# Data Analysis Portfolio

Prepared By:
Bansiben Prajapati

# Professional Background

Currently in my 3rd Year pursuing B.Tech CSE. I have secured 7.26 CGPA (till 5th sem) and have several skills including Machine Learning, Deep Learning, Data Analysis, Python.

I have done an AICTE Internship on Al: Transformative Learning with TechSaksham, a joint CSR initiative by Microsoft & SAP during the four-week internship it emphasized on developing essential skills in AI technologies through hands-on learning projects, the project aimed to detect plant diseases from leaf images, helping farmers identify issues early and take timely action using Convolutional Neural Networks (CNNs).

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

# Table Of Contents

# Project – 1
# Data Analytics Process

**Applications in Real Life Scenario**

1. **We book the tickets of movie on bookmyshow**

Plan
We decide which movie to watch before booking it. Whether a comedy, thriller, romantic genre we want to watch.
Prepare
Next we will check how much I am willing to spend.
Process
Then we need to check which movie we want to watch and how much the seat cost-normal seat or recliner seat.
Analyze
Then we will check which movie is recently released and what is it genre.
Share
Now we will decide the movie and see the available seats and decide the theatre which is nearby.
Act
We then finally book the tickets.

2. **We use Blink it to buy grocery**

Plan
We first decide things to buy from the application of blinkit.
Prepare
Next we check how much are we in need for and how much it costs.
Process
Then we check what is available such as vegetables, fruits, rice flour etc.
Analyze
Then we check what we need according to the pantry requirements and how much it costs in total.
Share
Now we add the items to the cart for final checkout.
Act
Then we finally buy it.

**3. We use google maps to go to the places whose path is unknown and which path takes minimum time to reach destination.**

Plan
We first decide a place where to go and by which means of transport we can go.
Prepare
Then we see which modes of transport are available, how much it costs and time taken to reach the destination.
Process
Then we compare the modes of transport which would be preferred according to number of passengers, time taken to reach the destination and cost of travel.
Analyze
According to the kilometers and traffic situations google map shows us the best route which takes minimum time to travel.
Share
We then finally decide the route to reach our destination according to the time taken to reach and cost.
Act
Then we finally start our journey to reach the destination.

**4. We use Ola to travel around the city.**

Plan
First we decide where we have to go and plan the destination.
Prepare
Then we see the routes from where we can go and how much time it will take to travel.
Process
Then we see the best suited vehicle according to number passengers and the route.
Analyze
Then we analyze the route what are the conditions and how much time it will take to reach the destination and what it will cost.
Share
Then we search the available vehicle and which is best suitable for the travel.
Act
Then we finally book it.

**5. What we should watch on Netflix**

Plan
We decide what we want to watch. Is it a fiction, drama, anime, sci-fi, horror etc.
Prepare
Then we check on what device we want to watch such as mobile, laptop or TV.
Process
Then we see what we want to watch from the available movies and series.
Analyze
Then according to the recent movies and series we watch which is freshly released and analyze what we want to watch according to our likings.
Share
Then we explore which is best suitable for us to watch.
Act
Then finally we start watching.

# Project – 2
# Instagram User Analytics

Project Description

The aim of this project is to analyze Instagram user engagement and take insights through which we will be able to take informed decisions about future directions of Instagram. By using MySQL workbench and SQL queries we will be able to handle tasks and answer key questions.

Approach

·First loaded the provided dataset into MySQL workbench.

·Then reviewed the table schema and the keys associated to it to understand the relationship between each table and columns.

·Further we analyzed the problem statement given and wrote SQL queries for each task and repeated the process of writing the SQL queries until we found an accurate answer.

·Verified accuracy of queries by validating results against expected patterns.

Tech-Stack Used

·Used MySQL workbench Version 8.0.41.0.

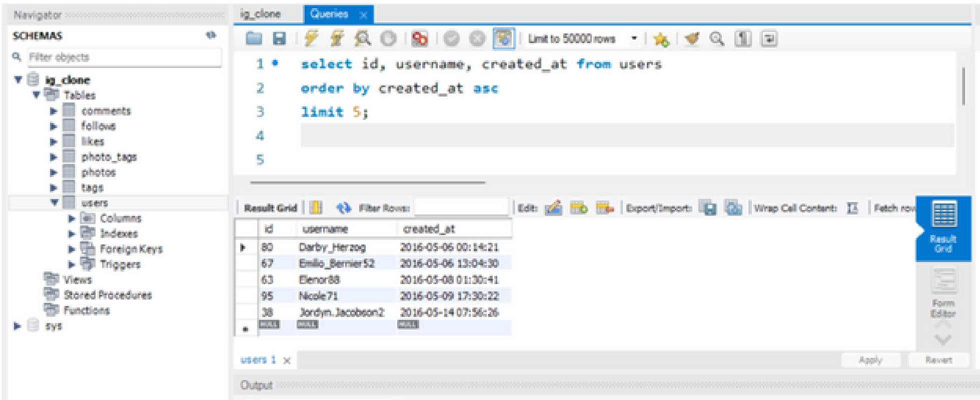
·Choose MySQL workbench as it can be used for database setup, querying, and performing commands for filtering, aggregating and joining data to have insights.

Insights

A)Marketing Analysis:

1. Loyal User Reward

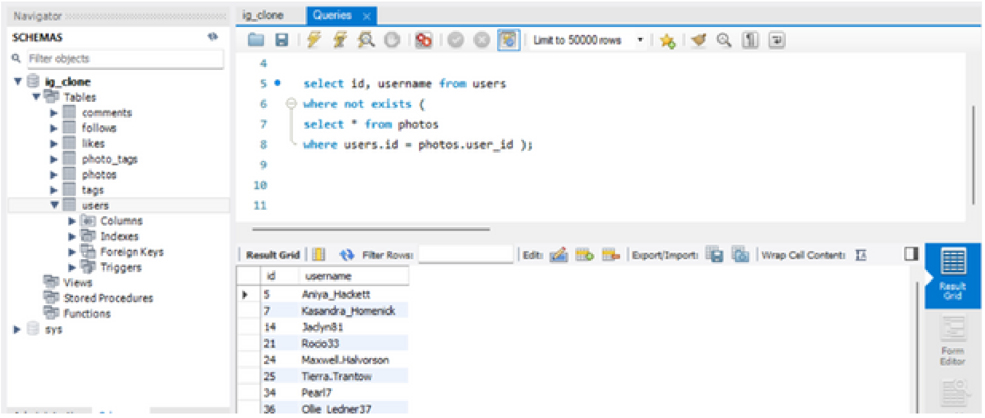
We found 5 most loyal users by accessing users table and arranging them in ascending order and by printing the 5 users who have made account first from the given dataset.

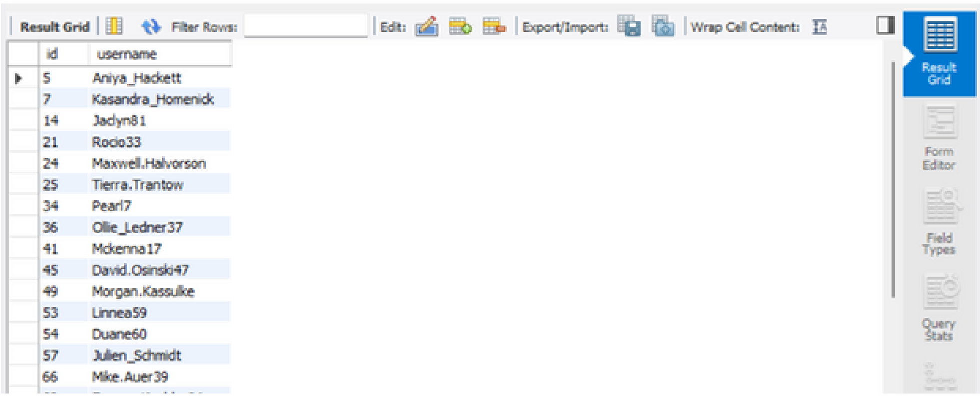


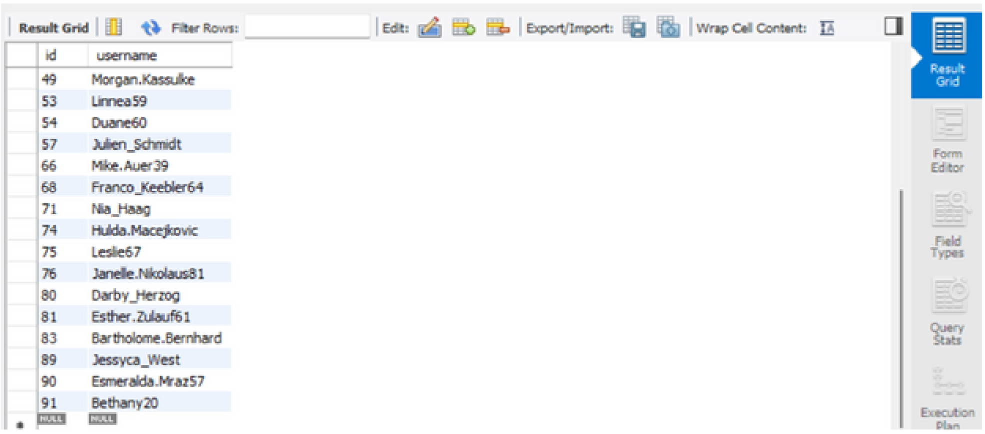
Insight- we found most loyal users by finding who has been using the Instagram for the longest time and found 5 oldest users on Instagram which are given above in output table.

