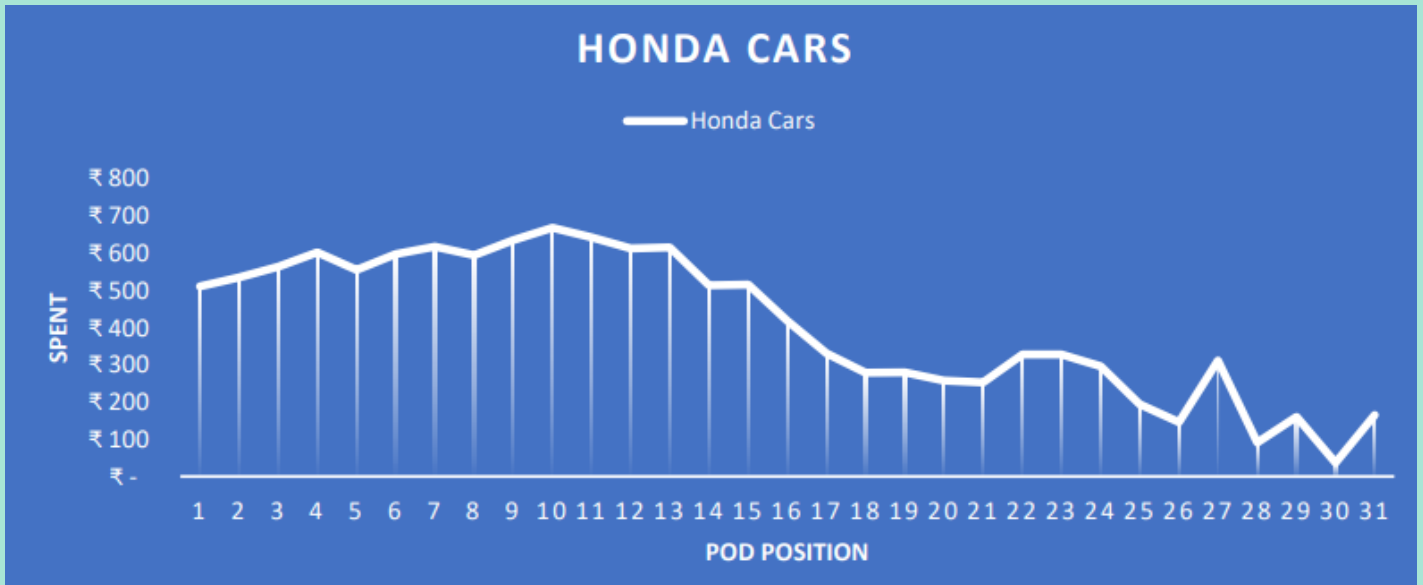
- **First importing the dataset provided.**
- **Then removing the duplicates and the blank cells.**
- **Improving the headers of each column with proper values.**

Finding and replacing the data with correct values. Tools used for visualization:

- **Excel**
- **PowerBI**

Finding-1

XYZ Ads Airing Report Analysis



Considering the pod position for our first analysis, pod position is the position of the ads in a particular period with proper duration. The below graph gives us the clear picture of the amount spent by the Honda car company in starting pod position and gradually decreasing in the last positions. from the above example, the Honda companies share on pod positions.

Finding-2

XYZ Ads Airing Report Analysis



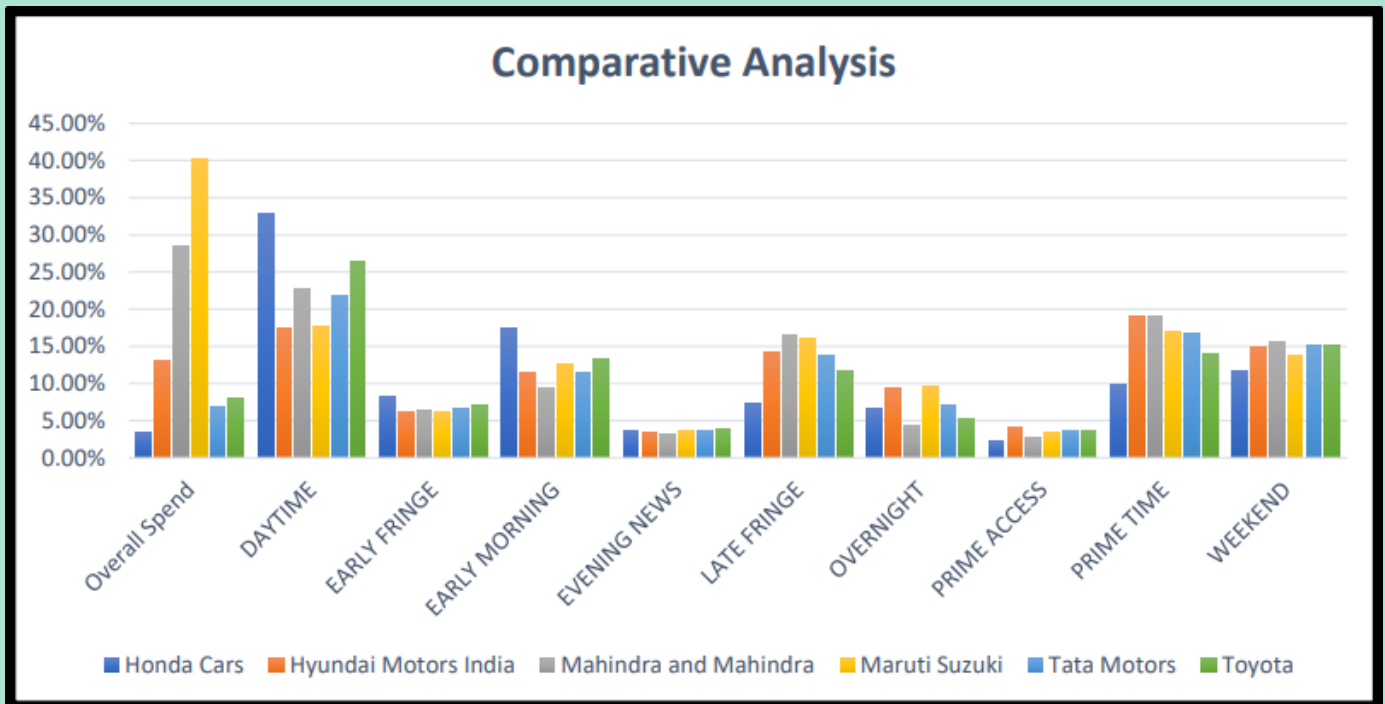
Column Labels <input type="button" value="v"/>				
Sum of Spend (\$)				
Row Labels <input type="button" value="v"/>	Q1	Q2	Q3	Q4
Honda Cars	3.80%	2.82%	3.90%	3.18%
Hyundai Motors India	13.62%	12.51%	12.55%	12.97%
Mahindra and Mahindra	27.63%	30.84%	30.18%	25.57%
Maruti Suzuki	40.39%	39.71%	39.53%	40.80%
Tata Motors	5.92%	5.01%	4.72%	12.60%
Toyota	8.64%	9.12%	9.13%	4.89%
Grand Total	100.00%	100.00%	100.00%	100.00%

