

Data Analytics Portfolio

HEMANT MURKUTE



ABOUT ME

Currently in Second year pursuing B.E. Electronics and Telecommunication Engineering. I have secured 9.59 CGPA (till 3rd Sem) and have several skills including Python, C, C++, Data Analytics, MATLAB, Multisim, Machine Learning, MySQL, Excel.

I have worked on various personal projects related to Excel and Python.

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

TABLE OF CONTENTS

About Me

Table Of Contents

Module -1 Project: Data Analytics Process

Module -2 Project: Instagram User Analytics

Module -3 Project: Operation & Metric Analytics

Module -4 Project: Hiring Process Analytics

Module -5 Project: IMDB Movie Analysis

Module -6 Project: Bank Loan Case Study

Module -7 Project: Impact of Car Features

Module -8 Project: ABC Call Volume Trend

Conclusion

Module -1 Project: Data Analytics Process

Data analytics is not just limited to businesses; we unknowingly use it in our daily lives. This project aimed to identify real-life scenarios where data analytics plays a crucial role and explain how structured analytical processes help in decision-making.

The main objective of this project was to explore how data analytics is applied in routine activities and to illustrate the different stages of the data analytics process: Plan, Prepare, Process, Analyze, Share, and Act.

Real-Life Scenarios Explored:

Travel Planning: I demonstrated how data analytics helps in choosing a destination, managing a budget, comparing accommodation options, and ensuring a smooth trip.

Home Renovation & Interior Design: This scenario focused on how data is used to budget for renovations, select materials, optimize space, and achieve the desired aesthetic appeal.

Approach & Tools Used:

The project was documented using Microsoft Word, where I structured real-world examples and explained how data analytics plays a role in improving decision-making.

Conclusion:

Through this project, I gained a deeper understanding of how data analytics simplifies complex decisions. Whether planning a trip or designing a home, following a systematic approach ensures better outcomes and efficient resource utilization.

Module -2 Project: Instagram User Analytics

The second project was Instagram user Analytics. This project was based on the queries raised by the product team of Instagram on various issues. They raised various queries and I used different SQL functions to solve and then used MS Office PowerPoint to prepare a detailed report on them.

The following queries were raised by the marketing team: -

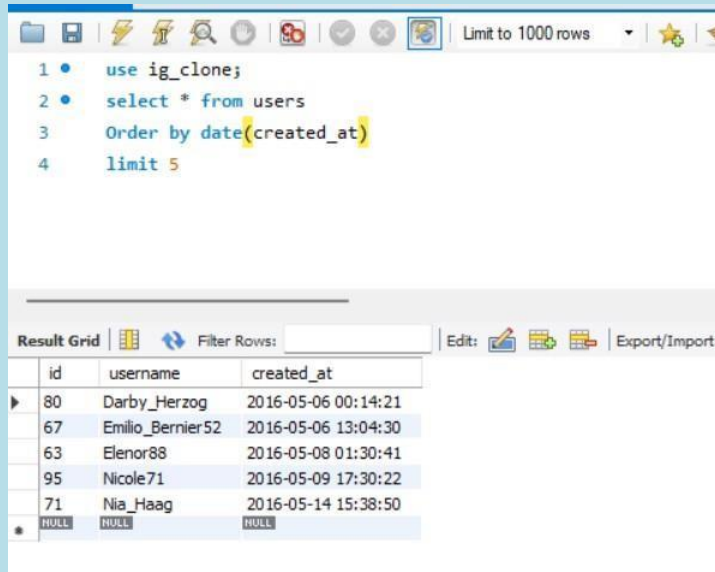
1. People who have been using platform for the longest time.
2. Who are the inactive users.
3. Which user has got the most likes on a photo.
4. Which are the most popular hashtags on Instagram.
5. Which is the best day to launch Ads.

The following queries were raised by the investors: -

1. How active are users on Instagram
2. Is Instagram crowded with fake and dummy accounts.

I first began with importing the data to SQL database. Then closely went through the data set by analyzing each column and type of values they store. After that I checked for any duplicate or null values or if any data cleaning is required. Then I began with the analysis and found out the answers.

1. People who have been using platform for the longest time.



```
1 • use ig_clone;
2 • select * from users
3 • Order by date(created_at)
4 • limit 5
```

	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
	71	Nia_Haag	2016-05-14 15:38:50
*	NULL	NULL	NULL

2. Who are the inactive users.

```

1 • use ig_clone;
2 • select *
3   from users
4   left join photos
5   on users.id = photos.user_id
6   where user_id is NULL

```

	id	username	created_at	id	image_url	user_id	created_at
▶	5	Aniya_Hackett	2016-12-07 01:04:39	NULL	NULL	NULL	NULL
	7	Kassandra_Homenick	2016-12-12 06:50:08	NULL	NULL	NULL	NULL
	14	Jadyn81	2017-02-06 23:29:16	NULL	NULL	NULL	NULL
	21	Rocio33	2017-01-23 11:51:15	NULL	NULL	NULL	NULL
	24	Maxwell.Halvorson	2017-04-18 02:32:44	NULL	NULL	NULL	NULL
	25	Tierra.Trantow	2016-10-03 12:49:21	NULL	NULL	NULL	NULL
	34	Pearl7	2016-07-08 21:42:01	NULL	NULL	NULL	NULL
	36	Ollie_Ledner37	2016-08-04 15:42:20	NULL	NULL	NULL	NULL
	41	Mckenna17	2016-07-17 17:25:45	NULL	NULL	NULL	NULL
	45	David.Osinski47	2017-02-05 21:23:37	NULL	NULL	NULL	NULL
	49	Morgan.Kassulke	2016-10-30 12:42:31	NULL	NULL	NULL	NULL
	53	Linnea59	2017-02-07 07:49:34	NULL	NULL	NULL	NULL
	54	Duane60	2016-12-21 04:43:38	NULL	NULL	NULL	NULL
	57	Julien_Schmidt	2017-02-02 23:12:48	NULL	NULL	NULL	NULL

3. Which user has got the most likes on a photo.

```

1 • use ig_clone;
2 • select users.username, photos.id, photos.image_url, count(*) as total_likes
3   from likes
4   join photos on photos.id=likes.photo_id
5   join users on users.id=likes.photo_id
6   group by photos.id
7   order by total_likes desc
8   limit 10;

```

	username	id	image_url	total_likes
▶	Kaley9	30	http://kenny.com	41
	Jayson65	61	https://dejon.name	41
	Zack_Kemmer93	52	https://hershel.com	41
	Tomas.Beatty93	97	https://carolanne.com	40
	Alexandro35	13	https://fred.com	40
	Javonte83	100	https://brook.com	39
	Ressie_Stanton46	62	https://rigoberto.net	39
	Seth46	44	http://golden.org	39
	Mike.Auer39	66	http://lionel.net	39
	Harley Lind18	3	http://vickv.biz	38

4. Which is the best day to launch Ads.

```
1 • use ig_clone;
2 • SELECT
3     DAYNAME(created_at) AS day, COUNT(*) AS total
4 FROM
5     users
6 GROUP BY day
7 ORDER BY total DESC
```

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

	day	total
	Thursday	16
	Sunday	16
▶	Friday	15
	Tuesday	14
	Monday	14
	Wednesday	13
	Saturday	12

The following queries were raised by the investors: -

1. How active are users on Instagram

SQL File 4* x

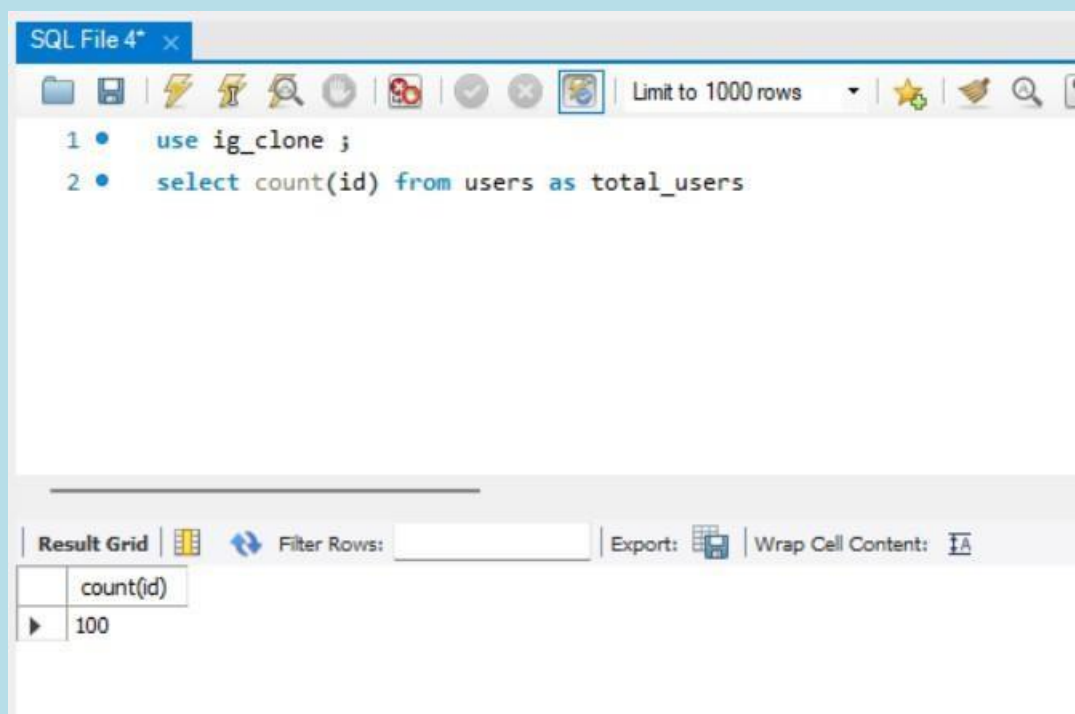
Limit to 1000 rows

```
1 • use ig_clone ;
2 • SELECT((SELECT COUNT(*)FROM photos)/(SELECT COUNT(*) FROM users)) as avg_post_time;
```

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

	avg_post_time
▶	2.5700

Total no of users on Instagram



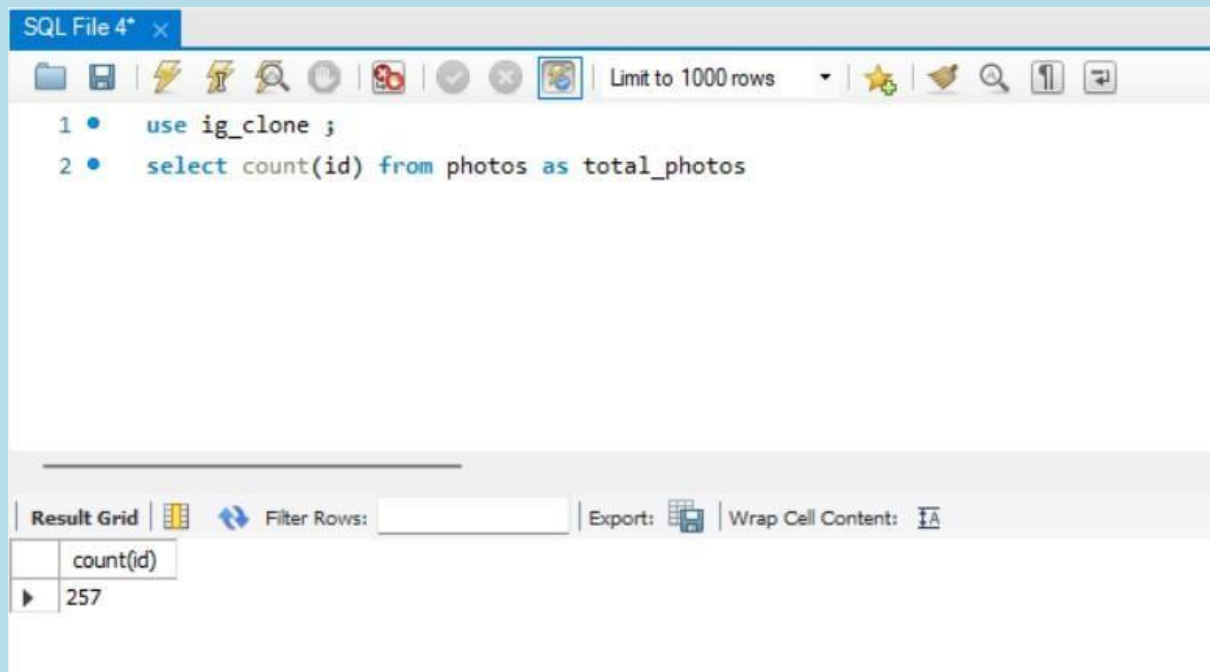
The screenshot shows a SQL editor window titled "SQL File 4*" with a toolbar and a query editor. The query is as follows:

```
1 • use ig_clone ;  
2 • select count(id) from users as total_users
```

Below the query editor, the "Result Grid" is displayed, showing the result of the query:

count(id)
100

Total no of photos



The screenshot shows a SQL editor window titled "SQL File 4*" with a toolbar and a query editor. The query is as follows:

```
1 • use ig_clone ;  
2 • select count(id) from photos as total_photos
```

Below the query editor, the "Result Grid" is displayed, showing the result of the query:

count(id)
257

Module -3 Project: Operation & Metric Analytics

The third project was Operation Analytics and investigating metric spikes. Operation Analytics is the analysis done for the complete end to end operations of a company. This helps the company to understand in which areas they can improve. On the other hand, metric spike is done to understand certain trends like daily increase or dip in engagement, sales figures etc.

I first imported the data to SQL database and used certain SQL functions to get desired answers and then prepared a presentation in MS Office PowerPoint with relevant explanation.