2. Inactive User Engagement

We found inactive user by finding who have not posted any photo. We found it from photos table where we searched which user id was not present.



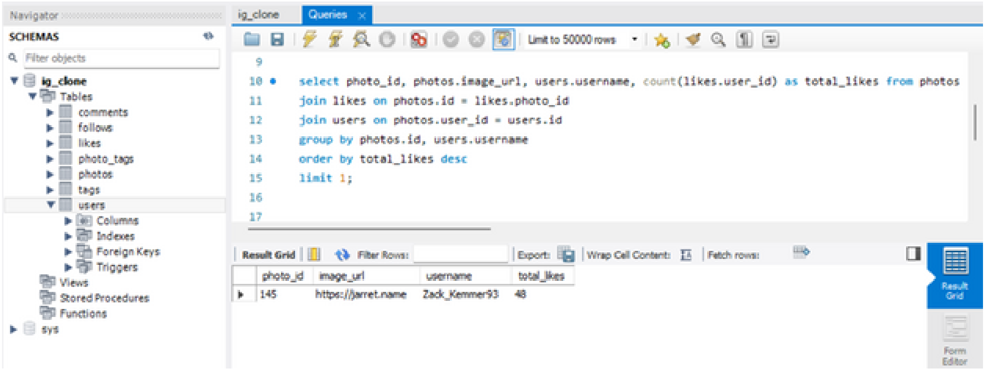



Insight - list of users who have never posted a single photo on Instagram this helps team to encourage inactive users to start engaging. The list is given above in the output table.

3. Contest Winner Declaration

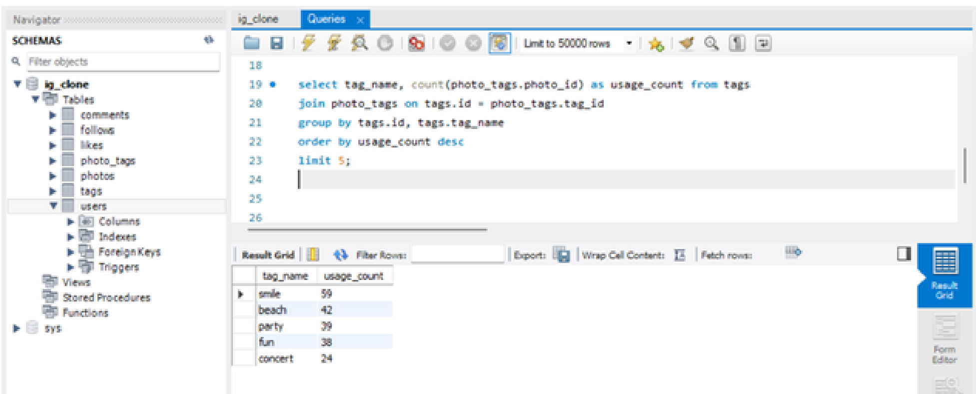
We found contest winner by finding the most liked photo from likes table and joining it with photos table that gives as number of likes per photo and then we return the details of user with most liked photos from users table.




Insight - found the person who has most likes on the post and identified the details about them. The person is Zack having username Zack_Kemmer93 and he has 48 likes.

4. Hashtag Research

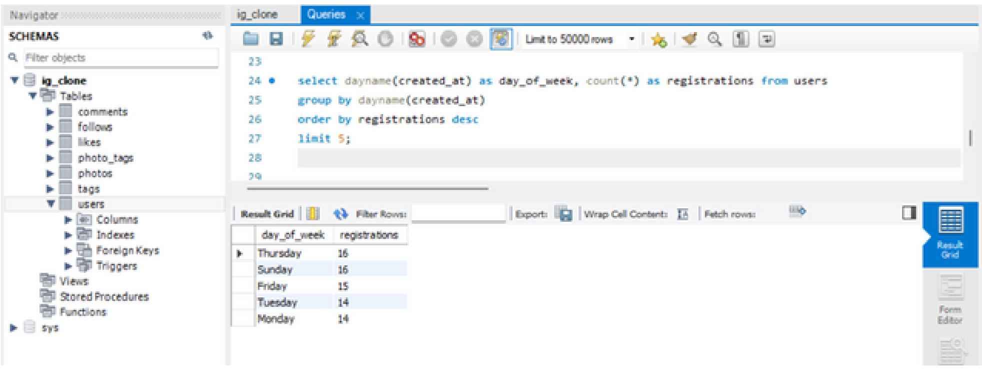
We found five most commonly used photo by accessing photo_tags table where we counted count tags by photo_id from photo_tags table and then ordering it in descending order and finding the top 5 hashtags.



Insight - identified five most used hashtags. This will help the users to maximize their reach on post by using most used hashtags. The hashtags are smile, beach, party, fun and concert.

5. Ad Campaign Launch

We found most users registration from users table where we counted the created_at which is day at which a user created account and then arrange it in descending order and got two most registered days.

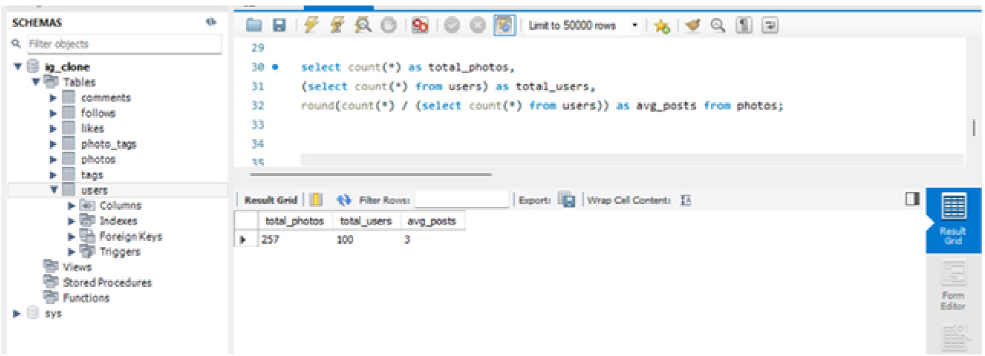



Insight - found the day on which the maximum registration was received which will help the team to schedule ad campaign. The days are Thursday and Sunday.

B) Investor Metrics:

1. User Engagement

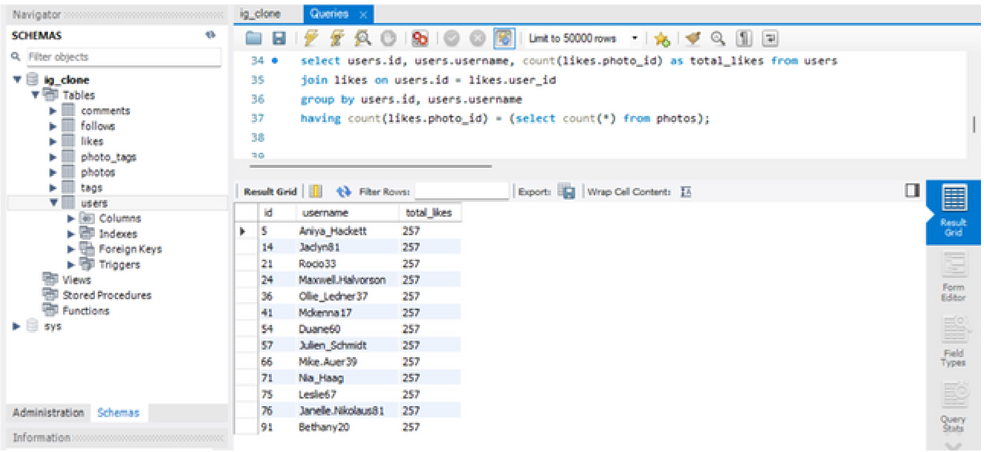
We found average number of posts per user by finding the count of total photos divided by count of total number of users.




Insight –found average number of posts on Instagram which is 3.