Count of Id				
Q1	Q2	Q3	Q4	
12.44%	9.77%	12.99%	11.29%	
10.48%	9.84%	9.17%	9.23%	
19.71%	24.01%	22.05%	13.57%	
38.78%	37.31%	36.55%	41.10%	
10.12%	7.62%	8.03%	20.93%	
8.46%	11.45%	11.21%	3.87%	
100.00%	100.00%	100.00%	100.00%	

The table shows us how much amount/share has been spent on each number of ads given (quarterly Q1, Q2, Q3, Q4). From this we can say that Maruthi Suzuki has spent from on their ads in each quarter.

Finding-3

XYZ Ads Airing Report Analysis



Competitive analysis for the brands and define advertisement strategy of different brands and how it differs across the brands. Considering the below chart, we can conclude that the amount spent more is by the brand Maruti Suzuki and we can observe that the brands spending more in various timelines in a particular day.

Finding-4

XYZ Ads Airing Report Analysis



Count of Duration	Column Labels <input type="button" value="v"/>			
Row Labels <input type="button" value="v"/>	DAYTIME	EARLY FRINGE	EARLY MORNING	EVENING NEWS
Honda Cars	17.63%	14.75%	16.50%	12.01%
Hyundai Motors India	7.85%	9.13%	9.04%	9.44%
Mahindra and Mahindra	21.10%	19.51%	15.53%	17.81%
Maruti Suzuki	31.32%	35.80%	38.92%	39.03%
Tata Motors	11.16%	11.17%	10.24%	11.78%
Toyota	10.94%	9.64%	9.77%	9.93%

LATE FRINGE	OVERNIGHT	PRIME ACCESS	PRIME TIME	WEEKEND
5.88%	10.33%	8.01%	6.95%	9.62%
9.73%	11.97%	12.50%	11.19%	10.11%
23.32%	11.71%	16.99%	23.45%	22.16%
43.10%	49.22%	39.86%	39.55%	36.95%
10.61%	10.52%	12.75%	11.23%	11.65%
7.35%	6.25%	9.90%	7.63%	9.51%

This is the additional analysis done considering the brands and the average duration of ads given by each company in each part of the day.

Analysis

XYZ Ads Airing Report Analysis



Considering the given data set of the ads with various columns like Id, Brand, Product, network, network type, show, date, time, pod position, day parts, duration, EQ units, spend broadcast year, broadcast month, quarter, broadcast week, day of week, hour of day, time zone our main agenda is to answer all the appropriate questions given.

Using the 5 Whys approach I am finding the root cause:-

“Why is the pod positions?” -To know the reach of the people.

“Why is the analysis done in quarters?”-To get an idea on how much they are spending.

"Why is the comparative analysis?" -To know the competition between brands.

Conclusion



In conclusion, I would like to tell that after doing a thorough analysis we were able to derive the insights from the data and was able to plot various graphs using that data. The data which once looked useless gave some very useful insights.

Knowledge on each and advanced concepts gave us the deeper idea of using those tools in industrial standards.

References



Google Sheets Link for Trainity Project:-
<https://drive.google.com/drive/folders/1O3SI2axlbZWNajIBR9bUe26E1n5yCcE>

PowerBI Link for Instagram and Operation Metric Project:-
<https://drive.google.com/drive/folders/1O66IO2tcQzGq0DDEnT0mlzPtrtTa-mzr>

Google Sheets Link for XYZ Ads Airing Report Analysis:-
<https://drive.google.com/drive/folders/1OCgE4foUd4h9OcAHTnV6CAGSelOyZ3sb>

Tableau Link for Udemy GPU Specs Project:-
https://drive.google.com/drive/folders/1OF_Bhtna5UNgU6thdij8Uz-2Sti_-YZf



*Thank
You*