



DATA ANALYTICS

GOKUL KRISHNAN

gokulkrishnan7211@gmail.com

ABOUT ME



I am a post graduate, passionate about data analytics / data science field. I have certified skills in SQL, python, power BI, R programming and machine learning. Also I have completed self paced data science course from Great Learning, currently I'm pursuing data analytics course from Trainity. Now I am seeking a responsible career opportunity to fully utilize my training and skills, while making a significant contribution to the success of the company.

As a fresher It would be great to work in the environment that throws challenges to my knowledge and capabilities, so that my skill and talents utilized to the fullest extent and be contributing to the success and profitability of the organization.

TABLE OF CONTENT

01

ABOUT ME

02

PROJECTS

03

LEARNINGS

04

APPENDIX

05

PERSONAL DETAILS

02. PROJECTS



TOOLS



- SQL, MY SQL WORK BENCH 8.0 CE
- EXCEL, MICROSOFT EXCEL 2021
- MICROSOFT WORD

01 **DATA ANALYSIS PROCESS**

DESCRIPTION OF THE PROJECT

Data analysis is used by every human being in the work we do in our day-to-day life. We use Data Analytics in everyday life without even knowing it. Our task is to give the example of such a real-life situation where we use Data Analytics and link it with the data analytics process.

FINDINGS

The data analytics process or framework is what drives the entire analysis cycle from asking the right business questions, to collecting, cleaning, analyzing and visualizing insights powered by data.

There are six step in data analytics process that is Plan, Prepare, Process, Analyze, Share and Act. Now a days every real life scenario we can link it with data analytics. Data analysis helps to understand the data better so it helps make wise decision, also it help to reduce the time and money.

02

INSTAGRAM

USER

ANALYTICS

DESCRIPTION OF THE PROJECT

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. You are working with the product team of Instagram and the product manager has asked you to provide insights on the questions asked by the management team.

FINDINGS

	id	username
▶	95	Nicole71
	80	Darby_Herzog
	67	Emilio_Bernier52
	63	Elenor88
	38	Jordyn.Jacobson2
●	NULL	NULL

fig 2.1: 5 oldest users of the instagram

- I found out that the 5 oldest users of the Instagram from the database provided for rewarding the most loyal users

	id	username
▶	5	Aniya_Hackett
	7	Kassandra_Homenick
	14	Jadyn81
	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

fig 2.2: Inactive users

- I found out the users who have never posted a single photo on Instagram
- Remind Inactive Users to Start Posting By sending them promotional emails to post their 1st photo

FINDINGS

	user_id	username	photo_id	likes_get
▶	52	Zack_Kemmer93	145	48
	46	Malinda_Streich	127	43
	65	Adelle96	182	43

fig 2.3: users who get most like on a single photo

	tag_id	tag_name	most_used_hashtags
▶	21	smile	59
	20	beach	42
	17	party	39
	13	fun	38
	18	concert	24

fig 2.4: most commonly used hash tag

- I found the user who gets the most likes on a single photo. so the team can Declaring the Contest Winner, the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.

	days	most_registered
▶	Thursday	16
	Sunday	16
	Friday	15
	Tuesday	14
	Monday	14
	Wednesday	13
	Saturday	12

fig 2.5: most registered

- I found out the top 5 most commonly used hashtags on the platform, because A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.

- I found out that Thursday and Sunday when most of the user registered that is 16 and Friday 14 people registered, Tuesday and Monday 14 users registered.
- The team wants to know, which day would be the best day to launch ADs. Schedule an ad campaign on Sundays which is a holiday and most of the users registered on this day. So the Sunday is the perfect day for ad campaign

FINDINGS

	post	no_of_times
▶	1	18
	5	14
	4	13
	2	13
	3	9
	8	2

fig 2.6: post and number of times

user engagement : from the above findings

- 74 Active users who have posted at least once.
- 100 Total users (as per the data)
- 257 Total posts made.
- Total Photos/Total users = $257/100 = 2.57$
- so the average will be $257/74 = 3.47$, Based on the data
- we can say that an average user posts 3-4 times.

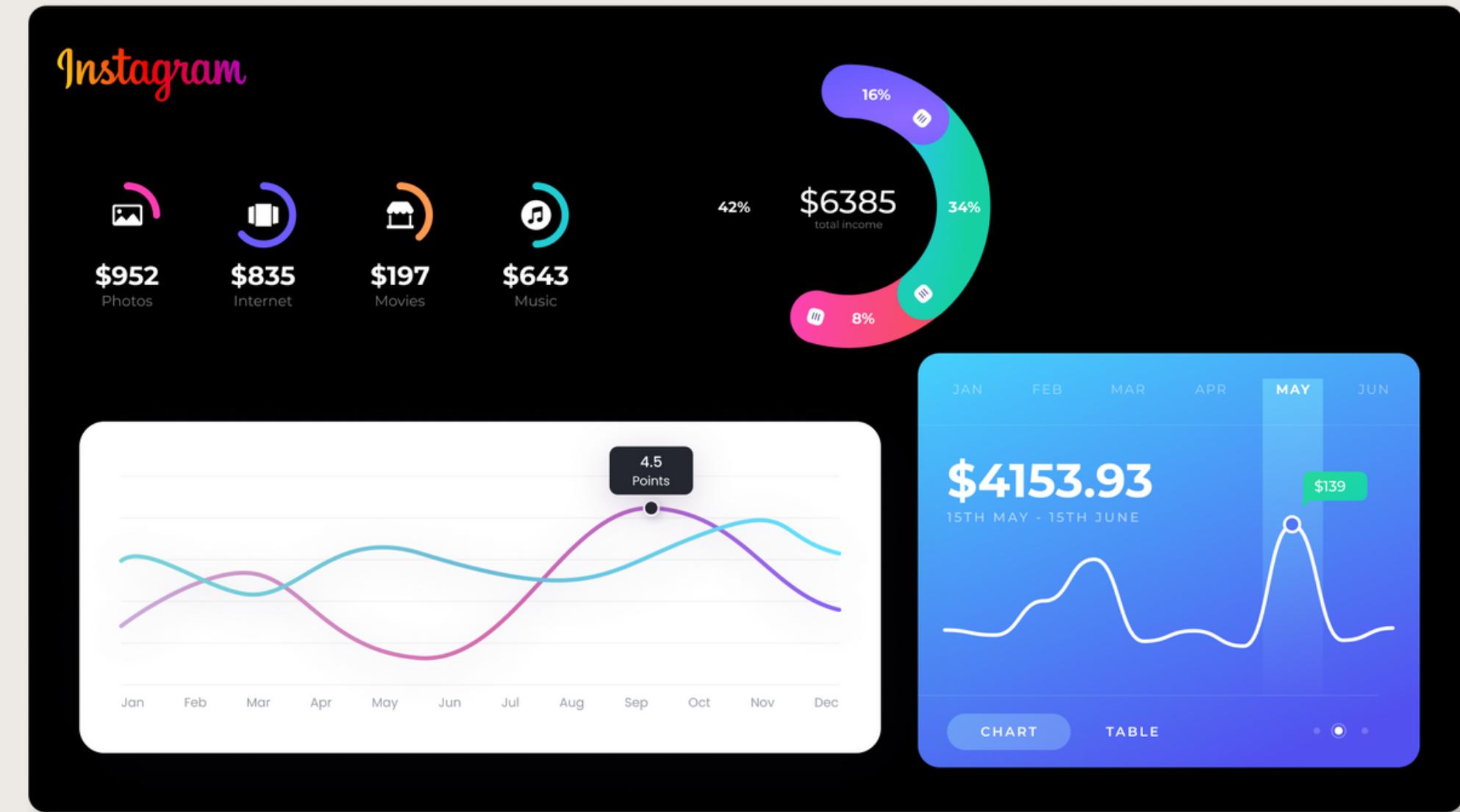
	username	user_id	likes_given
▶	Jadyn81	14	257
	Rocio33	21	257
	Maxwell.Halvorson	24	257
	Ollie_Ledner37	36	257
	Mckenna17	41	257
	Duane60	54	257
	Julien_Schmidt	57	257
	Mike.Auer39	66	257
▶	Nia_Haag	71	257
	Leslie67	75	257
	Janelle.Nikolaus81	76	257
	Bethany20	91	257

fig 2.7: users like every single photo

- I have found the users who have liked every single photo on the site, The investors want to know if the platform is crowded with fake and dummy accounts

CONCLUSION

From this project instagram user analytics, analyze the user activity on instagram how user engage and interact with the application. And provide the insights to marketing and business development team for improve the user experience and helps the business to grow. The marketing team want to launch a new marketing campaign. There are several tasks given in order to complete I used My SQL work bench 8.0 CE and used several SQL queries to get the answer.



03 OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

DESCRIPTION OF THE PROJECT

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.

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.

FINDINGS

	ds	no_of_jobs	total_seconds
▶	2020-11-30	2	40
	2020-11-29	1	20
	2020-11-28	2	33
	2020-11-27	1	104
	2020-11-26	1	56
	2020-11-25	1	45

fig 2.9: number of job reviewed

- using some of queries I have found out the number of jobs reviewed per hour per day for November 2020.

	event_	ds	total	rollingavg
▶	transfer	2020-11-25	45	45.0000
	skip	2020-11-26	56	50.5000
	decision	2020-11-27	104	68.3333
	transfer	2020-11-28	22	56.7500
	decision	2020-11-28	11	47.6000
	decision	2020-11-29	20	43.0000
	skip	2020-11-30	15	39.0000
	transfer	2020-11-30	25	36.1429

fig 2.10: 7 days rolling avg

- by using the window function over(). I have found out the 7 day rolling average of throughput

	language_	percentage(%)
▶	English	12.5000
	Arabic	12.5000
	Persian	37.5000
	Hindi	12.5000
	French	12.5000
	Italian	12.5000

fig 2.11: percentage share of each language

- with the help of some of the queries we have found out the percentage share of each language in the last 30 days.

