



Portfolio Project

By Malavika Aiyer

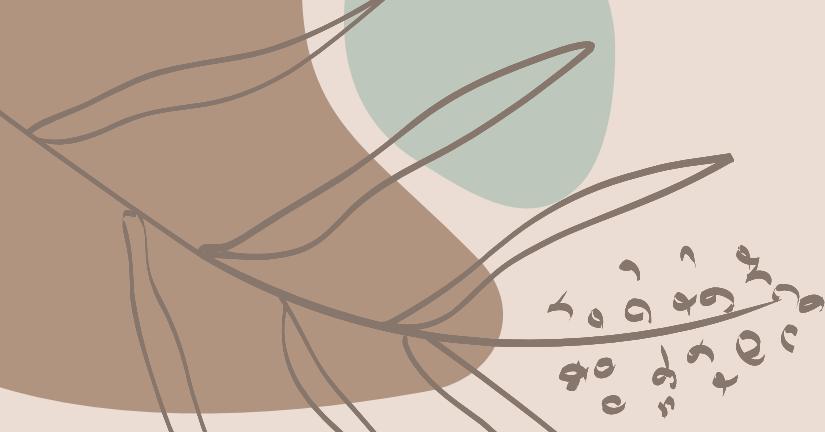
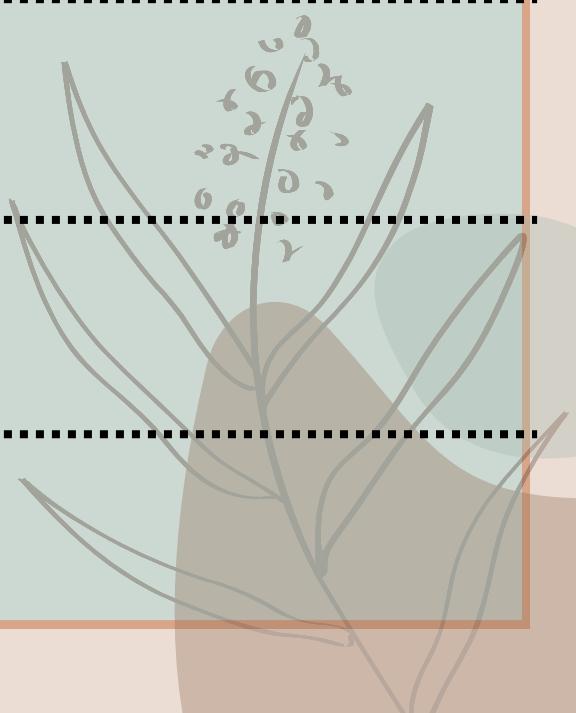


Table of Contents



- | | |
|----|---|
| 1. | Professional Background |
| 2. | Basics of an analysis topic |
| 3. | Design of all the attached projects |
| 4. | Project 1- Instagram user Analytics |
| 5. | Project 2- Operation and Metric Analysis |
| 6. | Project 3- Hiring Process Analytics |
| 7. | Project 4- IMDB Movie Analysis |
| 8. | Project 5- Bank Loan Case Study |

Table of Contents

9. **Project 6- XYZ Ads Airing Report**

10. **Project 7- ABC Call Volume Trend Analysis**

11. **Project 8- Pharmaceutical Marketing and Drug Consumption Analytics**

Professional Background

Currently pursuing my Integrated MTech (pre-final year) in Chemical engineering and have secured a CGPA of 9.6/10(. I have enrolled for data analytics to enhance my extra skills and implement this in data driven decision making.

I have had a hands on experience in building projects in C++ , SQL, Tableau and aware of basics of coding in C++, Python and MATLAB.

One of my Research papers 'Network Pharmacological approach in action of a plant phytochemical and human protein' is under progress, which also includes data driven decision making in choosing the right protein analysis. As a fresher, I'm looking for better prospective in data analytics and IT field and would like to work in this and explore my skills and enhance my knowledge.

Basics of an analysis topic

Started with the basics of data analytics by choosing a topic and creating a roadmap

Example 2

Travel Itinerary

Plan –

- Exploring various travel options and destinations of my interest
- Deciding the season of my visit and shortlisting places depending on the season

Prepare –

- Determining the cost I would be willing to spend on this trip, keeping a safe amount of money as savings.
- Organizing tours and travel package offers from various sites depending on my budget.
- Charting out the major tourist attraction spots in the place of visit and aligning it with the number of days of visit

Process –

- Deciding the place of stay in the required destination
- Checking the availability of best and affordable means of flights/trains to the destination

Analyze –

- Analyzing the best travel package/ offer which fits your dream destination within the required budget
- Determining if the travel package offers a stay at your required hotel
- Analyzing if there are any last-minute deals for your destination
- Verifying the places which the tours and travels covers, whether they cover the tourist spots which you want to visit

Share –

- Communicate with the travel agent regarding the itinerary and budget
- Read reviews from the internet and try communicating with the people who had already visited for some tips and planning from their experiences

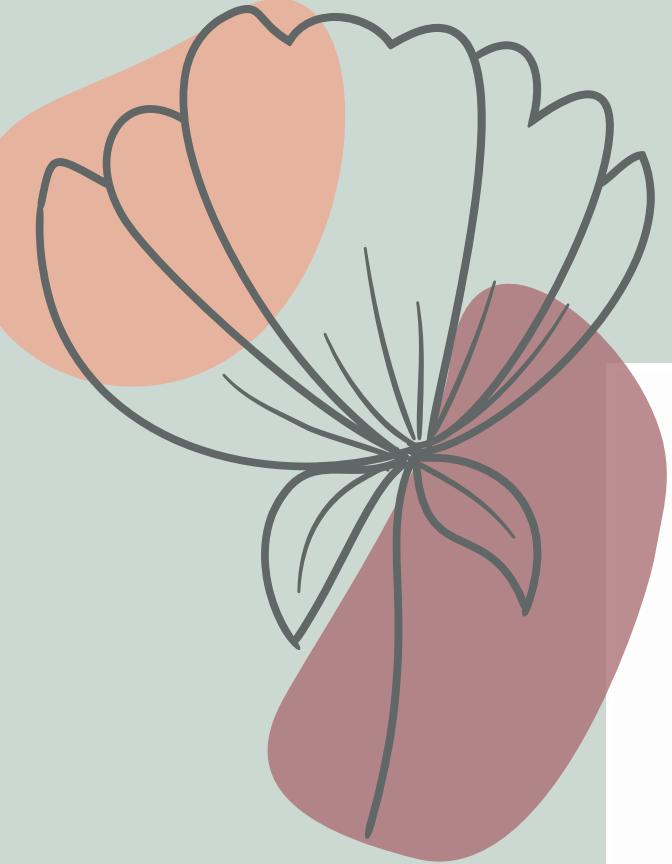
Act –

- Booking the tickets and hotels, and organizing a self-made travel plan
- Else if tours and travels complies to the budget, then choosing a proper package

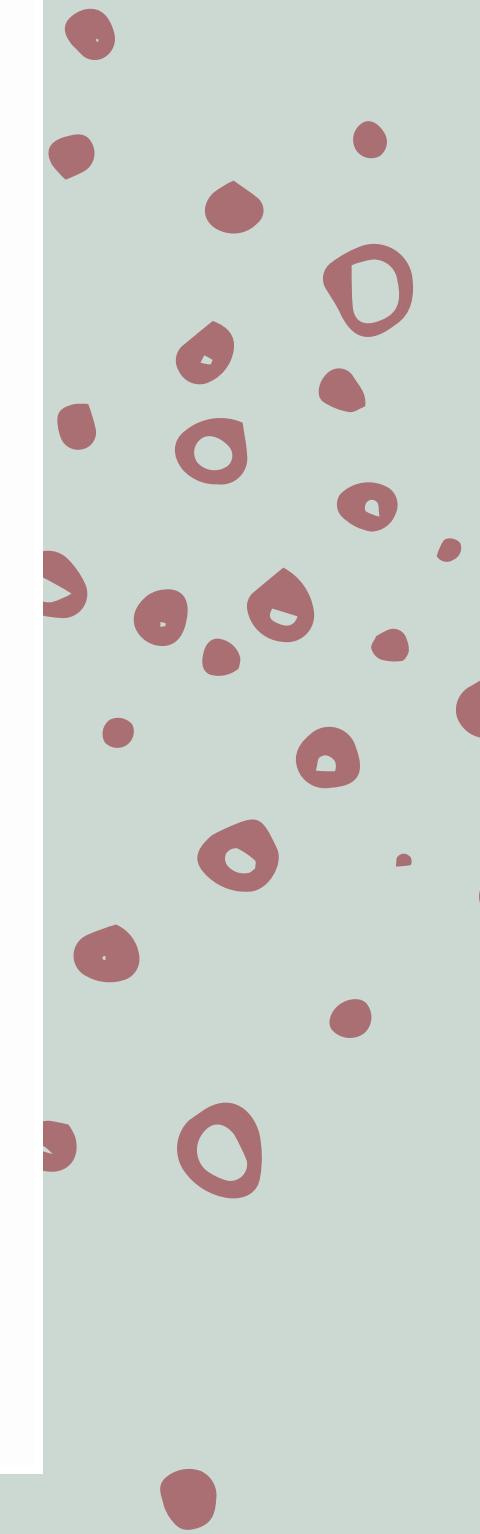
Design of all the attached projects

There are certain steps taken into count for all the attached projects before analysis

1. Importing the dataset and if required merging them.
2. Removing the Duplicates, the null columns, and unwanted data from the dataset
3. Checking the units (currency, time, percentage or the type of number)
4. Splitting or merging the cells based on the existing gaps/delimiters
5. Changing the case of names and highlighting the headers and formatting the data properly
6. If visualization required then exporting the data to Tableau



Project 1- Instagram User Analytics



Description

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.

Analysis and Insights 1

The screenshot shows a MySQL Workbench interface. At the top, there's a toolbar with various icons. Below it is a code editor window containing the following SQL query:

```
1 • select *
2   from users
3   order by created_at asc
4   limit 5;
5
```

Below the code editor is a result grid titled "Result Grid". It displays the following data:

	id	username	created_at
▶	80	Darby_Herzog	2016-05-06 00:14:21
	67	Emilio_Bernier52	2016-05-06 13:04:30
	63	Elenor88	2016-05-08 01:30:41
	95	Nicole71	2016-05-09 17:30:22
	38	Jordyn.Jacobson2	2016-05-14 07:56:26
*	HULL	HULL	HULL

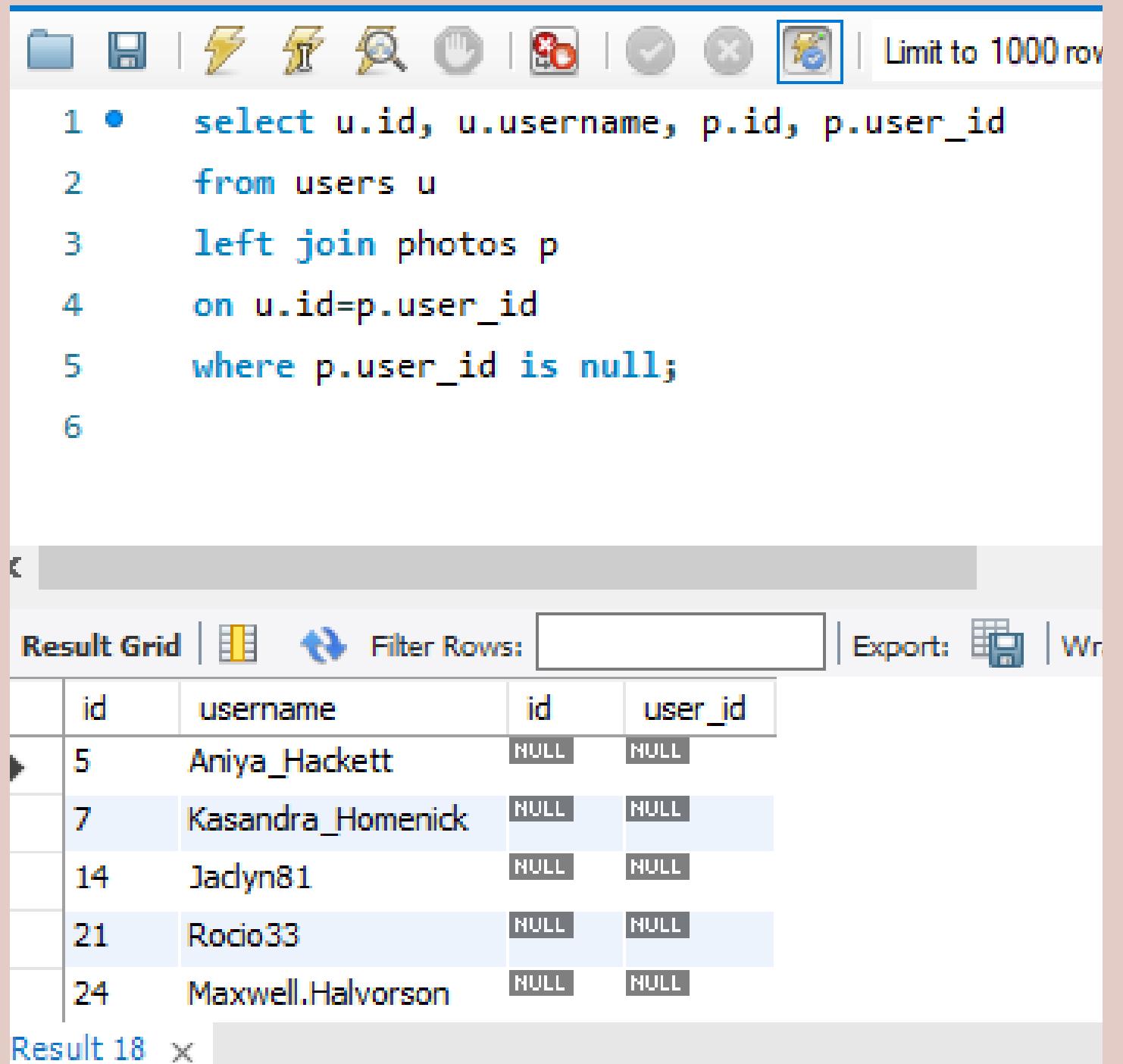
At the bottom of the result grid, it says "users 15" and has a close button.

The code gives us the top 5 oldest users of instagram.

The users with id 80,67,63,95 and 38 created the account at early stages of instagram.



Analysis and Insights 1



The screenshot shows a MySQL query editor interface. At the top, there are various icons for file operations, search, and refresh. A dropdown menu is open with the option "Limit to 1000 rows". Below the toolbar, the SQL query is displayed:

```
1 • select u.id, u.username, p.id, p.user_id
  2   from users u
  3   left join photos p
  4     on u.id=p.user_id
  5   where p.user_id is null;
  6
```

The "Result Grid" tab is selected at the bottom. The results show five rows of data, each with a red arrow pointing to the "user_id" column, which contains "NULL".

	id	username	id	user_id
▶	5	Aniya_Hackett	NULL	NULL
	7	Kassandra_Homenick	NULL	NULL
	14	Jadlyn81	NULL	NULL
	21	Rocio33	NULL	NULL
	24	Maxwell.Halvorson	NULL	NULL

Result 18 ×



A screenshot of a MySQL query editor showing a list of users with null user_id. The results are presented in a grid format with two columns: "id" and "username". An arrow points from the text "These are the analysis result of the users inactive on instagram and the app needs to send promotional email to them." to the "user_id" column in the first screenshot.

id	username
5	Aniya_Hackett
7	Kassandra_Homenick
14	Jadlyn81
21	Rocio33
24	Maxwell.Halvorson
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

These are the analysis result of the users inactive on instagram and the app needs to send promotional email to them.

Analysis and Insights 1

```
1 •  SELECT
2      u.username,
3      p.id as Photo_ID,
4      u.id as User_ID,
5      COUNT(*) AS numof_likes
6  FROM photos p
7  INNER JOIN likes l
8      ON l.photo_id = p.id
9  INNER JOIN users u
10     ON p.user_id = u.id
11 GROUP BY p.id
12 ORDER BY numof_likes DESC
13 LIMIT 1;
```

Result Grid | Filter Rows: Export:

	username	Photo_ID	User_ID	numof_likes
▶	Zack_Kemmer93	145	52	48

A red arrow points from the explanatory text below to the "username" column of the result grid, specifically highlighting the entry "Zack_Kemmer93".

This code result gives us the contest winner with the most liked instagram post. The person with username Zack_Kemmer93 won the contest

Analysis and Insights 1

```
1 • select t.tag_name, t.id, count(*) as nooftags  
2   from tags t  
3   inner join photo_tags pt on pt.tag_id=t.id  
4   group by t.id  
5   order by nooftags desc  
6   limit 5;
```

Result Grid | Filter Rows: [] | Export: [] | Wrap C

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

Result 4

This code gives the maximum used hashtag on Instagram for a better reach. These hashtags are- #smile, #beach, #party, #fun and concert



Analysis and Insights 1

The screenshot shows a MySQL Workbench interface with multiple tabs at the top: SQL File 3*, SQL File 5*, SQL File 7*, and SQL File 8* (selected). Below the tabs is a toolbar with various icons for file operations like Open, Save, and Execute. A dropdown menu says "Limit to 2000 rows". The main area contains a SQL query:

```
1 • SELECT
2     DAYNAME(created_at) AS days, COUNT(*) AS noofdays
3 FROM
4     users
5 GROUP BY days
6 ORDER BY noofdays DESC
7
```

Below the query is a "Result Grid" button and a "Filter Rows:" input field. The results table has columns "days" and "noofdays". The data is:

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

The app marketing team wants to know the best day to launch more advertisements on the app for new users.

As per the results Thursday and Sunday are the best days to launch the ad.

Analysis and Insights 1

```
1 • SELECT
2     COUNT(DISTINCT photos.id) AS noofphotos,
3     COUNT(DISTINCT username) AS noofusers,
4     COUNT(DISTINCT photos.id) / COUNT(DISTINCT username) AS averageasperdata
5 FROM
6     photos,
7     users
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: []

noofphotos	noofusers	averageasperdata
257	100	2.5700

User Engagement-
There are 257
photos and 100
users. So the
average photos per
user is 2.57

```
1 • SELECT count(u.id) as inactive_users
2     -- u.id, u.username, p.id, p.user_id
3 FROM
4     users u
5         LEFT JOIN
6             photos p ON u.id = p.user_id
7 WHERE
8     p.user_id IS NULL;
```

Result Grid | Filter Rows: [] | Export: []

inactive_users
26

But there are also 26 inactive
users so the new average
changes to $257/(100-26)=257/74=3.47$

Analysis and Insights 1

The screenshot shows a MySQL query editor interface. At the top, there are tabs for "SQL File 3*", "SQL File 5*", "SQL File 7*", "SQL File 8*", "SQL File 9*", and "SQL File 10*". Below the tabs is a toolbar with various icons. The main area contains the following SQL code:

```
1 • sELECT users.id, username, COUNT(likes.photo_id) As totallikes
2   FROM users
3   inner JOIN likes ON users.id = likes.user_id
4   GROUP BY users.id
5   HAVING totallikes = (SELECT COUNT(*) FROM photos)
```

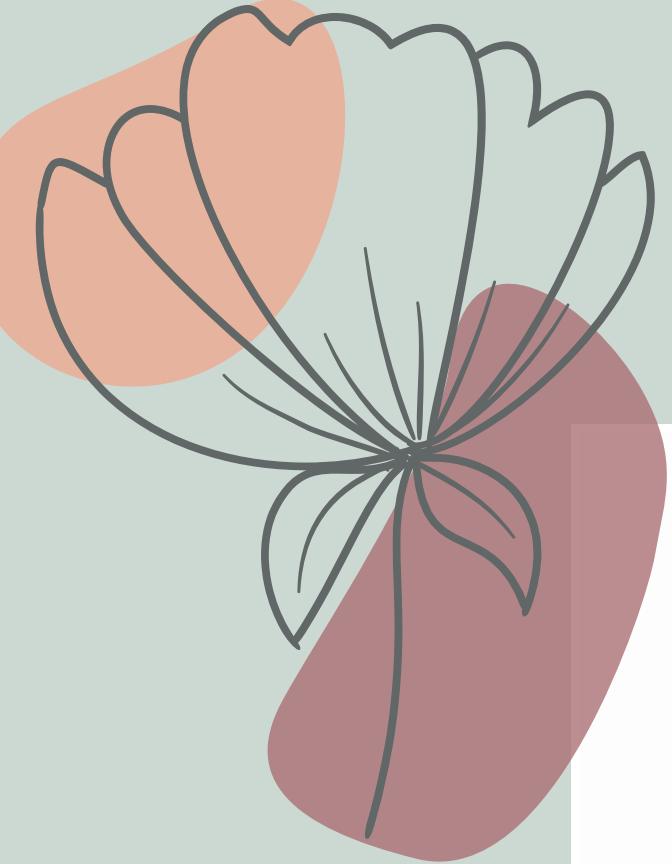
Below the code is a "Result Grid" table with columns: id, username, and totallikes. The data is identical to the one shown in the "Result 8" table below. The "Result 8" table has a header row and 10 data rows. The "Result Grid" table has a header row and 7 data rows.

	id	username	totallikes
▶	5	Aniya_Hackett	257
	14	Jadyn81	257
	21	Rocio33	257
	24	Maxwell.Halvorson	257
	36	Ollie_Ledner37	257
	41	Mckenna17	257
	54	Duane60	257

Result 8 ×

id	username	totallikes
5	Aniya_Hackett	257
14	Jaclyn81	257
21	Rocio33	257
24	Maxwell.Halvorson	257
36	Ollie_Ledner37	257
41	Mckenna17	257
54	Duane60	257
57	Julien_Schmidt	257
66	Mike.Auer39	257
71	Nia_Haag	257
75	Leslie67	257
76	Janelle.Nikolaus81	257
91	Bethany20	257

The suspected list of fake accounts/bots on the app

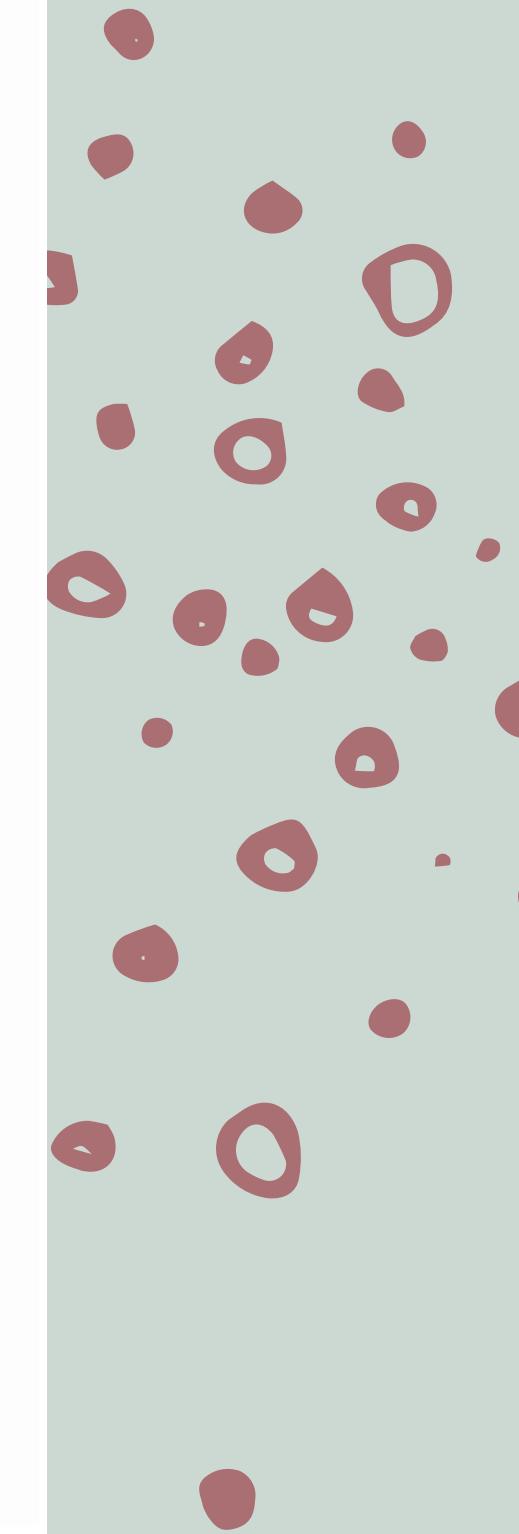


Project 2-Operation and metric analysis

Description

Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. this kind of analysis is further used to predict the overall growth or decline of a company's fortune.

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



Analysis and Insights 2

SQL File 4* X

File | New | Open | Save | Print | Run | Don't Limit | Help | Favorites | Search | Options |

```
1 •   select ds,a.job_number, a.hoursspent, job_number/hoursspent as answer
2   from
3   (select ds, count(*) as job_number,
4       sum(time_spent)/3600 as hoursspent
5   from job_datatable
6   where ds between '11/01/2020' and '11/30/2020'
7   group by ds)a
```

Result Grid | Filter Rows: [] | Export: | Wrap Cell Content: []

	ds	job_number	hoursspent	answer
▶	11/30/2020	2	0.0111	180.1802
	11/29/2020	1	0.0056	178.5714
	11/28/2020	2	0.0092	217.3913
	11/27/2020	1	0.0289	34.6021
	11/26/2020	1	0.0156	64.1026
	11/25/2020	1	0.0125	80.0000

The number of jobs reviewed per hour per day is calculated. Indirect working efficiency of an employee is being depicted

Analysis and Insights 2

SQL File 4* SQL File 7* x SQL File 4* SQL File 3* SQL File 5* SQL File 6*

File | New | Open | Save | Print | Database | Help | Don't Limit | Favorites | Search | Help |