There were two different case studies with different data sets for each one – Job Data and Investigating metric spike.

A brief description of both case studies and output required from both of them are given below: -

CASE STUDY 01 (JOB DATA)

1. Amount of jobs reviewed over time.
2. No. of events happening per second.
3. Share of each language for different contents.
4. Rows that have same values present in them.

CASE STUDY 02 (INVESTIGATING METRIC SPIKE)

5. To measure the activeness of a user.
6. Amount of users growing over time for a product.
7. Users getting retained weekly after signing up for a product.
8. To measure the activeness of a user weekly.
9. Users engaging with the email service.

I went through the data set closely, analyzed every column and checked for any duplicates or if any data cleaning is required. Then I began with finding solutions for above mentioned queries. Solutions to them are as below –

CASE STUDY 01 (JOB DATA)

1. Amount of jobs reviewed over time.

```
1 • use opr_n_ma ;
2 • SELECT ((COUNT(job_id))/sum(time_spent)/3600) as job_reviewed
3   FROM job_data
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

job_reviewed
0.00000746

2. No. of events happening per second.

```
1 • select ds , job_nums, total_time,
2   sum(job_nums) over (order by ds rows between 6 preceding and current row)/sum(total_time) as 7day_rolling_average
3   from
4   (
5     select ds ,count(job_id) as job_nums,sum(time_spent) as total_time
6     from job_data where ds >='2020-11-01' and ds <='2020-11-30'
7     group by ds
8   ) a
9   group by
10  ds
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

ds	job_nums	total_time	7day_rolling_average
2020-11-25	1	45	0.0222
2020-11-26	1	56	0.0357
2020-11-27	1	104	0.0288
2020-11-28	2	33	0.1515
2020-11-29	1	20	0.3000
2020-11-30	2	40	0.2000

3. Share of each language for different contents.

```
1 -- select * from job_data
2 • select language, lang_count,
3    sum(lang_count)/(sum(lang_count)over(order by language rows between unbounded preceding and unbounded following )) * 100.0 as perc_language
4 from
5 (
6   select language, count(language) as lang_count
7   from job_data
8   group by language
9 ) a
10
```

language	lang_count	perc_language
Arabic	1	12.50000
English	1	12.50000
French	1	12.50000
Hindi	1	12.50000
Italian	1	12.50000
Persian	3	37.50000

4. Rows that have same values present in them.

```
1 • select job_id, count(job_id) as duplicate_count
2 from job_data
3 group by job_id
4 having duplicate_count > 1
```

job_id	duplicate_count
23	3

CASE STUDY 02 (INVESTIGATING METRIC SPIKE)

5. To measure the activeness of a user.

The screenshot shows a SQL IDE window titled "SQL File 4*" with a tab for "events". The query is as follows:

```
1 • use opr_n_ma ;
2
3 • SELECT
4     extract(year from occurred_at) as year,
5     extract(week from occurred_at) as weeknum,
6     count(distinct user_id) as user_engagement
7 FROM
8     events
9 GROUP BY
10    year, weeknum
11 order by
12    year, weeknum
```

The "Result Grid" at the bottom displays the following data:

	year	weeknum	user_engagement
▶	2014	17	663
	2014	18	1068
	2014	19	1113
	2014	20	1154
	2014	21	1121

Result 2

6. Amount of users growing over time for a product.

The screenshot shows a SQL IDE window titled "SQL File 4*" with a tab for "events". The query is as follows:

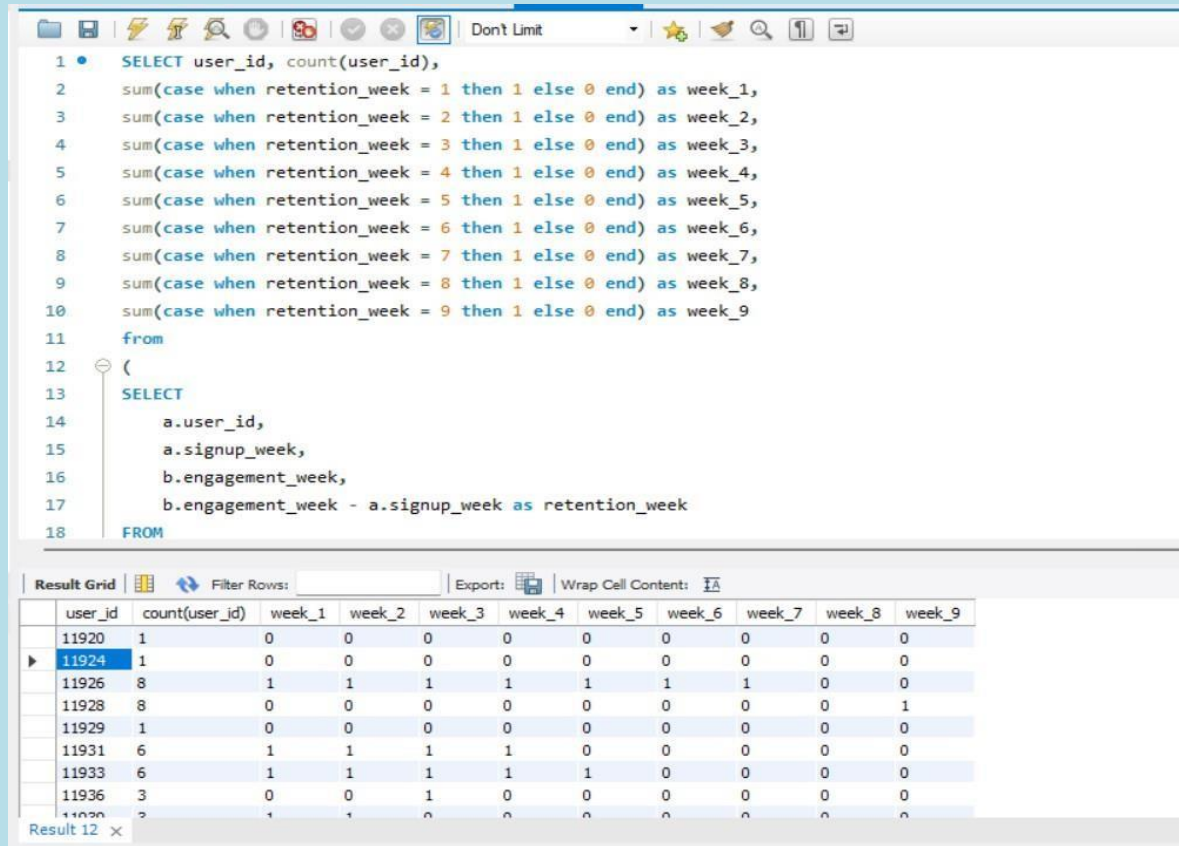
```
1 • use opr_n_ma ;
2
3 • select
4     year,
5     weeknum,
6     new_active_user,
7     sum(new_active_user)over(order by year,weeknum rows between unbounded preceding and current row) as cum_active_user
8 FROM
9     (
10    SELECT
11        extract(year from activated_at) as year,
12        extract(week from activated_at) as weeknum,
13        count(distinct user_id) as new_active_user
14    FROM
15        users
16    where
17        state = 'active'
18    group by
19        year,
20        weeknum
21    ) a
22
```

The "Result Grid" at the bottom displays the following data:

	year	weeknum	new_active_user	cum_active_user
▶	2013	0	23	23
	2013	1	30	53
	2013	2	48	101
	2013	3	36	137
	2013	4	30	167

Result 47

7. Users getting retained weekly after signing up for a product.



The screenshot shows a SQL IDE with a query to calculate user retention. The query uses a subquery to determine the retention week for each user based on their signup and engagement dates. The main query then counts the number of users retained for each week from week 1 to week 9.

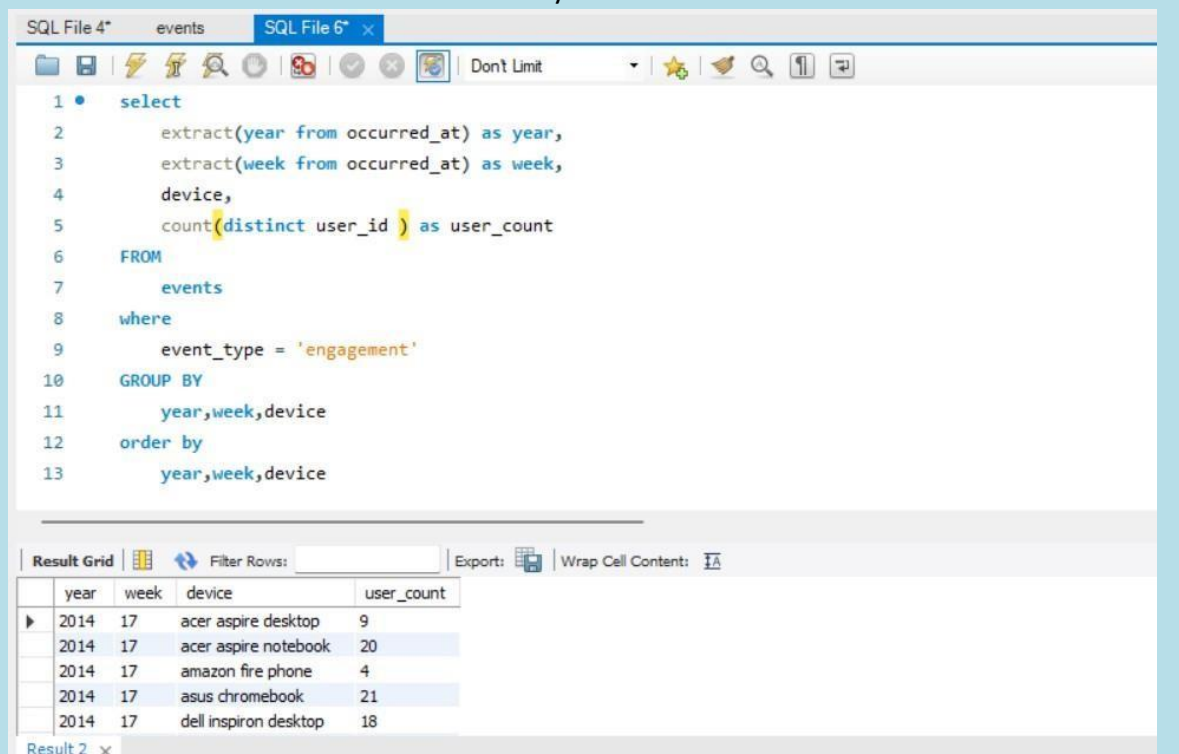
```

1 • SELECT user_id, count(user_id),
2     sum(case when retention_week = 1 then 1 else 0 end) as week_1,
3     sum(case when retention_week = 2 then 1 else 0 end) as week_2,
4     sum(case when retention_week = 3 then 1 else 0 end) as week_3,
5     sum(case when retention_week = 4 then 1 else 0 end) as week_4,
6     sum(case when retention_week = 5 then 1 else 0 end) as week_5,
7     sum(case when retention_week = 6 then 1 else 0 end) as week_6,
8     sum(case when retention_week = 7 then 1 else 0 end) as week_7,
9     sum(case when retention_week = 8 then 1 else 0 end) as week_8,
10    sum(case when retention_week = 9 then 1 else 0 end) as week_9
11 from
12 (
13     SELECT
14         a.user_id,
15         a.signup_week,
16         b.engagement_week,
17         b.engagement_week - a.signup_week as retention_week
18     FROM
  
```

The result grid displays the following data:

	user_id	count(user_id)	week_1	week_2	week_3	week_4	week_5	week_6	week_7	week_8	week_9
	11920	1	0	0	0	0	0	0	0	0	0
▶	11924	1	0	0	0	0	0	0	0	0	0
	11926	8	1	1	1	1	1	1	1	0	0
	11928	8	0	0	0	0	0	0	0	0	1
	11929	1	0	0	0	0	0	0	0	0	0
	11931	6	1	1	1	1	0	0	0	0	0
	11933	6	1	1	1	1	0	0	0	0	0
	11936	3	0	0	1	0	0	0	0	0	0
	11937	2	1	1	0	0	0	0	0	0	0

8. To measure the activeness of a user weekly.



The screenshot shows a SQL IDE with a query to measure user activeness. The query extracts the year and week from the 'occurred_at' date, counts the number of distinct users for each week, and filters for 'engagement' events. The results are grouped by year, week, and device.