FINDINGS

event_type	user_engagement	weeks
engagement	8019	17
engagement	17341	18
engagement	17224	19
engagement	17911	20
engagement	17151	21
engagement	18413	22
engagement	18280	23
engagement	19052	24
engagement	18642	25
engagement	19061	26
engagement	19881	27
engagement	20776	28
engagement	20067	29
engagement	21533	30
engagement	18556	31
engagement	16612	32
engagement	16145	33
engagement	16127	34
engagement	784	35

fig 2.12: weekly user engagement

- To measure the activeness of a user. we have found out the weekly user engagement.

users	weeks	event_type
385	17	signup_flow
901	18	signup_flow
954	19	signup_flow
955	20	signup_flow
961	21	signup_flow
1042	22	signup_flow
1065	23	signup_flow
1158	24	signup_flow
1075	25	signup_flow
1065	26	signup_flow
1140	27	signup_flow
1132	28	signup_flow
1166	29	signup_flow
1242	30	signup_flow
1029	31	signup_flow
1260	32	signup_flow
1300	33	signup_flow
1339	34	signup_flow
88	35	signup_flow

fig 2.13: weekly sign up

- I have found out the weekly retention of users-sign up cohort.

event_type	count	device	weeks
engagement	67	acer aspire desktop	17
engagement	206	acer aspire notebook	17
engagement	83	amazon fire phone	17
engagement	251	asus chromebook	17
engagement	187	dell inspiron desktop	17
engagement	503	dell inspiron notebook	17
engagement	132	hp pavilion desktop	17
engagement	190	htc one	17
engagement	338	ipad air	17
engagement	205	ipad mini	17
engagement	217	iphone 4s	17
engagement	706	iphone 5	17
engagement	473	iphone 5s	17
engagement	57	kindle fire	17
engagement	793	lenovo thinkpad	17
engagement	59	mac mini	17
engagement	490	macbook air	17
engagement	1516	macbook pro	17
engagement	145	nexus 10	17
engagement	382	nexus 5	17
engagement	448	phone 4s	18
engagement	145	nexus 10	18
engagement	382	nexus 5	18
engagement	141	amazon fire phone	19
engagement	268	asus chromebook	19
engagement	444	dell inspiron desktop	19
engagement	1193	dell inspiron notebook	19
engagement	376	hp pavilion desktop	19
engagement	270	htc one	19
engagement	595	ipad air	19
engagement	381	ipad mini	19
engagement	546	iphone 4s	19
engagement	1190	iphone 5	19
engagement	964	iphone 5s	19
engagement	225	kindle fire	19
engagement	2143	lenovo thinkpad	19
engagement	255	mac mini	19
engagement	1331	macbook air	19
engagement	3159	macbook pro	19
engagement	232	nexus 10	19
engagement	944	nexus 5	19
engagement	334	nexus 7	19
engagement	215	nokia lumia 635	19

fig 2.14: weekly engagement per device

- above shows the device that are used by users for weekly engagement we can see that, the devices used by users for weekly engagement

FINDINGS

action_	user_engagement	w	action_	user_engagement	weeks
► email_clickthrough	166	17	email_clickthrough	38	35
email_clickthrough	430	18	email_open	310	17
email_clickthrough	477	19	email_open	912	18
email_clickthrough	507	20	email_open	972	19
email_clickthrough	443	21	email_open	1004	20
email_clickthrough	488	22	email_open	1014	21
email_clickthrough	538	23	email_open	987	22
email_clickthrough	554	24	email_open	1075	23
email_clickthrough	530	25	email_open	1155	24
email_clickthrough	556	26	email_open	1096	25
email_clickthrough	621	27	email_open	1165	26
email_clickthrough	599	28	email_open	1228	27
email_clickthrough	590	29	email_open	1250	28
email_clickthrough	630	30	email_open	1219	29
email_clickthrough	445	31	email_open	1383	30
email_clickthrough	418	32	email_open	1351	31
email_clickthrough	490	33	email_open	1337	32
email_clickthrough	490	34	email_open	1432	33
email_clickthrough	38	35	email_open	1528	34
email open	310	17	email open	41	35

action_	user_engagement	weeks	action_	user_engagement	weeks
email_open	1528	34	sent_reengagement_email	48	35
email_open	41	35	sent_weekly_digest	908	17
sent_reengagement_email	73	17	sent_weekly_digest	2602	18
sent_reengagement_email	157	18	sent_weekly_digest	2665	19
sent_reengagement_email	173	19	sent_weekly_digest	2733	20
sent_reengagement_email	191	20	sent_weekly_digest	2822	21
sent_reengagement_email	164	21	sent_weekly_digest	2911	22
sent_reengagement_email	192	22	sent_weekly_digest	3003	23
sent_reengagement_email	197	23	sent_weekly_digest	3105	24
sent_reengagement_email	226	24	sent_weekly_digest	3207	25
sent_reengagement_email	196	25	sent_weekly_digest	3302	26
sent_reengagement_email	219	26	sent_weekly_digest	3399	27
sent_reengagement_email	213	27	sent_weekly_digest	3499	28
sent_reengagement_email	213	28	sent_weekly_digest	3592	29
sent_reengagement_email	213	29	sent_weekly_digest	3706	30
sent_reengagement_email	231	30	sent_weekly_digest	3793	31
sent_reengagement_email	222	31	sent_weekly_digest	3897	32
sent_reengagement_email	200	32	sent_weekly_digest	4012	33
sent_reengagement_email	264	33	sent_weekly_digest	4111	34
sent reengagement_email	261	34			

- from the table we can see that the email engagement metrics. here there are 4 actions taken by the users like email_clickthrough, email_open, sent_reengagement_email, sent_weekly_digest.

fig 2.15: email engagement

CONCLUSION

This project mainly focused on operation analytics. Operation analytics helps companies to find which area they have to improve. Here in this project there are two different kinds of data sets were provided. One was job data and another was to find investigation metric spike. There are certain number of task was given in each case studies. That required SQL queries, this project helped me to improve my existing skills and knowledge. The data sets given were bit complex and that required more understanding. I used some more function and queries to find answer, so it help me to familiar with some function like over(), count(), etc.



04

HIRING PROCESS ANALYTICS

DESCRIPTION OF THE PROJECT

Hiring process is the fundamental and the most important function of a company. Here, the MNCs get to know about the major underlying trends about the hiring process. Trends such as- number of rejections, number of interviews, types of jobs, vacancies etc. are important for a company to analyse before hiring freshers or any other individual. Thus, making an opportunity for a Data Analyst job here too!

FINDINGS

	Hired	Rejected	Grand Total
Row Labels			
Female	1854	819	2673
Male	2562	1521	4083
Grand Total	4416	2340	6756

fig 2.16: number of male and female hired

After removing outliers and missing values. We can see that there are 1854 females and 2562 males were hired.

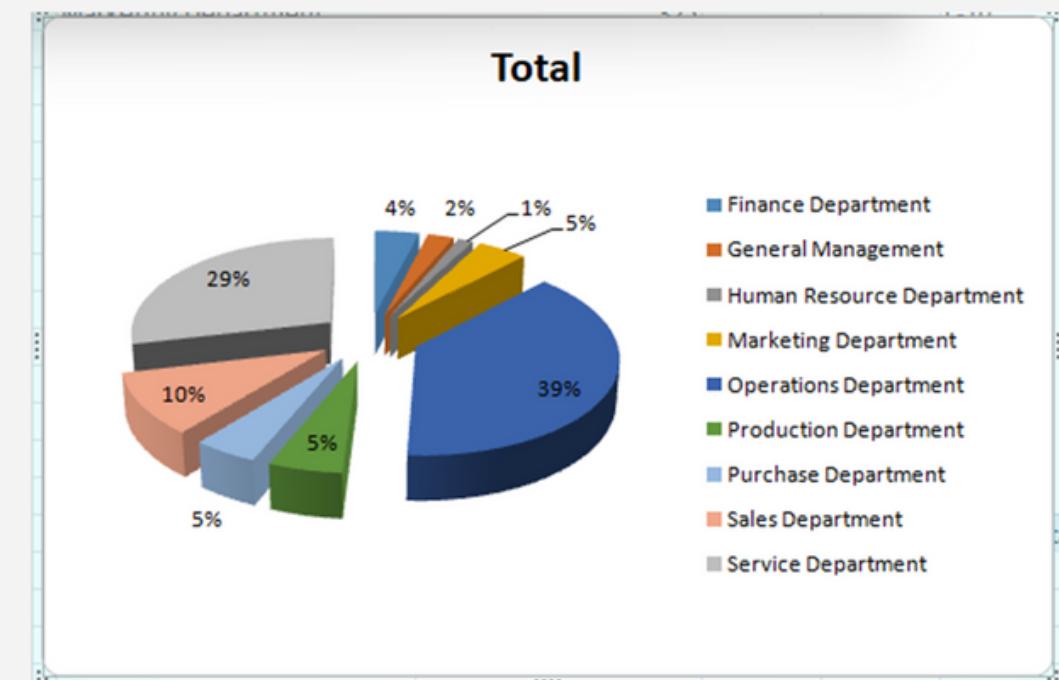


fig 2.17: department

- From the above graph we can see that majority of people works in the operation department that is 39% of people from the total. Just below the operation department where people work the most is sales department. we can also see that only 1% of people are works in the human resource department.

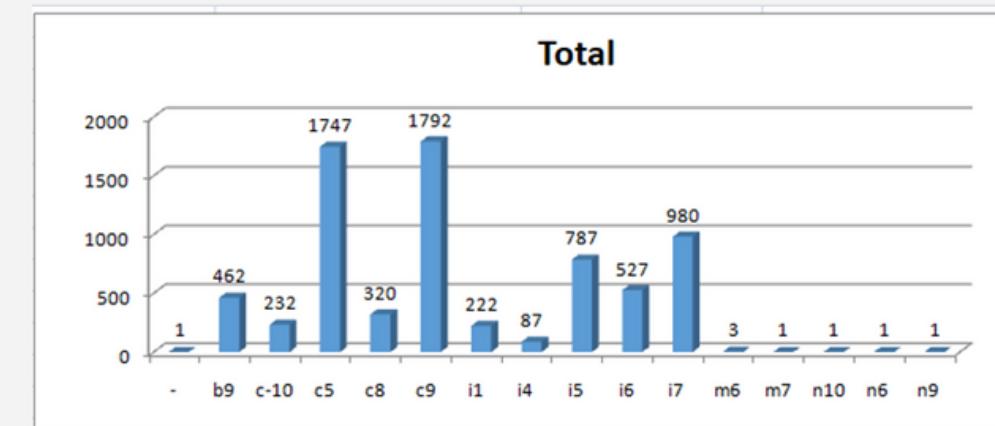


fig 2.18: post in different department

- From the above graph we can see the different post in different departments.