```
1 •   SELECT ds, number_of_events,
2     AVG(number_of_events)OVER(ORDER BY ds ROWS BETWEEN 6 PRECEDING AND
3     CURRENT ROW)AS 7_day_rolling_avg FROM
4     (SELECT ds, COUNT(event) AS number_of_events
5      FROM job_datatable
6     GROUP BY ds
7     ORDER BY ds)a
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

	ds	number_of_events	7_day_rolling_avg
▶	11/25/2020	1	1.0000
	11/26/2020	1	1.0000
	11/27/2020	1	1.0000
	11/28/2020	2	1.2500
	11/29/2020	1	1.2000
	11/30/2020	2	1.3333

A 7 day rolling average of the throughput is being calculated here

Analysis and Insights 2

SQL File 4* SQL File 4* SQL File 3* SQL File 5* ×

File | ⚡ ⚡ 🔎 ⚡ ⚡ | 📁 | 🌐 | ⚡ | Don't Limit

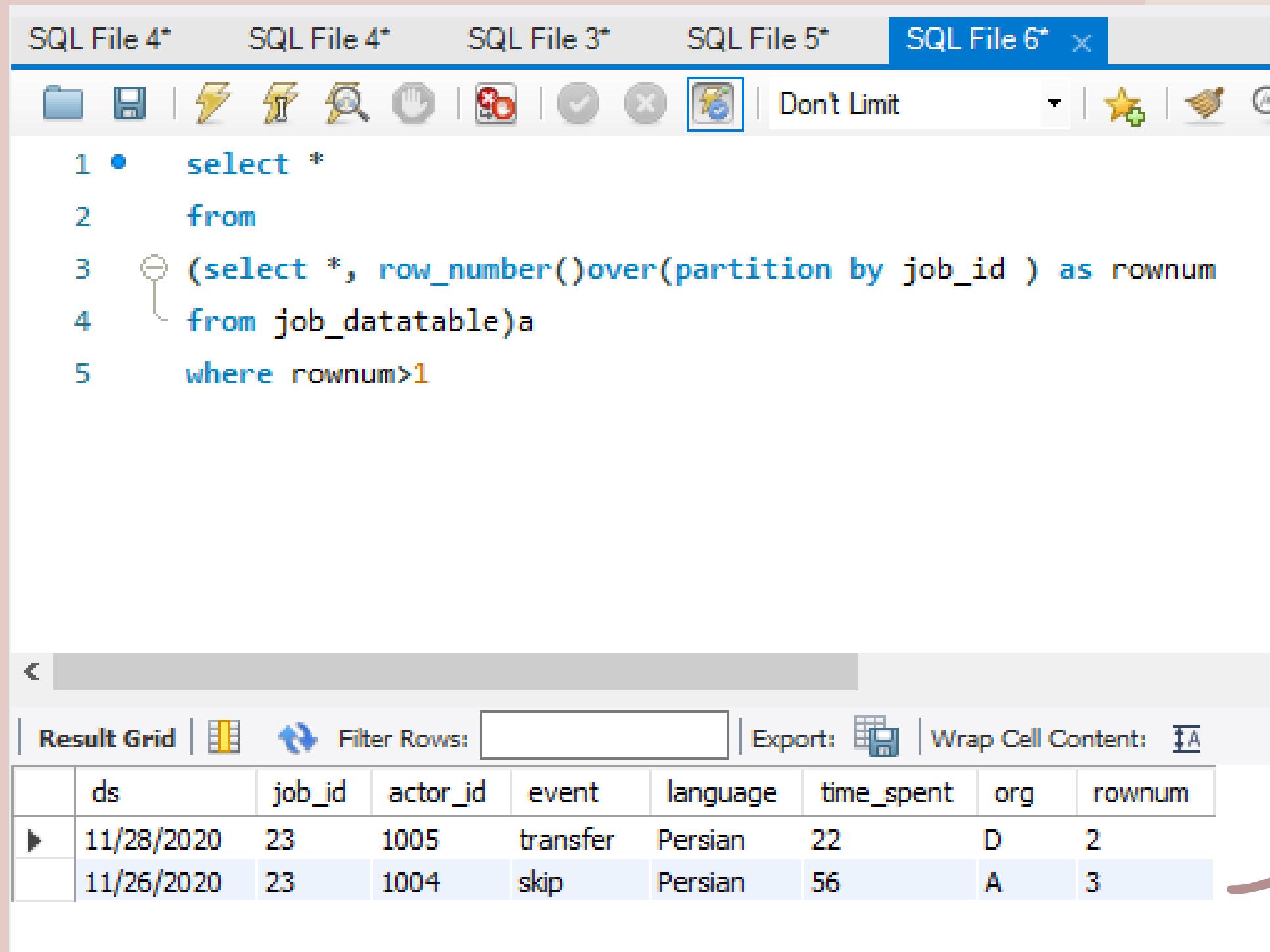
```
1 • select a.language, truncate((nooflang*100)/total,3)
2   from
3     (select language,
4       count(language) as nooflang
5       from job_datatable
6       group by language)a
7   cross join
8     (select count(job_id) as total
9      from job_datatable)b
10
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

	language	truncate((nooflang*100)/total,3)
▶	English	12.500
	Arabic	12.500
	Persian	37.500
	Hindi	12.500
	French	12.500
	Italian	12.500

% share of each
language in the last 30
days

Analysis and Insights 2



The screenshot shows a SQL editor interface with multiple tabs at the top labeled "SQL File 4*", "SQL File 4*", "SQL File 3*", "SQL File 5*", and "SQL File 6*". The "SQL File 6*" tab is active, containing the following SQL code:

```
1 • select *
2   from
3     (select *, row_number()over(partition by job_id ) as rounum
4      from job_datatable)a
5   where rounum>1
```

Below the code is a "Result Grid" showing the output of the query. The grid has columns: ds, job_id, actor_id, event, language, time_spent, org, and rounum. There are two rows of data:

	ds	job_id	actor_id	event	language	time_spent	org	rounum
▶	11/28/2020	23	1005	transfer	Persian	22	D	2
	11/26/2020	23	1004	skip	Persian	56	A	3

The duplicate values
in the table are
displayed.

Analysis and Insights 2

```
1 select extract(week from a.occurred_at) as weeknum,  
2 count(distinct user_id) as users  
3 from tutorial.yammer_events a  
4 group by weeknum
```

	Fields	Source
	weeknum	users
1	18	791
2	19	1244
3	20	1270
4	21	1341
5	22	1293
6	23	1366
7	24	1434
8	25	1462
9	26	1443
10	27	1477
11	28	1556
12	29	1556
13	30	1593
.	.	.

Measuring the number of active users in a week to check up the quality of product/services

Analysis and Insights 2

Run Limit 100 Format View history

```
1 select year,week,
2 sum(noofusers)
3 over(order by year,week rows between unbounded preceding and current row)
4 as total_active_users from
5 (select extract(year from created_at) as year,
6 extract(week from created_at) as week,
7 count(distinct user_id) as noofusers
8 from tutorial.yammer_users
9 where state in('active')
10 group by year, week
11 order by year,week)a
```

Data Fields Source

	year	week	total_active_users
1	2013	1	67
2	2013	2	96
3	2013	3	143
4	2013	4	179
5	2013	5	209
6	2013	6	257
7	2013	7	298
8	2013	8	337

<< < Page 1 of 1 > >> Showing rows 1-87 of 87 Columns 3 Size

Measuring the activity of the users after the weekly app update to check if the customers are really satisfied with the update or not

Analysis and Insights 2

```
1 SELECT
2   COUNT(user_id)as users,
3   SUM(CASE WHEN retention_week = 1 THEN 1 ELSE 0 END ) AS week_1,
4   SUM(CASE WHEN retention_week = 2 THEN 1 ELSE 0 END ) AS week_2,
5   SUM(CASE WHEN retention_week = 3 THEN 1 ELSE 0 END ) AS week_3,
6   SUM(CASE WHEN retention_week = 4 THEN 1 ELSE 0 END ) AS week_4,
7   SUM(CASE WHEN retention_week = 5 THEN 1 ELSE 0 END ) AS week_5
8   FROM
9   (
10   SELECT a.user_id,
11   a.sign_up_week,
12   b.engagement_week,
13   b.engagement_week - a.sign_up_week as retention_week
14   FROM
15   (
16   (select distinct user_id, extract(week from occurred_at) as sign_up_week
17   from tutorial.yammer_events
18   where event_type='signup_flow'
19   and event_name='complete_signup'
20   and extract(week from occurred_at)=18
21   )a
22   left JOIN
23   (
24   select distinct user_id,
25   extract(week from occurred_at) as engagement_week
26   from tutorial.yammer_events
27   where event_type='engagement'
28   )b
29   on a.user_id=b.user_id
30   )
31   order by
32   a.user_id )a
```

users	week_1	week_2	week_3	week_4	week_5
317	64	27	19	15	19

Calculating the weekly retention of users signup cohort. As we can see that the number of users signing up in week 1 is 64 and at week 5 is 19. We need to give the users a survey form to improve upon the updates

Analysis and Insights 2

```
1 select device, extract(week from occurred_at) as weeknum,  
2 count(distinct user_id) as users  
3 from tutorial.yammer_events  
4 where event_type='engagement'  
5 group by device, weeknum  
6 order by device, weeknum
```

	Data	Fields	Source
	device	weeknum	users
1	acer aspire desktop	18	10
2	acer aspire desktop	19	26
3	acer aspire desktop	20	22
4	acer aspire desktop	21	23
5	acer aspire desktop	22	28
6	acer aspire desktop	23	25
7	acer aspire desktop	24	21
8	acer aspire desktop	25	23
9	acer aspire desktop	26	29
10	acer aspire desktop	27	28
11	acer aspire desktop	28	29

Weekly engagement per device is calculated to improve the device features based on the user performance

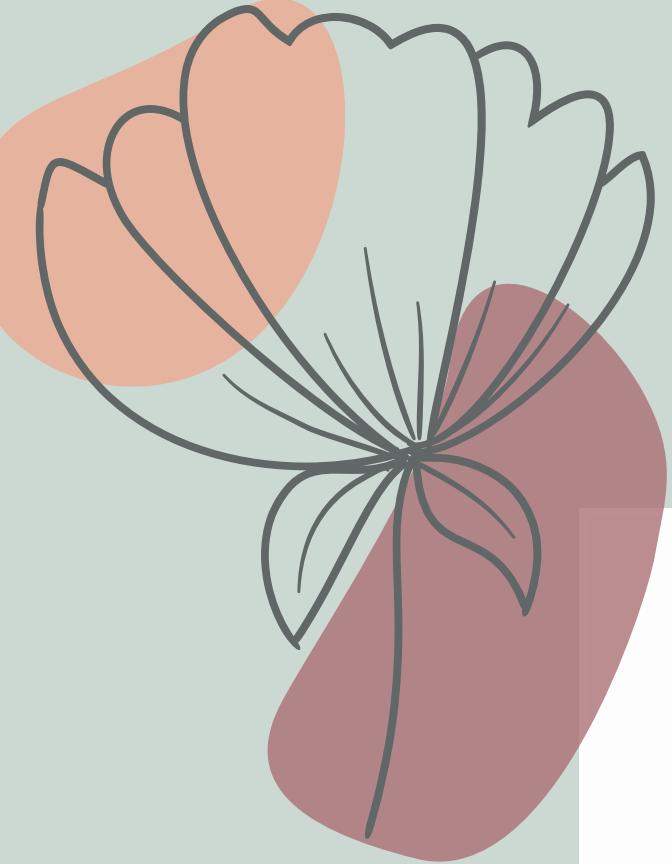
Analysis and Insights 2



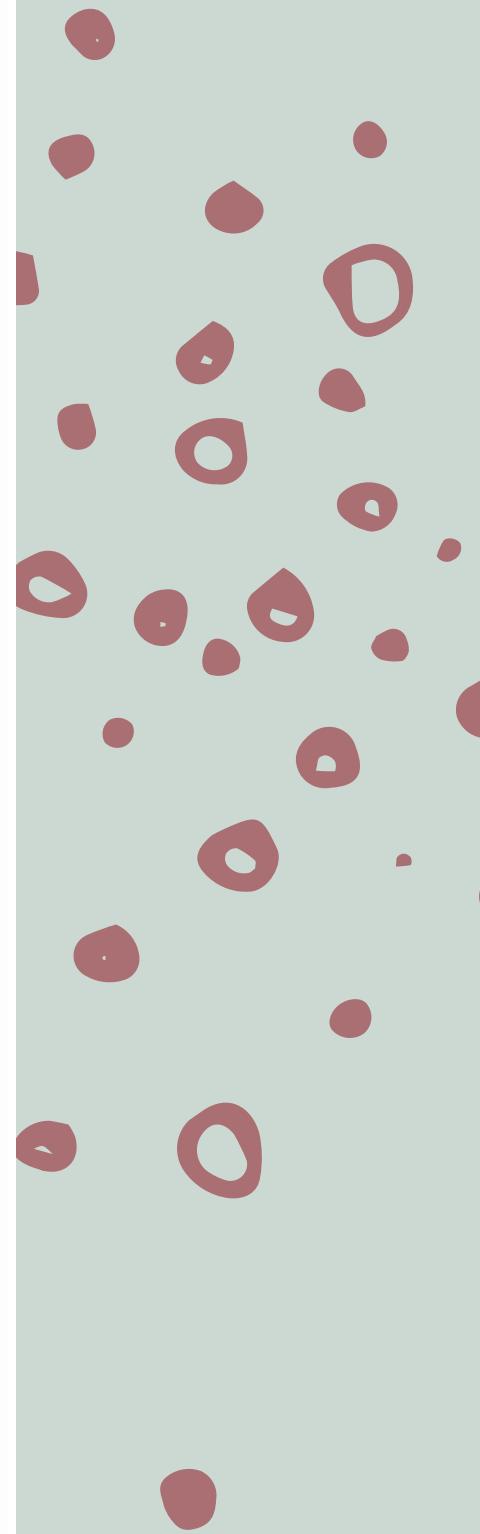
```
1 select
2 SUM(case when email_cat ='email_open' then 1 else 0 end ) as users_opened_email,
3 SUM(case when email_cat='email_sent' then 1 else 0 end )as users_sent_email,
4 SUM(case when email_cat ='email_clicked' then 1 else 0 end ) as users_clicked_email
5 FROM
6 (
7 SELECT
8 *,
9 CASE
10 WHEN action in('sent_weekly_digest','sent_reengagement_email')
11 | then 'email_sent'
12 WHEN action in('email_open')
13 | then 'email_open'
14 WHEN action in('email_clickthrough')
15 | then 'email_clicked'
16 end as email_cat
17 from tutorial.yammer_emails
18 ) a
```

Email engagement metrics to follow up on the emails received/sent/replied by the user.

Data	Fields	Source
	users_opened_email	users_sent_email
1	20459	57267
		users_clicked_email
		9010



Project 3- Hiring Process Analytics



Description

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.

Our job is to go through these trends and draw insights out of it for hiring department to work upon.

Analysis and Insights 3

	Hired	Total Applied
Female	1856	2675
Male	2563	4085
TOTAL	4419	6760

```
=COUNTIFS(C:C,"Hired",D:D,"Female")
```

```
=COUNTIFS(D:D,"Male",C:C,"Hired")
```

The number of female and male candidates hired

Sum of salary	358228369
Average of salary	49983.029

