

AVINASH H



Data Analyst

[My Resume](#)

Education :

Bachelors of business administration

Skills :

Advance excel, Power BI, SQL , MIS, Power Point.

Functional area :

Data Analysis, Data Visualization, business Intelligence, MIS Reporting , Statistics

Virtual Experience :

3 virtual Experience [PwC, JP Morgan, Trainity]

Internship :

1 Internship [Kulturehire]

Projects :

14 Industry Level Projects
[Excel, Power BI, SQL, Power Point]



[My LinkedIn profile](#)



EXCEL Portfolio



IMDb Movie Analysis

MOVIE ANALYSIS

Description

IMDB is a movie rating and review company with the global scope .analyzing the movies ,directors and actors

Performance based on the review given in imdb

It will help to suggestion of movies and types of movies which are interest specified audience

Approach

First analyze the data which are required means (column and rows)and deleted
Some of the unwanted data which is not required for the analysing and
visualization purpose and started analysis

Tech-Stack

excel PowerPoint notepad

INSIGHT

In the imdb movie analysis I got so many information like which is highest earned movie and decade wise movie list as well no.of movies released in particular year and top actor and the best director based on the imdb score And the actors information like in which movies they acted and year of release, Popular genre because which genre is most attracting the audience and I learnt the advanced excel skills like VLOOKUP pivot table and using of table in advance and complex problems .

Cleaning the data

Cleaning the data is very important process in data analytics

Before cleaning the data there are 5044 rows and 28 columns are their

And cleaned the some of the unwanted column and rows that was not important for analysis

After cleaning data 16 column and 3813

Rows

director_name
num_critic_for_reviews
duration
gross
genres
actor_1_name
movie_title
num_voted_users
actor_3_name
plot_keywords
num_user_for_reviews
language
country
budget
title_year
imdb_score

Movies with highest profit

TOP 10 MOVIES BASED ON THE PROFIT

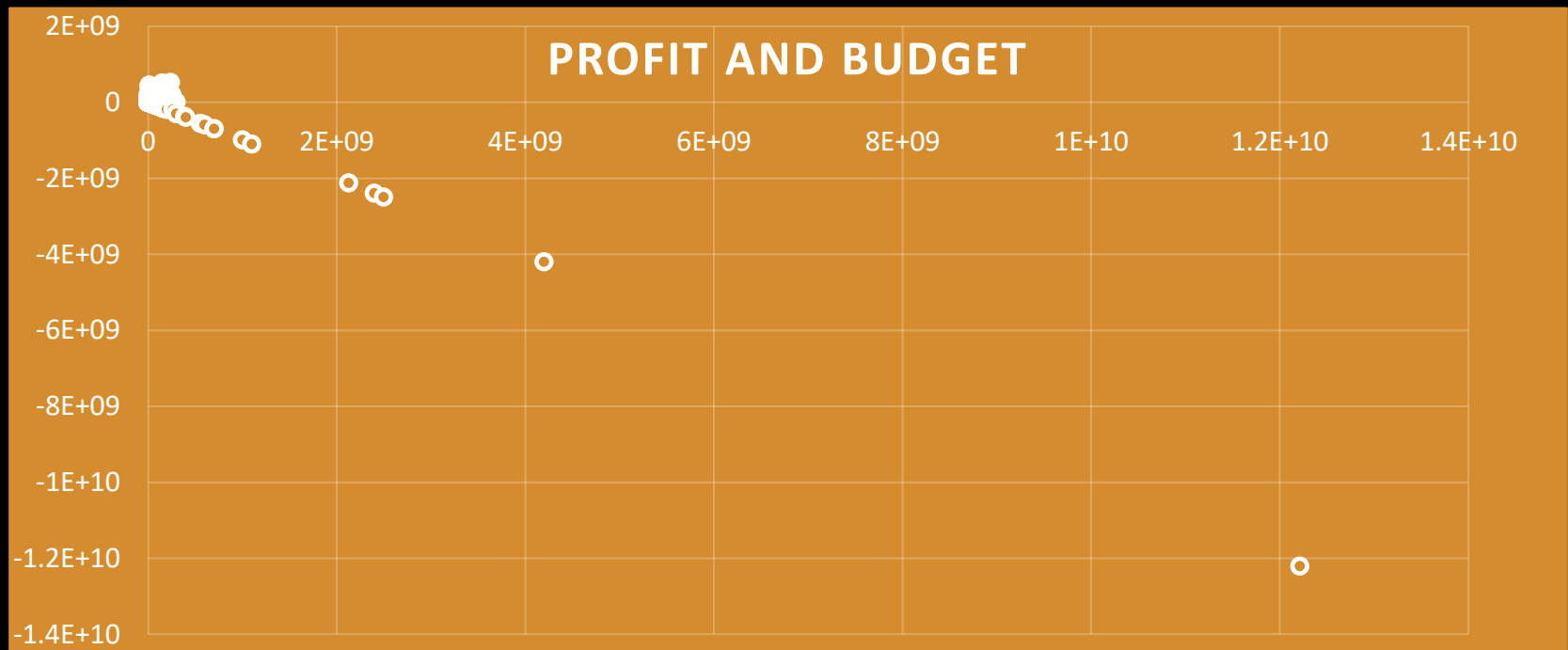
Profit = gross – budget

Avatar is top movie in terms of profit

Jurassic world is after the avatar

moviename	profit
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Star Wars: Episode IV - A New Hope	449935665
E.T. the Extra-Terrestrial	424449459
The Avengers	403279547
The Lion King	377783777
The Jungle Book	375290282
Star Wars: Episode I - The Phantom Menace	359544677
The Dark Knight	348316061

SCATTER CHART OF PROFIT AND BUDGET



TOP 250 MOVIES BASED IMDB SCORE

1	movie name	imdb_score	num_voted_users	rank
2	The Shawshank Redemption	9.3	1689764	1
3	The Godfather	9.2	1155770	2
4	The Godfather: Part II	9	790926	3
5	The Dark Knight	9	1676169	3
6	The Good, the Bad and the Ugly	8.9	503509	5
7	The Lord of the Rings: The Return of the King	8.9	1215718	5
8	Schindler's List	8.9	865020	5
9	Pulp Fiction	8.9	1324680	5
10	The Lord of the Rings: The Fellowship Ring	8.8	1238746	9
11	Star Wars: Episode V - The Empire Strikes Back	8.8	837759	9
12	Inception	8.8	1468200	9
13	Forrest Gump	8.8	1251222	9
14	Fight Club	8.8	1347461	9
15	The Matrix	8.7	1217752	14
16	The Lord of the Rings: The Two Towers	8.7	1100446	14
17	Seven Samurai	8.7	229012	14
18	Star Wars: Episode IV - A New Hope	8.7	91097	14
19	One Flew Over the Cuckoo's Nest	8.7	680041	14
20	Goodfellas	8.7	728685	14
21	City of God	8.7	533200	14
22	The Usual Suspects	8.6	740918	21
23	Se7en	8.6	1023511	21
24	The Silence of the Lambs	8.6	887467	21
25	Saving Private Ryan	8.6	881236	21
26	Spirited Away	8.6	417971	21
27	Modern Times	8.6	143086	21
28	Interstellar	8.6	928227	21
29	American History X	8.6	782437	21
30	The Prestige	8.5	844052	29
31	Terminator 2: Judgment Day	8.5	744891	29
32	The Green Mile	8.5	782610	29
33	The Lives of Others	8.5	259379	29
34	The Pianist	8.5	497946	29
35	Whiplash	8.5	399138	29
36	Raiders of the Lost Ark	8.5	661017	29
37	Samsara	8.5	22457	29
38	The Lion King	8.5	644348	29
39	Psycho	8.5	422432	29
40	The Departed	8.5	873649	29
41	The Dark Knight Rises	8.5	1144337	29
42	Back to the Future	8.5	732212	29
43	Memento	8.5	845580	29
44	Alien	8.5	563827	29
45	Children of Heaven	8.5	27882	29
46	Apocalypse Now	8.5	450676	29

43	The Other Dream Team	8.4	3086	48
50	Once Upon a Time in America	8.4	221000	48
51	WALL-E	8.4	718837	48
52	Requiem for a Dream	8.4	573541	48
53	Princess Mononoke	8.4	221552	48
54	Reservoir Dogs	8.4	664719	48
55	Oldboy	8.4	356181	48
56	Star Wars: Episode VI - Return of the Jedi	8.4	681857	48
57	Lawrence of Arabia	8.4	192775	48
58	Amélie	8.4	534262	48
59	Das Boot	8.4	168203	48
60	Bahubali: The Beginning	8.4	62756	48
61	American Beauty	8.4	822500	48
62	Braveheart	8.4	736638	48
63	A Separation	8.4	151812	48
64	Aliens	8.4	488537	48
65	Toy Story 3	8.3	544884	64
66	Room	8.3	161288	64
67	Unforgiven	8.3	277505	64
68	No End in Sight	8.3	7314	64
69	The Sting	8.3	175607	64
70	Scarface	8.3	537442	64
71	Toy Story	8.3	623757	64
72	Snatch	8.3	600996	64
73	Up	8.3	665575	64
74	Some Like It Hot	8.3	175196	64
75	Raging Bull	8.3	235133	64
76	The Hunt	8.3	170155	64
77	Intolerable Cruelty	8.3	885175	64
78	Downfall	8.3	248354	64
79	Monty Python and the Holy Grail	8.3	382240	64
80	2001: A Space Odyssey	8.3	427357	64
81	Indiana Jones and the Last Crusade	8.3	515306	64
82	Metropolis	8.3	111841	64
83	Inside Out	8.3	345198	64
84	Batman Begins	8.3	980946	64
85	L.A. Confidential	8.3	414219	64
86	Hoop Dreams	8.3	18980	64
87	Eternal Sunshine of the Spotless Mind	8.3	666937	64
88	Amadeus	8.3	270790	64
89	Good Will Hunting	8.3	604904	64
90	The Thing	8.2	258078	89
91	Pan's Labyrinth	8.2	467234	89
92	Trainspotting	8.2	469561	89
93	On the Waterfront	8.2	100890	89
94	The Secret in Their Eyes	8.2	131831	89

109	Lock, Stock and Two Smoking Barrels	8.2	414976	89
110	A Beautiful Mind	8.2	610568	89
111	Howl's Moving Castle	8.2	214091	89
112	Die Hard	8.2	592582	89
113	Gone with the Wind	8.2	215340	89
114	The Sixth Sense	8.1	704766	113
115	Nothing But a Man	8.1	891	113
116	The Wizard of Oz	8.1	291875	113
117	The Celebration	8.1	65951	113
118	Rush	8.1	312629	113
119	Stand by Me	8.1	271794	113
120	The Avengers	8.1	995415	113
121	The Grand Budapest Hotel	8.1	475518	113
122	There Will Be Blood	8.1	372990	113
123	The Help	8.1	318955	113
124	The Bourne Ultimatum	8.1	491077	113
125	The Imitation Game	8.1	467613	113
126	Shutter Island	8.1	786092	113
127	Pirates of the Caribbean: The Curse of the Black Pearl	8.1	809474	113
128	The Terminator	8.1	600266	113
129	Rocky	8.1	375240	113
130	The Truman Show	8.1	667983	113
131	Woodstock	8.1	12631	113
132	The Best Years of Our Lives	8.1	40359	113
133	Spotlight	8.1	195333	113
134	Sin City	8.1	656640	113
135	The Princess Bride	8.1	294163	113
136	No Country for Old Men	8.1	612060	113
137	The Revenant	8.1	406020	113
138	Tae Guk Gi: The Brotherhood of Swords	8.1	31943	113
139	The Sea Inside	8.1	64556	113
140	Platoon	8.1	291603	113
141	The Martian	8.1	472488	113
142	Prisoners	8.1	383591	113
143	Hotel Rwanda	8.1	264533	113
144	Groundhog Day	8.1	437418	113
145	Jurassic Park	8.1	613473	113
146	In the Shadow of the Moon	8.1	5475	113
147	Deadpool	8.1	479047	113
148	12 Years a Slave	8.1	439176	113
149	Million Dollar Baby	8.1	482064	113
150	Akira	8.1	106160	113
151	Kill Bill: Vol. 1	8.1	735784	113
152	Gone Girl	8.1	569841	113
153	Guardians of the Galaxy	8.1	682155	113
154	Amores Perros	8.1	173551	113

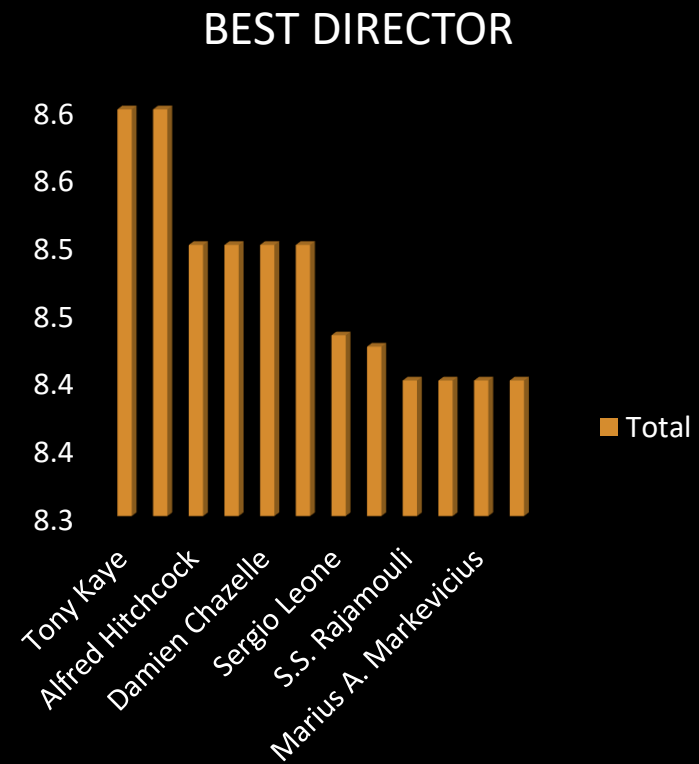
205	Dances with Wolves	8	186485	161
206	Catch Me If You Can	8	525801	161
207	Bowling for Columbine	8	123090	161
208	Aladdin	8	260939	161
209	Boys n the Hood	8	266020	161
210	Central Station	8	28951	161
211	Before Sunset	8	168398	161
212	Jaws	8	412454	161
213	Fiddler on the Roof	8	29839	161
214	Doctor Zhivago	8	55816	161
215	The Remains of the Day	7.9	45703	214
216	The Hatful of Eight	7.9	272839	214
217	The Blues Brothers	7.9	142448	214
218	The Untouchables	7.9	219008	214
219	Shrek	7.9	467113	214
220	The Chorus	7.9	44151	214
221	Walk the Line	7.9	188637	214
222	The Bourne Identity	7.9	407601	214
223	The Right Stuff	7.9	45271	214
224	The Hobbit: The Desolation of Smaug	7.9	483540	214
225	The Notebook	7.9	396396	214
226	The World's Fastest Indian	7.9	44198	214
227	The Insider	7.9	133526	214
228	The Wrestler	7.9	251349	214
229	Veer-Zaara	7.9	34449	214
230	The Second Mother	7.9	7025	214
231	Once	7.9	90827	214
232	Toy Story 2	7.9	385871	214
233	Taken	7.9	483756	214
234	The Fighter	7.9	275869	214
235	Straight Outta Compton	7.9	119328	214
236	Nightcrawler	7.9	293304	214
237	Nine Queens	7.9	38215	214
238	The Hobbit: An Unexpected Journey	7.9	637246	214
239	Little Miss Sunshine	7.9	355810	214
240	Children of Men	7.9	361767	214
241	Almost Famous	7.9	207287	214
242	Glory	7.9	203777	214
243	Letters from Iwo Jima	7.9	132149	214
244	Do the Right Thing	7.9	59524	214
245	Ernest & Celestine	7.9	12029	214
246	Crash	7.9	361169	214
247	4 Months, 3 Weeks and 2 Days	7.9	44763	214
248	Crouching Tiger, Hidden Dragon	7.9	217740	214
249	Moon	7.9	260607	214
250	Iron Man	7.9	696338	214