2. Bots & Fake Accounts

We found dummy and fake accounts by finding out the user who has liked every single photo we did it by joining user and likes table and counted number of likes per user and compared it with total number of photos and the accounts which were equal are printed.




Insight – we found user who have liked every post on the site this helps us to understand that the platform is crowded with 13 fake and dummy accounts.

Conclusion Insight

Found that most users who have not posted photos are dummy and fake users.

Result

Through this project we discovered user engagement patterns and behavior which helped us to understand the marketing and investor goals for future Instagram growth. It helped to gain deeper knowledge to analyze the situation and gain more knowledge in SQL querying and relational database.

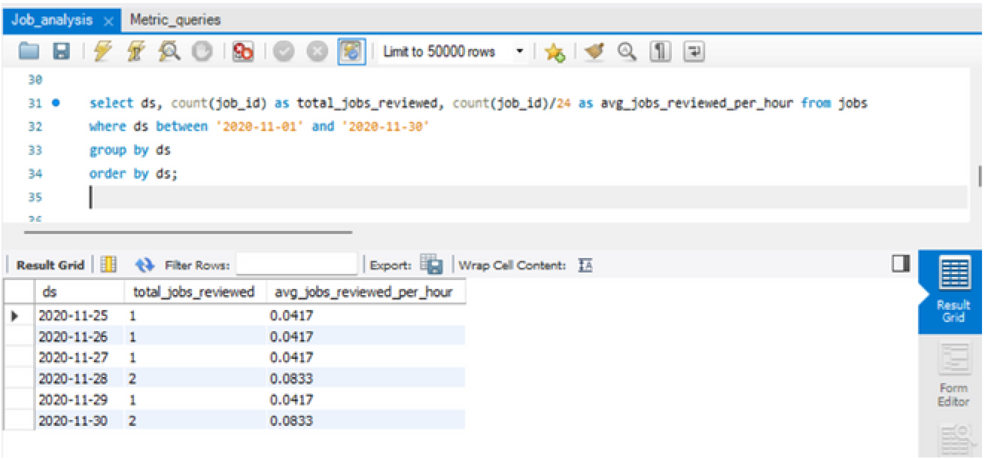
# Project – 3 Operation Analytics and Investigating Metric Spike

Project Description
The project aims to perform operational analytics which helps organizations to analyse their performance and end to end operations. The Job Data Analysis and Investigating Metric Spikes case studies will be used to perform dataset analyses using advanced SQL queries. This will help to extract insights to help decision making in marketing and operations.

Approach
·First we load the dataset provided by importing CSV files using MySQL workbench.
·Then we observe the tables to understand the relationship between each columns and table.
·We then analysed the data as per the situations given by performing optimized SQL queries. It helps to derive key business insights.
·Finally we verify the queries by comparing it to expected outcome and checking its accuracy.

Tech-Stack Used
·Used MySQL workbench Version 8.0.41.0 for database management and query execution.
·Excel sheet for data visualization.

Insights
Case Study 1: Job Data Analysis
A.Jobs reviewed over time
We have calculated total number of jobs reviewed per day in November 2020 and the average jobs reviewed per hour by dividing the count by 24. Then it is grouped by data ds and it is then sorted in the ascending order.




```sql
select ds, count(job_id) as total_jobs_reviewed, count(job_id)/24 as avg_jobs_reviewed_per_hour from jobs
where ds between '2020-11-01' and '2020-11-30'
group by ds
order by ds;
```

| ds | total_jobs_reviewed | avg_jobs_reviewed_per_hour |
|---|---|---|
| 2020-11-25 | 1 | 0.0417 |
| 2020-11-26 | 1 | 0.0417 |
| 2020-11-27 | 1 | 0.0417 |
| 2020-11-28 | 2 | 0.0833 |
| 2020-11-29 | 1 | 0.0417 |
| 2020-11-30 | 2 | 0.0833 |

Insights- The above table shows the average jobs reviewed per hour.\

B.Throughput analysis

Daily throughput is calculated by the number of events per total time spent. Then the rolling average throughput is calculated for 7 days considering the current day and previous 6 days. At the end the results are ordered by data ds To have an overview of trends overtime.

```
with daily_throughput as(
select ds, count(*) as total_events, sum(time_spent) as total_time_spent from jobs group by ds),
throughput as (select ds, total_events, total_time_spent, total_events / total_time_spent as throughput from daily_throughput)
select ds, throughput, avg(throughput)
over (order by ds rows between 6 preceding and current row) as rolling_avg_throughput
from throughput
order by ds;
```

| ds | throughput | rolling_avg_throughput |
|---|---|---|
| 2020-11-25 | 0.0222 | 0.02220000 |
| 2020-11-26 | 0.0179 | 0.02005000 |
| 2020-11-27 | 0.0096 | 0.01656667 |
| 2020-11-28 | 0.0606 | 0.02757500 |
| 2020-11-29 | 0.0500 | 0.03206000 |
| 2020-11-30 | 0.0500 | 0.03505000 |

Insights - The above table shows the rolling average throughput as per the days.

C.Language share analysis

We have calculated percentage share of each language by counting the number of jobs for each language la and total number of jobs tc. Then we have computed share of each language in percentage by dividing job count to total bs and into 100, then we have rounded it of to two decimal places and at the end we have printed it in descending order.

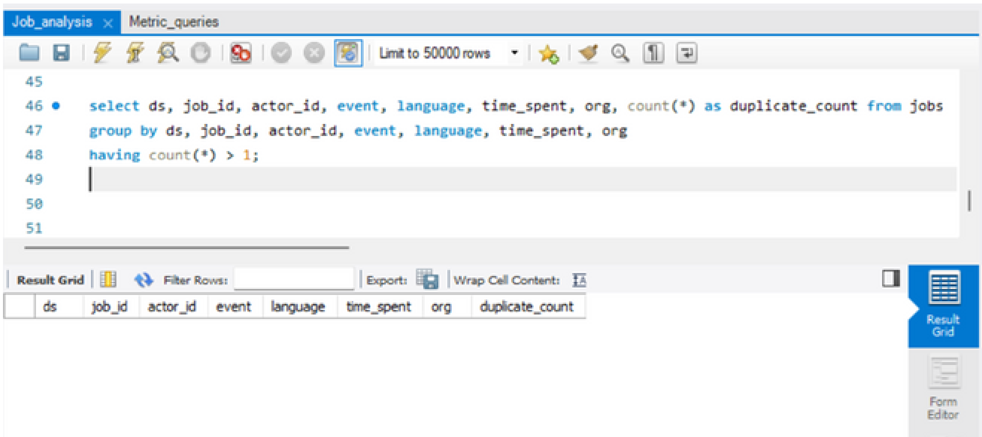
```
select la.language, round((la.job_count * 100.0 / tc.total_jobs), 2) as percentage_share
from (select language, count(job_id) as job_count from jobs where ds >= '2020-11-01' and ds <= '2020-11-30'
    group by language) la,
   (select count(job_id) as total_jobs
    from jobs where ds >= '2020-11-01' and ds <= '2020-11-30') tc
order by percentage_share desc;
```

| language | percentage_share |
|---|---|
| Persian | 37.50 |
| English | 12.50 |
| Arabic | 12.50 |
| Hindi | 12.50 |
| French | 12.50 |
| Italian | 12.50 |

Insights- The above table shows the percentage of language share and shows that persian has the most percentage share.

D.Duplicate rows detection

We have write this query to identify the duplicate rows in jobs table. We have first grouped the records by columns. We have then counted the number of the group has occurred and then filter it out. We have given the filter to only having more than one occurrence. Through this we are able to find duplicate count.




Insights- The above output shows that according to the dataset there is no duplicate rows currently present in the dataset.

Case Study 2: Investigating Metric Spike

A.Weekly user engagement

Here we have aggregated weekly active users by year and month for events which have classified as engagement through this we can extract year and month from the column occurred_at. We have used floor function to calculate the week of the month. Weekly active users are counted by distinct user_id for each week. In the end we have grouped and ordered by year, month and week of month.