```
=SUM(G:G)
```

```
=AVERAGE(G:G)
```

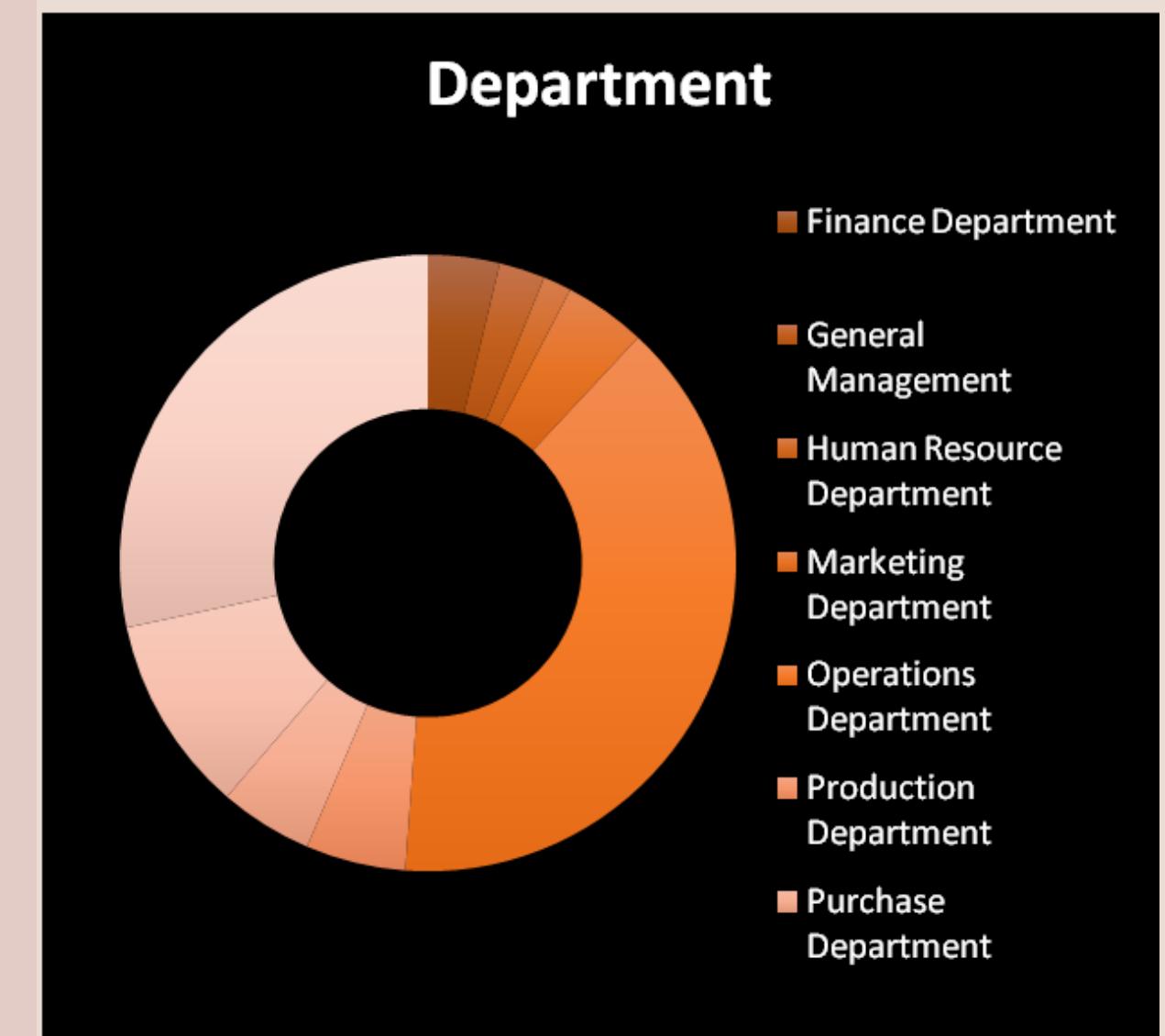
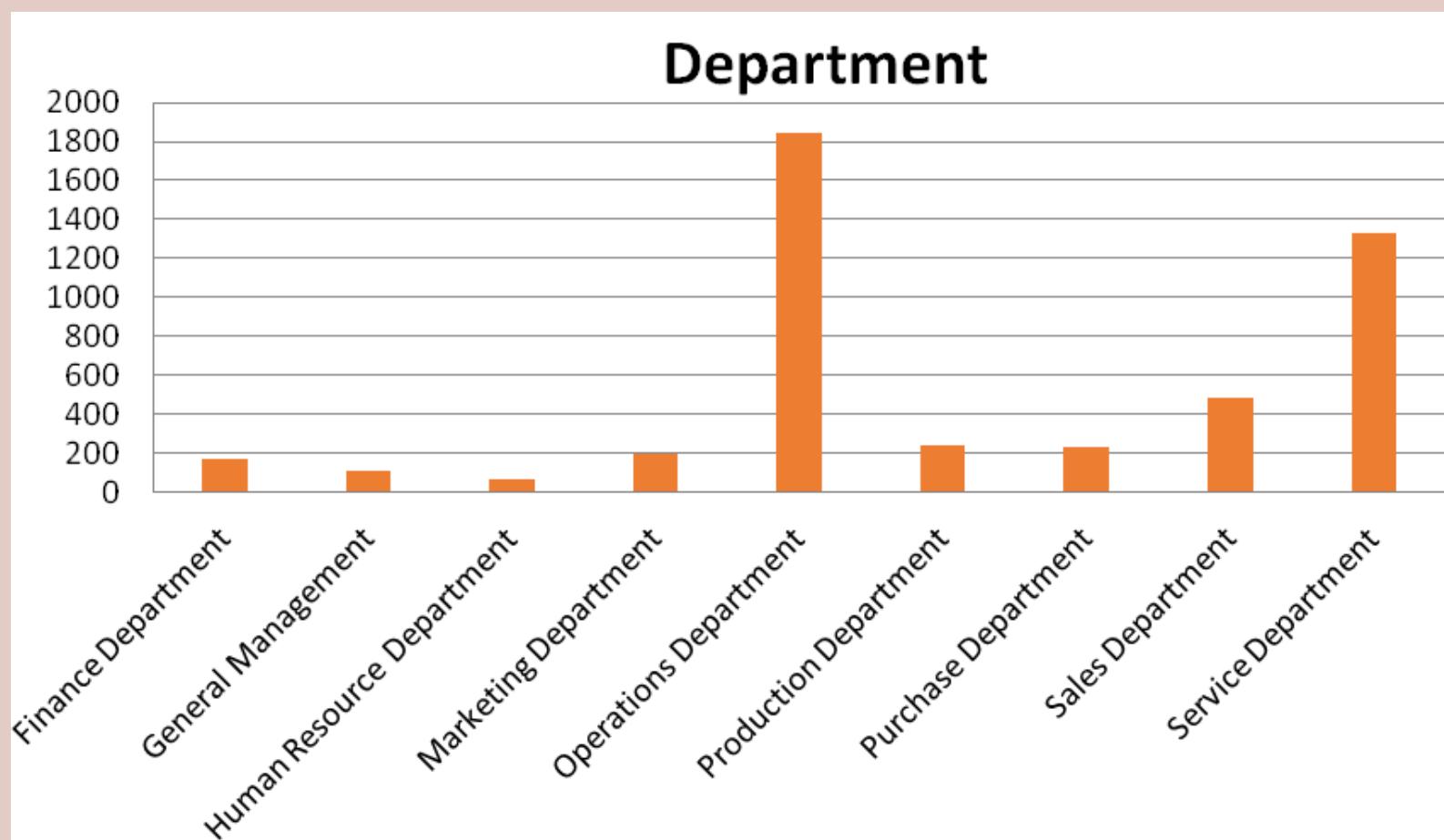
The sum and average of salary offered.

Analysis and Insights 3



Row Labels	Column Labels	
	Department	Grand Total
Finance Department	176	176
General Management	113	113
Human Resource Department	70	70
Marketing Department	202	202
Operations Department	1843	1843
Production Department	246	246
Purchase Department	230	230
Sales Department	485	485
Service Department	1332	1332
Grand Total	4697	4697

The people hired are maximum in the finance dept. and minimum in the HR Dept.



Analysis and Insights 3

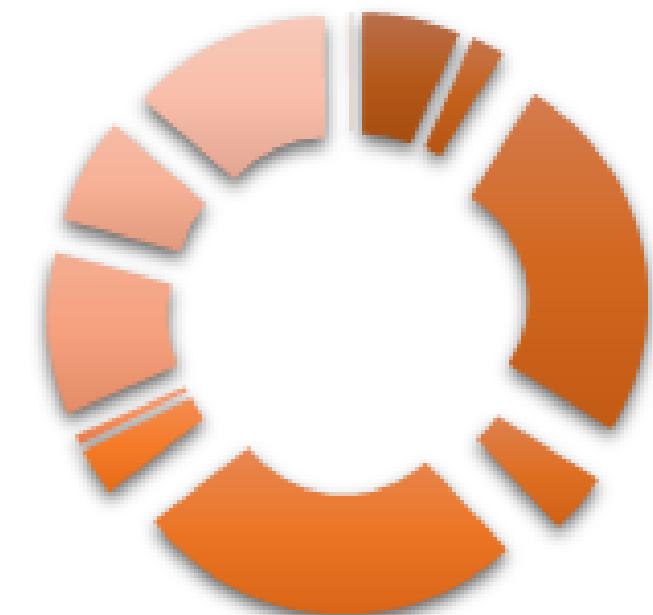


People hired for c9 post are higher whereas for i4 post is the minimum.

Row Labels	Column Labels		Grand Total
	People Hired		
-	1	1	
b9	308	308	
c-10	105	105	
c5	1182	1182	
c8	193	193	
c9	1239	1239	
i1	151	151	
i4	32	32	
i5	511	511	
i6	337	337	
i7	635	635	
m6	2	2	
n6	1	1	
Grand Total	4697	4697	



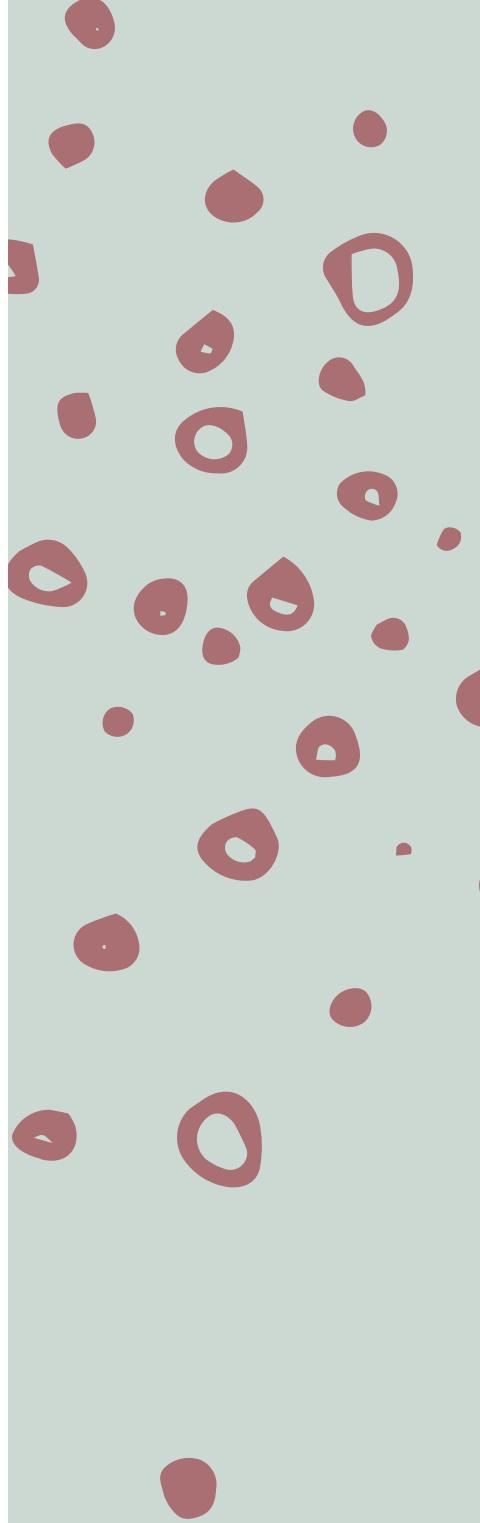
People Hired



-
- b9
- c-10
- c5
- c8
- c9
- i1
- i4
- i5
- i6
- i7
- m6
- n6



Project 4-IMDB movie analysis



Description

The movie review analytics helps us to understand the needs of the user and depending upon the most and least watched movies its easier to categorise and draw insights and analyze. Determining the business analytics and insights based on the IMDB rating of different movies and reviews.

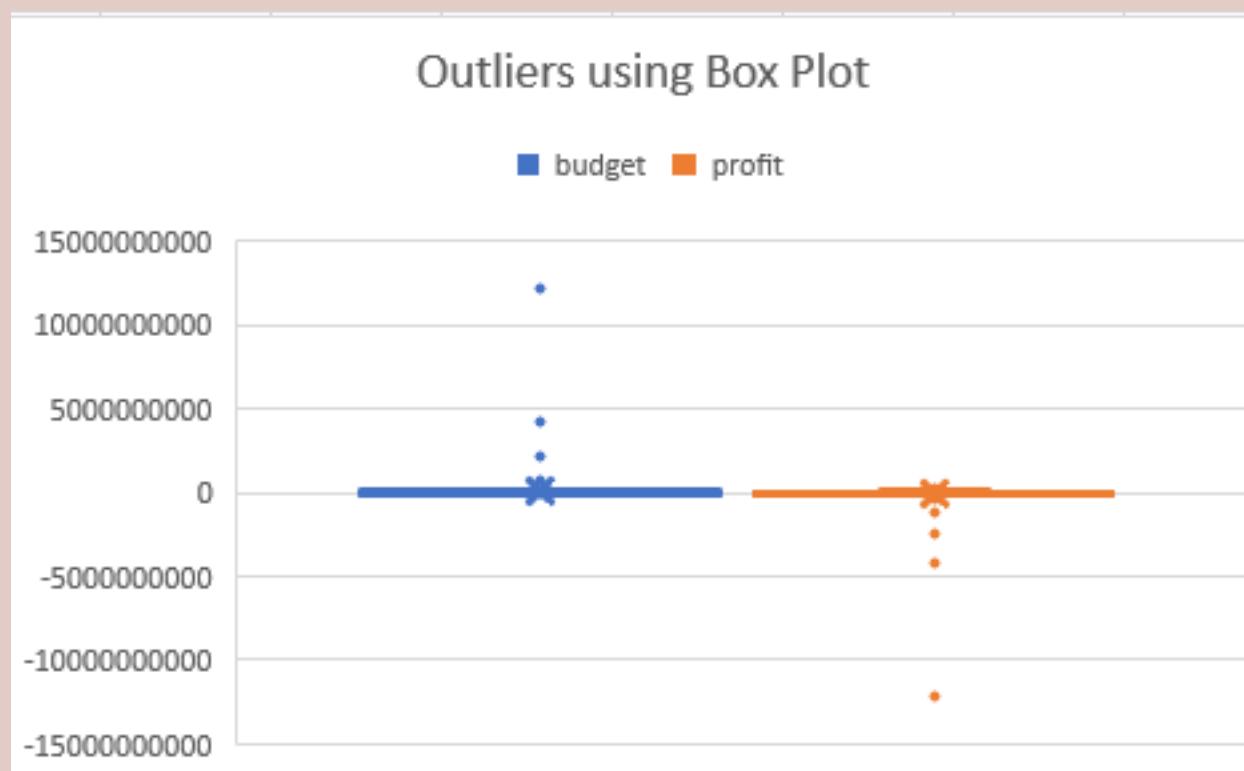
Whys technique to determine its root cause by repeatedly asking the question “Why”.

It's also called the Root Cause Analysis, developed by Sakichi Toyoda, founder of Toyota Industries.

Analysis and Insights 4

budget	profit
237000000	523505847
300000000	9404152
245000000	-44925825
250000000	198130642
0	0
263700000	-1.91E+08
258000000	78530303
260000000	-59192738
250000000	208991599
250000000	51956980
250000000	80249062
209000000	-8930592
200000000	-31631573
225000000	198032628
215000000	-1.26E+08
225000000	66021565
225000000	-83385977

MOVIE TITLE	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
Star Wars: Episode I - The Phantom Menace	359544677
The Dark Knight	348316061
The Hunger Games	329999255



The profit is calculated from gross and budget difference and the movies which made the highest profit is displayed. Also the outliers of budget and profit is shown here.

Analysis and Insights 4

MOVIE TITLE	IMDB SCORE	num_voted_users
The Shawshank	9.3	1689764
The Godfather	9.2	1155770
The Dark Knight	9	1676169
The Godfather	9	790926
The Lord of the	8.9	1215718
Schindler's Lis	8.9	865020
Pulp Fiction	8.9	1324680
The Good, the	8.9	503509
12 Angry Men	8.9	447785
Inception	8.8	1468200
The Lord of the	8.8	1238746
Fight Club	8.8	1347461
Forrest Gump	8.8	1251222
Star Wars: Epis	8.8	837759
The Lord of the	8.7	1100446
The Matrix	8.7	1217752
Goodfellas	8.7	728685
Star Wars: Epis	8.7	911097
One Flew Ove	8.7	680041
City of God	8.7	533200
Interstellar	8.6	928227
Saving Private	8.6	881236
Se7en	8.6	1023511
The Silence of	8.6	887467
