DATA ANALYSIS PORTFOLIO

SOWMIYA R

PROFESSIONAL BACKGROUND

Currently in final year pursuing B.Tech. Information Technology. I have secured 8.78 CGPA (till 6th Sem) and have several skills including Data Analysis, Python, MySQL, Excel. I have worked on various personal projects related to Data Analysis, Excel and Python.

Also, I have participated in various challenges and competitions on Hacker rank. Also, I have worked on Department Student Record, Counselling and Mentor system as part of mini project in Semester 5. Currently, I am working on implementation of mini project which is a group project for the Final year.

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'm excited to embark on a career that allows me to apply my knowledge and skills in a meaningful way. In my free time. And I believe by putting significant efforts, I will learn.

TABLE OF CONTENTS

Catalog

Click here to enter text.

IMDB Movie Analysis

Bank Loan Case Study

Impact Of Car Features

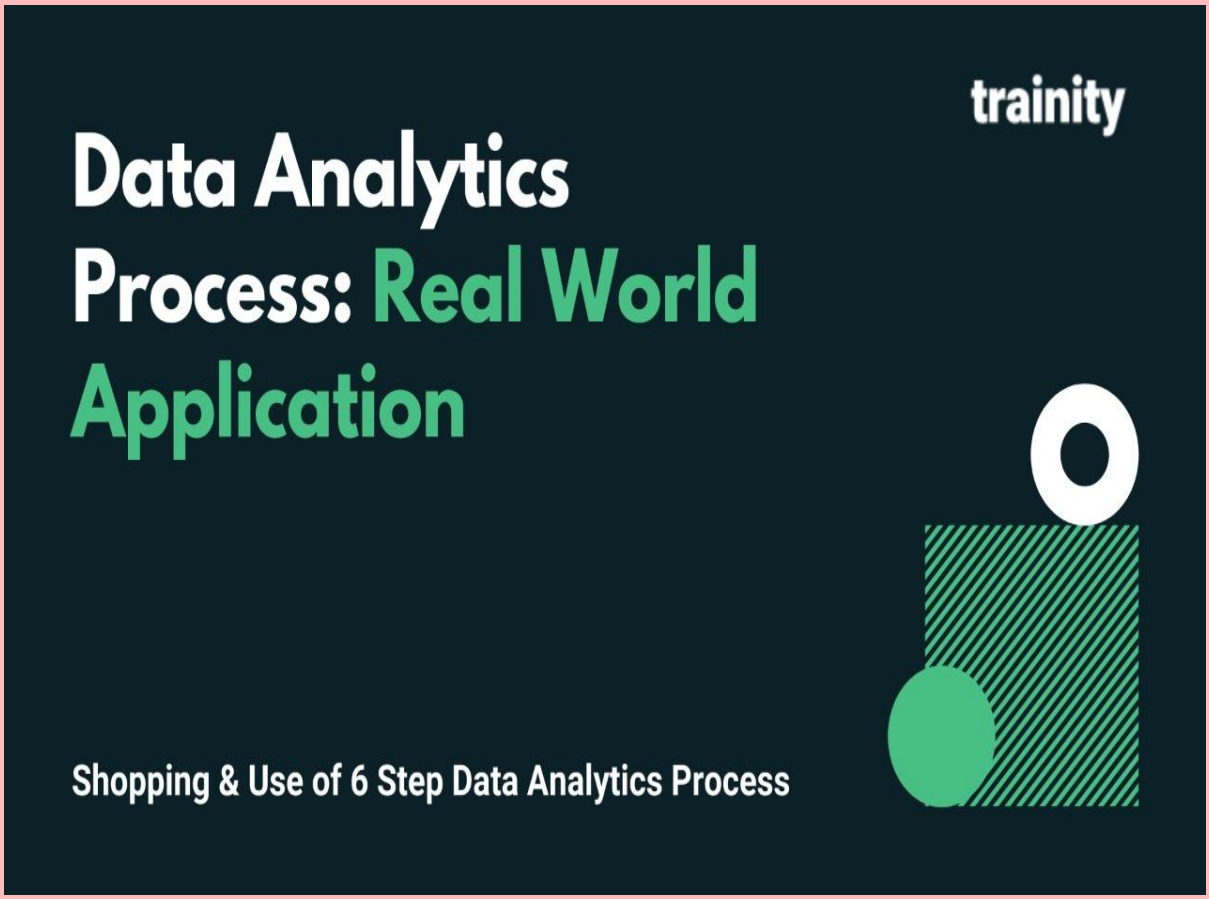
ABC Call Volume Trend

1. DATA ANALYTICS PROCESS

DESCRIPTION

We use Data Analytics in everyday life without even knowing it.

For eg : Going to a market to buy something .



DESIGN

Tour package suppliers are continually looking for ways to improve customer experiences, optimize services, and keep ahead of market trends in an increasingly competitive travel business. By employing data-driven insights to make educated decisions, data analytics has emerged as a valuable tool for achieving these aims.

HERE THE PROCESS ARE

Planning:

• Analyze the needs of the customer, including where they are satisfied, costeffective, for lodging, sightseeing, and food.
• Where is it safe to travel?
• Which package is both affordable and fun?

Prepare:

Collecting the information linked to customer requirement to serve the purpose :

• By searching the good location
• By finding the good accomodation
• By searching the transportation and food

Which are affordable to customer.

Process:

By using these information I need to process all of this information and make judgments after receiving it.

This includes by comparing the:

- cost and travel distance
- Quality of service
- Quantity of service
- Safety measures

So these are things used in the process method.

Analyze:

Here we are analyzing the data by comparing the place by using the data table:

PLACE COST TRAVEL TIME DISTANCE NUMBER OF SIGHTSEEING

PLACE

1. From chennai to Wayanad 3500 11hr 16min 614km 8
2. From chennai to Alapuzha 4000 14hr 33min 736km 10
3. From chennai to goa 4500 16hr 36min 911km 9
4. From chennai to bangalore 3000 7hr 13min 346km 5
5. From chennai to manali 8000 48hr 2687km 13

Share:

By sharing the analyzed data to customer to choose their tour package and also need to get feedback.

Act:

Now they can choose their tour package which they can afford easily with the good quality of a service and ending with the payment for the tour package by using UPI option, g pay or any other online payment option.

Conclusion:

- Finally customer satisfies with good quality of service.
- They found the affordable and enjoyable quantity of service.
- They found satisfied tour package.

Data Analytics Steps

i. Plan

ii. Prepare

iii. Process

iv. Analyze

v. Share

vi. Act

INSTAGRAM USER ANALYTICS

DESCRIPTION

Imagine you're a data analyst working with the product team at Instagram. Your role involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

User analysis involves tracking how users engage with a digital product, such as a software application or a mobile app. The insights derived from this analysis can be used by various teams within the business. For example, the marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience.

In this project, you'll be using SQL and MySQL Workbench as your tool to analyze Instagram user data and answer questions posed by the management team. Your insights will help the product manager and the rest of the team make informed decisions about the future direction of the Instagram app.

Remember, the goal of this project is to use your SQL skills to extract meaningful insights from the data. Your findings could potentially influence the future development of one of the world's most popular social media platforms.

The Problem

A) Marketing Analysis:

1. **Rewarding Most Loyal Users:** Finding the 5 oldest users of the Instagram from the database provided with the joined date.

2. **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo. Find the users who have never posted a single photo on Instagram.

3. **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner. Identify the winner of the contest and provide their details to the team.

4. **Hash tag Researching:** A partner brand wants to know, which hash tags to use in the post to reach the most people on the platform. Identify and suggest the top 5 most commonly used hash tags on the platform.

5. **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs. What day of the week

do most users register on? Provide insights on when to schedule an ad campaign.

B) Investor Metrics:

1. **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts. Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.
2. **Bots & Fake Accounts**: The investors want to know if the platform is crowded with fake and dummy accounts. Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

DESIGN

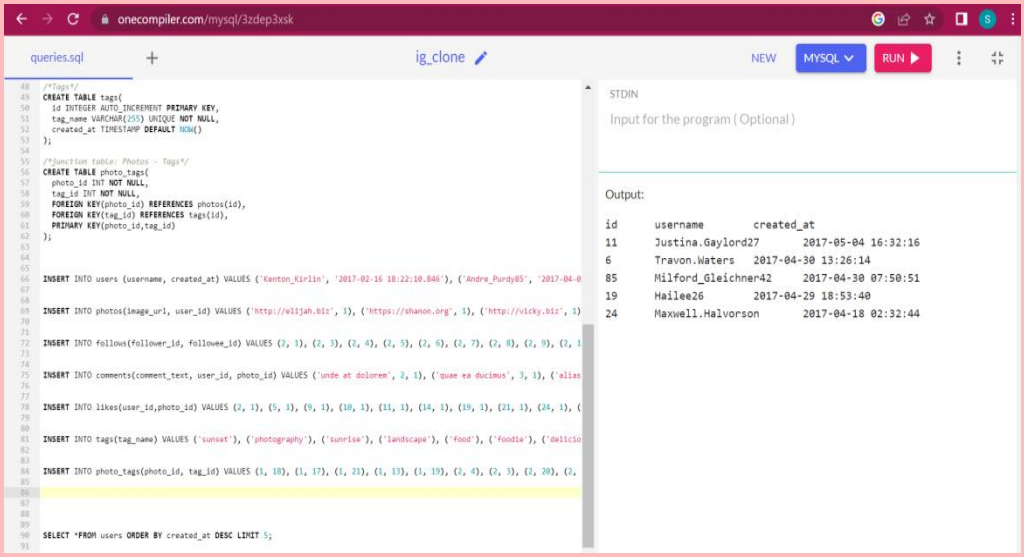
- Using the 'create db' function of MySQL create a database
- Then add tables and column names
- Then add the values into them using the 'insert into' function of MySQL
- By using the 'select' command we can query the desired output

--> MySQL Workbench 8.0 CE

Findings 1

1.**Rewarding Most Loyal Users:** Finding the 5 oldest users of the Instagram from the database provided with the joined date.

```
SELECT *FROM users

ORDER BY created_at DESC LIMIT 5;
```



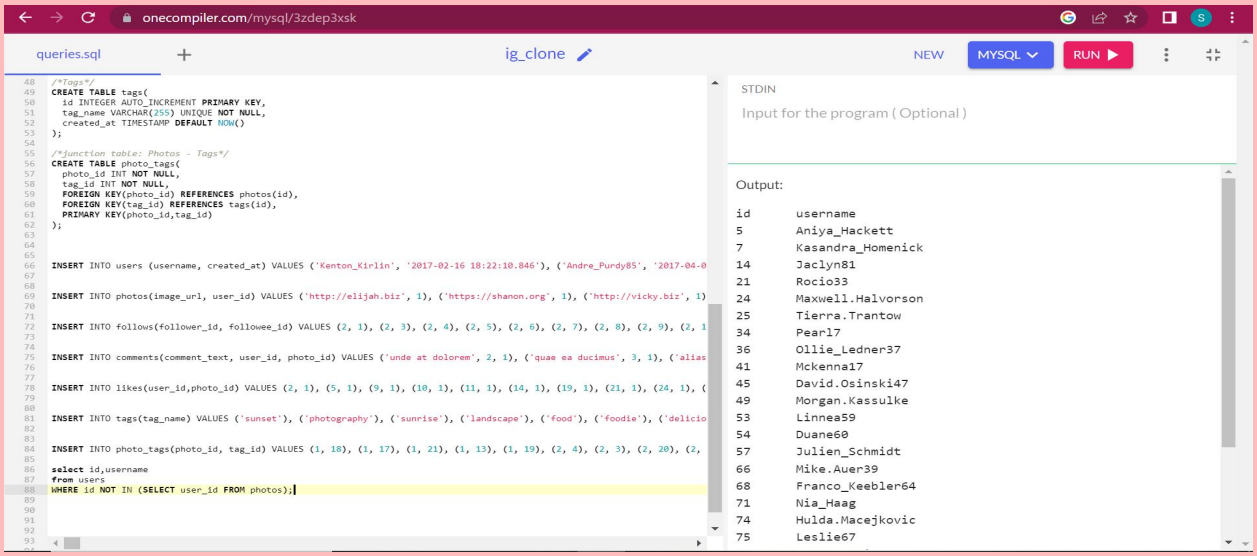
Oldest 5 users joining date with their username and user id:

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

Findings 2

2.Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo. Find the users who have never posted a single photo on Instagram.

```
select id,username

from users

WHERE id NOT IN (SELECT user_id FROM photos);
```



There are total 26 users who never posted a photo on the platform:

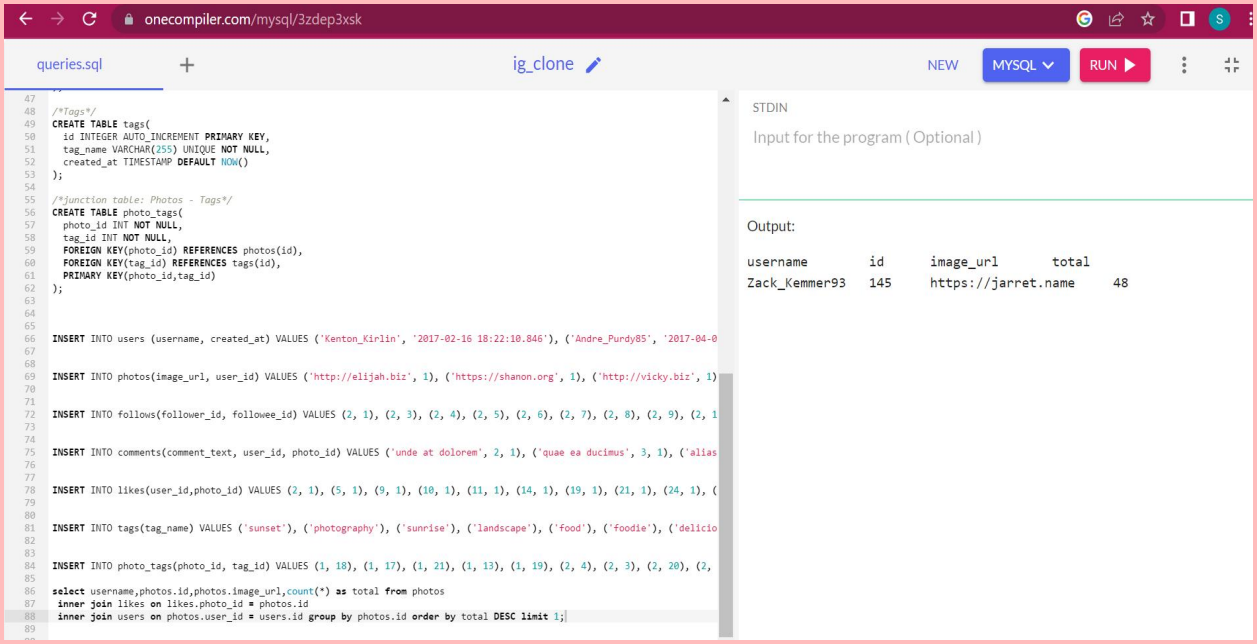
id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59

54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

Findings 3

3. **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner. Identify the winner of the contest and provide their details to the team.

```
SELECT username, photos.id, photos.image_url, count(*) as total
FROM photos inner join likes on likes.photo_id = photos.id

inner join users on photos.user_id = users.id

GROUP BY photos.id

ORDER BY total DESC LIMIT 1;
```



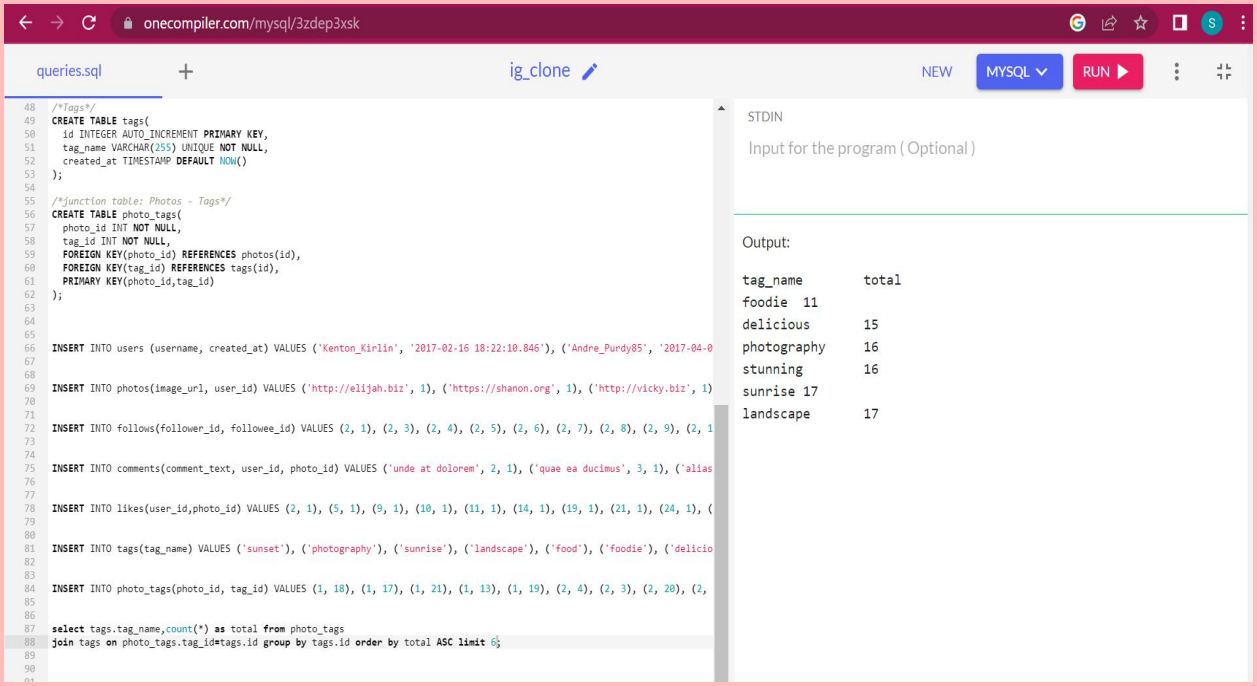
User with ID: Zack_kemmer93 has won the contest with 48 likes for a single photo he had posted.

username	id	image_url	total
Zack_Kemmer93	145	https://jarret.name	48

Findings 4

4. **Hash tag Researching:** A partner brand wants to know, which hash tags to use in the post to reach the most people on the platform. Identify and suggest the top 5 most commonly used hash tags on the platform.

```
SELECT tags.tag_name,count(*) as total FROM photo_tags

join tags on photo_tags.tag_id=tags.id

GROUP BY tags.id

ORDER BY total ASC LIMIT 6;
```

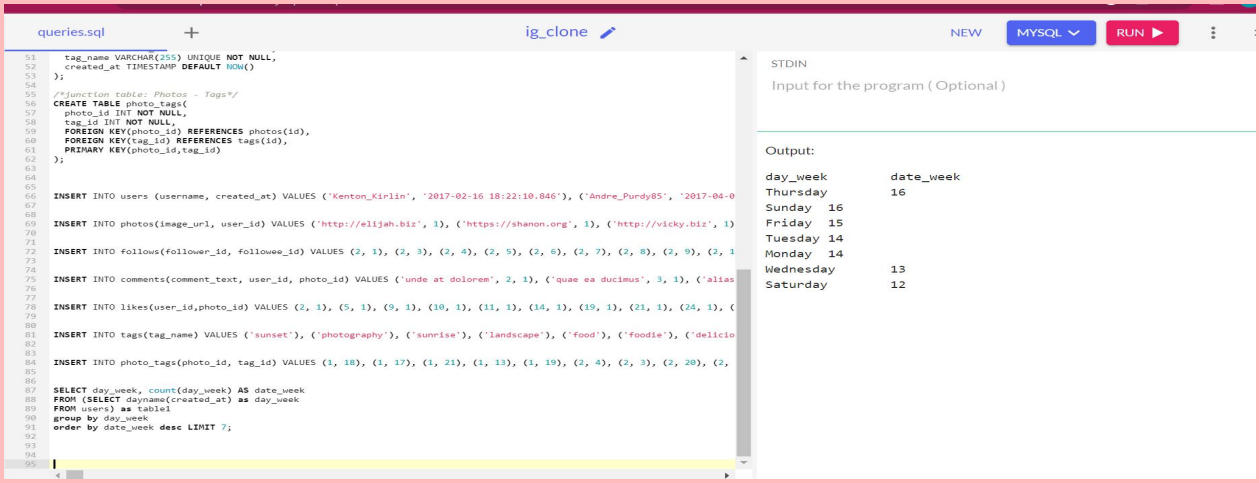


tag_name	total
foodie	11
delicious	15
photography	16
stunning	16
sunrise	17
landscape	17

Findings 5

5. **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs. What day of the week do most users register on? Provide insights on when to schedule an ad campaign.

```
SELECT day_week, count(day_week) AS date_week
FROM (SELECT dayname(created_at) AS day_week
FROM users) as table1
GROUP BY day_week
ORDER BY date_week DESC LIMIT 7;
```



day_week	date_week
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

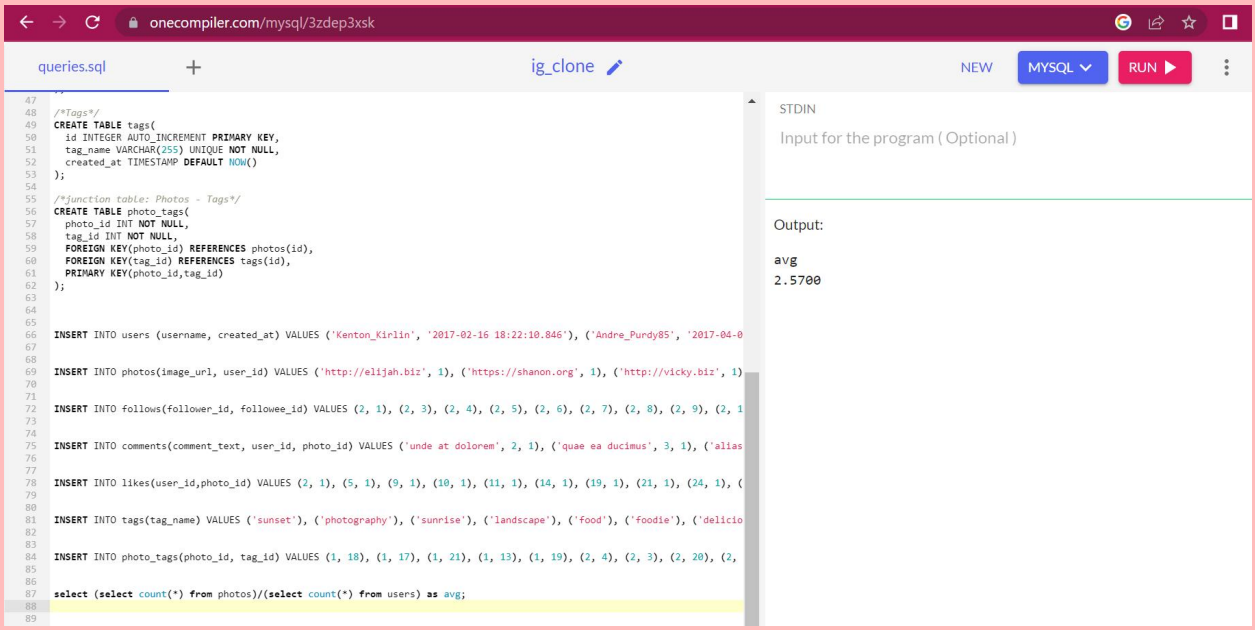
Thursdays and Sundays are two days with most user signup on Instagram.

Findings 6

1. **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts. Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.

Total number of photos on Instagram / Total number of users

```
SELECT (SELECT count(*) FROM photos)/(SELECT count(*) FROM users) AS average;
```

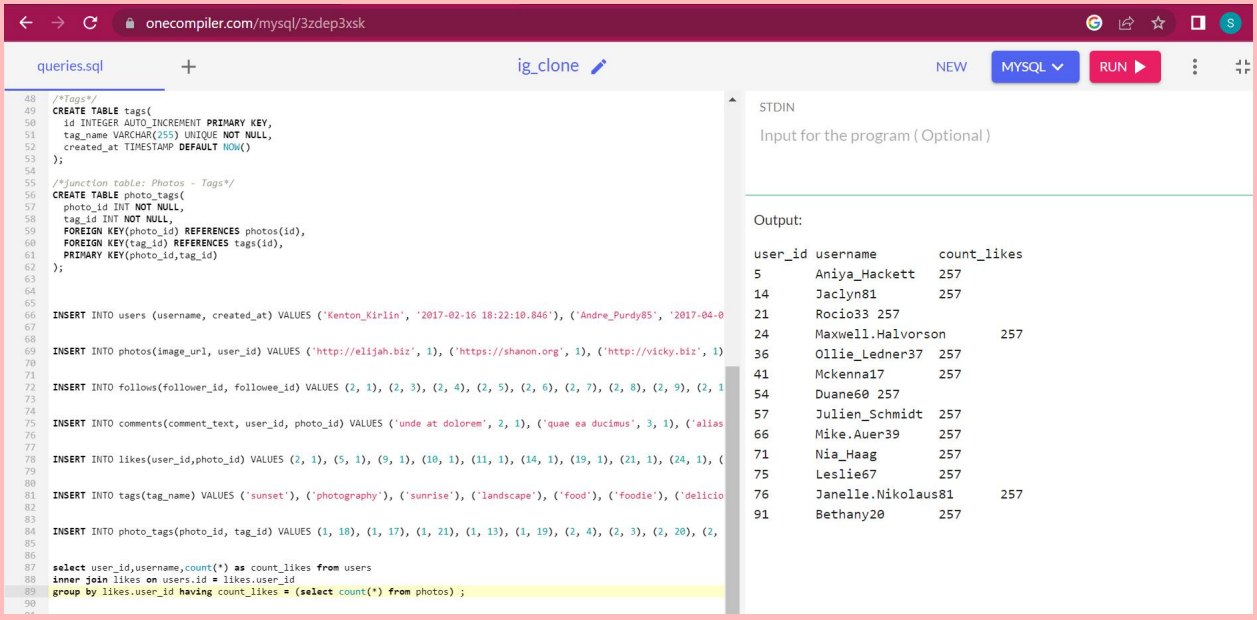


average
2.5700

Findings 7

2. **Bots & Fake Accounts**: The investors want to know if the platform is crowded with fake and dummy accounts. Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

```
SELECT user_id, username, count(*) AS count_likes FROM users

inner join likes on users.id = likes.user_id

GROUP BY likes.user_id

HAVING count_likes = (SELECT count(*) FROM photos) ;
```



user_id	username	count_likes
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

CONCLUSION

I would like to conclude that not only Instagram but many other social media and commercial firms use such Analysis to find the insights from their customer data which in turn help the firms to find the customers who will be an Asset to the firm in the future and not some Liability.

Such Analysis and sorting of the customer base is done at an weekly, monthly, quarterly or yearly basis as per the needs of the business firms so as to maximize their profits in future with minimal cost to the company.

Instagram 

Operation Analytics and Investigating Metric Spike

(Advanced sql)

DESCRIPTION

Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. As a Data Analyst, you'll work closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect.

One of the key aspects of Operational Analytics is investigating metric spikes. This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales. As a Data Analyst, you'll need to answer these questions daily, making it crucial to understand how to investigate these metric spikes.

In this project, you'll take on the role of a Lead Data Analyst at a company like Microsoft. You'll be provided with various datasets and tables, and your task will be to derive insights from this data to answer questions posed by different departments within the company. Your goal is to use your advanced SQL skills to analyze the data and provide valuable insights that can help improve the company's operations and understand sudden changes in key metrics.

The Problem

Case Study 1 (Job Data)

- **Number of jobs reviewed:** Amount of jobs reviewed over time.
- **Throughput:** It is the no. of events happening per second.
- **Percentage share of each language:** Share of each language for different contents.
- **Duplicate rows:** Rows that have the same value present in them.

Case Study 2 (Investigating metric spike)

- **User Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service.
- **User Growth:** Amount of users growing over time for a product.
- **Weekly Retention:** Users getting retained weekly after signing-up for a product.
- **Weekly Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.
- **Email Engagement:** Users engaging with the email service.

Design

- Using the 'create db' function of MySQL create a database
- Then add tables and column names
- Then add the values into them using the 'insert into' function of MySQL
- By using the 'select' command we can query the desired output

Steps taken to load the data into the data base

Software used for querying the results

--> MySQL Workbench 8.0 CE

Software used for analyzing using Bar plots

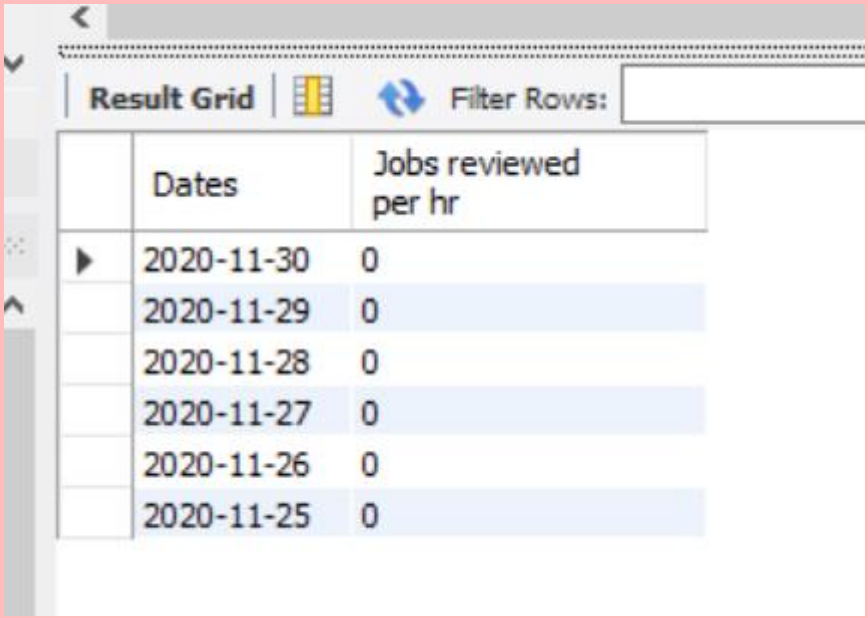
--> Microsoft Excel

Findings 1

1. **Number of jobs reviewed:** Amount of jobs reviewed over time.