CONCLUSION



The project Hiring process analytics is the fundamental and the most important function of a company. They need to know number of rejections, number of interviews, types of jobs, vacancies etc. that are important for a company to Analyse before hiring fresher's or any other individual.

To do this project I have to used Microsoft excel for the given tasks. Microsoft excel work really well for simple calculations and tracking almost any kind of information. This project help me to gain more knowledge about excel it improve my skills, help to gain more knowledge about the formula and function used in the excel. More importantly I get to know more about pivot table. I used pivot table more in this projects

05

IMDB

MOVIE

ANALYSIS

DESCRIPTION OF THE PROJECT

For the Final Project, we are providing you with dataset having various columns of different IMDB Movies. You are required to Frame the problem. For this task, you will need to define a problem you want to shed some light on. This project is also used root cause analysis, in this analysis certain questions will be asked as a data analysis we have to answer this questions.

FINDINGS

	gross	movie_title	budget	PROFIT
1	761 m	Avatar	237 M	524 M
2	652 m	Jurassic World	150 M	502 M
3	659 m	Titanic	200 M	459 M
4	461 m	Star Wars: Episode IV - A New Hope	11 M	450 M
5	435 m	E.T. the Extra-Terrestrial	11 M	424 M
6	623 m	The Avengers	220 M	403 M
7	423 m	The Lion King	45 M	378 M
8	475 m	Star Wars: Episode I - The Phantom Menace	115 M	360 M
9	533 m	The Dark Knight	185 M	348 M
10	408 m	The Hunger Games	78 M	330 M
11	363 m	Deadpool	58 M	305 M
12	425 m	The Hunger Games: Catching Fire	130 M	295 M
13	357 m	Jurassic Park	63 M	294 M
14	368 m	Despicable Me 2	76 M	292 M
15	350 m	American Sniper	59 M	291 M
16	381 m	Finding Nemo	94 M	287 M
17	436 m	Shrek 2	150 M	286 M
18	377 m	The Lord of the Rings: The Return of the King	94 M	283 M
19	309 m	Star Wars: Episode VI - Return of the Jedi	33 M	277 M
20	330 m	Forrest Gump	55 M	275 M

fig 2.19 movies with highest profit



fig 2.10

- form the given datasets I found out the movies with the highest profit Here we can see movies which make highest profit
- Top 1 position we have avatar a James Cameroon movie. and in the second position we have Jurassic world.

FINDINGS

	A	B	C	D	E	F	G	H	I	J	K
1	director_name	gross	movie_title	num_v	language	content	budget	title_ye	imdb_score	IMDb_Top_250	Rank
2	Frank Darabont	28341469	The Shawshank Rede	1689764	English	R	25000000	1994	9.3 The Shawshank Rede	1	
3	Francis Ford Copp	134821952	The GodfatherÂ	1155770	English	R	6000000	1972	9.2 The GodfatherÂ	2	
4	Christopher Nolar	533316061	The Dark KnightÂ	1676169	English	PG-13	1.85E+08	2008	9 The Dark KnightÂ	3	
5	Francis Ford Copp	57300000	The Godfather: Par	790926	English	R	13000000	1974	9 The Godfather: Part I	4	
6	Peter Jackson	377019252	The Lord of the Rin	1215718	English	PG-13	94000000	2003	8.9 The Lord of the Rings	5	
7	Quentin Tarantinc	107930000	Pulp FictionÂ	1324680	English	R	8000000	1994	8.9 Pulp FictionÂ	6	
8	Steven Spielberg	96067179	Schindler's ListÂ	865020	English	R	22000000	1993	8.9 Schindler's ListÂ	7	
9	Sergio Leone	6100000	The Good, the Bad	503509	Italian	Approved	1200000	1966	8.9 The Good, the Bad an	8	
10	Robert Zemeckis	329691196	Forrest GumpÂ	1251222	English	PG-13	55000000	1994	8.8 Forrest GumpÂ	9	
11	Irvin Kershner	290158751	Star Wars: Episode	837759	English	PG	18000000	1980	8.8 Star Wars: Episode V	10	
12	Peter Jackson	313837577	The Lord of the Rin	1238746	English	PG-13	93000000	2001	8.8 The Lord of the Rings	11	
13	Christopher Nolar	292568851	InceptionÂ	1468200	English	PG-13	1.6E+08	2010	8.8 InceptionÂ	12	
14	David Fincher	37023395	Fight ClubÂ	1347461	English	R	63000000	1999	8.8 Fight ClubÂ	13	
15	George Lucas	460935665	Star Wars: Episode	911097	English	PG	11000000	1977	8.7 Star Wars: Episode IV	14	
16	Peter Jackson	340478898	The Lord of the Rin	1100446	English	PG-13	94000000	2002	8.7 The Lord of the Rings	15	
17	Lana Wachowski	171383253	The MatrixÂ	1217752	English	R	63000000	1999	8.7 The MatrixÂ	16	
18	Milos Forman	112000000	One Flew Over the	680041	English	R	4400000	1975	8.7 One Flew Over the C	17	
19	Martin Scorsese	46836394	GoodfellasÂ	728685	English	R	25000000	1990	8.7 GoodfellasÂ	18	
20	Fernando Meirelle	7563397	City of GodÂ	533200	Portuguese	R	3300000	2002	8.7 City of GodÂ	19	
21	Akira Kurosawa	269061	Seven SamuraiÂ	229012	Japanese	Unrated	2000000	1954	8.7 Seven SamuraiÂ	20	

	A	B	C	D	E	F	G	H	I	J	K	L
235	Darren Aronofsky	26236603	The WrestlerÂ	251349	English	R	6000000	2008	7.9 The WrestlerÂ	234		
236	Edgar Wright	23618786	Hot FuzzÂ	352695	English	R	8000000	2007	7.9 Hot FuzzÂ	235		
237	James Ivory	22954968	The Remains of the	45703	English	PG	11500000	1993	7.9 The Remains of the D	236		
238	Paul Thomas Ande	26384919	Boogie NightsÂ	189032	English	R	15000000	1997	7.9 Boogie NightsÂ	237		
239	Quentin Tarantinc	54116191	The Hateful EightÂ	272839	English	R	44000000	2015	7.9 The Hateful EightÂ	238		
240	John Carney	9437933	OnceÂ	90827	English	R	180000	2007	7.9 OnceÂ	239		
241	Edward Zwick	26830000	GloryÂ	101888	English	R	18000000	1989	7.9 GloryÂ	240		
242	Edward Zwick	26830000	GloryÂ	101889	English	R	18000000	1989	7.9 GloryÂ	241		
243	Richard Linklater	8114507	Before MidnightÂ	95362	English	R	3000000	2013	7.9 Before MidnightÂ	242		
244	Cristian Mungiu	1185783	4 Months, 3 Weeks	44763	Romanian	Not Rated	590000	2007	7.9 4 Months, 3 Weeks ar	243		
245	Duncan Jones	5009677	MoonÂ	260607	English	R	5000000	2009	7.9 MoonÂ	244		
246	FabiÃn Bielinsky	1221261	Nine QueensÂ	38215	Spanish	R	1500000	2000	7.9 Nine QueensÂ	245		
247	Christophe Barrat	3629758	The ChorusÂ	44151	French	PG-13	5500000	2004	7.9 The ChorusÂ	246		
248	Clint Eastwood	13753931	Letters from Iwo Jir	132149	Japanese	R	19000000	2006	7.9 Letters from Iwo Jim	247		
249	Philip Kaufman	21500000	The Right StuffÂ	45271	English	PG	27000000	1983	7.9 The Right StuffÂ	248		
250	Michael Haneke	225377	AmourÂ	70382	French	PG-13	8900000	2012	7.9 AmourÂ	249		
251	Tim Burton	5887457	Ed WoodÂ	142416	English	R	18000000	1994	7.9 Ed WoodÂ	250		
986												

fig 2.21: top rated movies

- From the table that shows top 250 movies on the basis of imdb rating and users voting. Ranking also provided here.
- So here you can see on the top one position in rank is 'The Shawshank Redemption'. which got imdb rating 9.3. directed by Frank Darabont

FINDINGS

1	top 10 director	Average of imdb_score
2	Akira Kurosawa	8.7
3	Alfred Hitchcock	8.5
4	Asghar Farhadi	8.4
5	Charles Chaplin	8.6
6	Christopher Nolan	8.425
7	Damien Chazelle	8.5
8	Majid Majidi	8.5
9	Richard Marquand	8.4
10	Ron Fricke	8.5
11	Sergio Leone	8.433333333
12	Tony Kaye	8.6

fig 2.22: top 10 director

Row Labels	Sum of PROFIT	Average of imdb_score
Action Adventure Fantasy Sci-Fi	2394717100	6.891666667
Action Adventure Sci-Fi	2205209191	6.652083333
Action Adventure Thriller	1150988238	6.748888889
Adventure Animation Comedy Family Fantasy	1309538719	6.42962963
Comedy	3003940524	5.849275362
Comedy Drama	1414093589	6.56884058
Comedy Drama Romance	1672899623	6.486394558
Comedy Romance	2692838991	5.938167939
Drama Romance	1172699733	6.973913043
Horror	1194653981	5.929545455
Grand Total	18211579689	6.375502959

fig 2.23: genres mostly preferred

- Above table shows the top 10 director on the basis of mean imdb score. Here on the top position we have Akira Kurosawa (8.7)

- From the above table we can see that action and adventure which is the genres is mostly preferred by people, also the average imdb score is not bad. Least popular is horror

FINDINGS

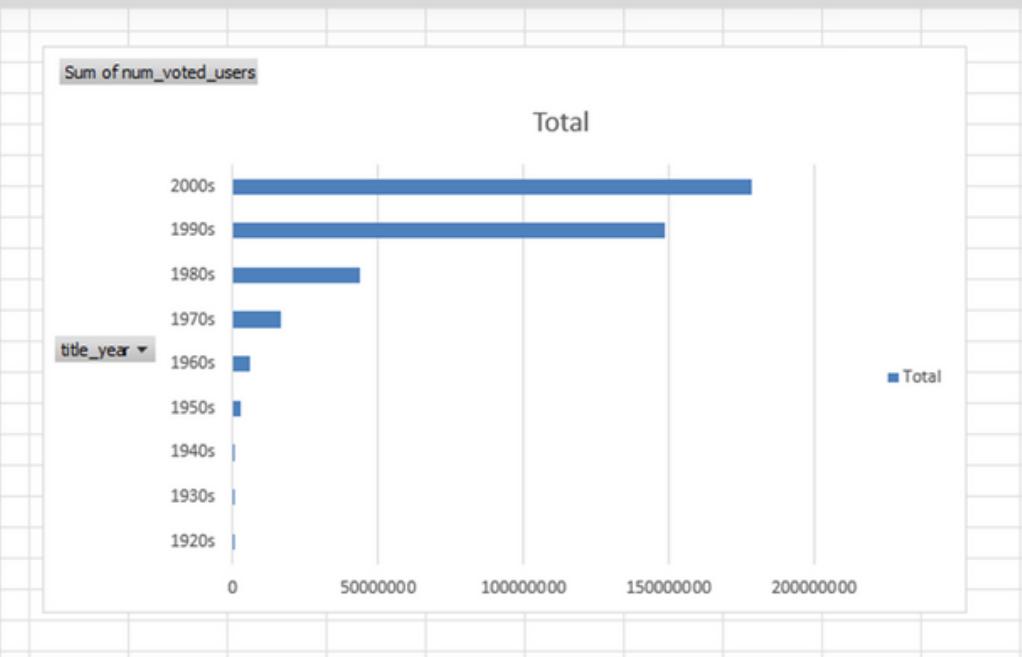


fig 2.24:number of voters over decades

- Observe the change in number of voted users over decades using.
- From the above graph we can see that number of user reviews getting increased over the decades.