Spirited Away	8.6	417971

The Prestige	8.5	844052
The Pianist	8.5	497946
Apocalypse Now	8.5	450676
Raiders of the	8.5	661017
Psycho	8.5	422432
Back to the Fu	8.5	732212
Alien	8.5	563827
Memento	8.5	845580
Whiplash	8.5	399138
Dr. Strangelov	8.5	342585
WALL·E	8.4	718837
Braveheart	8.4	736638
Amélie	8.4	534262
Star Wars: Epis	8.4	681857
The Shining	8.4	610333
Aliens	8.4	488537
American Beau	8.4	822500
Requiem for a	8.4	573541
Oldboy	8.4	356181
Reservoir Dog	8.4	664719
Toy Story 3	8.3	544884
Up	8.3	665575
Inside Out	8.3	345198
Batman Begin	8.3	980946
Inglourious Ba	8.3	885175
Indiana Jones	8.3	515306

The list shows the movie title and the imdb score arranged in descending order . The Shawshank redemption is the movie with highest imdb rating.

Analysis and Insights 4

LANGUAGE	MOVIE NAME	IMDB RATING
Polish	Dekalog	9.1
Polish	Dekalog	9.1
Italian	The Good, the Bad	8.9
Italian	Gomorrah	8.7
Portuguese	City of God	8.7
Japanese	Seven Samurai	8.7
Japanese	Spirited Away	8.6
Hindi	Airlift	8.5
None	Samsara	8.5
German	The Lives of Others	8.5
Persian	Children of Heaven	8.5
French	Amélie	8.4
Telugu	Baahubali: The Beginning	8.4
Japanese	Princess Mononoke	8.4
German	Das Boot	8.4
Hindi	Rang De Basanti	8.4
Korean	Oldboy	8.4
Persian	A Separation	8.4
French	The Returned	8.3
German	Metropolis	8.3
German	Downfall	8.3
Danish	The Hunt	8.3
Japanese	Godzilla Resurgence	8.2
Icelandic	Trapped	8.2

French	Irreversible	7.4
Hindi	Gandhi, My Father	7.4
Czech	I Served the King	7.4
German	Summer Storm	7.4
Polish	Ida	7.4
Mandarin	Silver Medallist	7.4
Spanish	Under the Same Moon	7.4
Vietnamese	Journey from the	7.4
Hindi	Monsoon Wedding	7.4
Arabic	Ajami	7.4
Dari	Osama	7.4
French	The Case of the	7.4
French	Une Femme Mariée	7.4
Swahili	Stories of Our Lives	7.4
Japanese	The Cure	7.4
French	Molière	7.3
Mongolian	Mongol: The Rise	7.3
French	Paris, je t'aime	7.3
French	The Widow of Saïgon	7.3
Cantonese	Shaolin Soccer	7.3
Korean	The Good, the Bad	7.3
Italian	Loose Cannons	7.3
Danish	Flame and Citron	7.3
French	Polisse	7.3
Hindi	Namastey London	7.3
French	Far from Men	7.3

Urdu	Karachi se Lahore	7
Hindi	Faith Connections	7
Mandarin	Bodyguards and Assassins	6.9
French	Bon voyage	6.9
Japanese	Steamboy	6.9
Mandarin	Reign of Assassins	6.9
Icelandic	Of Horses and Men	6.9
Hindi	Jab Tak Hai Jaan	6.9
Hindi	Yeh Jawaani Hai Deewani	6.9
French	Alias Betty	6.9
Zulu	Faith Like Potatoe	6.9
Spanish	The King of Najayo	6.9
Italian	The Beyond	6.9
French	When the Cat's Away	6.9
French	Clean	6.9
Spanish	El Mariachi	6.9
French	A Monster in Paris	6.8
French	The Chambermaid	6.8
Spanish	La otra conquista	6.8
French	High Tension	6.8
Spanish	El crimen del padrino	6.8
Spanish	Heli	6.8
Cantonese	The Motel	6.8
German	3	6.8
Filipino	The Great Raid	6.7
French	Coco Before Chanel	6.7

The language movies having highest imdb rating are depicted out of which a polish movie titled "Dekalog" has a rating of 9.1

Analysis and Insights 4

Row Labels	Average of imdb
Action Adventure Crime Drama Sci-Fi Thriller	8.8
Action Adventure Biography Drama History	8.6
Crime Drama Fantasy Mystery	8.5