OTHER LANGUAGE TOP MOVIES

movie_title	language	num_voted_us	imdb_score	rank
The Good, the Bad and the Ugly	Italian	503509	8.9	5
City of God	Portuguese	533200	8.7	14
Seven Samurai	Japanese	229012	8.7	14
Spirited Away	Japanese	417971	8.6	21
The Lives of Others	German	259379	8.5	29
Children of Heaven	Persian	27882	8.5	29
Samsara	None	22457	8.5	29
A Separation	Persian	151812	8.4	48
Oldboy	Korean	356181	8.4	48
Das Boot	German	168203	8.4	48
Baahubali: The Beginning	Telugu	62756	8.4	48
Amélie	French	534262	8.4	48
Princess Mononoke	Japanese	221552	8.4	48
The Hunt	Danish	170155	8.3	64
Metropolis	German	111841	8.3	64
Downfall	German	248354	8.3	64
Pan's Labyrinth	Spanish	467234	8.2	89
The Secret in Their Eyes	Spanish	131831	8.2	89
Incendies	French	80429	8.2	89
The Act of Killing	Indonesian	23836	8.2	89
Howl's Moving Castle	Japanese	214091	8.2	89
Amores Perros	Spanish	173551	8.1	113
The Celebration	Danish	65951	8.1	113
Elite Squad	Portuguese	81644	8.1	113
The Sea Inside	Spanish	64556	8.1	113
Tae Guk Gi: The Brotherhood of Blood	Korean	31943	8.1	113
Akira	Japanese	106160	8.1	113
A Fistful of Dollars	Italian	147566	8	161
Central Station	Portuguese	28951	8	161
Waltz with Bashir	Hebrew	46107	8	161
Persepolis	French	70194	8	161
My Name Is Khan	Hindi	69759	8	161
Crouching Tiger, Hidden Dragon	Mandarin	217740	7.9	215
4 Months, 3 Weeks and 2 Days	Romanian	44763	7.9	215
Nine Queens	Spanish	38215	7.9	215
The Chorus	French	44151	7.9	215
The Second Mother	Portuguese	7025	7.9	215
Veer-Zaara	Hindi	34449	7.9	215
Letters from Iwo Jima	Japanese	132149	7.9	215

Best Directors

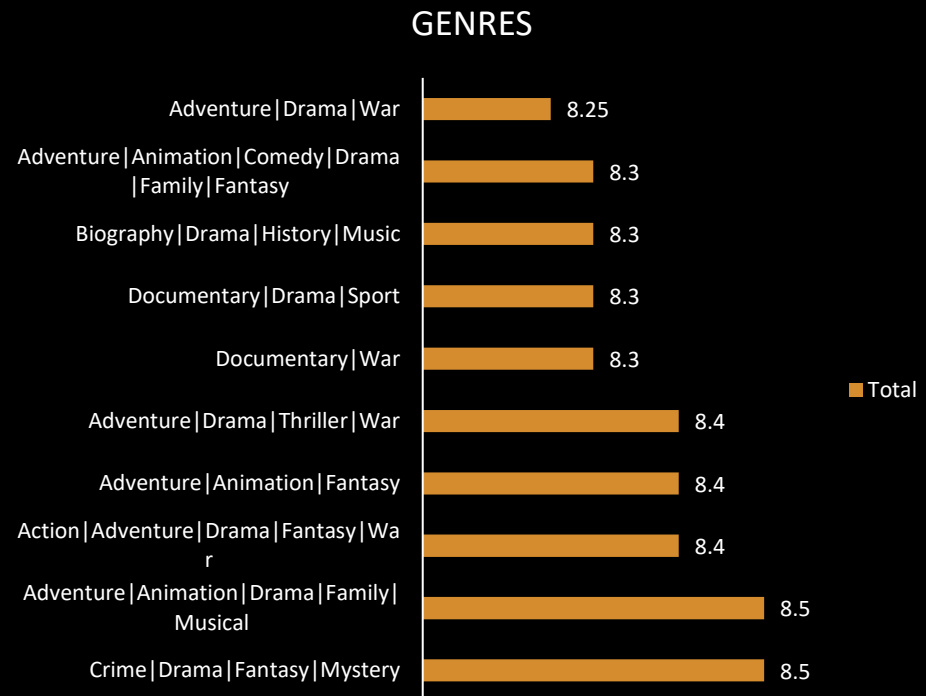
DIRECTOR_NAME	AVERAGE OF IMDB_SCORE
TONY KAYE	8.6
CHARLES CHAPLIN	8.6
ALFRED HITCHCOCK	8.5
RON FRICKE	8.5
DAMIEN CHAZELLE	8.5
MAJID MAJIDI	8.5
SERGIO LEONE	8.4
CHRISTOPHER NOLAN	8.4
S.S. RAJAMOULI	8.4
RICHARD MARQUAND	8.4
MARIUS A. MARKEVICIUS	8.4
ASGHAR FARHADI	8.4



Popular Genres:

CINEMA DRAMA FANTASY MYSTERY AND ADEVENTUE ANIMATION DRAMA
FAMILY MUSICAL HAD MORE IMDB SCORE

genres	Average of imdb_score
Crime Drama Fantasy Mystery	8.5
Adventure Animation Drama Family Musical	8.5
Action Adventure Drama Fantasy War	8.4
Adventure Animation Fantasy	8.4
Adventure Drama Thriller War	8.4
Documentary War	8.3
Documentary Drama Sport	8.3
Biography Drama History Music	8.3
Adventure Animation Comedy Drama Family Fantasy	8.3
Adventure Drama War	8.25



Creating new table

Grouping the movies of Meryl Streep, Leonardo DiCaprio, and Brad Pitt

Using the excel I done the grouping of these actors movies which they acted as lead

Meryl Streep	Leonardo DiCaprio	Brad Pitt
Out of AfricaÂ	The Quick and the DeadÂ	True RomanceÂ
The River WildÂ	Romeo + JulietÂ	Interview with the Vampire: The Vampire ChroniclesÂ
One True ThingÂ	Marvin's RoomÂ	Seven Years in TibetÂ
The HoursÂ	TitanicÂ	Fight ClubÂ
The Devil Wears PradaÂ	The Man in the Iron MaskÂ	Ocean's ElevenÂ
A Prairie Home CompanionÂ	The BeachÂ	Spy GameÂ
Lions for LambsÂ	Catch Me If You CanÂ	Sinbad: Legend of the Seven SeasÂ
Julie & JuliaÂ	Gangs of New YorkÂ	Ocean's TwelveÂ
It's ComplicatedÂ	The AviatorÂ	TroyÂ
The Iron LadyÂ	The DepartedÂ	Mr. & Mrs. SmithÂ
Hope SpringsÂ	Blood DiamondÂ	BabelÂ
	Revolutionary RoadÂ	The Assassination of Jesse James by the Coward Robert FordÂ
	Body of LiesÂ	The Curious Case of Benjamin ButtonÂ
	InceptionÂ	The Tree of LifeÂ
	Shutter IslandÂ	Killing Them SoftlyÂ
	J. EdgarÂ	FuryÂ
	Django UnchainedÂ	By the SeaÂ
	The Great GatsbyÂ	
	The Great GatsbyÂ	
	The Wolf of Wall StreetÂ	
	The RevenantÂ	

Finding the mean of the actors

Using pivot table I extract the average of Anum users for review and numb critic for review of the extract and extracted table is below

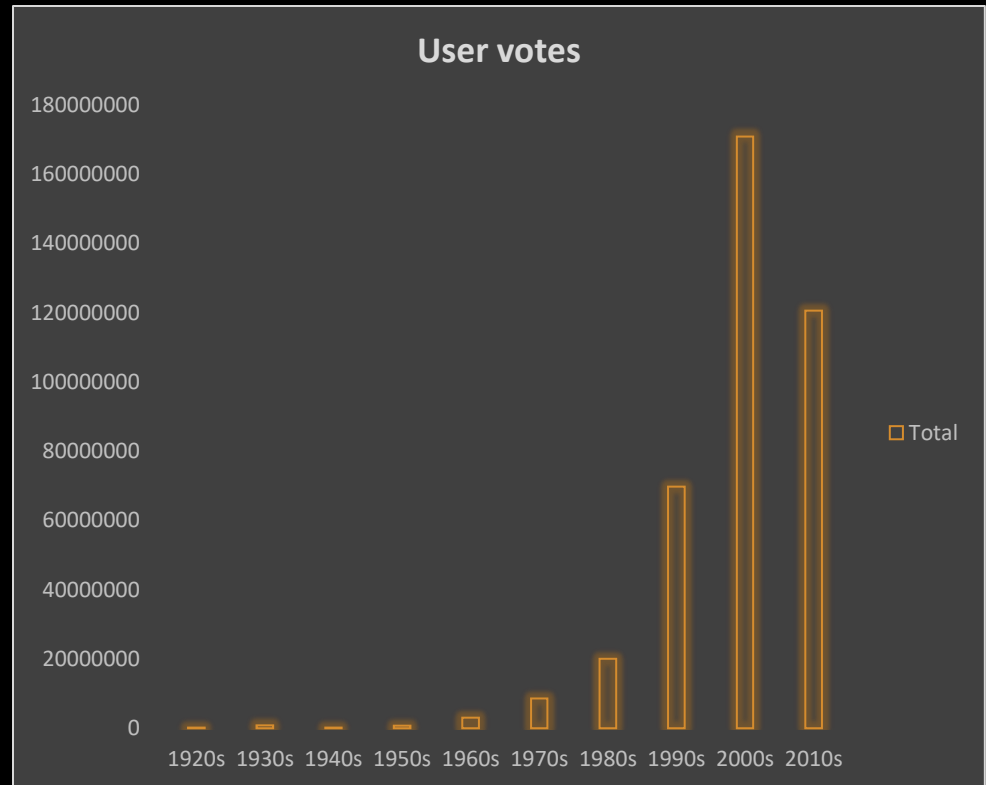
actor name	Average of num_user_for_reviews	Average of num_critic_for_reviews
Albert Finney	1498	750
Phaldut Sharma	1885	738
Peter Capaldi	995	654
Craig Stark	1018	596
BÃ©rÃ©nice Bejo	583	576
Suraj Sharma	755	552
Ellar Coltrane	836	548
Mike Howard	405	546
Lou Taylor Pucci	789	543
Joel Courtney	849	539
Maika Monroe	631	533
Tim Holmes	511	525
Elina Alminas	611	489
Kurt Fuller	509	487
Iko Uwais	316	481
QuvenzhanÃ© Wallis	392.3333333	478.6666667
Edgar Arreola	461	478
Sharlto Copley	1262	472
Cory Hardict	326	452
Matt Frewer	1229	451
Aidan Turner	894.25	447
Elizabeth McGovern	801	447
Michael Fassbender	837.5	434.3333333
Wood Harris	588	432
Jennifer Lawrence	583.2142857	418.9285714
Anil Kapoor	1017	418
Jessica Barden	438	417
Sally Hawkins	384	415
Danielle Kottch	524	411
Micah Sloat	1189	409
Ivana Baquero	1083	406
Christo Jivkov	2814	406

actor who had highest average num_user_for_reviews	Heather Donahue
actor who hsd highest average of num_critic_for_reviews	Albert Finney

Analysing the decades

To analyze the decade ,first done grouping of decades using the VLOOKUP
And used the bar chart for further analyze

decades	number of user voted
1920s	116387
1930s	804839
1940s	159517
1950s	678336
1960s	2983442
1970s	8523299
1980s	19987476
1990s	69735679
2000s	170876429
2010s	120535554



Grouping the movies based on decade

In the extracted data made the group of movies based on decades

1920s	1930s	1940s	1950s	1970s	1980s	1990s	2020s
Metropolis	42nd Street	Duel in the Sun	Annie Get Your Gun	A Bridge Too Far	3 Men and a Baby	10 Things I Hate About You	While We're Young
The Broadway Melody	Gone with the Wind	Pinocchio	On the Waterfront	Alien	9½ Weeks	20 Dates	Whiplash
	Modern Times	The Best Years of Our Lives	Seven Samurai	American Graffiti	A Nightmare on Elm Street	200 Cigarettes	White House Down
	Snow White and the Seven Dwarfs	The Lady from Shanghai	Some Like It Hot	Animal House	A Nightmare on Elm Street	3 Ninjas Kick Back	Why Did I Get Married Too?
	The Wizard of Oz	The Pirate	The Beast from 20,000 Fathoms	Annie Hall	A Nightmare on Elm Street	54	Wild Target
	Top Hat		The Bridge on the River Kwai	Apocalypse Now	A Nightmare on Elm Street	8 Heads in a Duffel Bag	Wild
			The Greatest Show on Earth	Benji	A Nightmare on Elm Street	8MM	Winnie the Pooh
			The Robe	Beyond the Valley of the Drums	A Passage to India	A Bug's Life	Winter's Bone
				Blazing Saddles	A Room with a View	A Civil Action	Winter's Tale
				Caravans	A View to a Kill	A Dog of Flanders	Wish I Was Here
				Close Encounters of the Third Kind	Action Jackson	A Few Good Men	Woman in Gold
				Darling Lili	Airplane!	A League of Their Own	World War Z
				Diamonds Are Forever	Akira	A Low Down Dirty Shame	Wrath of the Titans
				Escape from Alcatraz	Aliens	A Night at the Roxbury	Wreck-It Ralph
				Fiddler on the Roof	Amadeus	A Simple Plan	X-Men: Apocalypse
				Grease	American Ninja 2: The Conquest	A Simple Wish	X-Men: Days of Future Past
				Halloween	April Fool's Day	A Thin Line Between Love and Hate	X-Men: First Class
				Jaws 2	Back to the Future Part II	A Time to Kill	Yogi Bear
				Jaws	Back to the Future	A Walk on the Moon	You Again
				Kingdom of the Spiders	Batman	Absolute Power	You Will Meet a Tall Dark Stranger
				Live and Let Die	Beetlejuice	Ace Ventura: Pet Detective	Young Adult
				Logan's Run	Beverly Hills Cop II	Ace Ventura: When Nature Calls	Your Highness
				March or Die	Beverly Hills Cop	Aimee & Jaguar	Your Sister's Sister

Bank Loan Case Study

Bank Loan Case Study

DESCRIPTION

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.

Business Understanding:

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers. You must use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

Approach

first downloaded the two csv file and imported into excel for the analysis this is the large set of data and it takes more time for cleaning ,after wards I started the analysis (eda) of given data

Tech-Stack Used

excel ,power point ,notepad

Insight

BY DOING THIS ANALYSIS I GOT KNOW ABOUT THE EDA AND IT IS VERY IMPORTANT FOR EVERY DATA ANALYST AND ALSO LEARNED THE ADVANCE EXCEL TECHNIQUES AS WELL FORMULA OF THE EXCEL I GOT MORE CONFIDENCE IN CLEANING THE DATA AND VISUALIZING THE DATA AND MADE THE CORRELATION AND DIFFERENT TYPE OF THE ANALYSIS

Present the overall approach of the **analysis**. Mention the problem statement and the analysis approach briefly

There are three data sets namely **application_data.csv** contains all the information of the client at the time of application.

The data is about whether a client has payment difficulties.

previous_application.csv contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

columns_description.csv is data dictionary which describes the meaning of the variables.