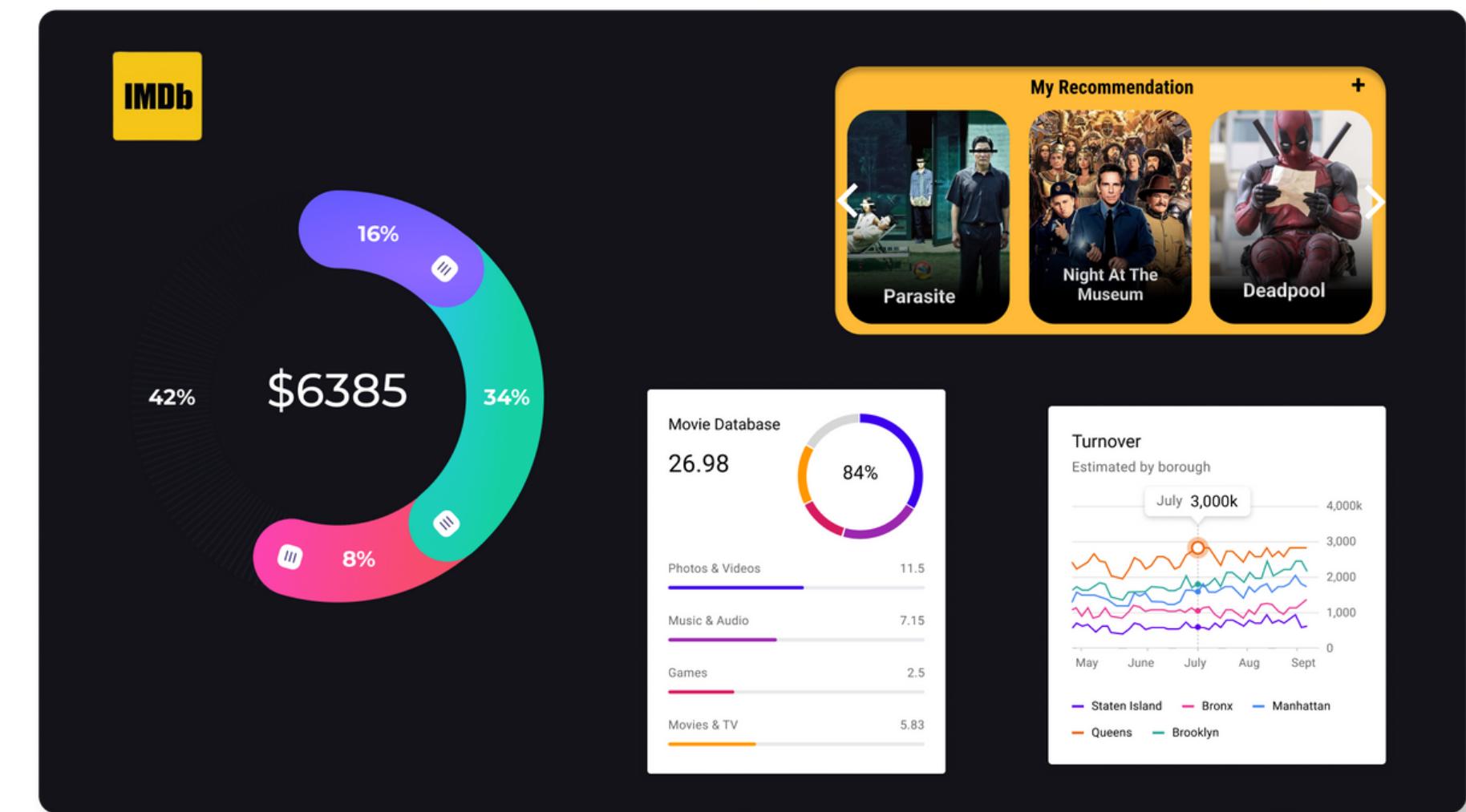
Row Labels	Average of num_critic_for_reviews	Average of num_user_for_reviews
Brad Pitt	245	742.3529412
Leonardo DiCaprio	322.2	922.55
Meryl Streep	181.4545455	297.1818182
Grand Total	262.6041667	715.4166667

fig 2.25: number of critic reviews and user reviews.

- From the above table that shows the average score of number of critic reviews and user reviews.
- On the basis of number of critics Leonardo DiCaprio is in top, second is Brad Pitt and last Meryl Streep

CONCLUSION

In this project IMDB movie analysis I have to use Microsoft office excel. This is used because Microsoft excel work really well for simple calculations and tracking almost any kind of information. Doing this project I have learned lot this technical knowledge vice and movie vice. For technical vice my skills in data analyse have improved, also I have learned more about excel formulas and tools function. Pivot table help me to sorting the data on the basis of average of some values and sum of some values etc. After doing this project I have gain some knowledge about movies also like which is the high gross movie, which actor audience like the most, who is the best director. Which movie get high score in imdb. IMDB movie analysis this project improve my skill and knowledge



06

BANK LOAN CASE STUDY

DESCRIPTION OF THE PROJECT

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also 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

Findings

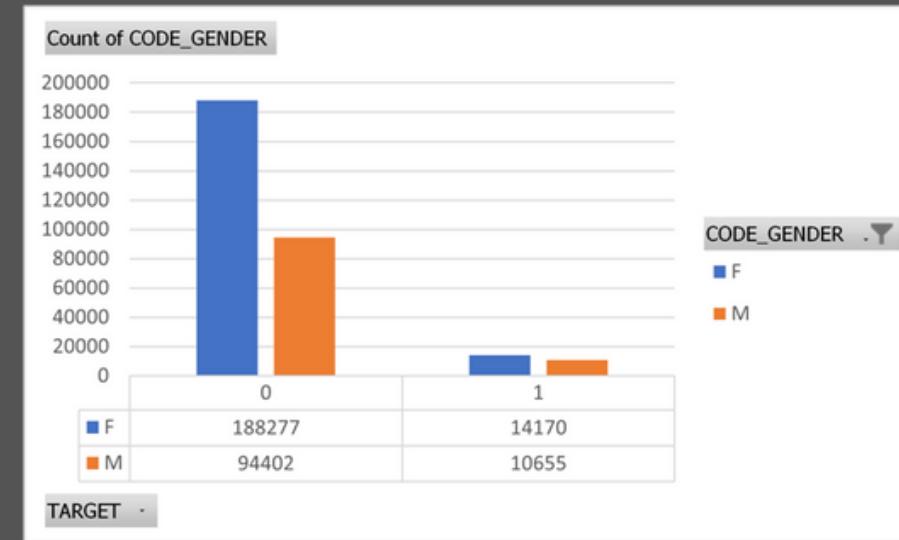


fig 2.26: payment difficulties by gender

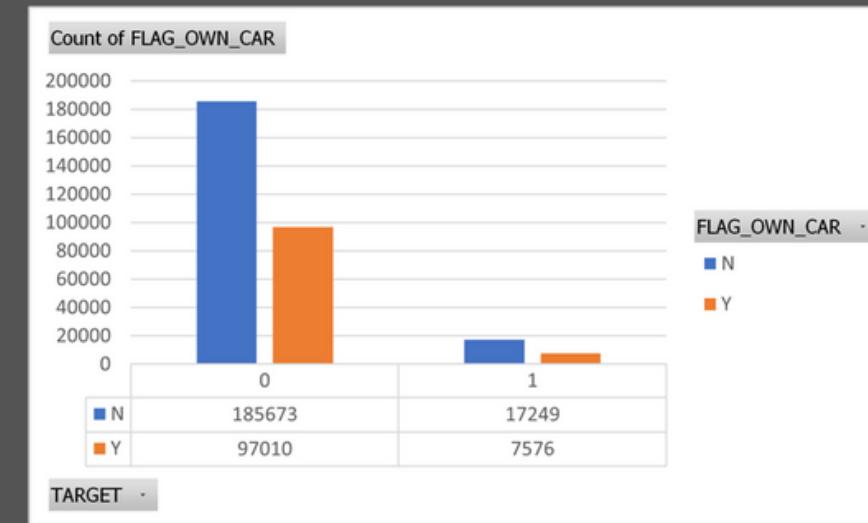


fig 2.27: clients have car or not

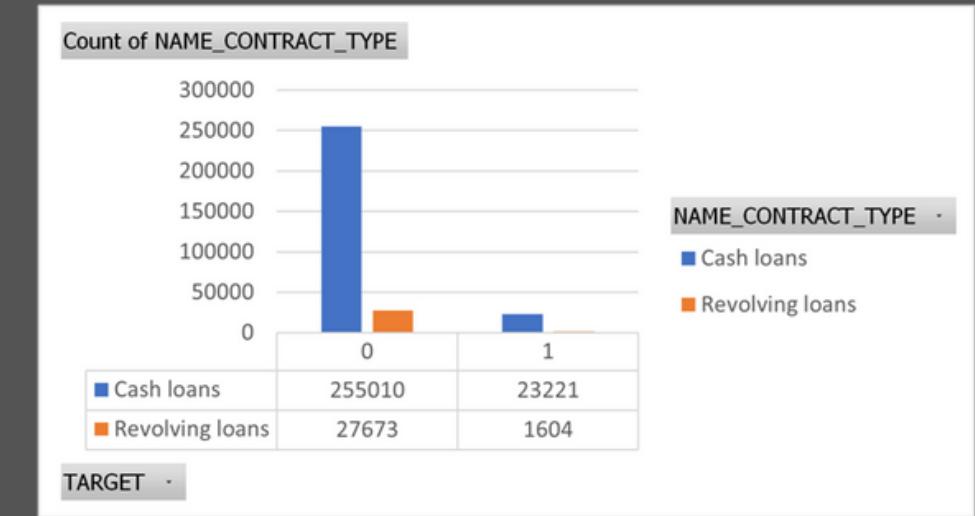


fig 2.28: loan type

- From the above graph we can see that both female and male are almost equal in payment difficulties.
- In the other hand number of female client is more than male for not having payment difficulties

- From the above graph we can infer that most of the client who not having payment difficulties they don't have a car

- From the figure that shows most of the client with on payment difficulties uses cash loans. Revolving loans type are very low compared to cash loans.