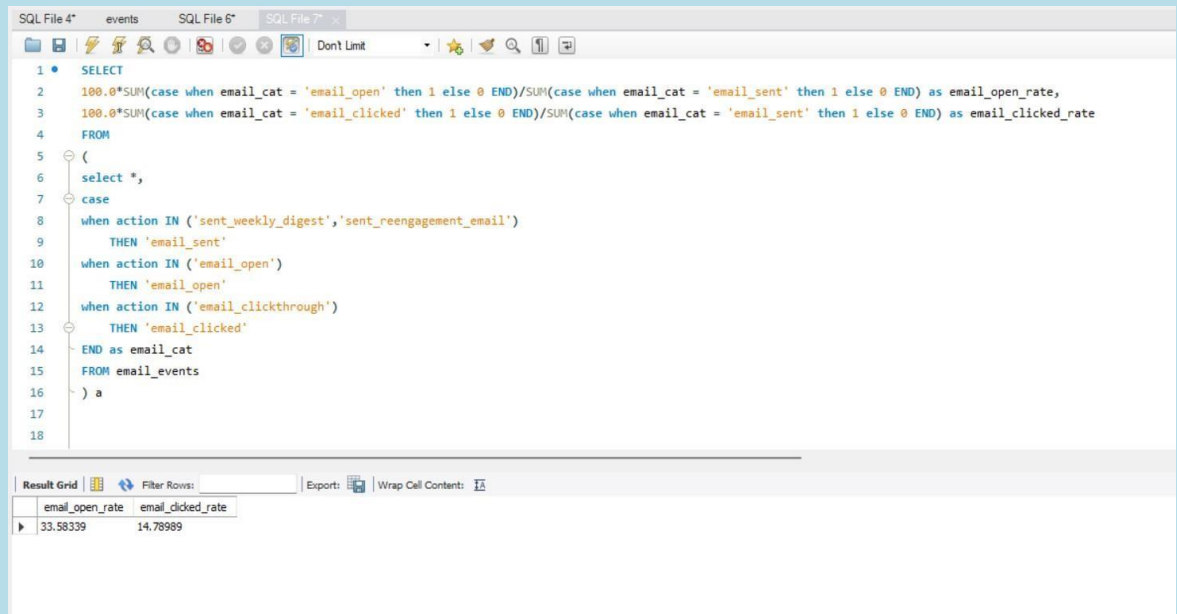
```

1 • select
2     extract(year from occurred_at) as year,
3     extract(week from occurred_at) as week,
4     device,
5     count(distinct user_id ) as user_count
6 FROM
7     events
8 where
9     event_type = 'engagement'
10 GROUP BY
11     year, week, device
12 order by
13     year, week, device
  
```

The result grid displays the following data:

	year	week	device	user_count
▶	2014	17	acer aspire desktop	9
	2014	17	acer aspire notebook	20
	2014	17	amazon fire phone	4
	2014	17	asus chromebook	21
	2014	17	dell inspiron desktop	18

9. Users engaging with the email service.



The screenshot shows a SQL IDE interface with a query editor and a results grid. The query is as follows:

```
1 SELECT
2 100.0*SUM(case when email_cat = 'email_open' then 1 else 0 END)/SUM(case when email_cat = 'email_sent' then 1 else 0 END) as email_open_rate,
3 100.0*SUM(case when email_cat = 'email_clicked' then 1 else 0 END)/SUM(case when email_cat = 'email_sent' then 1 else 0 END) as email_clicked_rate
4 FROM
5 (
6 select *,
7 case
8 when action IN ('sent_weekly_digest','sent_reengagement_email')
9 THEN 'email_sent'
10 when action IN ('email_open')
11 THEN 'email_open'
12 when action IN ('email_clickthrough')
13 THEN 'email_clicked'
14 END as email_cat
15 FROM email_events
16 ) a
17
18
```

The results grid shows the following data:

email_open_rate	email_clicked_rate
33.58339	14.78989

After thorough analysis, we found out the answers for every question asked.

Module -4 Project: Hiring Process Analytics

The fourth project I did was hiring process analytics. It is the most important function of a company. Companies get to know about number of hirings, number of resignations, number of rejections, interviews, types of jobs, vacancies etc.

A brief description of the case study and output required is given below: -

1. How many males and females are Hired?
2. What is the average salary offered in this company?
3. Draw the class intervals for salary in the company?
4. Draw Pie Chart / Bar Graph (or any other graph) to show proportion of people working different department?
5. Represent different post tiers using chart/graph?

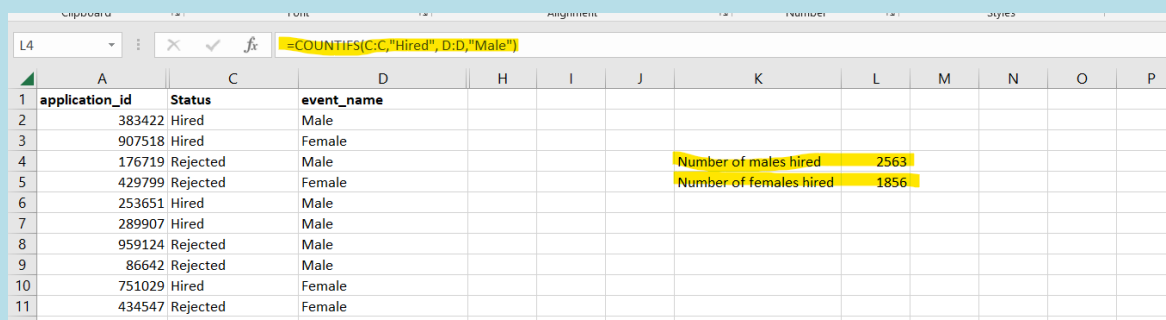
Before beginning with the analysis, I first went through the data, tried to understand every column, then checked for any duplicate data or if any data cleaning is required. After verifying all above mentioned processes I began with getting solutions for above mentioned problem statements.

The data was present in excel so performed the analysis in MS Excel and prepared a detailed report in MS PowerPoint. In order to make the report more presentable, used Pivot Tables to draw graphs. The answers are as follows -

1. How many males and females are Hired?

The formula used in excel was as follows: -

- COUNTIFS (C:C,"Hired", D:D,"Male")
- COUNTIFS (C:C,"Hired", D:D, "Female")



The screenshot shows an Excel spreadsheet with the following data:

	A	C	D	H	I	J	K	L	M	N	O	P
1	application_id	Status	event_name									
2	383422	Hired	Male									
3	907518	Hired	Female									
4	176719	Rejected	Male									
5	429799	Rejected	Female									
6	253651	Hired	Male									
7	289907	Hired	Male									
8	959124	Rejected	Male									
9	86642	Rejected	Male									
10	751029	Hired	Female									
11	434547	Rejected	Female									
12	548854	Rejected	Male									

Formulas and Results:

- Formula bar: `=COUNTIFS(C:C,"Hired",D:D,"Male")`
- Cell K4: `=COUNTIFS(C:C,"Hired",D:D,"Male")` (Result: 2563)
- Cell L4: `=COUNTIFS(C:C,"Hired",D:D,"Female")` (Result: 1856)

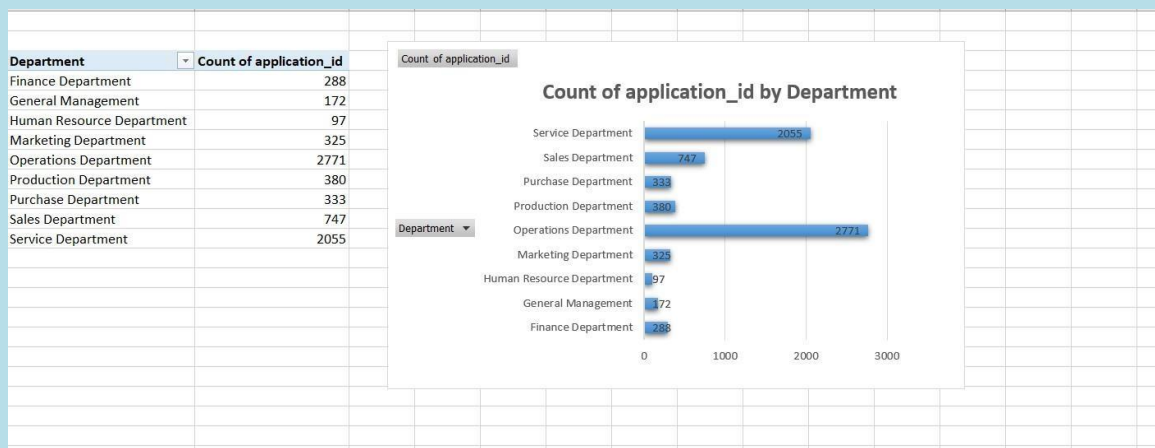
2. What is the average salary offered in this company?

L4											
=AVERAGE(G:G)											
1	application_id	Status	Offered Salary	H	I	J	K	L	M	N	O
2	383422	Hired	56553								
3	907518	Hired	22075								
4	176719	Rejected	70069								
5	429799	Rejected	3207								
6	253651	Hired	29668								
7	289907	Hired	85914								
8	959124	Rejected	69904								
9	86642	Rejected	11758								
10	751029	Hired	15156								
11											

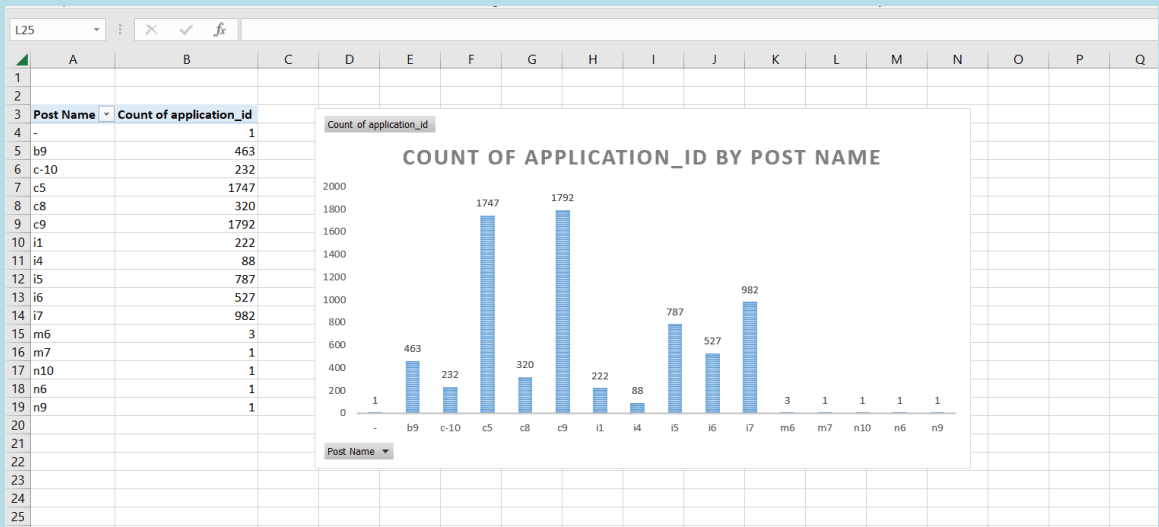
3. Draw the class intervals for salary in the company?

L5											
=MAX(G:G)-MIN(G:G)											
1	application_id	Offered Salary	H	I	J	K	L	M	N	O	P
2	383422	56553									
3	907518	22075									
4	176719	70069									
5	429799	3207									
6	253651	29668									
7	289907	85914									
8	959124	69904									
9	86642	11758									
10	751029	15156									
11	434547	49515									
12	518854	26990									

4. Draw Pie Chart / Bar Graph (or any other graph) to show proportion of people working different department?



5. Represent different post tiers using chart/graph?



Module -5 Project: IMDB

Movie Analysis

This is the fifth project I did. This was based on the movie rating platform IMDB. IMDB collects large amount of data of each movie like actors, directors, producers' names, public rating, budget, earnings, stores them in their databases and then analyses them based on their requirements. This project gives a brief idea about the movie analysis done by IMDB to give ratings based on different criteria.

Output required is given below: -

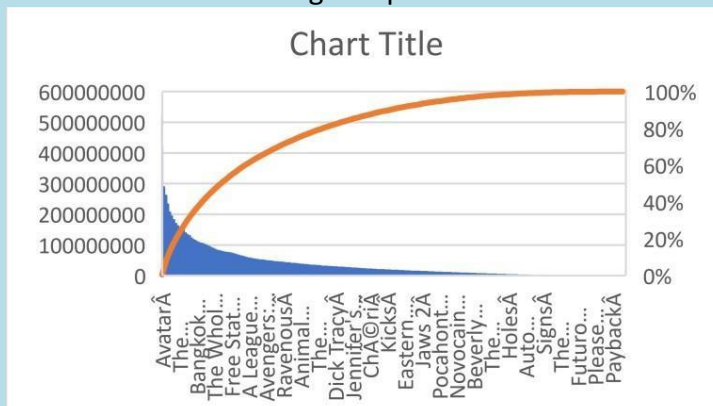
1. Clean the data
2. Find the movies with the highest profit?
3. Find IMDB Top 250
4. Find the best directors
5. Find popular genres
6. Find the critic-favorite and audience-favorite actors

The data was given in excel format, so I first tried to understand the data and variables stored in each column, then went on with data cleaning and finally beginning with the analysis. I used MS Excel to study and analyze data and then used MS PowerPoint to prepare a detailed report for the same.

The insights drawn from the data sets are as follows –

1. Clean the data
2. Find the movies with the highest profit?

Ans – Movie with the highest profit is Avatar.