Action Drama History Thriller War	8.5
Adventure Animation Drama Family Musical	8.5
Action Adventure Drama Fantasy War	8.4
Comedy Drama History Romance	8.4
Action Animation Crime Sci-Fi Thriller	8.4
Adventure Drama Thriller War	8.4
Comedy Mystery Thriller	8.3
Adventure Animation Comedy Drama Family Fantasy	8.3
Adventure Drama History Thriller War	8.3
Biography Drama History Music	8.3
Documentary Drama Sport	8.3
Adventure Drama Fantasy Mystery	8.3

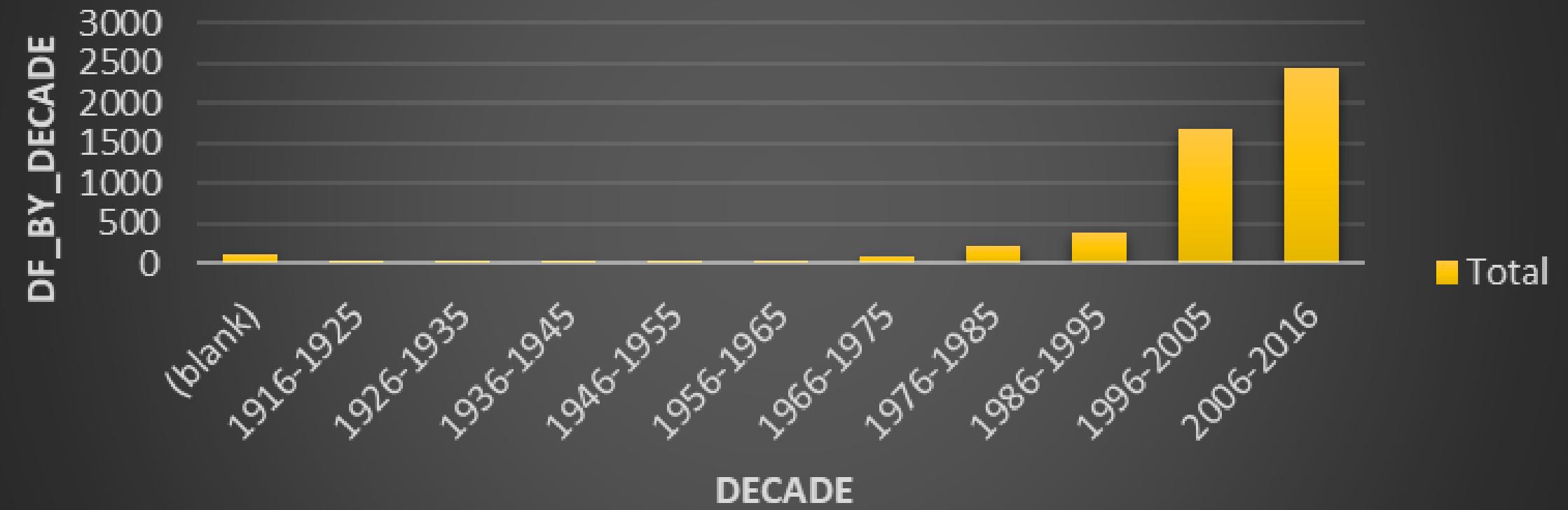
The most popular genre of movies are action, Adventure, Crime, Drama, Sci-fi, Thriller

director name	imdb
James Cameron	7.9
Gore Verbinski	7.1
Sam Mendes	6.8
Christopher Nolan	8.5
Doug Walker	7.1
Andrew Stanton	6.6
Sam Raimi	6.2
Nathan Greno	7.8
Joss Whedon	7.5
David Yates	7.5
Zack Snyder	6.9
Bryan Singer	6.1
Marc Forster	6.7
Gore Verbinski	7.3
Gore Verbinski	6.5
Zack Snyder	7.2
Andrew Adamson	6.6
Joss Whedon	8.1
Rob Marshall	6.7
Barry Sonnenfeld	6.8
Peter Jackson	7.5

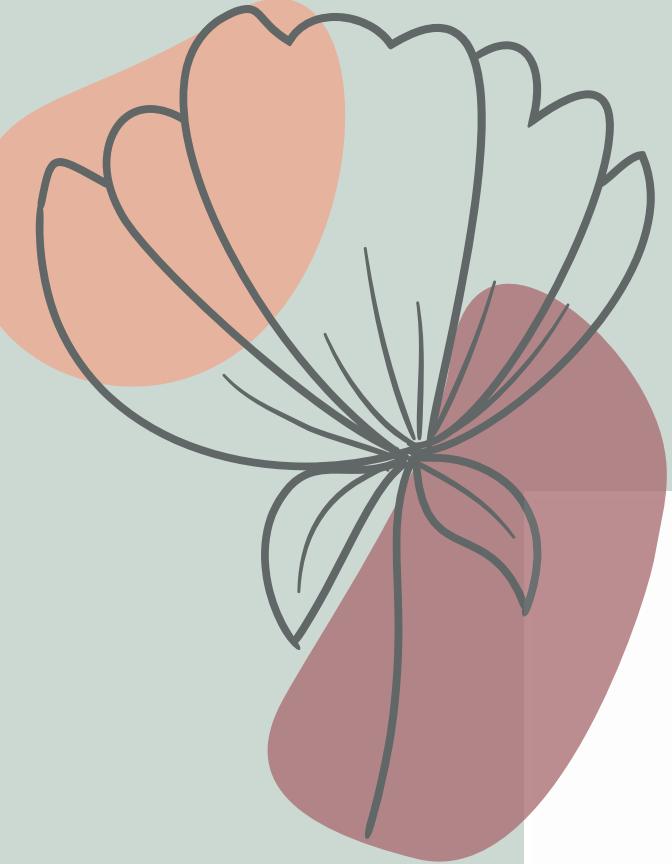
James Cameron is by far the best director based on imdb rating

Analysis and Insights 4

df_by_decade vs Decade



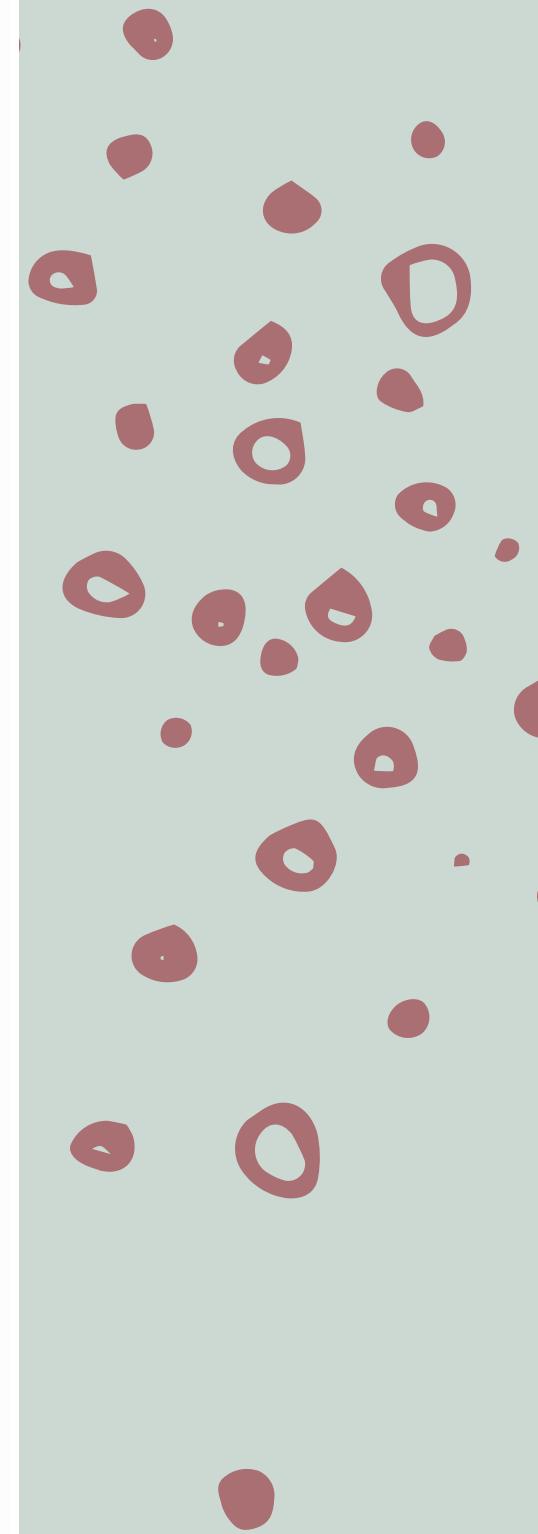
Number of people voting in a decade is displayed here through a graph.



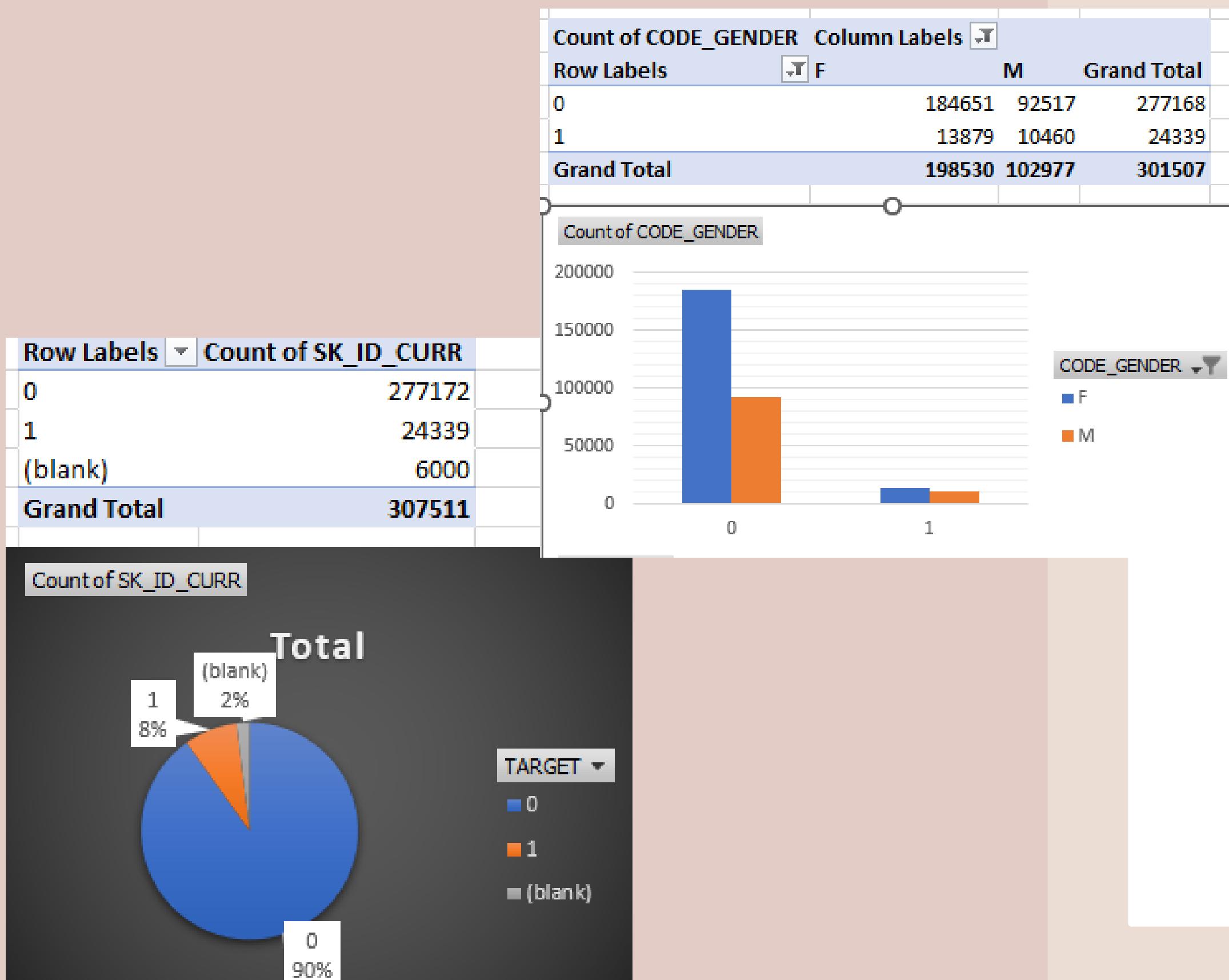
Project 5- Bank Loan Case Study

Problem

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 specialises in lending various types of loans to urban customers. You 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.

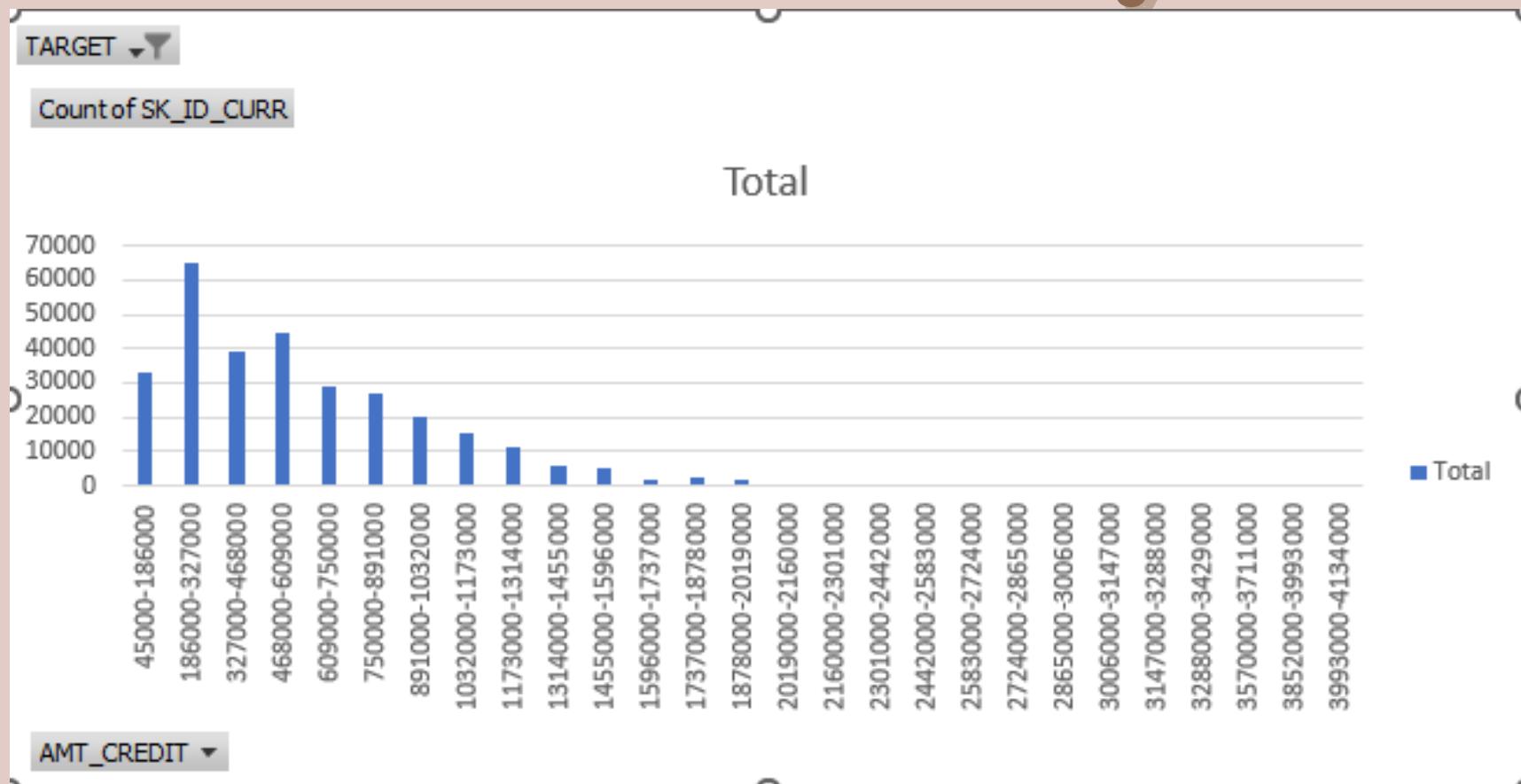


Analysis and Insights 5

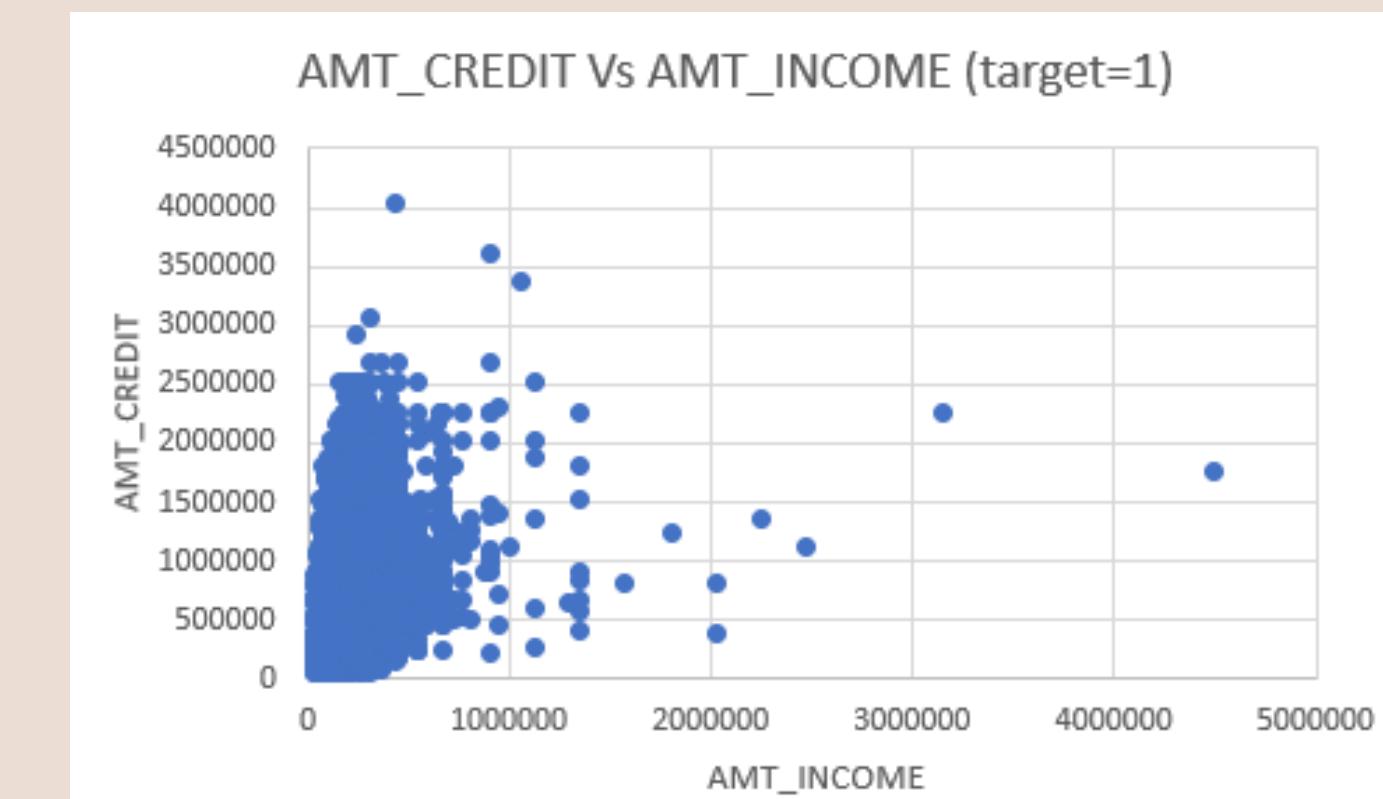
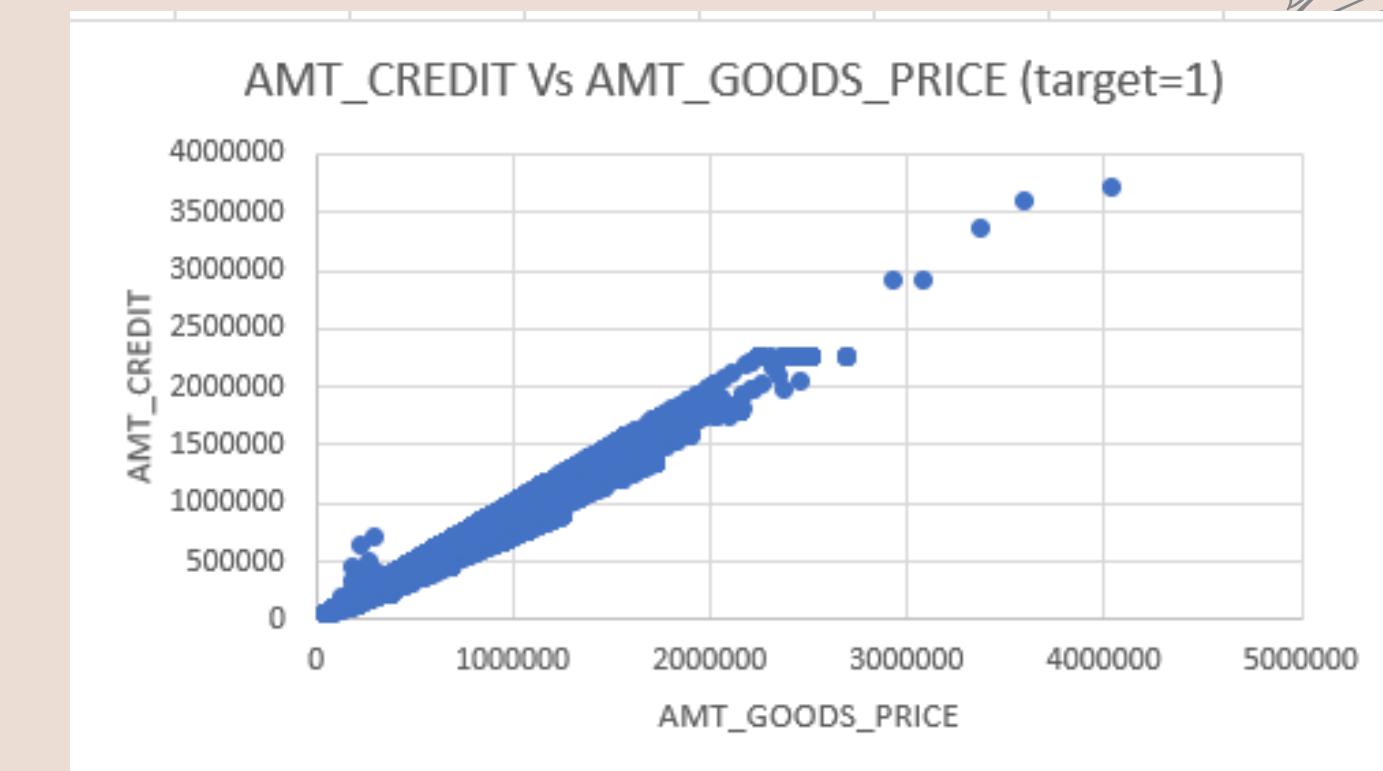


0 depicts the loan repayer and 1 defaulter. From the graph 90% are loan repayers whereas 8% are defaulters. Addition to this, Female defaulters and female repayers are higher.

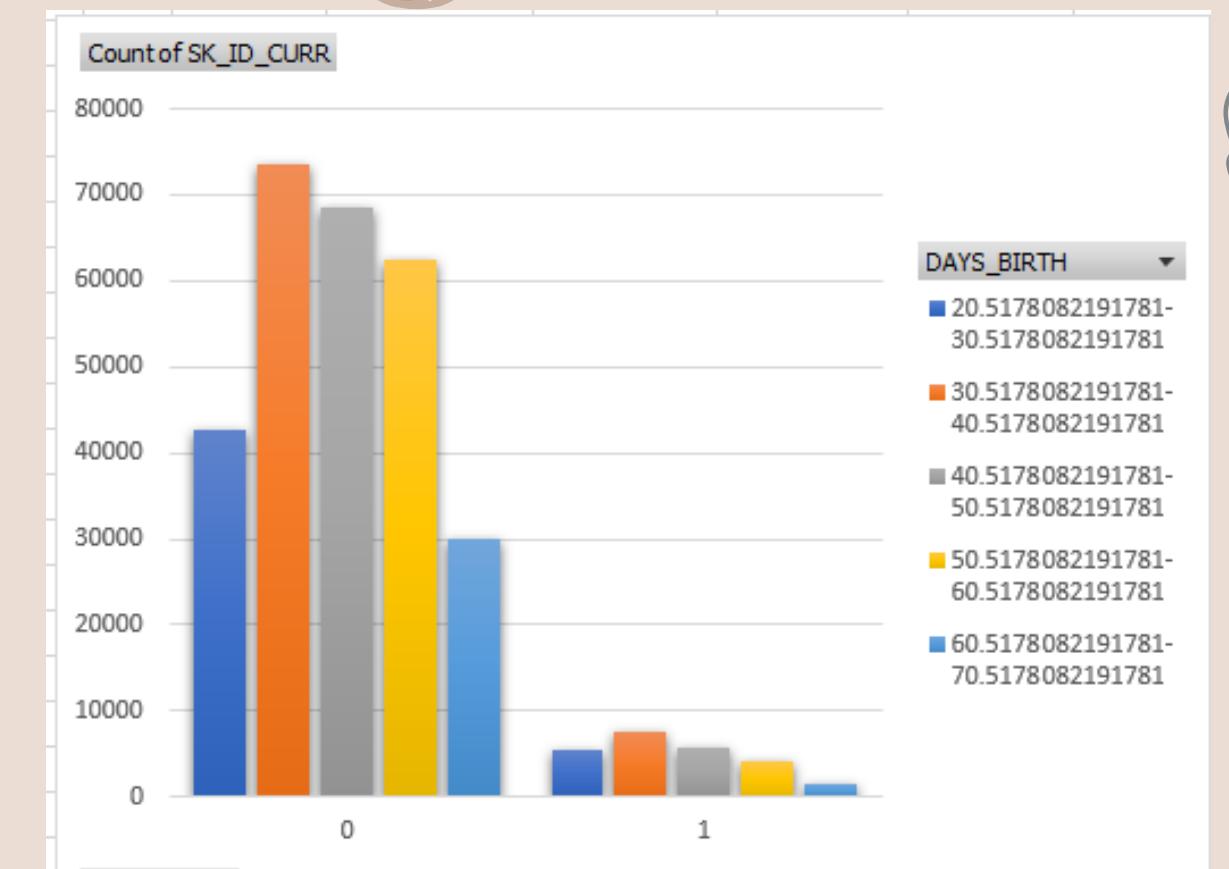
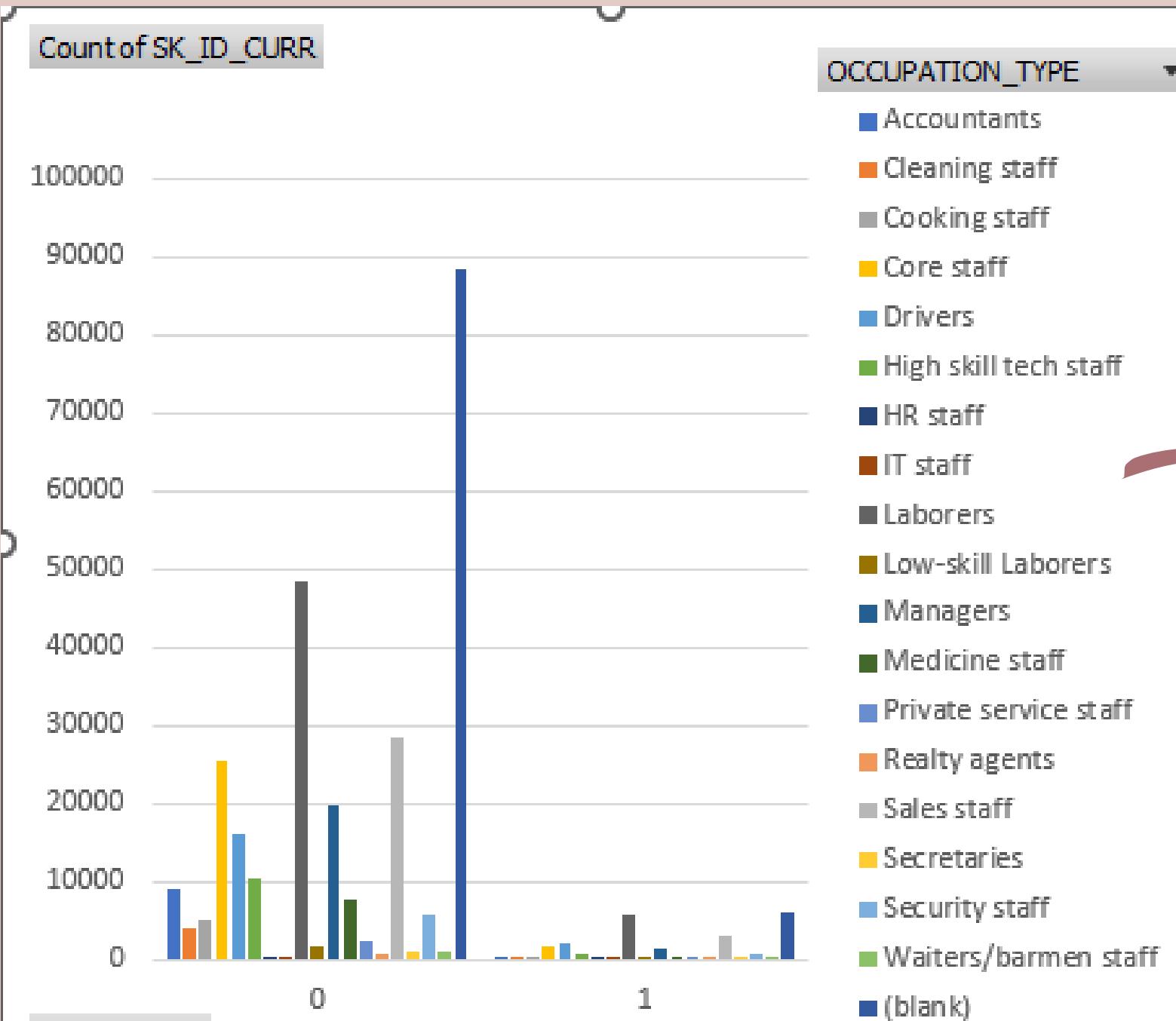
Analysis and Insights 5



People having the amount credit 186000-327000 are higher. Also if the amount credit increases, goods price increases and amount income decreases

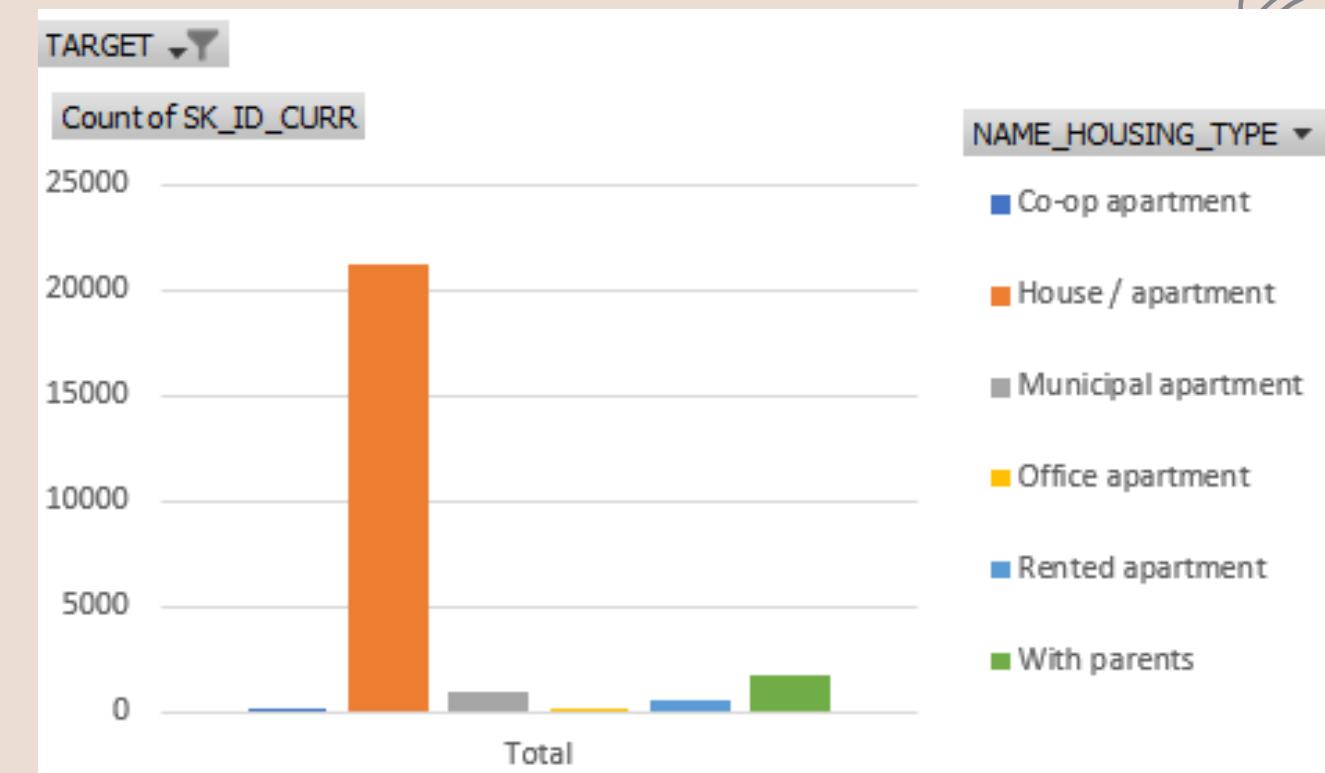
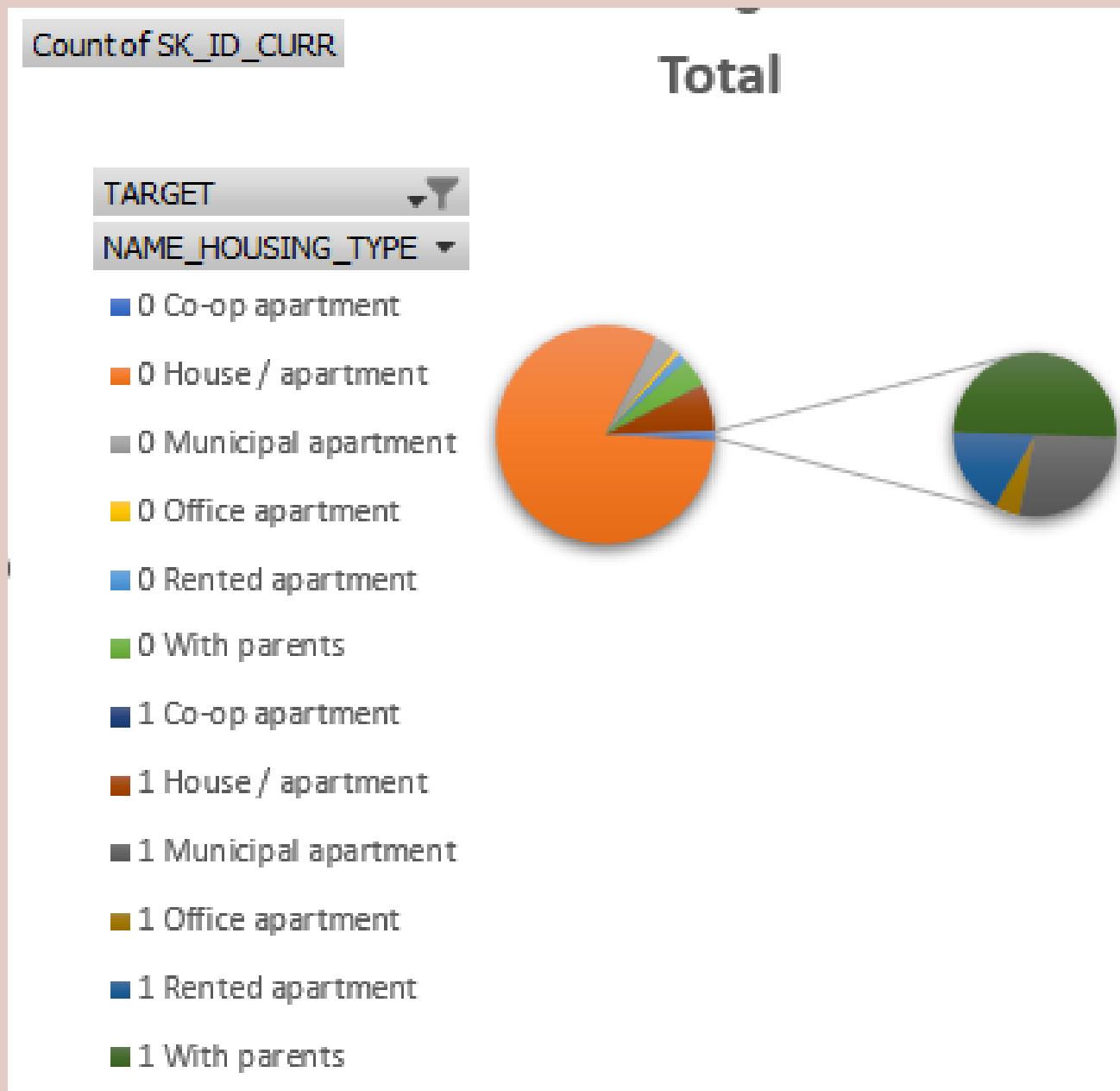


Analysis and Insights 5



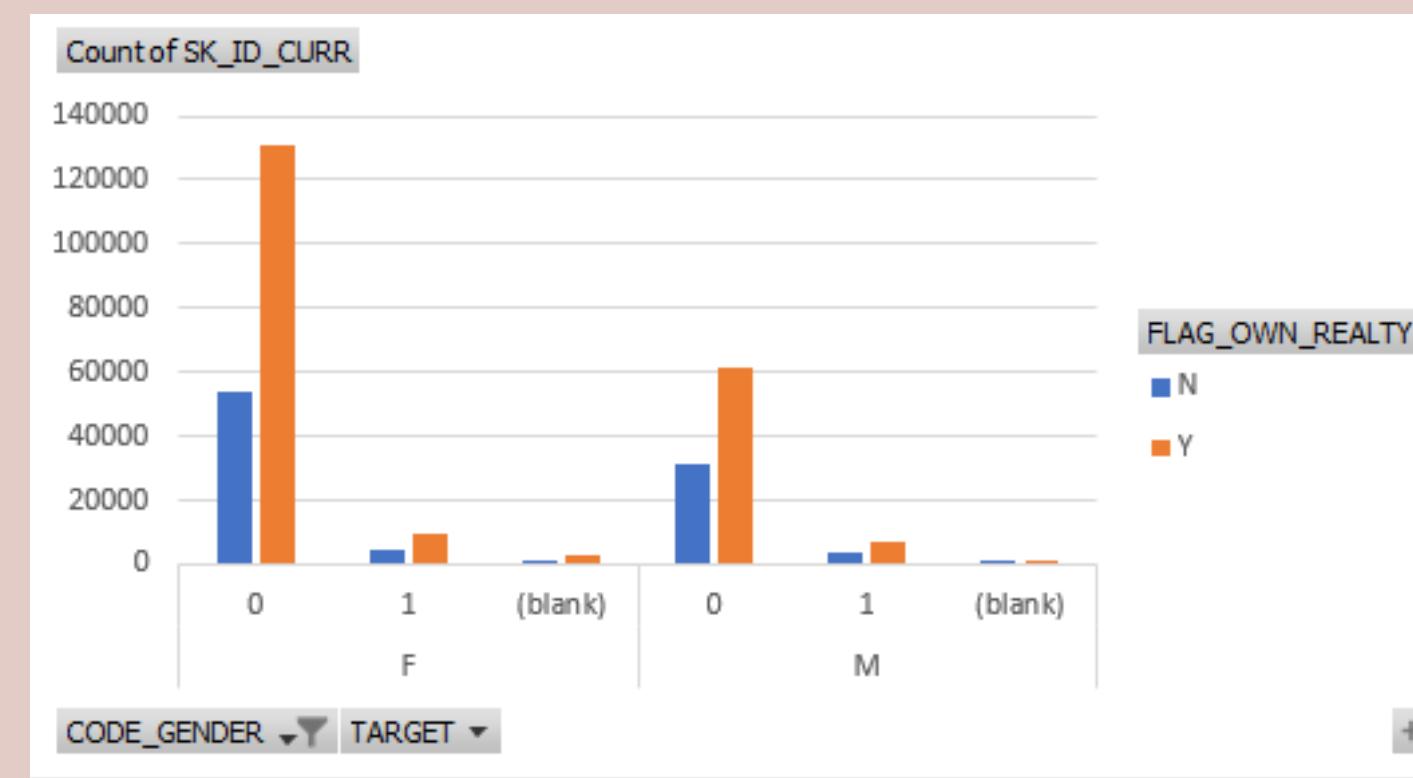
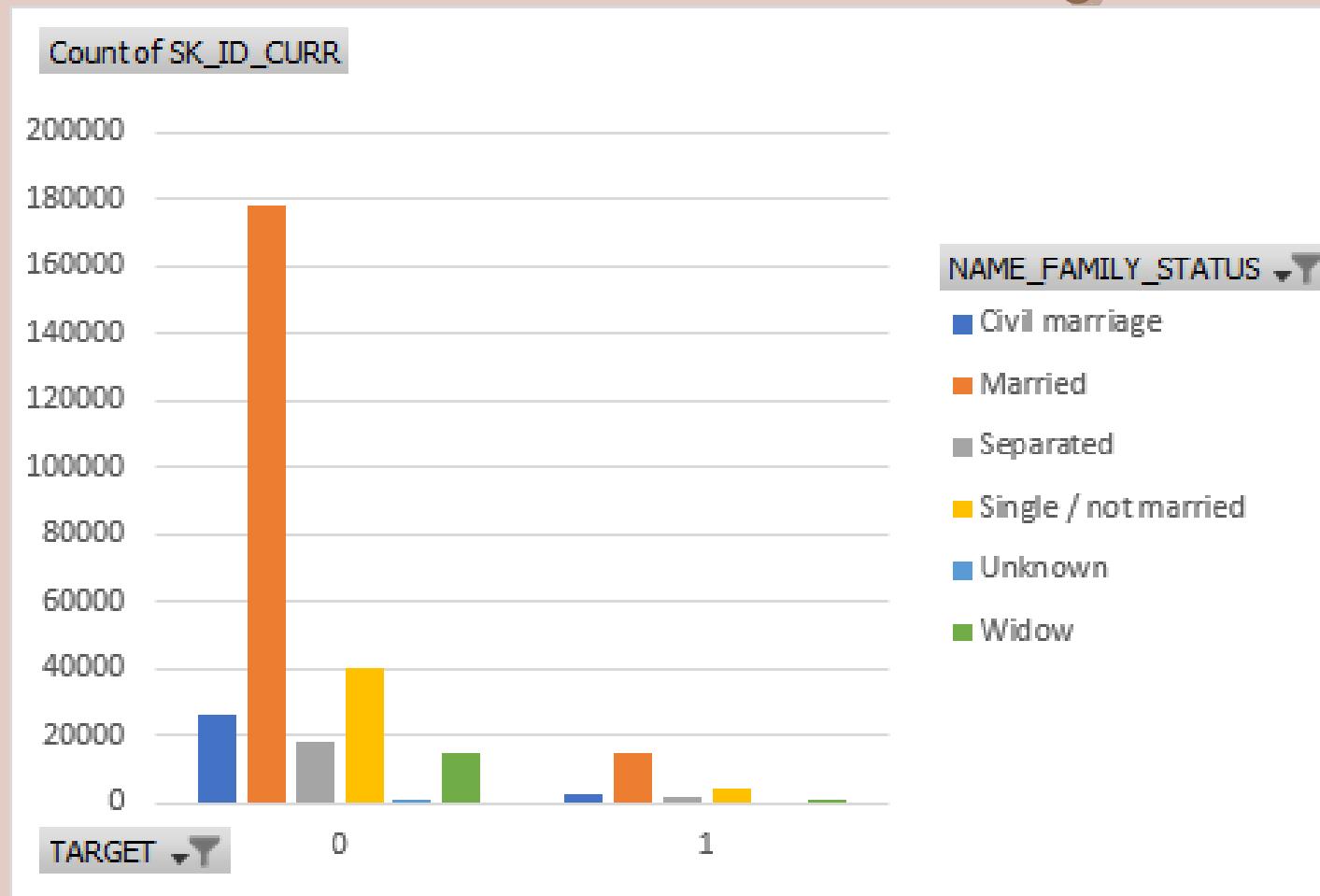
The age group 30-40 years tends to take more loans. And the maximum loan appliers are labourers, medical staff, core staff, sales staff.

Analysis and Insights 5



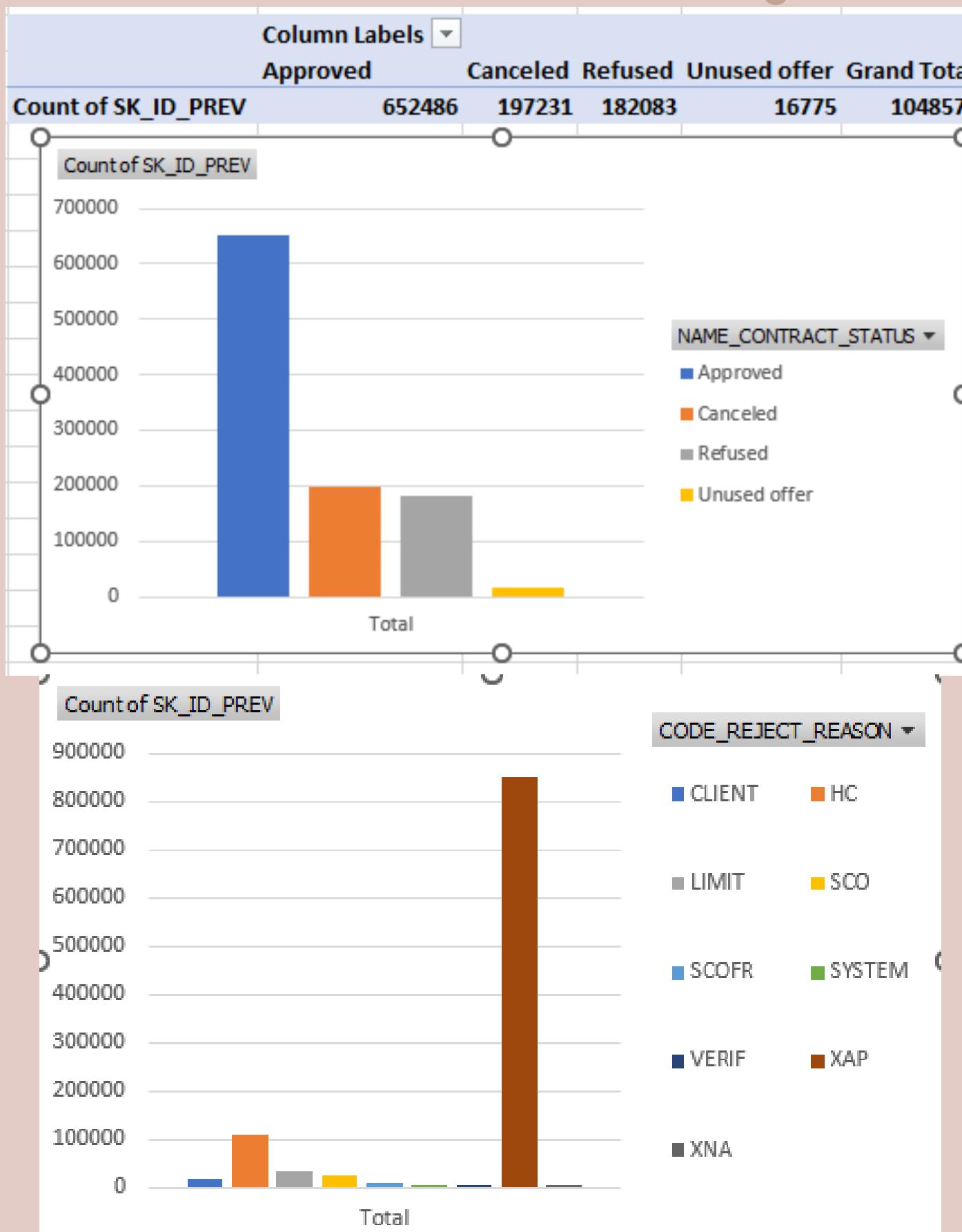
Maximum
Defaulters and
maximum
Repayers both
stay in
apartment/house.

Analysis and Insights 5



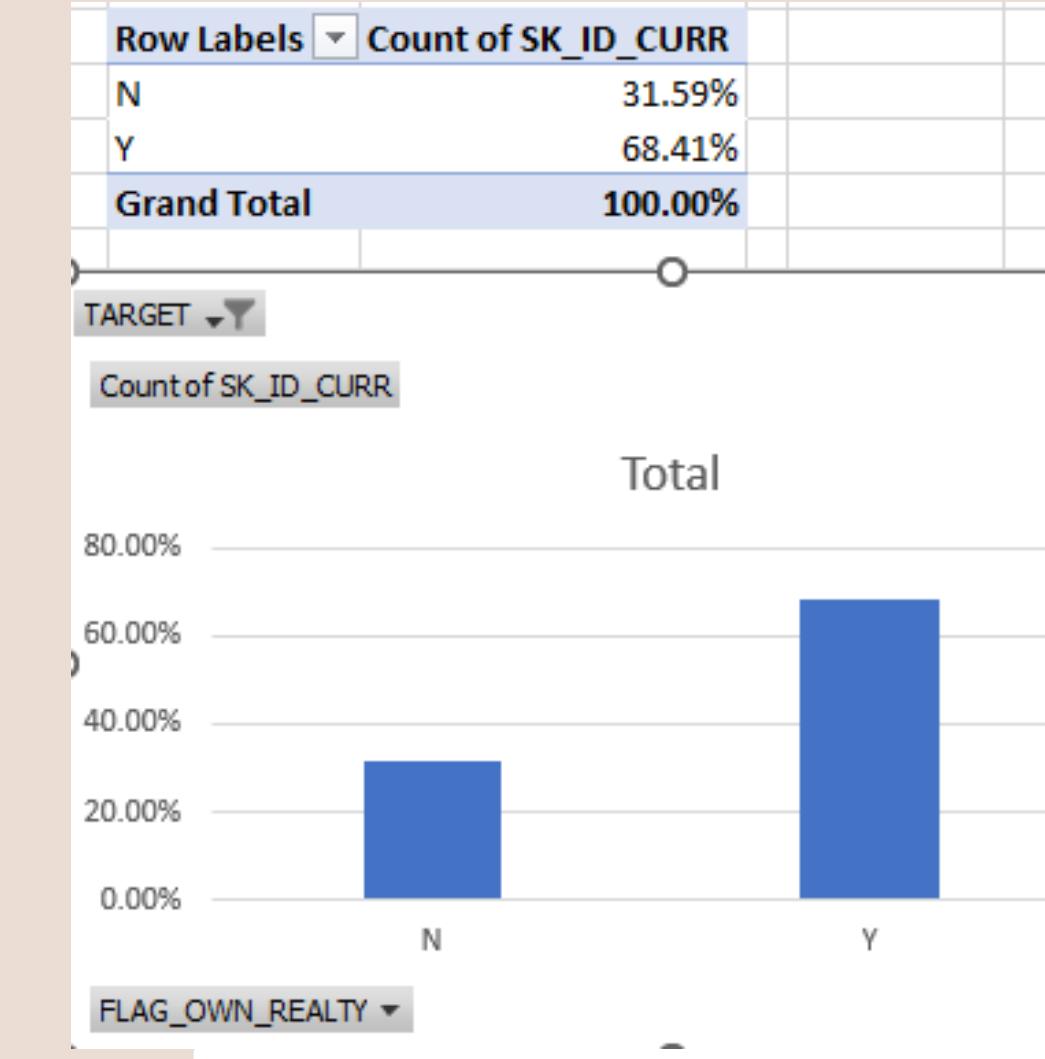
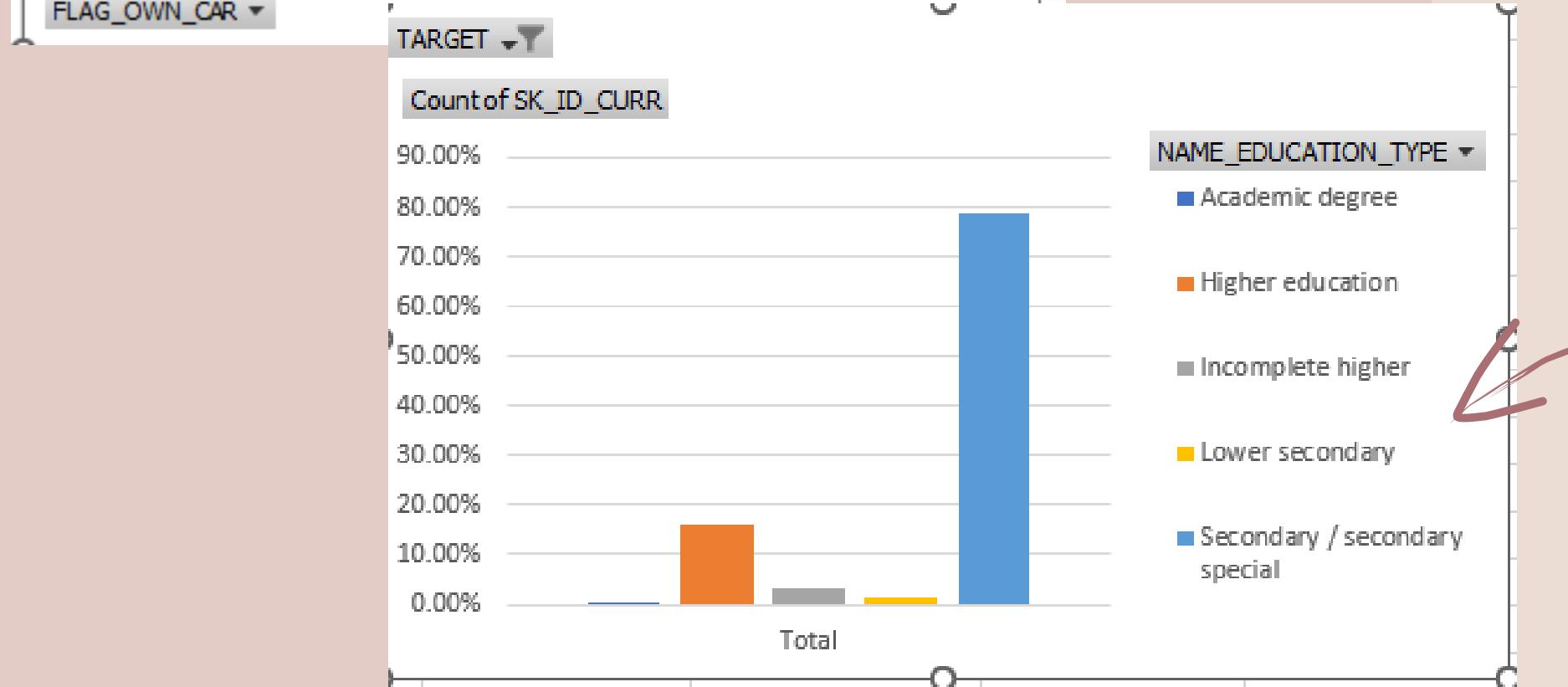
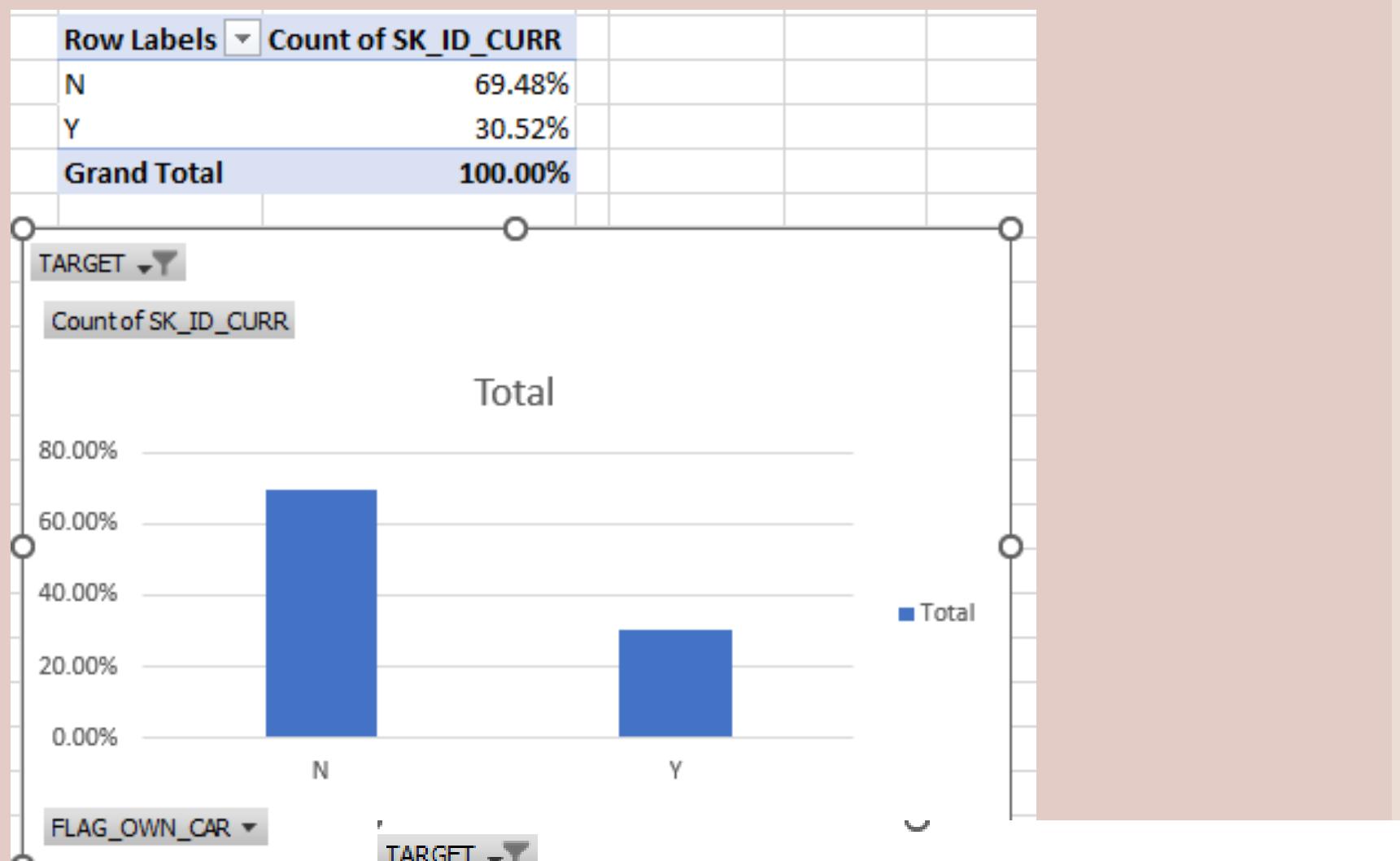
Married people tend to take more loans.
Maximum female and male repayers and defaulters own a reality.

Analysis and Insights 5

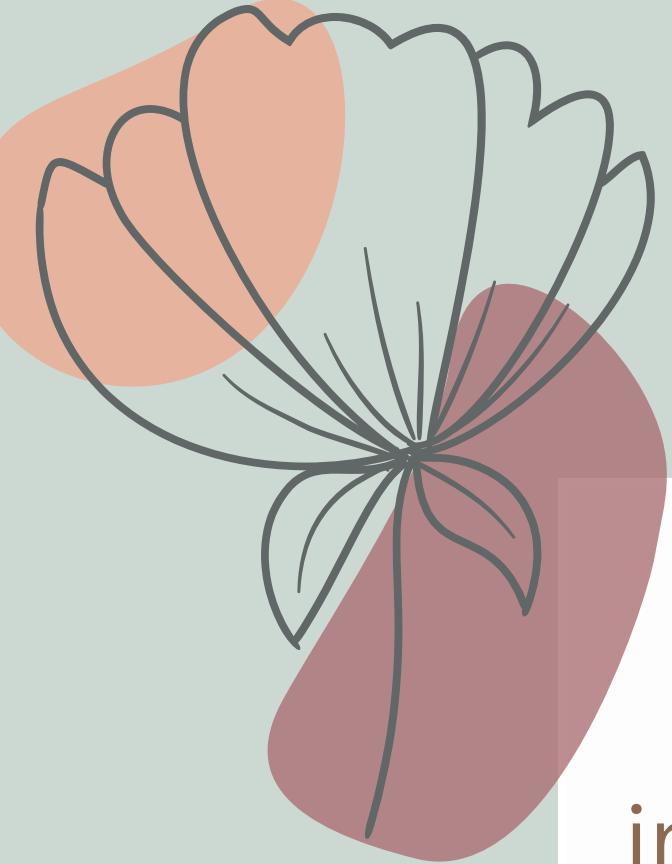


Maximum loans are approved and some of the loans are rejected and the reason for that is XAP

Analysis and Insights 5



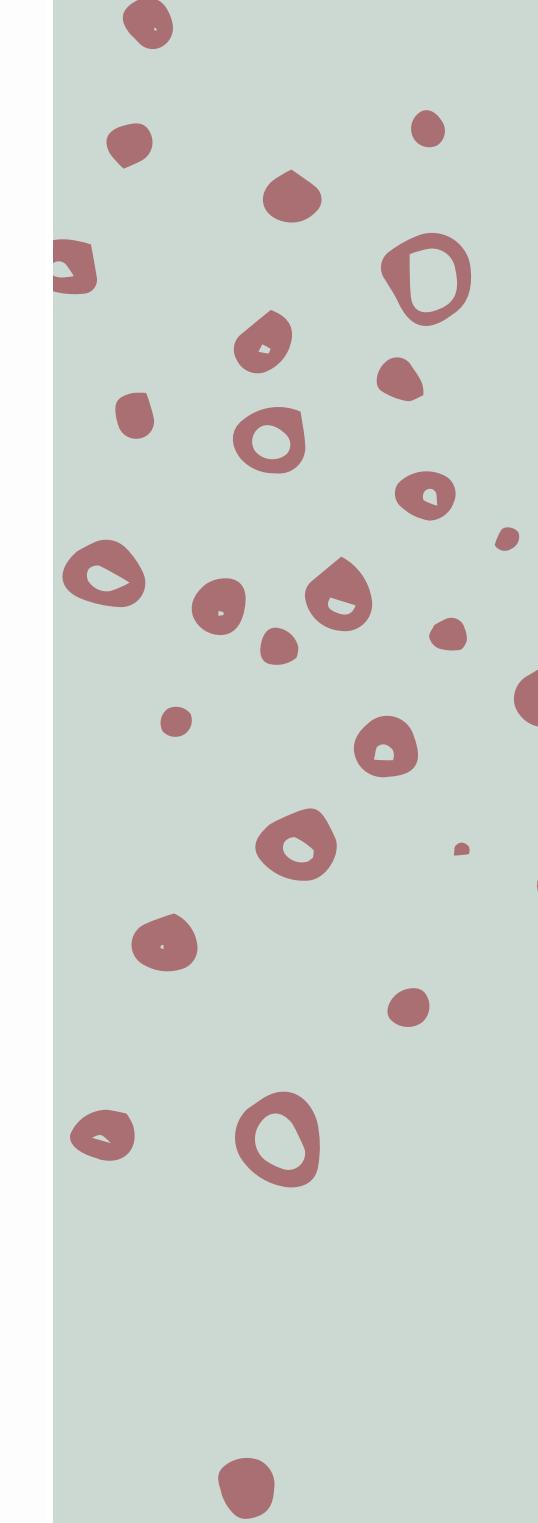