There are so many unwanted rows and column are present in the data(in two data sets) first I removed the unwanted data from the given data after that I started the EDA

the given datasets are

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	SK_ID_PRI	SK_ID_CUI	NAME_CC	AMT_ANNUITY	AMT_APP	AMT_CREDIT	AMT_DOWRY	AMT_GOC	WEEKDAY	HOUR_AP	FLAG_LAS	NFLAG_LAS	RATE_DOWRY	RATE_INTEREST	RATE_INTEREST	NAME_CA	NAME_CC	DAYS_DEC	NAME_PA	CODE_REJ	NAME
2	2030495	271877	Consumer	1730.43	17145	17145	0	17145	SATURDAY	15 Y		1	0	0.182832	0.867336	XAP	Approved	-73	Cash thro	XAP	
3	2802425	108129	Cash loan:	25188.62	607500	679671		607500	THURSDAY	11 Y		1				XNA	Approved	-164	XNA	XAP	Unac
4	2523466	122040	Cash loan:	15060.74	112500	136444.5		112500	TUESDAY	11 Y		1				XNA	Approved	-301	Cash thro	XAP	Spot
5	2819243	176158	Cash loan:	47041.34	450000	470790		450000	MONDAY	7 Y		1				XNA	Approved	-512	Cash thro	XAP	
6	1784265	202054	Cash loan:	31924.4	337500	404055		337500	THURSDAY	9 Y		1				Repairs	Refused	-781	Cash thro	HC	
7	1383531	199383	Cash loan:	23703.93	315000	340573.5		315000	SATURDAY	8 Y		1				Everyday	Approved	-684	Cash thro	XAP	Fami
8	2315218	175704	Cash loans		0	0			TUESDAY	11 Y		1				XNA	Canceled	-14	XNA	XAP	
9	1656711	296299	Cash loans		0	0			MONDAY	7 Y		1				XNA	Canceled	-21	XNA	XAP	
10	2367563	342292	Cash loans		0	0			MONDAY	15 Y		1				XNA	Canceled	-386	XNA	XAP	
11	2579447	334349	Cash loans		0	0			SATURDAY	15 Y		1				XNA	Canceled	-57	XNA	XAP	
12	1715995	447712	Cash loan:	11368.62	270000	335754		270000	FRIDAY	7 Y		1				XNA	Approved	-735	Cash thro	XAP	Unac
13	2257824	161140	Cash loan:	13832.78	211500	246397.5		211500	FRIDAY	10 Y		1				XNA	Approved	-815	Cash thro	XAP	Unac
14	2330894	258628	Cash loan:	12165.21	148500	174361.5		148500	TUESDAY	15 Y		1				XNA	Approved	-860	Cash thro	XAP	Unac
15	1397919	321676	Consumer	7654.86	53779.5	57564	0	53779.5	SUNDAY	15 Y		1	0			XAP	Approved	-408	Cash thro	XAP	Unac
16	2273188	270658	Consumer	9644.22	26550	27252	0	26550	SATURDAY	10 Y		1	0			XAP	Approved	-726	Cash thro	XAP	
17	1232483	151612	Consumer	21307.46	126490.5	119853	12649.5	126490.5	TUESDAY	7 Y		1	0.103971			XAP	Approved	-699	Cash thro	XAP	Unac
18	2163253	154602	Consumer	4187.34	26955	27297	1350	26955	SATURDAY	12 Y		1	0.051324			XAP	Approved	-1473	Cash thro	XAP	Unac
19	1285768	142748	Revolving	9000	180000	180000		180000	FRIDAY	13 Y		1				XAP	Approved	-336	XNA	XAP	Unac
20	2393109	396305	Cash loan:	10181.7	180000	180000		180000	THURSDAY	14 Y		1				XNA	Approved	-700	Cash thro	XAP	Unac
21	1173070	199178	Cash loan:	4666.5	45000	49455		45000	SATURDAY	16 Y		1				Everyday	Refused	-584	XNA	HC	
22	1506815	166490	Cash loan:	25454.03	450000	491580		450000	MONDAY	6 Y		1				XNA	Refused	-401	Cash thro	HC	Unac
23	1182516	267782	Cash loan:	20361.6	405000	451777.5		405000	SATURDAY	4 Y		1				XNA	Approved	-686	Cash thro	XAP	
24	1172842	302212	Cash loans		0	0			TUESDAY	9 Y		1				XNA	Refused	-239	XNA	HC	
25	1172937	302212	Cash loan:	39475.31	1129500	1277105		1129500	THURSDAY	5 Y		1				XNA	Refused	-594	Cash thro	HC	
26	1555330	199353	Cash loans		0	0			SATURDAY	6 Y		1				XNA	Canceled	-202	XNA	XAP	
27	1543131	275707	Cash loan:	22619.52	229500	241920		229500	THURSDAY	8 Y		1				XNA	Approved	-370	Cash thro	XAP	Unac
1	SK_ID_CUI	TARGET	NAME_CC	CODE_GET	FLAG_OW	FLAG_OW	CNT_CHIL	AMT_INC	AMT_CREDIT	AMT_ANNUITY	AMT_GOC	NAME_TY	NAME_IN	NAME_ED	NAME_FA	NAME_HC	REGION_F	DAYS_BIR	DAYS_EMI	DAYS_REG	DAYS
2	100002	1	Cash loan: M		N	Y	0	202500	406597.5	24700.5	351000	Unaccom	Working	Secondary	Single / n	House / a	0.018801	-9461	-637	-3648	-2
3	100003	0	Cash loan: F		N	N	0	270000	1293503	35698.5	1129500	Family	State serv	Higher ed	Married	House / a	0.003541	-16765	-1188	-1186	-
4	100004	0	Revolving M		Y	Y	0	67500	135000	6750	135000	Unaccom	Working	Secondary	Single / n	House / a	0.010032	-19046	-225	-4260	-2
5	100006	0	Cash loan: F		N	Y	0	135000	312682.5	29686.5	297000	Unaccom	Working	Secondary	Civil marri	House / a	0.008019	-19005	-3039	-9833	-2
6	100007	0	Cash loan: M		N	Y	0	121500	513000	21865.5	513000	Unaccom	Working	Secondary	Single / n	House / a	0.028663	-19932	-3038	-4311	-3
7	100008	0	Cash loan: M		N	Y	0	99000	490495.5	27515.7	454500	Spouse, p	State serv	Secondary	Married	House / a	0.035792	-16941	-1588	-4970	-
8	100009	0	Cash loan: F		Y	Y	1	171000	1560726	41301	1395000	Unaccom	Commerci	Higher ed	Married	House / a	0.035792	-13778	-3130	-1213	-
9	100010	0	Cash loan: M		Y	Y	0	360000	1530000	42075	1530000	Unaccom	State serv	Higher ed	Married	House / a	0.003122	-18850	-449	-4597	-2
10	100011	0	Cash loan: F		N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary	Married	House / a	0.018634	-20099	365243	-7427	-3
11	100012	0	Revolving M		N	Y	0	135000	405000	20250	405000	Unaccom	Working	Secondary	Single / n	House / a	0.019689	-14469	-2019	-14437	-3
12	100014	0	Cash loan: F		N	Y	1	112500	652500	21177	652500	Unaccom	Working	Higher ed	Married	House / a	0.0228	-10197	-679	-4427	-
13	100015	0	Cash loan: F		N	Y	0	38419.16	148365	10678.5	135000	Children	Pensioner	Secondary	Married	House / a	0.015221	-20417	365243	-5246	-2
14	100016	0	Cash loan: F		N	Y	0	67500	80865	5881.5	67500	Unaccom	Working	Secondary	Married	House / a	0.031329	-13439	-2717	-311	-3
15	100017	0	Cash loan: M		Y	N	1	225000	918468	28966.5	697500	Unaccom	Working	Secondary	Married	House / a	0.016612	-14086	-3028	-643	-4
16	100018	0	Cash loan: F		N	Y	0	189000	773680.5	32778	679500	Unaccom	Working	Secondary	Married	House / a	0.010006	-14583	-203	-615	-2
17	100019	0	Cash loan: M		Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary	Single / n	Rented ap	0.020713	-8728	-1157	-3494	-1
18	100020	0	Cash loan: M		N	N	0	108000	509602.5	26149.5	387000	Unaccom	Working	Secondary	Married	House / a	0.018634	-12931	-1317	-6392	-3
19	100021	0	Revolving F		N	Y	1	81000	270000	13500	270000	Unaccom	Working	Secondary	Married	House / a	0.010966	-9776	-191	-4143	-2
20	100022	0	Revolving F		N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary	Widow	House / a	0.04622	-17718	-7804	-8751	-1
21	100023	0	Cash loan: F		N	Y	1	90000	544491	17563.5	454500	Unaccom	State serv	Higher ed	Single / n	House / a	0.015221	-11348	-2038	-1021	-3
22	100024	0	Revolving M		Y	Y	0	135000	427500	21375	427500	Unaccom	Working	Secondary	Married	House / a	0.015221	-18252	-4286	-298	-1
23	100025	0	Cash loan: F		Y	Y	1	202500	1132574	37561.5	927000	Unaccom	Commerci	Secondary	Married	House / a	0.025164	-14815	-1652	-2299	-2
24	100026	0	Cash loan: F		N	N	1	450000	497520	32521.5	450000	Unaccom	Working	Secondary	Married	Rented ap	0.020713	-11146	-4306	-114	-2
25	100027	0	Cash loan: F		N	Y	0	83250	239850	23850	225000	Unaccom	Pensioner	Secondary	Married	House / a	0.006296	-24827	365243	-9012	-3
26	100029	0	Cash loan: M		Y	N	2	135000	247500	12703.5	247500	Unaccom	Working	Secondary	Married	House / a	0.026392	-11286	-746	-108	-3
27	100030	0	Cash loan: F		N	Y	0	90000	225000	11074.5	225000	Unaccom	Working	Secondary	Married	House / a	0.028663	-19334	-3494	-2419	-2
28	100031	0	Cash loan: F		N	Y	0	112500	513000	21865.5	513000	Unaccom	Working	Secondary	Single / n	House / a	0.028663	-19932	-3038	-4311	-3
application data																					

application_data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		Table	Row	Descriptio	Special																
2		1 applicatio	SK_ID_CUR	ID of loan in our sample																	
3		2 applicatio	TARGET	Target variable (1 - client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample, 0 - all other cases)																	
4		5 applicatio	NAME_CC	Identification if loan is cash or revolving																	
5		6 applicatio	CODE_GEO	Gender of the client																	
6		7 applicatio	FLAG_OW	Flag if the client owns a car																	
7		8 applicatio	FLAG_OW	Flag if client owns a house or flat																	
8		9 applicatio	CNT_CHILD	Number of children the client has																	
9		10 applicatio	AMT_INCOME	Income of the client																	
10		11 applicatio	AMT_CREDIT	Credit amount of the loan																	
11		12 applicatio	AMT_ANNUITY	Loan annuity																	
12		13 applicatio	AMT_GOODS_PRICE	For consumer loans it is the price of the goods for which the loan is given																	
13		14 applicatio	NAME_TYPE	Who was accompanying client when he was applying for the loan																	
14		15 applicatio	NAME_INCOME_TYPE	Clients income type (businessman, working, maternity leave,...)																	
15		16 applicatio	NAME_EDUCATION	Level of highest education the client achieved																	
16		17 applicatio	NAME_FAMILY_STATUS	Family status of the client																	
17		18 applicatio	NAME_HOUSING	What is the housing situation of the client (renting, living with parents, ...)																	
18		19 applicatio	REGION_FEDERATION	Normalize normalized																	
19		20 applicatio	DAYS_BIRTH	Client's age time only relative to the application																	
20		21 applicatio	DAYS_EMPLOYMENT	How many time only relative to the application																	
21		22 applicatio	DAYS_REGISTRATION	How many time only relative to the application																	
22		23 applicatio	DAYS_ID_CREDIT	How many time only relative to the application																	
23		24 applicatio	OWN_CAR	Age of client's car																	
24		25 applicatio	FLAG_MOBILE_PHONE	Did client provide mobile phone (1=YES, 0=NO)																	
25		26 applicatio	FLAG_WORK_PHONE	Did client provide work phone (1=YES, 0=NO)																	
26		27 applicatio	FLAG_HOME_PHONE	Did client provide home phone (1=YES, 0=NO)																	
27		28 applicatio	FLAG_PHONE_REACHABLE	Was mobile phone reachable (1=YES, 0=NO)																	
28		29 applicatio	FLAG_PHONE_REACHABLE	Was mobile phone reachable (1=YES, 0=NO)																	

For clean data using some of the statistics method for analysis like iqr, quartiles and removing the Unnecessary data which are not needed

Identify the missing data and use appropriate method to deal with it.

Application_data.csv

Before cleaning the data there are 307521 rows and 122 column are present

AND deleted some of the unwanted column means more then 40% of blank data where present in the data

column name	missing	non misssing	total rows	no.of.%
OCCUPATION_TYPE	96391	211120	307511	31.3455
EXT_SOURCE_3	60965	246546	307511	19.8253
AMT_REQ_CREDIT_BUREAU_HOUR	41519	265992	307511	13.5016
AMT_REQ_CREDIT_BUREAU_DAY	41519	265992	307511	13.5016
AMT_REQ_CREDIT_BUREAU_WEEK	41519	265992	307511	13.5016
AMT_REQ_CREDIT_BUREAU_MON	41519	265992	307511	13.5016
AMT_REQ_CREDIT_BUREAU_QRT	41519	265992	307511	13.5016
AMT_REQ_CREDIT_BUREAU_YEAR	41519	265992	307511	13.5016
NAME_TYPE_SUITE	1292	306219	307511	0.42015
OBS_30_CNT_SOCIAL_CIRCLE	1021	306490	307511	0.33202
DEF_30_CNT_SOCIAL_CIRCLE	1021	306490	307511	0.33202
OBS_60_CNT_SOCIAL_CIRCLE	1021	306490	307511	0.33202
DEF_60_CNT_SOCIAL_CIRCLE	1021	306490	307511	0.33202
EXT_SOURCE_2	660	306851	307511	0.21463
AMT_GOODS_PRICE	278	307233	307511	0.0904
AMT_ANNUITY	12	307499	307511	0.0039
CNT_FAM_MEMBERS	2	307509	307511	0.00065
DAYS_LAST_PHONE_CHANGE	1	307510	307511	0.00033
SK_ID_CURR	0	307511	307511	0
TARGET	0	307511	307511	0
NAME_CONTRACT_TYPE	0	307511	307511	0
CODE_GENDER	0	307511	307511	0
FLAG_OWN_CAR	0	307511	307511	0
FLAG_OWN_REALTY	0	307511	307511	0
CNT_CHILDREN	0	307511	307511	0

only less then 40 % and needed column are used

and used the **count** and **count blank and counta** formula for this cleaning and I not replaced any value in this data

In the below 40 percent missing value column there are some unwanted data which are not needed ONLY 40 column are needed to analysis

previous_application.csv

Before cleaning data there are 37 column and 1048576 are data are present important notice that original data is 17670216 but in excel their no option for that much bigdata also it is like analyzing the sample data

In this data there are so many column having 40% of null values so I removed that column also deleted the unwanted data which is not useful for the analysis for this analysis I

Used the count and counta and count blank function

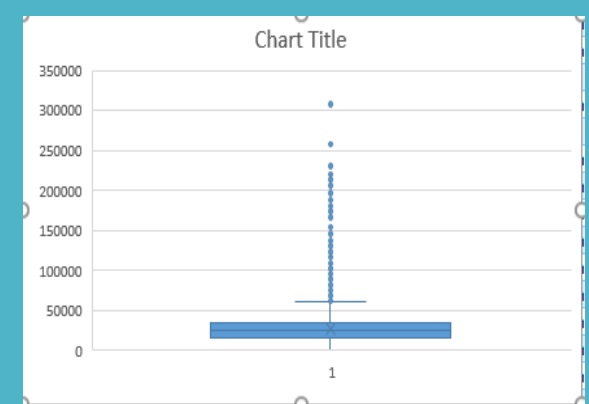
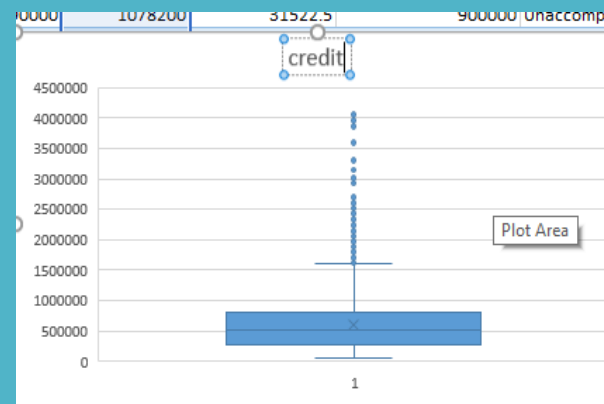
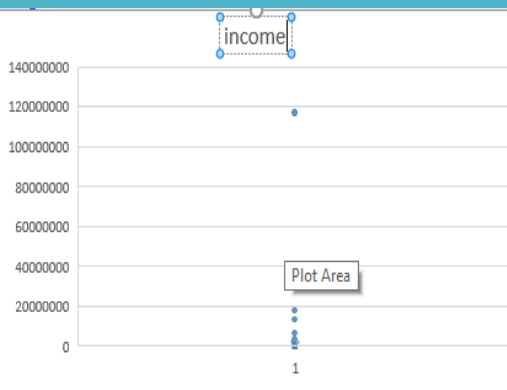
And also, I not replaced any value and only 14 column are needed