3. Find IMDB Top 250

The Shawshank Redemption										
	A	B	C	D	E	F	G	H	I	J
1	movie_title	imdb_score	num_voted_users (applied filter > 25000)	IMDB_top_250 (num_voted_users>25000)						
2	The Shawshank Redem	9.3	1689764	The Shawshank Redemption						
3	The Godfather	9.2	1155770	The Godfather						
4	The Dark Knight	9	1676169	The Dark Knight						
5	The Godfather: Part II	9	790926	The Godfather: Part II						
6	Fargo	9	170055	Fargo						
7	The Lord of the Rings:	8.9	1215718	The Lord of the Rings: The Return of the King						
8	Schindler's List	8.9	865020	Schindler's List						
9	Pulp Fiction	8.9	1324680	Pulp Fiction						
10	The Good, the Bad and	8.9	503509	The Good, the Bad and the Ugly						
11	12 Angry Men	8.9	447785	12 Angry Men						
12	Inception	8.8	1468200	Inception						
13	The Lord of the Rings:	8.8	1238746	The Lord of the Rings: The Fellowship of the Ring						
14	Daredevil	8.8	213483	Daredevil						
15	The Shawshank Redem	9.3	1689764	The Shawshank Redemption						

4. Find the best directors

The top 10 directors are as below: -

director_name	imdb_score	top 10 directors
John Blanchard Average	9.5	John Blanchard Average
John Blanchard	9.5	Frank Darabont Average
Frank Darabont Average	9.3	Francis Ford Coppola Average
Frank Darabont	9.3	John Stockwell Average
Francis Ford Coppola Average	9.2	Christopher Nolan Average
Francis Ford Coppola	9.2	Francis Ford Coppola Average
John Stockwell Average	9.1	Peter Jackson Average
John Stockwell	9.1	Steven Spielberg Average
Christopher Nolan Average	9	Quentin Tarantino Average
Christopher Nolan	9	Sergio Leone Average
Francis Ford Coppola Average	9	
Francis Ford Coppola	9	
Peter Jackson Average	8.9	

5. Find popular genres

genres	imdb_score	top 10 genres
Comedy	9.5	Comedy Average
Comedy Average	9.5	Crime Drama Average
Crime Drama	9.3	Drama Average
Crime Drama Average	9.3	Action Average
Crime Drama	9.2	Action Crime Drama Thriller Average
Crime Drama Average	9.2	Crime Drama Thriller Average
Drama	9.1	Biography Drama History Average
Drama Average	9.1	Western Average
Drama	9.1	Action Adventure Sci-Fi Thriller Average
Drama Average	9.1	Action Adventure Drama Fantasy Average
Action	9.1	
Action Average	9.1	
Action Crime Drama Thriller	9	
Action Crime Drama Thriller Average	9	

6. Find the critic-favorite and audience-favorite actors

	B	C	D	E	F	G	H	I	J
1	Meryl Streep	Leonardo DiCaprio	Brad Pitt	Combined column (appended)				help column	actor_1_name
2	It's Complicated	Titanic	The Curious Case of Benjamin Button	It's Complicated				CCH Pounder:1	CCH Pounder
3	The River Wild	The Great Gatsby	Troy	The River Wild				Johnny Depp:1	Johnny Depp
4	Julie & Julia	Inception	Ocean's Twelve	Julie & Julia				Christoph Waltz:1	Christoph Waltz
5	The Devil Wears Prada	The Revenant	Mr. & Mrs. Smith	The Devil Wears Prada				Tom Hardy:1	Tom Hardy
6	Lions for Lambs	The Aviator	Spy Game	Lions for Lambs				Doug Walker:1	Doug Walker
7	Out of Africa	Django Unchained	Ocean's Eleven	Out of Africa				Daryl Sabara:1	Daryl Sabara
8	Hope Springs	Blood Diamond	Fury	Hope Springs				J.K. Simmons:1	J.K. Simmons
9	One True Thing	The Wolf of Wall Street	Seven Years in Tibet	One True Thing				Brad Garrett:1	Brad Garrett
10	Florence Foster Jenkins	Gangs of New York	Fight Club	Florence Foster Jenkins				Chris Hemsworth:1	Chris Hemsworth
11	The Hours	The Departed	Sinbad: Legend of the Seven Seas	The Hours				Alan Rickman:1	Alan Rickman
12	The Iron Lady	Shutter Island	Interview with the Vampire: The Vampire Chronicles	The Iron Lady				Henry Cavill:1	Henry Cavill
13	A Prairie Home Companion	Body of Lies	The Tree of Life	A Prairie Home Companion				Kevin Spacey:1	Kevin Spacey
14	Julia	Catch Me If You Can	The Assassination of Jesse James by the Coward Robert Ford	Julia				Giancarlo Giannini:1	Giancarlo Giannini
15	The Beach	The Beach	Babel	Titanic				Johnny Depp:2	Johnny Depp
16	Revolutionary Road	By the Sea	The Great Gatsby	The Great Gatsby				Johnny Depp:3	Johnny Depp
17	The Man in the Iron Mask	Killing Them Softly	Inception	Inception				Henry Cavill:2	Henry Cavill
18	J. Edgar	True Romance	The Revenant	The Revenant				Peter Dinklage:1	Peter Dinklage
19	The Quick and the Dead	Johnny Suede	The Aviator	The Aviator				Chris Hemsworth:2	Chris Hemsworth
20	Marvin's Room		Django Unchained	Django Unchained				Johnny Depp:4	Johnny Depp
21	Romeo + Juliet		Blood Diamond	Blood Diamond				Will Smith:1	Will Smith
22	The Great Gatsby		The Wolf of Wall Street	The Wolf of Wall Street				Aidan Turner:1	Aidan Turner

This completes our analysis as all questions have now been answered.

Module -6 Project: Bank Loan Case Study

This project is about a case study related to a Bank Loan. We have to carry out an EDA (Exploratory Data Analysis). Based on our analysis, we will get the solution for required questions.

I first analyzed the data. While analyzing, I found out that data had a lot of missing values. So, my first task was to get the missing values by performing mean, median and mode functions as required. So, I began by cleaning the data and then finding the outliers so as to make the data standardized. To perform the analysis, I used MS Excel 2019 for analysis and used MS Word to prepare the report.

So, let's begin with analysis.....

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

First, we imported the data to excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE
2	100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied
3	100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family
4	100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied
5	100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied
6	100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied
7	100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner
8	100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied
9	100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied
10	100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children
11	100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied
12	100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied
13	100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children
14	100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied
15	100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied
16	100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied
17	100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family
18	100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied
19	100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied
20	100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A
21	100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied
22	100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied
23	100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied
24	100026	0	Cash loans	F	N	N	0	450000	497520	32521.5	450000	Unaccompanied
25	100027	0	Cash loans	F	N	Y	1	83250	239850	23850	225000	Unaccompanied
26	100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied
27	100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied
28	100031	1	Cash loans	F	N	Y	0	112500	979992	27076.5	702000	Unaccompanied
29	100032	0	Cash loans	M	N	Y	1	112500	327024	23827.5	270000	Family

Column1	Table	Row	Description	Special
1	application_data	SK_ID_CURR	ID of loan in our sample	
2	application_data	TARGET	Target variable (1 - client with payment difficulties: he/she had late payment more than X days on at	
4	application_data	NAME_CONTRACT_TYPE	Identification if loan is cash or revolving	
6	application_data	CODE_GENDER	Gender of the client	
7	application_data	FLAG_OWN_CAR	Flag if the client owns a car	
8	application_data	FLAG_OWN_REALTY	Flag if client owns a house or flat	
9	application_data	CNT_CHILDREN	Number of children the client has	
10	application_data	AMT_INCOME_TOTAL	Income of the client	
11	application_data	AMT_CREDIT	Credit amount of the loan	
12	application_data	AMT_ANNUITY	Loan annuity	
13	application_data	AMT_GOODS_PRICE	For consumer loans it is the price of the goods for which the loan is given	
14	application_data	NAME_TYPE_SUITE	Who was accompanying client when he was applying for the loan	
15	application_data	NAME_INCOME_TYPE	Clients income type (businessman, working, maternity leave,...)	
16	application_data	NAME_EDUCATION_TYPE	Level of highest education the client achieved	
17	application_data	NAME_FAMILY_STATUS	Family status of the client	
18	application_data	NAME_HOUSING_TYPE	What is the housing situation of the client (renting, living with parents, ...)	
19	application_data	REGION_POPULATION_RELATIVE	Normalized population of region where client lives (higher number means the client lives in more po normalized	
20	application_data	DAYS_BIRTH	Client's age in days at the time of application	time only relative to the application
21	application_data	DAYS_EMPLOYED	How many days before the application the person started current employment	time only relative to the application
22	application_data	DAYS_REGISTRATION	How many days before the application did client change his registration	time only relative to the application
23	application_data	DAYS_ID_PUBLISH	How many days before the application did client change the identity document with which he applie	time only relative to the application
24	application_data	OWN_CAR_AGE	Age of client's car	
25	application_data	FLAG_MOBIL	Did client provide mobile phone (1=YES, 0=NO)	
26	application_data	FLAG_EMP_PHONE	Did client provide work phone (1=YES, 0=NO)	
27	application_data	FLAG_WORK_PHONE	Did client provide home phone (1=YES, 0=NO)	
28	application_data	FLAG_CONT_MOBILE	Was mobile phone reachable (1=YES, 0=NO)	
29	application_data	FLAG_PHONE	Did client provide home phone (1=YES, 0=NO)	
30	application_data	FLAG_EMAIL	Did client provide email (1=YES, 0=NO)	

SK_ID_CURR	SK_ID_PREV	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	HOURL_APPR_PROCESS_START
2030495	271877	Consumer loans	1730.43	17145	17145	0	17145	SATURDAY	
2802425	108129	Cash loans	25188.615	607500	679671		607500	THURSDAY	
2523466	122040	Cash loans	15060.735	112500	136444.5		112500	TUESDAY	
2819243	176158	Cash loans	47041.335	450000	470790		450000	MONDAY	
1784265	202054	Cash loans	31924.395	337500	404055		337500	THURSDAY	
1383531	199383	Cash loans	23703.93	315000	340573.5		315000	SATURDAY	
2315218	175704	Cash loans		0	0			TUESDAY	
1656711	296299	Cash loans		0	0			MONDAY	
2367563	342292	Cash loans		0	0			MONDAY	
2579447	334349	Cash loans		0	0			SATURDAY	
1715995	447712	Cash loans	11368.62	270000	335754		270000	FRIDAY	
2257824	161140	Cash loans	13832.775	211500	246397.5		211500	FRIDAY	
2330894	258628	Cash loans	12165.21	148500	174361.5		148500	TUESDAY	
1397919	321676	Consumer loans	7654.86	53779.5	57564	0	53779.5	SUNDAY	
2273188	270658	Consumer loans	9644.22	26550	27252	0	26550	SATURDAY	
1232483	151612	Consumer loans	21307.455	126490.5	119853	12649.5	126490.5	TUESDAY	
2163253	154602	Consumer loans	4187.34	26955	27297	1350	26955	SATURDAY	
1285768	142748	Revolving loans	9000	180000	180000		180000	FRIDAY	
2393109	396305	Cash loans	10181.7	180000	180000		180000	THURSDAY	
1173070	199178	Cash loans	4666.5	45000	49455		45000	SATURDAY	
1506815	166490	Cash loans	25454.025	450000	491580		450000	MONDAY	
1182516	267782	Cash loans	20361.6	405000	451777.5		405000	SATURDAY	
1172842	302212	Cash loans		0	0			TUESDAY	
1172937	302212	Cash loans	39475.305	1129500	1277104.5		1129500	THURSDAY	
1555330	199353	Cash loans		0	0			SATURDAY	
1543131	275707	Cash loans	22619.52	229500	241920		229500	THURSDAY	
2536650	338725	Cash loans	16708.32	369000	369000		369000	WEDNESDAY	
1676258	433469	Cash loans	22242.825	247500	268083	0	247500	THURSDAY	

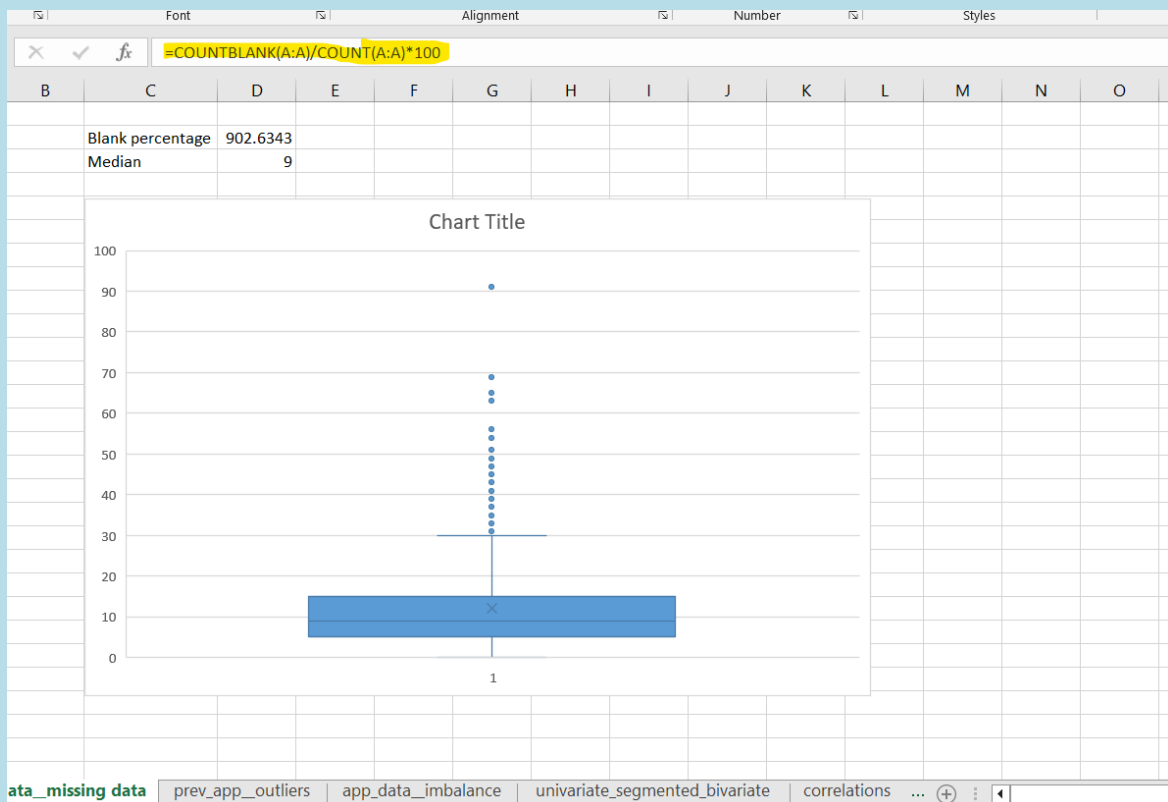
Then, in order to clean data I, highlighted the blank cells first .