Findings

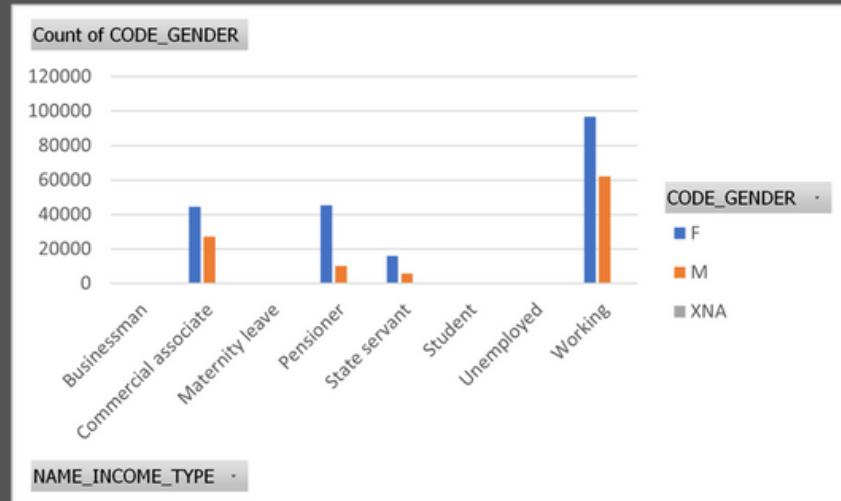


fig 2.29: income type gender vise

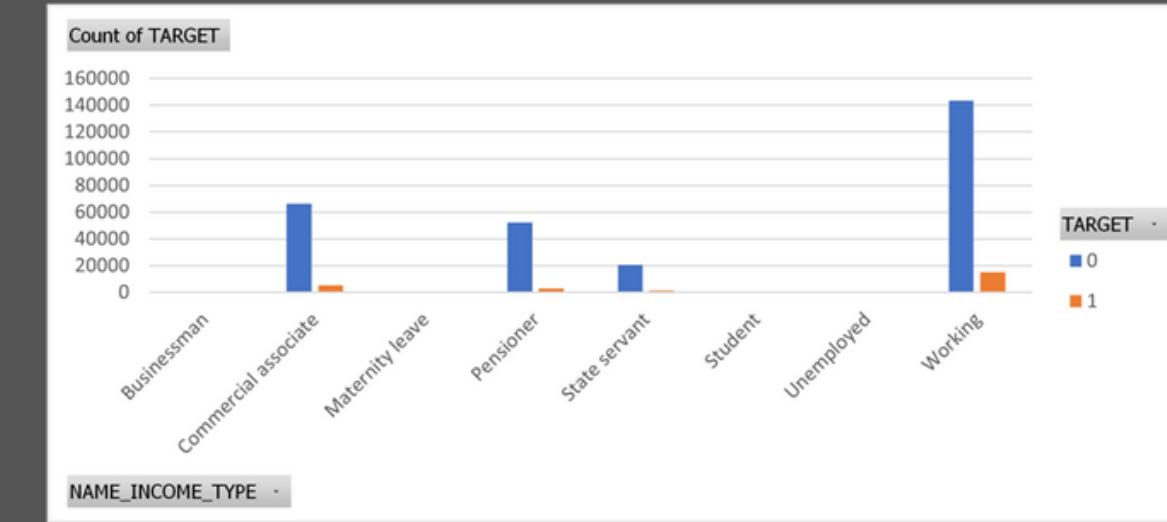


fig 2.30: income type on the basis of payment difficulties

- Above column chart shows income type vs gender. Here we can see that number of female is more than male from the income type of working
- Also we can see that female more in commercial associate and pensioner. Because we already seen that female client are client are higher than male.

- From the above figure we can get that most of the client with no payment difficulties has income from working.

Findings

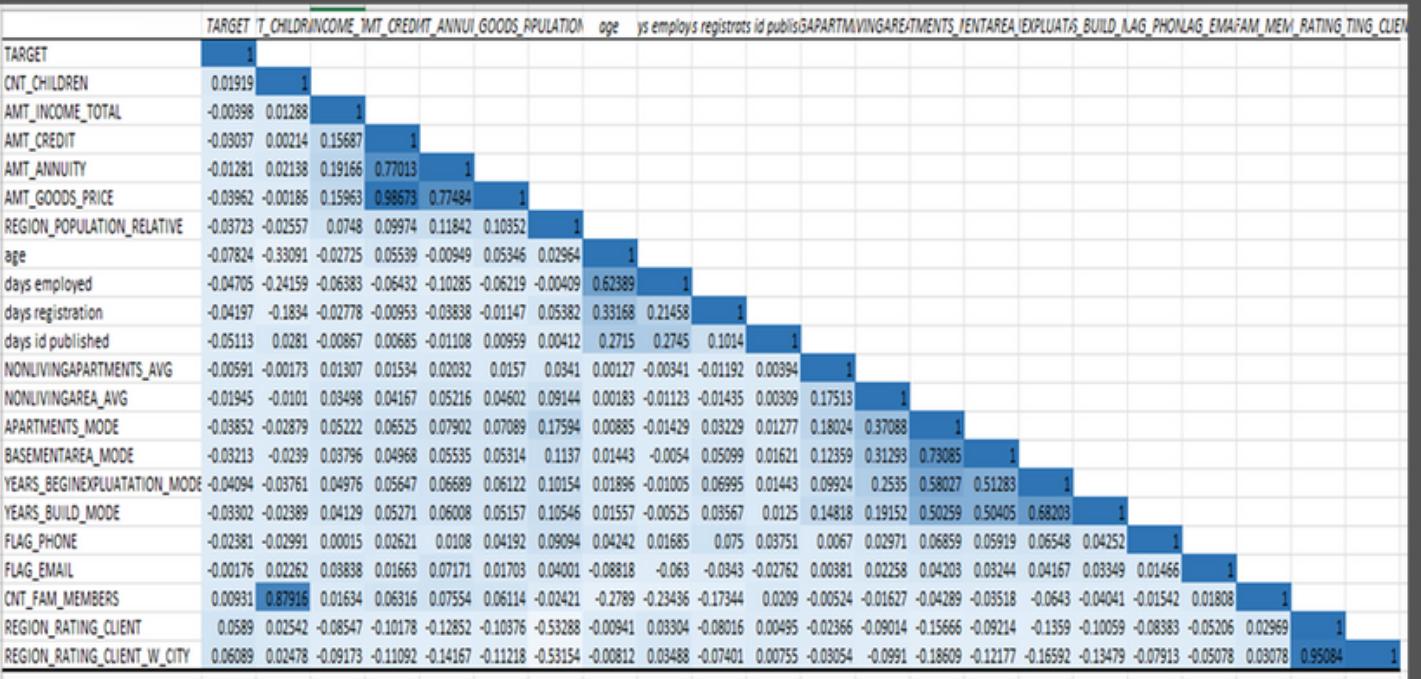


fig 2.31: correlation

- Here are some top positive correlation.
 - AMT_ANNUITY to AMT_CREDIT
 - AMT_GOODS_PRICE to AMT_CREDIT
 - AMT_GOODS_PRICE to AMT_ANNUITY
 - days employed to age
 - BASEMENTAREA_MODE to APARTMENTS_MODE
 - YEARS_BUILD_MODE to YEARS_BEGINEXPLUATATION_MODE
 - CNT_FAM_MEMBERS to CNT_CHILDREN
 - REGION_RATING_CLIENT to REGION_RATING_CLIENT_W_CITY
 - some negative correlations are
 - REGION_RATING_CLIENT to REGION_POPULATION_RELATIVE
 - REGION_POPULATION_RELATIVE to REGION_RATING_CLIENT_W_CITY
 - age to CNT_CHILDREN

CONCLUSION

Loan Case Study



IN THIS PROJECT BANK LOAN CASE STUDY WE ANALYSIS THE DATASETS AND CHECK IF A CLIENT HAS DIFFICULTY PAYING THEIR INSTALMENTS WHICH MAY BE USED FOR TAKING ACTIONS SUCH AS DENYING THE LOAN, REDUCING THE AMOUNT OF LOAN, LENDING AT A HIGHER INTEREST RATE, ETC. THIS WILL ENSURE THAT THE CONSUMERS CAPABLE OF REPAYING THE LOAN ARE NOT REJECTED. IN THIS PROJECT I HAVE USE MICROSOFT EXCEL FOR APPLYING EDA(EXPLORATORY DATA ANALYSIS) WHICH INCREASE MY EXCEL SKILLS AND KNOWLEDGE. I HAVE MOSTLY USE PIVOT TABLE AND PIVOT CHART FOR ANALYSING THE DATA'S. THE TABLES ARE CONTAIN 3 LAKH AND 10 LAKH PLUS ROWS SO IT MAKE DIFFICULT TO ANALYSIS EACH COLUMN. WHEN I USE PIVOT TABLE FOR ANALYSING THE TABLE IT MAKE ME MUCH EASIER FOR ANALYSING EACH COLUMNS AND REDUCE THE TIME ALSO. WE HAVE TO USE EDA TO ANALYSE THE PATTERNS PRESENT IN THE DATA. THIS WILL ENSURE THAT THE APPLICANTS CAPABLE OF REPAYING THE LOAN ARE NOT REJECTED. THIS WILL IMPROVE OUR ANALYTICAL SKILLS AND VISUALIZATION SKILLS

07

XYZ ADS AIRING REPORT ANALYSIS

DESCRIPTION OF THE PROJECT

For your Final Project, we are providing you 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.