COLUMNNAME				PERCENT
RATE_INTEREST_PRIMARY	3734	1044841	1048575	99.6438977
RATE_INTEREST_PRIVILEGED	3734	1044841	1048575	99.6438977
AMT_DOWN_PAYMENT	489182	559393	1048575	53.3479246
RATE_DOWN_PAYMENT	489182	559393	1048575	53.3479246
NAME_TYPE_SUITE	533438	515137	1048575	49.1273395
DAYS_FIRST_DRAWING	627871	420704	1048575	40.1214982
DAYS_FIRST_DUE	627871	420704	1048575	40.1214982
DAYS_LAST_DUE_1ST_VERSION	627871	420704	1048575	40.1214982
DAYS_LAST_DUE	627871	420704	1048575	40.1214982
DAYS_TERMINATION	627871	420704	1048575	40.1214982
NFLAG_INSURED_ON_APPROVAL	627871	420704	1048575	40.1214982
AMT_GOODS_PRICE	807612	240963	1048575	22.9800443
AMT_ANNUITY	815567	233008	1048575	22.2213957
CNT_PAYMENT	815570	233005	1048575	22.2211096
PRODUCT_COMBINATION	1048351	224	1048575	0.02136233
SK_ID_CURR	1048575	0	1048575	0
NAME_CONTRACT_TYPE	1048575	0	1048575	0
AMT_APPLICATION	1048575	0	1048575	0
AMT_CREDIT	1048575	0	1048575	0
WEEKDAY_APPR_PROCESS_START	1048575	0	1048575	0
HOUR_APPR_PROCESS_START	1048575	0	1048575	0
FLAG_LAST_APPL_PER_CONTRAC	1048575	0	1048575	0

Identify if there are outliers in the dataset

Finding the outliers using the IQR method and finding the upper bound and lower bound and DOING EDA

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE
COUNT	307515	307515	307503	307237
AVG	168797.7236	599020.2079	27110.22133	538391.1997
Q3	202500	808650	34596	679500
Q1	112500	270000	16524	238500
MEDIAN	147150	513531	24903	450000
S.D	237121.8732	402491.0906	14511.4141	369446.4769
IQR	90000	538650	18072	441000
UPPER BOUND	337500	1616625	61704	1341000
LOWER BOUND	-22500	-537975	-10584	-423000



Previous data

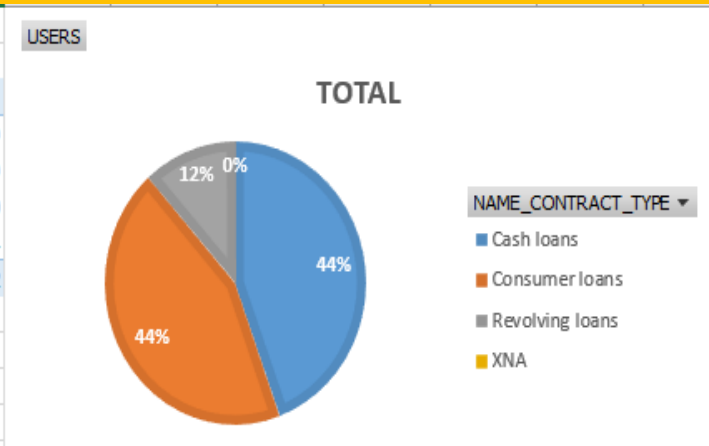
Using the q1 and q3 got the IQR and find the upper bound and lower bound as well and it is very important to remove the outliers of this given data and it also contains the null values and there are many data that are excluding the lower and upper bound

	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE
count	815554	1048562	1048562	807599
average	15891.29777	174270.9554	195001.046	226291.0418
Q1	6301.35	18900	24277.5	50580
Q2	11250	70821	80253	111523.5
Q3	20523.105	180000	215244	229500
MIN	0	0	0	0
MAX	418058.145	6905160	6905160	6905160
IQR	14221.755	161100	190966.5	178920
LOWER BOUND	-15031.2825	-222750	-262172.25	-217800
UPPER BOUND	41855.7375	421650	501693.75	497880

CONTRACT TYPE

THERE ARE 4 DIFFERENT CONTRACT TYPES AND TOTAL THIS TYPE OF LOAN ARE THEIR

1		
2		
3	CONTRACT TYRE	USERS
4	Cash loans	465859
5	Consumer loans	461039
6	Revolving loans	121440
7	XNA	224
8	Grand Total	1048562
9		
10		
11		
12		

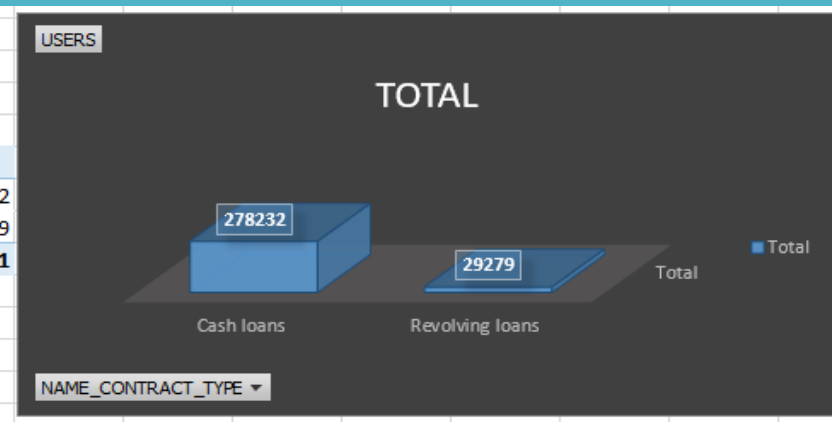


TYPES OF LOANS

THEIR 2 TYPE OF LOANS

CASH LOANS AND REVOLVING LOANS ARE SANCTIONED

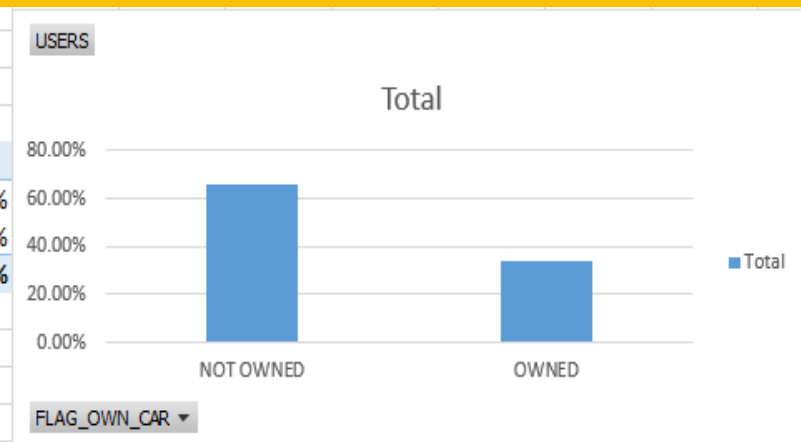
CONTACT TYPE	USERS
Cash loans	278232
Revolving loans	29279
Grand Total	307511



CAR OWNED

IN 100 PERCENT CUSTOMERS THERE ARE
66% ARE NOT OWNED CAR AND 34 % ARE
OWNED CAR

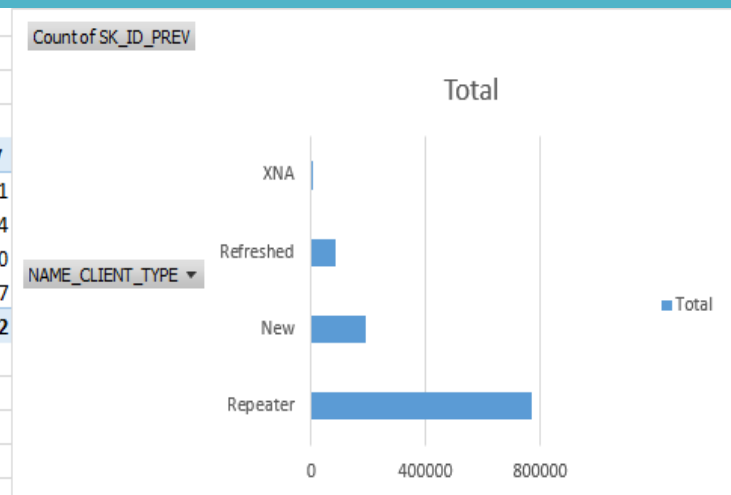
2		
3		
4		
5		
6	CAR OWNIN <input type="button" value="v"/>	USERS
7	NOT OWNED	65.97%
8	OWNED	34.03%
9	Grand Total	100.00%
10		
11		
12		
13		



CLIENT TYPE

THERE ARE 4 TYPES OF CLIENT TYPE IN THE BANK, REPETER ARE MORE COMPARED TO OTHER

Row Labels	Count of SK_ID_PREV
Repeater	771951
New	189984
Refreshed	85410
XNA	1217
Grand Total	1048562

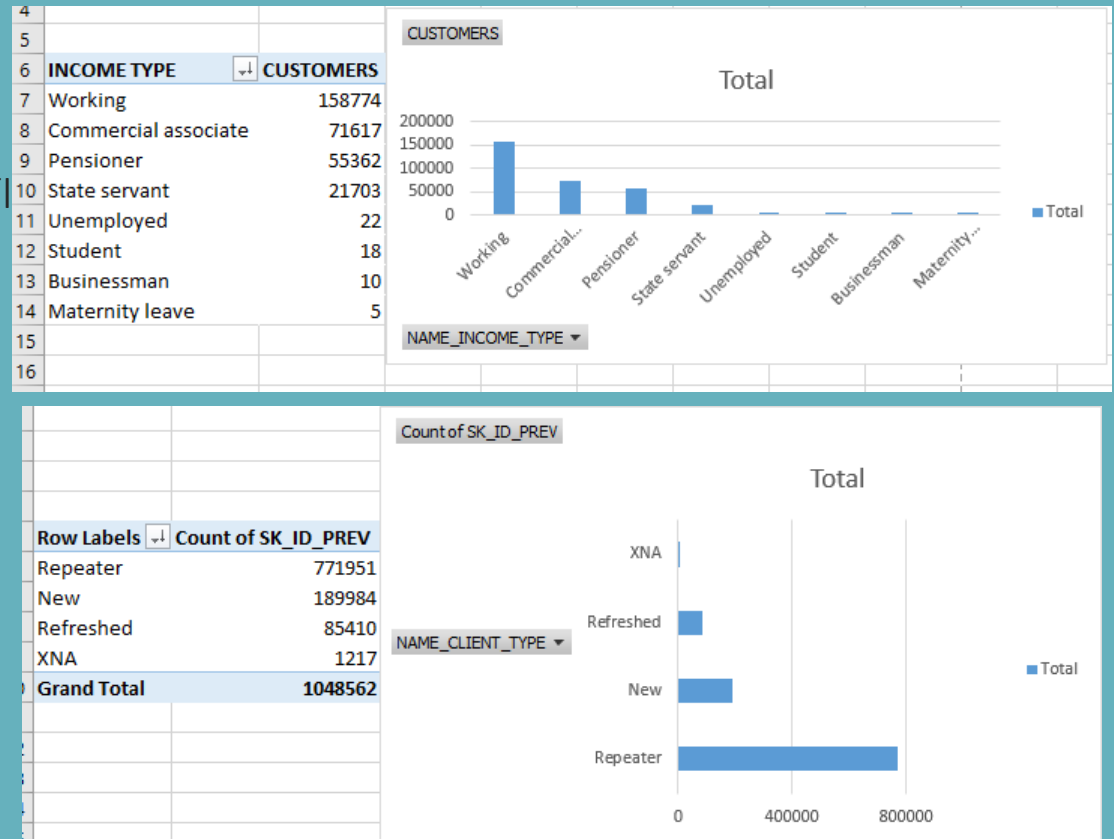


results of univariate, segmented univariate, bivariate analysis

UNIVARIATE NUMERICAL ANALYSIS

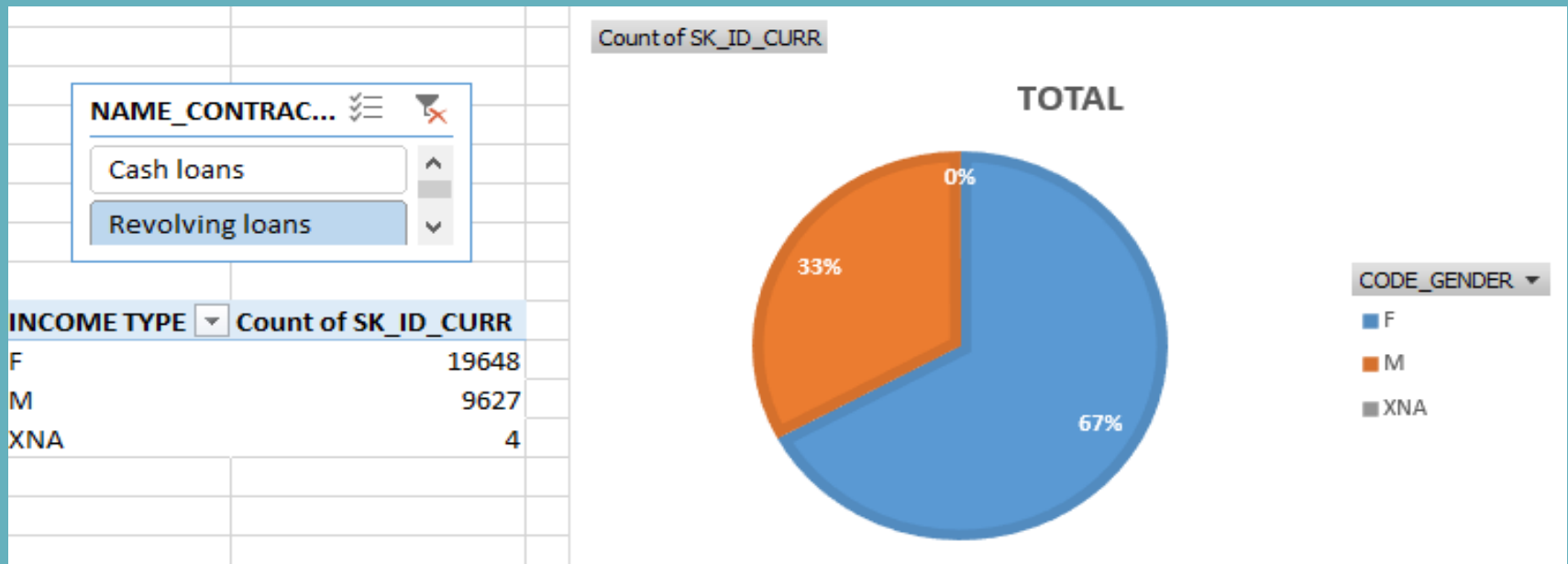
IN THIS ANALYSIS FIRST PICTURE SHOWS INCOME TYPES AND LOAN SANCTIONED AND 2ND PICTURE CLIENT TYPE WHO GOT THE LOAN