Our task is to Calculate the number of jobs reviewed per hour per day for November 2020.

```
SELECT ds AS Dates,
(round(count(job_id)/sum(time_spent))*3600) as 'Jobs reviewed per hr'

from `sql project-1 table (1)`

where ds between '2020-11-01' and '2020-11-30'

group by ds;
```



Dates	Jobs reviewed per hr
2020-11-30	0
2020-11-29	0
2020-11-28	0
2020-11-27	0
2020-11-26	0
2020-11-25	0

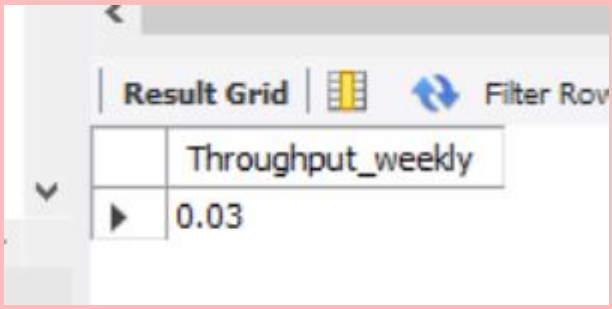
Findings 2

2.Throughput: It is the no. of events happening per second.

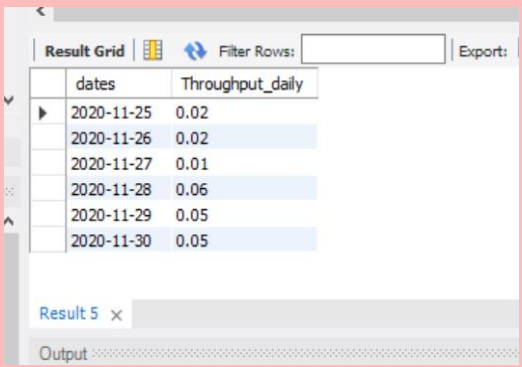
Our task is to Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?

```
SELECT (ROUND(COUNT(event)/sum(time_spent),2)) as
Throughput_weekly

FROM `sql project-1 table (1)`;
```

The weekly throughput is 0.03




```
select ds as dates,  (ROUND(COUNT(event)/sum(time_spent),2))
as Throughput_daily

FROM `sql project-1 table (1)`

group by ds

order by ds;
```



Findings 3

3.Percentage share of each language: Share of each language for different contents.

Our task is to calculate the percentage share of each language in the last 30 days.

```
SELECT  'language' AS Languages, (ROUND (100*
COUNT(*)/total,2)) as percentage_of_lang

FROM  `sql project-1 table (1)`

CROSS JOIN (SELECT COUNT(*) as total FROM `sql project-1
table (1)`) sub_total

GROUP BY languages;
```

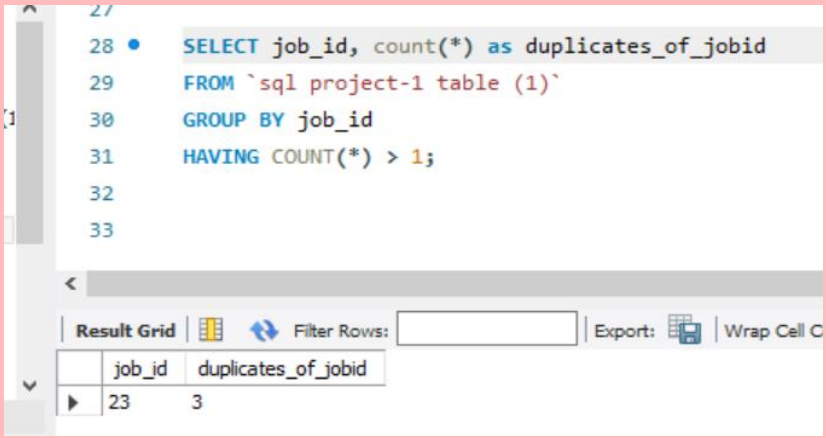
Languages	Percentage
English	12.50
Arabic	12.50
Persian	37.50
Hindi	12.50
French	12.50
Italian	12.50

Findings 4

4.Duplicate rows: Rows that have the same value present in them.

Our task is to the display duplicates from the table.

```
SELECT job_id, count(*) as duplicates_of_jobid
FROM `sql project-1 table (1)`
GROUP BY job_id
HAVING COUNT(*) > 1;
```

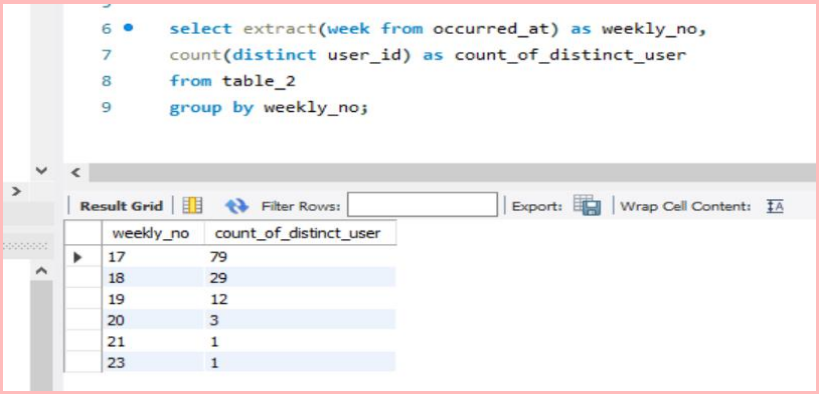


Findings 5

1. **User Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service.

Our task is to calculate the weekly user engagement.

```
SELECT  EXTRACT(week FROM occurred_at) AS weekly_no,
COUNT(distinct user_id) AS count_of_distinct_user
FROM table_2
GROUP BY weekly_no;
```

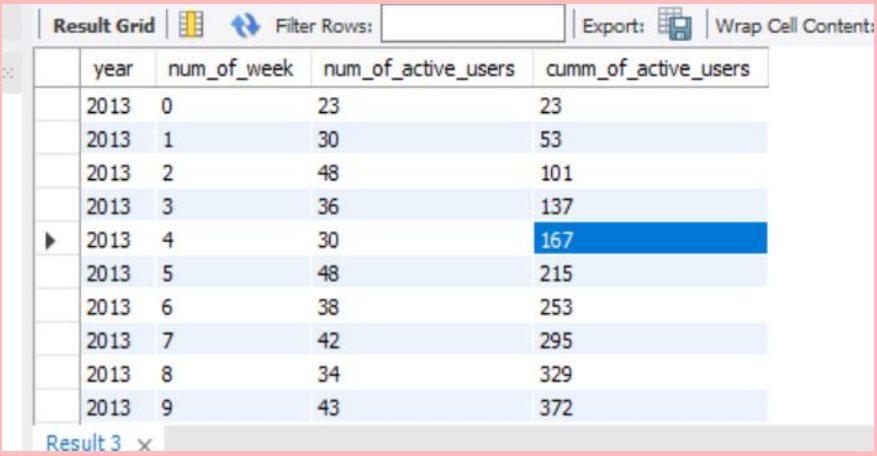


Findings 6

2.User Growth: Amount of users growing over time for a product.

Our task is to calculate the user growth for product.

```
SELECT year, num_of_week, num_of_active_users,

SUM(num_of_active_users) OVER(ORDER BY  year,
num_of_week rows BETWEEN unbounded preceding and current
row)

AS cumm_of_active_users

FROM

(SELECT

    EXTRACT(year FROM a.activated_at) AS year,

    EXTRACT(week FROM a.activated_at)AS num_of_week,

    COUNT(distinct user_id) AS num_of_active_users

FROM table_1 a

WHERE state='active'

GROUP BY year, num_of_week

ORDER BY year, num_of_week

)a;
```



year	num_of_week	num_of_active_users	cumm_of_active_users
2013	0	23	23
2013	1	30	53
2013	2	48	101
2013	3	36	137
2013	4	30	167
2013	5	48	215
2013	6	38	253
2013	7	42	295
2013	8	34	329
2013	9	43	372

Findings 7

3.Weekly Retention: Users getting retained weekly after signing-up for a product.

Our task is to calculate the weekly retention of users-sign up cohort.

```
select count(user_id),
      sum(case when retention_week = 1 then 1 else 0 end) as
retention_per_week
from
(
select (a.user_id,
      a.sign_up_week,
      b.engagement_week,
      b.engagement_week - a.sign_up_week) as
retention_week

from
(
(select distinct user_id, extract(week from occured_at) as
week_signup
from table_2
where event_type = 'signup_flow'
and event_name = 'complete_signup'
and extract(week from occured_at)=18)a
left join
(select distinct user_id, extract(week from occured_at) as
engagement_week
from table_2
where event_type = 'engagement')b
on a.user_id = b.user_id
)
)
group by user_id
order by user_id;
```

year	week	device	count(distinct user_id)
NULL	NULL	acer aspire desktop	198
NULL	NULL	acer aspire notebook	338
NULL	NULL	amazon fire phone	89
NULL	NULL	asus chromebook	355
NULL	NULL	dell inspiron desktop	360
NULL	NULL	dell inspiron notebook	677
NULL	NULL	hp pavilion desktop	339
NULL	NULL	htc one	196
NULL	NULL	ipad air	478
NULL	NULL	ipad mini	292
NULL	NULL	iphone 4s	409
NULL	NULL	iphone 5	1025
NULL	NULL	iphone 5s	626
NULL	NULL	kindle fire	205
NULL	NULL	lenovo thinkpad	1309
NULL	NULL	mac mini	150
NULL	NULL	macbook air	950
NULL	NULL	macbook pro	1952
NULL	NULL	nexus 10	273
NULL	NULL	nexus 5	621
NULL	NULL	nexus 7	355
NULL	NULL	nokia lumia 635	211
NULL	NULL	samsumg galaxy tablet	107
NULL	NULL	samsung galaxy note	119
NULL	NULL	samsung galaxy s4	803
NULL	NULL	windows surface	182

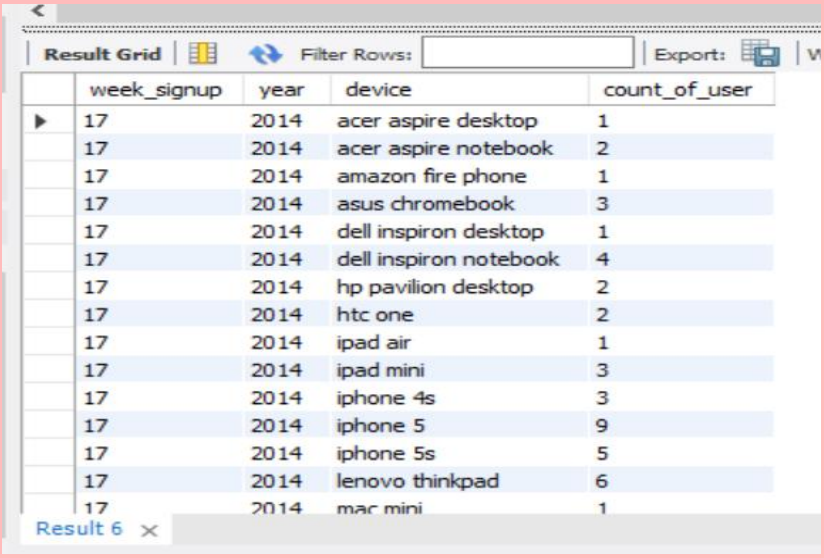
Findings 8

4.Weekly Engagement: To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.

Our task is to calculate the weekly engagement per device.

```
select extract(week from occurred_at)as week_signup, extract(year from occurred_at)as year, device,

count(distinct user_id)as 'count_of_user'

from table_2

where event_type='engagement'
```

```
group by 1,2,3
order by 1,2,3;
```



Result Grid | Filter Rows: | Export:

week_signup	year	device	count_of_user
17	2014	acer aspire desktop	1
17	2014	acer aspire notebook	2
17	2014	amazon fire phone	1
17	2014	asus chromebook	3
17	2014	dell inspiron desktop	1
17	2014	dell inspiron notebook	4
17	2014	hp pavilion desktop	2
17	2014	htc one	2
17	2014	ipad air	1
17	2014	ipad mini	3
17	2014	iphone 4s	3
17	2014	iphone 5	9
17	2014	iphone 5s	5
17	2014	lenovo thinkpad	6
17	2014	mac mini	1

Result 6

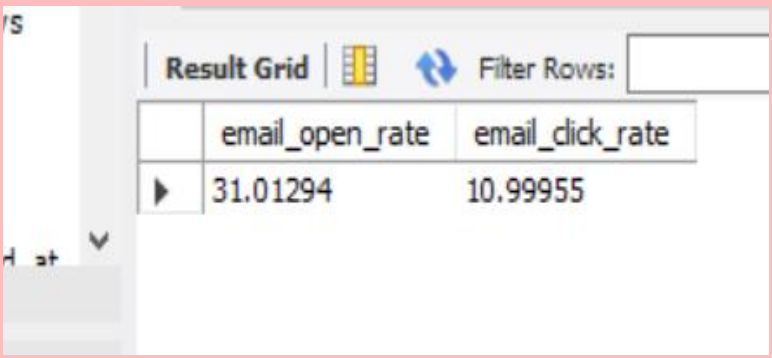
Findings 9

5.Email Engagement: Users engaging with the email service.

Our task is to calculate the email engagement metrics.

```
select
100.0 * sum(case when email_cat = 'email_opened' then 1 else 0
end)
    /sum(case when email_cat = 'email_sent' then 1 else 0 end)
as email_open_rate,
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_click_rate
from
(
select *,
case when action in ('sent_weekly_digest',
'sent_reengagement_email')
    then 'email_sent'
    when action in ('email_open')
    then 'email_opened'
    when action in ('email_clickthrough')
    then 'email_clicked'
end as email_cat
from email_events
)a;
```



Conclusion

I would like to conclude that Operation Analytics and Investigating Metric Spike are very necessary and they must be done on daily, weekly, Monthly, Quarterly or Yearly basis based on the Business needs of the firm.

Also, any firm/entity must focus on the Email Engagement with the customers; the firm must use catchy headings along with reasonable discounts and coupons so as to increase their existing customer base.

Also any firm must have a separate department(if possible) so as to hear out to the problems of those Visitors who had left the Sign-up Process in between, the firm must guide them so as to convert them from Visitors to Customers.

Hiring Process Analytics

Description

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

Being a Data Analyst, your job is to go through these trends and draw insights out of it for hiring department to work upon.

You are working for a MNC such as Google as a lead Data Analyst and the company has provided with the data records of their previous hiring and have asked you to answer certain questions making sense out of that data.

The Problem

Data Analytics Tasks:

After downloading the dataset, use Excel to answer the below questions:

A. Hiring Analysis: The hiring process involves bringing new individuals into the organization for various roles.

Your Task: Determine the gender distribution of hires. How many males and females have been hired by the company?

B. Salary Analysis: The average salary is calculated by adding up the salaries of a group of employees and then dividing the total by the number of employees.

Your Task: What is the average salary offered by this company? Use Excel functions to calculate this.

C. Salary Distribution: Class intervals represent ranges of values, in this case, salary ranges. The class interval is the difference between the upper and lower limits of a class.

Your Task: Create class intervals for the salaries in the company. This will help you understand the salary distribution.

D. Departmental Analysis: Visualizing data through charts and plots is a crucial part of data analysis.

Your Task: Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.

E. Position Tier Analysis: Different positions within a company often have different tiers or levels.

Your Task: Use a chart or graph to represent the different position tiers within the company. This will help you understand the distribution of positions across different tiers.

Design

Before starting the actual analysis I have:-

- Firstly I made a copy of the raw data where I can perform the Analysis so that what ever changes I made it will not affect the original data.

- Secondly I looked for blank spaces and NULL values if any.

- Then I had imputed the numerical blank and NULL cells with mean of the column(if no outliers existed for that particular column) or with median (if outliers existed for that column)
- Then I looked for if any outliers exists and replaced them with the median of the particular column where the outlier existed.

- Then for blank cells of categorical variables I had replaced with the variable with the highest count Then I looked for duplicate rows and removed them if any.

- Then I removed the irrelevant columns(data) from the dataset which was not necessary for doing the analysis.

Software used for doing the overall Analysis:-

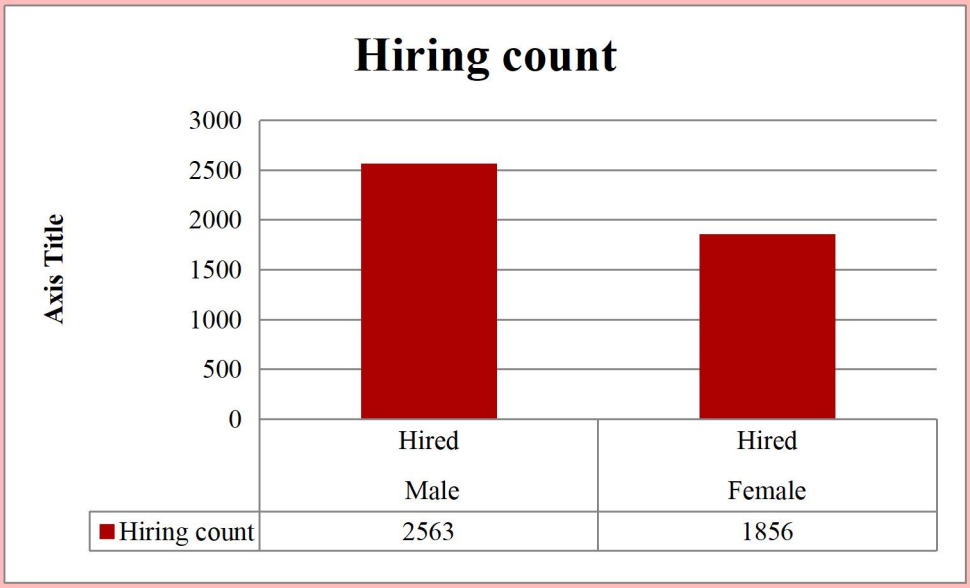
----> **Microsoft Excel**

Findings 1

1. **Hiring:** Process of intaking of people into an organization for different kinds of positions.

Our Task: How many males and females are Hired ?

event_name	status	Hiring count
Male	Hired	2563
Female	Hired	1856



From the above bar plot , we can observe that there are 2563 males and 1856 females have been hired for different positions of organizations of the company.

Findings 2

2. **Average Salary:** Adding all the salaries for a select group of employees and then dividing the sum by the number of employees in the group.

Our Task : What is the average salary offered in this company ?

Average salary	49983.03
Median of salary	49625
minimum salary	100
maximum salary	400000

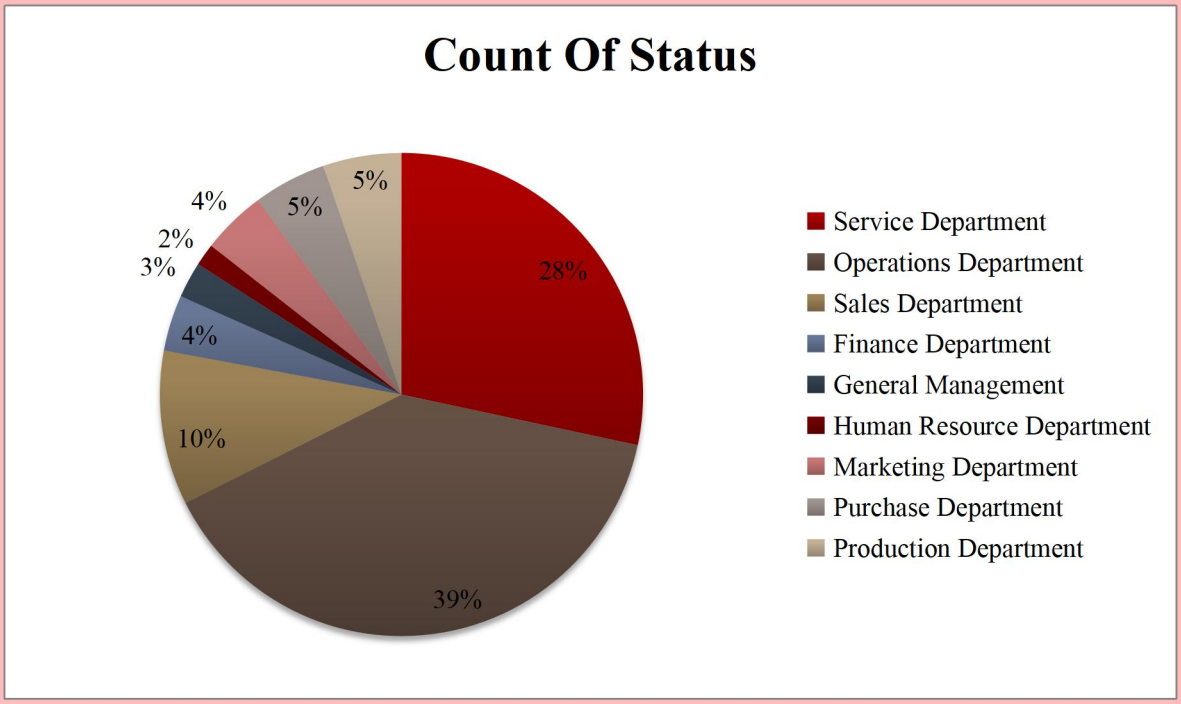
From above , the average salary offered by this company was 49983.03

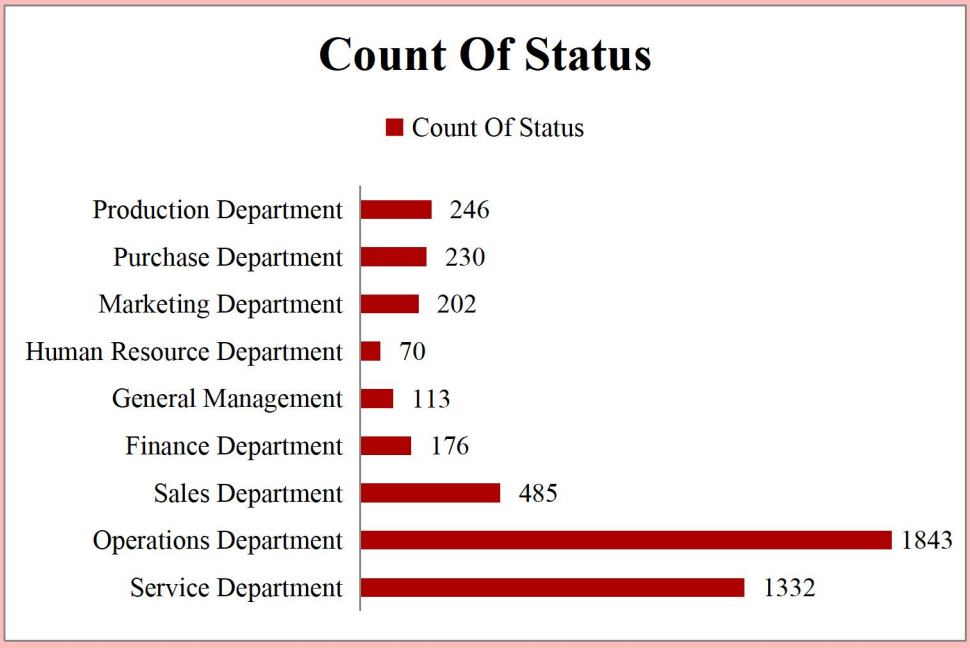
Findings 3

3. Charts and Plots: This is one of the most important part of analysis to visualize the data.

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

Department	Count Of Status
Service Department	1332
Operations Department	1843
Sales Department	485
Finance Department	176
General Management	113
Human Resource Department	70
Marketing Department	202
Purchase Department	230
Production Department	246





From the above pie chart and bar plot I have analyzed that most of the people are working on the

Operations Department (i.e) people hired on this particular department and Least number of people hired for Human Resource Department.

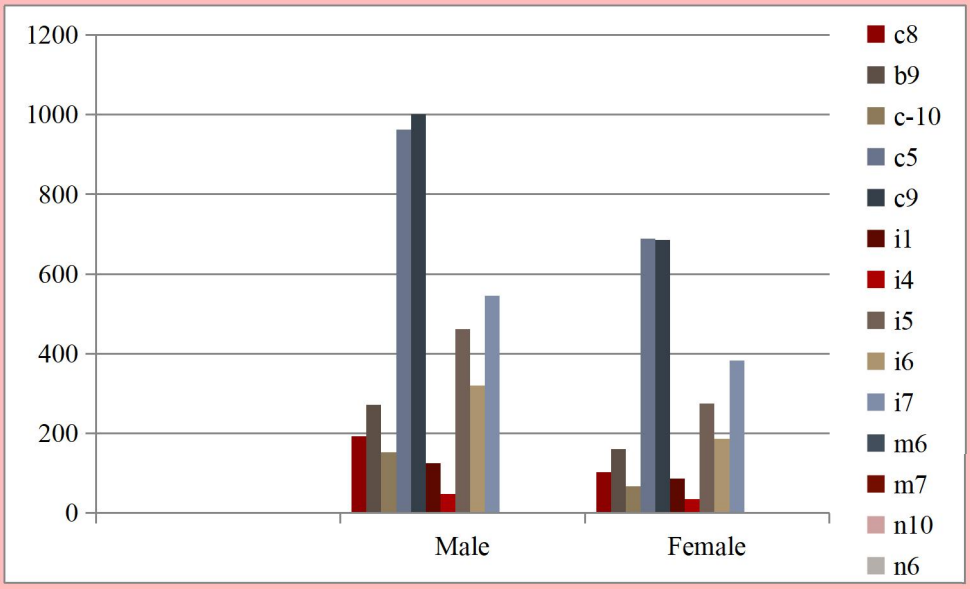
39% hired people that is ,(1843) people are working in the Operations Department and 1% of hired people that is, (70) only people working in the Human Resource Department.

Findings 4

4. Charts: Use different charts and graphs to perform the task representing the data.

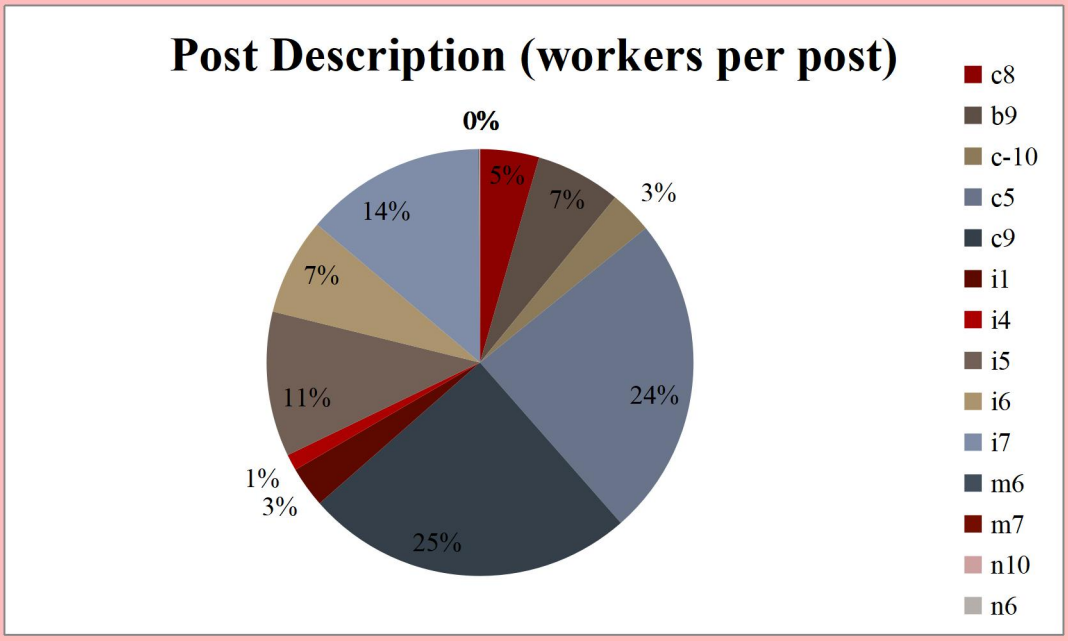
Represent different post tiers using chart/graph?

Event_name	c8	b9	c-10	c5	c9	i1	i4	i5	i6	i7	m6	m7	n10	n6	n9	
Male		193	271	152	963	1001	125	48	461	320	546	2	1	1	0	0
Female		103	161	67	689	686	86	35	275	187	383	1	0	0	1	1



From above column chart, Post Description (Gender wise distribution per post) is then male has the highest post tiers than the female employees.