SK_ID_CURR	SK_ID_PREV	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	HOURL_APPR_PROCESS_START
2030495	271877	Consumer loans	1730.43	17145	17145	0	17145	SATURDAY	
2802425	108129	Cash loans	25188.615	607500	679671		607500	THURSDAY	
2523466	122040	Cash loans	15060.735	112500	136444.5		112500	TUESDAY	
2819243	176158	Cash loans	47041.335	450000	470790		450000	MONDAY	
1784265	202054	Cash loans	31924.395	337500	404055		337500	THURSDAY	
1383531	199383	Cash loans	23703.93	315000	340573.5		315000	SATURDAY	
2315218	175704	Cash loans		0	0			TUESDAY	
1656711	296299	Cash loans		0	0			MONDAY	
2367563	342292	Cash loans		0	0			MONDAY	
2579447	334349	Cash loans		0	0			SATURDAY	
1715995	447712	Cash loans	11368.62	270000	335754		270000	FRIDAY	
2257824	161140	Cash loans	13832.775	211500	246397.5		211500	FRIDAY	
2330894	258628	Cash loans	12165.21	148500	174361.5		148500	TUESDAY	
1397919	321676	Consumer loans	7654.86	53779.5	57564	0	53779.5	SUNDAY	
2273188	270658	Consumer loans	9644.22	26550	27252	0	26550	SATURDAY	
1232483	151612	Consumer loans	21307.455	126490.5	119853	12649.5	126490.5	TUESDAY	
2163253	154602	Consumer loans	4187.34	26955	27297	1350	26955	SATURDAY	
1285768	142748	Revolving loans	9000	180000	180000		180000	FRIDAY	
2393109	396305	Cash loans	10181.7	180000	180000		180000	THURSDAY	
1173070	199178	Cash loans	4666.5	45000	49455		45000	SATURDAY	
1506815	166490	Cash loans	25454.025	450000	491580		450000	MONDAY	
1182516	267782	Cash loans	20361.6	405000	451777.5		405000	SATURDAY	
1172842	302212	Cash loans		0	0			TUESDAY	
1172937	302212	Cash loans	39475.305	1129500	1277104.5		1129500	THURSDAY	
1555330	199353	Cash loans		0	0			SATURDAY	
1543131	275707	Cash loans	22619.52	229500	241920		229500	THURSDAY	
2536650	338725	Cash loans	16708.32	369000	369000		369000	WEDNESDAY	
1676258	433469	Cash loans	22242.825	247500	268083	0	247500	THURSDAY	

	Q	R	S	T	U	V	W	X	Y	Z	AA
1	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	OWN_CAR_AGE	FLAG_MOBIL	FLAG_EMP_PHONE	FLAG_WORK_PHONE	FLAG_CONT_MOBILE	FLAG_PHONE
2	0.018801	-9461	-637	-3648	-2120		1	1	0	1	
3	0.003541	-16765	-1188	-1186	-291		1	1	0	1	
4	0.010032	-19046	-225	-4260	-2531	26	1	1	1	1	
5	0.008019	-19005	-3039	-9833	-2437		1	1	0	1	
6	0.028663	-19932	-3038	-4311	-3458		1	1	0	1	
7	0.035792	-16941	-1588	-4970	-477		1	1	1	1	
8	0.035792	-13778	-3130	-1213	-619	17	1	1	0	1	
9	0.003122	-18850	-449	-4597	-2379	8	1	1	1	1	
10	0.018634	-20099	365243	-7427	-3514		1	0	0	1	
11	0.019689	-14469	-2019	-14437	-3992		1	1	0	1	
12	0.0228	-10197	-679	-4427	-738		1	1	0	1	
13	0.015221	-20417	365243	-5246	-2512		1	0	0	1	
14	0.031329	-13439	-2717	-311	-3227		1	1	1	1	
15	0.016612	-14086	-3028	-643	-4911	23	1	1	0	1	
16	0.010006	-14583	-203	-615	-2056		1	1	0	1	
17	0.020713	-8728	-1157	-3494	-1368	17	1	1	0	1	
18	0.018634	-12931	-1317	-6392	-3866		1	1	0	1	
19	0.010966	-9776	-191	-4143	-2427		1	1	0	1	
20	0.04622	-17718	-7804	-8751	-1259		1	1	0	1	
21	0.015221	-11348	-2038	-1021	-3964		1	1	1	1	
22	0.015221	-18252	-4286	-298	-1800	7	1	1	0	1	
23	0.025164	-14815	-1652	-2299	-2299	14	1	1	0	1	
24	0.020713	-11146	-4306	-114	-2518		1	1	0	1	
25	0.006296	-24827	365243	-9012	-3684		1	0	0	1	
26	0.026392	-11286	-746	-108	-3729	7	1	1	0	1	
27	0.028663	-19334	-3494	-2419	-2893		1	1	0	1	
28	0.018029	-18724	-2628	-6573	-1827		1	1	0	1	
29	0.019101	-15948	-1234	-5782	-3153		1	1	0	1	

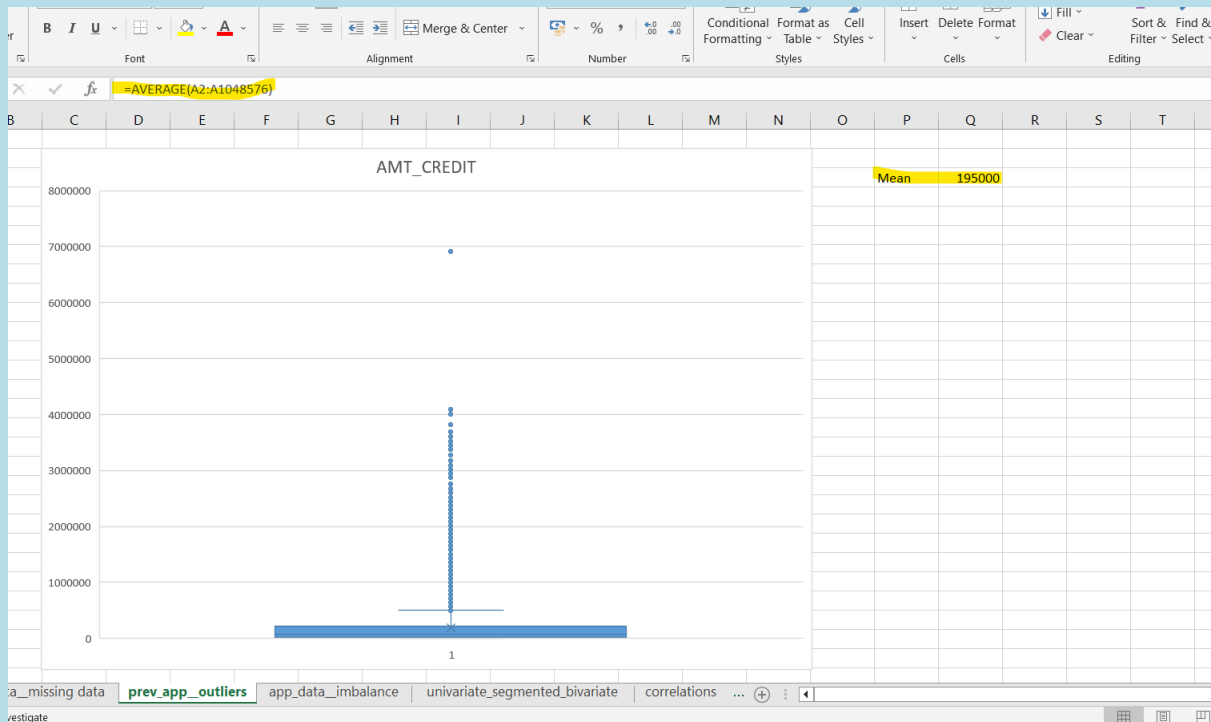
2. Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value)

I found out the blank percentage and median of the column and filled the empty spaces there. (This is just for one table. Actual cleaning and filling of data is shown in excel file attached for other columns).

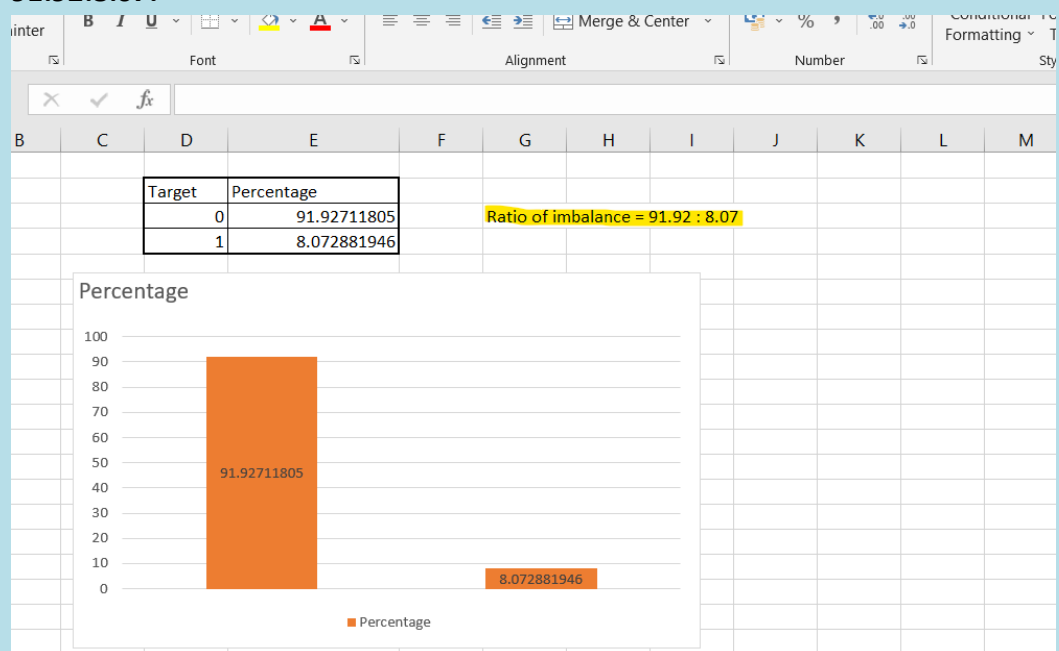


3. Identify if there are outliers in the dataset. Also, mention why do you think it is an outlier.

For Numerical columns, I found out the outliers and chose the value for the upper whisker as shown below. The credit amount value above 195000 is considered to be an upper whisker.



4. Identify if there is data imbalance in the data. Find the ratio of data imbalance. The ratio of imbalance for Target Table came out to be 91.92:8.07.

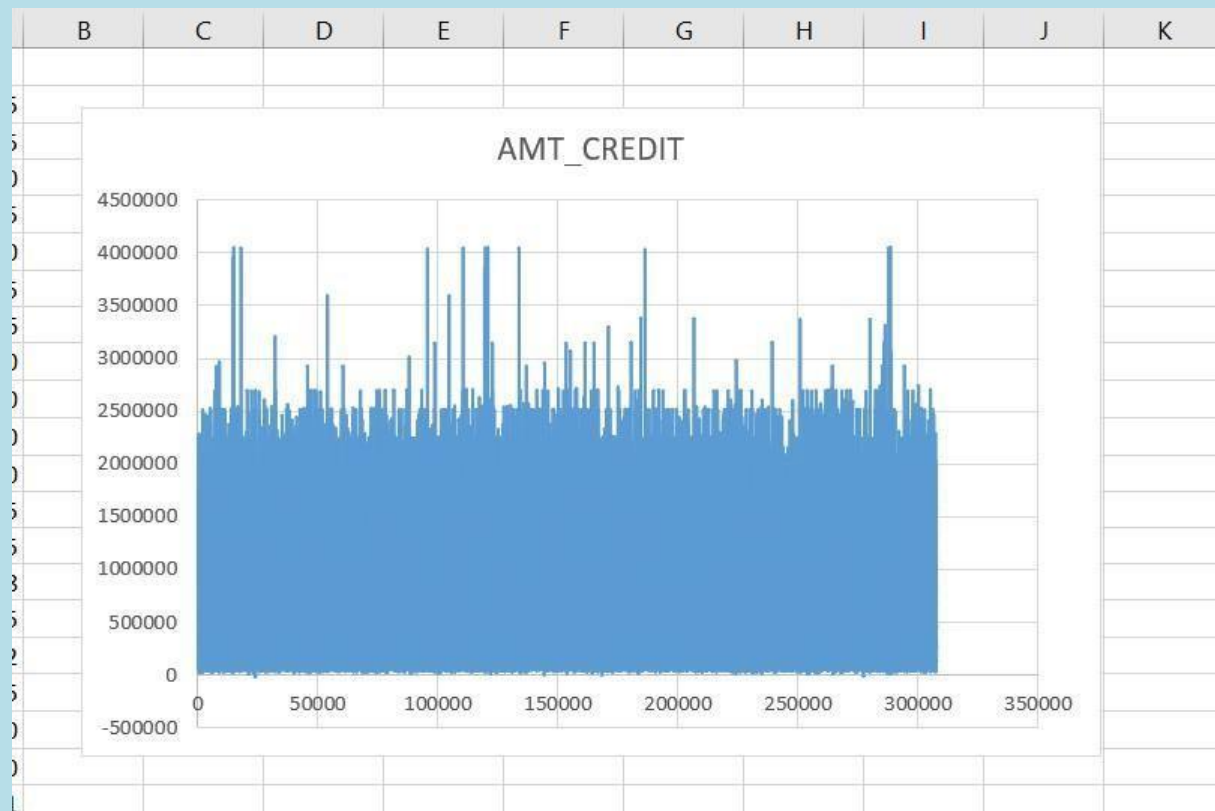


5. **Explain the results of univariate, segmented univariate, bivariate analysis, etc. in business terms.**