FINDINGS

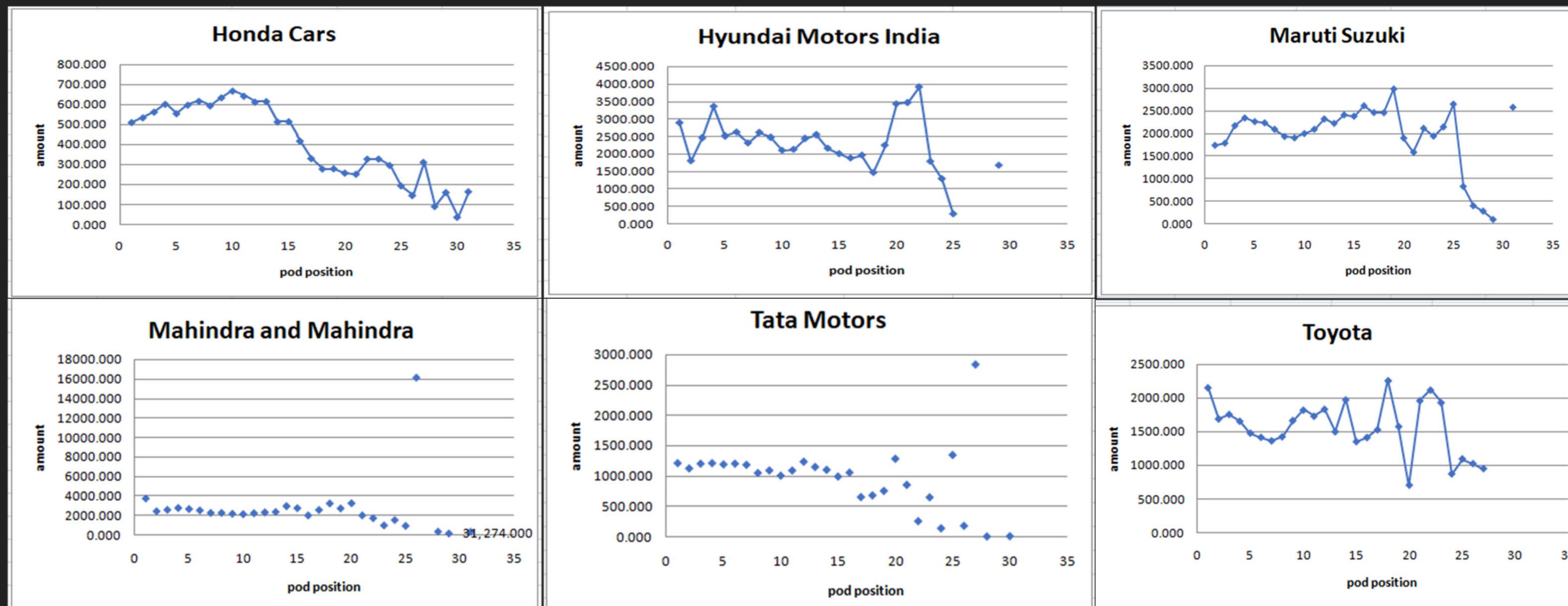


fig 2.32: pod position and amount spend
of different brands

From the graphs we can infer that Pod position number affect the amount spent on Ads for a specific period of time by a company . also From the above graphs we can say that when decrease the pod position spending money will increase. And if the pod position increase spending money will decrease.

FINDINGS

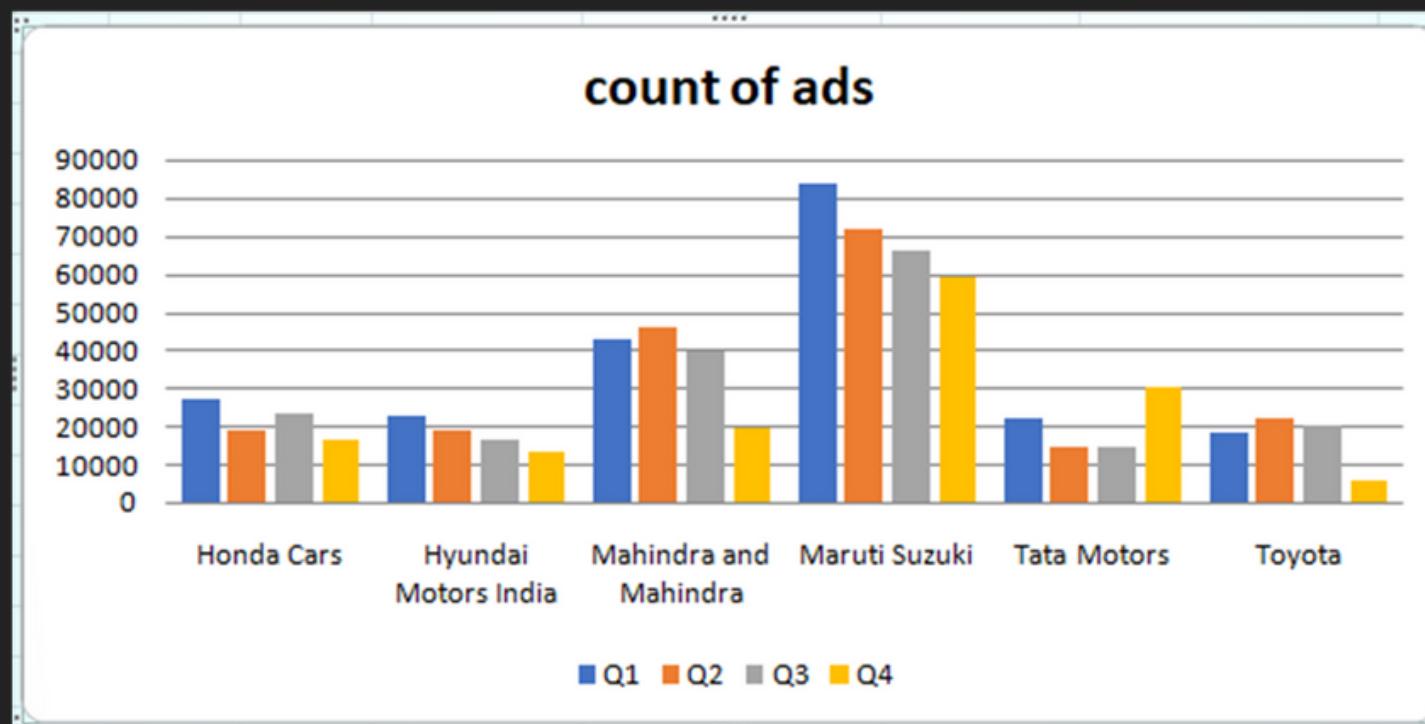


fig 2.33: number of ads of different brands

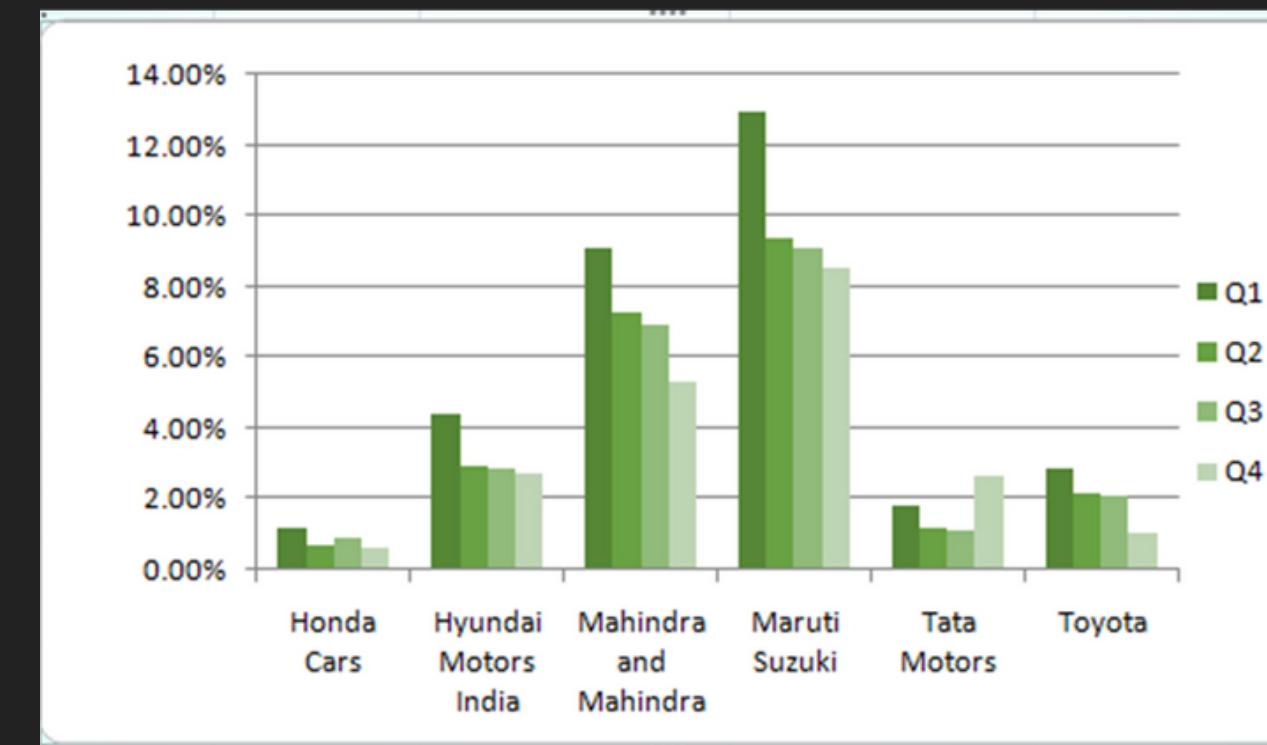


fig 2.34: amount spend by different brands

- Maruti Suzuki has the highest number of ads among the other company
- The share of ads decreases from Q1 to Q4 for Honda cars, Hyundai motors, Mahindra and Mahindra, Maruti Suzuki, Toyota

- Maruti Suzuki has the highest share in amount spent on ads aired followed by Mahindra and Mahindra company
- only Tata motors who increase their share of amount spent on ads aired.
- Honda car has the lowest share of amount spent