69.48% of defaulters do not own a car whereas 68.41% defaulters own a reality. People having secondary education tend to default more.



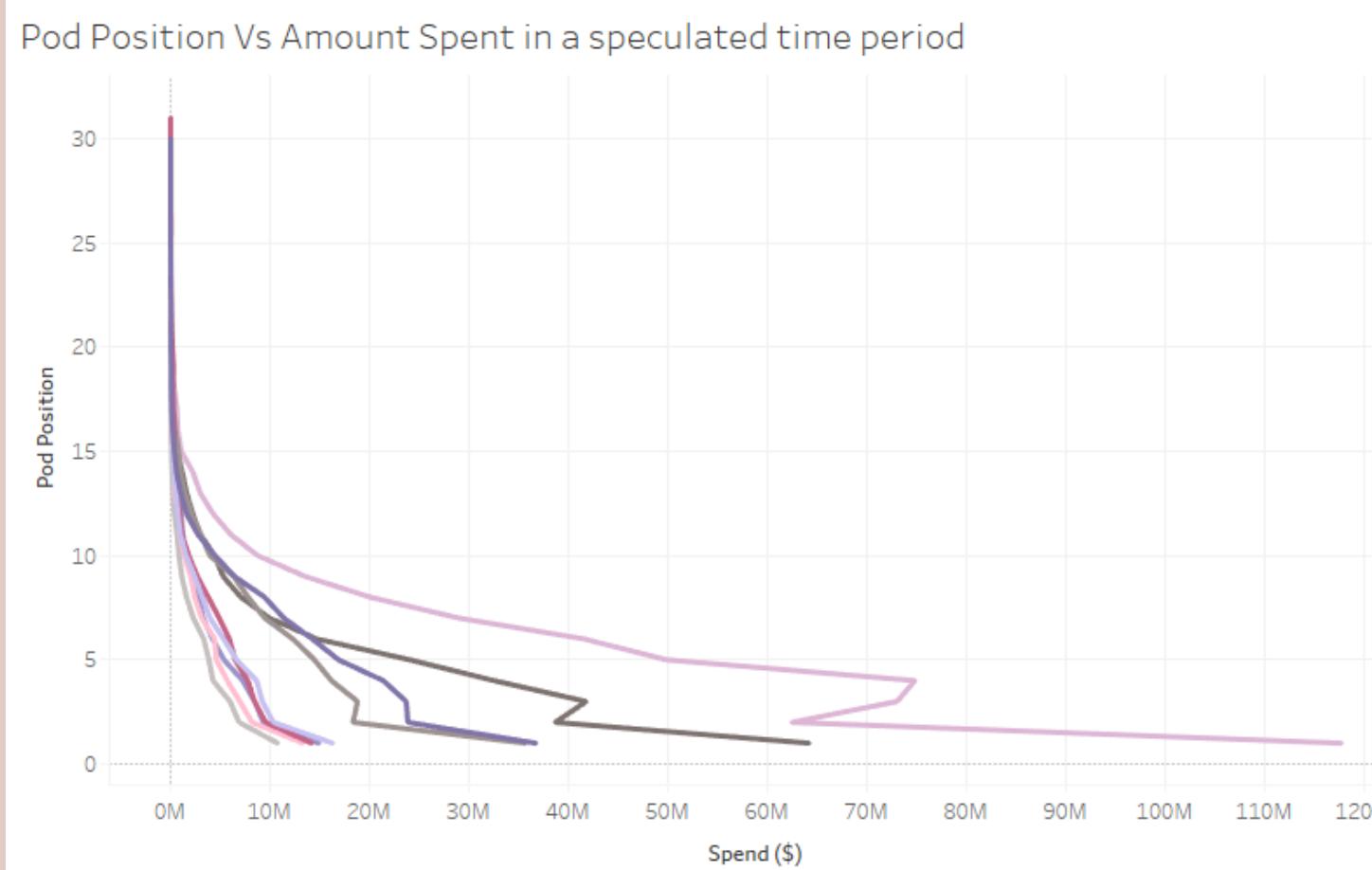
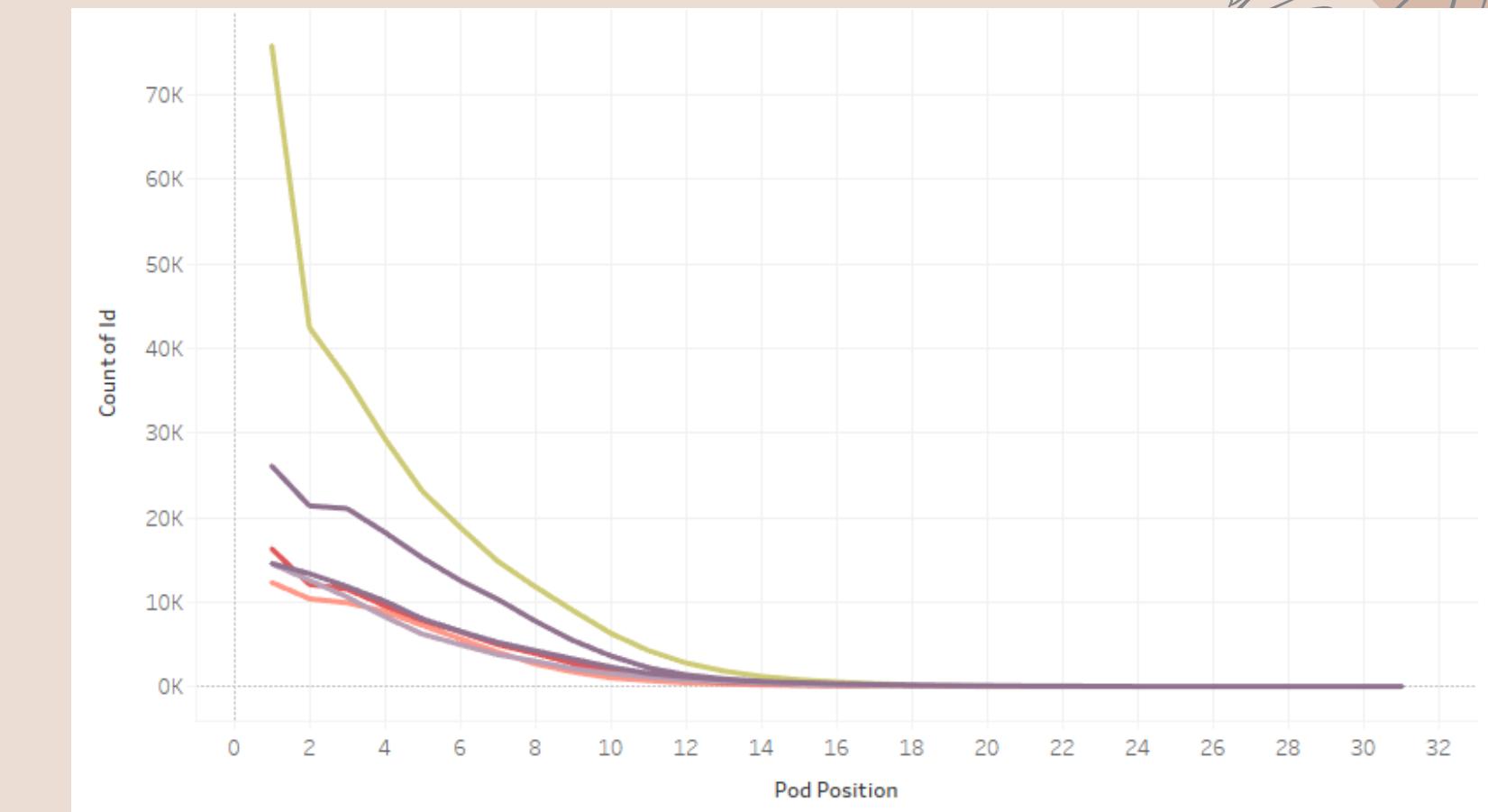
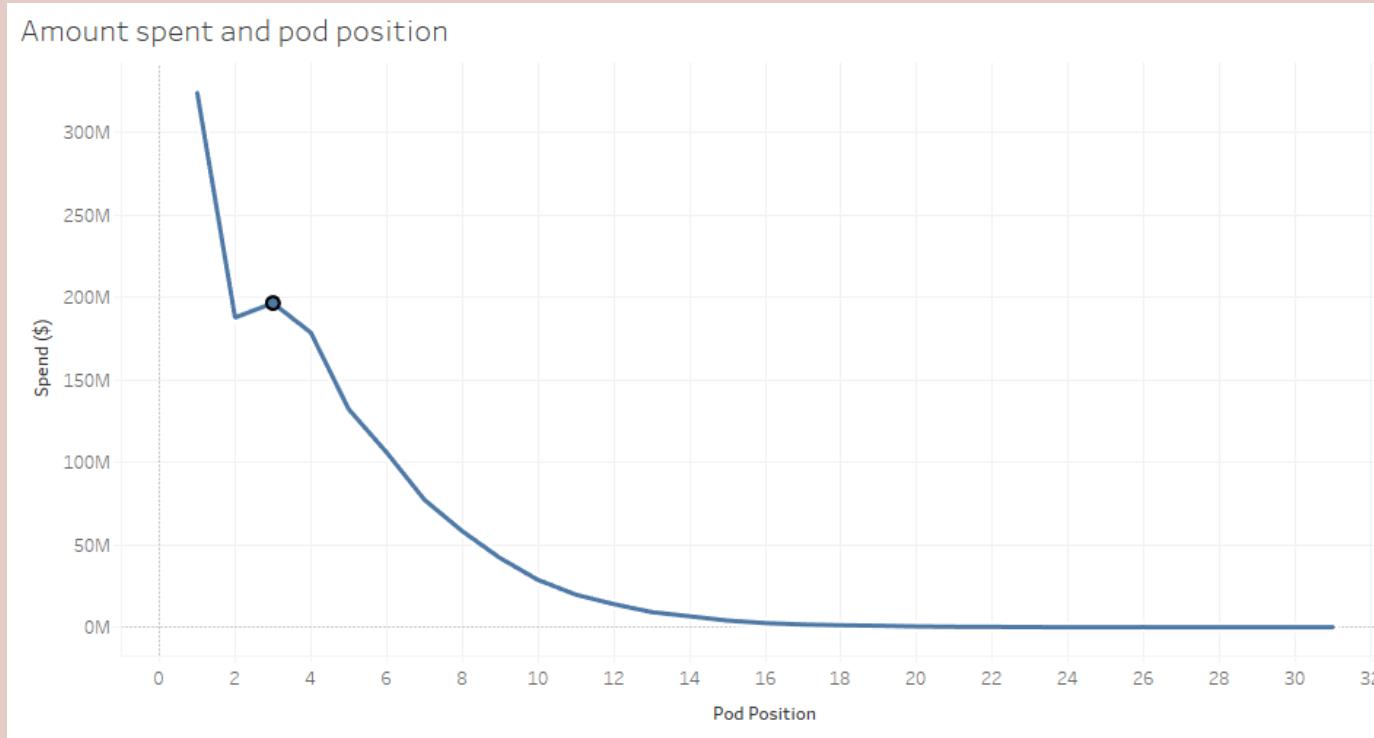
Project 6- XYZ Ads Airing Report

Description

Advertising is a way of marketing your business in order to increase sales or make your audience aware of your products or services. 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. Target audience for businesses could be local, regional, national or international or a mixture. So they use different ways for advertisement.

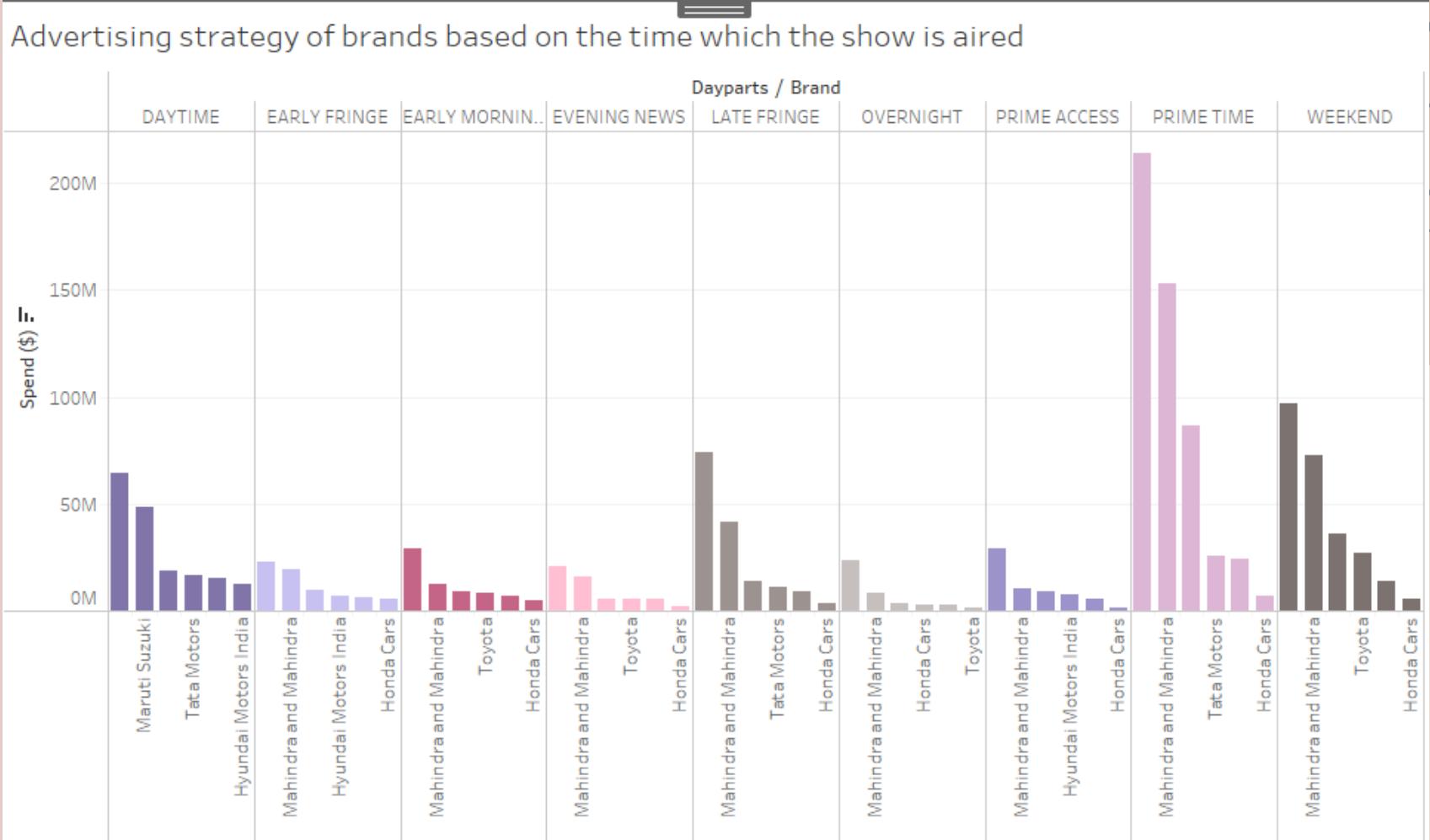
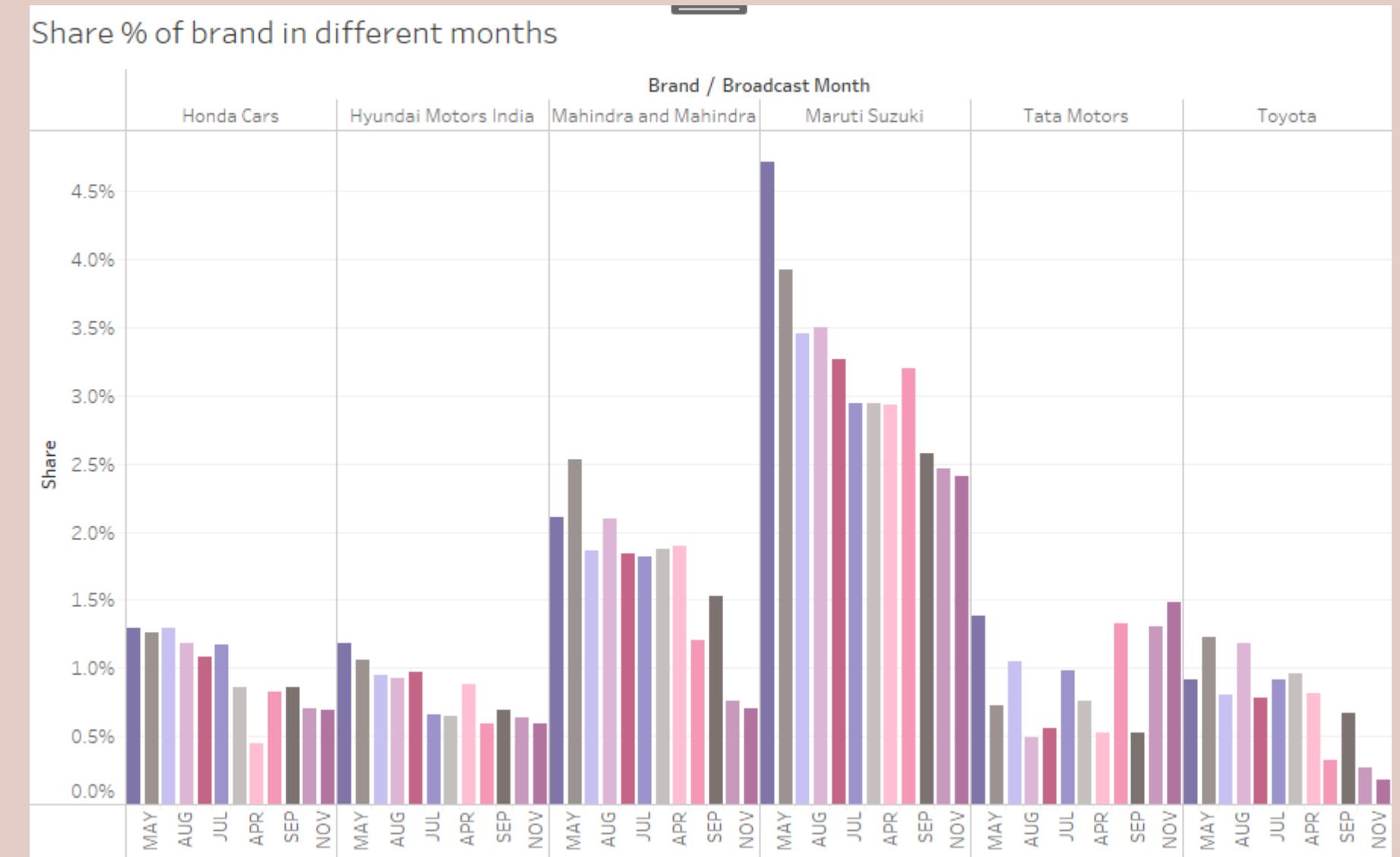


Analysis and Insights 6



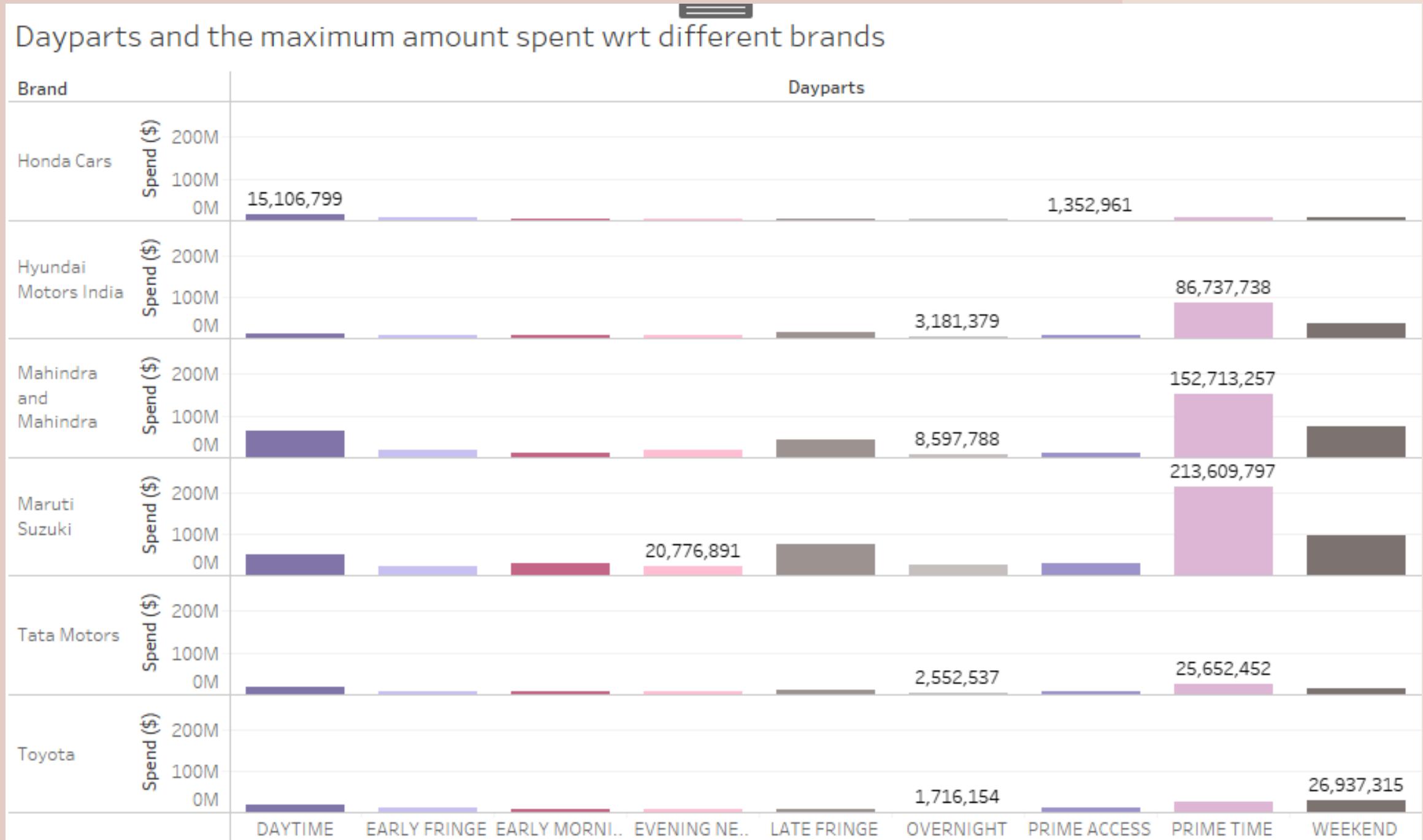
Higher is the amount spent, less is the pod position depicted in the analysis. Also maximum ads are aired when the pod position is the least

Analysis and Insights 6



Maruti Suzuki has the highest share of all the brands. The maximum ads are aired during the prime time.

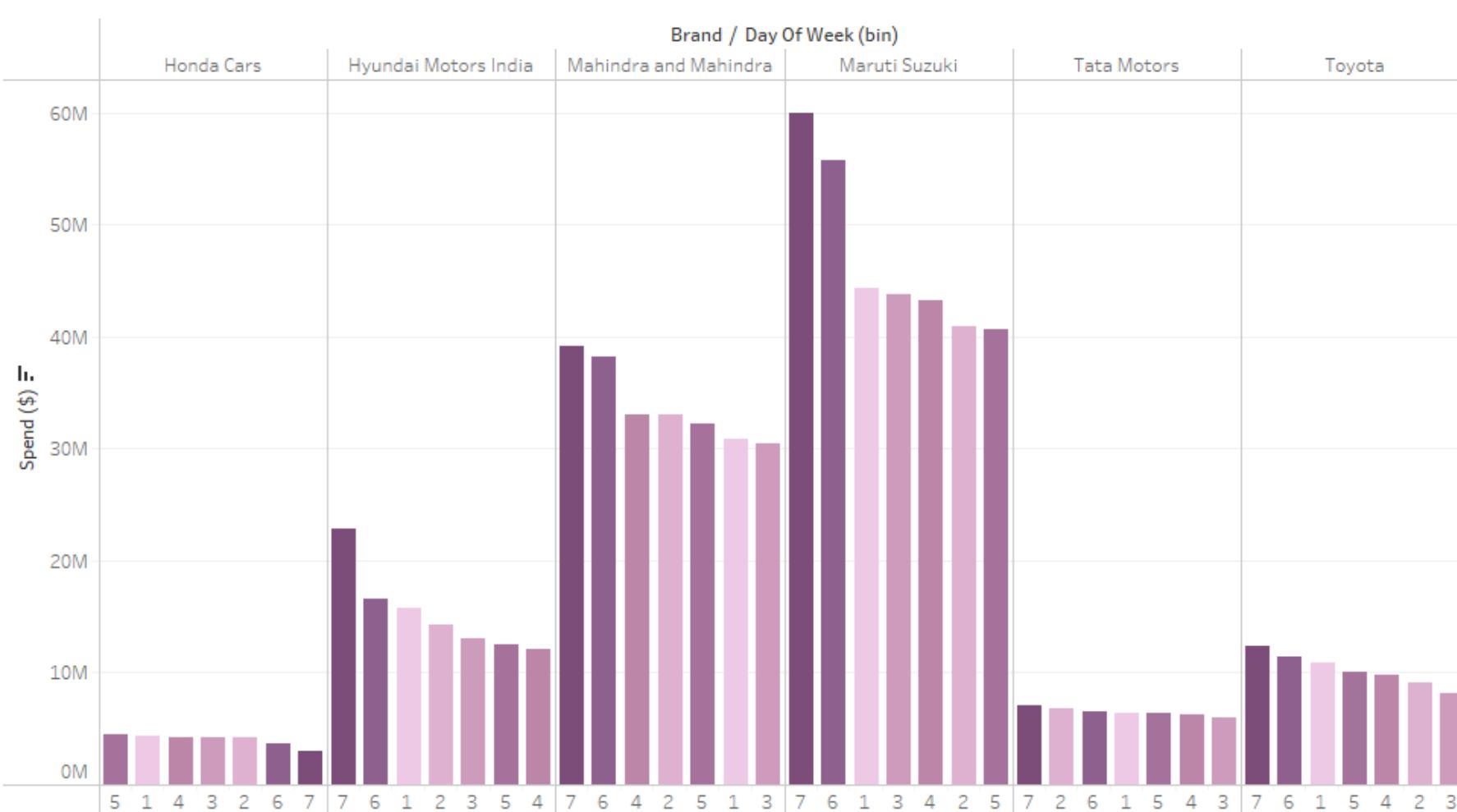
Analysis and Insights 6



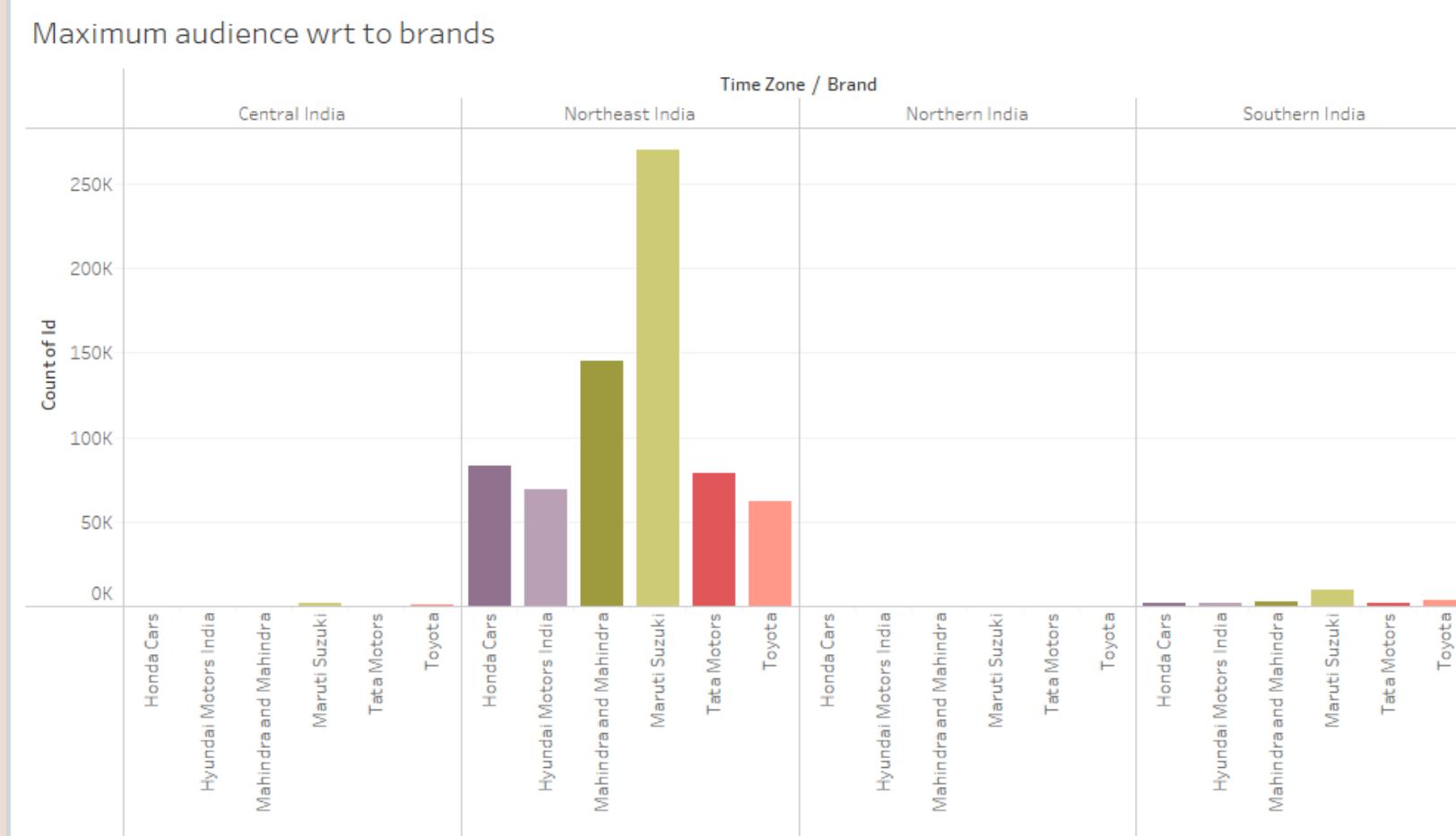
In the previous analysis, we concluded that maximum ads are aired during prime time. But the advertising strategy for Honda and Toyota is different. Honda cars spend the maximum in airing ads at daytime, whereas Toyota motors in the weekend

Analysis and Insights 6

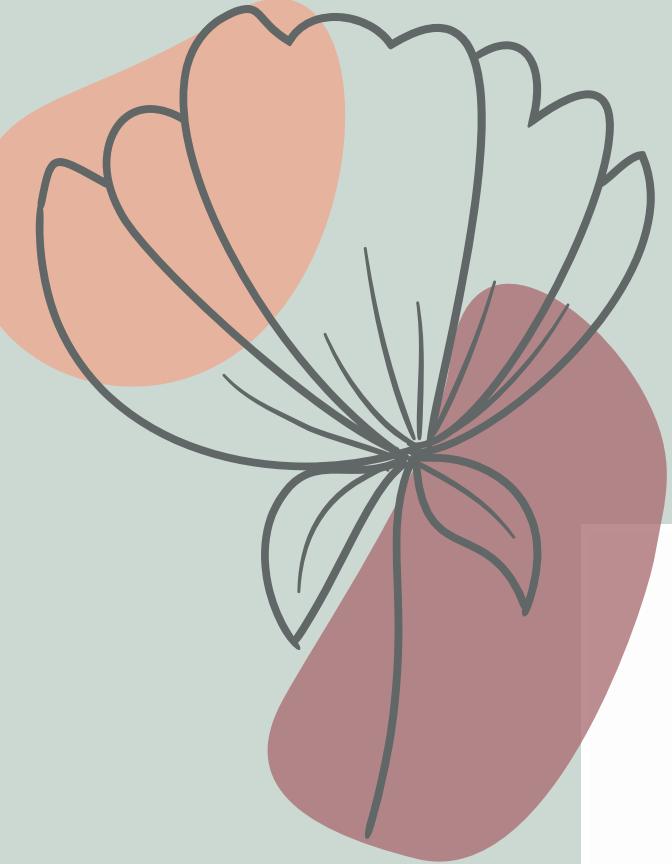
At what day of the week does different brands spend max and min money on ads



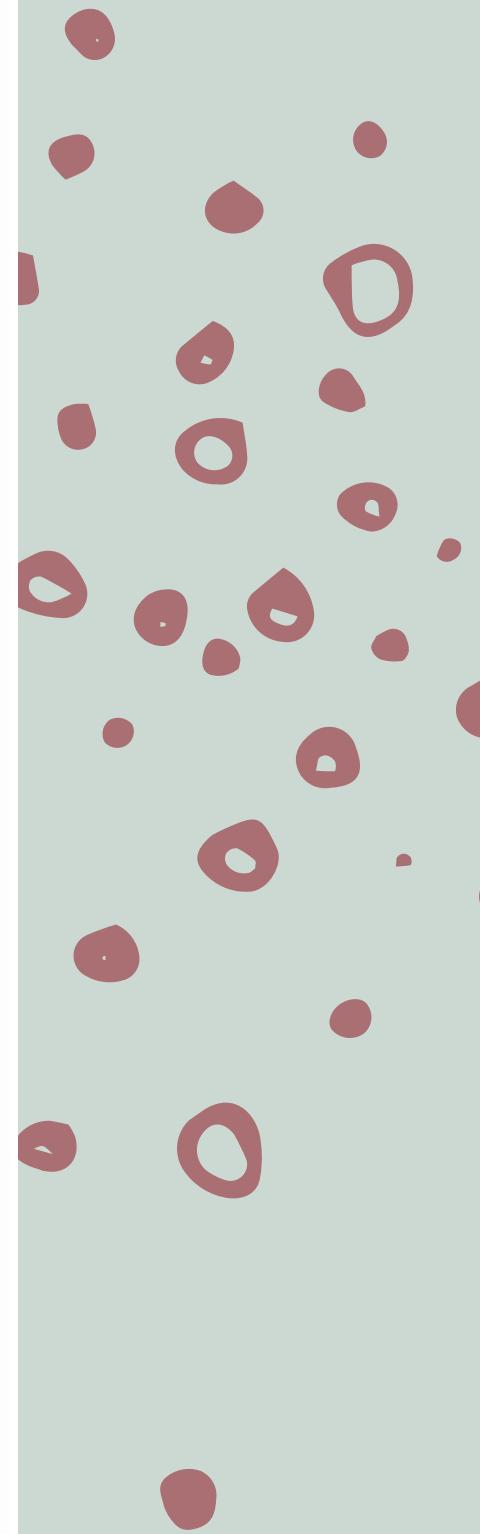
Most of the brands spend the maximum at day 6 and day 7 of the week



Northeast India has the highest audience of viewing the ads and North India has the least ad viewing statistics



Project 7- ABC Call Volume Trend Analysis



Description

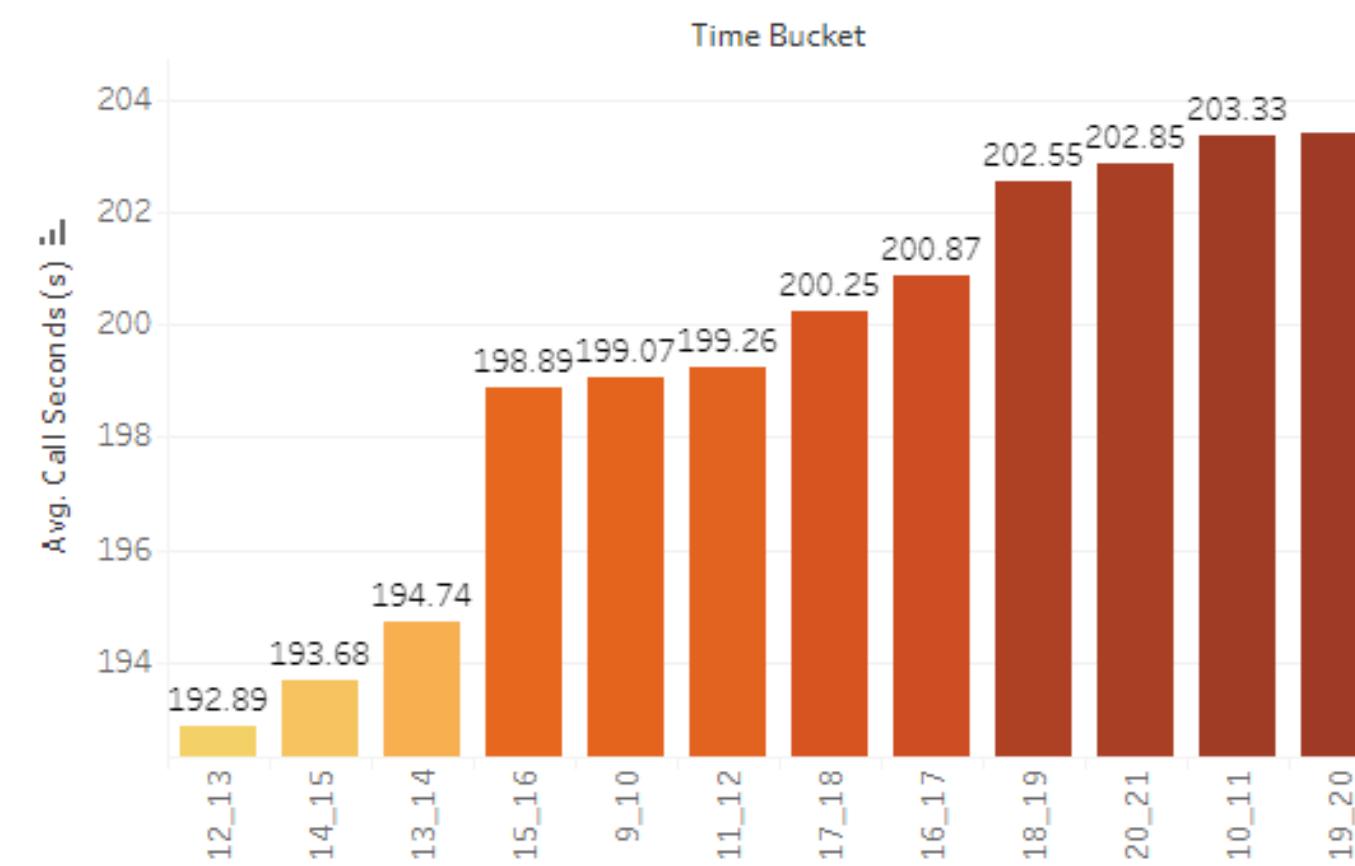
A customer experience (CX) team consists of professionals who analyze customer feedback and data, and share insights with the rest of the organization. Typically, these teams fulfil various roles and responsibilities such as: Customer experience programs (CX programs), Digital customer experience, Design 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.

Analysis and Insights 7

Average Seconds of call answered in a time bucket

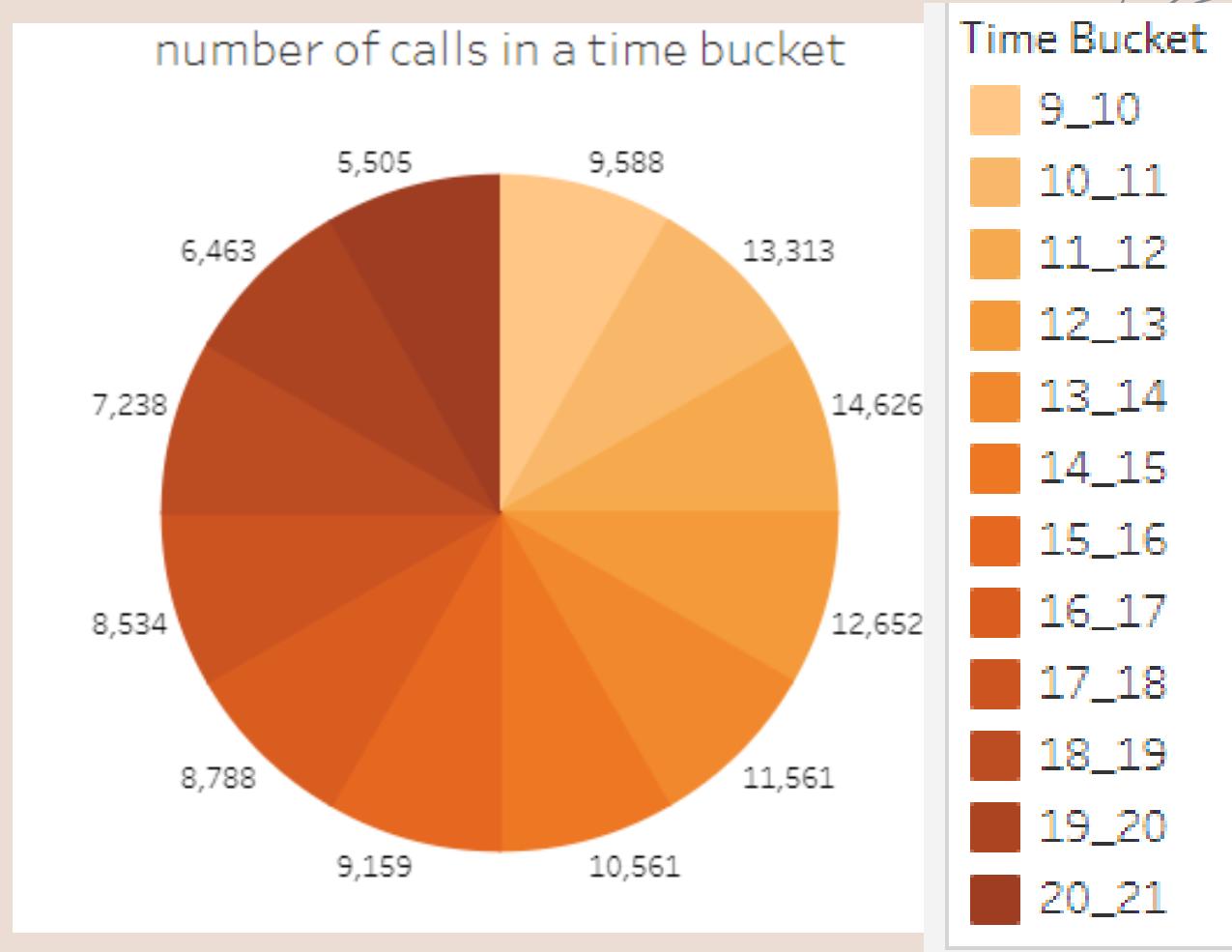
Time Bucket	Avg. Call Seconds (s)
9_10	199.07
10_11	203.33
11_12	199.26
12_13	192.89
13_14	194.74
14_15	193.68
15_16	198.89
16_17	200.87
17_18	200.25
18_19	202.55
19_20	203.41
20_21	202.85

Average Call time duration in each time bucket



Maximum calls attended are in the time bucket 10-11 am and 7-8pm. And the average call time duration is 198.6 seconds.

number of calls in a time bucket



Maximum incoming calls are during 11-12 noon and minimum incoming incoming calls are during 8-9pm

Analysis and Insights 7

III. 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%.

The total working hours of the agent lunch and snack break		9 hrs