Job_analysis | Metric_queries

```sql
select year(occurred_at) as year, month(occurred_at) as month,
floor((day(occurred_at) - 1) / 7) + 1 as week_of_month,
count(distinct user_id) as weekly_active_users
from events where event_type = 'engagement'
group by year, month, week_of_month
order by year, month, week_of_month;
```

| year | month | week_of_month | weekly_active_users |
|---|---|---|---|
| 2014 | 5 | 1 | 1058 |
| 2014 | 5 | 2 | 1068 |
| 2014 | 5 | 3 | 1126 |
| 2014 | 5 | 4 | 1135 |
| 2014 | 5 | 5 | 681 |
| 2014 | 6 | 1 | 1186 |
| 2014 | 6 | 2 | 1232 |
| 2014 | 6 | 3 | 1275 |
| 2014 | 6 | 4 | 1264 |
| 2014 | 6 | 5 | 416 |
| 2014 | 7 | 1 | 1310 |
| 2014 | 7 | 2 | 1368 |

Result 24

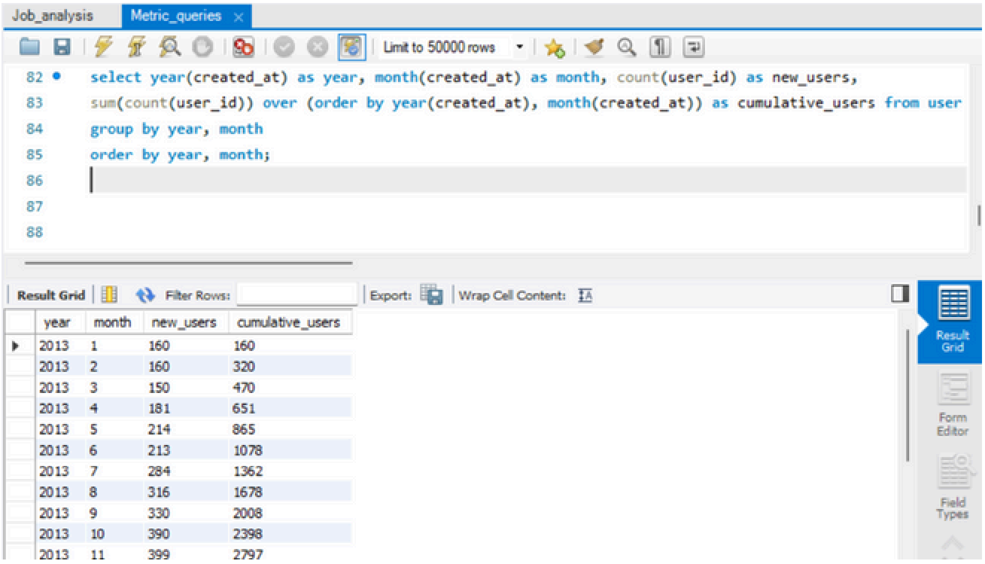
| year | month | week_of_month | weekly_active_users |
|---|---|---|---|
| 2014 | 6 | 5 | 416 |
| 2014 | 7 | 1 | 1310 |
| 2014 | 7 | 2 | 1368 |
| 2014 | 7 | 3 | 1373 |
| 2014 | 7 | 4 | 1411 |
| 2014 | 7 | 5 | 994 |
| 2014 | 8 | 1 | 1313 |
| 2014 | 8 | 2 | 1233 |
| 2014 | 8 | 3 | 1201 |
| 2014 | 8 | 4 | 1175 |
| 2014 | 8 | 5 | 608 |

Result 24

Insights- The above table shows the weekly active users according to the month.

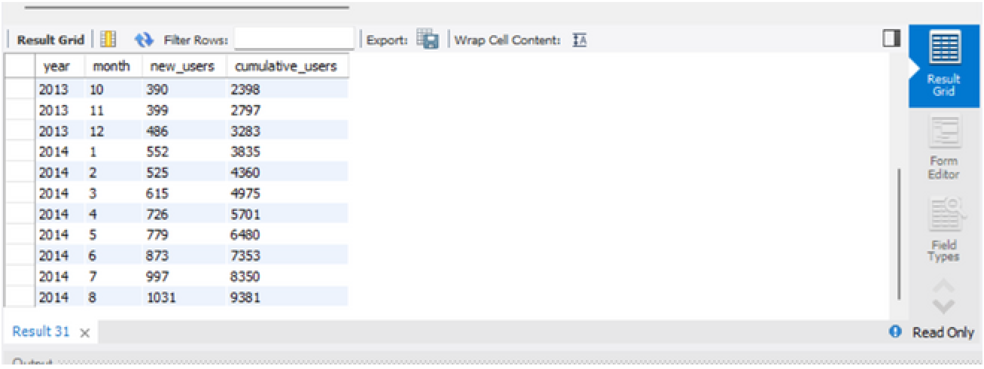
B. User growth analysis

In this we have retrieved number of users added in each month as an active user and then we have cumulate number of users over time to find user growth. We first have extracted the year and month from the created_at column from user table. To find monthly user we have counted user_id for each month. We have then used sum window function to calculate cumulative total which is ordered by year and month. In the end the result is grouped and ordered by year and month.



```sql
select year(created_at) as year, month(created_at) as month, count(user_id) as new_users,
sum(count(user_id)) over (order by year(created_at), month(created_at)) as cumulative_users from user
group by year, month
order by year, month;
```

| year | month | new_users | cumulative_users |
|---|---|---|---|
| 2013 | 1 | 160 | 160 |
| 2013 | 2 | 160 | 320 |
| 2013 | 3 | 150 | 470 |
| 2013 | 4 | 181 | 651 |
| 2013 | 5 | 214 | 865 |
| 2013 | 6 | 213 | 1078 |
| 2013 | 7 | 284 | 1362 |
| 2013 | 8 | 316 | 1678 |
| 2013 | 9 | 330 | 2008 |
| 2013 | 10 | 390 | 2398 |
| 2013 | 11 | 399 | 2797 |

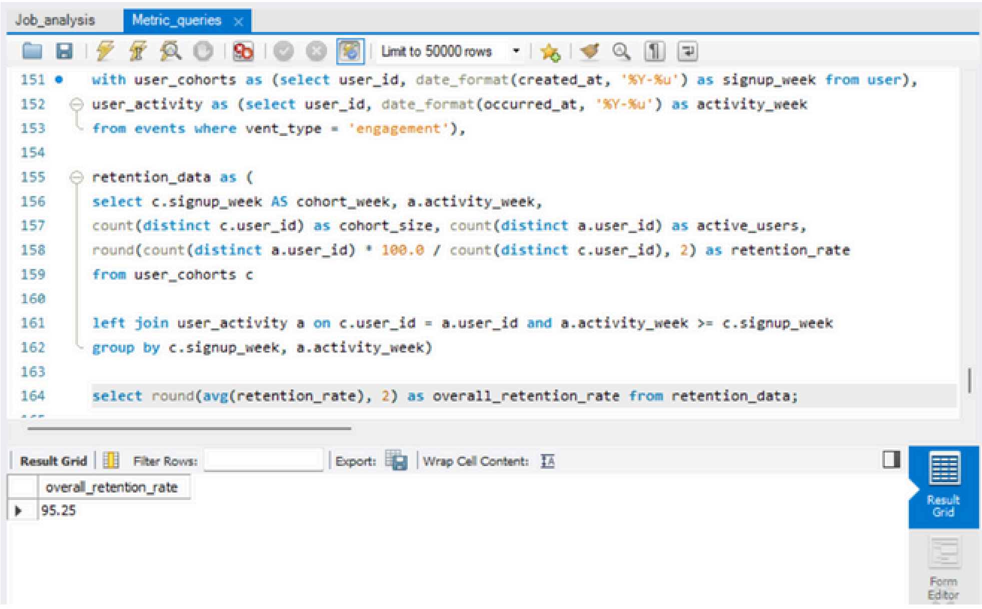


| year | month | new_users | cumulative_users |
|---|---|---|---|
| 2013 | 10 | 390 | 2398 |
| 2013 | 11 | 399 | 2797 |
| 2013 | 12 | 486 | 3283 |
| 2014 | 1 | 552 | 3835 |
| 2014 | 2 | 525 | 4360 |
| 2014 | 3 | 615 | 4975 |
| 2014 | 4 | 726 | 5701 |
| 2014 | 5 | 779 | 6480 |
| 2014 | 6 | 873 | 7353 |
| 2014 | 7 | 997 | 8350 |
| 2014 | 8 | 1031 | 9381 |

Insights- The above table shows the number of new users as per the month and then we have added the users to find the growth analysis and 9381 are number of users the growth has been seen.