THIS ANALYSIS IS MADE USING EXCEL AND USING PIVOT TABLE AND THIS IS THE UNIVARIATE ANALYSIS

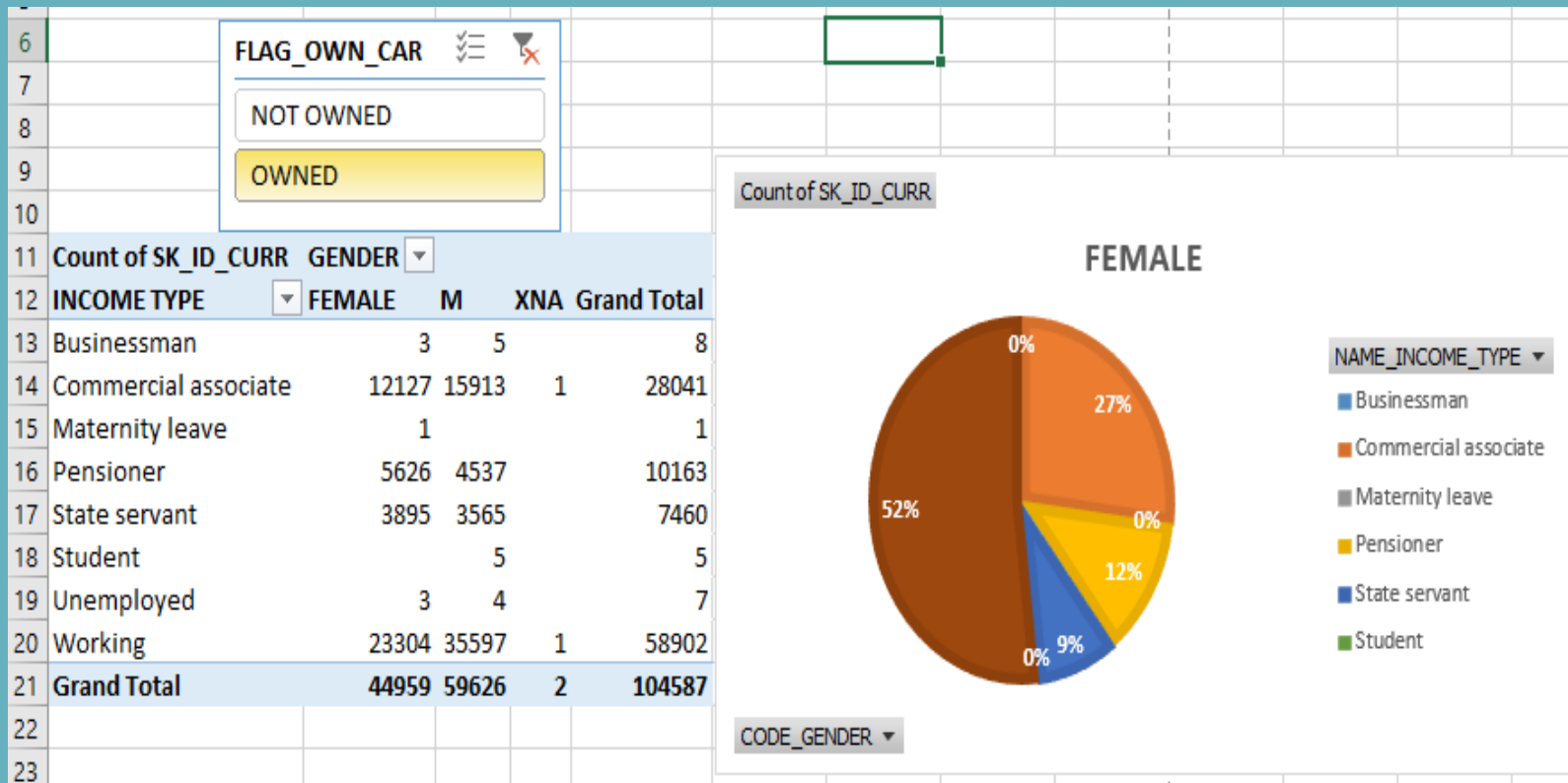


BIVARIATE ANALYSIS

IN THIS DATA ,IT GIVING THE INFORMATION OF LOAN GOT FOR MALE AND FEMALE AND ALSO THE CONTRACT THAT IS CASH AND REVOLVING LOAN THIS IS THE BIVARIATE ANALYSIS

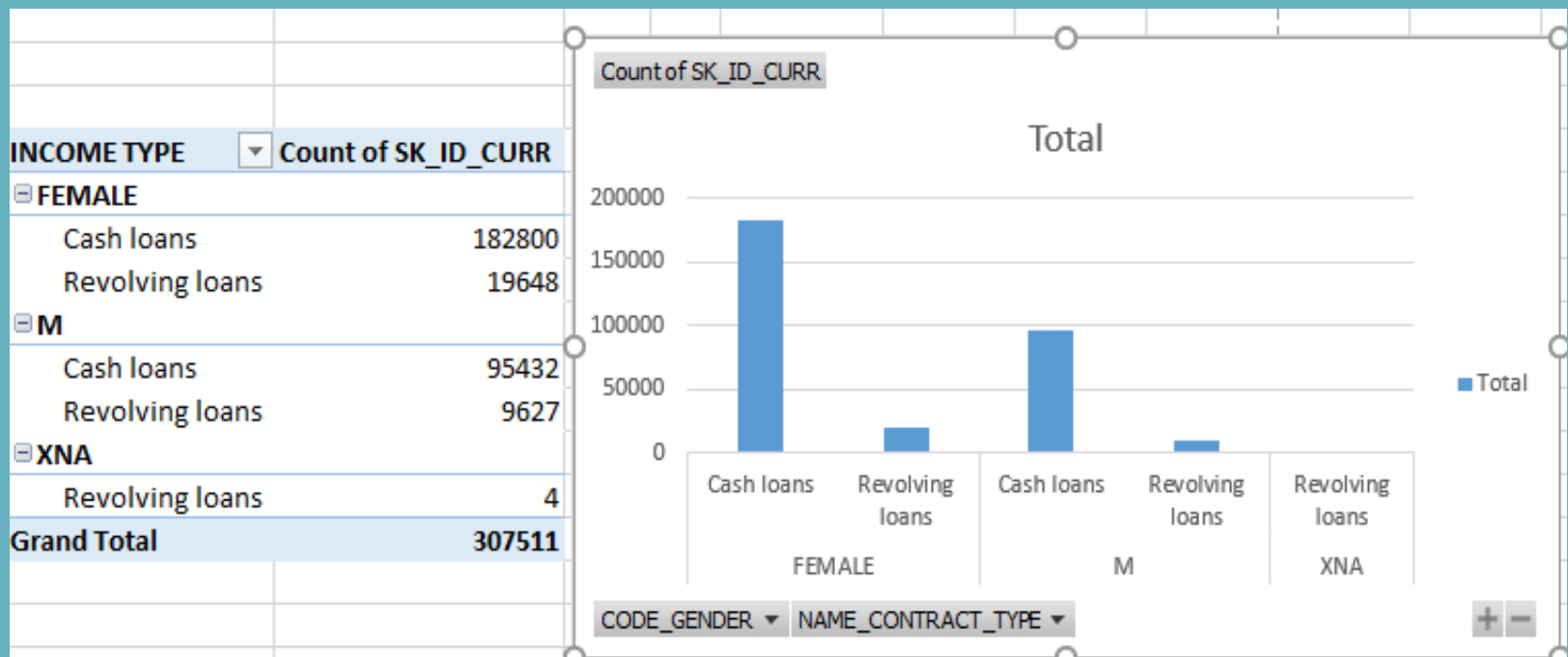


ANOTHER VISUALIZATION OF THE BIVARIATE ANALYSIS CUSTOMER INCOME TYPE AND WHO OWNED THE CAR AND NOT AND WHO GOT THE LOAN AND THIS IS MADE BY USING PIVOT TABLE



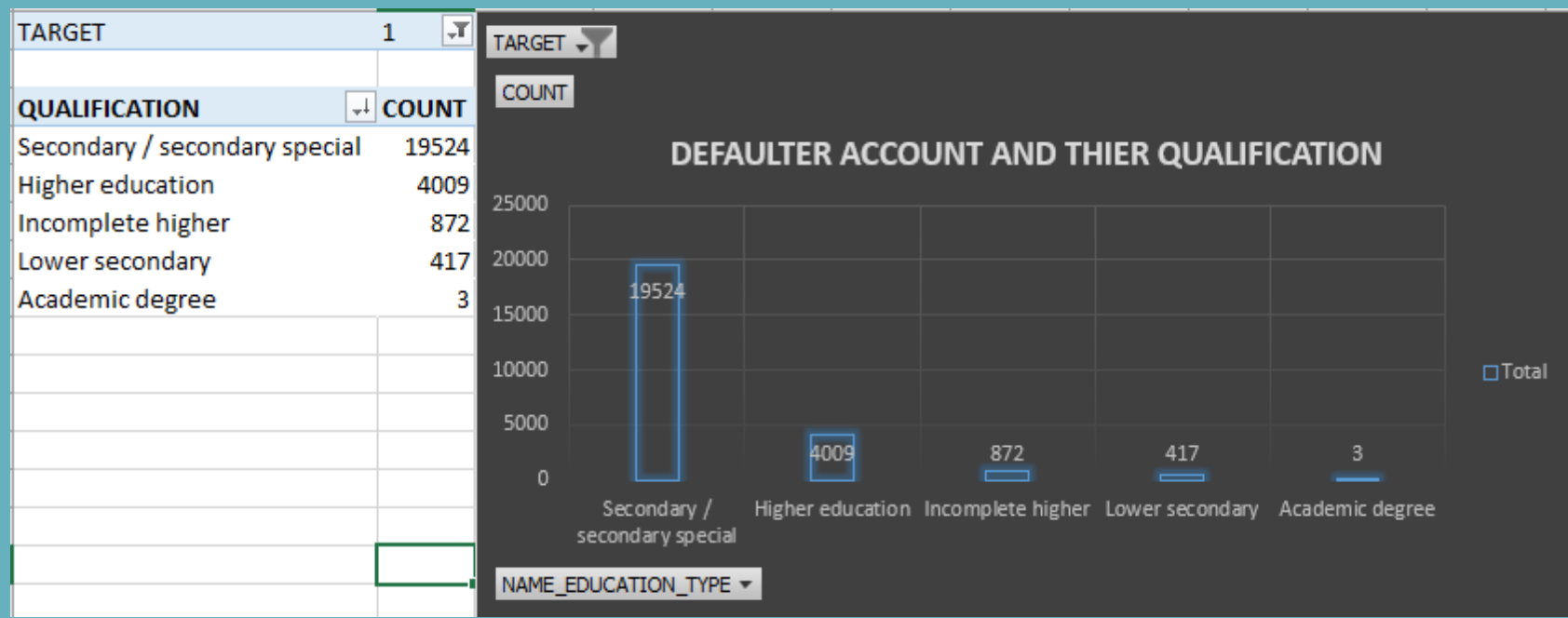
SEGMENTED UNIVARIATE ANALYSIS

IN THE SEGMENTED TYPE OF ANALYSIS GETTING THE INFORMATION OF MULTIPLE DATA IN SINGLE VISUALIZATION GENDER AS WELL TYPE OF LOAN WHO GOT THE LOAN





correlation for the Client with payment difficulties and all other case

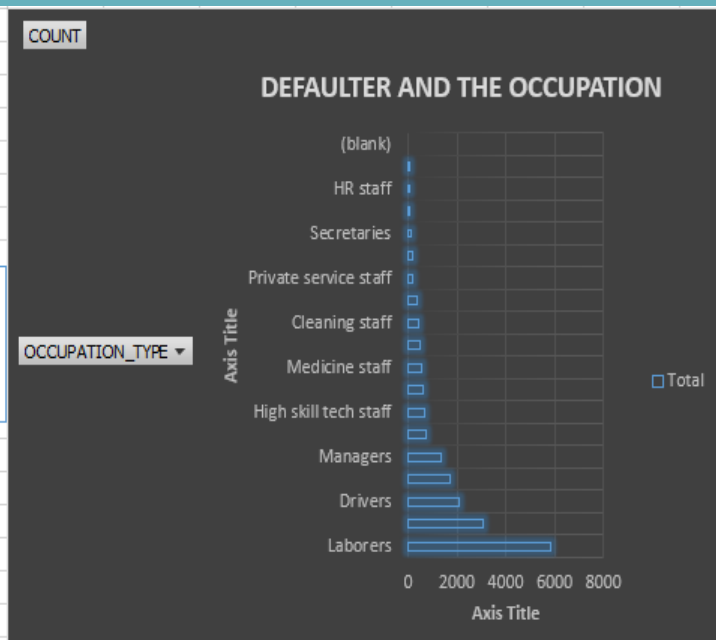
I MADE THIS USING THE PIVOT TABLE AND ALSO MADE FILTER FOR THE TARGET COULMN AND GIVING THE RELATION OF DEFAULTER BASED ON THE THEIR QUALIFICATION MORE ARE SECONDARY OR SECONDARY SPECIAL



OCCUPATION OF THE DEFALTER THERE ARE MANY OCCUPATION IN THAT LABORERES ARE MORE

WE CAN SAY THAT DEFALTER ARE MORE IN THE LABORER SETION AND I NOT INCLUDED BLANK CELL FOR THIS ANALYSIS AND ALSO ADDED SLICER FOR TARGET U CAN CHECK ANY

3	OCCUPATION	COUNT		
4	Laborers	5838		
5	Sales staff	3092		
6	Drivers	2107		
7	Core staff	1738		
8	Managers	1328		
9	Security staff	722		
10	High skill tech staff	701		
11	Cooking staff	621	TARGET	 
12	Medicine staff	572	<input type="text" value="0"/>	
13	Accountants	474		
14	Cleaning staff	447	<input type="text" value="1"/>	
15	Low-skill Laborers	359		
16	Private service staff	175		
17	Waiters/barmen staff	152		
18	Secretaries	92		
19	Realty agents	59		
20	HR staff	36		
21	IT staff	34		



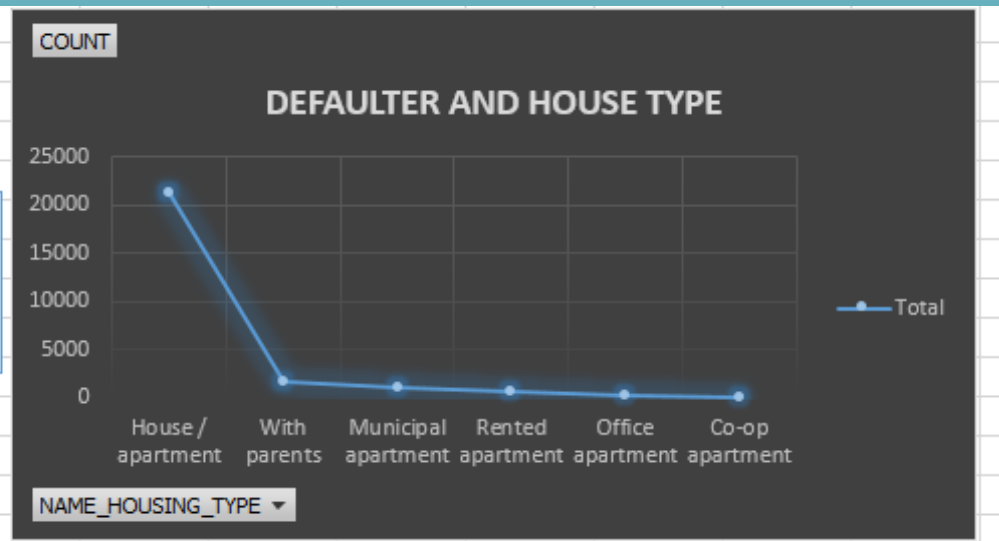
HOUSE TYPE OF DEFAULTER MORE ARE HOUSE OR APARTMENT AND THERE IS DRASTICALLY DOWN FOR OTHER HOUSE TYPE AND THI THE RELATIONS AND THIS IS MADE USING POVOT CHART AND PIVOT TABLE

HOUSE TYPE	COUNT
House / apartment	21272
With parents	1736
Municipal apartment	955
Rented apartment	601
Office apartment	172
Co-op apartment	89

TARGET

0

1



Ads Airing Report Analysis

XYZ Ads Airing Report Analysis

Description

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 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.