Working hours	1.5 hrs	
The employee works	7.5 hrs	
	60% of the working hours	
	60% of 7.5 hrs	
	4.5 hrs	
Date & T	1/1/2022	

Row Labels	Count of numbers	Sum of Call_Seconds (s)	Sum of hours
abandon	684	0	0
answered	3883	670279	186.1886111
transfer	77	6385	1.773611111
Grand Total	4644	676664	187.9622222

The total call hours in a day are 187.96 hours and each employee works 4.5 hours to attend these calls

Therefore number of employees required are

$187.96/(4.5) = 41.76938272$ employees in a day are currently working

the calls answered with 42 employees are 60%

then for 90% what is the employee required

$41.76938272 \times 1.5 = 62.65407407$

? $\times 1.5 = 90\%$

$62.65407407 \times 1.5 = 93.98110505$

that is approx 63 employees are required

(For 100% call_seconds 63 employees are required then for x% calls in each time bucket:- $x * 63$



(Percentage of call duration in seconds)			
Row Labels	x% call seconds	x% * 63	number of employees in each time bucket
10_11	7.8783%	4.963	5
11_12	10.3752%	6.536	7
12_13	11.1222%	7.007	7
13_14	10.5013%	6.616	7
14_15	9.4280%	5.940	6
15_16	9.4519%	5.955	6
16_17	9.6852%	6.102	6
17_18	9.3164%	5.869	6
18_19	7.6642%	4.828	5
19_20	5.6759%	3.576	4
20_21	3.5428%	2.232	2
9_10	5.3586%	3.376	3
Grand Total	100.00%	63	

Analysis and Insights 7

Manpower plan for night

Row Labels	Count of numbers	Sum of hours
1/10/202	4983	216.3163889
1/11/202	4637	217.2547222
1/12/202	4643	197.2038889
1/13/202	4123	192.0333333
1/14/202	3155	156.7297222
1/15/202	3058	154.5186111
1/16/202	5142	187.3316667
1/17/202	22347	262.6708333
1/18/202	5774	221.3244444
1/19/202	4703	208.4083333
1/20/202	4322	211.0036111
1/21/202	3675	177.7375
1/22/202	3291	172.6602778
1/23/202	3225	153.8608333
1/1/2022	4644	187.9622222
1/2/2022	3351	159.4452778
1/3/2022	4789	225.7952778
1/4/2022	5113	239.4294444
1/5/2022	4790	235.2216667
1/6/2022	4951	230.2888889
1/7/2022	4948	210.2830556
1/8/2022	4672	204.29
1/9/2022	3652	150.3186111
(blank)		
Grand Total	117988	4572.088611
Average calls	5129.913043	198.7864614

Average number of calls in a day 5130
For 100 calls at day, there are 30 calls at night, hence for 5130 cal
100 30
5130 ?

1539 calls at night

Therefore additional agents required assuming each works for 4.5 hours in a morning shift =
 $76.483 / 4.5$

16.99624245 agents are required
approx 17 additional agents are required in a night shift
The distribution of agents are given below:-



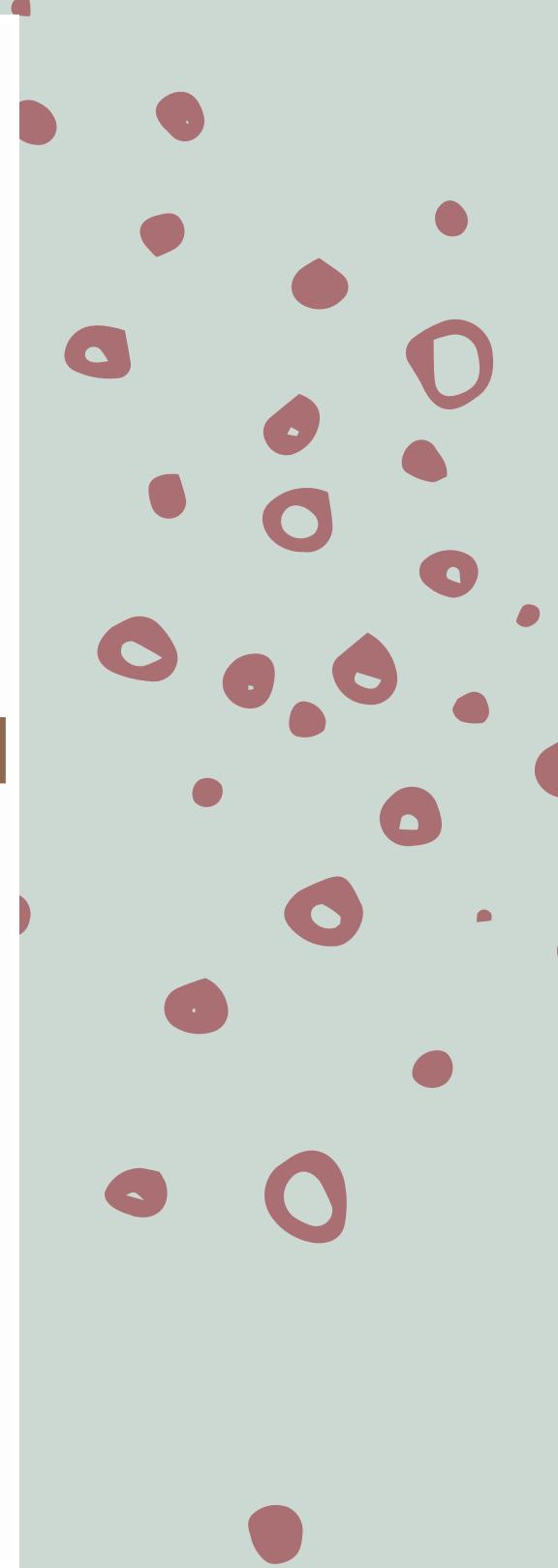
Time bucket (9pm-9am)	Time Distribution	% of time distribution	%time distribution*17	agents required
9pm-10pm	3	10.00%	1.7	2
10pm-11pm	3	10.00%	1.7	2
11pm-12pm	2	6.67%	1.133333333	1
12pm-1am	2	6.67%	1.133333333	1
1am-2am	1	3.33%	0.566666667	1
2am-3am	1	3.33%	0.566666667	1
3am-4am	1	3.33%	0.566666667	1
4am-5am	1	3.33%	0.566666667	1
5am-6am	3	10.00%	1.7	2
6am-7am	4	13.33%	2.266666667	2
7am-8am	4	13.33%	2.266666667	2
8am-9am	5	16.67%	2.833333333	3
	30			19



Project 8- Pharmaceutical Marketing and Consumption Analysis

Description