No. of People	post_name
320	c8
463	b9
232	c-10
1747	c5
1792	c9
222	i1
88	i4
787	i5
527	i6
982	i7
3	m6
1	m7
1	n10
1	n6
1	n9
1	-



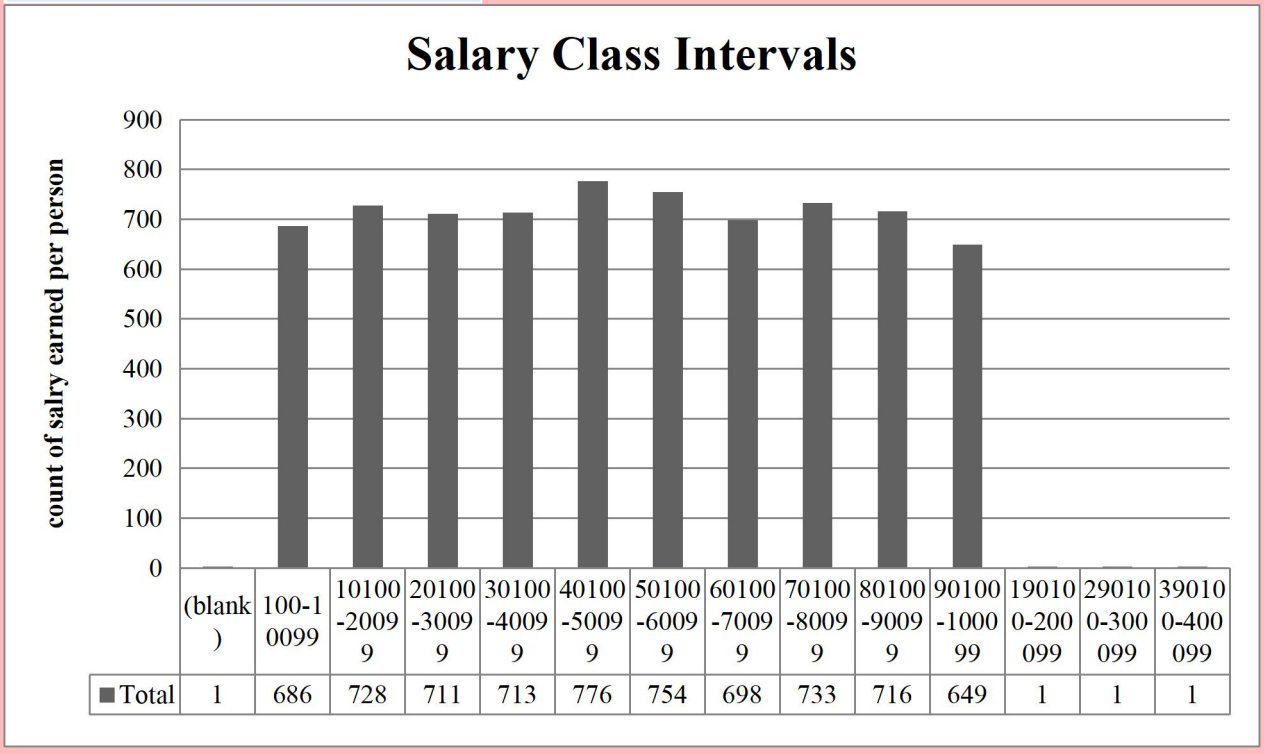
From above pie chart , I have inferred a post description (workers per post) then there c9 post of 25% is the most workers working in c9 post.

Findings 5

5. Class Intervals: The class interval is the difference between the upper class limit and the lower class limit.

Draw the class intervals for salary in the company ?

Row Labels	Count of Status
(blank)	1
100-10099	686
10100-20099	728
20100-30099	711
30100-40099	713
40100-50099	776
50100-60099	754
60100-70099	698
70100-80099	733
80100-90099	716
90100-100099	649
190100-200099	1
290100-300099	1
390100-400099	1
Grand Total	**7168**

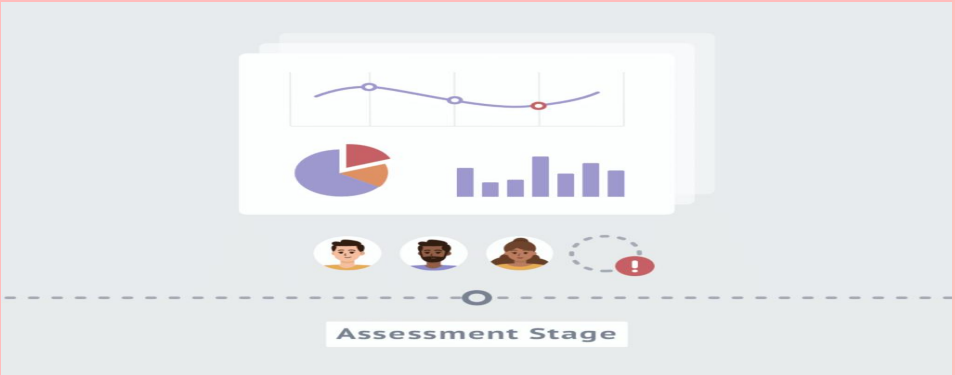


From the above class interval column chart represents that most of the people are on the intervals of 40100 – 50099 i.e, (776) are getting salary more than 40100.

Conclusion

In the conclusion part, I would like to conclude that Hiring Process Analytics plays an important part for all the companies and firms to decide the job openings for the near future.

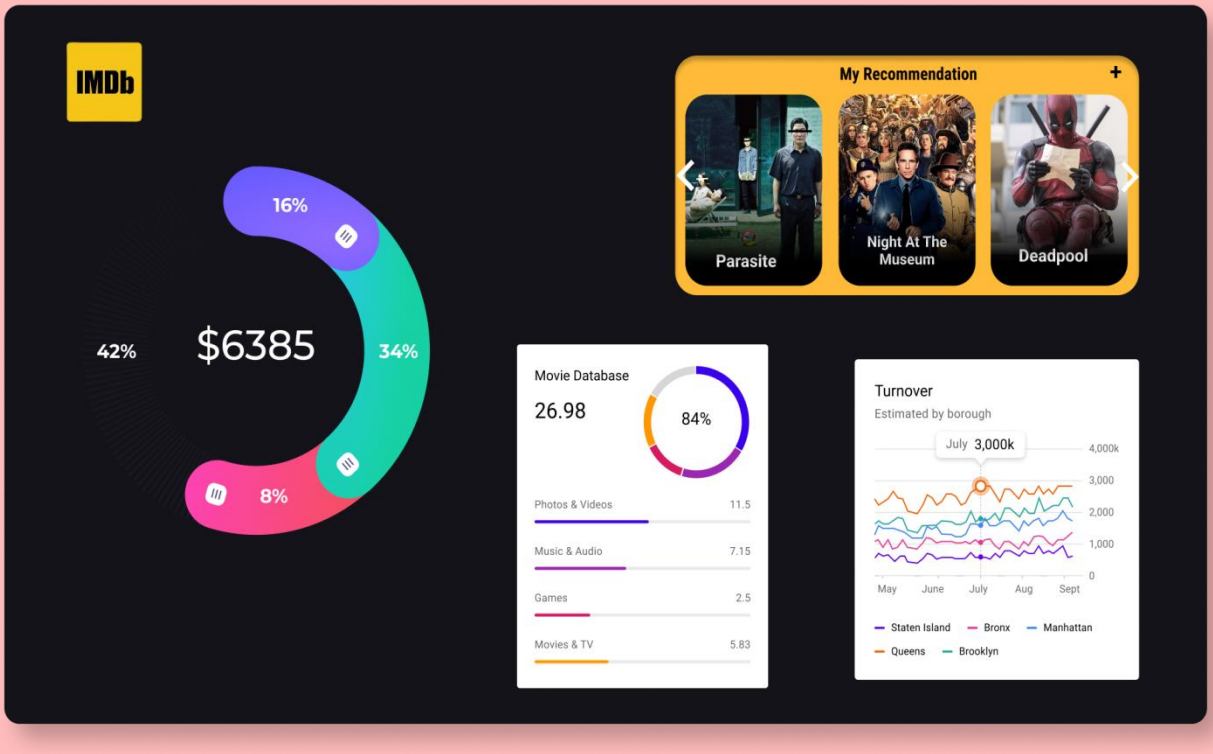
Hiring Process Analytics is done on monthly, quarterly or yearly basis as per the needs and policies of the companies For any company the Operations Department has the highest number of workforce due to the workload on this department as this department acts as a central hub for all the executive tasks carried out For any company there will some employees who have high salary packages compared to other employees, and this is due to the fact that they have some special skills and years of experience in their particular field of work Hiring Process Analytics helps the company to decide the salaries for new freshers joining the company; also it tells requirement of workforce by each department; it also helps the company decide the appraisals and increment for it's current employees.



IMDB Movie Analysis

Final Project-1

(Project – 5)



DESCRIPTION

Problem Statement: The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

Data Cleaning: This step involves preprocessing the data to make it suitable for analysis. It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.

Data Analysis: Here, you'll explore the data to understand the relationships between different variables. You might look at the correlation between movie ratings and other factors like genre, director, budget, etc. You might also want to consider the year of release, the actors involved, and other relevant factors.

The Problem

✓ Clean the data

(Dropping columns, removing null values, etc)

✓ Find the movies with the highest profit?

(observing the outliers using the appropriate chart type)

✓ Find IMDB Top 250

(Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film.)

✓ Find the best directors

(Top 10 director by highest mean of IMDB score)

✓ Find popular genres

(Perform this by step while performing previous steps)

✓ Find the critic-favorite and audience-favorite actors

(Finding the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean) and with decade of new data frame called df_by_decade.

Design

1 . Firstly I made a copy of the raw data where I can perform the Analysis so that what ever changes I made it will not affect the original data.

2. Then dropping the columns which have no use for the analysis that we will be doing Columns like ‘Color’ , ‘director_facebook_likes’ , ‘actor_3_facebook_likes, ‘actor_2_name’ , ‘actor_1_facebook_likes’ , ‘cast_total_facebook_likes, ‘actor_3_name’ , ‘facenumber_in_posts’ , ‘plot_keywords’ , ‘movie_imdb_link’ , ‘content_rating’ , ’actor_2_facebook_likes’ , ‘aspect_ratio’ , ‘movie_facebook_likes’ are the columns containing irrelevant data for the analysis tasks provided. So, these columns needs to be dropped.

3. After dropping the irrelevant columns now we need to remove the rows from the dataset having anyone of its column value as blank/NULL. Then we need to get rid off the duplicate values in the dataset which can be achieved by using the ‘Remove Duplicate Values/Cells’ available in the ‘Data’ tab

Findings 1

1. **Cleaning the data:** This is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.

Your task: Clean the data

- First of all I have analyzed that with the provided dataset , there are 5043 rows and 28 columns and 1 header row are present there.
- After then, I looked a project description that related and required columns can only be used for the upcoming analysis so then I have been used a data cleaning strategy using MySQL workbench and excel functions to solve the data cleaning task.
- So I need only the columns and they are :

- **director_name**
- **num_critic_for_reviews**
- **gross**
- **genres**
- **actor_1_name**
- **movie_title**
- **num_voted_users**
- **num_user_for_reviews**

- **language**
- **budget**
- **title_year**
- **imdb_score**
- **movie_facebook_likes**

❖ After these are the only included then I left with a **13 columns** . Then I go with a null values and null rows , this was get rid by using the excel function (press F5 – go to special – select blanks – press ok) , it will show the blank rows and then (data – remove duplicates) it will remove the duplicates and save this sheet . Then load this currently saved dataset in the MySQL workbench , queried for removing the extra null values precisely.

❖ And save the removed null values dataset and again load in the excel to see how many entries has been left .

❖ By performing these functions in the excel I found that the total number of rows is then 3789.

❖ Below this , I have attached with cleaned dataset as a picture of removing duplicates and with the blanks, required columns , and set the blank(language) column been default as 'English'.

❖ This is my attached excel file. https://docs.google.com/spreadsheets/d/1Ha--2RrVp2mce6NS6hj5dGjqWq4jOpiw/edit?usp=drive_link