The results of univariate, segmented univariate, bivariate analysis are as follows –

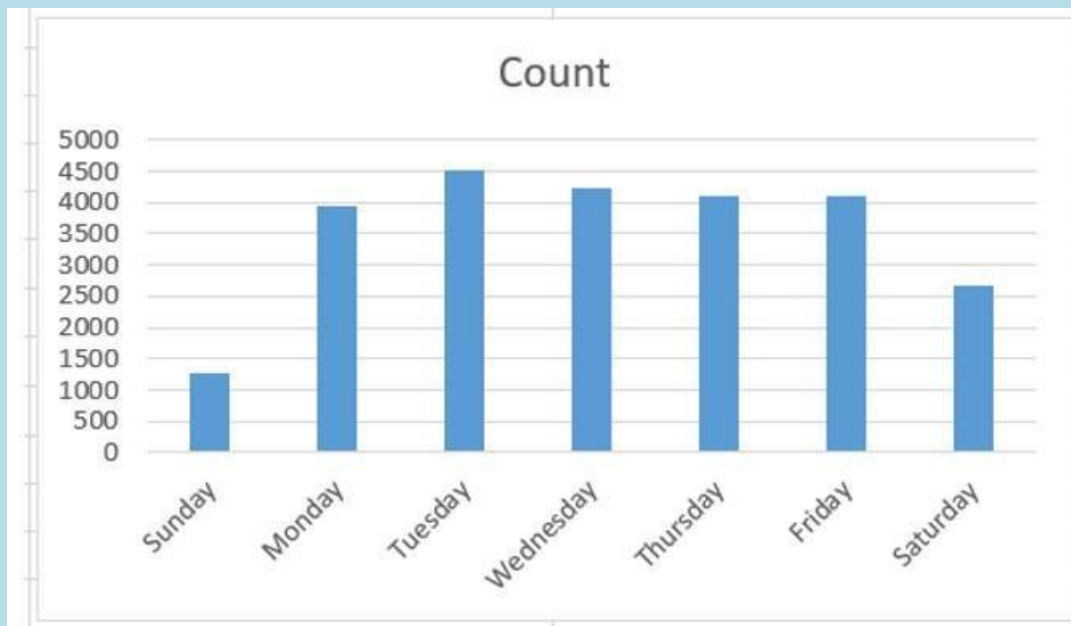
To perform the analysis, I first divided the data into two sets i.e. Target - 0 and Target – 1

AMT_CREDIT

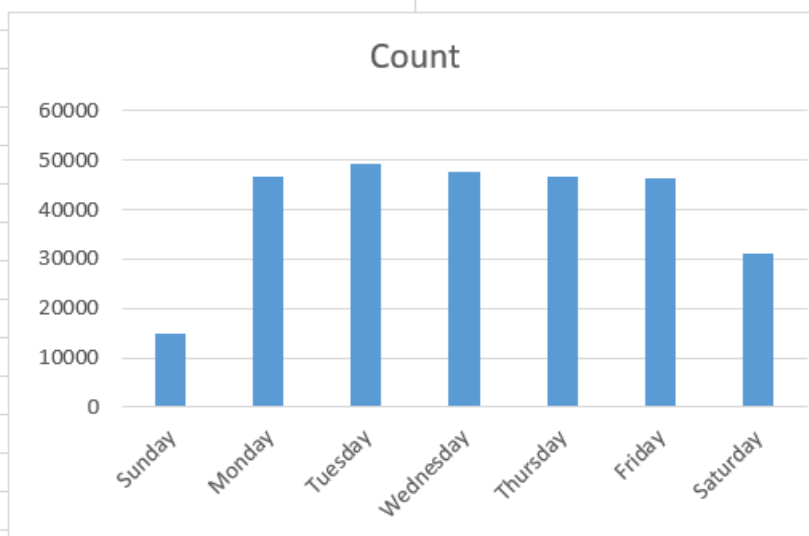


WEEKDAY_APPR_PROCESS_START

Target - 1	
WEEKDAY_APPR_PROCESS_START	Count
Sunday	1283
Monday	3934
Tuesday	4501
Wednesday	4238
Thursday	4098
Friday	4101
Saturday	2670



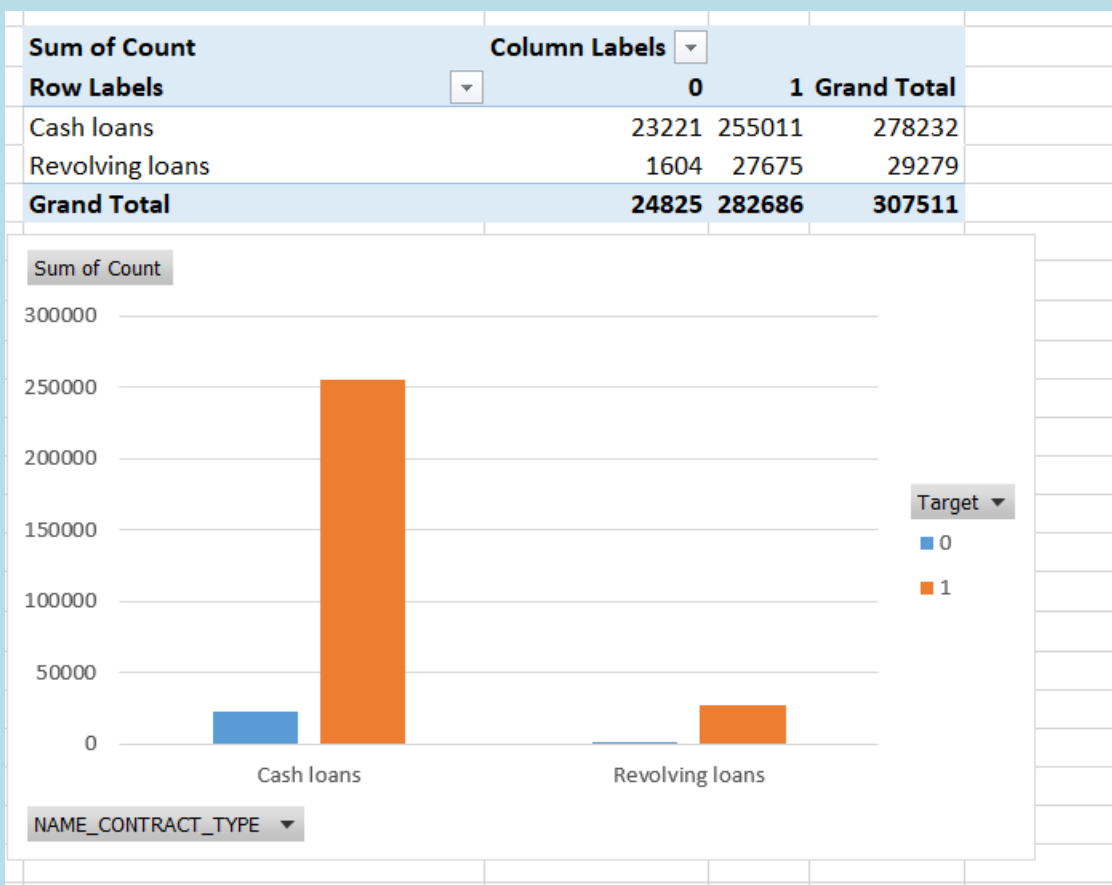
Target - 0	
WEEKDAY_APPR_PROCESS_START	Count
Sunday	14898
Monday	46780
Tuesday	49400
Wednesday	47696
Thursday	46493
Friday	46237
Saturday	31182



INSIGHTS – We can conclude that application starting process is less on Saturday and Sunday.

NAME_CONTRACT_TYPE

Target - 1			
NAME_CONTRACT_TYPE	Count		
Cash loans	23221		
Revolving loans	1604		
Target - 0			
NAME_CONTRACT_TYPE	Count		
Cash loans	255011		
Revolving loans	27675		
NAME_CONTRACT_TYPE	Count	Target	
Cash loans	255011	1	
Revolving loans	27675	1	
Cash loans	23221	0	
Revolving loans	1604	0	

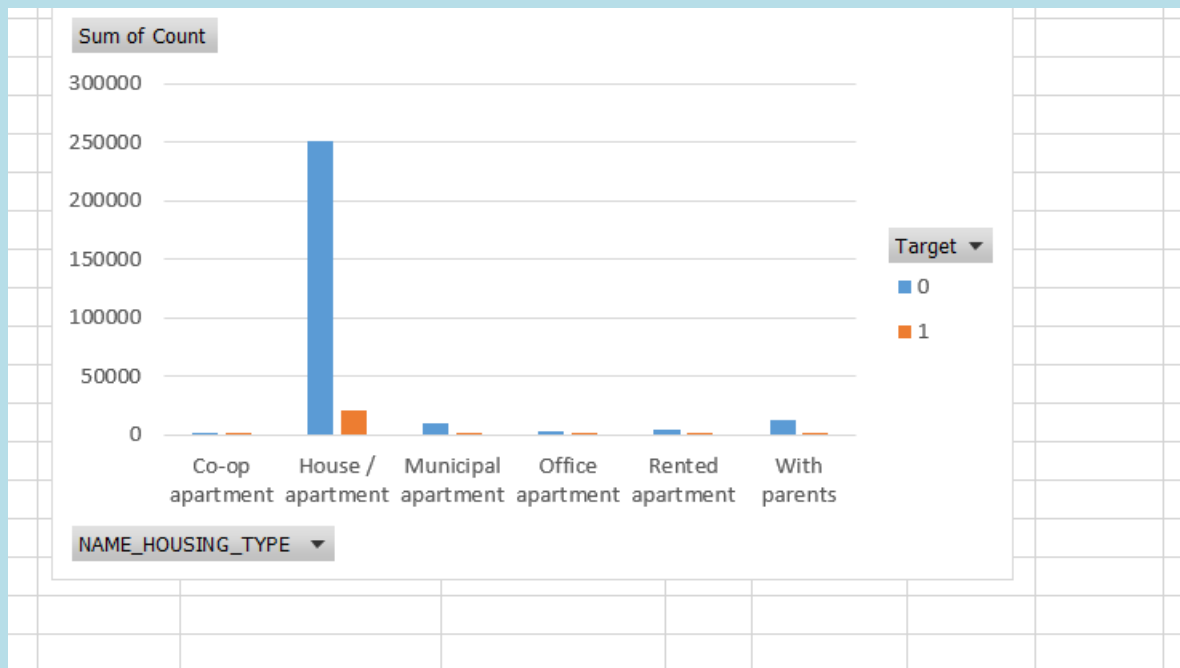


INSIGHTS – We can conclude that people prefer cash type loans more than other. People take more cash loans.

NAME_HOUSING_TYPE

Target - 1			
NAME_HOUSING_TY	Count		
House / apartment	21272		
Co-op apartment	89		
Municipal apartment	955		
Office apartment	172		
Rented apartment	601		
With parents	1736		
Target - 0			
NAME_HOUSING_TY	Count		
House / apartment	251596		
Co-op apartment	1033		
Municipal apartment	10228		
Office apartment	2445		
Rented apartment	4280		
With parents	13104		

NAME_HOUSING_TY	Count	Target		
House / apartment	21272	1		
Co-op apartment	89	1		
Municipal apartment	955	1		
Office apartment	172	1		
Rented apartment	601	1		
With parents	1736	1		
House / apartment	251596	0		
Co-op apartment	1033	0		
Municipal apartment	10228	0		
Office apartment	2445	0		
Rented apartment	4280	0		
With parents	13104	0		
Sum of Count		Column Labels		
Row Labels		0	1	Grand Total
Co-op apartment	1033	89		1122
House / apartment	251596	21272		272868
Municipal apartment	10228	955		11183
Office apartment	2445	172		2617
Rented apartment	4280	601		4881
With parents	13104	1736		14840
Grand Total	282686	24825		307511

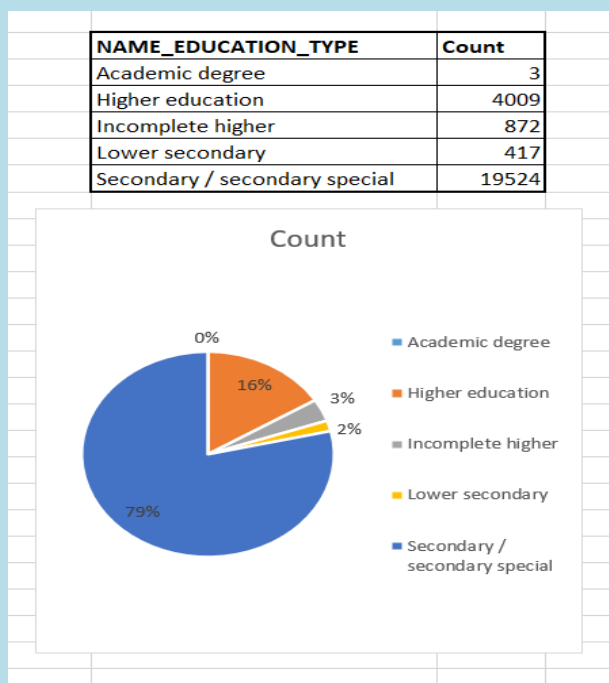


INSIGHTS – We can conclude that people living in houses fall in both the category of default loans and non-default loans.