All the business strategies include analytics to improve upon their product and marketing strategies. One such project includes my internship in Kremoint Pharmaceuticals on the marketing analytics of drugs and finding why a particular drug is being sold less and what are the customer consumption statistics depending on their age and their body reaction.



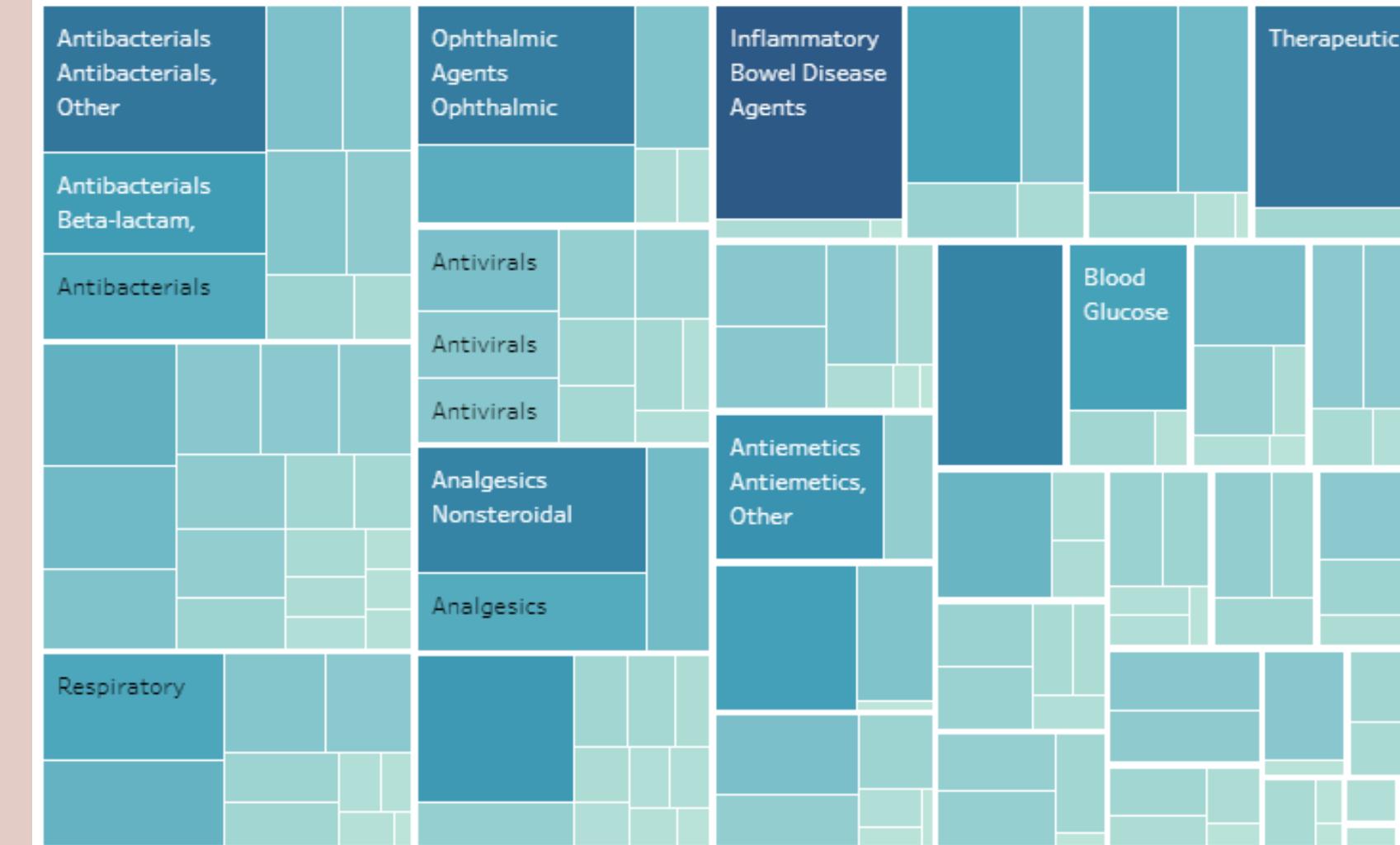
Analysis and Insights 8

highest and lowest purchased category of drugs



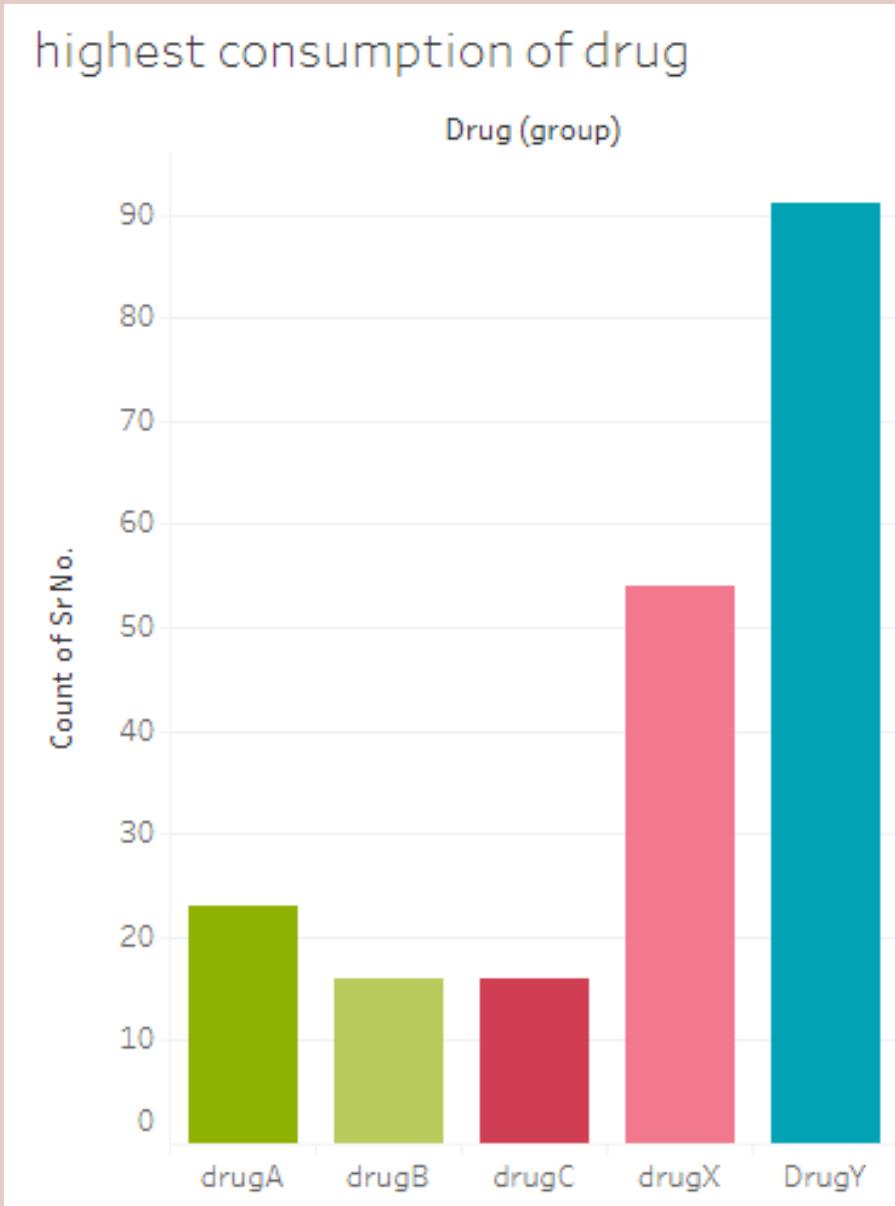
highest purchased
is antibacterials
and lowest
purchased is
antimyasthenic
agents

Category and class

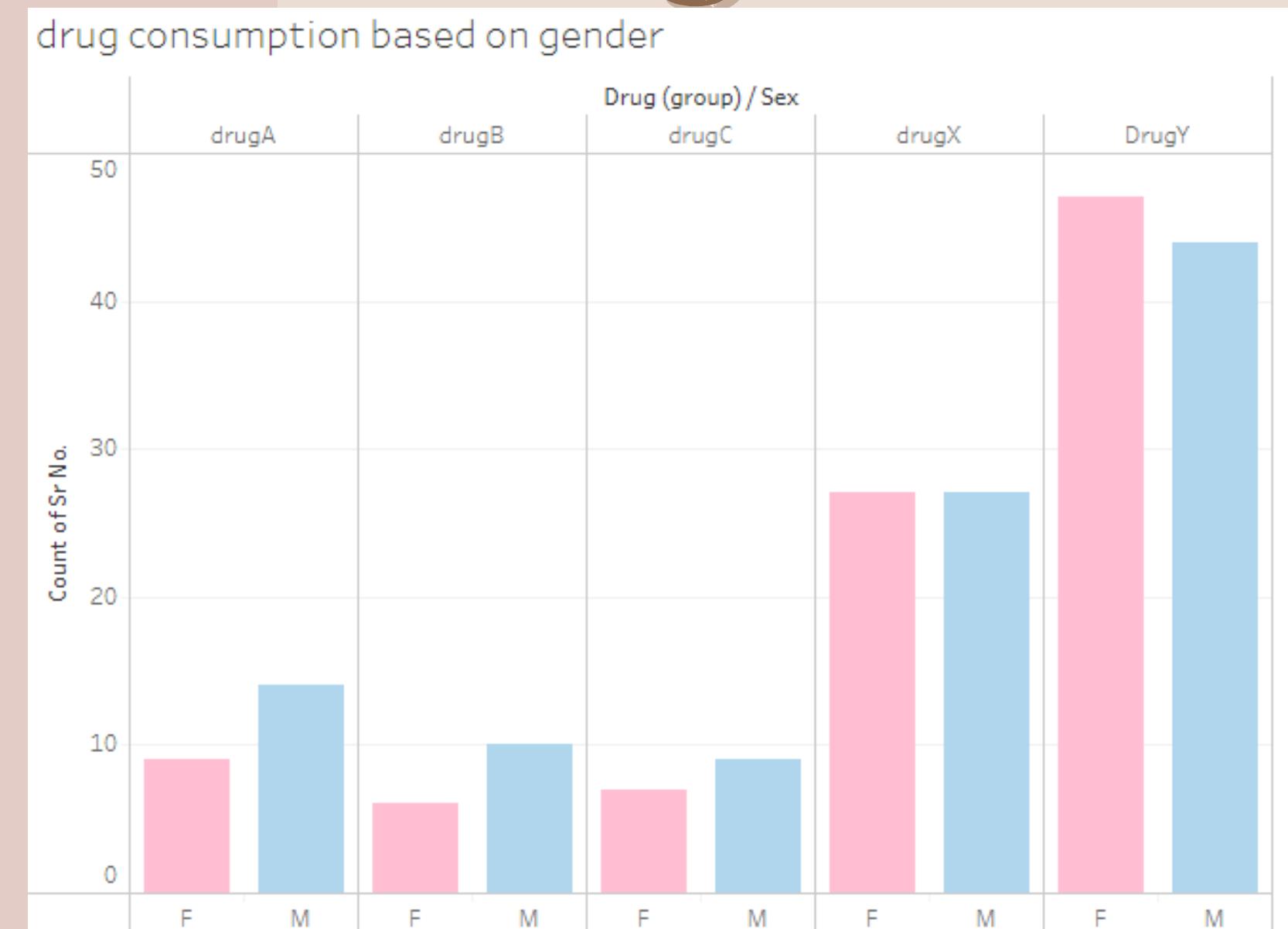


This chart gives us the idea and count of most and least purchased class and category of count along with a quantitative representation

Analysis and Insights 8



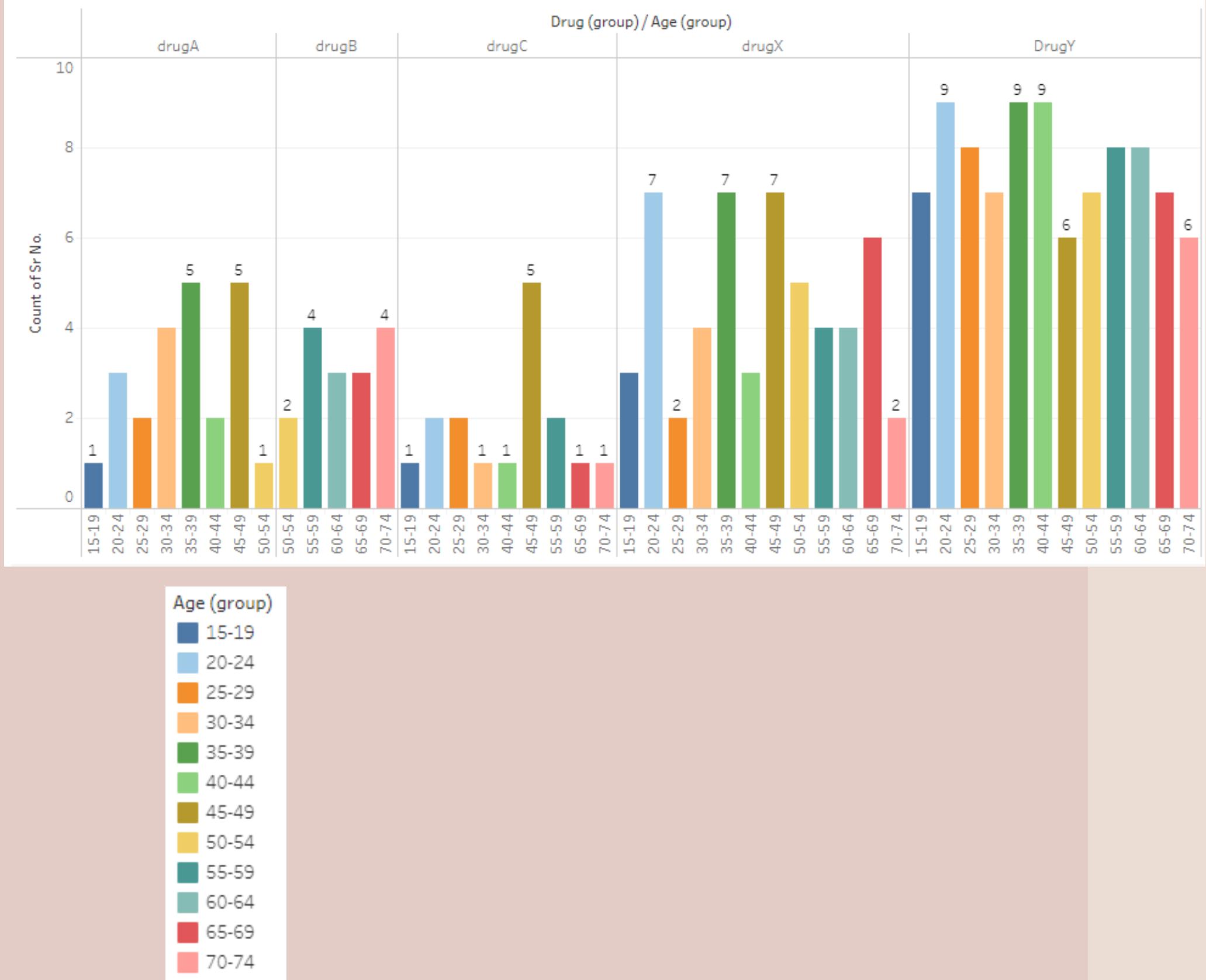
The highest consumed drug is drug Y and the least consumed is drug C



Females consume more of drug Y and males consume more of other drugs

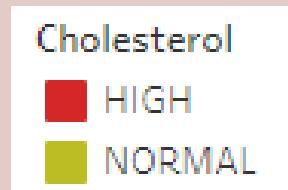
Analysis and Insights 8

Age group and drug consumption

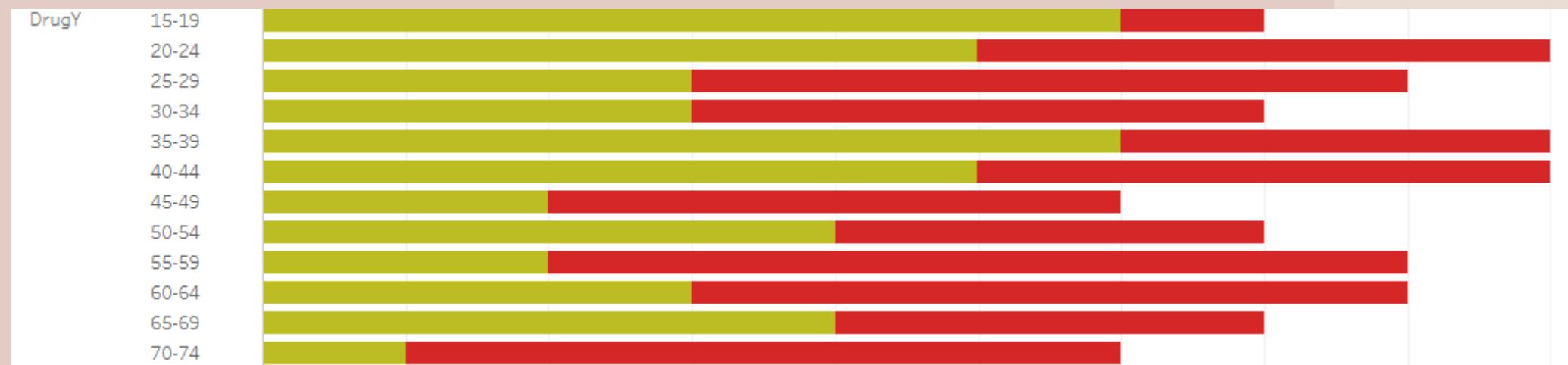


Drug Y is consumed by all the age groups whereas other drugX is more common in younger to middle aged and Drug B is more common in older aged people

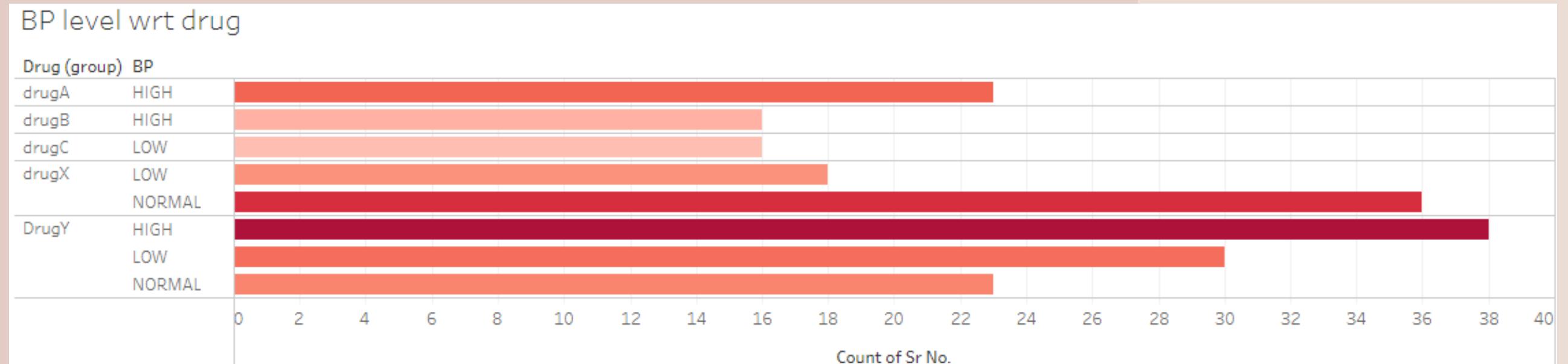
Analysis and Insights 8



Drug C is the least consumed because the cholesterol levels shoot up to very high whereas drug Y doesn't show significant changes in cholesterol level

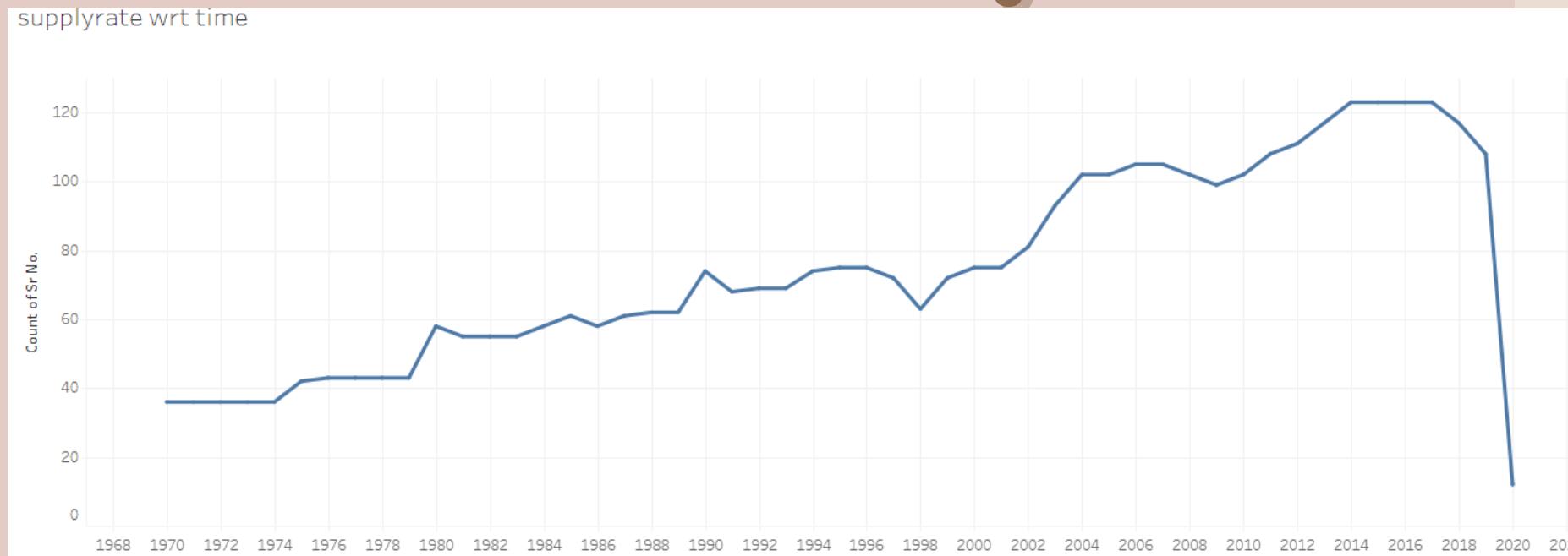


Analysis and Insights 8

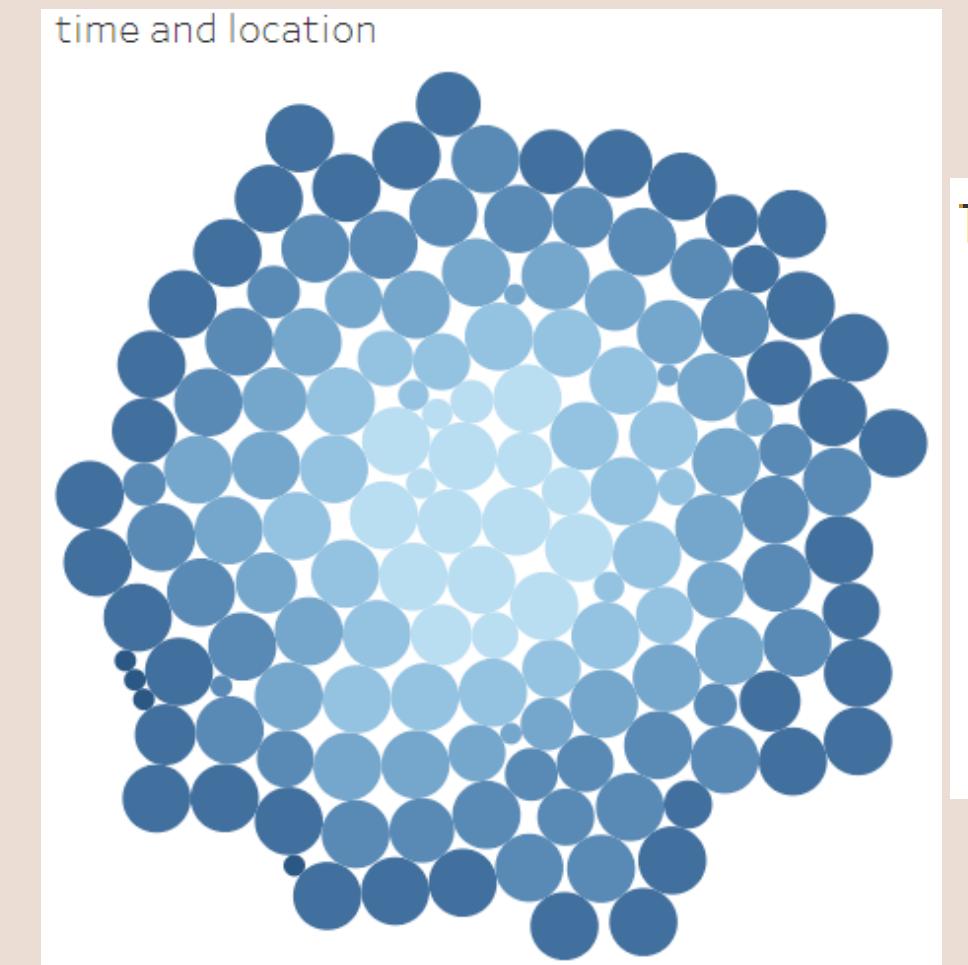


People consuming DrugC tends to record a lower BP significantly whereas the records are average for people consuming Drug Y, yet another reason for Drug C's lower consumption and Drug Y's higher consumption.

Analysis and Insights 8



The supply rate wrt time has increased till 2017 and then has significantly declined after 2017



The maximum drugs were supplied in the year 1970 and the least in the year 2020.

Conclusion Project 8

In conclusion, I would like to tell that after doing a thorough analysis we were able to derive the insights from the data and was able to plot various graphs using that data. The data that once looked useless became useful and helped to find out the courses that were a burden for Udemy to continue providing. Analyzing the data proved helpful in finding various issues among the courses. It also helped us improve business and marketing strategies to produce and decrease the production of resources.

Working excel /tableau and pdf link of the projects made

[Project 1-pdf link](#)

[Project 2-pdf link](#)

[Project 3-pdf link](#)

[Project 4-pdf link](#)

[Project 5-pdf link](#)

[Project 6-pdf link](#)

[Project 7-pdf link](#)

[tabelau for project - Ads Airing Analysis](#)

[tableau for project 8](#)

[excel link for call Volume trend Analysis](#)

[excel link for bank loan case study](#)

[excel link for hiring process analytics](#)

[excel for IMDB movie analysis](#)

*Thank
you*