director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_reviews	language	budget	title_year	imdb_score	movie_facebook_likes
James Cameron	723	760505847	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder	AvatarÂ	886204	3054	English	237000000	2009	7.9	33000
Gore Verbinski	302	309404152	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: At World's EndÂ	471220	1238	English	300000000	2007	7.1	0
Sam Mendes	602	200074175	Action|Adventure|Thriller	Christoph Waltz	SpectreÂ	275868	994	English	245000000	2015	6.8	85000
Christopher Nolan	813	448130642	Action|Thriller	Tom Hardy	The Dark Knight RisesÂ	1144337	2701	English	250000000	2012	8.5	164000
Andrew Stanton	462	73058679	Action|Adventure|Sci-Fi	Daryl Sabara	John CarterÂ	212204	738	English	263700000	2012	6.6	24000
Sam Raimi	392	336530303	Action|Adventure|Romance	J.K. Simmons	Spider-Man 3Â	383056	1902	English	258000000	2007	6.2	0
Nathan Greno	324	200807262	Adventure|Animation|Comedy|Family|Fantasy|Musical|Romance	Brad Garrett	TangledÂ	294810	387	English	260000000	2010	7.8	29000
Joss Whedon	635	458991599	Action|Adventure|Sci-Fi	Chris Hemsworth	Avengers: Age of UltronÂ	462669	1117	English	250000000	2015	7.5	118000
David Yates	375	301956980	Adventure|Family|Fantasy|Mystery	Alan Rickman	Harry Potter and the Half-Blood PrinceÂ	321795	973	English	250000000	2009	7.5	10000
Zack Snyder	673	330249062	Action|Adventure|Sci-Fi	Henry Cavill	Batman v Superman: Dawn of JusticeÂ	371639	3018	English	250000000	2016	6.9	197000
Bryan Singer	434	200069408	Action|Adventure|Sci-Fi	Kevin Spacey	Superman ReturnsÂ	240396	2367	English	209000000	2006	6.1	0
Marc Forster	403	168368427	Action|Adventure	Giancarlo Giannini	Quantum of SolaceÂ	330784	1243	English	200000000	2008	6.7	0
Gore Verbinski	313	423032628	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: Dead Man's ChestÂ	522040	1832	English	225000000	2006	7.3	5000
Gore Verbinski	450	89289910	Action|Adventure|Western	Johnny Depp	The Lone RangerÂ	181792	711	English	215000000	2013	6.5	48000
Zack Snyder	733	291021565	Action|Adventure|Fantasy|Sci-Fi	Henry Cavill	Man of SteelÂ	548573	2536	English	225000000	2013	7.2	118000
Andrew Adamson	258	141614023	Action|Adventure|Family|Fantasy	Peter Dinklage	The Chronicles of Narnia: Prince CaspianÂ	149922	438	English	225000000	2008	6.6	0
Joss Whedon	703	623279547	Action|Adventure|Sci-Fi	Chris Hemsworth	The AvengersÂ	995415	1722	English	220000000	2012	8.1	123000
Rob Marshall	448	241063875	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: On Stranger TidesÂ	370704	484	English	250000000	2011	6.7	58000
Barry Sonnenfeld	451	179020854	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi	Will Smith	Men in Black 3Â	268154	341	English	225000000	2012	6.8	40000
Peter Jackson	422	255108370	Adventure|Fantasy	Aidan Turner	The Hobbit: The Battle of the Five ArmiesÂ	354228	802	English	250000000	2014	7.5	65000
Marc Webb	599	262030663	Action|Adventure|Fantasy	Emma Stone	The Amazing Spider-ManÂ	451803	1225	English	230000000	2012	7	56000
Ridley Scott	343	105219735	Action|Adventure|Drama|History	Mark Addy	Robin HoodÂ	211765	546	English	200000000	2010	6.7	17000
Peter Jackson	509	258355354	Adventure|Fantasy	Aidan Turner	The Hobbit: The Desolation of SmaugÂ	483540	951	English	225000000	2013	7.9	83000
Chris Weitz	251	70083519	Adventure|Family|Fantasy	Christopher Lee	The Golden CompassÂ	149019	666	English	180000000	2007	6.1	0
Peter Jackson	446	218051260	Action|Adventure|Drama|Romance	Naomi Watts	King KongÂ	316018	2618	English	207000000	2005	7.2	0
James Cameron	315	658672302	Drama|Romance	Leonardo DiCaprio	TitanicÂ	793059	2528	English	200000000	1997	7.7	26000
Anthony Russo	516	407197282	Action|Adventure|Sci-Fi	Robert Downey Jr.	Captain America: Civil WarÂ	272670	1022	English	250000000	2016	8.2	72000
Peter Berg	377	65173160	Action|Adventure|Sci-Fi|Thriller	Liam Neeson	BattleshipÂ	202382	751	English	209000000	2012	5.9	44000
Colin Trevorrow	644	652177271	Action|Adventure|Sci-Fi|Thriller	Bryce Dallas Howard	Jurassic WorldÂ	418214	1290	English	150000000	2015	7	150000
Sam Mendes	750	304360277	Action|Adventure|Thriller	Albert Finney	SkyfallÂ	522030	1498	English	200000000	2012	7.8	80000
Sam Raimi	300	373377893	Action|Adventure|Fantasy|Romance	J.K. Simmons	Spider-Man 2Â	411164	1303	English	200000000	2004	7.3	0
Shane Black	608	408992272	Action|Adventure|Sci-Fi	Robert Downey Jr.	Iron Man 3Â	557489	1187	English	200000000	2013	7.2	95000
Tim Burton	451	334185206	Adventure|Family|Fantasy	Johnny Depp	Alice in WonderlandÂ	306320	736	English	200000000	2010	6.5	24000
Brett Ratner	334	234360014	Action|Adventure|Fantasy|Sci-Fi|Thriller	Hugh Jackman	X-Men: The Last StandÂ	383427	1912	English	210000000	2006	6.8	0
Dan Scanlon	376	268488329	Adventure|Animation|Comedy|Family|Fantasy	Steve Buscemi	Monsters UniversityÂ	235025	265	English	200000000	2013	7.3	44000
Michael Bay	366	402076689	Action|Adventure|Sci-Fi	Glenn Morshower	Transformers: Revenge of the FallenÂ	323207	1439	English	200000000	2009	6	0
Michael Bay	378	245428137	Action|Adventure|Sci-Fi	Bingbing Li	Transformers: Age of ExtinctionÂ	242420	918	English	210000000	2014	5.7	56000
Sam Raimi	525	234903076	Adventure|Family|Fantasy	Tim Holmes	Oz the Great and PowerfulÂ	175409	511	English	215000000	2013	6.4	60000
Marc Webb	495	202853933	Action|Adventure|Fantasy|Sci-Fi	Emma Stone	The Amazing Spider-Man 2Â	321227	1067	English	200000000	2014	6.7	41000
Joseph Kosinski	469	172051787	Action|Adventure|Sci-Fi	Jeff Bridges	TRON: LegacyÂ	264185	663	English	170000000	2010	6.8	30000
John Lasseter	304	191450875	Adventure|Animation|Comedy|Family|Sport	Joe Mantegna	Cars 2Â	101178	285	English	200000000	2011	6.3	10000
Martin Campbell	436	116593191	Action|Adventure|Sci-Fi	Ryan Reynolds	Green LanternÂ	233595	550	English	200000000	2011	5.6	24000
Lee Unkrich	453	414984497	Adventure|Animation|Comedy|Family|Fantasy	Tom Hanks	Toy Story 3Â	544884	733	English	200000000	2010	8.3	30000
McG	422	125320003	Action|Adventure|Sci-Fi	Christian Bale	Terminator SalvationÂ	286093	874	English	200000000	2009	6.6	0
James Wan	424	350034110	Action|Crime|Thriller	Jason Statham	Furious 7Â	278232	657	English	190000000	2015	7.2	94000
Marc Forster	654	202351611	Action|Adventure|Horror|Sci-Fi|Thriller	Peter Capaldi	World War ZÂ	465019	933	English	190000000	2013	7	129000
Bryan Singer	559	233914986	Action|Adventure|Fantasy|Sci-Fi|Thriller	Jennifer Lawrence	X-Men: Days of Future PastÂ	514125	732	English	200000000	2014	8	82000
J.J. Abrams	590	228756232	Action|Adventure|Sci-Fi	Benedict Cumberbatch	Star Trek Into DarknessÂ	393575	1171	English	190000000	2013	7.8	92000
Bryan Singer	558	65171860	Adventure|Fantasy	Eddie Marsan	Jack the Giant SlayerÂ	106416	203	English	195000000	2013	6.3	22000
Baz Luhrmann	490	144812796	Drama|Romance	Leonardo DiCaprio	The Great GatsbyÂ	362912	753	English	105000000	2013	7.3	115000
Mike Newell	306	90755643	Action|Adventure|Fantasy|Romance	Jake Gyllenhaal	Prince of Persia: The Sands of TimeÂ	222403	435	English	200000000	2010	6.6	25000
Guillermo del Toro	575	101785482	Action|Adventure|Sci-Fi	Charlie Hunnam	Pacific RimÂ	381148	1106	English	190000000	2013	7	85000
Michael Bay	428	352358779	Action|Adventure|Sci-Fi	Glenn Morshower	Transformers: Dark of the MoonÂ	326180	899	English	195000000	2011	6.3	49000
Steven Spielberg	470	317011114	Action|Adventure|Fantasy	Harrison Ford	Indiana Jones and the Kingdom of the Crystal SkullÂ	333847	2054	English	185000000	2008	6.2	5000
Mark Andrews	488	237282182	Adventure|Animation|Comedy|Family|Fantasy	Kelly Macdonald	BraveÂ	275336	428	English	185000000	2012	7.2	32000
Justin Lin	322	130468626	Action|Adventure|Sci-Fi|Thriller	Sofia Boutella	Star Trek BeyondÂ	35607	432	English	185000000	2016	7.5	30000
Andrew Stanton	421	223806889	Adventure|Animation|Family|Sci-Fi	John Ratzenberger	WALLÂ·EÂ	718837	1043	English	180000000	2008	8.4	16000
Brett Ratner	162	140080850	Action|Comedy|Crime|Thriller	Tzi Ma	Rush Hour 3Â	121084	221	English	140000000	2007	6.2	0
Roland Emmerich	367	166112167	Action|Adventure|Sci-Fi	Oliver Platt	2012Â	283418	1033	English	200000000	2009	5.8	15000
Robert Zemeckis	240	137850096	Animation|Drama|Family|Fantasy	Robin Wright	A Christmas CarolÂ	72809	249	English	200000000	2009	6.8	0
Lana Wachowski	384	47375327	Action|Adventure|Sci-Fi	Channing Tatum	Jupiter AscendingÂ	153395	720	English	176000000	2015	5.4	44000
David Yates	248	124051759	Action|Adventure|Drama|Romance	Christoph Waltz	The Legend of TarzanÂ	42372	239	English	180000000	2016	6.6	29000
Andrew Adamson	284	291709845	Adventure|Family|Fantasy	Jim Broadbent	The Chronicles of Narnia: The Lion, the Witch and the WardrobeÂ	286506	1463	English	180000000	2005	6.9	0
Bryan Singer	396	154985087	Action|Adventure|Sci-Fi	Jennifer Lawrence	X-Men: ApocalypseÂ	148379	622	English	178000000	2016	7.3	54000
Christopher Nolan	645	533316061	Action|Crime|Drama|Thriller	Christian Bale	The Dark KnightÂ	1676169	4667	English	185000000	2008	9	37000
Pete Docter	408	292979556	Adventure|Animation|Comedy|Family	John Ratzenberger	UpÂ	665387	704	English	175000000	2009	8.3	27000
Rob Letterman	219	198332128	Action|Adventure|Animation|Comedy|Family|Sci-Fi	Amy Poehler	Monsters vs. AliensÂ	114355	187	English	175000000	2009	6.5	0
Jon Favreau	486	318298180	Action|Adventure|Sci-Fi	Robert Downey Jr.	Iron ManÂ	696338	1055	English	140000000	2008	7.9	10000
Martin Scorsese	682	73820094	Adventure|Drama|Family|Mystery	ChloÃ« Grace Moretz	HugoÂ	245555	678	English	170000000	2011	7.5	42000
Barry Sonnenfeld	85	113745408	Action|Comedy|Sci-Fi|Western	Will Smith	Wild Wild WestÂ	129601	648	English	170000000	1999	4.8	0
Rob Cohen	264	102176165	Action|Adventure|Fantasy|Horror|Thriller	Jet Li	The Mummy: Tomb of the Dragon EmperorÂ	117927	501	English	145000000	2008	5.2	0
David Ayer	418	161087183	Action|Adventure|Comedy|Sci-Fi	Will Smith	Suicide SquadÂ	118992	971	English	175000000	2016	6.9	80000
Tom Shadyac	186	100289690	Comedy|Family|Fantasy	Jimmy Bennett	Evan AlmightyÂ	115089	237	English	175000000	2007	5.4	2000
Doug Liman	585	100189501	Action|Adventure|Sci-Fi	Tom Cruise	Edge of TomorrowÂ	431620	741	English	178000000	2014	7.9	77000
Kevin Reynolds	91	88246220	Action|Adventure|Sci-Fi|Thriller	Jeanne Tripplehorn	WaterworldÂ	144557	309	English	175000000	1995	6.1	0
Stephen Sommers	250	150167630	Action|Adventure|Sci-Fi|Thriller	Joseph Gordon-Levitt	G.I. Joe: The Rise of CobraÂ	174578	554	English	175000000	2009	5.8	0
Pete Docter	536	356454367	Adventure|Animation|Comedy|Drama|Family|Fantasy	Amy Poehler	Inside OutÂ	345198	773	English	175000000	2015	8.3	118000
Jon Favreau	370	362645141	Adventure|Drama|Family|Fantasy	Scarlett Johansson	The Jungle BookÂ	106072	398	English	175000000	2016	7.8	65000
Jon Favreau	453	312057433	Action|Adventure|Sci-Fi	Robert Downey Jr.	Iron Man 2Â	522571	725	English	200000000	2010	7	18000
Rupert Sanders	416	155111815	Action|Adventure|Drama|Fantasy	Chris Hemsworth	Snow White and the HuntsmanÂ	228554	710	English	170000000	2012	6.1	33000
Robert Stromberg	401	241407328	Action|Adventure|Family|Fantasy|Romance	Angelina Jolie Pitt	MaleficentÂ	232257	634	English	180000000	2014	7	39000
Matt Reeves	521	208543795	Action|Adventure|Drama|Sci-Fi	Gary Oldman	Dawn of the Planet of the ApesÂ	317342	620	English	170000000	2014	7.6	45000
Carl Rinsch	218	38297305	Action|Adventure|Drama|Fantasy	Keanu Reeves	47 RoninÂ	116994	324	English	175000000	2013	6.3	33000
Anthony Russo	376	259746958	Action|Adventure|Sci-Fi	Scarlett Johansson	Captain America: The Winter SoldierÂ	496749	742	English	170000000	2014	7.8	33000
Mike Mitchell	226	238371987	Adventure|Animation|Comedy|Family|Fantasy	Jon Hamm	Shrek Forever AfterÂ	128961	173	English	165000000	2010	6.4	0
Brad Bird	443	93417865	Action|Adventure|Family|Mystery|Sci-Fi	Judy Greer	TomorrowlandÂ	128306	497	English	190000000	2015	6.5	37000
Don Hall	384	222487711	Action|Adventure|Animation|Comedy|Drama|Family|Sci-Fi	Damon Wayans Jr.	Big Hero 6Â	279093	435	English	165000000	2014	7.9	41000
Rich Moore	377	189412677	Adventure|Animation|Comedy|Family|Sci-Fi	Jack McBrayer	Wreck-It RalphÂ	272554	545	English	165000000	2012	7.8	40000
Robert Zemeckis	188	665426	Adventure|Animation|Family|Fantasy	Tom Hanks	The Polar ExpressÂ	120798	444	English	165000000	2004	6.6	10000
Roland Emmerich	286	102315545	Action|Adventure|Sci-Fi	Vivica A. Fox	Independence Day: ResurgenceÂ	38157	520	English	165000000	2016	5.5	67000
Dean DeBlois	288	217387997	Adventure|Animation|Family|Fantasy	Gerard Butler	How to Train Your DragonÂ	480450	492	English	165000000	2010	8.2	35000
Jonathan Mostow	280	150350192	Action|Sci-Fi	Nick Stahl	Terminator 3: Rise of the MachinesÂ	305340	1676	English	200000000	2003	6.4	0
James Gunn	653	333130696	Action|Adventure|Sci-Fi	Bradley Cooper	Guardians of the GalaxyÂ	682155	1097	English	170000000	2014	8.1	96000
Christopher Nolan	712	187991439	Adventure|Drama|Sci-Fi	Matthew McConaughey	InterstellarÂ	928227	2725	English	165000000	2014	8.6	349000
Christopher Nolan	642	292568851	Action|Adventure|Sci-Fi|Thriller	Leonardo DiCaprio	InceptionÂ	1468200	2803	English	160000000	2010	8.8	175000
Peter Jackson	645	303001229	Adventure|Fantasy	Aidan Turner	The Hobbit: An Unexpected JourneyÂ	637246	1367	English	180000000	2012	7.9	166000
Rob Cohen	187	144512310	Action|Crime|Thriller	Paul Walker	The Fast and the FuriousÂ	272223	988	English	38000000	2001	6.7	14000
David Fincher	362	127490802	Drama|Fantasy|Romance	Brad Pitt	The Curious Case of Benjamin ButtonÂ	459346	822	English	150000000	2008	7.8	23000
Matthew Vaughn	500	146405371	Action|Adventure|Sci-Fi	Jennifer Lawrence	X-Men: First ClassÂ	513337	938	English	160000000	2011	7.8	54000
Francis Lawrence	339	281666058	Adventure|Sci-Fi	Jennifer Lawrence	The Hunger Games: Mockingjay - Part 2Â	189157	385	English	160000000	2015	6.6	38000
Jon Turteltaub	255	63143812	Action|Adventure|Family|Fantasy	Nicolas Cage	The Sorcerer's ApprenticeÂ	124183	238	English	150000000	2010	6.1	11000
Wolfgang Petersen	251	60655503	Action|Adventure|Drama|Thriller	Jimmy Bennett	PoseidonÂ	82380	629	English	160000000	2006	5.6	0
James Bobin	218	76846624	Adventure|Family|Fantasy	Johnny Depp	Alice Through the Looking GlassÂ	21352	131	English	170000000	2016	6.4	30000
Chris Miller	227	320706665	Adventure|Animation|Comedy|Family|Fantasy	Justin Timberlake	Shrek the ThirdÂ	211971	326	English	160000000	2007	6.1	0
Duncan Jones	273	46978995	Action|Adventure|Fantasy	Dominic Cooper	WarcraftÂ	111609	781	English	160000000	2016	7.3	39000
Alan Taylor	474	89732035	Action|Adventure|Sci-Fi	J.K. Simmons	Terminator GenisysÂ	188457	867	English	155000000	2015	6.6	82000
Michael Apted	228	104383624	Adventure|Family|Fantasy	Bruce Spence	The Chronicles of Narnia: The Voyage of the Dawn TreaderÂ	106446	227	English	155000000	2010	6.3	11000
Michael Bay	191	198539855	Action|Drama|History|Romance|War	Jennifer Garner	Pearl HarborÂ	254111	1999	English	140000000	2001	6.1	0
Michael Bay	386	318759914	Action|Adventure|Sci-Fi	Zack Ward	TransformersÂ	515138	1782	English	150000000	2007	7.1	8000
Oliver Stone	248	34293771	Action|Adventure|Biography|Drama|History|Romance|War	Anthony Hopkins	AlexanderÂ	138695	1590	English	155000000	2004	5.5	0
David Yates	329	292000866	Adventure|Family|Fantasy|Mystery	Robert Pattinson	Harry Potter and the Order of the PhoenixÂ	383157	1108	English	150000000	2007	7.5	0
Mike Newell	293	289994397	Adventure|Family|Fantasy|Mystery	Robert Pattinson	Harry Potter and the Goblet of FireÂ	385670	1099	English	150000000	2005	7.6	0
Peter Berg	318	227946274	Action|Drama	Will Smith	HancockÂ	345848	590	English	150000000	2008	6.4	0
Francis Lawrence	323	256386216	Drama|Horror|Sci-Fi	Will Smith	I Am LegendÂ	530870	1415	English	150000000	2007	7.2	11000
Tim Burton	276	206456431	Adventure|Comedy|Family|Fantasy	Johnny Depp	Charlie and the Chocolate FactoryÂ	320284	1361	English	150000000	2005	6.7	0
Brad Bird	318	206435493	Animation|Comedy|Family|Fantasy	Janeane Garofalo	RatatouilleÂ	475337	626	English	150000000	2007	8	10000
Christopher Nolan	478	205343774	Action|Adventure	Christian Bale	Batman BeginsÂ	980946	2685	English	150000000	2005	8.3	15000
Eric Darnell	167	179982968	Action|Adventure|Animation|Comedy|Family	Bernie Mac	Madagascar: Escape 2 AfricaÂ	146019	119	English	150000000	2008	6.7	0
Shawn Levy	183	177243721	Adventure|Comedy|Family|Fantasy	Robin Williams	Night at the Museum: Battle of the SmithsonianÂ	130272	209	English	150000000	2009	5.9	2000
Gavin Hood	330	179883016	Action|Adventure|Fantasy|Sci-Fi|Thriller	Hugh Jackman	X-Men Origins: WolverineÂ	361924	641	English	150000000	2009	6.7	0
Lana Wachowski	243	139259759	Action|Sci-Fi	Essie Davis	The Matrix RevolutionsÂ	364948	2121	English	150000000	2003	6.7	0
Chris Buck	406	400736600	Adventure|Animation|Comedy|Family|Fantasy|Musical	Josh Gad	FrozenÂ	421658	904	English	150000000	2013	7.6	58000
Lana Wachowski	273	281492479	Action|Sci-Fi	Steve Bastoni	The Matrix ReloadedÂ	421818	2789	English	150000000	2003	7.2	0
Alan Taylor	436	206360018	Action|Adventure|Fantasy	Chris Hemsworth	Thor: The Dark WorldÂ	414070	532	English	170000000	2013	7.1	65000
George Miller	739	153629485	Action|Adventure|Sci-Fi|Thriller	Tom Hardy	Mad Max: Fury RoadÂ	552503	1588	English	150000000	2015	8.1	191000
Ron Howard	298	133375846	Mystery|Thriller	Tom Hanks	Angels & DemonsÂ	207859	453	English	150000000	2009	6.7	0
Kenneth Branagh	516	181015141	Action|Adventure|Fantasy	Chris Hemsworth	ThorÂ	536514	738	English	150000000	2011	7	65000
Byron Howard	225	114053579	Adventure|Animation|Comedy|Drama|Family	ChloÃ« Grace Moretz	BoltÂ	146766	178	English	150000000	2008	6.9	0
Hoyt Yeatman	145	119420252	Action|Adventure|Animation|Comedy|Family|Fantasy|Sci-Fi	Kelli Garner	G-ForceÂ	55042	90	English	150000000	2009	5.1	0
Jonathan Liebesman	510	83640426	Action|Adventure|Fantasy	Liam Neeson	Wrath of the TitansÂ	152826	235	English	150000000	2012	5.8	19000
Tim Burton	526	79711678	Comedy|Fantasy|Horror	Johnny Depp	Dark ShadowsÂ	199059	479	English	100000000	2012	6.2	32000
Christopher McQuarrie	463	195000874	Action|Adventure|Thriller	Tom Cruise	Mission: Impossible - Rogue NationÂ	232187	440	English	150000000	2015	7.4	47000
Joe Johnston	337	61937495	Drama|Fantasy|Horror|Thriller	Anthony Hopkins	The WolfmanÂ	89442	452	English	150000000	2010	5.8	0
Steve Hickner	184	126597121	Adventure|Animation|Comedy|Family	Matthew Broderick	Bee MovieÂ	103902	206	English	150000000	2007	6.2	0
Jennifer Yuh Nelson	284	165230261	Action|Adventure|Animation|Comedy|Family	Angelina Jolie Pitt	Kung Fu Panda 2Â	182718	203	English	150000000	2011	7.3	20000
M. Night Shyamalan	280	131564731	Action|Adventure|Family|Fantasy	Seychelle Gabriel	The Last AirbenderÂ	118351	1582	English	150000000	2010	4.2	18000
J.J. Abrams	310	133382309	Action|Adventure|Thriller	Philip Seymour Hoffman	Mission: Impossible IIIÂ	256693	871	English	150000000	2006	6.9	0
Roland Emmerich	339	73103784	Action|Drama|Thriller	Channing Tatum	White House DownÂ	164258	434	English	150000000	2013	6.4	55000
Simon Wells	152	21379315	Action|Adventure|Animation|Comedy|Family|Sci-Fi	Elisabeth Harnois	Mars Needs MomsÂ	17390	112	English	150000000	2011	5.4	0
David Bowers	153	64459316	Adventure|Animation|Comedy|Family	Hugh Jackman	Flushed AwayÂ	85086	122	English	149000000	2006	6.7	0
Joe Wright	256	34964818	Adventure|Family|Fantasy	Hugh Jackman	PanÂ	39956	186	English	150000000	2015	5.8	24000
Rob Minkoff	196	111505642	Adventure|Animation|Comedy|Family|Sci-Fi	Ty Burrell	Mr. Peabody & ShermanÂ	47900	130	English	145000000	2014	6.9	11000
Wolfgang Petersen	220	133228348	Adventure	Brad Pitt	TroyÂ	381672	1694	English	175000000	2004	7.2	0
Eric Darnell	211	216366733	Adventure|Animation|Comedy|Family	Jada Pinkett Smith	Madagascar 3: Europe's Most WantedÂ	119215	134	English	145000000	2012	6.9	17000
Lee Tamahori	264	160201106	Action|Adventure|Thriller	Toby Stephens	Die Another DayÂ	169914	1135	English	142000000	2002	6.1	0
Paul Feig	464	118099659	Action|Comedy|Fantasy|Sci-Fi	Ed Begley Jr.	GhostbustersÂ	69757	1211	English	144000000	2016	5.5	62000
Michael Bay	167	201573391	Action|Adventure|Sci-Fi|Thriller	Bruce Willis	ArmageddonÂ	322395	1171	English	140000000	1998	6.6	11000
Barry Sonnenfeld	208	190418803	Action|Adventure|Comedy|Family|Fantasy|Mystery|Sci-Fi	Will Smith	Men in Black IIÂ	270207	606	English	140000000	2002	6.1	2000
Robert Zemeckis	287	82161969	Action|Adventure|Animation|Fantasy	Robin Wright	BeowulfÂ	142440	305	English	150000000	2007	6.5	5000
Alessandro Carloni	210	143523463	Action|Adventure|Animation|Comedy|Family	J.K. Simmons	Kung Fu Panda 3Â	64522	145	English	145000000	2016	7.2	24000
Brad Bird	452	209364921	Action|Adventure|Thriller	Tom Cruise	Mission: Impossible - Ghost ProtocolÂ	363104	512	English	145000000	2011	7.4	35000
Peter Ramsey	256	103400692	Adventure|Animation|Family|Fantasy	Hugh Jackman	Rise of the GuardiansÂ	125355	174	English	145000000	2012	7.3	23000
Dean Parisot	155	110332737	Comedy|Crime	John Michael Higgins	Fun with Dick and JaneÂ	110788	238	English	100000000	2005	6.1	2000
Edward Zwick	190	111110575	Action|Drama|History|War	Tom Cruise	The Last SamuraiÂ	317166	928	English	140000000	2003	7.7	0
Ridley Scott	314	65007045	Action|Adventure|Drama	Christian Bale	Exodus: Gods and KingsÂ	128682	657	English	140000000	2014	6.1	51000
J.J. Abrams	513	257704099	Action|Adventure|Sci-Fi	Chris Hemsworth	Star TrekÂ	504419	1339	English	150000000	2009	8	19000
Sam Raimi	291	403706375	Action|Adventure|Fantasy|Romance	J.K. Simmons	Spider-ManÂ	544663	2012	English	139000000	2002	7.3	5000
Dean DeBlois	292	176997107	Action|Adventure|Animation|Comedy|Family|Fantasy	Gerard Butler	How to Train Your Dragon 2Â	221128	343	English	145000000	2014	7.9	49000
Alex Proyas	184	31141074	Action|Adventure|Fantasy	Gerard Butler	Gods of EgyptÂ	51892	275	English	140000000	2016	5.5	24000
Rob Cohen	143	31704416	Action|Adventure|Sci-Fi|Thriller	Sam Shepard	StealthÂ	45455	388	English	135000000	2005	5	0
Zack Snyder	451	107503316	Action|Drama|Mystery|Sci-Fi	Matt Frewer	WatchmenÂ	392474	1229	English	130000000	2009	7.7	18000
Richard Donner	141	129734803	Action|Crime|Thriller	Jet Li	Lethal Weapon 4Â	127497	287	English	140000000	1998	6.6	0
Ang Lee	267	132122995	Action|Sci-Fi	Kevin Rankin	HulkÂ	212106	1445	English	137000000	2003	5.7	0
Jon M. Chu	331	122512052	Action|Adventure|Sci-Fi|Thriller	Channing Tatum	G.I. Joe: RetaliationÂ	146352	288	English	130000000	2013	5.8	42000
Breck Eisner	165	68642452	Action|Adventure|Comedy|Thriller	Matthew McConaughey	SaharaÂ	77673	463	English	130000000	2005	6	0
Hironobu Sakaguchi	186	32131830	Action|Adventure|Animation|Fantasy|Romance|Sci-Fi	Steve Buscemi	Final Fantasy: The Spirits WithinÂ	72259	738	English	137000000	2001	6.4	0
Joe Johnston	510	176636816	Action|Adventure|Sci-Fi	Chris Evans	Captain America: The First AvengerÂ	303813	679	English	140000000	2011	6.9	46000
Michael Apted	197	126930660	Action|Adventure|Thriller	Colin Salmon	The World Is Not EnoughÂ	157319	685	English	135000000	1999	6.4	2000
Peter Weir	244	93926386	Action|Adventure|Drama|History|War	James D'Arcy	Master and Commander: The Far Side of the WorldÂ	168207	684	English	150000000	2003	7.4	0
Bill Condon	322	292298923	Adventure|Drama|Fantasy|Romance	Robert Pattinson	The Twilight Saga: Breaking Dawn - Part 2Â	185394	329	English	120000000	2012	5.5	65000
George Miller	136	63992328	Animation|Comedy|Family|Musical	Robin Williams	Happy Feet 2Â	32589	79	English	135000000	2011	5.9	0
Louis Leterrier	334	134518390	Action|Adventure|Sci-Fi	Ty Burrell	The Incredible HulkÂ	326286	643	English	150000000	2008	6.8	0
Steven Spielberg	252	52792307	Adventure|Family|Fantasy	Mark Rylance	The BFGÂ	12572	106	English	140000000	2016	6.8	27000
Alejandro G. IÃ±Ã¡rritu	556	183635922	Adventure|Drama|Thriller|Western	Leonardo DiCaprio	The RevenantÂ	406020	1188	English	135000000	2015	8.1	190000
David Soren	186	83024900	Adventure|Animation|Comedy|Family|Sport	Ryan Reynolds	TurboÂ	62424	90	English	135000000	2013	6.5	15000
Gore Verbinski	362	123207194	Adventure|Animation|Comedy|Family|Western	Johnny Depp	RangoÂ	185203	557	English	135000000	2011	7.2	26000
Eric Darnell	133	83348920	Adventure|Animation|Comedy|Family	Benedict Cumberbatch	Penguins of MadagascarÂ	60250	118	English	132000000	2014	6.7	11000
Paul Greengrass	329	227137090	Action|Mystery|Thriller	Matt Damon	The Bourne UltimatumÂ	491077	820	English	110000000	2007	8.1	0
Mark Osborne	266	215395021	Action|Adventure|Animation|Comedy|Family	Angelina Jolie Pitt	Kung Fu PandaÂ	307029	360	English	130000000	2008	7.6	6000
Peyton Reed	517	180191634	Action|Adventure|Comedy|Sci-Fi	Judy Greer	Ant-ManÂ	315866	549	English	130000000	2015	7.4	61000
Francis Lawrence	502	424645577	Adventure|Sci-Fi|Thriller	Jennifer Lawrence	The Hunger Games: Catching FireÂ	488587	706	English	130000000	2013	7.6	82000
Tim Johnson	163	177343675	Adventure|Animation|Comedy|Family|Fantasy|Sci-Fi	Jim Parsons	HomeÂ	70121	214	English	135000000	2015	6.7	26000
Steven Spielberg	401	234277056	Adventure|Sci-Fi|Thriller	Tom Cruise	War of the WorldsÂ	354545	2741	English	132000000	2005	6.5	0
Michael Bay	94	138396624	Action|Comedy|Crime|Thriller	Will Smith	Bad Boys IIÂ	178126	511	English	130000000	2003	6.6	0
Chris Miller	246	149234747	Action|Adventure|Animation|Comedy|Family|Fantasy	Salma Hayek	Puss in BootsÂ	114287	137	English	130000000	2011	6.7	16000
Phillip Noyce	350	118311368	Action|Crime|Mystery|Thriller	Angelina Jolie Pitt	SaltÂ	245621	514	English	110000000	2010	6.4	25000
Darren Aronofsky	454	101160529	Action|Adventure|Drama	Anthony Hopkins	NoahÂ	200022	1240	English	125000000	2014	5.8	71000
Steven Spielberg	440	77564037	Action|Adventure|Family|Mystery	Toby Jones	The Adventures of TintinÂ	177585	447	English	135000000	2011	7.4	44000
Alfonso CuarÃ³n	274	249358727	Adventure|Family|Fantasy|Mystery	Daniel Radcliffe	Harry Potter and the Prisoner of AzkabanÂ	382255	1304	English	130000000	2004	7.8	11000
Baz Luhrmann	243	49551662	Adventure|Drama|Romance|War	Essie Davis	AustraliaÂ	102558	430	English	130000000	2008	6.6	0
M. Night Shyamalan	349	60522097	Action|Adventure|Sci-Fi	Will Smith	After EarthÂ	158720	744	English	130000000	2013	4.9	37000
Eric Leighton	143	137748063	Adventure|Animation|Family|Thriller	Alfre Woodard	DinosaurÂ	38458	241	English	127500000	2000	6.5	0
Shawn Levy	134	113733726	Adventure|Comedy|Family|Fantasy	Robin Williams	Night at the Museum: Secret of the TombÂ	67225	126	English	127000000	2014	6.2	11000
Tom McGrath	255	148337537	Action|Animation|Comedy|Family|Sci-Fi	J.K. Simmons	MegamindÂ	172754	187	English	130000000	2010	7.3	15000
Chris Columbus	258	317557891	Adventure|Family|Fantasy	Daniel Radcliffe	Harry Potter and the Sorcerer's StoneÂ	444685	1371	English	125000000	2001	7.5	16000
Robert Schwentke	208	33592415	Action|Comedy|Fantasy	Ryan Reynolds	R.I.P.D.Â	91640	210	English	130000000	2013	5.6	20000
Gore Verbinski	271	305388685	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: The Curse of the Black PearlÂ	883474	2113	English	140000000	2003	8.1	10000
Francis Lawrence	403	337103873	Adventure|Sci-Fi|Thriller	Jennifer Lawrence	The Hunger Games: Mockingjay - Part 1Â	305008	591	English	125000000	2014	6.7	32000
Ron Howard	294	217536138	Mystery|Thriller	Tom Hanks	The Da Vinci CodeÂ	314255	1986	English	125000000	2006	6.6	0
Carlos Saldanha	159	131536019	Adventure|Animation|Comedy|Family|Musical	Miguel Ferrer	Rio 2Â	58498	99	English	103000000	2014	6.4	0
Bryan Singer	289	214948780	Action|Adventure|Fantasy|Sci-Fi|Thriller	Hugh Jackman	X-Men 2Â	405975	1035	English	110000000	2003	7.5	0
Justin Lin	342	209805005	Action|Crime|Thriller	Paul Walker	Fast FiveÂ	284792	566	English	125000000	2011	7.3	34000
Guy Ritchie	382	186830669	Action|Adventure|Crime|Mystery|Thriller	Robert Downey Jr.	Sherlock Holmes: A Game of ShadowsÂ	358653	412	English	125000000	2011	7.5	39000
Louis Leterrier	344	163192114	Action|Adventure|Fantasy	Liam Neeson	Clash of the TitansÂ	238679	657	English	125000000	2010	5.8	15000
Paul Verhoeven	196	119412921	Action|Sci-Fi	Ronny Cox	Total RecallÂ	240241	591	English	65000000	1990	7.5	0
John McTiernan	85	32694788	Action|Adventure|History	Tony Curran	The 13th WarriorÂ	101411	546	English	85000000	1999	6.6	0
Tony Gilroy	456	113165635	Action|Adventure|Thriller	Jeremy Renner	The Bourne LegacyÂ	229825	504	English	125000000	2012	6.7	51000
Joel Schumacher	185	107285004	Action	Michael Gough	Batman & RobinÂ	189855	1018	English	125000000	1997	3.7	0
Ron Howard	173	260031035	Comedy|Family|Fantasy	Clint Howard	How the Grinch Stole ChristmasÂ	141414	482	English	123000000	2000	6	0
Roland Emmerich	259	186739919	Action|Adventure|Sci-Fi|Thriller	Jake Gyllenhaal	The Day After TomorrowÂ	355248	1139	English	125000000	2004	6.4	19000
John Woo	257	215397307	Action|Adventure|Thriller	Tom Cruise	Mission: Impossible IIÂ	242188	1426	English	125000000	2000	6.1	0
Wolfgang Petersen	251	182618434	Action|Adventure|Drama|Thriller	Karen Allen	The Perfect StormÂ	155076	779	English	140000000	2000	6.4	0
Tim Story	262	131920333	Action|Adventure|Sci-Fi|Thriller	Chris Evans	Fantastic 4: Rise of the Silver SurferÂ	215273	439	English	130000000	2007	5.6	0
Ang Lee	352	124976634	Adventure|Drama|Fantasy	Suraj Sharma	Life of PiÂ	440034	733	English	120000000	2012	8	122000
Mark Steven Johnson	276	115802596	Action|Fantasy|Thriller	Nicolas Cage	Ghost RiderÂ	182661	631	English	110000000	2007	5.2	0
Paul Greengrass	267	108521835	Action|Thriller	Matt Damon	Jason BourneÂ	40125	297	English	120000000	2016	7.1	51000
McG	102	100685880	Action|Adventure|Comedy|Crime	Demi Moore	Charlie's Angels: Full ThrottleÂ	100821	534	English	120000000	2003	4.8	0
Ridley Scott	773	126464904	Adventure|Mystery|Sci-Fi	Michael Fassbender	PrometheusÂ	436260	2536	English	130000000	2012	7	97000
Rob Minkoff	71	64736114	Adventure|Animation|Comedy|Family|Fantasy	Nathan Lane	Stuart Little 2Â	56471	69	English	120000000	2002	5.4	459
Neill Blomkamp	476	93050117	Action|Drama|Sci-Fi|Thriller	Matt Damon	ElysiumÂ	358037	814	English	115000000	2013	6.6	61000
David Twohy	207	57637485	Action|Adventure|Sci-Fi|Thriller	Vin Diesel	The Chronicles of RiddickÂ	183909	666	English	105000000	2004	6.7	0
JosÃ© Padilha	492	58607007	Action|Crime|Sci-Fi|Thriller	Gary Oldman	RoboCopÂ	132399	650	English	100000000	2014	6.2	60000

Findings 2

2.Movies with highest profit: Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.

Your task: Find the movies with the highest profit?

- Now as the values in gross and budget columns are very large so I have used a excel to get rid of these large values is not readable that **(go to format cells – select custom – type #,,"million" – then select ok)** it will give us the readable values , i.e 274M.

- Next , we create a column called 'profit' as gross-budget difference in the MySQL workbench to get the desired results. The query can be as follows:

```
SELECT  director_name,
   movie_title,
   ROUND(budget, 0) AS budget,
   ROUND(gross - budget, 0) AS profit
FROM
   moviesp.cleanedm
ORDER BY gross - budget DESC
LIMIT 10;
```

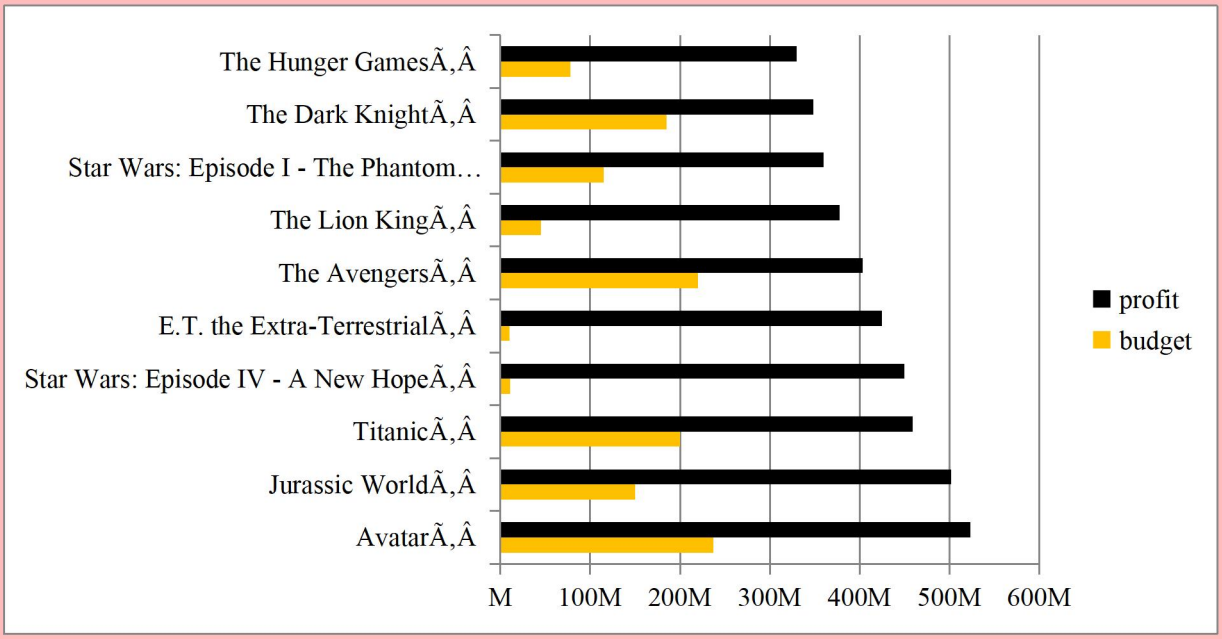
With this query , I have found out the result of movie_title are as follows:

AvatarÂ
Jurassic WorldÂ
TitanicÂ
Star Wars: Episode IV - A New HopeÂ
E.T. the Extra-TerrestrialÂ
The AvengersÂ
The Lion KingÂ
Star Wars: Episode I - The Phantom MenaceÂ
The Dark KnightÂ
The Hunger GamesÂ

director_name	movie_title	budget	profit
James Cameron	AvatarÃ,Â	237M	524M
Colin Trevorrow	Jurassic WorldÃ,Â	150M	502M
James Cameron	*TitanicÃ,Â*	200M	459M
George Lucas	Star Wars: Episode IV - A New HopeÃ,Â	11M	450M
Steven Spielberg	E.T. the Extra-TerrestrialÃ,Â	11M	424M
Joss Whedon	The AvengersÃ,Â	220M	403M
Roger Allers	The Lion KingÃ,Â	45M	378M
George Lucas	Star Wars: Episode I - The Phantom MenaceÃ,Â	115M	360M
Christopher Nolan	The Dark KnightÃ,Â	185M	348M
Gary Ross	The Hunger GamesÃ,Â	78M	330M

This the result of above query , that we have found that Movies with Highest profit . And the insights of this table are:

This is my attached excel file:
https://docs.google.com/spreadsheets/d/1vU6T3pqE0nJI688PbSpdxeA8LBAjKr0R/edit?usp=drive_link



From the above bar chart represents that the highest profit of respected movies have been found to be analyzed and gave insights.

Findings 3

3.Top 250: Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.

Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film. You can use your own imagination also!