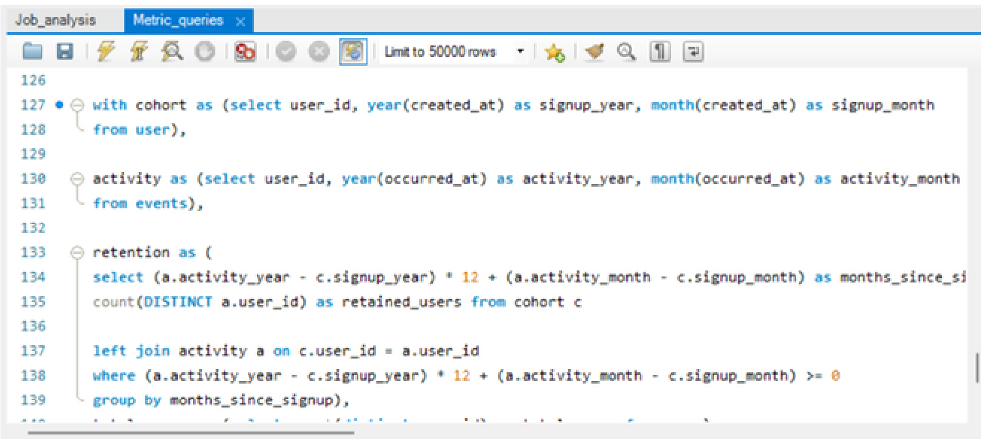
C. Weekly retention analysis

First to understand we have calculated the overall retention rate of user cohorts over time. By creating subqueries to determine the signup week and activity week for every user. Then the retention_data subquery is used to group user by signup week and then to finally calculate retention rates for each cohort by comparing the number of active users to chort size. Then at the end finally we have calculated the average retention rate across all cohorts.




Insights- The above output shows that the retention rate is 95.25.

Here we have calculated overall retention rate of users over the time. First we have calculated subqueries first to track user signups second to track activities of users and third retention is made to determine number of retained users per month since signup. Then to find total number of users another query total_users is made. Finally we have calculated overall retention rate by dividing number of retained users by total number of users. Finally we have grouped and ordered the result by months since their signup.



Job_analysis | Metric_queries

Limit to 50000 rows

```sql
count(DISTINCT a.user_id) as retained_users from cohort c

left join activity a on c.user_id = a.user_id
where (a.activity_year - c.signup_year) * 12 + (a.activity_month - c.signup_month) >= 0
group by months_since_signup),
total_users as (select count(distinct user_id) as total_users from user)

select r.months_since_signup, r.retained_users, t.total_users,
round ((r.retained_users / t.total_users) * 100, 2) as overall_retention_rate from retention r

cross join total_users t
order by r.months_since_signup;
```

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

| months_since_signup | retained_users | total_users | overall_retention_rate |
|---|---|---|---|
| 0 | 3680 | 9381 | 39.23 |
| 1 | 1699 | 9381 | 18.11 |
| 2 | 701 | 9381 | 7.47 |
| 3 | 587 | 9381 | 6.26 |
| 4 | 540 | 9381 | 5.76 |
| 5 | 480 | 9381 | 5.12 |

Result 44 | Read Only


```sql
total_users as (select count(distinct user_id) as total_users from user)
```


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

| months_since_signup | retained_users | total_users | overall_retention_rate |
|---|---|---|---|
| 5 | 480 | 9381 | 5.12 |
| 6 | 415 | 9381 | 4.42 |
| 7 | 365 | 9381 | 3.89 |
| 8 | 325 | 9381 | 3.46 |
| 9 | 335 | 9381 | 3.57 |
| 10 | 281 | 9381 | 3.00 |
| 11 | 251 | 9381 | 2.68 |
| 12 | 218 | 9381 | 2.32 |
| 13 | 187 | 9381 | 1.99 |
| 14 | 187 | 9381 | 1.99 |
| 15 | 182 | 9381 | 1.94 |
| 16 | 169 | 9381 | 1.80 |
| 17 | 119 | 9381 | 1.27 |
| 18 | 69 | 9381 | 0.74 |
| 19 | 29 | 9381 | 0.31 |

Result 44 | Read Only

Insights- The above table shows the retention rate as per the month since signup which will help to understand when the retention rate was high.

Here we have aimed to calculate retention time. By first creating subqueries to calculate cohort for tracking user signups then activity to track activities done by users and finally retention to calculate number of retained users per cohort. After this we have join the retention and cohert to obtain cohort size and then calculate retention rate which is obtained by dividing retained users by cohort size. Finally the result are then grouped and ordered by signup year and month and then months since signup.

Job_analysis | Metric_queries

Limit to 50000 rows

```sql
with cohort as ( select user_id, year(created_at) as signup_year, month(created_at) as signup_month,
monthname(created_at) as signup_month_name from user),

activity as ( select user_id, year(occurred_at) as activity_year, month(occurred_at) as activity_montl
monthname(occurred_at) as activity_month_name from events),

retention as (select c.signup_year, c.signup_month, c.signup_month_name,
(a.activity_year - c.signup_year) * 12 + (a.activity_month - c.signup_month) as months_since_signup,

count(distinct a.user_id) as retained_users from cohort c
left join activity a on c.user_id = a.user_id
where (a.activity_year - c.signup_year) * 12 + (a.activity_month - c.signup_month) >= 0
group by c.signup_year, c.signup_month, c.signup_month_name, months_since_signup
```

Job_analysis | Metric_queries

Limit to 50000 rows

```sql
group by c.signup_year, c.signup_month, c.signup_month_name, months_since_signup
)

select r.signup_year, r.signup_month, r.signup_month_name, r.months_since_signup, r.retained_users,
count(distinct c.user_id) as cohort_size,

round((r.retained_users / count(distinct c.user_id)) * 100, 2) as retention_rate from retention r

join cohort c on r.signup_year = c.signup_year and r.signup_month = c.signup_month

group by r.signup_year, r.signup_month, r.signup_month_name, r.months_since_signup
order by r.signup_year, r.signup_month, r.months_since_signup;
```

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

| signup_year | signup_month | signup_month_name | months_since_signup | retained_users | cohort_size | retention_rate |
|---|---|---|---|---|---|---|
| 2013 | 1 | January | 16 | 33 | 160 | 20.63 |
| 2013 | 1 | January | 17 | 37 | 160 | 23.13 |
| 2013 | 1 | January | 18 | 38 | 160 | 23.75 |
| 2013 | 1 | January | 19 | 29 | 160 | 18.13 |
| 2013 | 2 | February | 15 | 43 | 160 | 26.88 |

Result Grid

Result 41 | Read Only

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

| signup_year | signup_month | signup_month_name | months_since_signup | retained_users | cohort_size | retention_rate |
|---|---|---|---|---|---|---|
| 2013 | 1 | January | 16 | 33 | 160 | 20.63 |
| 2013 | 1 | January | 17 | 37 | 160 | 23.13 |
| 2013 | 1 | January | 18 | 38 | 160 | 23.75 |
| 2013 | 1 | January | 19 | 29 | 160 | 18.13 |
| 2013 | 2 | February | 15 | 43 | 160 | 26.88 |
| 2013 | 2 | February | 16 | 47 | 160 | 29.38 |
| 2013 | 2 | February | 17 | 49 | 160 | 30.63 |
| 2013 | 2 | February | 18 | 31 | 160 | 19.38 |
| 2013 | 3 | March | 14 | 39 | 150 | 26.00 |
| 2013 | 3 | March | 15 | 47 | 150 | 31.33 |
| 2013 | 3 | March | 16 | 44 | 150 | 29.33 |
| 2013 | 3 | March | 17 | 33 | 150 | 22.00 |
| 2013 | 4 | April | 13 | 51 | 181 | 28.18 |
| 2013 | 4 | April | 14 | 55 | 181 | 30.39 |

Result Grid | Form Editor | Field Types | Query Stats

Result 41 | Read Only

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

| signup_year | signup_month | signup_month_name | months_since_signup | retained_users | cohort_size | retention_rate |
|---|---|---|---|---|---|---|
| 2013 | 3 | March | 17 | 33 | 150 | 22.00 |
| 2013 | 4 | April | 13 | 51 | 181 | 28.18 |
| 2013 | 4 | April | 14 | 55 | 181 | 30.39 |
| 2013 | 4 | April | 15 | 52 | 181 | 28.73 |
| 2013 | 4 | April | 16 | 45 | 181 | 24.86 |
| 2013 | 5 | May | 12 | 50 | 214 | 23.36 |
| 2013 | 5 | May | 13 | 45 | 214 | 21.03 |
| 2013 | 5 | May | 14 | 56 | 214 | 26.17 |
| 2013 | 5 | May | 15 | 40 | 214 | 18.69 |
| 2013 | 6 | June | 11 | 45 | 213 | 21.13 |
| 2013 | 6 | June | 12 | 41 | 213 | 19.25 |
| 2013 | 6 | June | 13 | 52 | 213 | 24.41 |
| 2013 | 6 | June | 14 | 37 | 213 | 17.37 |
| 2013 | 7 | July | 10 | 66 | 284 | 23.24 |