- Find the top 10 **correlation** for the Client with payment difficulties and all other cases (Target variable).

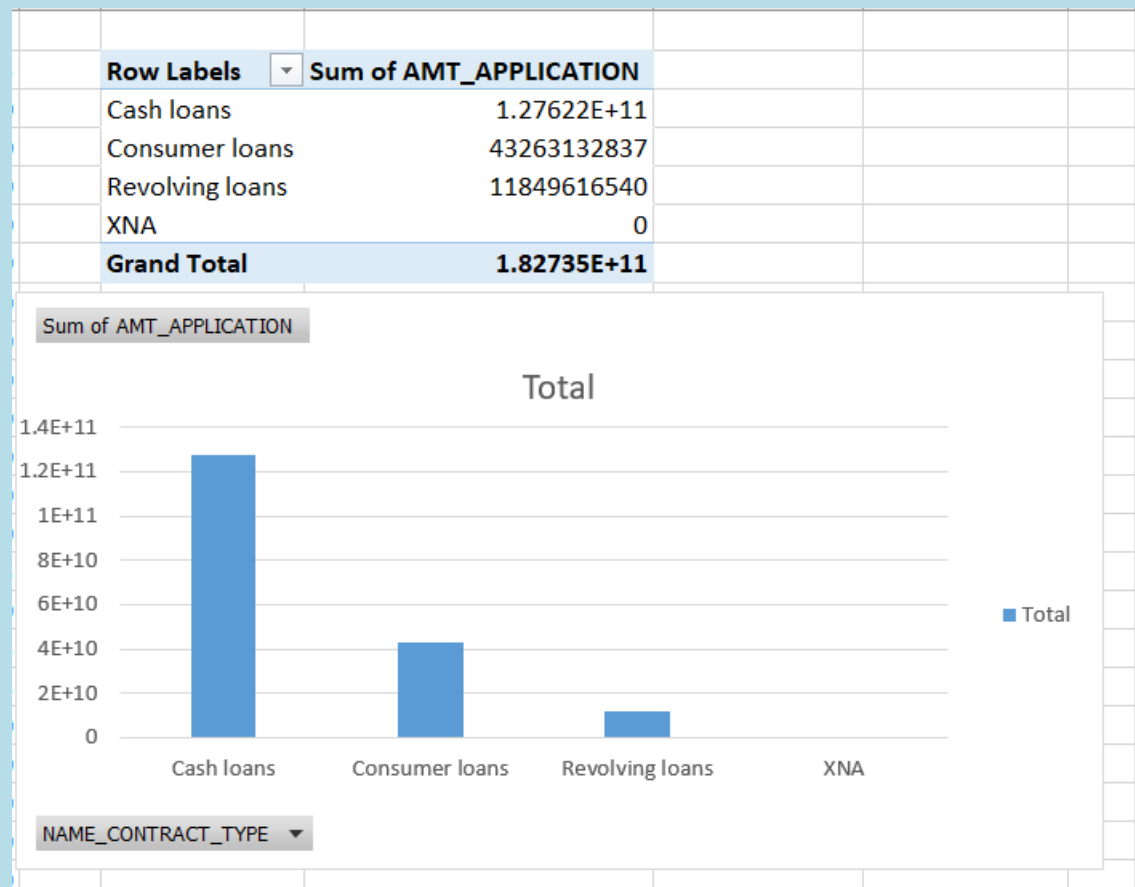
To find the correlation, we again divide the data into two sets based on Targets and consider Target – 1 as defaulters.

NAME_EDUCATION_TYPE



INSIGHTS – We can find that people with education type as Secondary/Secondary Special are more likely to default and people with education type Academic degree default the least.

NAME_CONTRACT_TYPE and AMT_APPLICATION



INSIGHTS – If we sum the total amount for loan in applications, we find that that people mostly take cash loans.

AMT_APPLICATION and AMT_CREDIT

AMT_APPLICATION	AMT_CREDIT			
17145	17145		Correlation Coefficient	
607500	679671		0.975777217	
112500	136444.5			
450000	470790			
337500	404055			
315000	340573.5			
0	0			
0	0			

INSIGHTS – We find that the correlation coefficient is 0.9758 using excel formula =CORREAL.

AMT_INCOME_TOTAL and AMT_ANNUITY

W	X	Y	Z	
AMT_INCOME_TOTAL	AMT_ANNUITY			
202500	24700.5		Correlation Coefficient	
270000	35698.5		0.191657428	
67500	6750			
135000	29686.5			
121500	21865.5			
99000	27517.5			
150000	31000			

INSIGHTS – We find that the correlation coefficient is 0.19166 using excel formula =CORREL.

CONCLUSION – From the above analysis, we can find out what kind of people and can repay loan, what kinds of loan people prefer to take, people taking loans come from what background, what is their source of income, for what type of people, the loan applications are refused and based on which conditions.

RESULTS: -

1. People with academic degree have less defaults.
2. People prefer cash loans more than any other type.
3. People with secondary/secondary special as education type have more chances of defaulting loans.
4. People who have less than 5 years of employment have high default rate.
5. Focused variable for application file – Target.
6. Focused variable for Previous application file – NAME_CONTRACT_STATUS.
7. Important fields to consider for loan repayment are –
8. NAME_EDUCATION_TYPE
9. AMT_INCOME_TOTAL
10. DAYS_EMPLOYED
11. AMT_CREDIT
12. People with lower total income are more likely to default.
13. People with high Credit amount are less likely to default.

Module -7 Project: Impact of Car Features

The automotive industry has been evolving rapidly, focusing on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition, manufacturers seek to optimize pricing and product development to maximize profitability while meeting consumer demand.

This project analyses various car features that impact manufacturers' pricing decisions. The dataset includes multiple car models with their specifications, such as engine power, transmission type, fuel efficiency, and body style. The objective is to identify key factors influencing car prices and how they vary across manufacturers and market segments.

The dataset used for analysis is available on Kaggle (provided by Cooper College, New York City) and contains **11,813 observations** in **CSV format**. The analysis was performed using **MS Excel 2019**, utilizing pivot tables, formulas, and visualization techniques.

Project Approach and Key Insights:

Q1. How does the popularity of car models vary across different market categories?

- The analysis reveals that **Hatchback and Flex Fuel** cars have the highest number of models, making them the most dominant in the dataset.
- This insight helps manufacturers focus on these market segments for potential growth.

Q2. What is the relationship between a car's engine power and its price?

- A **strong positive correlation** exists between engine power and price.
- High-powered cars generally have higher prices, indicating that performance plays a key role in pricing strategy.

Q3. Which car features most impact a car's price?

- **Engine Cylinders** have the strongest influence on the Manufacturer's Suggested Retail Price (MSRP).
- Cars with more cylinders tend to have higher prices, as they often offer greater power and performance.

Q4. How does the average car price vary across different manufacturers?

- **Bugatti has the highest average car price**, significantly above other brands.
- This suggests that premium and luxury car brands set prices based on exclusivity and high-performance features.

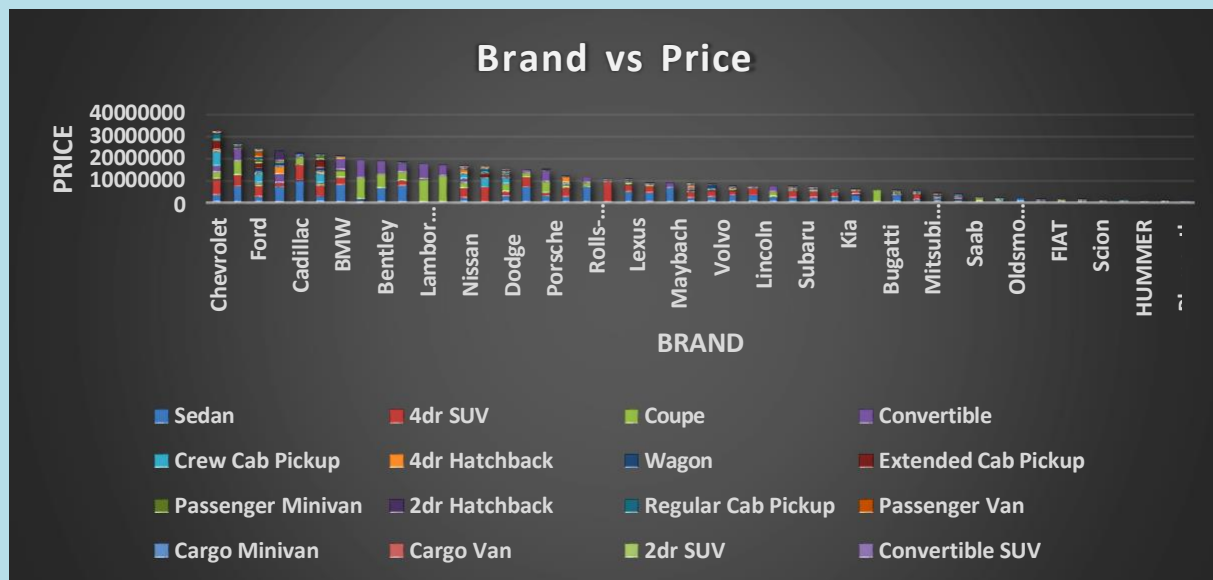
Q5. What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

- Fuel efficiency **decreases as the number of cylinders increases**.
- Cars with **4 cylinders** offer the best fuel efficiency, making them ideal for consumers prioritizing mileage over power.

Dashboard Insights & Data Visualization:

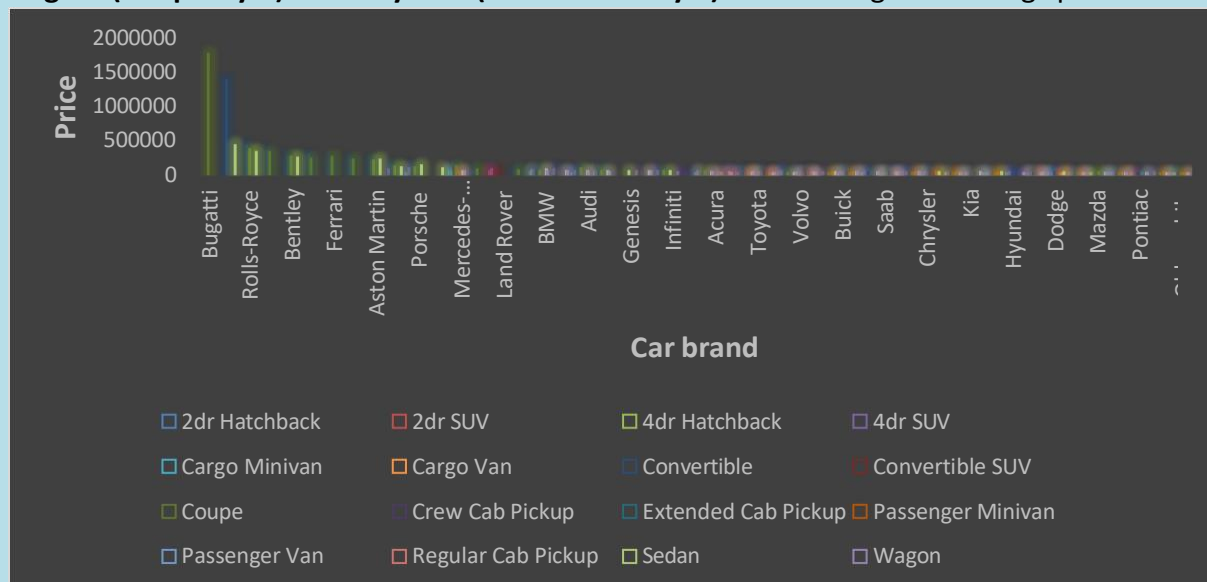
Task 1: Distribution of car prices by brand and body style

- Chevrolet and Mercedes-Benz contribute the most to the overall car price distribution.



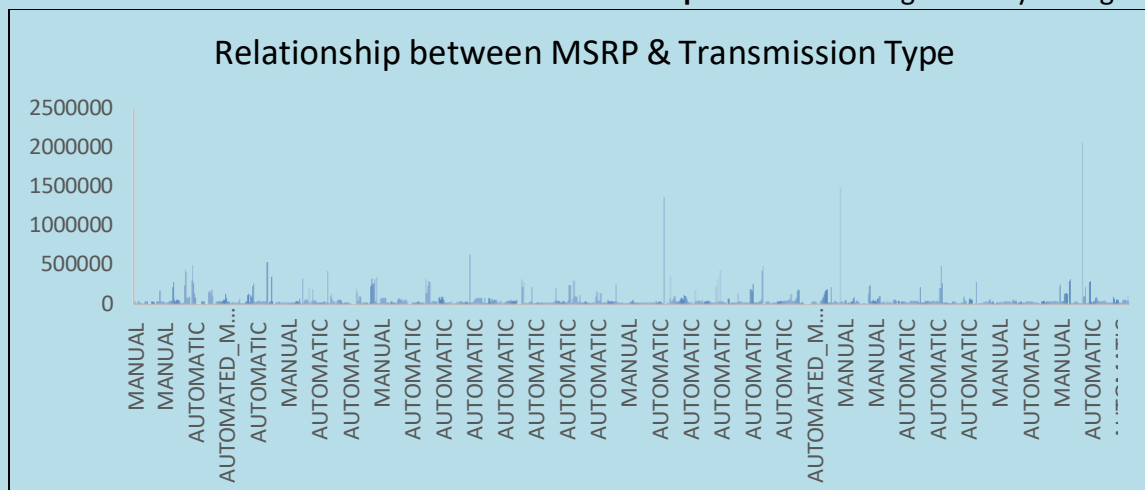
Task 2: Which brands have the highest and lowest average MSRPs, and how does this vary by body style?