Your task: Find IMDB Top 250

a) F

or the first part we should look for all the movies which has higher imdb scores user votes. So that I had solved this question by query with the user votes of higher than 30000 are as follows:

```
SELECT row_number() over(order by imdb_score DESC,
num_voted_users DESC) as ranking,
imdb_score, num_voted_users, movie_title AS IMDb_Top_250,
language
FROM moviesp.cleanedm
WHERE num_voted_users > 30000
LIMIT 250;
```

IMDB Top 250 Movies			
ranking	imdb_score	IMDb_Top_250	language
1	9.3	The Shawshank RedemptionÃ‚Â	English
2	9	The Dark KnightÃ‚Â	English
3	9	The Godfather: Part IIÃ‚Â	English
4	8.9	The Lord of the Rings: The Return of the KingÃ‚Â	English
5	8.9	Schindler's ListÃ‚Â	English
6	8.8	InceptionÃ‚Â	English
7	8.8	Fight ClubÃ‚Â	English
8	8.8	Forrest GumpÃ‚Â	English

9	8.8	The Lord of the Rings: The Fellowship of the RingÃ‚Â	English
10	8.8	Star Wars: Episode V - The Empire Strikes BackÃ‚Â	English
11	8.7	The MatrixÃ‚Â	English
12	8.7	The Lord of the Rings: The Two TowersÃ‚Â	English
13	8.7	Star Wars: Episode IV - A New HopeÃ‚Â	English
14	8.7	GoodfellasÃ‚Â	English
15	8.6	Se7enÃ‚Â	English
16	8.6	InterstellarÃ‚Â	English
17	8.6	The Silence of the LambsÃ‚Â	English
18	8.6	Saving Private RyanÃ‚Â	English
19	8.6	American History XÃ‚Â	English
20	8.6	Spirited AwayÃ‚Â	Japanese
21	8.5	The Dark Knight RisesÃ‚Â	English
22	8.5	GladiatorÃ‚Â	English
23	8.5	Django UnchainedÃ‚Â	English
24	8.5	The DepartedÃ‚Â	English
25	8.5	The PrestigeÃ‚Â	English
26	8.5	The Green MileÃ‚Â	English
27	8.5	Terminator 2: Judgment DayÃ‚Â	English
28	8.5	Back to the FutureÃ‚Â	English
29	8.5	Raiders of the Lost ArkÃ‚Â	English
30	8.5	The Lion KingÃ‚Â	English
31	8.5	The PianistÃ‚Â	English
32	8.5	Apocalypse NowÃ‚Â	English
33	8.5	PsychoÃ‚Â	English
34	8.4	American BeautyÃ‚Â	English
35	8.4	BraveheartÃ‚Â	English
36	8.4	WALLÃ‚Â·EÃ‚Â	English
37	8.4	Star Wars: Episode VI - Return of the JediÃ‚Â	English
38	8.4	AmÃƒÂ©lieÃ‚Â	French
39	8.4	AliensÃ‚Â	English
40	8.4	Once Upon a Time in AmericaÃ‚Â	English
41	8.4	Lawrence of ArabiaÃ‚Â	English
42	8.4	Das BootÃ‚Â	German
43	8.4	Baahubali: The BeginningÃ‚Â	Telugu
44	8.3	Batman BeginsÃ‚Â	English
45	8.3	Inglourious BasterdsÃ‚Â	English
46	8.3	Eternal Sunshine of the Spotless MindÃ‚Â	English
47	8.3	UpÃ‚Â	English
48	8.3	Toy StoryÃ‚Â	English
49	8.3	Good Will HuntingÃ‚Â	English
50	8.3	SnatchÃ‚Â	English
51	8.3	Toy Story 3Ã‚Â	English
52	8.3	ScarfaceÃ‚Â	English
53	8.3	Indiana Jones and the Last CrusadeÃ‚Â	English
54	8.3	2001: A Space OdysseyÃ‚Â	English

55	8.3	L.A. ConfidentialÃ,Â	English
56	8.3	Inside OutÃ,Â	English
57	8.3	UnforgivenÃ,Â	English
58	8.3	AmadeusÃ,Â	English
59	8.3	DownfallÃ,Â	German
60	8.3	Raging BullÃ,Â	English
61	8.3	RoomÃ,Â	English
62	8.3	MetropolisÃ,Â	German
63	8.2	V for VendettaÃ,Â	English
64	8.2	The Wolf of Wall StreetÃ,Â	English
65	8.2	Finding NemoÃ,Â	English
66	8.2	A Beautiful MindÃ,Â	English
67	8.2	Die HardÃ,Â	English
68	8.2	Gran TorinoÃ,Â	English
69	8.2	The Big LebowskiÃ,Â	English
70	8.2	How to Train Your DragonÃ,Â	English
71	8.2	Pan's LabyrinthÃ,Â	Spanish
72	8.2	Blade RunnerÃ,Â	English
73	8.2	Into the WildÃ,Â	English
74	8.2	CasinoÃ,Â	English
75	8.2	WarriorÃ,Â	English
76	8.2	Captain America: Civil WarÃ,Â	English
77	8.2	The ThingÃ,Â	English
78	8.2	Howl's Moving CastleÃ,Â	Japanese
79	8.1	The AvengersÃ,Â	English
80	8.1	Pirates of the Caribbean: The Curse of the Black PearlÃ,Â	English
81	8.1	Shutter IslandÃ,Â	English
82	8.1	Kill Bill: Vol. 1Ã,Â	English
83	8.1	The Sixth SenseÃ,Â	English
84	8.1	Guardians of the GalaxyÃ,Â	English
85	8.1	The Truman ShowÃ,Â	English
86	8.1	Sin CityÃ,Â	English
87	8.1	Jurassic ParkÃ,Â	English
88	8.1	No Country for Old MenÃ,Â	English
89	8.1	Monsters, Inc.Ã,Â	English
90	8.1	Gone GirlÃ,Â	English
91	8.1	Mad Max: Fury RoadÃ,Â	English
92	8.1	The Bourne UltimatumÃ,Â	English
93	8.1	Million Dollar BabyÃ,Â	English
94	8.1	DeadpoolÃ,Â	English
95	8.1	The Grand Budapest HotelÃ,Â	English
96	8.1	The MartianÃ,Â	English
97	8.1	The Imitation GameÃ,Â	English
98	8.1	12 Years a SlaveÃ,Â	English
99	8.1	Groundhog DayÃ,Â	English
100	8.1	The RevenantÃ,Â	English

101	8.1	PrisonersÃ,Â	English
102	8.1	There Will Be BloodÃ,Â	English
103	8.1	The HelpÃ,Â	English
104	8.1	RushÃ,Â	English
105	8.1	The Princess BrideÃ,Â	English
106	8.1	Hotel RwandaÃ,Â	English
107	8.1	SpotlightÃ,Â	English
108	8.1	The Sea InsideÃ,Â	Spanish
109	8.1	Tae Guk Gi: The Brotherhood of WarÃ,Â	Korean
110	8	Slumdog MillionaireÃ,Â	English
111	8	Black SwanÃ,Â	English
112	8	District 9Ã,Â	English
113	8	Catch Me If You CanÃ,Â	English
114	8	X-Men: Days of Future PastÃ,Â	English
115	8	Kill Bill: Vol. 2Ã,Â	English
116	8	Star TrekÃ,Â	English
117	8	The King's SpeechÃ,Â	English
118	8	The IncrediblesÃ,Â	English
119	8	RatatouilleÃ,Â	English
120	8	Casino RoyaleÃ,Â	English
121	8	Casino RoyaleÃ,Â	English
122	8	Life of PiÃ,Â	English
123	8	JawsÃ,Â	English
124	8	Blood DiamondÃ,Â	English
125	8	Rain ManÃ,Â	English
126	8	HerÃ,Â	English
127	8	The Perks of Being a WallflowerÃ,Â	English
128	8	Big FishÃ,Â	English
129	8	Mystic RiverÃ,Â	English
130	8	The Pursuit of HappynessÃ,Â	English
131	8	In BrugesÃ,Â	English
132	8	The ExorcistÃ,Â	English
133	8	Dead Poets SocietyÃ,Â	English
134	8	AladdinÃ,Â	English
135	8	SerenityÃ,Â	English
136	8	MagnoliaÃ,Â	English
137	8	Mulholland DriveÃ,Â	English
138	8	The ArtistÃ,Â	English
139	8	Dances with WolvesÃ,Â	English
140	8	True RomanceÃ,Â	English
141	8	BrazilÃ,Â	English
142	8	Cinderella ManÃ,Â	English
143	8	The Iron GiantÃ,Â	English
144	8	JFKÃ,Â	English
145	8	Dancer in the DarkÃ,Â	English
146	8	Doctor ZhivagoÃ,Â	English

147	7.9	AvatarÃ,Â	English
148	7.9	Iron ManÃ,Â	English
149	7.9	The Hobbit: An Unexpected JourneyÃ,Â	English
150	7.9	TakenÃ,Â	English
151	7.9	The Hobbit: The Desolation of SmaugÃ,Â	English
152	7.9	ShrekÃ,Â	English
153	7.9	Edge of TomorrowÃ,Â	English
154	7.9	The Bourne IdentityÃ,Â	English
155	7.9	The NotebookÃ,Â	English
156	7.9	Toy Story 2Ã,Â	English
157	7.9	Children of MenÃ,Â	English
158	7.9	Edward ScissorhandsÃ,Â	English
159	7.9	Hot FuzzÃ,Â	English
160	7.9	Captain PhillipsÃ,Â	English
161	7.9	E.T. the Extra-TerrestrialÃ,Â	English
162	7.9	Big Hero 6Ã,Â	English
163	7.9	The FighterÃ,Â	English
164	7.9	The Hateful EightÃ,Â	English
165	7.9	How to Train Your Dragon 2Ã,Â	English
166	7.9	The UntouchablesÃ,Â	English
167	7.9	Crouching Tiger, Hidden DragonÃ,Â	Mandarin
168	7.9	Almost FamousÃ,Â	English
169	7.9	Boogie NightsÃ,Â	English
170	7.9	Walk the LineÃ,Â	English
171	7.9	HalloweenÃ,Â	English
172	7.9	HeroÃ,Â	Mandarin
173	7.9	The Blues BrothersÃ,Â	English
174	7.9	Ed WoodÃ,Â	English
175	7.9	The InsiderÃ,Â	English
176	7.9	Letters from Iwo JimaÃ,Â	Japanese
177	7.9	Straight Outta ComptonÃ,Â	English
178	7.9	GloryÃ,Â	English
179	7.9	GloryÃ,Â	English
180	7.9	My Fair LadyÃ,Â	English
181	7.9	The Remains of the DayÃ,Â	English
182	7.9	The Right StuffÃ,Â	English
183	7.9	The World's Fastest IndianÃ,Â	English
184	7.8	The HangoverÃ,Â	English
185	7.8	GravityÃ,Â	English
186	7.8	Silver Linings PlaybookÃ,Â	English
187	7.8	SkyfallÃ,Â	English
188	7.8	X-Men: First ClassÃ,Â	English
189	7.8	Captain America: The Winter SoldierÃ,Â	English
190	7.8	The Curious Case of Benjamin ButtonÃ,Â	English
191	7.8	DriveÃ,Â	English
192	7.8	Ocean's ElevenÃ,Â	English

193	7.8	Star Trek Into DarknessÃ‚Â	English
194	7.8	Birdman or (The Unexpected Virtue of Ignorance)Ã‚Â	English
195	7.8	Harry Potter and the Prisoner of AzkabanÃ‚Â	English
196	7.8	The Bourne SupremacyÃ‚Â	English
197	7.8	Back to the Future Part IIÃ‚Â	English
198	7.8	The Girl with the Dragon TattooÃ‚Â	English
199	7.8	American GangsterÃ‚Â	English
200	7.8	TangledÃ‚Â	English
201	7.8	PredatorÃ‚Â	English
202	7.8	Wreck-It RalphÃ‚Â	English
203	7.8	Lucky Number SlevinÃ‚Â	English
204	7.8	The GameÃ‚Â	English
205	7.8	Being John MalkovichÃ‚Â	English
206	7.8	The Fault in Our StarsÃ‚Â	English
207	7.8	The Lego MovieÃ‚Â	English
208	7.8	3:10 to YumaÃ‚Â	English
209	7.8	Moonrise KingdomÃ‚Â	English
210	7.8	ApocalyptoÃ‚Â	Maya
211	7.8	Donnie BrascoÃ‚Â	English
212	7.8	O Brother, Where Art Thou?Ã‚Â	English
213	7.8	GattacaÃ‚Â	English
214	7.8	The FugitiveÃ‚Â	English
215	7.8	About TimeÃ‚Â	English
216	7.8	Office SpaceÃ‚Â	English
217	7.8	ChangelingÃ‚Â	English
218	7.8	Pride & PrejudiceÃ‚Â	English
219	7.8	AtonementÃ‚Â	English
220	7.8	The Big ShortÃ‚Â	English
221	7.8	Finding NeverlandÃ‚Â	English
222	7.8	What's Eating Gilbert GrapeÃ‚Â	English
223	7.8	South Park: Bigger Longer & UncutÃ‚Â	English
224	7.8	Remember the TitansÃ‚Â	English
225	7.8	Fantastic Mr. FoxÃ‚Â	English
226	7.8	The Boy in the Striped PajamasÃ‚Â	English
227	7.8	The Last of the MohicansÃ‚Â	English
228	7.8	The Jungle BookÃ‚Â	English
229	7.8	The Jungle BookÃ‚Â	English
230	7.8	Kung Fu HustleÃ‚Â	Cantonese
231	7.8	TombstoneÃ‚Â	English
232	7.8	NebraskaÃ‚Â	English
233	7.8	Glengarry Glen RossÃ‚Â	English
234	7.8	The Last EmperorÃ‚Â	English
235	7.8	The Best OfferÃ‚Â	English
236	7.8	The Conjuring 2Ã‚Â	English
237	7.8	The Color PurpleÃ‚Â	English
238	7.8	Black BookÃ‚Â	Dutch

239	7.8	The White RibbonÃ‚Â	German
240	7.8	HamletÃ‚Â	English
241	7.8	The Little PrinceÃ‚Â	English
242	7.8	The VerdictÃ‚Â	English
243	7.7	TitanicÃ‚Â	English
244	7.7	300Ã‚Â	English
245	7.7	The Social NetworkÃ‚Â	English
246	7.7	ArgoÃ‚Â	English
247	7.7	Kick-AssÃ‚Â	English
248	7.7	Minority ReportÃ‚Â	English
249	7.7	Cast AwayÃ‚Â	English
250	7.7	WatchmenÃ‚Â	English

From the above query , I got this **IMDb_Top_250** , I had attached as proof of this.

This is my attached excel file:

https://docs.google.com/spreadsheets/d/16NnCBQPx46wDCiuNatK76Pn8Ra47QAzb/edit?usp=drive_link

b)For the next part as , out of these top250 movies with a highest user votes , we search for all the **Top_Foreign_Lang_Film** using the below query :

```
SELECT  row_number() over(order by imdb_score DESC,
num_voted_users DESC) as ranking,
        imdb_score,
        num_voted_users,
        movie_title AS Top_non_english,
        language
FROM
        moviesp.cleanedm
```

```
WHERE
        num_voted_users > 30000  and  language <> 'English'
LIMIT 36;
```

ranking	imdb_score	num_voted_users	Top_non_english	language
1	8.6	417971	Spirited AwayÃ‚Â	Japanese
2	8.4	534262	AmÃƒÂ©lieÃ‚Â	French
3	8.4	168203	Das BootÃ‚Â	German
4	8.4	62756	Baahubali: The BeginningÃ‚Â	Telugu
5	8.3	248354	DownfallÃ‚Â	German
6	8.3	111841	MetropolisÃ‚Â	German
7	8.2	467234	Pan's LabyrinthÃ‚Â	Spanish
8	8.2	214091	Howl's Moving CastleÃ‚Â	Japanese
9	8.1	64556	The Sea InsideÃ‚Â	Spanish
10	8.1	31943	Tae Guk Gi: The Brotherhood of WarÃ‚Â	Korean
11	7.9	217740	Crouching Tiger, Hidden DragonÃ‚Â	Mandarin
12	7.9	149414	HeroÃ‚Â	Mandarin
13	7.9	132149	Letters from Iwo JimaÃ‚Â	Japanese
14	7.8	236000	ApocalyptoÃ‚Â	Maya
15	7.8	99353	Kung Fu HustleÃ‚Â	Cantonese
16	7.8	59507	Black BookÃ‚Â	Dutch
17	7.8	52958	The White RibbonÃ‚Â	German
18	7.7	85589	PonyoÃ‚Â	Japanese
19	7.7	62607	A Very Long EngagementÃ‚Â	French
20	7.7	55516	The Great BeautyÃ‚Â	Italian
21	7.6	92295	House of Flying DaggersÃ‚Â	Mandarin
22	7.6	68119	The Kite RunnerÃ‚Â	Dari
23	7.6	38690	The Flowers of WarÃ‚Â	Mandarin
24	7.5	77656	The Painted VeilÃ‚Â	Mandarin
25	7.5	41496	2046Ã‚Â	Cantonese
26	7.4	36894	Red CliffÃ‚Â	Mandarin
27	7.3	63084	Paris, je t'aimeÃ‚Â	French
28	7.3	37635	Mongol: The Rise of Genghis KhanÃ‚Â	Mongolian
29	7.2	55928	District B13Ã‚Â	French
30	7.1	179235	The Passion of the ChristÃ‚Â	Aramaic
31	7	36455	Curse of the Golden FlowerÃ‚Â	Mandarin
32	6.7	32003	Coco Before ChanelÃ‚Â	French
33	6.4	86152	The InterpreterÃ‚Â	Aboriginal
34	5.9	71574	The Legend of ZorroÃ‚Â	Spanish
35	4.3	31414	In the Land of Blood and HoneyÃ‚Â	Bosnian

- From the above query , I have found 35 movies list with respect to the **num_voted_users** for non English movies with the top 250 lists.
- Then with this results , have been queried with respect to the number of voters for the most favorable movies with ratings of imdb_score, that is non_english that is foreign language movies that have been listed using the query. So in this question Top 250 and Top_Foreign_Lang_Film(35 movies) , MySQL workbench had been used to get the desired results and with the results , I had transferred the results to excel to make table to be readable.
- This is my file attached here: https://drive.google.com/file/d/1sDk5gBzVoLVdENnGPDrv77P4AksNeHbW/view?usp=drive_link

Findings 4

1. **Best Directors:** Group the column using the director_name column.

Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.

Your task: Find the best directors

For this question ,want to find out the top 10 directors whom with the imdb_score of highest one . This will be found out by using the below query:

```
SELECT
    director_name AS Top_10_Director,
    ROUND(AVG(imdb_score), 2) AS avg_rating
FROM
    moviesp.cleanedm
GROUP BY director_name
ORDER BY avg_rating DESC , director_name
LIMIT 10;
```

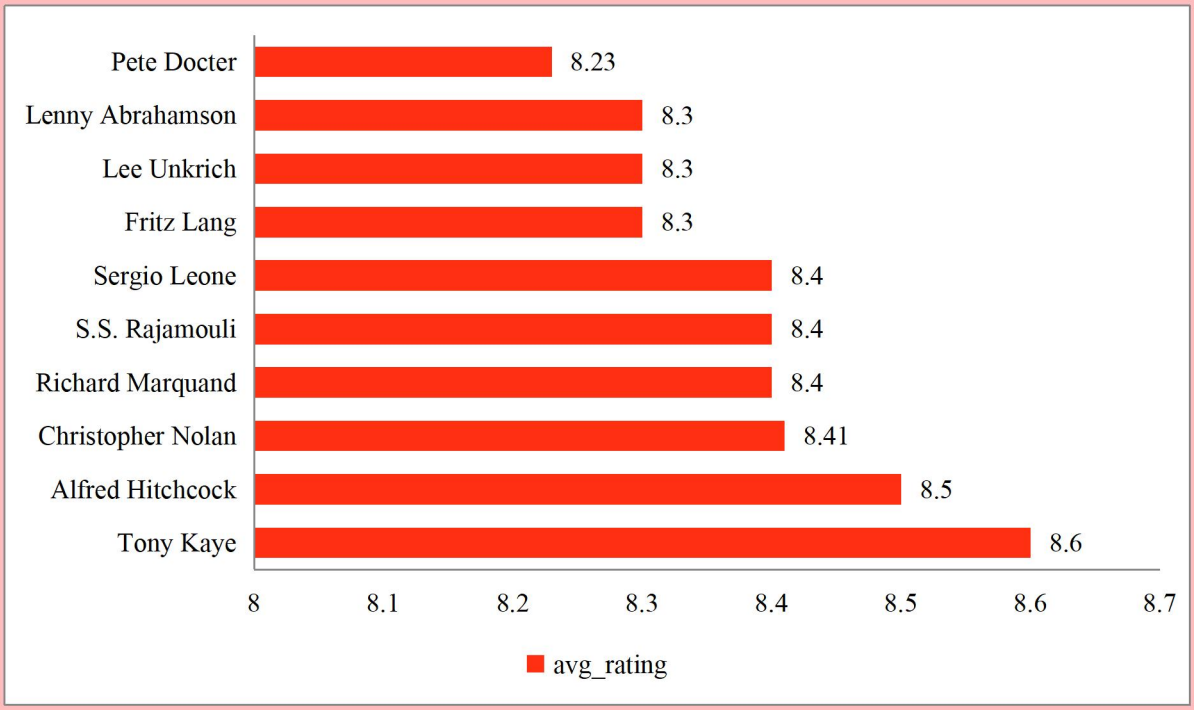
With this query , I have got the results…..

Top 10 Directors	
Top_10_Director	avg_rating
Tony Kaye	8.6
Alfred Hitchcock	8.5
Christopher Nolan	8.41
Richard Marquand	8.4
S.S. Rajamouli	8.4
Sergio Leone	8.4
Fritz Lang	8.3
Lee Unkrich	8.3
Lenny Abrahamson	8.3
Pete Docter	8.23

From the above results , I have inferred that the top 10 directors have been listed by using the sql query results with the below insights.

This is my excel file attached:

https://docs.google.com/spreadsheets/d/1kDmzbNlehAevghQAK8UPAuYi1rA3jlXj/edit?usp=drive_link



Findings 5

2. **Popular Genres:** Perform this step using the knowledge gained while performing previous steps.

Your task: Find popular genres

❖ For this question, we first split the genres of each movie into individual genres into 8 separate levels of genres. Then, we make a list of all genres by uniting all

levels of genres and then count the popularity of each genre.

- I have been using sql query to solve this description of separating 8 levels of genres with list of all genres the with mean of gross and the number of movies.

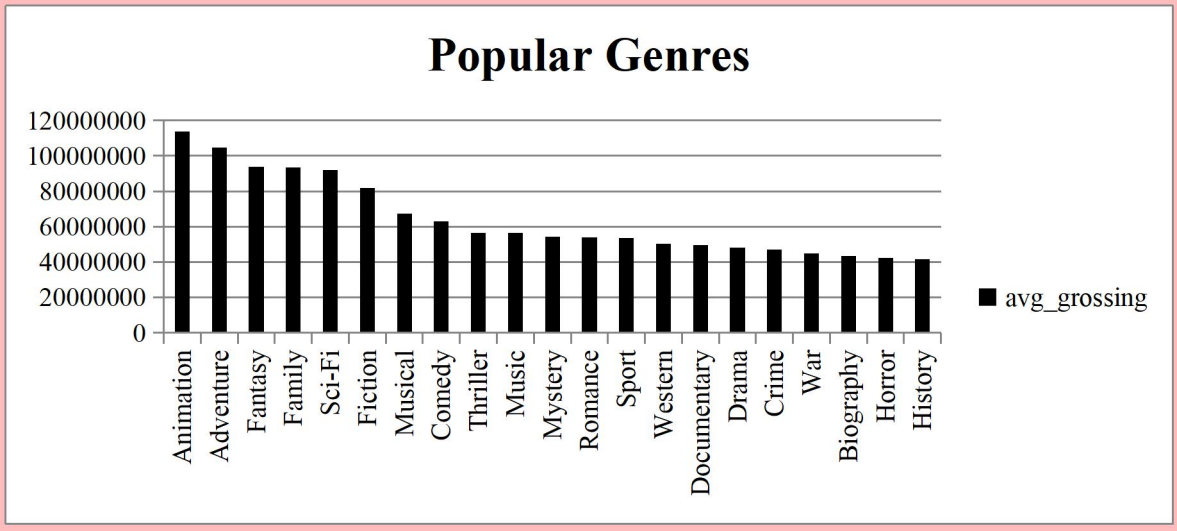
```
With d8 as
 (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) =
genre_1 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_2 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_3 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_4 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_5 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_6 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_7 THEN NULL
 ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) END) AS genre_8
 FROM
 (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) =
genre_1 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_2 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_3 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_4 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_5 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_6 THEN NULL
  ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) END) AS genre_7
  FROM
  (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) =
genre_1 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_2 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_3 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_4 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_5 THEN NULL
  ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) END) AS genre_6
  FROM
  (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) =
genre_1 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_2 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_3 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_4 THEN NULL
  ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) END) AS genre_5
  FROM(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1)
= genre_1 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) = genre_2 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) = genre_3 THEN NULL
  ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) END) AS genre_4
  FROM
  (SELECT *,(CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 3), '|', -1) =
genre_1 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 3), '|', -1) = genre_2 THEN NULL
  ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 3), '|', -1) END) AS genre_3
```

```
 FROM
 (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 2), '|', -1) =
genre_1 THEN NULL
 ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 2), '|', -1) END) AS genre_2
 FROM
 (SELECT
SUBSTRING(genres, 2) as genres,
SUBSTRING_INDEX(SUBSTRING_INDEX(SUBSTRING(genres, 2), '|', 1), '|', - 1) AS genre_1
 FROM moviesp.cleanedm) d1) d2) d3) d4) d5) d6) d7),
 list as
 (SELECT DISTINCT genre_1 AS genre_list
 FROM
 (
 (SELECT genre_1 FROM d8) UNION ALL
 (SELECT genre_2 FROM d8) UNION ALL
 (SELECT genre_3 FROM d8) UNION ALL
 (SELECT genre_4 FROM d8) UNION ALL
 (SELECT genre_5 FROM d8) UNION ALL
 (SELECT genre_6 FROM d8) UNION ALL
 (SELECT genre_7 FROM d8) UNION ALL
 (SELECT genre_8 FROM d8)) g12
 WHERE genre_1 IS NOT NULL ORDER BY genre_1)
 SELECT genre_list,round(AVG(DISTINCT CASE WHEN INSTR(genres, genre_list) THEN
gross ELSE NULL END), 2) AS avg_grossing,
 COUNT(DISTINCT CASE WHEN INSTR(genres, genre_list) THEN genres ELSE NULL END)
AS no_of_movies
 FROM cleanedm JOIN list
 GROUP BY genre_list
 ORDER BY avg_grossing DESC;

 SELECT actor_1_name, COUNT(movie_title) AS no_of_movies,
 ROUND(AVG(num_user_for_reviews), 2) AS user_reviews,
 ROUND(AVG(num_critic_for_reviews), 2) AS critic_reviews
 FROM
 (
 (SELECT actor_1_name, movie_title,num_user_for_reviews, num_critic_for_reviews
 FROM moviesp.cleanedm
 WHERE actor_1_name = ' CCH Pounder')
 UNION ALL
 (SELECT actor_1_name, movie_title,num_user_for_reviews, num_critic_for_reviews
 FROM moviesp.cleanedm
 WHERE actor_1_name = ' J.K. Simmons')
 UNION ALL
 (SELECT actor_1_name, movie_title, num_user_for_reviews, num_critic_for_reviews
 FROM moviesp.cleanedm
 WHERE actor_1_name = ' Jennifer Lawrence')) c
 GROUP BY actor_1_name
 ORDER BY user_reviews DESC , critic_reviews DESC;
```

Popular Genres		
genre_list	avg_grossing	no_of_movies
Animation	113547233.2	84
Adventure	104744835.7	238
Fantasy	93680359.4	170
Family	93247050.7	178

Sci-Fi	91891509.05	134
Fiction	81796457.68	235
Musical	67450846.54	53
Comedy	62876377.93	263
Thriller	56544554.87	179
Music	56385988.68	101
Mystery	54357927.28	97
Romance	53797743.97	171
Sport	53731204.96	39
Western	50326576.6	33
Documentary	49478940.43	4
Drama	47978352.15	310
Crime	46975720.56	121
War	44756370.57	60
Biography	43329712.5	63
Horror	42187374.12	70
History	41782106.2	51

From the above query results , There are 23 Genres got to inferred that the splitted genres with the mean of gross and number of movies match lists. And got the analysis of insights of column chart are as follows:



This is my attached excel file:
https://docs.google.com/spreadsheets/d/13gHz7H3oDULFNXcZJIRAB2g6VOb-UgVZ/edit?usp=drive_link

Findings 6

3. **Charts:** Create three new columns namely, **Meryl_Streep, Leo_Caprio, and Brad_Pitt** which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined.

Group the combined column using the **actor_1_name** column.
Find the mean of the **num_critic_for_reviews** and **num_users_for_review** and identify the actors which have the highest mean.

Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as

1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called **df_by_decade**.

Your task: Find the critic-favorite and audience-favorite actors

a) For this critic-favourite , the assigned favourite movies has been extracted which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors.

❖ So I have queried for extracting the provided favourite actors with the number of movies , user reviews (audience-favourite) and critic-reviews. The query for finding this as follows:

```
SELECT actor_1_name, COUNT(movie_title) AS no_of_movies,
  ROUND(AVG(num_user_for_reviews), 2) AS user_reviews,
  ROUND(AVG(num_critic_for_reviews), 2) AS critic_reviews
  FROM
  (
  (SELECT actor_1_name, movie_title,num_user_for_reviews,
num_critic_for_reviews
  FROM moviesp.cleanedm
  WHERE actor_1_name = 'Meryl Streep')
  UNION ALL
```

```
 (SELECT actor_1_name, movie_title,num_user_for_reviews,
num_critic_for_reviews
 FROM moviesp.cleanedm
 WHERE actor_1_name = 'Leonardo DiCaprio')
 UNION ALL
 (SELECT actor_1_name, movie_title, num_user_for_reviews,
num_critic_for_reviews
 FROM moviesp.cleanedm
 WHERE actor_1_name = 'Brad Pitt')) c
 GROUP BY actor_1_name
 ORDER BY user_reviews DESC , critic_reviews DESC;
```

Favourite actors			
actor_1_name	num_of_movies	user_reviews	critic_reviews
Leonardo DiCaprio	20	922.55	322.2
Brad Pitt	17	742.35	245
Meryl Streep	11	297.18	181.45

- From the above query , I got this results which inferred of extracting from the actor name 1 with he help of user reviews and critic reviews.
- And this query helped me to find out that Leonardo DiCaprio is the most favorite actor for both the audience and the critics.
- This is the file attached here: https://docs.google.com/spreadsheets/d/1046zg_F9vOST467Ewu6HOAwdWET-C-Bn/edit?usp=drive_link

b) Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called **df_by_decade.**

- For this decades per number user votes , we will bw finding out extracting the years from MySQL query workbench are as follows:

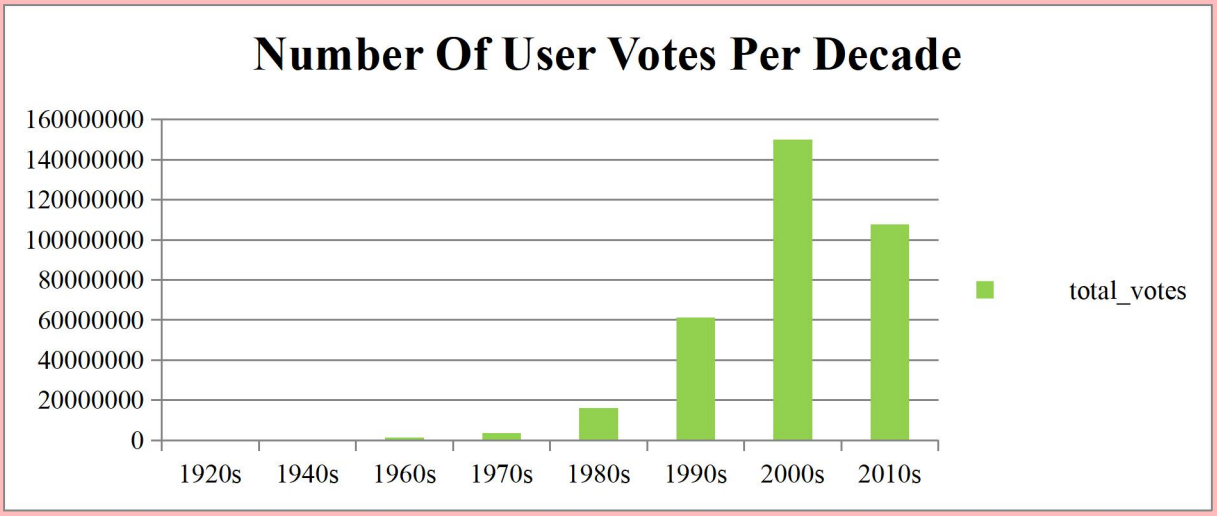
```
SELECT
    CONCAT(CONVERT( FLOOR(title_year / 10) * 10 ,
CHAR),
            's') AS decade,
    SUM(num_voted_users) AS total_votes
FROM
    moviesp.cleanedm
GROUP BY decade
ORDER BY decade;
```

This is excel attached here:
https://docs.google.com/spreadsheets/d/1B3LzOFwL59h8d2jBRb6YJ73Y3tmHTAiR/edit?usp=drive_link

Number Of User Votes Per Decade	
decade	total_votes
1920s	111841

1940s	90360
1960s	1218680
1970s	3679621
1980s	16136813
1990s	61213572
2000s	150036196
2010s	107714560

From the above query , I have inferred that the results of decades per number of user votes with the voting that are extracted using the query using aggregation function .