Result 41 × Read Only

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

| signup_year | signup_month | signup_month_name | months_since_signup | retained_users | cohort_size | retention_rate |
|---|---|---|---|---|---|---|
| 2013 | 6 | June | 14 | 37 | 213 | 17.37 |
| 2013 | 7 | July | 10 | 66 | 284 | 23.24 |
| 2013 | 7 | July | 11 | 64 | 284 | 22.54 |
| 2013 | 7 | July | 12 | 66 | 284 | 23.24 |
| 2013 | 7 | July | 13 | 39 | 284 | 13.73 |
| 2013 | 8 | August | 9 | 86 | 316 | 27.22 |
| 2013 | 8 | August | 10 | 79 | 316 | 25.00 |
| 2013 | 8 | August | 11 | 82 | 316 | 25.95 |
| 2013 | 8 | August | 12 | 61 | 316 | 19.30 |
| 2013 | 9 | September | 8 | 69 | 330 | 20.91 |
| 2013 | 9 | September | 9 | 82 | 330 | 24.85 |
| 2013 | 9 | September | 10 | 72 | 330 | 21.82 |
| 2013 | 9 | September | 11 | 60 | 330 | 18.18 |
| 2013 | 10 | October | 7 | 87 | 390 | 22.31 |

Result 41 × Read Only

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

| signup_year | signup_month | signup_month_name | months_since_signup | retained_users | cohort_size | retention_rate |
|---|---|---|---|---|---|---|
| 2013 | 10 | October | 7 | 87 | 390 | 22.31 |
| 2013 | 10 | October | 8 | 91 | 390 | 23.33 |
| 2013 | 10 | October | 9 | 95 | 390 | 24.36 |
| 2013 | 10 | October | 10 | 64 | 390 | 16.41 |
| 2013 | 11 | November | 6 | 80 | 399 | 20.05 |
| 2013 | 11 | November | 7 | 88 | 399 | 22.06 |
| 2013 | 11 | November | 8 | 100 | 399 | 25.06 |
| 2013 | 11 | November | 9 | 72 | 399 | 18.05 |
| 2013 | 12 | December | 5 | 108 | 486 | 22.22 |
| 2013 | 12 | December | 6 | 106 | 486 | 21.81 |
| 2013 | 12 | December | 7 | 99 | 486 | 20.37 |
| 2013 | 12 | December | 8 | 65 | 486 | 13.37 |
| 2014 | 1 | January | 4 | 128 | 552 | 23.19 |
| 2014 | 1 | January | 5 | 120 | 552 | 21.74 |

Result 41 × Read Only

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

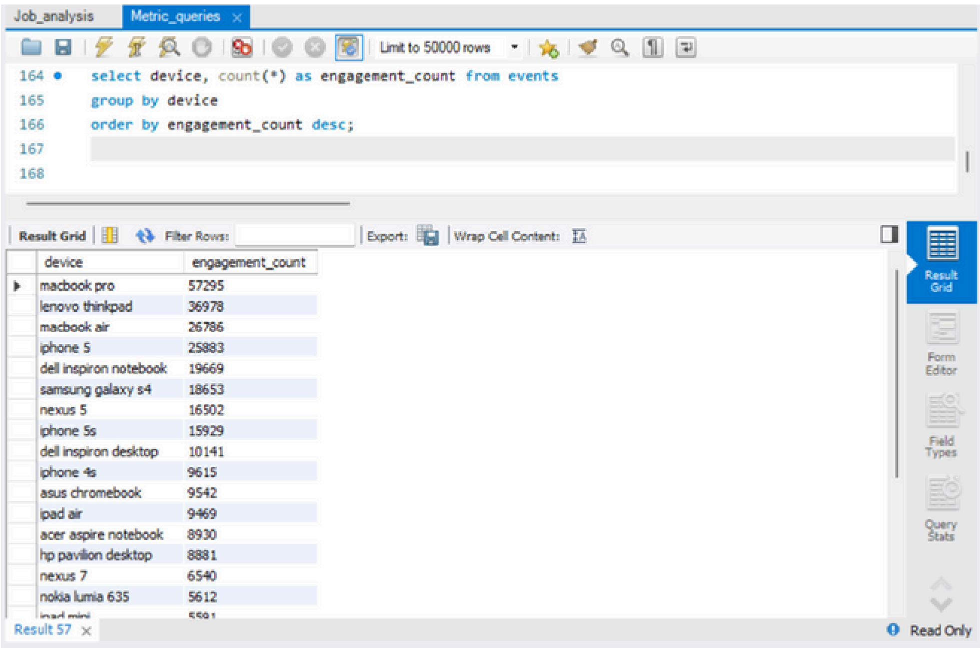
| signup_year | signup_month | signup_month_name | months_since_signup | retained_users | cohort_size | retention_rate |
|---|---|---|---|---|---|---|
| 2014 | 1 | January | 4 | 128 | 552 | 23.19 |
| 2014 | 1 | January | 5 | 120 | 552 | 21.74 |
| 2014 | 1 | January | 6 | 135 | 552 | 24.46 |
| 2014 | 1 | January | 7 | 91 | 552 | 16.49 |
| 2014 | 2 | February | 3 | 136 | 525 | 25.90 |
| 2014 | 2 | February | 4 | 126 | 525 | 24.00 |
| 2014 | 2 | February | 5 | 129 | 525 | 24.57 |
| 2014 | 2 | February | 6 | 94 | 525 | 17.90 |
| 2014 | 3 | March | 2 | 182 | 615 | 29.59 |
| 2014 | 3 | March | 3 | 141 | 615 | 22.93 |
| 2014 | 3 | March | 4 | 160 | 615 | 26.02 |
| 2014 | 3 | March | 5 | 123 | 615 | 20.00 |
| 2014 | 4 | April | 1 | 379 | 726 | 52.20 |
| 2014 | 4 | April | 2 | 168 | 726 | 23.14 |

Result 41 × Read Only

Insights- The table shows the retention rate as per the month and according to the month since sign up which helps to understand in which month and for how days until signup has retention rate.

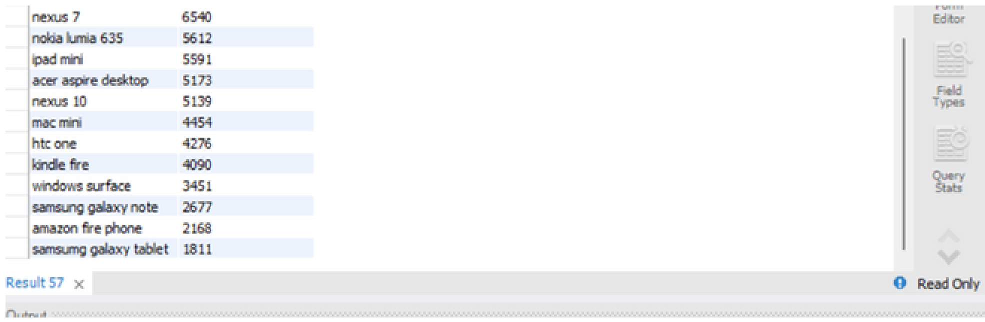
D.Weekly engagement per device

Here first to understand we have counted number of engagement events for each type of device. Then the result is grouped by device and is then ordered by engagement count to arrange in descending order to see the most used devices first.




```sql
select device, count(*) as engagement_count from events
group by device
order by engagement_count desc;
```

| device | engagement_count |
|---|---|
| macbook pro | 57295 |
| lenovo thinkpad | 36978 |
| macbook air | 26786 |
| iphone 5 | 25883 |
| dell inspiron notebook | 19669 |
| samsung galaxy s4 | 18653 |
| nexus 5 | 16502 |
| iphone 5s | 15929 |
| dell inspiron desktop | 10141 |
| iphone 4s | 9615 |
| asus chromebook | 9542 |
| ipad air | 9469 |
| acer aspire notebook | 8930 |
| hp pavilion desktop | 8881 |
| nexus 7 | 6540 |
| nokia lumia 635 | 5612 |




| | |
|---|---|
| nexus 7 | 6540 |
| nokia lumia 635 | 5612 |
| ipad mini | 5591 |
| acer aspire desktop | 5173 |
| nexus 10 | 5139 |
| mac mini | 4454 |
| htc one | 4276 |
| kindle fire | 4090 |
| windows surface | 3451 |
| samsung galaxy note | 2677 |
| amazon fire phone | 2168 |
| samsumg galaxy tablet | 1811 |

Insights- The table above table shows the total engagement count of users per device.