FINDINGS

Row Labels	Sum of Spend (\$)									Grand Total
	DAYTIME	EARLY FRINGE	EARLY MORNING	EVENING NEWS	LATE FRINGE	OVERNIGHT	PRIME ACCESS	PRIME TIME	WEEKEND	
Honda Cars	31%	12%	11%	7%	1%	6%	3%	15%	11%	100%
Hyundai Motors India	7%	4%	5%	3%	8%	2%	4%	48%	20%	100%
Mahindra and Mahindra	16%	5%	3%	4%	11%	2%	3%	38%	18%	100%
Maruti Suzuki	9%	4%	5%	4%	13%	4%	5%	38%	17%	100%
Tata Motors	17%	6%	8%	6%	12%	3%	6%	27%	15%	100%
Toyota	16%	9%	7%	5%	8%	2%	8%	21%	24%	100%
Grand Total	0.125945966	0.050754111	0.050517939	0.039859892	0.109838692	0.03050862	0.04536211	0.366159016	0.181053653	100%

fig 2.35: amount spend on dayparts by different brands

- Above table shows the amount spend on ads aired in each dayparts
- Taking Honda car which we can see that. they spend more money in daytime
- In case of Hyundai motors. They spend more money at prime time
- Mahindra and Mahindra company also spend more money in primetime
- In case of Maruti Suzuki which also more moneys in prime time
- Tata motors also spend more moneys in the prime time
- Toyota company spend more moneys in weekends and also in prime time.
- All brands spend least amount in overnight ads airing

FINDINGS

Count of Id	Column Labels									
Row Labels	DAYTIME	EARLY FRINGE	EARLY MORNING	EVENING NEWS	LATE FRINGE	OVERNIGHT	PRIME ACCESS	PRIME TIME	WEEKEND	Grand Total
Honda Cars	3.83%	0.96%	2.03%	0.42%	0.84%	0.78%	0.26%	1.15%	1.37%	11.65%
Hyundai Motors India	1.71%	0.60%	1.11%	0.33%	1.40%	0.90%	0.41%	1.85%	1.44%	9.74%
Mahindra and Mahindra	4.58%	1.28%	1.91%	0.62%	3.34%	0.89%	0.55%	3.87%	3.17%	20.21%
Maruti Suzuki	6.81%	2.34%	4.79%	1.36%	6.18%	3.72%	1.29%	6.52%	5.28%	38.30%
Tata Motors	2.42%	0.73%	1.26%	0.41%	1.52%	0.80%	0.41%	1.85%	1.66%	11.07%
Toyota	2.38%	0.63%	1.20%	0.35%	1.05%	0.47%	0.32%	1.26%	1.36%	9.02%
Grand Total	21.73%	6.54%	12.32%	3.49%	14.34%	7.56%	3.25%	16.49%	14.29%	100.00%

fig 2.36: number of ads and dayparts

Sum of Spend(₹)	Column Labels									
Row Labels	DAYTIME	EARLY FRINGE	EARLY MORNING	EVENING NEWS	LATE FRINGE	OVERNIGHT	PRIME ACCESS	PRIME TIME	WEEKEND	Grand Total
Honda Cars	31.30%	11.94%	10.76%	4.36%	7.09%	5.84%	2.80%	14.51%	11.39%	100.00%
Hyundai Motors India	6.84%	3.96%	4.82%	2.97%	7.55%	1.76%	4.27%	47.97%	19.88%	100.00%
Mahindra and Mahindra	16.15%	4.83%	3.05%	4.03%	10.52%	2.16%	2.59%	38.44%	18.23%	100.00%
Maruti Suzuki	8.71%	4.07%	5.18%	3.72%	13.26%	4.23%	5.19%	38.24%	17.40%	100.00%
Tata Motors	17.42%	6.39%	7.50%	6.15%	11.77%	2.69%	6.12%	27.06%	14.89%	100.00%
Toyota	16.48%	8.65%	7.36%	4.80%	7.87%	1.52%	7.97%	21.43%	23.91%	100.00%
Grand Total	12.59%	5.08%	5.05%	3.99%	10.98%	3.05%	4.54%	36.62%	18.11%	100.00%

fig 2.37: amount spend on dayparts by different brands

- from the tables we can see that 38.24% amount spend prime time ads airing. But there are only 19.14% of ads they are produced
- In daytime they have 22.51% number of ads aired but they only spend 16.15% of amount for ads aired. So they have to focus on daytime.
- may be increase the ads of mahindra and mahindra company will get chance of cover more audience

- The table shows dayparts used by each company. In this table we can simply that almost every company spend more moneys in the prime time.
- Target audience:** mahindra and mahindra company should target primetime, prime access, daytime and weekend audiences, because most of the audience watch tv in the evening and primetime, and there are audience on daytime and weekends also.
- if mahindra and mahindra can use daytime,prime access, prime time and weekend. May be that will change in the ads views

CONCLUSION



IN THIS PROJECT XYZ ADS AIRING REPORT ANALYSIS, WE HAVE SEVERAL TASKS TO PERFORM. TO PERFORM THE TASKS I HAVE USED EXCEL SO THE TASKS WHICH GIVEN IN THIS PROJECT THAT INCREASE MY EXCEL SKILLS AND KNOWLEDGE. THE PROJECT XYZ ADS AIRING REPORT ANALYSIS, WHICH ALSO HELPS TO IMPROVE OUR ANALYTICAL SKILLS, VISUALIZATION SKILLS, DATA ANALYTICAL SKILLS. THE DATASET HAVING DIFFERENT TV AIRING BRANDS, THEIR PRODUCT, THEIR CATEGORY ALSO IT INCLUDES THE NETWORK THROUGH WHICH ADS ARE AIRING, TYPES OF NETWORK LIKE CABLE/ BROADCAST AND THE SHOW NAME ALSO ON WHICH ADS GOT AIRED. IN THIS PROJECT WE HAVE TO ANALYSE THE ADVERTISEMENT STRATEGY OF EACH BRANDS.

08

ABC CALL VOLUME TREND ANALYSIS

DESCRIPTION OF THE PROJECT

For your final project we are providing you with a dataset of a Customer Experience (CX) Inbound calling team for 23 days. Data includes Agent_Name, Agent_ID, Queue_Time [duration for which customer have to wait before they get connected to an agent], Time [time at which call was made by customer in a day], Time_Bucket [for easiness we have also provided you with the time bucket], Duration [duration for which a customer and executives are on call], Call_Seconds [for simplicity we have also converted those time into seconds], call status (Abandon, answered, transferred).

FINDINGS

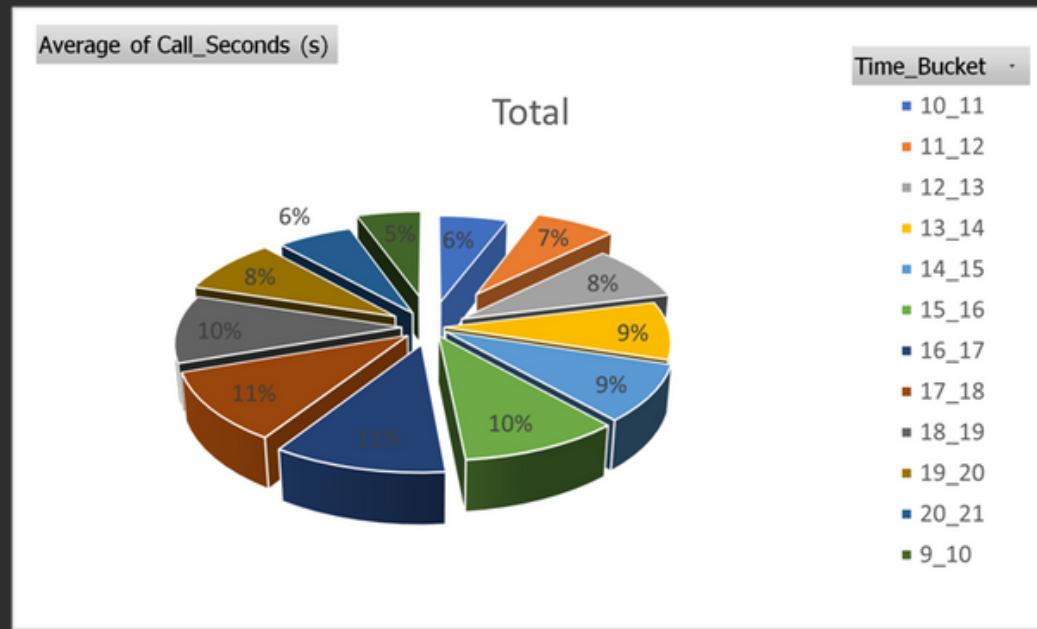


fig 2.38: avg call tiime duration in each time bucket

- ABOVE GRAPH AND SHOWS THE AVERAGE CALL TIME DURATION FOR ALL INCOMING CALLS RECEIVED BY AGENTS DURING EACH TIME BUCKET.
- FROM THE GRAPH WE CAN SEE THAT THEY MAKE A LONG CONVERSATION DURING 17-18 AND 18-19 TIME BUCKET.