From the above column chart , I have analyzed a insights of decades over the number of user votes.

Conclusion

In Conclusion, I would like to conclude that IMDB Movie Analysis or any such analysis is done not only by Movie makers before movie production, but it is also done by various investors, stake holders, theatre outlet owners.

Normal people would not mind to do such analysis but such analysis plays an crucial part during the pre-production phase of the movies and also during the post-production phase.

Also, it is not necessary that the movie with the highest IMDB rating will have the highest profit. Profit is calculated truly on the basis on the number of tickets sold by theatres all over the world .

Most of the people are tired with their daily lives and they prefer movies with Comedy/ Drama genre or both, and they would not go for movies with Action/Horror genre So, directors and production team must keep in mind the above points and shall do the pre-production analysis before the commencement of filming.

Bank Loan Case Study

Final Project-2

DESCRIPTION

Imagine you're a data analyst at a finance company that specializes in lending various types of loans to urban customers. Your company faces a challenge: some customers who don't have a sufficient credit history take advantage of this and default on their loans. Your task is to use Exploratory Data Analysis (EDA) to analyze patterns in the data and ensure that capable applicants are not rejected.

The Problem

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

It aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected.

Identification of such applicants using EDA is the aim of this case study. In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

To develop your understanding of the domain, you are advised to independently research a little about risk analytics – understanding the types of variables and their significance should be enough)

Design

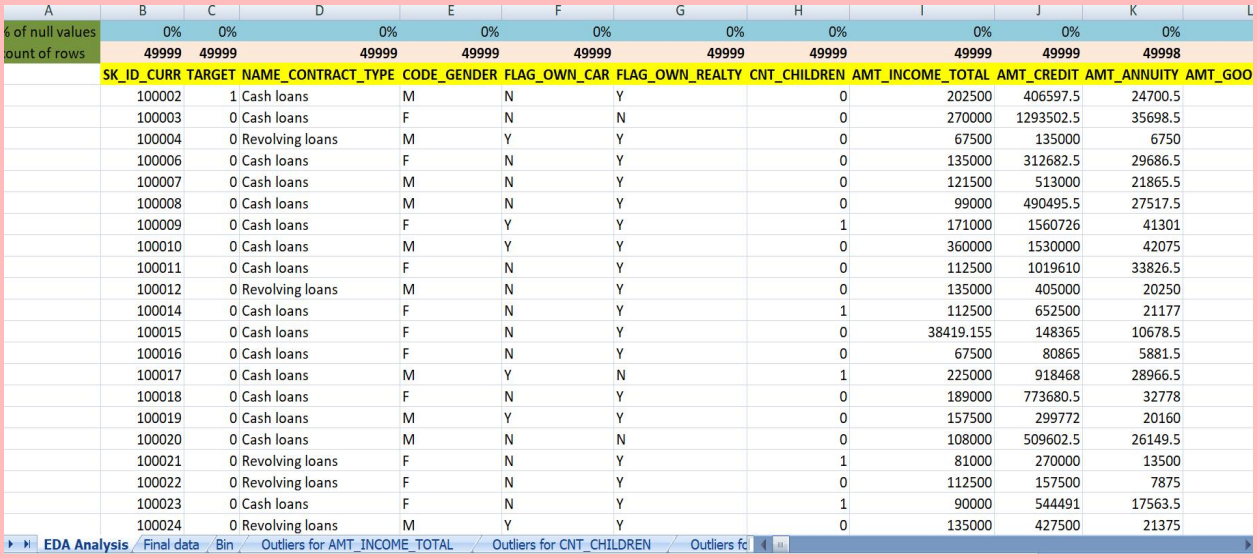
- First of all , in this project they have provided with a 2 datasets which are **current application** and **previous application.** So I have started cleaning the datasets and grouped as one, removed blanks, unwanted and unrelated columns and so on.

- I thought that in the given datasets unneeded columns are present there so I thought to remove the useless

columns which are not needed for risk analysis. I analyzed in given reference that helped me to analyze this risk assessment analysis.

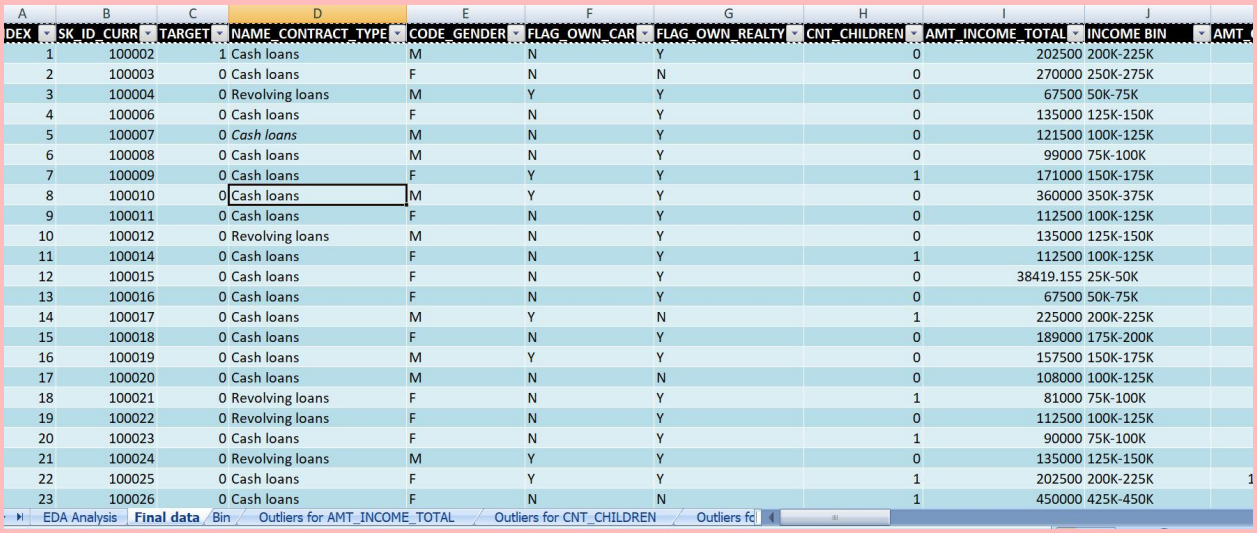
- Next, I am supposed to find the outliers then removed the outliers, then I count the blanks of making it percentage how much the blanks are present so the percentage helped me to remove the unwanted columns and blanks by using excel **count functions.**
- By finding and cleaning all these data and moving on to tasks , my excel file named as **Final data** helped me to calculate the tasks and visually given a various insights like bar chart, box plot, column chart by make using excel pivot tables.

Findings 1



	A	B	C	D	E	F	G	H	I	J	K	L
	% of null values	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
	ount of rows	49999	49999	49999	49999	49999	49999	49999	49999	49999	49998	
		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_GOO
		100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	
		100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	
		100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	
		100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	
		100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	
		100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	
		100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	
		100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	
		100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	
		100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	
		100014	0	Cash loans	F	N	Y	1	112500	652500	21177	
		100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	
		100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	
		100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	
		100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	
		100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	
		100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	
		100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	
		100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	
		100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	
		100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	

EDA Analysis | Final data | Bin | Outliers for AMT_INCOME_TOTAL | Outliers for CNT_CHILDREN | Outliers fc



	A	B	C	D	E	F	G	H	I	J
	DEX	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	INCOME BIN
1	1	100002	1	Cash loans	M	N	Y	0	202500	200K-225K
2	2	100003	0	Cash loans	F	N	N	0	270000	250K-275K
3	3	100004	0	Revolving loans	M	Y	Y	0	67500	50K-75K
4	4	100006	0	Cash loans	F	N	Y	0	135000	125K-150K
5	5	100007	0	*Cash loans*	M	N	Y	0	121500	100K-125K
6	6	100008	0	Cash loans	M	N	Y	0	99000	75K-100K
7	7	100009	0	Cash loans	F	Y	Y	1	171000	150K-175K
8	8	100010	0	Cash loans	M	Y	Y	0	360000	350K-375K
9	9	100011	0	Cash loans	F	N	Y	0	112500	100K-125K
10	10	100012	0	Revolving loans	M	N	Y	0	135000	125K-150K
11	11	100014	0	Cash loans	F	N	Y	1	112500	100K-125K
12	12	100015	0	Cash loans	F	N	Y	0	38419.155	25K-50K
13	13	100016	0	Cash loans	F	N	Y	0	67500	50K-75K
14	14	100017	0	Cash loans	M	Y	N	1	225000	200K-225K
15	15	100018	0	Cash loans	F	N	Y	0	189000	175K-200K
16	16	100019	0	Cash loans	M	Y	Y	0	157500	150K-175K
17	17	100020	0	Cash loans	M	N	N	0	108000	100K-125K
18	18	100021	0	Revolving loans	F	N	Y	1	81000	75K-100K
19	19	100022	0	Revolving loans	F	N	Y	0	112500	100K-125K
20	20	100023	0	Cash loans	F	N	Y	1	90000	75K-100K
21	21	100024	0	Revolving loans	M	Y	Y	0	135000	125K-150K
22	22	100025	0	Cash loans	F	Y	Y	1	202500	200K-225K
23	23	100026	0	Cash loans	F	N	N	1	450000	425K-450K

EDA Analysis | Final data | Bin | Outliers for AMT_INCOME_TOTAL | Outliers for CNT_CHILDREN | Outliers fc

From the above picture, I have represented the EDA analysis by percentage values and final dataset which are cleaned are pasted above as picture. Also then I have included a excel file as **EDA analysis** and **Final data**.

- SK_ID_PREV
- SK_ID_CURR
- NAME_CONTRACT_TYPE
- AMT_ANNUITY

- AMT_APPLICATION
- AMT_CREDIT
- AMT_GOODS_PRICE
- WEEKDAY_APPR_PROCESS_START
- HOUR_APPR_PROCESS_START
- NAME_CONTRACT_STATUS
- DAYS_DECISION
- NAME_PAYMENT_TYPE
- CODE_REJECT_REASON
- NAME_CLIENT_TYPE
- NAME_GOODS_CATEGORY
- CNT_PAYMENT
- PRODUCT_COMBINATION

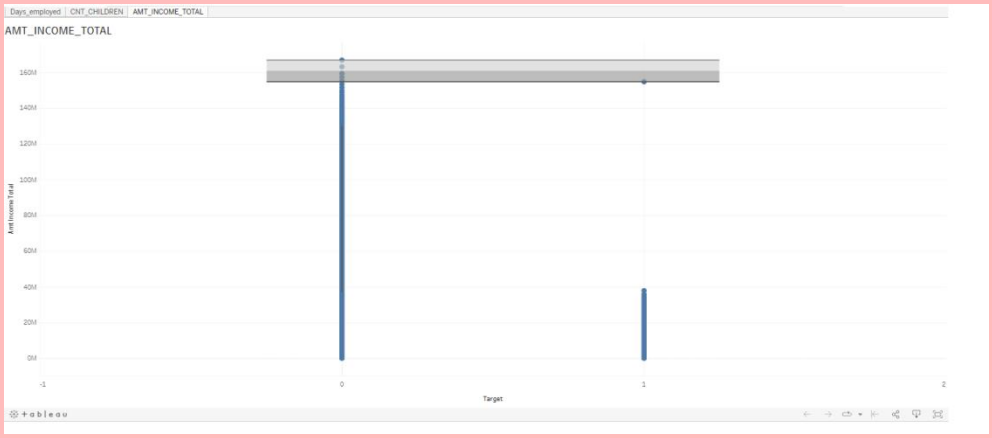
	A	B	C	D	E	F	G	H	I	J
	% of null values	0%	0%	0%	21%	0%	0%	50%	21%	0%
	Count of Rows	49999	49999	49999	39407	49999	49999	24801	39255	49999
		SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START
		2030495	271877	Consumer 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

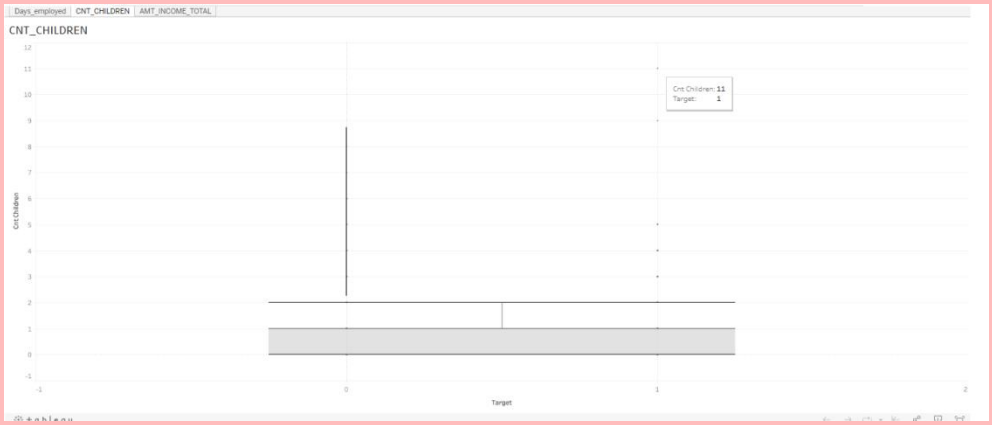
K_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	HOUR_AP	NAME_CONTRACT_STATUS
2030495	271877	Consumer loans	1730.43	17145	17145	17145	SATURDAY	15	Approved
2802425	108129	Cash loans	25188.615	607500	679671	607500	THURSDAY	11	Approved
2523466	122040	Cash loans	15060.735	112500	136444.5	112500	TUESDAY	11	Approved
2819243	176158	Cash loans	47041.335	450000	470790	450000	MONDAY	7	Approved
1784265	202054	Cash loans	31924.395	337500	404055	337500	THURSDAY	9	Refused
1383531	199383	Cash loans	23703.93	315000	340573.5	315000	SATURDAY	8	Approved
1715995	447712	Cash loans	11368.62	270000	335754	270000	FRIDAY	7	Approved
2257824	161140	Cash loans	13832.775	211500	246397.5	211500	FRIDAY	10	Approved
2330894	258628	Cash loans	12165.21	148500	174361.5	148500	TUESDAY	15	Approved
1397919	321676	Consumer loans	7654.86	53779.5	57564	53779.5	SUNDAY	15	Approved
2273188	270658	Consumer loans	9644.22	26550	27252	26550	SATURDAY	10	Approved
1232483	151612	Consumer loans	21307.455	126490.5	119853	126490.5	TUESDAY	7	Approved
2163253	154602	Consumer loans	4187.34	26955	27297	26955	SATURDAY	12	Approved
1285768	142748	Revolving loans	9000	180000	180000	180000	FRIDAY	13	Approved
2393109	396305	Cash loans	10181.7	180000	180000	180000	THURSDAY	14	Approved
1173070	199178	Cash loans	4666.5	45000	49455	45000	SATURDAY	16	Refused
1506815	166490	Cash loans	25454.025	450000	491580	450000	MONDAY	6	Refused
1182516	267782	Cash loans	20361.6	405000	451777.5	405000	SATURDAY	4	Approved
1172937	302212	Cash loans	39475.305	1129500	1277104.5	1129500	THURSDAY	5	Refused
1543131	275707	Cash loans	22619.52	229500	241920	229500	THURSDAY	8	Approved
2536650	338725	Cash loans	16708.32	369000	369000	369000	WEDNESDAY	13	Approved
1676258	433469	Cash loans	22242.825	247500	268083	247500	THURSDAY	14	Approved
2075578	418383	Consumer loans	7656.705	74610	65610	74610	MONDAY	14	Approved

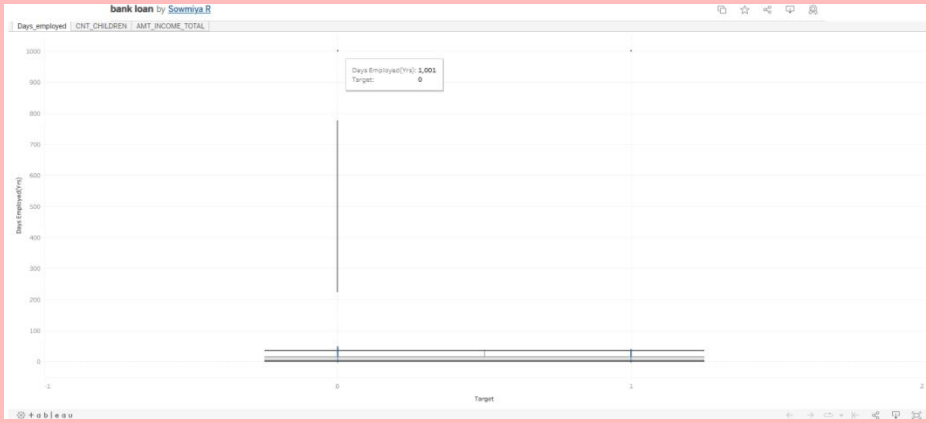
From the above 2 pictures , I had shown that counting the number of null rows with the percentage of every other columns and resulted to be 39256 entries and 17 header rows.

Findings 2

Through then **AMT_INCOME_TOTAL, CNT_CHILDREN, DAYS_EMPLOYED(yrs)** has some number of outliers.



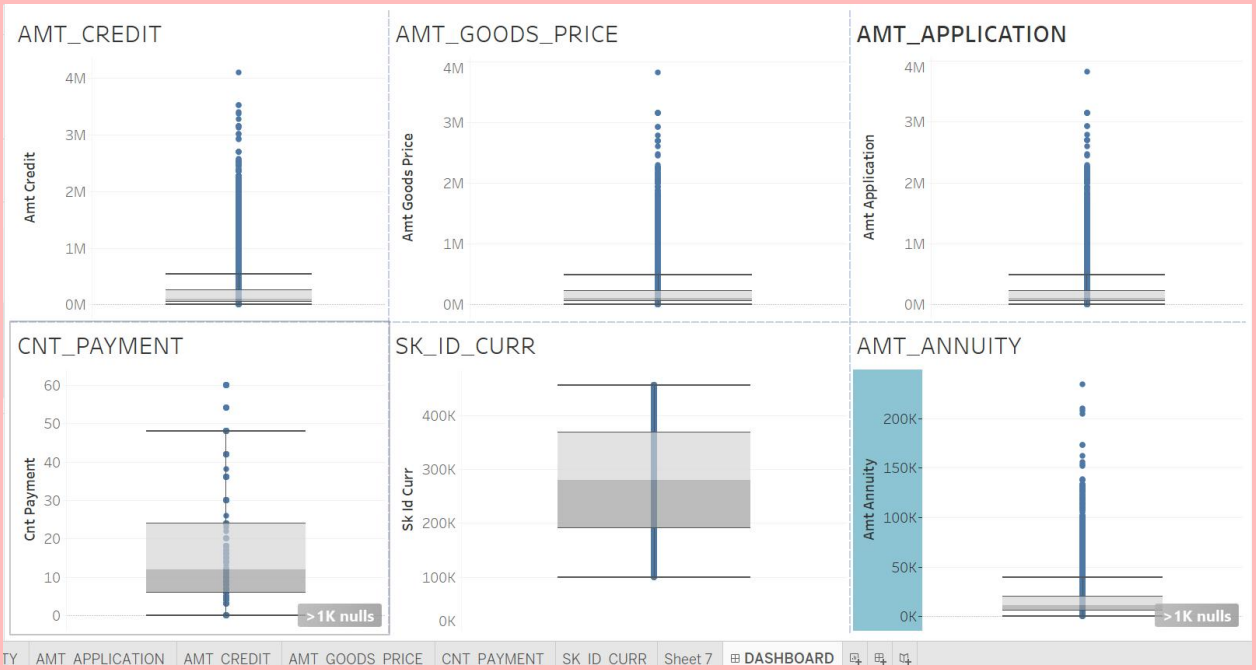




I have used Tableau for finding out the outliers box plot which are visually represented in the above 3 images .

Quartile 1
87750
Quartile 3
248273
Inter Quartile Range
160523
Upper Limit
489057
Lower Limit
328534

This is the findings of outliers quartile ranges for amt_income_total which are shown above.



From the above dashboard is then ouliers of amt_credit, amt_goods_price, amt_application, cnt_payment, sk_id_curr, amt_annuity and inferred that number of outliers with the relation between the previous application.

Findings 3

In this data imbalance tasks, I have inferred that accuracy of data in the representation of visual charts as counting off the applicants which are grouped as **0** and **1** .

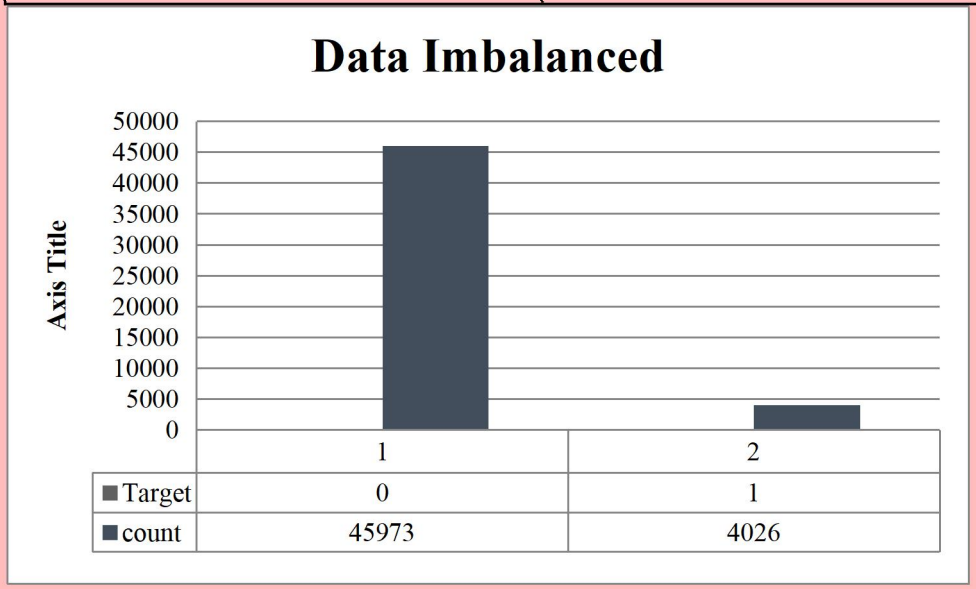
Target	count
0	45973
1	4026

Grand Total	49999

The above table says that the total number of applicants as Target (0 and 1) in the count which is total of 49999.

	Cnt of 0's and 1's	Ratio
0	45,973	11.42
1	4,026	

Target	Contribution
0	92%
1	8%



From the above chart represents the data imbalancement of target as 0 and 1 (applicants) and the count of those referred to the repayment of loan status.

Findings 4

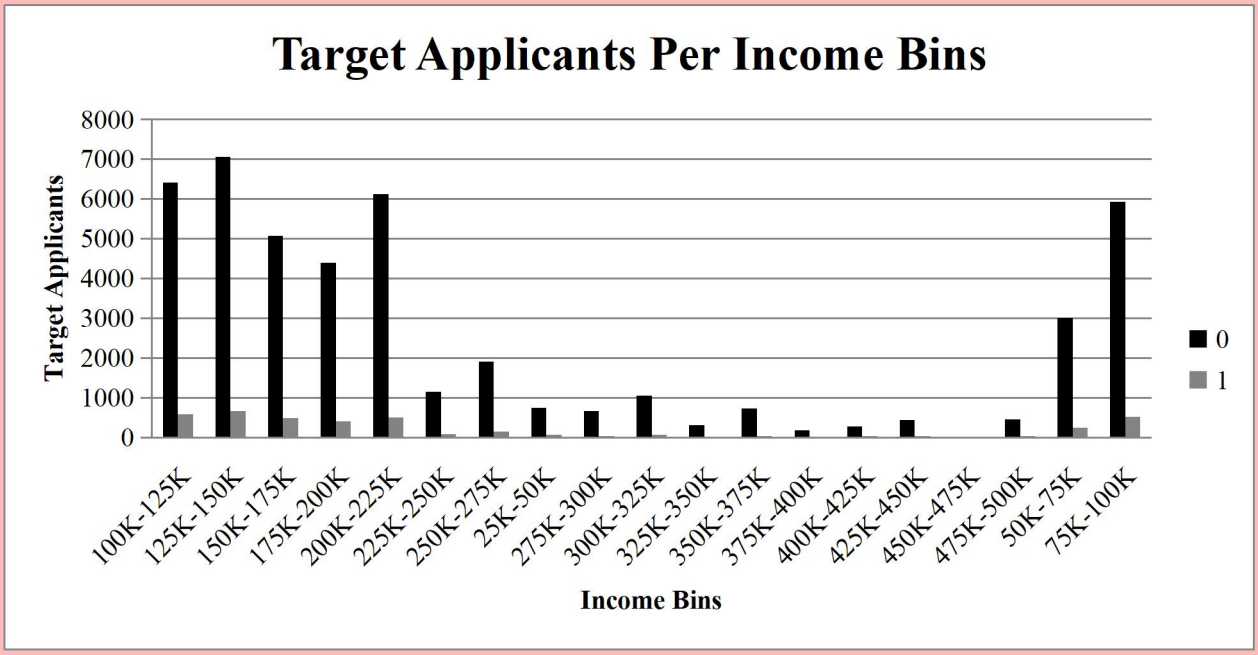
In this univariate and bivariate analysis , I had found out the class intervals of income bins and credit bins .

Income	Income Bins
0	0-25K
25,001	25K-50K
50,001	50K-75K
75,001	75K-100K
1,00,001	100K-125K
1,25,001	125K-150K
1,50,001	150K-175K
1,75,001	175K-200K
2,00,001	200K-225K
2,25,001	225K-250K
2,50,001	250K-275K
2,75,001	275K-300K
3,00,001	300K-325K
3,25,001	325K-350K
3,50,001	350K-375K
3,75,001	375K-400K
4,00,001	400K-425K
4,25,001	425K-450K
4,50,001	450K-475K
4,75,001	475K-500K
5,00,001	5 Lacs and above

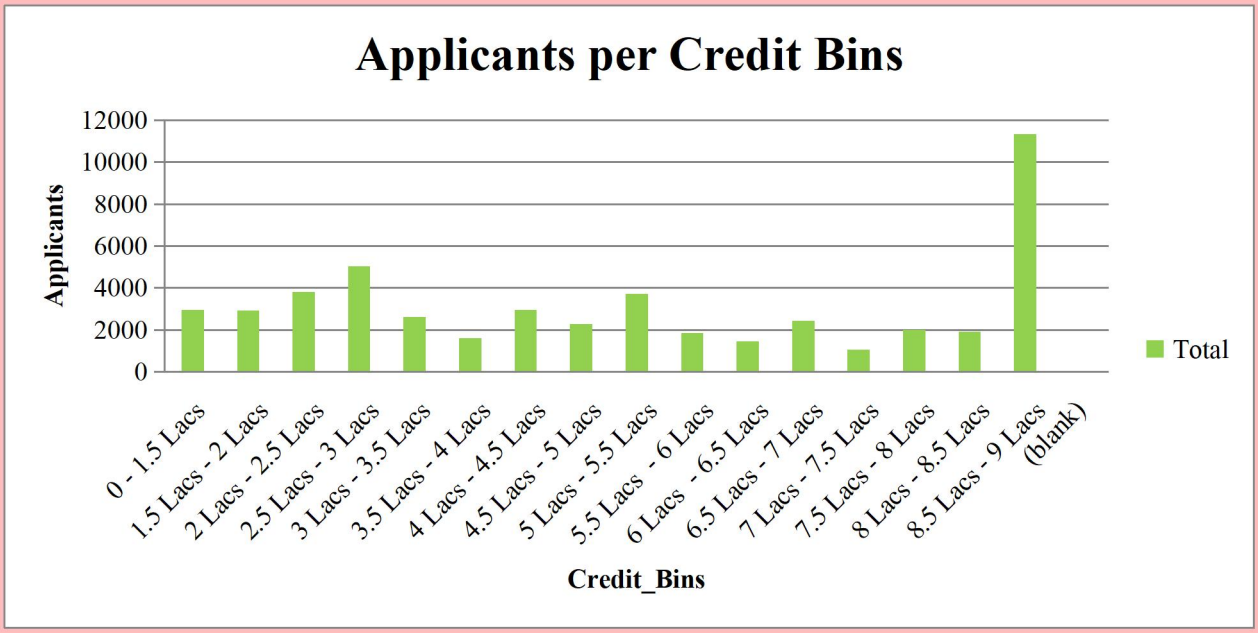
Credit	Credit Bins
0	0 - 1.5 Lacs
1,50,001	1.5 Lacs - 2 Lacs
2,00,001	2 Lacs - 2.5 Lacs
2,50,001	2.5 Lacs - 3 Lacs
3,00,001	3 Lacs - 3.5 Lacs
3,50,001	3.5 Lacs - 4 Lacs
4,00,001	4 Lacs - 4.5 Lacs
4,50,001	4.5 Lacs - 5 Lacs
5,00,001	5 Lacs - 5.5 Lacs
5,50,001	5.5 Lacs - 6 Lacs
6,00,001	6 Lacs - 6.5 Lacs
6,50,001	6.5 Lacs - 7 Lacs
7,00,001	7 Lacs - 7.5 Lacs
7,50,001	7.5 Lacs - 8 Lacs
8,00,001	8 Lacs - 8.5 Lacs
8,50,001	8.5 Lacs - 9 Lacs
9,00,001	9 Lacs and above

These are the class intervals which are to found in the univariate , segmented univariate, bivariate analysis.