Business Understanding

Advertising is a way of marketing your business in order to increase sales or make your audience aware of your products or services. Until a customer deals with you directly and actually buys your products or services, your advertising may help to form their first impressions of your business. Target audience for businesses could be local, regional, national or international or a mixture. So they use different ways for advertisement

Approach

The given data is cleaned and there is no need of doing the cleaning of the data and I used the excel ,power point to do this project and also used pivot table and some of the formulas of excel

Tech-Stack Used

excel PowerPoint and notepad

INSIGHT

- I got the information about advertising techniques and how much the amount spends for advertising for their brand promotion
- And also, I got information of which day of the week and month the ad is more broadcasting
- I got the information of day part
- What are the aspect they considering while broad casting the ad
- Which company is spending more money on add and also less spending company

Working file (data)

5	Id	Brand	Product	Network	Network Typ	Show	Date	Time	Pod Positior	Dayparts	Duration	EQ Units	Spend (\$)	Broadcast Y	Broadcast M	Broadca	Day Of Week	Hour Of Day	TimeZon
6	1000001	Honda Cars	Honda Jazz	N-10079	cable	S-16736	01/10/2021	10:19:08 PM	1	PRIME TIME	30	1	178	2021	JAN	2	7	22	Northea
7	1000002	Maruti Suzu	Maruti Suzu	N-10096	cable	S-17450	01/16/2021	7:28:13 PM	3	WEEKEND	30	1	514	2021	JAN	3	6	19	Northea
8	1000003	Maruti Suzu	Maruti Suzu	N-10044	cable	S-12299	01/05/2021	1:09:26 PM	2	DAYTIME	30	1	2313	2021	JAN	2	2	13	Northea
9	1000004	Maruti Suzu	Maruti Suzu	N-10060	cable	S-14354	01/06/2021	8:55:49 AM	1	EARLY MORN	30	1	308	2021	JAN	2	3	8	Northea
10	1000005	Hyundai Mo	Hyundai I20	N-10123	cable	S-15146	01/02/2021	11:07:43 PM	2	LATE FRINGE	30	1	1885	2021	JAN	1	6	23	Northea
11	1000006	Hyundai Mo	Hyundai I20	N-10116	cable	S-14314	01/15/2021	11:43:47 AM	2	DAYTIME	30	1	2683	2021	JAN	3	5	11	Northea
12	1000007	Maruti Suzu	Maruti Suzu	N-10135	cable	S-15889	01/06/2021	7:12:36 PM	3	PRIME ACCE	30	1	1545	2021	JAN	2	3	19	Northea
13	1000008	Hyundai Mo	Hyundai I20	N-10003	cable	S-11802	01/14/2021	5:59:07 AM	6	EARLY MORN	30	1	575	2021	JAN	3	4	5	Northea
14	1000009	Honda Cars	Honda Jazz	N-10121	cable	S-12232	01-08-2021	8:42:40 AM	9	EARLY MORN	15	0.5	1383	2021	JAN	2	5	8	Northea
15	1000010	Toyota	Toyota Inno	N-10005	cable	S-20089	01/01/2021	7:16:41 AM	6	EARLY MORN	15	0.5	49	2021	JAN	1	5	7	Northea
16	1000011	Maruti Suzu	Maruti Suzu	N-10103	cable	S-19117	01/07/2021	3:24:03 AM	1	OVERNIGHT	30	1	55	2021	JAN	2	4	3	Northea
17	1000012	Tata Motors	Tata Tiago	N-10026	cable	S-19337	01/13/2021	1:37:24 PM	10	DAYTIME	30	1	1126	2021	JAN	3	3	13	Northea
18	1000013	Honda Cars	Honda Civic	N-10110	cable	S-18885	01/13/2021	10:55:51 AM	3	DAYTIME	30	1	20	2021	JAN	3	3	10	Northea
19	1000014	Hyundai Mo	Hyundai I20	N-10054	cable	S-10822	01/10/2021	8:15:29 AM	8	EARLY MORN	15	0.5	327	2021	JAN	2	7	8	Northea
20	1000015	Maruti Suzu	Maruti Suzu	N-10030	broadcast	S-21631	01/04/2021	12:15:30 AM	11	LATE FRINGE	30	1	3954	2021	JAN	2	1	0	Northea
21	1000016	Hyundai Mo	Hyundai I20	N-10026	cable	S-20756	01/11/2021	6:53:50 PM	1	EVENING NE	30	1	2435	2021	JAN	3	1	18	Northea
22	1000017	Tata Motors	Tata Safari	N-10066	cable	S-16893	01/07/2021	9:53:49 AM	7	DAYTIME	15	0.5	702	2021	JAN	2	4	9	Northea
23	1000018	Hyundai Mo	Hyundai I20	N-10088	cable	S-17278	01/04/2021	9:43:12 AM	8	DAYTIME	30	1	111	2021	JAN	2	1	9	Northea
24	1000019	Maruti Suzu	Maruti Suzu	N-10089	broadcast	S-10684	01/12/2021	7:09:29 PM	2	PRIME ACCE	30	1	40422	2021	JAN	3	2	19	Northea
25	1000020	Maruti Suzu	Maruti Suzu	N-10080	cable	S-17787	01/02/2021	11:27:43 PM	1	LATE FRINGE	30	1	745	2021	JAN	1	6	23	Northea
26	1000021	Toyota	Toyota Inno	N-10014	cable	S-19909	01/08/2021	6:40:42 AM	6	EARLY MORN	30	1	32	2021	JAN	2	5	6	Northea
27	1000022	Hvundai Mo	Hvundai I20	N-10044	cable	S-13134	01/16/2021	11:37:33 PM	2	LATE FRINGE	15	0.5	2027	2021	JAN	3	6	23	Northea

There are 19 column and
731789 rows are them
And it is cleaned no require
of cleaning data

Id
Brand
Product
Network
Network Type
Show
Date
Time
Pod Position
Dayparts
Duration
EQ Units
Spend (\$)
Broadcast Year
Broadcast Month
Broadcast Week
Day Of Week
Hour Of Day
TimeZone

POD POSITION

MEANING

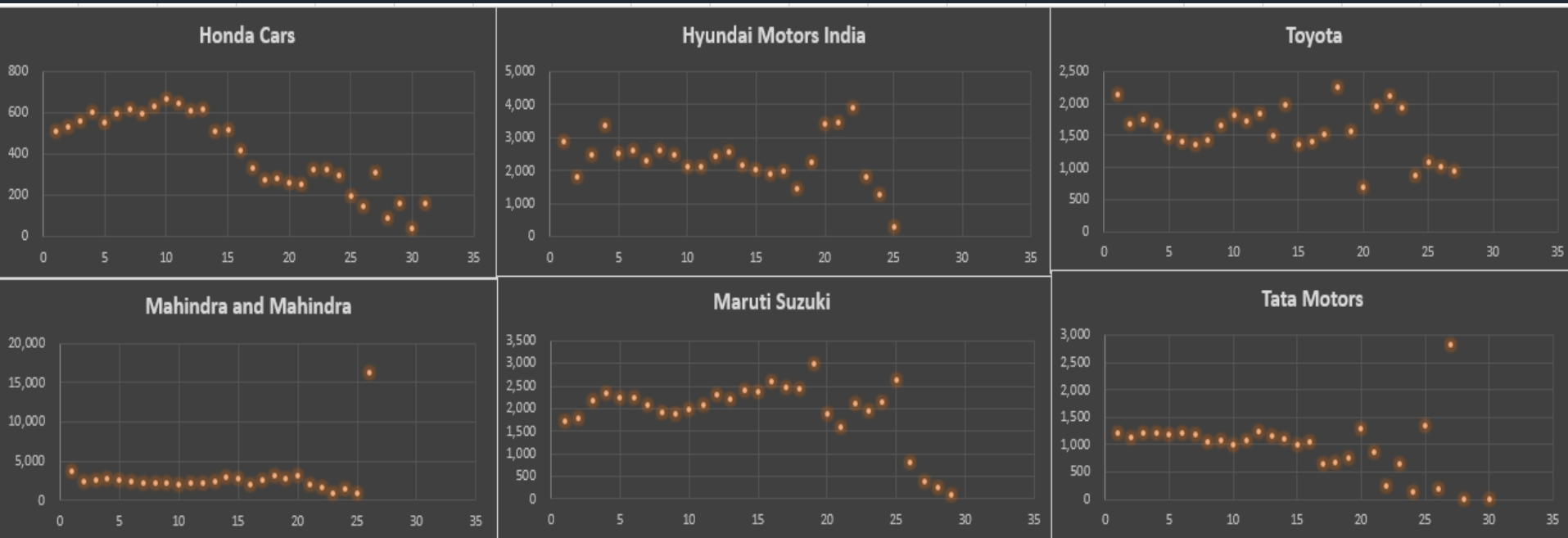
The position of an individual advertisement within a certain commercial pod OR basically the Sequence in which the ads are played on TV, Like which ad plays in the beginning followed by which one and so on

Explain in Details with examples from the dataset provided

BRAND NAME	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
Honda Cars	512	535	564	603	556	599	618	595	635	669	643	614	617	514	516	418	331	279	280	258	253	328	329	297	194	146	312	91	161	37	165
Hyundai Motors India	2,888	1,811	2,467	3,359	2,518	2,621	2,324	2,611	2,477	2,110	2,139	2,446	2,551	2,170	2,015	1,891	1,971	1,469	2,258	3,429	3,467	3,917	1,793	1,291	276				1,679		
Mahindra and Mahindra	3,723	2,394	2,543	2,734	2,632	2,490	2,228	2,229	2,140	2,094	2,192	2,277	2,329	2,929	2,720	1,965	2,528	3,204	2,694	3,223	1,972	1,687	935	1,497	897	16,175		294	101		274
Maruti Suzuki	1,730	1,776	2,163	2,334	2,252	2,226	2,082	1,924	1,892	1,988	2,083	2,311	2,213	2,400	2,372	2,603	2,453	2,449	2,973	1,890	1,575	2,105	1,932	2,136	2,639	819	394	274	90		2,569
Tata Motors	1,209	1,124	1,202	1,210	1,190	1,202	1,183	1,049	1,091	1,005	1,090	1,234	1,149	1,100	991	1,057	653	681	754	1,283	854	259	650	141	1,344	182	2,830	7			11
Toyota	2,155	1,691	1,758	1,656	1,478	1,415	1,362	1,424	1,665	1,822	1,730	1,835	1,499	1,976	1,350	1,412	1,528	2,261	1,575	700	1,962	2,121	1,932	868	1,093	1,021	947				

Honda cars the maximum price reached is for pod position 10 but it then starts decreasing as pod position increases. there is increasing also and drastic change also

And used pivot table and also pivot chart for the visualization



Share of various brands in TV airings and how has it changed from Q1 to Q4 in 2021

I made this table using the pivot table

Row Labels	Sum of Spend (\$)				Count of Id				Total Sum of Spend (\$)		Total Count of Id	
	Column Labels											
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4				
Honda Cars	3.80%	2.82%	3.90%	3.18%	12.44%	9.77%	12.99%	11.29%	3.47%		11.65%	
Hyundai Motors India	13.62%	12.51%	12.55%	12.97%	10.48%	9.84%	9.17%	9.23%	12.98%		9.74%	
Mahindra and Mahindra	27.63%	30.84%	30.18%	25.57%	19.71%	24.01%	22.05%	13.57%	28.53%		20.21%	
Maruti Suzuki	40.39%	39.71%	39.53%	40.80%	38.78%	37.31%	36.55%	41.10%	40.12%		38.30%	
Tata Motors	5.92%	5.01%	4.72%	12.60%	10.12%	7.62%	8.03%	20.93%	6.81%		11.07%	
Toyota	8.64%	9.12%	9.13%	4.89%	8.46%	11.45%	11.21%	3.87%	8.09%		9.02%	

Honda cars as lower percent of spend on ads

Maruti Suzuki spend more compared to another brand

Tata motors are increased as spend in q4 more then 100% percent compared to q1

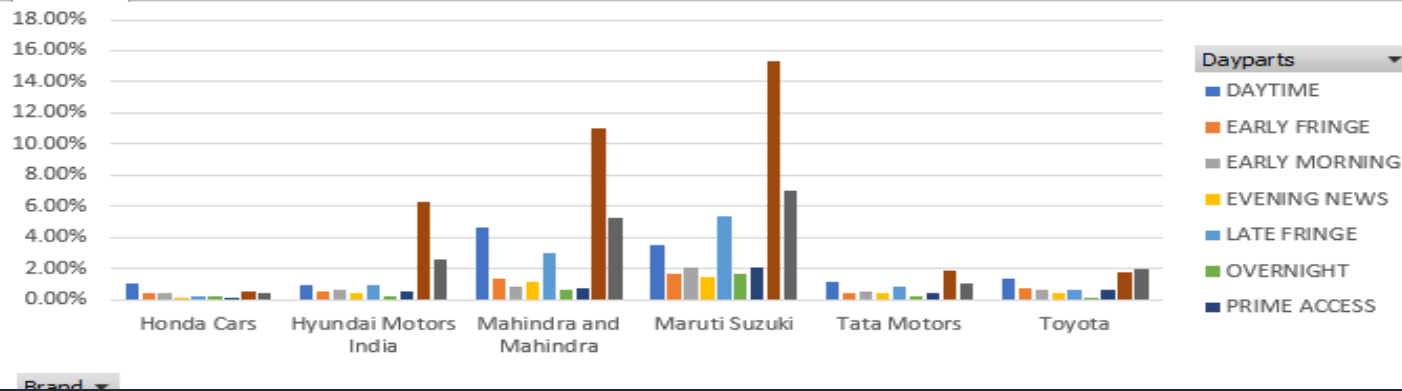
Toyotas drastically decreased the ad spend from 9 to 5