Here we have calculated average weekly engagement for every device. First we have created weekly _engagement subquery to group events by device, year and week. It is done to count engagement. Then we have averaged out the weekly engagement counts for each device. At the end we have ordered it by average weekly engagement in descending orde

Job_analysis | Metric_queries

```sql
with weekly_engagements as (
select device, year(occurred_at) as year, week(occurred_at, 1) as week, count(*) as engagement_count
from events

group by device, year(occurred_at), week(occurred_at, 1))

select device, avg(engagement_count) as avg_weekly_engagement
from weekly_engagements

group by device
order by avg_weekly_engagement desc;
```

| device | avg_weekly_engagement |
|---|---|
| macbook pro | 3183.0556 |
| lenovo thinkpad | 2054.3333 |
| macbook air | 1488.1111 |
| iphone 5 | 1437.9444 |
| dell inspiron notebook | 1092.7222 |
| samsung galaxy s4 | 1036.2778 |
| nexus 5 | 916.7778 |
| iphone 5s | 884.9444 |

Result 63

| device | avg_weekly_engagement |
|---|---|
| macbook pro | 3183.0556 |
| lenovo thinkpad | 2054.3333 |
| macbook air | 1488.1111 |
| iphone 5 | 1437.9444 |
| dell inspiron notebook | 1092.7222 |
| samsung galaxy s4 | 1036.2778 |
| nexus 5 | 916.7778 |
| iphone 5s | 884.9444 |
| dell inspiron desktop | 563.3889 |
| iphone 4s | 534.1667 |
| asus chromebook | 530.1111 |
| ipad air | 526.0556 |
| acer aspire notebook | 496.1111 |
| hp pavilion desktop | 493.3889 |
| nexus 7 | 363.3333 |
| nokia lumia 635 | 311.7778 |

Result 63

| device | avg_weekly_engagement |
|---|---|
| asus chromebook | 530.1111 |
| ipad air | 526.0556 |
| acer aspire notebook | 496.1111 |
| hp pavilion desktop | 493.3889 |
| nexus 7 | 363.3333 |
| nokia lumia 635 | 311.7778 |
| ipad mini | 310.6111 |
| acer aspire desktop | 287.3889 |
| nexus 10 | 285.5000 |
| mac mini | 247.4444 |
| htc one | 237.5556 |
| kindle fire | 227.2222 |
| windows surface | 191.7222 |
| samsung galaxy note | 148.7222 |
| amazon fire phone | 120.4444 |
| samsumg galaxy tablet | 100.6111 |

Result 63

Insights- The above table shows the average user engagement per week. This will help to understand the growth and take operational decision.

In this we have calculated number of engagement events for each device on weekly basis. Then we have grouped the data by device, year and week. After that number of engagements for each group is counted. At the end the result is ordered by year, device and week to observe the trend of engagement over the time period for each device.

```sql
select device, year(occurred_at) as year, week(occurred_at, 1) as week, count(*) as engagement_count
from events
group by device, year(occurred_at), week(occurred_at, 1)
order by device, year, week;
```

| device | year | week | engagement_count |
|---|---|---|---|
| acer aspire desktop | 2014 | 18 | 71 |
| acer aspire desktop | 2014 | 19 | 300 |
| acer aspire desktop | 2014 | 20 | 252 |
| acer aspire desktop | 2014 | 21 | 228 |
| acer aspire desktop | 2014 | 22 | 318 |
| acer aspire desktop | 2014 | 23 | 261 |
| acer aspire desktop | 2014 | 24 | 258 |
| acer aspire desktop | 2014 | 25 | 279 |
| acer aspire desktop | 2014 | 26 | 278 |
| acer aspire desktop | 2014 | 27 | 308 |
| acer aspire desktop | 2014 | 28 | 326 |
| acer aspire desktop | 2014 | 29 | 268 |
| acer aspire desktop | 2014 | 30 | 231 |
| acer aspire desktop | 2014 | 31 | 408 |
| acer aspire desktop | 2014 | 32 | 370 |
| acer aspire desktop | 2014 | 33 | 377 |

| device | year | week | engagement_count |
|---|---|---|---|
| acer aspire desktop | 2014 | 33 | 377 |
| acer aspire desktop | 2014 | 34 | 362 |
| acer aspire desktop | 2014 | 35 | 278 |
| acer aspire notebook | 2014 | 18 | 215 |
| acer aspire notebook | 2014 | 19 | 386 |
| acer aspire notebook | 2014 | 20 | 399 |
| acer aspire notebook | 2014 | 21 | 487 |
| acer aspire notebook | 2014 | 22 | 480 |
| acer aspire notebook | 2014 | 23 | 407 |
| acer aspire notebook | 2014 | 24 | 474 |
| acer aspire notebook | 2014 | 25 | 526 |
| acer aspire notebook | 2014 | 26 | 599 |
| acer aspire notebook | 2014 | 27 | 327 |
| acer aspire notebook | 2014 | 28 | 605 |
| acer aspire notebook | 2014 | 29 | 590 |
| acer aspire notebook | 2014 | 30 | 539 |

| device | year | week | engagement_count |
|---|---|---|---|
| acer aspire notebook | 2014 | 30 | 539 |
| acer aspire notebook | 2014 | 31 | 662 |
| acer aspire notebook | 2014 | 32 | 554 |
| acer aspire notebook | 2014 | 33 | 604 |
| acer aspire notebook | 2014 | 34 | 482 |
| acer aspire notebook | 2014 | 35 | 594 |
| amazon fire phone | 2014 | 18 | 84 |
| amazon fire phone | 2014 | 19 | 179 |
| amazon fire phone | 2014 | 20 | 149 |
| amazon fire phone | 2014 | 21 | 104 |
| amazon fire phone | 2014 | 22 | 26 |
| amazon fire phone | 2014 | 23 | 48 |
| amazon fire phone | 2014 | 24 | 193 |
| amazon fire phone | 2014 | 25 | 145 |
| amazon fire phone | 2014 | 26 | 125 |
| amazon fire phone | 2014 | 27 | 137 |

| device | year | week | engagement_count |
|---|---|---|---|
| asus chromebook | 2014 | 24 | 704 |
| asus chromebook | 2014 | 25 | 442 |
| asus chromebook | 2014 | 26 | 461 |
| asus chromebook | 2014 | 27 | 609 |
| asus chromebook | 2014 | 28 | 528 |
| asus chromebook | 2014 | 29 | 547 |
| asus chromebook | 2014 | 30 | 570 |
| asus chromebook | 2014 | 31 | 494 |
| asus chromebook | 2014 | 32 | 677 |
| asus chromebook | 2014 | 33 | 590 |
| asus chromebook | 2014 | 34 | 616 |
| asus chromebook | 2014 | 35 | 563 |
| dell inspiron desktop | 2014 | 18 | 198 |
| dell inspiron desktop | 2014 | 19 | 700 |
| dell inspiron desktop | 2014 | 20 | 426 |
| dell inspiron desktop | 2014 | 21 | 516 |

Insights- The above table shows the number of engagement count per week according to the sign up week and according to the device.

E.Email engagement analysis

In this we have exam

| device | year | week | engagement_count |
|---|---|---|---|
| amazon fire phone | 2014 | 27 | 137 |
| amazon fire phone | 2014 | 28 | 109 |
| amazon fire phone | 2014 | 29 | 52 |
| amazon fire phone | 2014 | 30 | 94 |
| amazon fire phone | 2014 | 31 | 185 |
| amazon fire phone | 2014 | 32 | 147 |
| amazon fire phone | 2014 | 33 | 163 |
| amazon fire phone | 2014 | 34 | 117 |
| amazon fire phone | 2014 | 35 | 111 |
| asus chromebook | 2014 | 18 | 286 |
| asus chromebook | 2014 | 19 | 498 |
| asus chromebook | 2014 | 20 | 286 |
| asus chromebook | 2014 | 21 | 467 |
| asus chromebook | 2014 | 22 | 565 |
| asus chromebook | 2014 | 23 | 639 |
| asus chromebook | 2014 | 24 | 704 |

ine email events to determine unique users and number of actions for each action type. Then we have counted total number of occurrences of each action as action_count. Unique_users are printed as distinct users performing each action. At the end the results asre grouped and ordered by action count in descending order as to see most frequent actions first in place.

```sql
select action, count(*) as action_count, count(distinct user_id) as unique_users
from emailevents
group by action
order by action_count desc;
```

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

| action | action_count | unique_users |
|---|---|---|
| sent_weekly_digest | 57267 | 4111 |
| email_open | 20459 | 5927 |
| email_clickthrough | 9010 | 5277 |
| sent_reengagement_email | 3653 | 3653 |

Insights – The above table shows the total count of action of email as per user. The action is counted per user and the unique user. This will help to analyse the email engagement and it would be help to analyse the email performance.

Here we have analyzed user engagement of email events dataset. Here we have extracted occurred_at for month and year. Then they are grouped by month-year and action. For each group the number of distinct engaged users, engagement rate and total actions are calculated. At and they are ordered by month year and action.