SEGMENTED UNIVARIATE

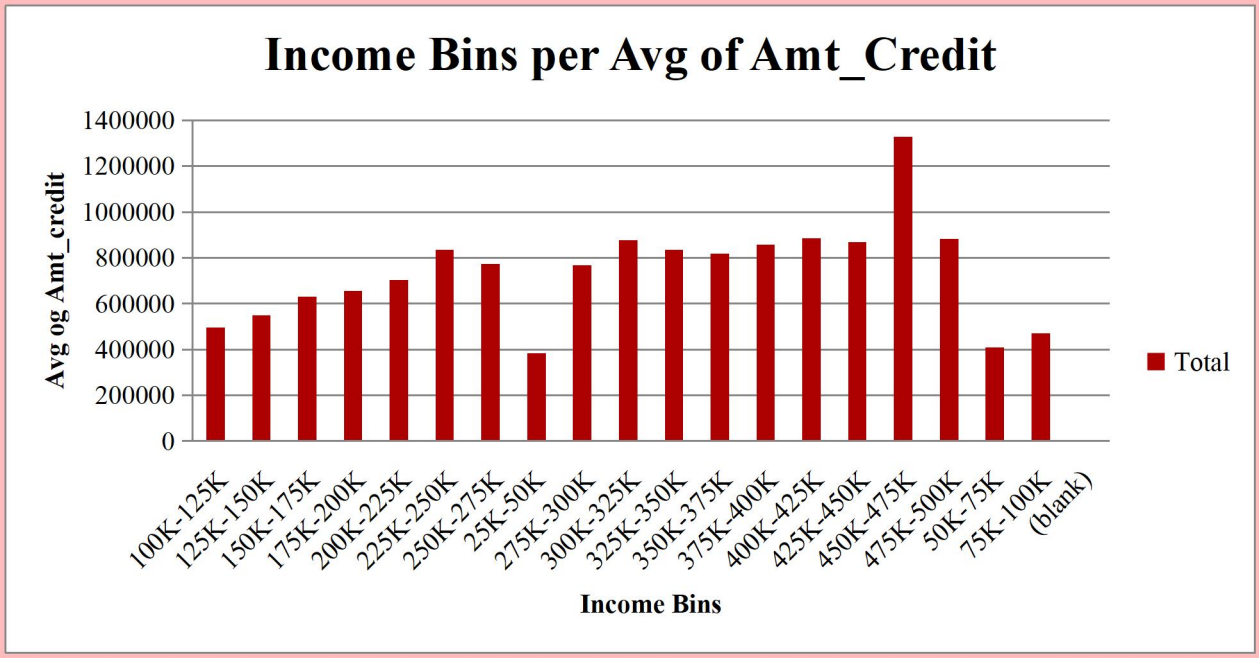


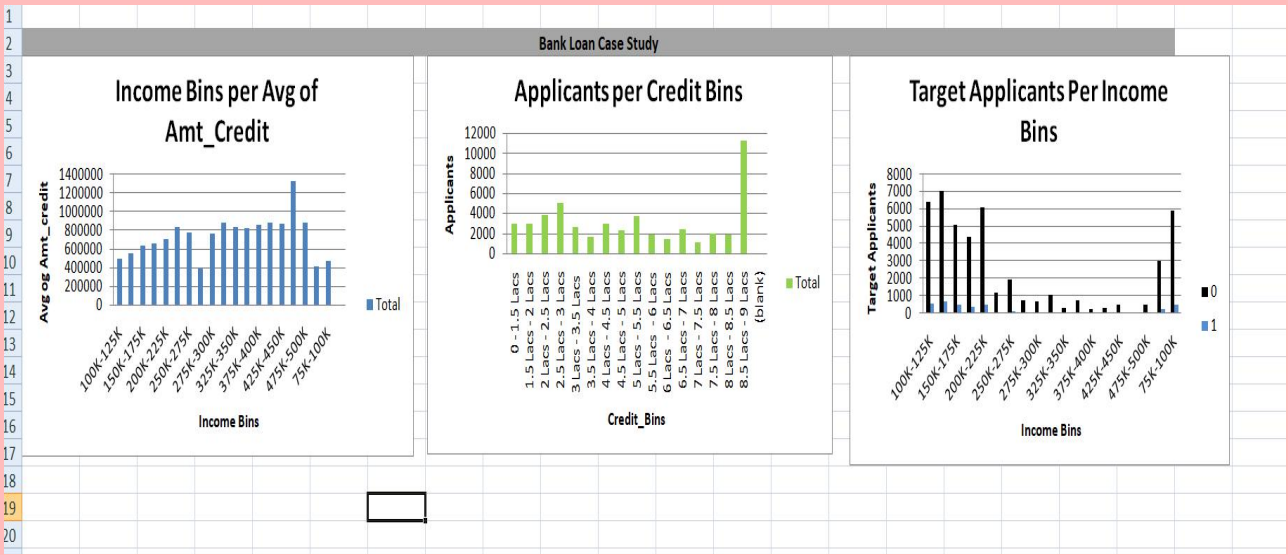
UNIVARIATE ANALYSIS



From the above chart , I analyzed that the credit bins is the score of applicants pay the loan or not.

BIVARIATE ANALYSIS





Findings 5

i) So I have taken up of **target 0** as one correlation. Then I included these columns alone from the final dataset and performed correlation between them.

TARGET, CNT_CHILDREN, AMT_INCOME_TOTAL, AMT_CREDIT, REGION_POPULATION_RELATIVE, DAYS_BIRTH(yrs), DAYS_EMPLOYED(YRS), DAYS_ID_PUBLISH(YRS), REGION_RATING_CLIENT

	CORRELATION FOR APPLICANTS WITH PAYMENT MADE ON TIME							
CNT_CHILDREN	1	0.047	0.011	-0.026	-0.322	-0.250	0.044	0.011
AMT_INCOME_TOTAL	0.047	1	0.406	0.175	-0.073	-0.183	-0.033	-0.231
AMT_CREDIT	0.011	0.406	1	0.070	0.052	-0.083	0.019	-0.103
REGION_POPULATION_RELATIVE	-0.026	0.175	-0.026	1	0.033	-7E-03	-0.001	-0.534
DAYS_BIRTH(YEARS)	-0.322	-0.073	0.052	0.033	1	0.633	0.262	0.003
DAYS_EMPLOYED(YEARS)	-0.250	-0.183	-0.083	7E-05	0.633	1	0.259	0.043
DAYS_ID_PUBLISH(YEARS)	0.044	-0.033	0.019	-0.001	0.262	0.259	1	0.015
REGION_RATING_CLIENT	0.011	-0.231	-0.103	-0.534	0.003	0.043	0.015	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH(YEARS)	DAYS_EMPLOYED(YEARS)	DAYS_ID_PUBLISH(YEARS)	REGION_RATING_CLIENT

So from the above selected columns I have correlated and inferred to create a heat map for this **Target 0.** In the above heat map itself says the answer that is the correlation of each and every different scenarios.

ii) For the **Target 1** correlation, the same as above the calculations as including the columns of particular needed headed rows and performing calculation of correlation of different scenarios.

TARGET, CNT_CHILDREN, AMT_INCOME_TOTAL, AMT_CREDIT, REGION_POPULATION_RELATIVE, DAYS_BIRTH(yrs), DAYS_EMPLOYED(YRS), DAYS_ID_PUBLISH(YRS), REGION_RATING_CLIENT

With this columns I have inferred a **correlation of target 1**in the kind of heat maps.

CORRELATION FOR APPLICANTS WITH PAYMENT DIFFICULTIES								
CNT_CHILDREN	1	-0.068	0.053	-0.009	-0.235	-0.162	0.100	-0.025
AMT_INCOME_TOTAL	-0.068	1	0.379	0.144	0.039	-0.108	-0.019	-0.144
AMT_CREDIT	0.053	0.379	1	0.061	0.165	-0.043	0.095	-0.015
REGION_POPULATION_RELATIVE	-0.009	0.144	0.061	1	-0.052	-0.114	0.022	-0.498
DAYS_BIRTH(Years)	-0.235	0.039	0.165	-0.052	1	0.545	0.288	0.100
DAYS_EMPLOYED (Years)	-0.162	-0.108	-0.043	-0.114	0.545	1	0.224	0.090
DAYS_ID_PUBLISH(Years)	0.100	-0.019	0.095	0.022	0.288	0.224	1	0.019
REGION_RATING_CLIENT	-0.025	-0.144	-0.015	-0.498	0.100	0.090	0.019	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH(Years)	DAYS_EMPLOYED (Years)	DAYS_ID_PUBLISH(Years)	REGION_RATING_CLIENT

From the above heat map , I have analyzed that the correlation between the different scenarios of **correlation for target 1** and above map itself calculated for this tasks for correlation of each and every other loan scenarios.

Conclusion

- The proportion/percentage of the defaulters(target = 1) is around 8% and that of non-defaulters(target = 0) is around 92% The Bank generally lends more loan to

Female clients as compared to Males clients as the count of Female clients in the defaulter's list is less than that of Males. Still Bank can look for more Male clients if their credit amount is satisfied.

- Also the clients who belong to Working class tend to pay their loans on time followed by the clients who fall under Commercial Associate

- Clients having Education status like Secondary/ Higher Secondary or more tend to pay loan on time so bank can prefer lending loans to clients having such Education Status Clients who fall in the Age Group 31-40 have the highest count for paying off their loans on time followed by the clients who fall in the Age Groups 41-60

- Clients having LOW credit amount range tend to pay off their loans on time than compared to HIGH and MEDIUM credit range

- Clients living with their Parents tend to pay off their loans quickly as compared to other housing type. So Bank can lend loan to clients having housing type → Living with Parents

- Clients taking loan for purchasing New Home i.e. clients taking Home Loans or purchasing New Car i.e. Car Loans and clients who have a income type as State Servant tend to pay their loans on time and hence Bank should prefer clients having such background

- The Bank should be more cautious when lending money to clients with Repairs purpose because they have high count of Defaulters along with High count of Defaulters

Analyzing the Impact of Car Features on Price and Profitability

Final Project-3

DESCRIPTION

The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars.

In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline-powered cars remain dominant in the market, with varying fuel types and grades available to consumers.

For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?

This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances

consumer demand with profitability, and identify which product features to focus on in future product development efforts. This could help the manufacturer improve its competitiveness in the market and increase its profitability over time.

The Problem

TASKS: ANALYSIS

1. How does the popularity of a car model vary across different market categories?

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

3. Which car features are most important in determining a car's price?

4. How does the average price of a car vary across different manufacturers?

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

BUILDING THE DASHBOARD:

1. How does the distribution of car prices vary by brand and body style?

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

3. How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

4. How does the fuel efficiency of cars vary across different body styles and model years?

5. How does the car's horsepower, MPG, and price vary across different Brands?

Design

Perform statistical analysis to understand the distribution, variability, and basic characteristics of the collected data. Visualize data patterns, correlations, and anomalies using graphs and plots. Identify initial insights that can guide subsequent analysis. Evaluate model performance using metrics like accuracy, precision, recall, and F1-score. Showcase how predictive models can optimize maintenance schedules and reduce downtime.

Demonstrate how real-time alerts and warnings can be generated for drivers and relevant authorities. Evaluate the potential reduction in accidents and overall enhancement in road safety. Analyze driving patterns and vehicle performance data to uncover factors influencing fuel efficiency. Develop a model to predict fuel efficiency based on driving behavior, road conditions, and vehicle characteristics.

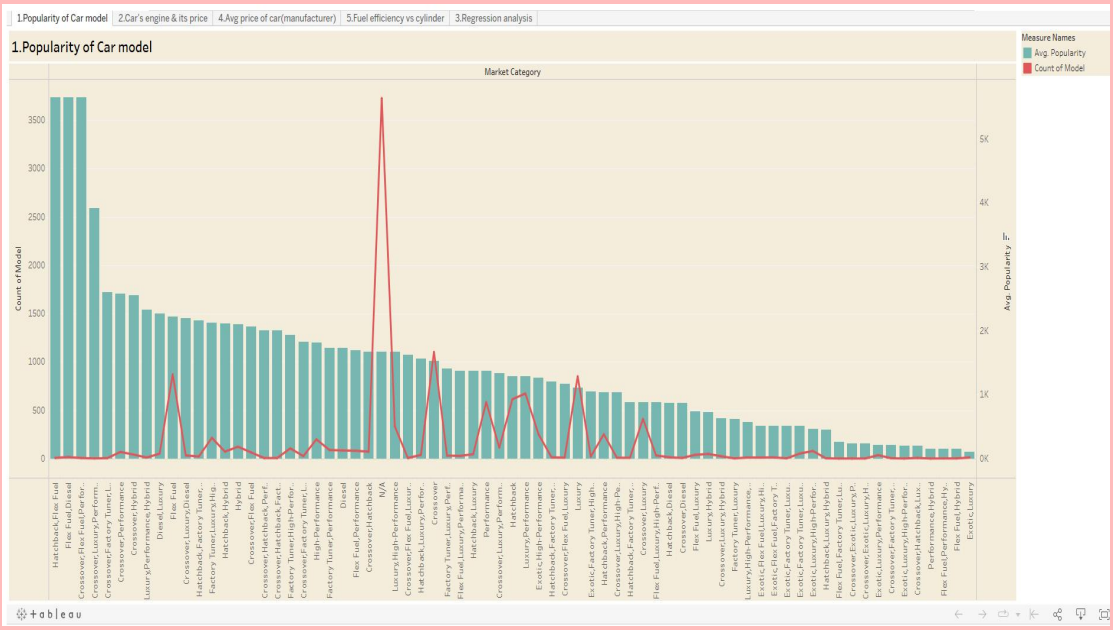
The project aims to provide a comprehensive understanding of the impact of car data analysis on the automotive industry, offering insights, solutions, and guidance for industry stakeholders to harness the power of data for positive transformations.

Findings 1

The solution for the task 1 of A is that of created a pivot table with the help of showing the number of car models in each market category with respect to the popularity scores which is of creating a make or manufacturers with the popularity scores of maintaining the market category of impact of car trends.

Popularity of a car model vary across different market categories		
	Values	
Market Category	Average of Popularity	Count of Model
Crossover	1545	1110
Crossover,Diesel	873	7
Crossover,Exotic,Luxury,High-Performance	238	1
Crossover,Exotic,Luxury,Performance	238	1
Crossover,Factory Tuner,Luxury,High-Performance	1823	26
Crossover,Factory Tuner,Luxury,Performance	2607	5
Crossover,Factory Tuner,Performance	210	4
Crossover,Flex Fuel	2074	64
Crossover,Flex Fuel,Luxury	1173	10
Crossover,Flex Fuel,Luxury,Performance	1624	6
Crossover,Flex Fuel,Performance	5657	6
Crossover,Hatchback	1676	72
Crossover,Hatchback,Factory Tuner,Performance	2009	6
Crossover,Hatchback,Luxury	204	7
Crossover,Hatchback,Performance	2009	6
Crossover,Hybrid	2563	42
Crossover,Luxury	885	410
Crossover,Luxury,Diesel	2149	34
Crossover,Luxury,High-Performance	1037	9
Crossover,Luxury,Hybrid	631	24
Crossover,Luxury,Performance	1345	113
Crossover,Luxury,Performance,Hybrid	3916	2
Crossover,Performance	2586	69
Diesel	1731	84
Diesel,Luxury	2275	51
Exotic,Factory Tuner,High-Performance	1046	21
Exotic,Factory Tuner,Luxury,High-Performance	518	52

Exotic,Factory Tuner,Luxury,Performance	520	3
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	520	13
Exotic,Flex Fuel,Luxury,High-Performance	520	11
Exotic,High-Performance	1271	261
Exotic,Luxury	113	12
Exotic,Luxury,High-Performance	467	79
Exotic,Luxury,High-Performance,Hybrid	204	1
Exotic,Luxury,Performance	217	36
Exotic,Performance	1391	10
Factory Tuner,High-Performance	1941	106
Factory Tuner,Luxury	617	2
Factory Tuner,Luxury,High-Performance	2133	215
Factory Tuner,Luxury,Performance	1413	31
Factory Tuner,Performance	1696	92
Flex Fuel	2217	872
Flex Fuel,Diesel	5657	16
Flex Fuel,Factory Tuner,Luxury,High-Performance	258	1
Flex Fuel,Hybrid	155	2
Flex Fuel,Luxury	747	39
Flex Fuel,Luxury,High-Performance	879	33
Flex Fuel,Luxury,Performance	1380	28
Flex Fuel,Performance	1680	87
Flex Fuel,Performance,Hybrid	155	2
Hatchback	1319	641
Hatchback,Diesel	873	14
Hatchback,Factory Tuner,High-Performance	1205	13
Hatchback,Factory Tuner,Luxury,Performance	887	9
Hatchback,Factory Tuner,Performance	2159	22
Hatchback,Flex Fuel	5657	7
Hatchback,Hybrid	2121	72
Hatchback,Luxury	1380	46
Hatchback,Luxury,Hybrid	454	3
Hatchback,Luxury,Performance	1566	38
Hatchback,Performance	1040	252
High-Performance	1821	199
Hybrid	2106	123
Luxury	1103	855
Luxury,High-Performance	1668	334
Luxury,High-Performance,Hybrid	569	12
Luxury,Hybrid	674	52
Luxury,Performance	1293	673
Luxury,Performance,Hybrid	2333	11
Performance	1349	601
Performance,Hybrid	155	1
Grand Total	**1499**	**8172**



From the above combo chart tasks, I have created with a bar chart and combination of line chart, it varies as colors for both line and bar chart. Here I have insisted that the popularity of car is based on the average of popularity and count of each model.

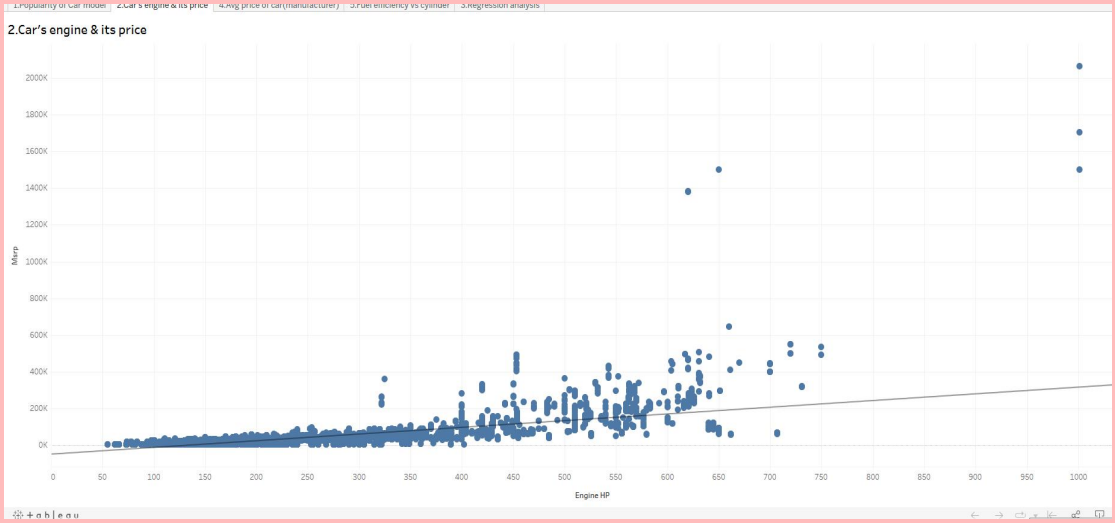
Findings 2

the relationship between car's engine and its price as MSRP , found out by separating from the final dataset of cleaned data which help me out to categorize the particular tasks and have been visualized with the trend line chart.

B1 MSRP

	A	B	C	D
1	Engine HP	MSRP		
2	335	$ 46,135		
3	300	$ 40,650		
4	300	$ 36,350		
5	230	$ 29,450		
6	230	$ 34,500		
7	230	$ 31,200		
8	300	$ 44,100		
9	300	$ 39,300		
10	230	$ 36,900		
11	230	$ 37,200		
12	300	$ 39,600		
13	230	$ 31,500		
14	300	$ 44,400		
15	230	$ 37,200		
16	230	$ 31,500		
17	320	$ 48,250		
18	320	$ 43,550		
19	172	$ 2,000		
20	172	$ 2,000		
21	172	$ 2,000		
22	172	$ 2,000		
23	172	$ 2,000		
24	172	$ 2,000		

The above picture depicts that the separated data of car's **Engine HP** and **MSRP** , to find the relationship between the car's engine power and its price.



From the above picture depicts that the scatter plot and combination of shown trend lines for finding the relationship between Engine HP and MSRP with respect to it.

Then I have taken for finding the visualization of scatter plot and with a trend lines is then taken Engine HP and MSRP to solve the relationship between them.

And also Engine HP will be taken as x-axis and MSRP taken as y-axis and plotting it accordingly and gone through the relation between , in the analysis it will be shown as trend lines - show trend lines and then edit the axis according to it.

Findings 3

Here I have learned how to solve the regression analysis using the excel data analysis functions, it gives the anova table , summary output with the residual and regression which inferred the regression data of car features on determining a car's price.

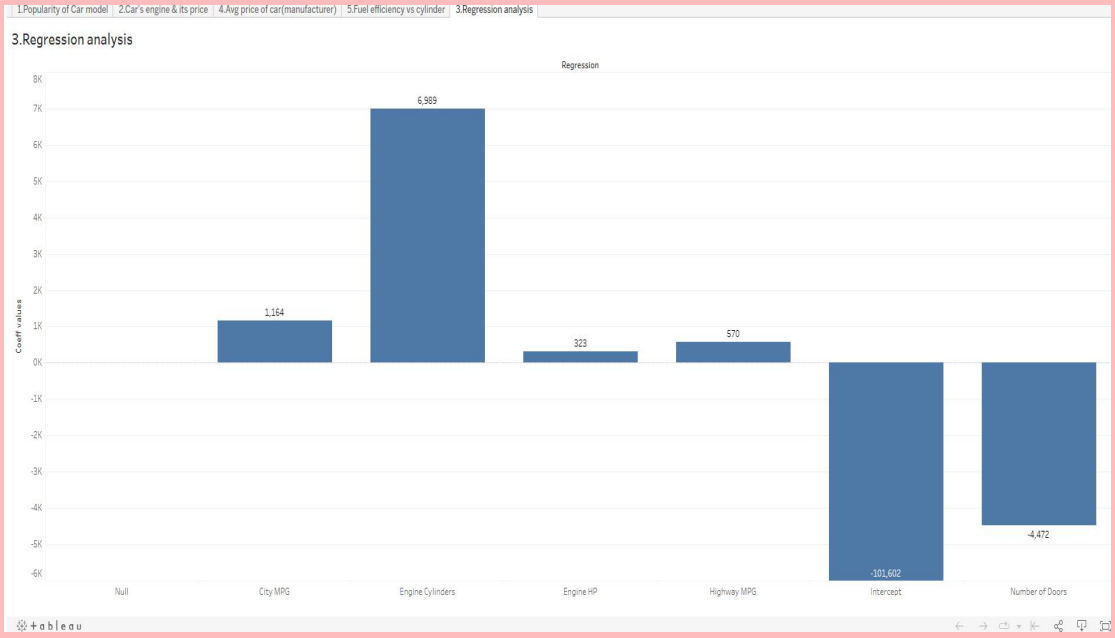
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.680708139
R Square	0.46336357
Adjusted R Square	0.463136297
Standard Error	44170.77827
Observations	11812

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	1.98891E+13	3.97782E+12	2038.799457	0
Residual	11806	2.30342E+13	1951057653		
Total	11811	4.29233E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-101601.736	3684.351697	-27.57655738	2.765E-162	-108823.673	-94379.79896	-108823.673	-94379.79896
Engine HP	322.7465574	6.01767382	53.63310924	0	310.9509241	334.5421906	310.9509241	334.5421906
Engine Cylinders	6989.177662	439.6449924	15.89732121	2.53591E-56	6127.400961	7850.954363	6127.400961	7850.954363
Number of Doors	-4472.158125	465.7180593	-9.602715711	9.35015E-22	-5385.042338	-3559.273912	-5385.042338	-3559.273912
Highway MPG	570.1808088	105.7839778	5.390048859	7.17937E-08	362.826764	777.5348535	362.826764	777.5348535
City MPG	1163.755457	121.9978136	9.539150113	1.72109E-21	924.61962	1402.891294	924.61962	1402.891294

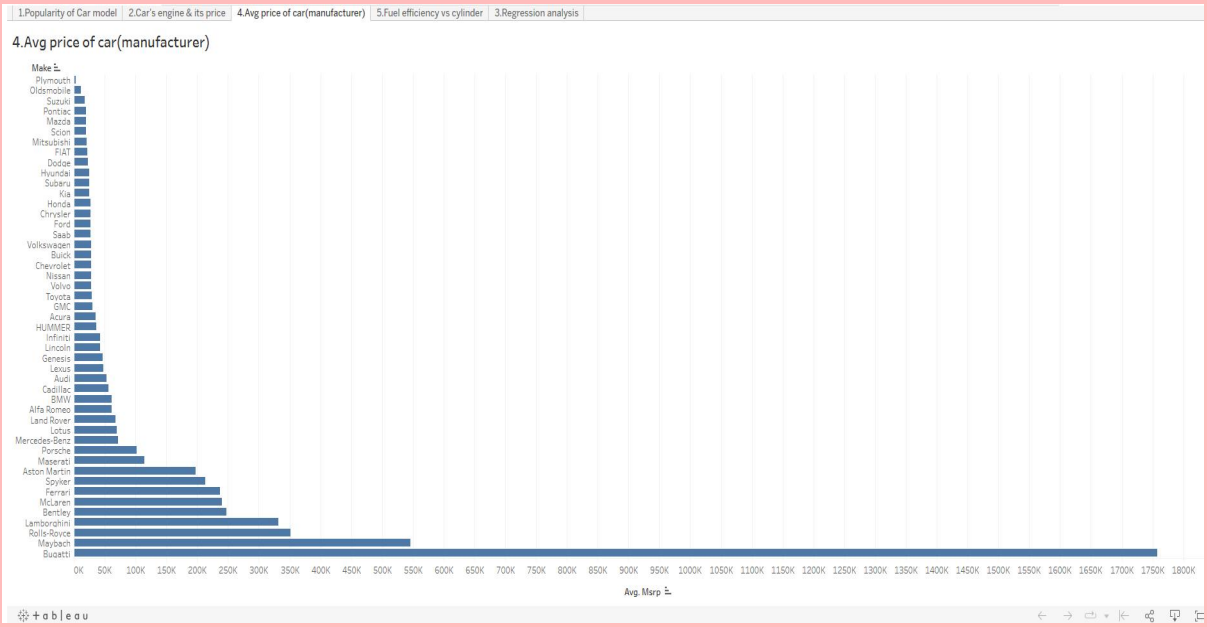


Findings 4

So I have insisted on creating a pivot table with the help of manufacturers and MSRP for price of each make. And then created the pivot table with the row labels of make that is manufacturer and MSRP as Average of car price, that are included as pivot table and then insisted as the changing of count to average of car price with the help of value field settings in the pivot table.