M and m increased in q2 and q3 but decreed in q4

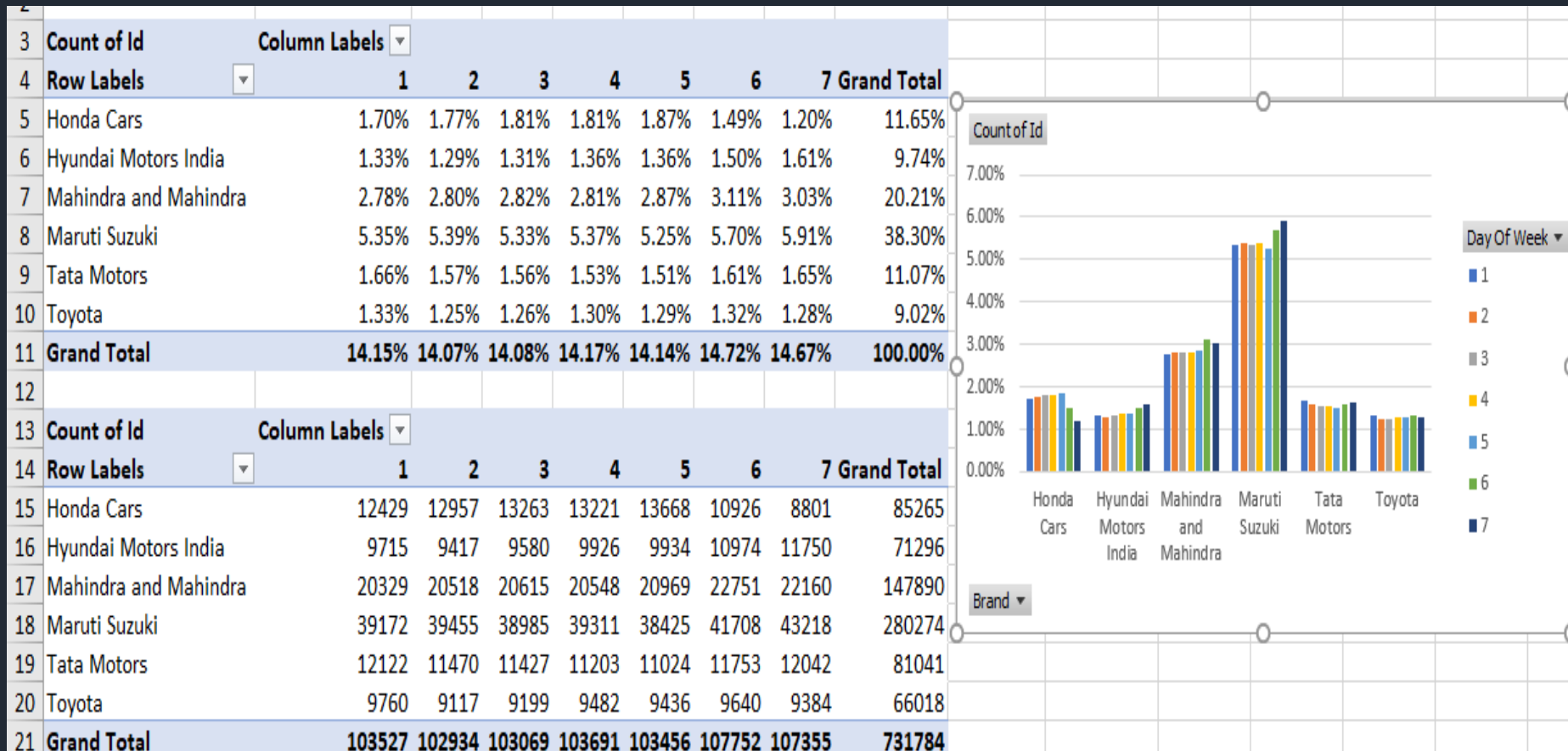
Competitive analysis for the brands on the spends

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	1.08%	0.41%	0.37%	0.15%	0.25%	0.20%	0.10%	0.50%	0.39%	3.47%
Hyundai Motors India	0.89%	0.51%	0.63%	0.39%	0.98%	0.23%	0.55%	6.23%	2.58%	12.98%
Mahindra and Mahindra	4.61%	1.38%	0.87%	1.15%	3.00%	0.62%	0.74%	10.97%	5.20%	28.53%
Maruti Suzuki	3.50%	1.63%	2.08%	1.49%	5.32%	1.70%	2.08%	15.34%	6.98%	40.12%
Tata Motors	1.19%	0.44%	0.51%	0.42%	0.80%	0.18%	0.42%	1.84%	1.01%	6.81%
Toyota	1.33%	0.70%	0.60%	0.39%	0.64%	0.12%	0.64%	1.73%	1.93%	8.09%
Grand Total	12.59%	5.08%	5.05%	3.99%	10.98%	3.05%	4.54%	36.62%	18.11%	100.00%

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	15106799	5763471	5190376	2105762	3421197	2820096	1352961	7002902	5494776	48258340
Hyundai Motors India	12360920	7156835	8708318	5364194	13648569	3181379	7711727	86737738	35939076	180808756
Mahindra and Mahindra	64154402	19204408	12119383	16018235	41781609	8597788	10299276	152713257	72417297	397305655
Maruti Suzuki	48678486	22745305	28920899	20776891	74069950	23614157	29021227	213609797	97209760	558646472
Tata Motors	16513542	6058611	7110565	5829272	11161135	2552537	5799904	25652452	14112209	94790227
Toyota	18560894	9744570	8294798	5409054	8863807	1716154	8979945	24146575	26937315	112653112
Grand Total	175375043	70673200	70344339	55503408	152946267	42482111	63165040	509862721	252110433	1392462562



Brand broad casting their ads in difference days of week



Summary of competitive analysis

- Maruti Suzuki spend more in the primetime for the add and also it is spending more money compared to other brand
- Honda cars spending less amount in the prime access
- In the over night brands are spending less money because of less watching
- We can observe that all day of week brand are adding their advertise and there is no more changes
- Maruti Suzuki broadcasting more ads in 7th day of week
- Prime time and weekend days are more then other days the company is spending the money
- Honda cars spending less amount
- Toyota doesn't have a specific preference when it comes to selecting a specific time to show its ads

Media plan to the CMO of Mahindra and Mahindra

- Mahindra and Mahindra spend less amount comparing to the maruthi suzuki
- Mahindra and Mahindra should focus on the weekend and primetime because another brand take over that position and also this these 2 daypart very important because of many are watching advertise in this time

Row Labels	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Grand Total
Mahindra and Mahindra	18.18%	19.81%	21.71%	25.38%	23.59%	23.29%	21.41%	22.38%	22.39%	16.09%	11.70%	12.34%	20.21%

We can observe that their increasing in the broadcast and drastically decreasing it not good for any company and also q1 of 2022 getting tougher for the ad

- And also they want to target daytime and primetime

Call Volume Trend Analysis

ABC Call Volume Trend Analysis

Description:

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). 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.

ABC Call Volume Trend Analysis

Business Understanding:

Advertising is a way of marketing your business to increase sales or make your audience aware of your products or services. Until a customer deals with you directly and buys your products or services, your advertising may help to form their first impressions of your business. Target audience for businesses could be local, regional, national or international or a mixture. So, they use different ways for advertisement. Some of the types of advertisement are Internet/online directories, Trade and technical press, Radio, Cinema, Outdoor advertising, National papers, magazines and TV. Advertising business is very competitive as a lot of players bid a lot of money in a single segment of business to target the same audience. Here comes the analytical skills of the company to target those audiences from those types of media platforms where they convert them to their customers at a low cost.

Approach

The given data is cleaned there is not required clean the data for analysis
117989 rows are there I done some of the analysis with the help of pivot table
And some formulas it is very important understand the learn the concept excel in the analytics

Tech stack used

power point ,excel

Insight

I got more information about usage of excel the real power as well of excel in the Analytics and got the information about call trends means customer support Concept of Timeslot, Call Status, IVR Duration , etc. understand clearly

REQUIREMENTS

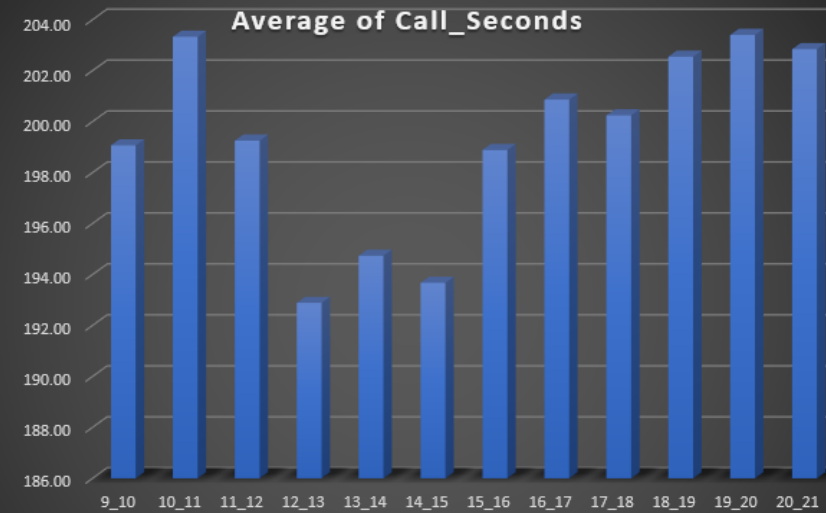
- Calculate the average call time duration for all incoming calls received by agents (in each Time Bucket).
- Show the total volume/ number of calls coming in via charts/ graphs [Number of calls v/s Time]. You can select time in a bucket form (i.e., 1-2, 2-3,)
- As you can see current abandon rate is 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.)
- Let's say customers also call this ABC insurance company in night but didn't get answer as there are no agents to answer, this creates a bad customer experience for this Insurance company. Suppose every 100 calls that customer made during 9 Am to 9 Pm, customer also made 30 calls in night between interval [9 Pm to 9 Am] and distribution of those 30 calls are as follows:

Average call time duration

TIME_BUCKET	Average of Call_Seconds	TIME_BUCKET	Average of Call_Seconds
10_11	203.33	9_10	199.07
11_12	199.26	10_11	203.33
12_13	192.89	11_12	199.26
13_14	194.74	12_13	192.89
14_15	193.68	13_14	194.74
15_16	198.89	14_15	193.68
16_17	200.87	15_16	198.89
17_18	200.25	16_17	200.87
18_19	202.55	17_18	200.25
19_20	203.41	18_19	202.55
20_21	202.85	19_20	203.41
9_10	199.07	20_21	202.85

Call_Status

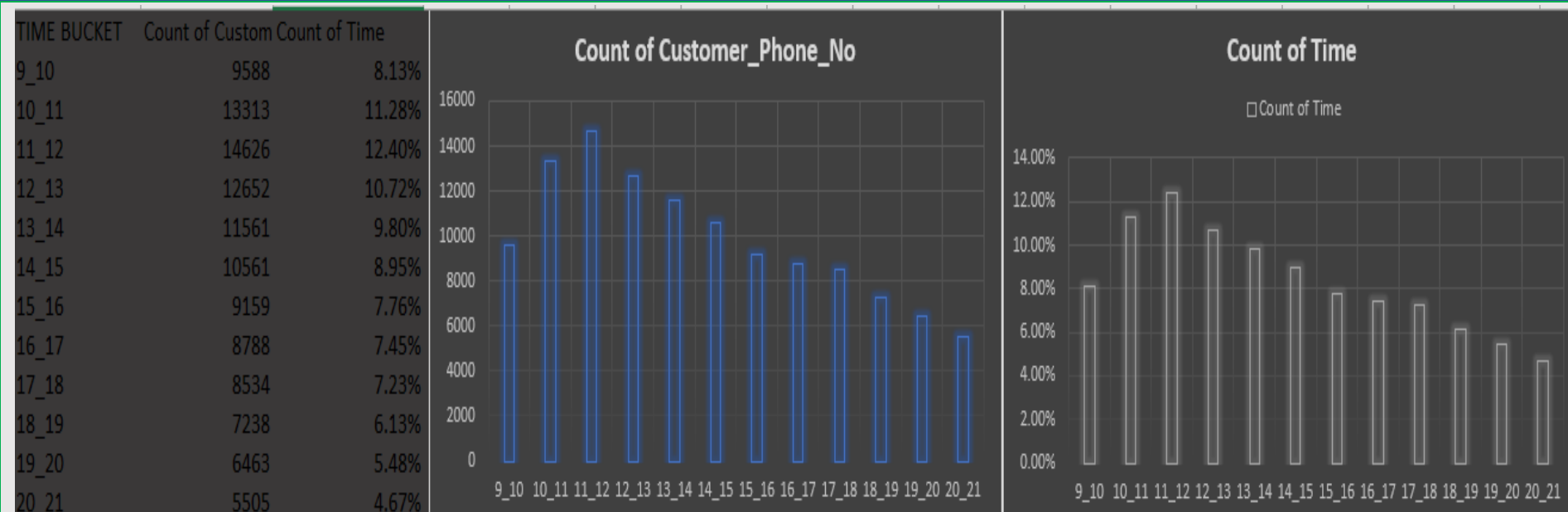
- abandon
- answered**
- transfer



Maximum number of call in 10-11 time bucket

Average call time for all answered call is 198.6

total volume/ number of calls coming in via charts/ graphs



Maximum count of call in 11_12 bucket

And least in 20_21

We can observe that there is drastically down after 11_12-time bucket

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.)

DATE	abandon	answered	transfer	Grand Total
01-Jan	684	3883	77	4644
02-Jan	356	2935	60	3351
03-Jan	599	4079	111	4789
04-Jan	595	4404	114	5113
05-Jan	536	4140	114	4790
06-Jan	991	3875	85	4951
07-Jan	1319	3587	42	4948
08-Jan	1103	3519	50	4672
09-Jan	962	2628	62	3652
10-Jan	1212	3699	72	4983
11-Jan	856	3695	86	4637
12-Jan	1299	3297	47	4643
13-Jan	738	3326	59	4123
14-Jan	291	2832	32	3155
15-Jan	304	2730	24	3058
16-Jan	1191	3910	41	5142
17-Jan	16636	5706	5	22347
18-Jan	1738	4024	12	5774
19-Jan	974	3717	12	4703
20-Jan	833	3485	4	4322
21-Jan	566	3104	5	3675
22-Jan	239	3045	7	3291
23-Jan	381	2832	12	3225
AVERAGE	1495.78261	3584.8696	49.26087	5129.913043
%	29%	70%	1%	100%

Assumption: An agent work for 6 days a week; On an average total unplanned leaves per agent is 4 days a month; An agent total working hrs is 9 Hrs out of which 1.5 Hrs goes into lunch and snacks in the office. On average an agent occupied for 60% of his total actual working Hrs (i.e 60% of 7.5 Hrs) on call with customers/ users. Total days in a month is 30 days.

[illegible]

TIME TAKEN ON AN AVG TO ANSWER THE CALL	198.6
TIME REQUIRE TO ANSWER 90%OF CALL	254.7002
TOTAL WORKING PERSON REQUIRED PER DAY	57

Average duration spent in a call is 198.6 sec

Time required to answer 70% of the calls

$$198.6 * 5130 * 0.7 / 3600 = 198.10 \text{ hours.}$$

A person works 4.5 hours a day,

Total number of people need to work in a day to get a 70%

$$198.10/4.5$$

$$= 44 \text{ People.}$$

Time required to answer 90% of the calls

$$= 198.6 * 5130 * 0.9 / 3600$$

$$= 254.700 \text{ hours.}$$

Total number of people need to work in a day to get 90% answering rate

$$= 254.700/4.5$$

$$= 57 \text{ People}$$

Let's say customers also call this ABC insurance company in night but didn't get answer as there are no agents to answer, this creates a bad customer experience for this Insurance company. Suppose every 100 calls that customer made during 9 Am to 9 Pm, customer also made 30 calls in night between interval [9 Pm to 9 Am] and distribution of those 30 calls are as follows:

CALL VOLUME DAILY 9AM TO 9PM						5129.913
IF WE PROVIDE SUPPORT IN NIGHT 9PM TO 9AM						1539
ADDITIONAL HOUR REQUIRED						76.41135
ADDITIONAL HEAD COUNT						17
TOTAL HEAD COUNT						74
		TIME DIS	76.4114	REQUIREMENT	CALL DISTRIBUTION	
21_22	3	10%	7.64114	13	8	
22_23	3	10%	7.64114	13	8	
23_24	2	7%	5.09409	8	6	
00_01	2	7%	5.09409	8	6	
01_02	1	3%	2.54705	4	3	
2_3	1	3%	2.54705	4	3	
3_4	1	3%	2.54705	4	3	
4_5	1	3%	2.54705	4	3	
5_6	3	10%	7.64114	13	8	
6_7	4	13%	10.1882	17	11	
7_8	4	13%	10.1882	17	11	
8_9	5	17%	12.7352	21	14	
		30				

A person works effectively 4.5 hours a day

Average duration spent in a call 198.6 sec

Call volume daily 5129.913

Provide support in night 1539

Additional hour requires 76

Additional headcount 17

Total head count $17+57=74$

Need 17 more man-power to support at night each day and also keeping the abandon rate to 10%.