- Bugatti (Coupe style) and Maybach (Convertible style) have the highest average prices.



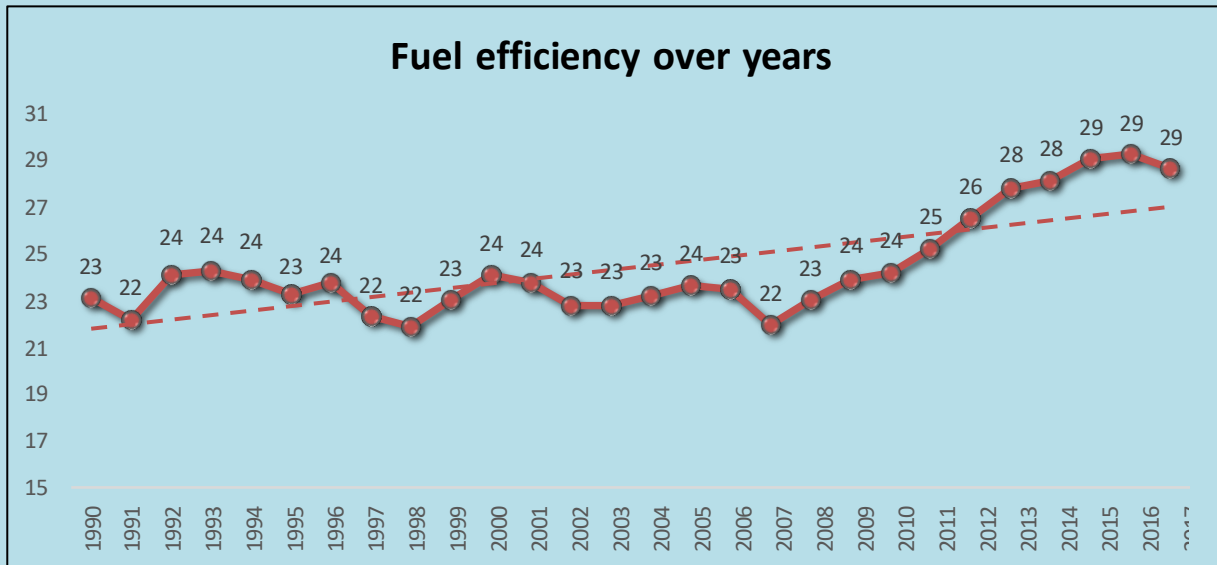
Task 3: Impact of transmission type on MSRP across body styles

- Automated convertible and automated manual coupes contribute significantly to high MSRPs.



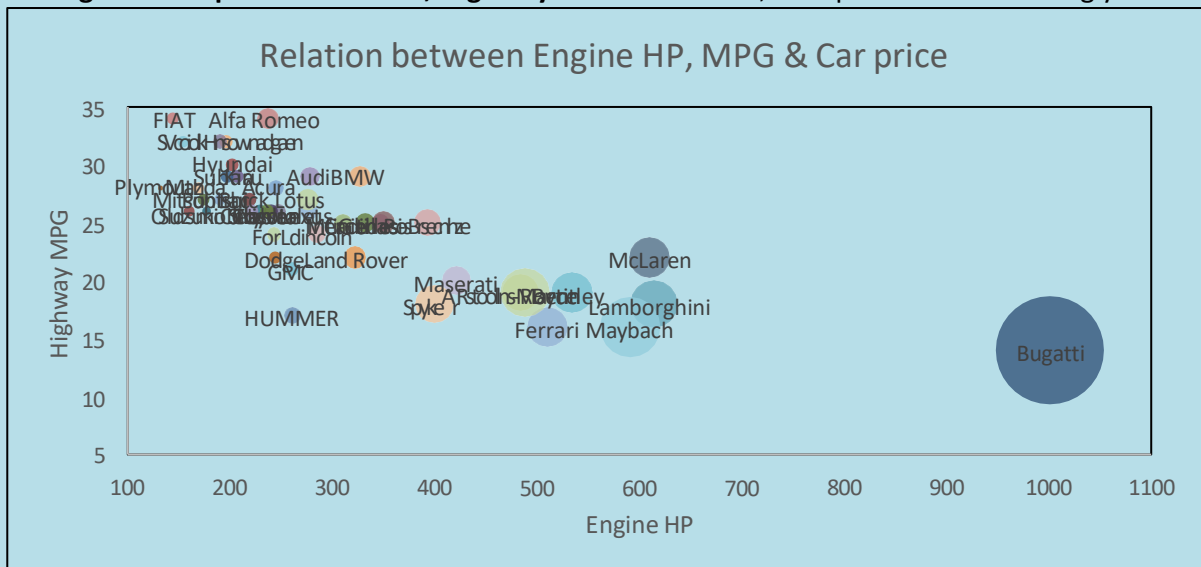
Task 4: Fuel efficiency trends over time

- Fuel efficiency (**Highway MPG**) has improved over time, highlighting advancements in automotive technology.



Task 5: How horsepower, MPG, and price vary across different brands

- As **engine horsepower increases**, **highway MPG decreases**, and prices rise accordingly.



Final Results & Recommendations:

- Coupe body style** has the highest contribution to MSRP.
- Automated Manual Transmission** significantly impacts pricing due to its dual benefits of manual and automatic gear systems.
- Fuel efficiency is a crucial factor** for affordability; manufacturers should focus on balancing power and mileage.
- Companies should prioritize fuel-efficient models to cater to a broader market while maintaining profitability.

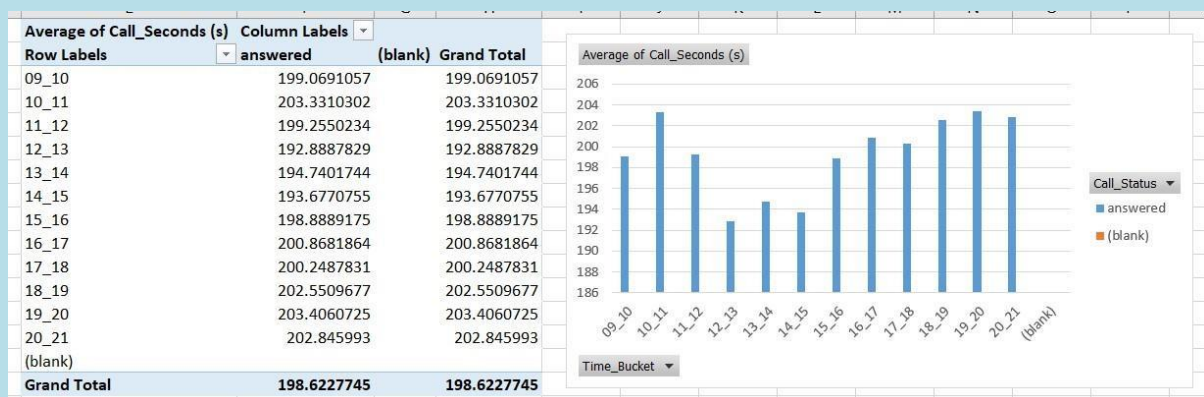
Module -8 Project: ABC Call Volume Trend

This project is based on call volume report analysis of a call center. ABC is a call center which has a separate team for voice process. In this project we are provided with dataset having details of agents, call duration, time duration, details on calls answered, abandoned and transferred. Data set also contains customer phone number, queue time, IVR time, date and time of call.

Here we have to analyze the rate of call which went unanswered and how many more agents are required to answer the call in both day and night shifts.

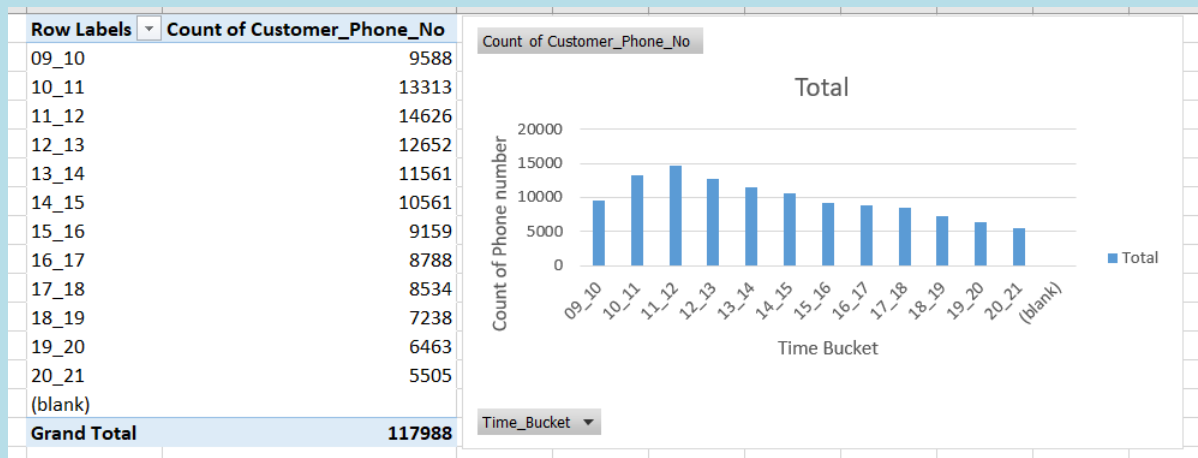
In this dataset, I first went through the data set to understand the details of the different variables and columns. I checked for any null values, missing or blank cells, duplicate data or if any data cleaning is required. After checking all these fields, I went up to perform the data analysis and answer the required questions. To perform the analysis, I used MS Excel 2019 for analysis and used MS Word to prepare the report.

Q1.) Calculate the average call time duration for all incoming calls received by agents (in each Time_Bucket).



- Average of call time duration answered by agents is 198.62 seconds.
- Average of call duration is highest between 10 to 11 am and 7 to 8 pm.

Q2.) 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,)



- The number of calls increases from 9 am to 12 noon and then decreases.

Q3.) 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 have to calculate minimum number of agents required in each time bucket so that at least 90 calls should be answered out of 100.)

Description	unit	
Total no of working days per week	days	6
Total unplanned leaves per month	days	4
Total working hours	hrs	9
Time spent in lunch and snacks	hrs	1.5
Actual working hours	perc	60%
Avg time agent is occupied (60% of 7.5 hrs) daily	hrs	4.5
Total no. of days in a month	days	30

Time taken on average to answer calls (sec)	199.314176
Time requirement to answer 90% of the calls (hrs)	255.6204308
Total working person required per day	57

Day Calls	Avg answered calls	total calls	total calls in %	No. of agents required
09_10	199.0691057	9588	8.13%	5
10_11	203.3310302	13313	11.28%	6
11_12	199.2550234	14626	12.40%	7
12_13	192.8887829	12652	10.72%	6
13_14	194.7401744	11561	9.80%	6
14_15	193.6770755	10561	8.95%	5
15_16	198.8889175	9159	7.76%	4
16_17	200.8681864	8788	7.45%	4
17_18	200.2487831	8534	7.23%	4
18_19	202.5509677	7238	6.13%	3
19_20	203.4060725	6463	5.48%	3
20_21	202.845993	5505	4.67%	3
Average	199.314176		100.00%	57

Count of Duration(hh:r		Call Status			
Days	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	
Grand Total	34403	82452	1133	117988	
	1496	3585	49	5130	
	29.16%	69.88%	0.96%		
	30.00%	70%	1%		

- Total agents required to answer 90% of calls per day is 57.
- The amount of answered calls are 70%, abandon are 30% and transferred are 1% approximately.

Q4.) 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:

Distribution of 30 calls coming in night for every 100 calls coming in between 9am - 9pm (i.e. 12 hrs slot)											
9pm- 10pm	10pm - 11pm	11pm- 12am	12am- 1am	1am - 2am	2am - 3am	3am - 4am	4am - 5am	5am - 6am	6am - 7am	7am - 8am	8am - 9am
3	3	2	2	1	1	1	1	3	4	4	5

Now propose a manpower plan required during each time bucket in a day. Maximum Abandon rate assumption would be same 10%.

Call volume daily (9 AM - 9pm)	5130	
If we provide support in night, (9 PM - 9 AM)	1539	
Additional hours required	76.68613	
Additional HC	17	
Total HC	74	
Note - all values are taken from previous sheet		
4.5 - avg time agent is occupied - taken from prev sheet		

Nights Call	Calls Distribution	Time Distribution	Agents Required
21_22	3	10%	2
22_23	3	10%	2
23_24	2	7%	1
00_01	2	7%	1
01_02	1	3%	1
2_3	1	3%	1
3_4	1	3%	1
4_5	1	3%	1
5_6	3	10%	2
6_7	4	13%	2
7_8	4	13%	2
8_9	5	17%	3
	30	100%	17

- First calculated the Time Distribution by dividing each calls distribution by total calls i.e. 30.
- Total agents required to answer 90% of calls at night is 17.

CONCLUSION

After performing all the projects, I came to know about the real-world application of this data. I came to know how these big product companies use data driven insights to get best results in minimum time. I came to know about the various applications and importance of tools like SQL, MS Excel, MS Word, MS PowerPoint, use of formulas, logics, commands, Pivot Tables, graphs etc. I also came to understand how meaningful results can be drawn from such huge amount of data if they are properly sorted and cleaned. They can give the most accurate insights about the operations of a company and what improvements can be made in order to grow.

THANK YOU!