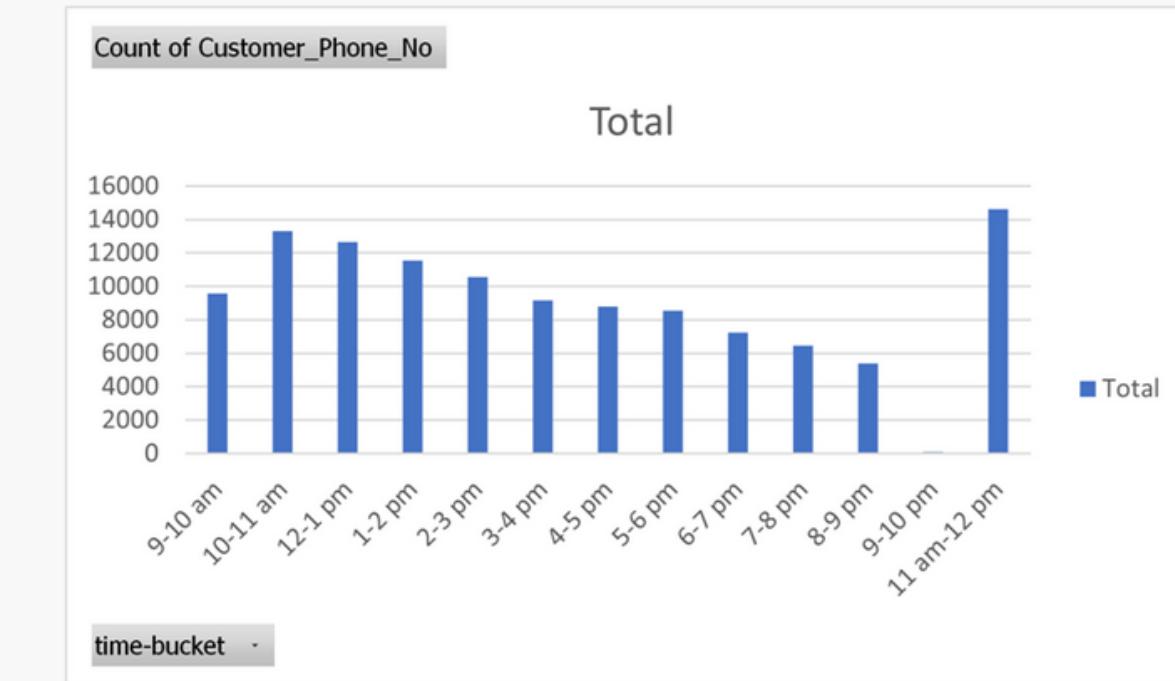


fig 2.39: number of calls in each time bucket

- ABOVE SHOWS THE GRAPHICAL REPRESENTATION OF NUMBER OF CALLS MAKES ON EACH TIME BUCKET
- WE CAN INFER FROM THE GRAPH THAT PEOPLE MAKES MORE CALLS 11 AM -12 PM TIME BUCKET. ALSO IN 10 – 11 AM.

FINDINGS

time-bucket	avg calls	total call time(hrs)	number of agent required
10-11 am	578.826087	20.19090098	4
11-12 am	635.9130435	22.18223674	5
1-2 pm	502.6521739	17.53376446	4
12-1 pm	550.0869565	19.18840826	4
2-3 pm	459.173913	16.01713402	4
3-4 pm	398.2173913	13.89081815	3
4-5 pm	382.0869565	13.32814826	3
5-6 pm	371.0434783	12.94292413	3
6-7 pm	314.6956522	10.97737109	2
7-8 pm	281	9.8019825	2
8-9 pm	234.3043478	8.173121413	2
9-10 am	416.8695652	14.54145261	3
9-10 pm	5.043478261	0.17592913	1
grand total	5129.913043	178.9441917	40

fig 2.41: number of agent required in each time buck

- ABOVE GRAPH AND SHOWS THE AVERAGE CALL TIME DURATION FOR ALL INCOMING CALLS RECEIVED BY AGENTS DURING EACH TIME BUCKET.

time-bucket	avg call	total call time(hrs)	no of man required
10-11 am	173.6478	6.057270293	1
11-12 am	190.7739	6.654671022	1
1-2 pm	150.7957	5.260129337	1
12-1 pm	165.0261	5.756522478	1
2-3 pm	137.7522	4.805140207	1
3-4 pm	119.4652	4.167245446	1
4-5 pm	114.6261	3.998444478	1
5-6 pm	111.313	3.882877239	1
6-7 pm	94.4087	3.293211326	1
7-8 pm	84.3	2.94059475	1
8-9 pm	70.2913	2.451936424	1
9-10 am	125.0609	4.362435783	1
9-10 pm	1.513043	0.052778739	0
grand total	1539	53.68325752	12

fig 2.40: number of agent required in night

- HERE WE CAN SEE THAT NUMBER OF MAN POWER REQUIRED IN EACH TIME BUCKET IN THE NIGHT TIME



CONCLUSION

In this project ABC Call Volume Trend Analysis we analysis customer experience. A customer experience team consists of professionals who analyse customer feedback and data, and share insights with the rest of the organization. Keep a good customer relationship that help business for future growth. The project done through using Microsoft excel. So it help me to increase my technical skills and knowledge in excel. In this analysis project I have analysis the customer relationship, to solve customer problem that keep good relationship between the business and customer. After doing this project this help me to improve my data analytical skills, visualization skills etc.

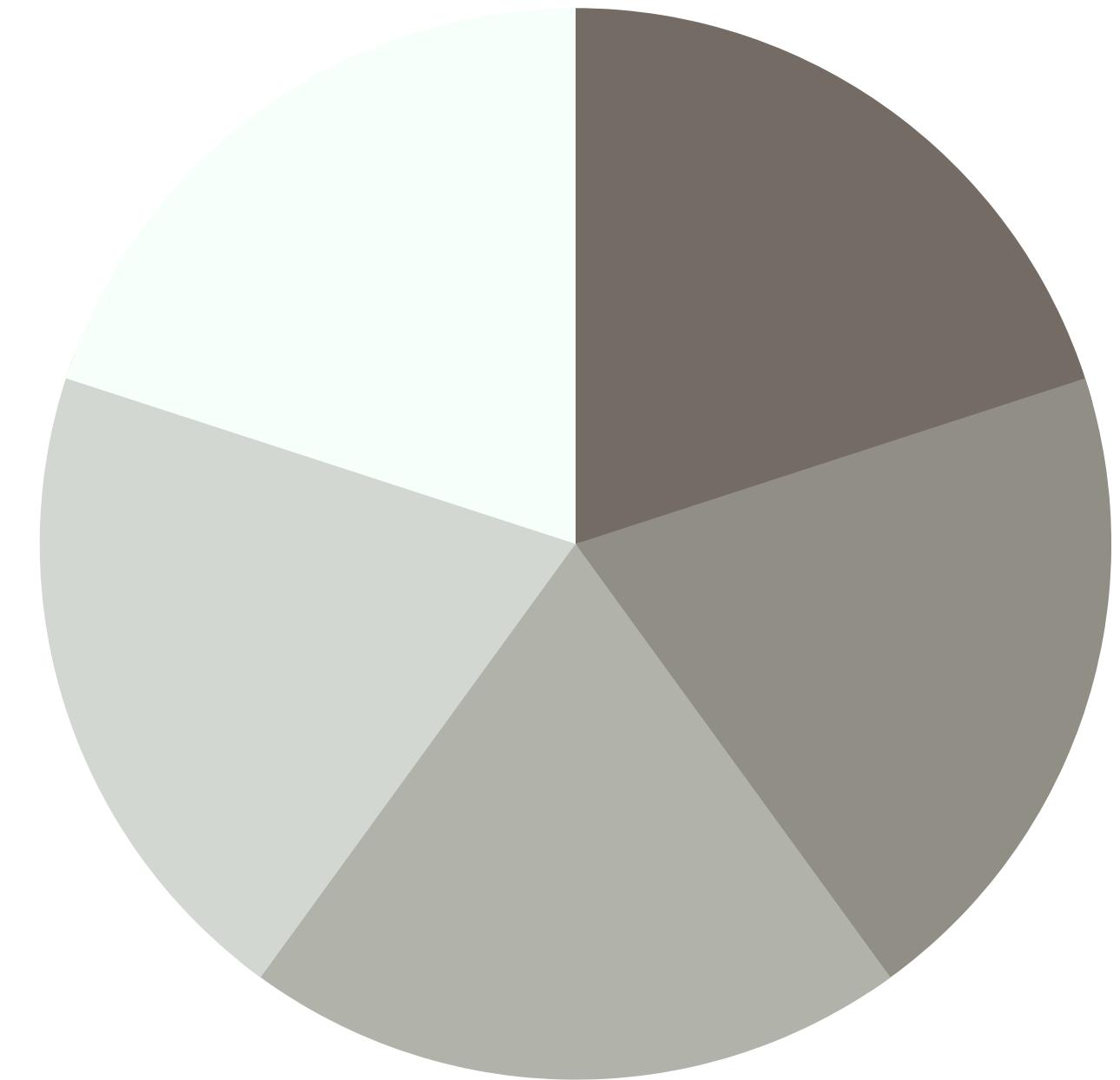


03. LEARNIGS



Data analytics is the science of analyzing raw data in order to make conclusions about that information. Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system.

I have learned a lot by doing this projects. from the first project we can understand what exactly is data analytics, actually we data analysis in our daily life situations. from these projects I have earned skills like data analysis, data visualization, predictive analytics, mathematical skills, presentation skill, data base management etc. for doing this projects I have use tools like SQL, excel, My SQL, charts and graph etc. it improve my skill sets, I learned more about excel and pivot table and chart. I have also familiar with SQL queries.



04. APPENDIX



01 DATA ANALYSIS PROCESS

https://drive.google.com/file/d/1tPUR6UY20ottNDWGB9-rMOvGl_Dzs_Hk/view?usp=share_link

02 INSTAGRAM USER ANALYTICS

https://drive.google.com/file/d/10PU4ICKGL1vnfuqIF5aMc1wD7hNyYyln/view?usp=share_link

03 OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

https://drive.google.com/file/d/158YCqDPRfoFzkc2KnMmK3MH_Q5mCWftH/view?usp=share_link

04 HIRING PROCESS ANALYTICS

https://drive.google.com/file/d/1NXYthgw40GxXKX0a0OuvqMq2R6pvioVr/view?usp=share_link

05 IMDB MOVIE ANALYSIS

https://drive.google.com/file/d/16kNahPcLI6FfUxRlVZSp2bNBPRzNSaC/view?usp=share_link

06 BANK LOAN CASE STUDY

https://drive.google.com/file/d/1fp4KQVIQDjsCVJ2sYVCXfSvSVQarNcu-/view?usp=share_link

07 XYZ ADS AIRING REPORT ANALYSIS

https://drive.google.com/file/d/16FGST13nYAvpzH4LRJiHujxf5rqJ7L6h/view?usp=share_link

<https://www.loom.com/share/db788ae48b7c4b94922739700f837bb5>

08 ABC CALL VOLUME TREND ANALYSIS

https://drive.google.com/file/d/1QUm9iyXOT6KEHaPrKY7-WIrl-pVdMMbO/view?usp=share_link

<https://www.loom.com/share/aeda50680d894b49adcd04643faa022b>

05. PERSONAL DETAILS:

Address

kumbalakode (H),
amarakuni (PO), pulpally,
wayanad, pin 673579, Kerala, India

contact

+91-8281904639

+91-8086311768

WhatsApp +91-80831178

e-mail

gokulkrishnan7211@gmail.com

THANK YOU