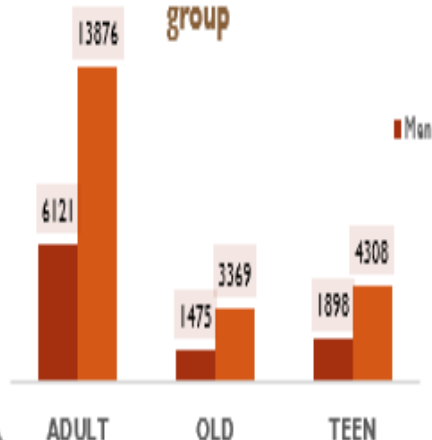
MY LEARNING AND EXPERIENCE

In starting while doing the project I don't know anything about power point and how project is doing and difficulties but now from these Project, I got some confidence in the data analytics, and I got the information about how data is cleaning ,analyzing ,visualizing
I got more knowledge in the SQL ,ADVANCE EXCEL,VISUALIZATION
Now I'm ready to the end-to-end project with the use of SQL and excel
Understand some of business terminology
Got the information about how data analyst doing their work

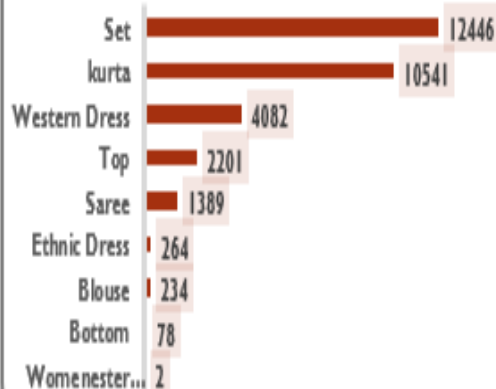
Sales Report

Vrinda Store Sales Report

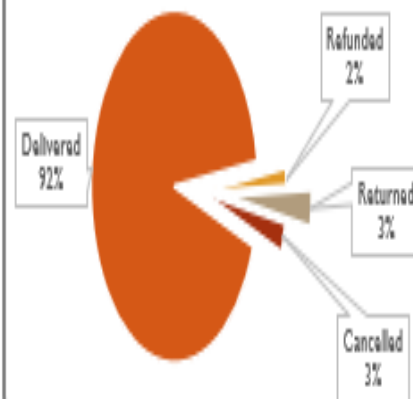
Male VS Female in each age group



Total Quantity sold in each category



Delivery Status



TOTAL ORDER

28471

Total Customers

28437

Category

9

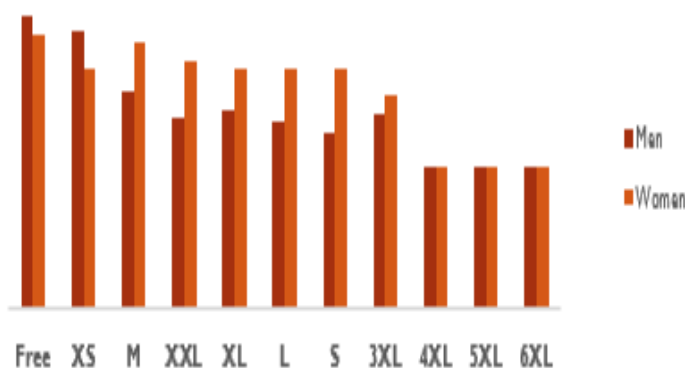
Total Channel

7

Sales Trend Year 2022



Average Dress size of Male VS Female



Products Order Report

Order Report

Total Order	Total Revenue	Average Revenue	Average Discount
-------------	---------------	-----------------	------------------

794

239042

301

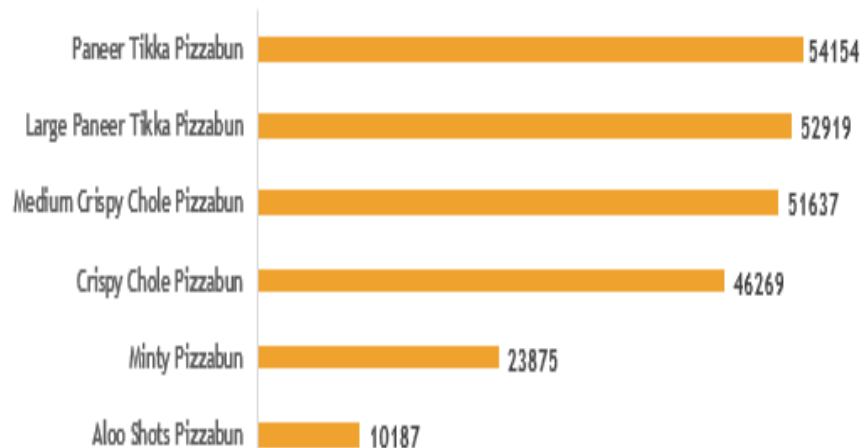
46%

Order Type

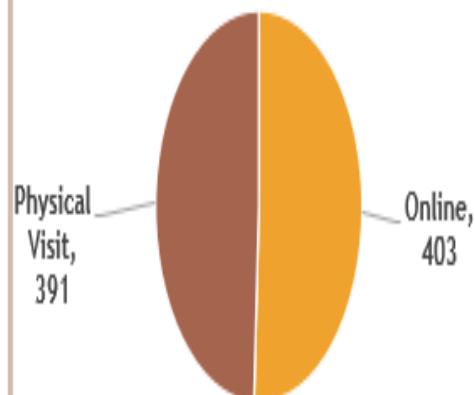
Online

Physical Visit

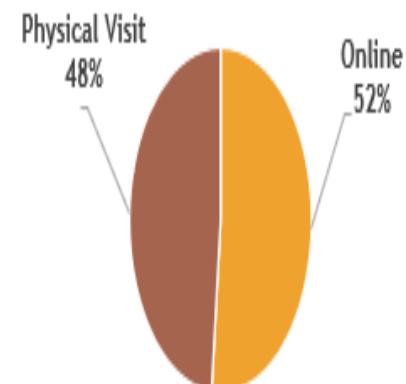
Trend of Revenue Generated for Each Product



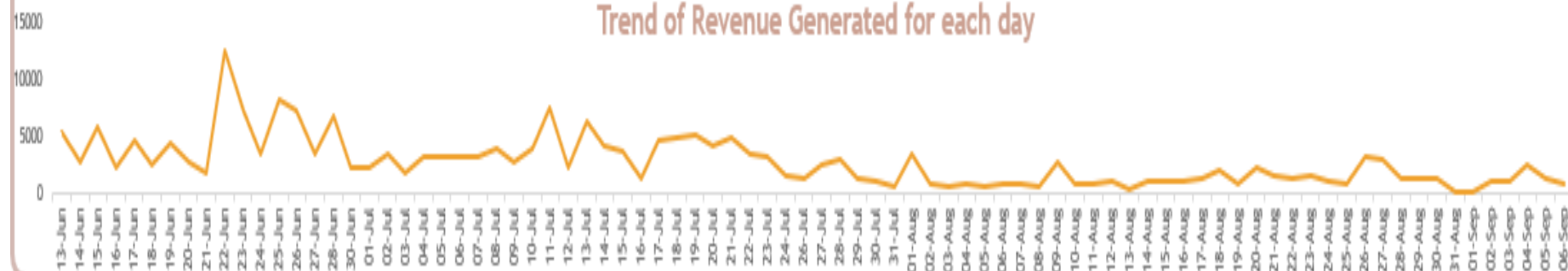
Total order



Total Revenue



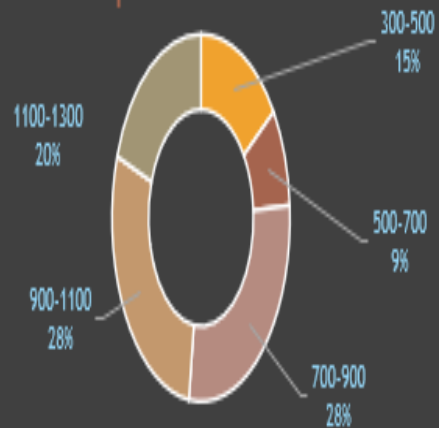
Trend of Revenue Generated for each day



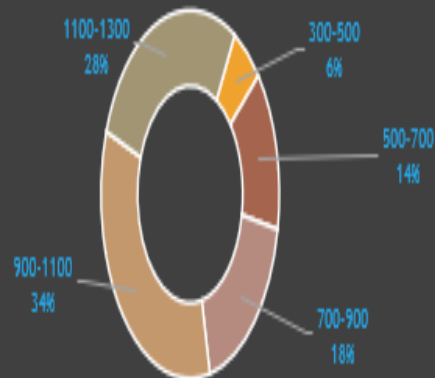
Financial Dashboard

Financial Dashboard

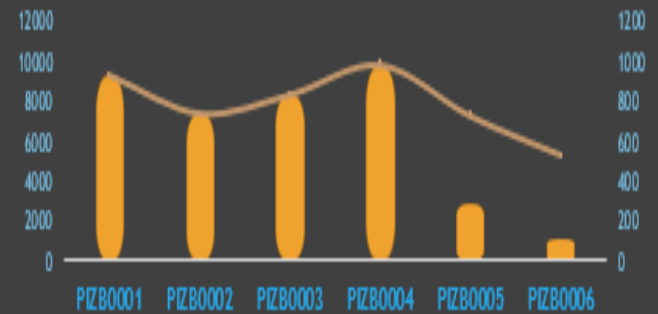
sold products - buckets



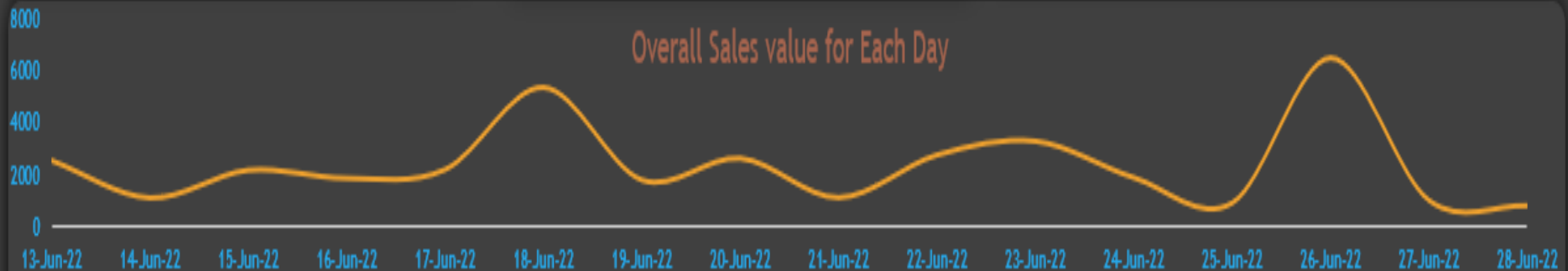
Sales Amount-buckets



product's sales value and their average prices



Overall Sales value for Each Day



Customer Service Dashboard

Total customer

50

Total Agents

3

Customer Service Report

Qtr

2022

APR

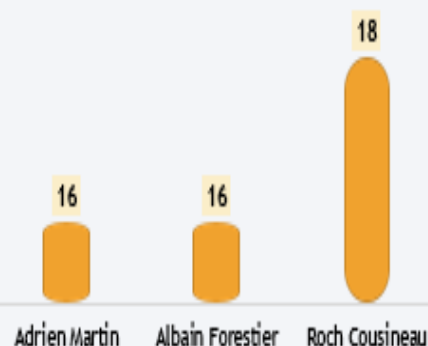
MAY

JUN

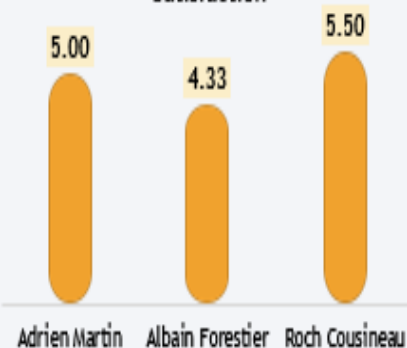
JUL

AUG

Agent's No of Interactions with customers



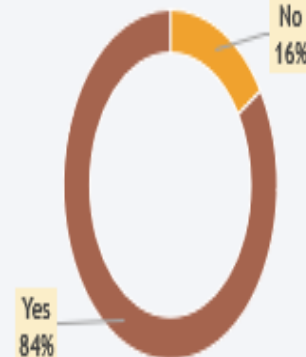
Agent's Average Customer Satisfaction



Contact Types's No of Interactions



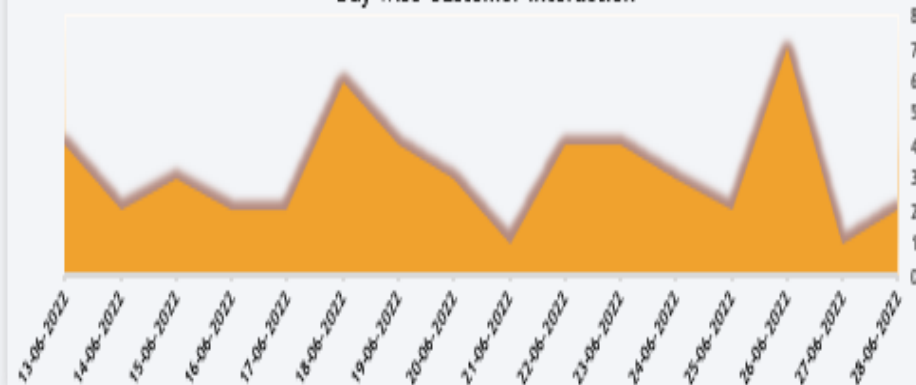
Is It for an Order ?



Day wise Customer Satisfaction



Day wise Customer Interaction



Finance Formulas

(in USD Millions)	2022	2021
Total Current Assets	1,35,405	1,34,836
Inventories	4,946	6,580
Total Current Liabilities	1,53,982	1,25,481
Quick Ratio	0.847235391	1.07455312
TOTAL CUREENT ASSET-INVENTORIES/TOTAL CURRENT LIABILITY		

(in USD Millions)	2022	2021
Total Current Assets	1,35,405	1,34,836
Inventories	4,946	6,580
Total Current Liabilities	1,53,982	1,25,481
Quick Ratio	0.847235391	1.07455312
TOTAL CUREENT ASSET-INVENTORIES/TOTAL CURRENT LIABILITY		

2021 YEAR HAD GOOD LIQIDITY

(in USD Millions)	2022	2021
Inventories	4,946	6,580
Cost of Goods Sold (COGS) incl. D&A	2,23,546	2,12,981
Inventory Turnover	38.79	
COGS /AVERAGE OF INVENTORIES		

(in USD Millions)	2022	2021
Sales/Revenue	3,94,328	3,65,817
Total Customers	2,119	1,921
ARPU	186.09	190.43

SALES OR REVE /TOTAL CUSTOMER

(in USD Millions)	2022	2021	2020	2019	2018
Sales/Revenue	3,94,328	3,65,817	2,74,150	2,59,968	2,65,809
CAGR	8.21%	RRI(NO.OF YEAR ,PRESENT VALUR,FUTURE VALUE)			
Test	394328				

Amount Borrowed	\$10,000.00
Rate of Interest	7%
Number of Years	5
Total owing at the end	\$14,025.52

principlal amount *(1 + rate of intrest)^ no.of year

(in USD Millions)	2022	2021
Sales/Revenue	3,94,328	3,65,817
Cost of Goods Sold (COGS) incl. D&A	2,23,546	2,12,981
Gross Income	1,70,782	1,52,836
SG&A Expense	51,345	43,887
EBIT	1,19,437	1,08,949
Pretax Income	1,19,103	1,09,207
Income Tax	19,300	14,527
Net Income	99,803	94,680

Net Profit Margin	25%	26%
Net Profit Margin	net income/total revenue	

(in USD Millions)	2022	2021
Sales/Revenue	3,94,328	3,65,817
Cost of Goods Sold (COGS) incl. D&A	2,23,546	2,12,981
Net Income	99,803	94,680

Revenue Growth	7.79%	REVENUE2022 / REVENUE2021 -1
COGS Growth	4.96%	COGS2022 / COGS2021 -1
Net Income Growth	5.41%	Net income2022 / Netincome2021 -1