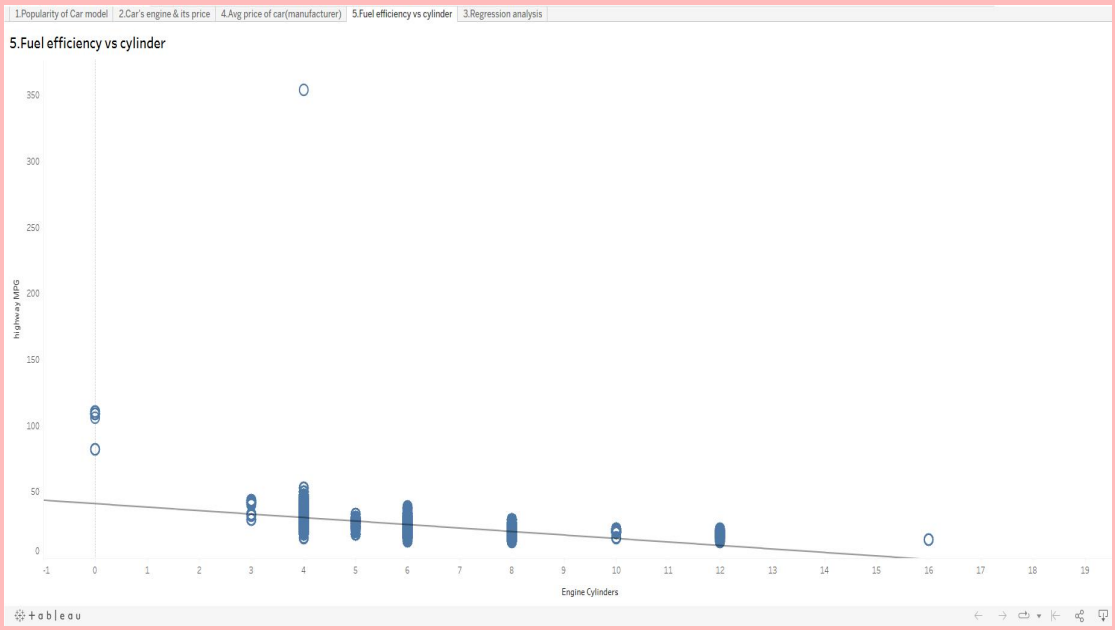
Manufacturer	Average of MSRP
Acura	₹ 34,887.59
Alfa Romeo	₹ 61,600.00
Aston Martin	₹ 1,97,910.38
Audi	₹ 53,452.11
Bentley	₹ 2,47,169.32
BMW	₹ 61,546.76
Bugatti	₹ 17,57,223.67
Buick	₹ 28,206.61
Cadillac	₹ 56,231.32
Chevrolet	₹ 28,273.36
Chrysler	₹ 26,722.96
Dodge	₹ 22,390.06
Ferrari	₹ 2,37,383.82
FIAT	₹ 22,206.02
Ford	₹ 27,393.42
Genesis	₹ 46,616.67
GMC	₹ 30,493.30
Honda	₹ 26,629.82
HUMMER	₹ 36,464.41
Hyundai	₹ 24,597.04
Infiniti	₹ 42,394.21
Kia	₹ 25,112.39
Lamborghini	₹ 3,31,567.31
Land Rover	₹ 67,823.22
Lexus	₹ 47,549.07
Lincoln	₹ 42,494.37
Lotus	₹ 69,188.28
Maserati	₹ 1,14,207.71
Maybach	₹ 5,46,221.88
Mazda	₹ 19,719.06
McLaren	₹ 2,39,805.00

Mercedes-Benz	₹	71,537.81
Mitsubishi	₹	21,215.47
Nissan	₹	28,513.37
Oldsmobile	₹	11,542.54
Plymouth	₹	3,122.90
Pontiac	₹	19,321.55
Porsche	₹	1,01,622.40
Rolls-Royce	₹	3,51,130.65
Saab	₹	27,413.50
Scion	₹	19,932.50
Spyker	₹	2,13,323.33
Subaru	₹	24,827.50
Suzuki	₹	17,900.96
Toyota	₹	28,946.15
Volkswagen	₹	28,076.20
Volvo	₹	28,541.16
Grand Total	**₹**	**40,559.94**



From the above pasted picture represents that the horizontal a stacked bar chart for insights of average price of car with respect to the manufacturer. Taking Average of price that is in the header row of MSRP is been taken as x-axis and Manufacturer will be taken as y-axis .

Findings 5



The above picture represents the scatter plot of finding the fuel efficiency versus number of cylinder and with a trend lines for the reference of insisting the correlation of fuel efficiency with each of the cylinders.

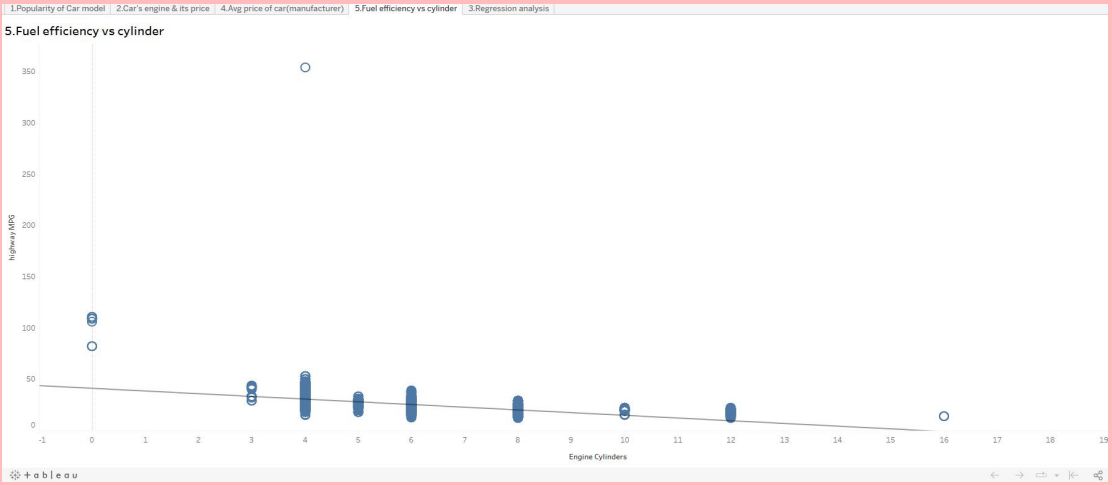
I found the solution that the correlation of coefficient between number of cylinders and the highway MPG played a role in this tasks. It quantifises the strength of each of the cylinders numbered and a direction of relationship between them.

Clipboard | Font | Alignment

E6

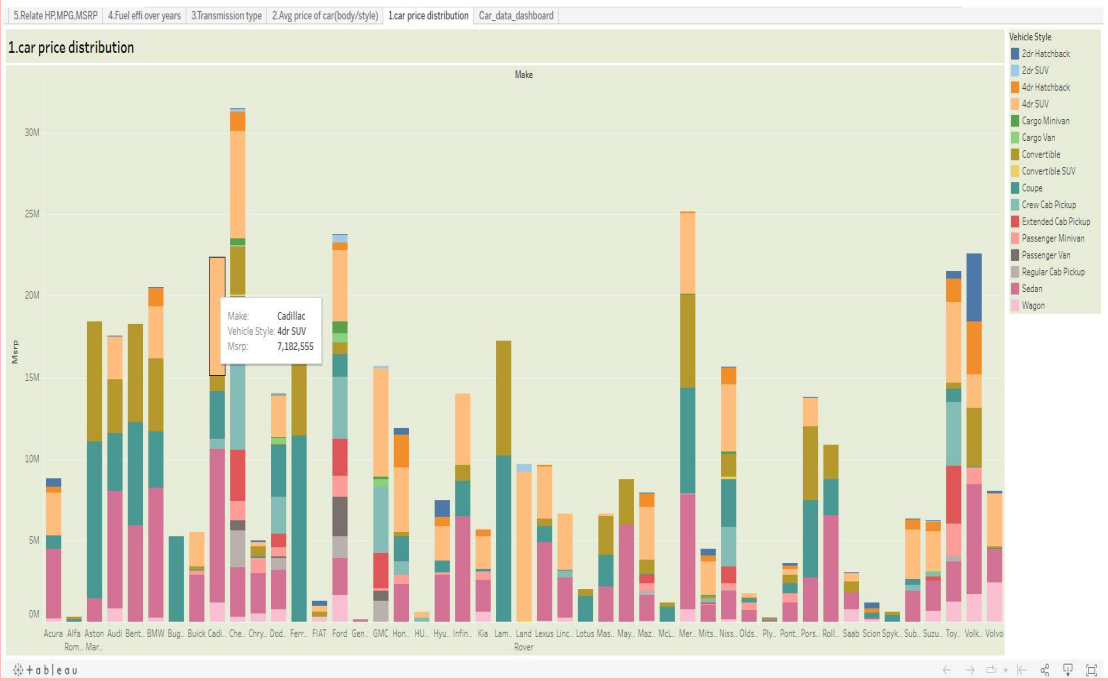
	A	B	C	D	E	F
1	Engine Cylinders	highway MPG				
2	6	26			Correlation, r=	-0.62031
3	6	28				
4	6	28				
5	6	28				
6	6	28				
7	6	28				
8	6	26				
9	6	28				
10	6	28				
11	6	27				
12	6	28				
13	6	28				
14	6	28				
15	6	28				
16	6	28				
17	6	25				
18	6	28				
19	6	24				
20	6	24				
21	6	20				
22	6	24				
23	6	21				
24	6	24				

From the above picture is my analysis of finding the fuel efficiency with the number of cylinders and found out that coefficient values as **r=-0.620312551.**



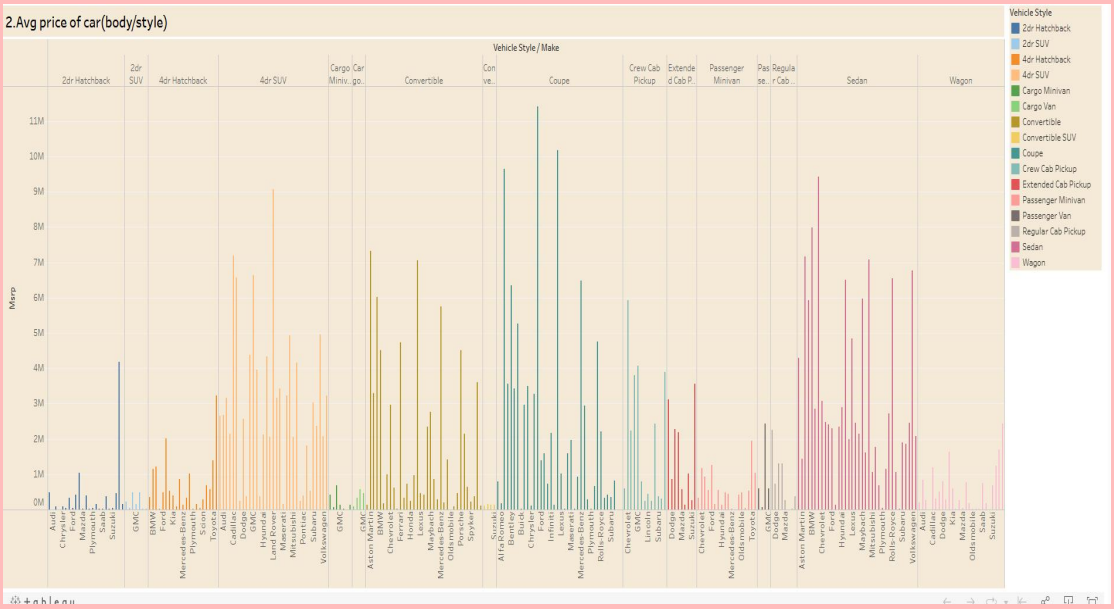
Building dashboards

Findings 1



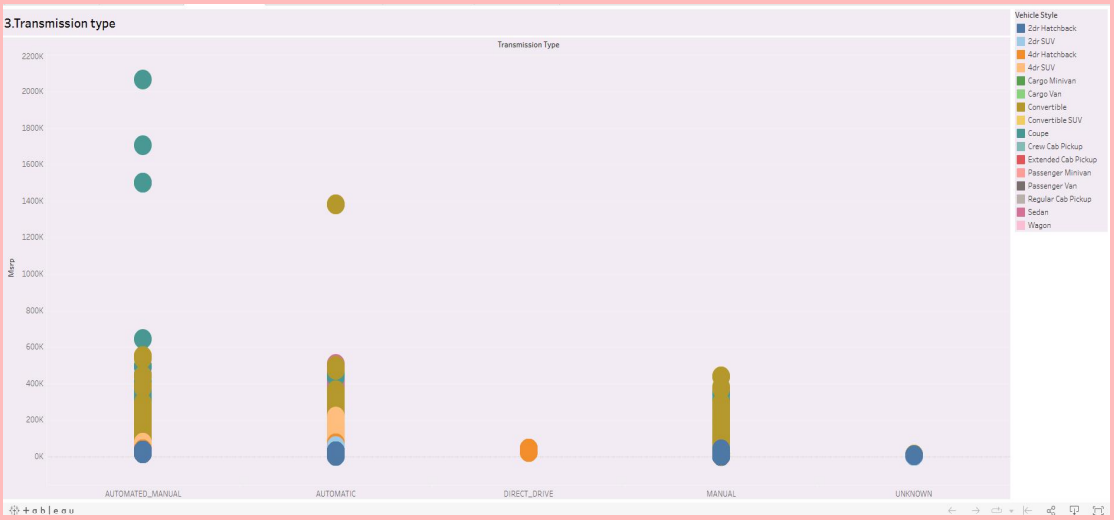
In this dashboard analysis, I have inferred that the distribution of car price with respect to the vehicle style and manufacturer for each of the car. Now here what is the tasks is that to make a stacked column chart for making the visualizations of finding the distribution of car price based on branch and vehicle style.

Findings 2



For this dashboard analysis , I have insisted on finding the average price car based on both body and style position. So I have created a column chart to visualize the values here I have done both the body and style together with each of the brand with its price that is MSRP.

Findings 3



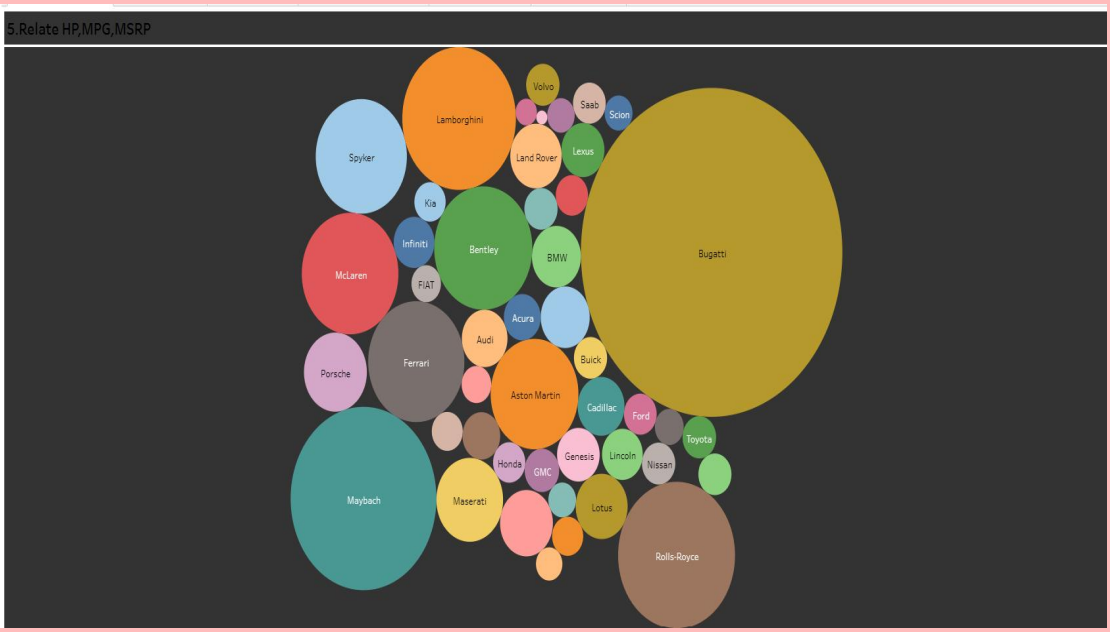
For this dashboard analysis, I have inferred that the transmission type of vehicle styles as a feature that affects the MSRP of car to shown below the pivot tables. Here I have taken vehicle style as x-axis and msrp as y-axis to get the designed scatter plot to visualize the relationship between the msrp and vehicle style, and also legends shown as for the reference of each of the color given to vehicle styles.

Findings 4

For this dashboard analysis, it is make a Line chart to show the trend of fuel efficiency (MPG) over time for each body style. It represents that the fuel efficiency of finding the average number of highway MPG with respect to the year.

And calculated the average MPG for each combination of body style and model year with the visualization of line chart with the trend lines.

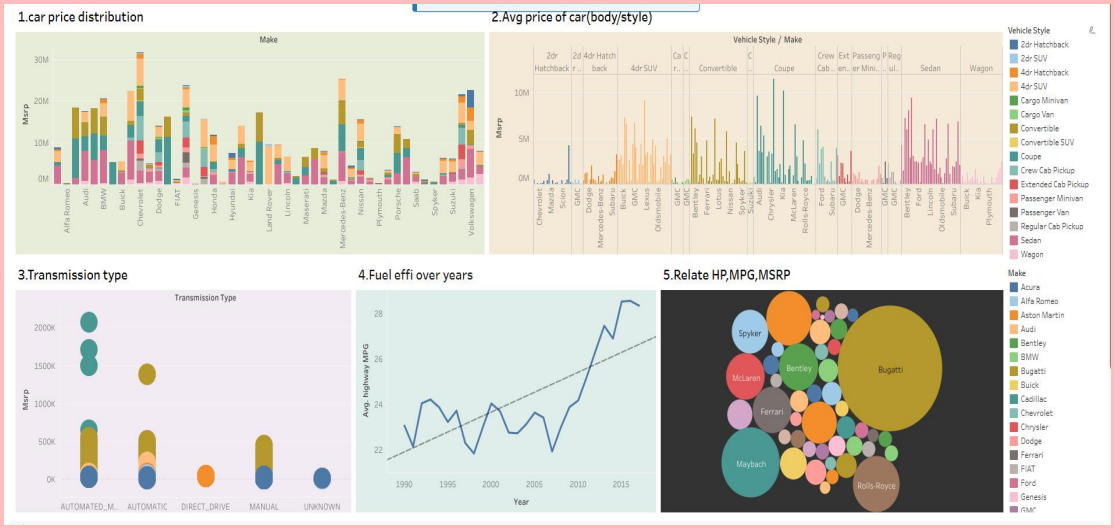
Findings 5



For this dashboard analysis, I have insisted on making a bubble chart with the help of relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles

with the car model name. And have calculated the average horsepower, MPG, and MSRP for each car brand.

DASHBOARD(overall)



Conclusion

In this project , I have learnt how to make a analysis of each of the dashboards, I was very thrilled to making a beautiful insights as a data analyst. I have used tableau public and excel to find the tasks and dashboard analysis.

Future directions continue to be improvements in software and algorithms. Increasing the level of parallelism is desired for dealing with large scale computational and memory requirements.

ABC CALL VOLUME TREND ANALYSIS

Final Project4

DESCRIPTION

In this project, you'll be diving into the world of Customer Experience (CX) analytics, specifically focusing on the inbound calling team of a company. You'll be provided with a dataset that spans 23 days and includes various details such as the agent's name and ID, the queue time (how long a customer had to wait before connecting with an agent), the time of the call, the duration of the call, and the call status (whether it was abandoned, answered, or transferred).

A Customer Experience (CX) team plays a crucial role in a company. They analyze customer feedback and data, derive insights from it, and share these insights with the rest of the organization. This team is responsible for a wide range of tasks, including managing customer experience programs, handling internal communications, mapping customer journeys, and managing customer data, among others.

In the current era, several AI-powered tools are being used to enhance customer experience. These include Interactive Voice Response (IVR), Robotic Process Automation (RPA), Predictive Analytics, and Intelligent Routing.

One of the key roles in a CX team is that of the customer service representative, also known as a call center agent. These agents handle various types of support, including email, inbound, outbound, and social media support.

The Problem

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

Show the total volume/ number of calls coming in via charts/ graphs [Number of calls v/s Time]. You can select time in a bucket form (i.e. 1-2, 2-3, …..)

As you can see current abandon rate is approximately 30%. Propose a manpower plan required during each time bucket [between 9am to 9pm] to reduce the abandon rate to 10%. (i.e. You have to calculate minimum number of agents required in each time bucket so that at least 90 calls should be answered out of 100.)

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

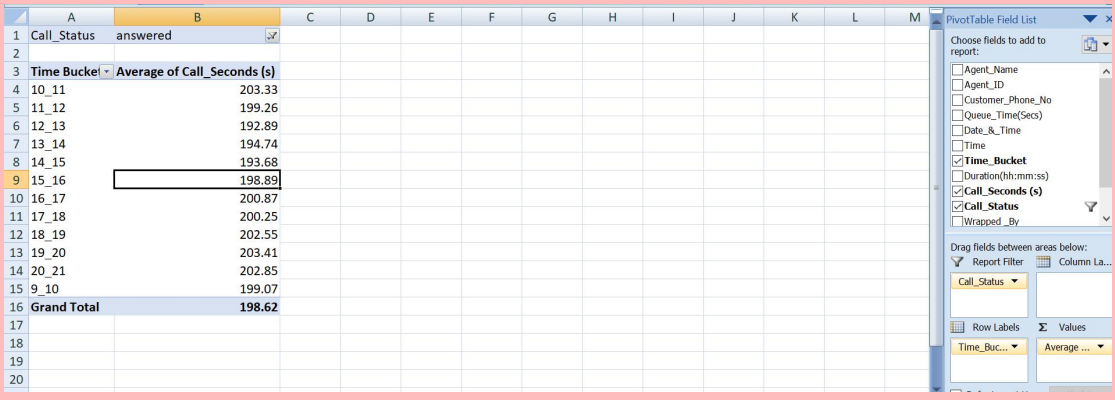
Assumption: An agent work for 6 days a week; On an average total unplanned leaves per agent is 4 days a month; An agent total working hrs is 9 Hrs out of which 1.5 Hrs goes into lunch and snacks in the office. On average an agent occupied for 60% of his total actual working Hrs (i.e 60% of 7.5 Hrs) on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

Findings 1

For this tasks , that is, average duration of calls for each time bucket have found out to be average of call duration for each of the time bucket. So I have been analyzed in the cleaned data and created a pivot table based on the understanding of the question.

Call_Status	answered

Time Bucket	Average of Call_Seconds (s)
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
9_10	199.07
Grand Total	**198.62**



From the above excel picture represents that the pivot table calculation for finding the average of duration of calls with

each of the time bucket and analyzed with pivot table calculation of finding the average of call duration.

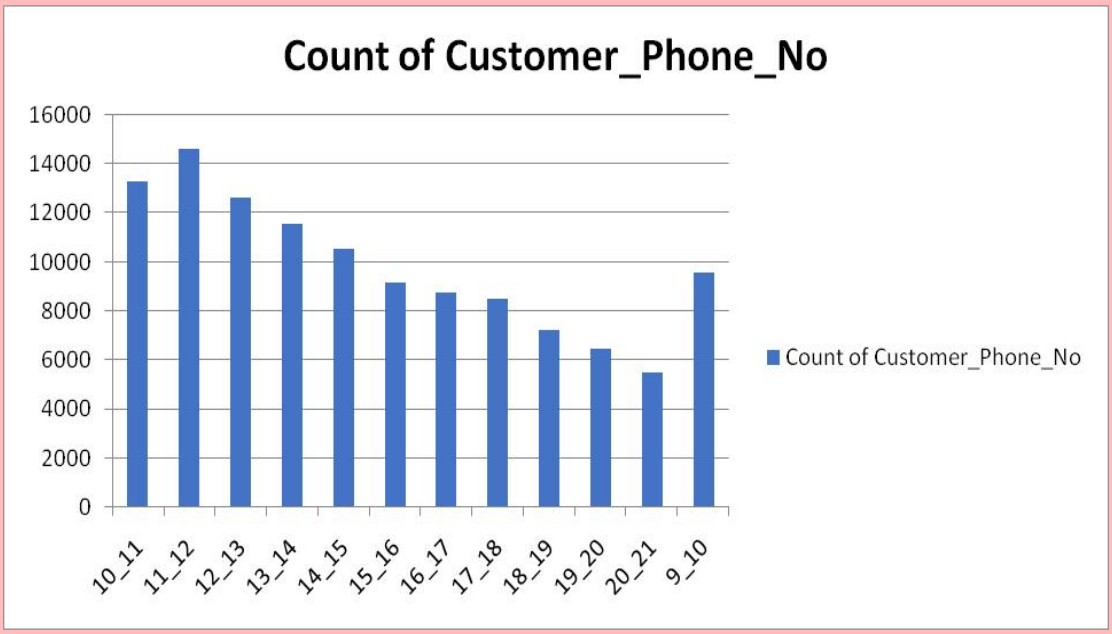
So with that , the average of call duration for each time bucket is 198.62

Findings 2

For this tasks, I created the pivot table about the number of calls received with each of the time bucket, so taken with a time bucket, number of time and customer phone number.

	Values	
Time Bucket	Count of Customer_Phone_No	Count of Time
10_11	13313	11%
11_12	14626	12%
12_13	12652	11%
13_14	11561	10%
14_15	10561	9%
15_16	9159	8%
16_17	8788	7%
17_18	8534	7%
18_19	7238	6%
19_20	6463	5%
20_21	5505	5%
9_10	9588	8%
Grand Total	**117988**	**100%**

From the above pivot table chart represents that count of time and with each of the customer phone number with the time bucket of 10_11 to 9_10.



The above bar chart depicts that the time bucket with each of the customer phone number also that grand total of 117988 with count of time would be 100%.

Findings 3

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

From this assumptions, I have been analyzed through the simple mathematical operations in the microsoft excel, then delivered a distribution of time buckets of 30 calls coming in night for every 100 calls.

107018	23-01-2022 20:58		1		1			
107019	23-01-2022 20:58		1		1			
107020	23-01-2022 20:59		1		1			
107021	**Grand Total**	**34403**	**82452**	**1133**	**117988**			
107022		1496	3585	49	5130			
107023		29%	70%	1%	100%			
107024								
107025	Working Hrs / agent	4.5						
107026	Avg call duration	198.62						
107027								
107028								
107029	For 90% hrs needed	254.73015						
107030	No of agents need is	57						
107031					9am-9pm	57	90%	
107032					9pm-9am	17	90%	
107033	Avg no of calls in night is	1539			overall agents needed is	74		
107034	To increase call rate to 90% ni8 is	76						
107035	Num of agents needed in night	17						

So above picture is the calculation for finding the minimum number of agents required in each time bucket . Then I have calculated with the pivot table calculations and picked a

abandon (total value/no of values)

Answered(total value/no of values)

Transfer(total value/no of values)

With the count of duration(hh:mm:ss).

Working Hrs / agent	4.5
Avg call duration	198.62

This is calculated for finding the working hours per each of the agent is 4.5 and Average of call duration is 198.62(previous calculation). This is been calculated as …

Working hours per agent=(60/100)*7.5

=4.5

Average of call duration=198.62

Time Bucket	Average of Call_Seconds (s)

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
9_10	199.07
Grand Total	**198.62**

For 90% hrs needed	254.73015
No of agents need is	57

From this we can calculate the number of agents needed for 90Hrs is 57.This can be calculated as …..

For 90hrs needed =5130*198.62*0.9/3600

=254.73015

No of agents needed =254.73015/4.5

=57

Findings 4

For this tasks , I have used previously calculated pivot table and used with respect to the analysis of tasks, This creates a poor customer experience. Assume that for every 100 calls that customers make between 9 am and 9 pm, they also make 30 calls at night between 9 pm and 9 am. So I have done calculations based on the night shift between 9pm-9am.

Avg no of calls in night is	1539
To increase call rate to 90% ni8 is	76
Num of agents needed in night	17

In this excel sheet calculation, I have analyzed with a average number of calls in night is 1539 and to increase the call rate to 90% in night is 76 and number of agents have calculated for the night shift.

The calculations be….

Average number of calls in night =0.3*5130

=1539

To increase the call rate to 90% in night=198.62*1539*0.9/3600

=76

Number of agents needed in night=76/4.5

=17

9am-9pm	57	90%
9pm-9am	17	90%
overall agents needed is	74	

Then the overall agents is been listed above and the calculations would be….

9am-9pm = 57 (90%)

9pm-9am = 17 (90%)

And the overall agents needed = 57+17

=74

Conclusion

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

Appendix

1. Data Analytics Process:-

---> Link for the shared PDF on Google Drive:

https://drive.google.com/file/d/1KMCuiG0fkwXOrVK9fDC1cKvg3u-zLhUJ/view?usp=drive_link

2.Instagram User Analytics:-

----> Link for the shared file on Google Drive :

https://drive.google.com/file/d/10-vCx1VXrsek946TH1n2bWMbeyWHhGEV/view?usp=drive_link

3.Operation Analytics and Investigating Metric Spike Analysis:-

-----> Link for the shared file on Google Drive:

https://drive.google.com/file/d/1Bs5mmoI7THWJmUfB-hHOZ5MKXZ-e9n0a/view?usp=drive_link

4.Hiring Process Analytics:-

----> Link for shared PDF on google drive:

https://drive.google.com/file/d/13hIf2XhhMIQtwI70n5BKBuNt8afX5DaD/view?usp=drive_link

5.IMDB Movie Analysis-

---> Link for the shared PDF on Google Drive:

https://drive.google.com/drive/folders/1L1Lmf1OXD2d6N2DZx08B4saOG2RrdUwi?usp=drive_link

6.Bank Loan Case Study:-

----> Link for the shared file on Google Drive:

https://drive.google.com/drive/folders/1BPXUmps2cb8Q4bD1HmBHuwsYSGuSYyl5?usp=drive_link

7.Analyzing the Impact of Car Features on Price and Profitability:-

----> Link for the shared file on Google Drive:

https://drive.google.com/drive/folders/1teT0DcMzWfXiVx_33PFkRoUtyh_v9gJK?usp=drive_link

8.ABC Call Volume Trend Analysis:-

-----> Link for the shared file on Google Drive:

https://drive.google.com/drive/folders/19utlO7Ypv0pzRLC-4Q3vXB-lw4zpOQU0?usp=drive_link

Link to GitHub Portfolio:-

https://github.com/Sowmiyar1512

Link to LinkedIn Profile:-

https://www.linkedin.com/in/sowmiya-r-58767a215/