Job_analysis | Metric_queries | Limit to 50000 rows

```sql
select date_format(occurred_at, '%Y-%M') as month_year, action,
count(distinct user_id) as engaged_users, count(*) as total_actions,
round(count(distinct user_id) / (select count(distinct user_id) from emailevents) * 100, 2) as engagement_rate
from emailevents

group by month_year, action
order by month_year, action;
```

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

| month_year | action | engaged_users | total_actions | engagement_rate |
|---|---|---|---|---|
| 2014-August | email_clickthrough | 1804 | 1992 | 29.20 |
| 2014-August | email_open | 3879 | 5978 | 62.78 |
| 2014-August | sent_reengagement_email | 1073 | 1073 | 17.37 |
| 2014-August | sent_weekly_digest | 4111 | 16480 | 66.53 |
| 2014-July | email_clickthrough | 2267 | 2721 | 36.69 |
| 2014-July | email_open | 3457 | 5611 | 55.95 |
| 2014-July | sent_reengagement_email | 933 | 933 | 15.10 |
| 2014-July | sent_weekly_digest | 3685 | 15902 | 59.64 |
| 2014-June | email_clickthrough | 1915 | 2274 | 30.99 |
| 2014-June | email_open | 3037 | 4658 | 49.15 |
| 2014-June | sent_reengagement_email | 889 | 889 | 14.39 |
| 2014-June | sent_weekly_digest | 3231 | 13155 | 52.29 |
| 2014-May | email_clickthrough | 1703 | 2023 | 27.56 |
| 2014-May | email_open | 2681 | 4212 | 43.39 |
| 2014-May | sent_reengagement_email | 758 | 758 | 12.27 |

Result 66 | Read Only

Insights- the above table shows the action and according to that total actions are given to find engagement rate is given. It helps to increase the performance of email as a platform.

Result

Through this project we learned to analyse the operations related to retention, user engagement, user activity and email interactions. Using MySQL workbench we found:

- User growth per month of new users
- Retention rate based on sign up cohorts weekly
- Usage of device per active user
- Average weekly activity engagement per device
- User monthly growth analysis
- Email engagement with email services

It helped to gain deeper knowledge to analyze the situation and gain more knowledge in SQL querying and relational database.

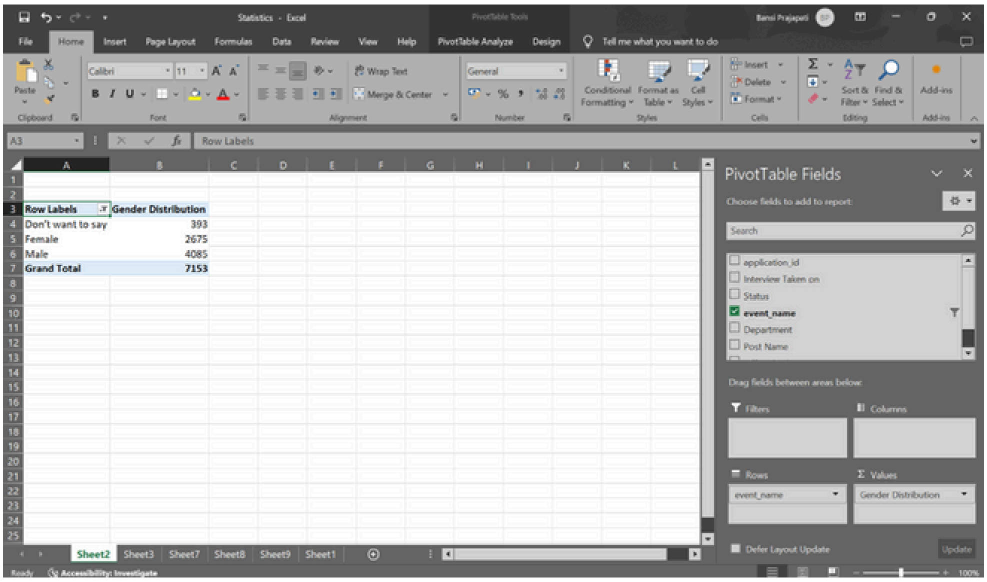
# Project – 4
# Hiring Process Analytics

Project Description
This project aims to analyse the hiring process within an organization. Here in this project we have analysed gender distribution, salary analyses, salary distribution understanding, department properties visualization and position tier distribution visualization. In conclusion we have gained insights into company's salary structure and hiring trends using excel as data analytics tool.

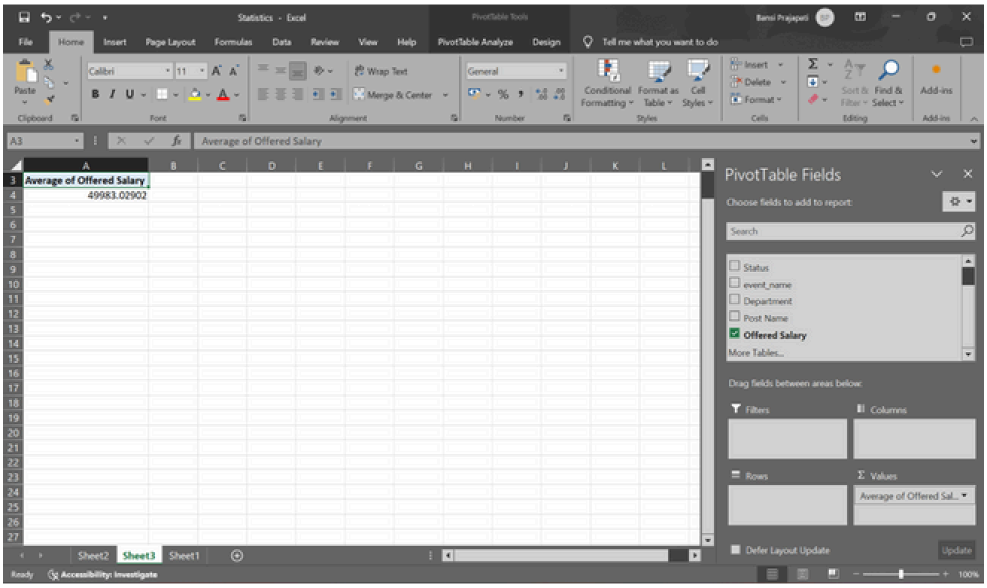
Approach
·Dataset was downloaded and imported into Microsoft excel
·Ensured dataset was clean and free of missing values
·Then we started with analysis using pivot tables and its function.
·For gender distribution we used pivot tables to count and compare the number of females and males hired.
·For salary analysis we calculated average salary using pivot tables.
·Then for salary distribution we used if function to first form salary range column and then by using pivot table we calculated the salary distribution.
·Then we analysed proportion of employees in different departments using pivot graphs and we plot the pie chart to show the distribution.
·To show the position tier analysis we used pivot graph to the position across different tiers using bar graph.

Tech-Stack Used
Microsoft® Word 2019 MSO (Version 2501 Build 16.0.18429.20132) 64-bit for data analysis, pivot table creation and chart generation.

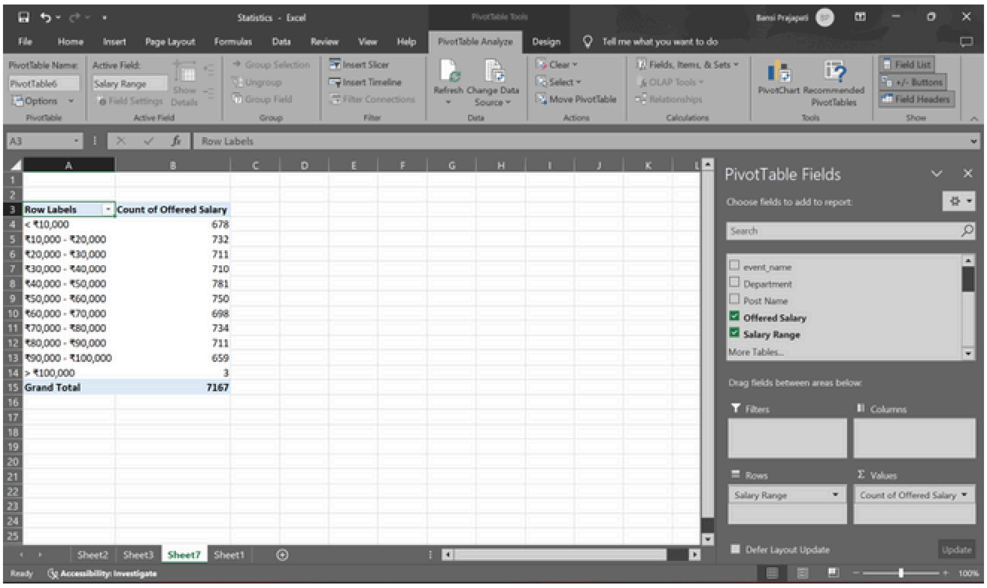
Insights
A.Hiring Analysis
Analysed proportion of males and females hired. Here 2675 female and 4085 males are hired revealing the gender imbalances.




B.Salary Analysis
Average salary offered by the company is 49983/-.

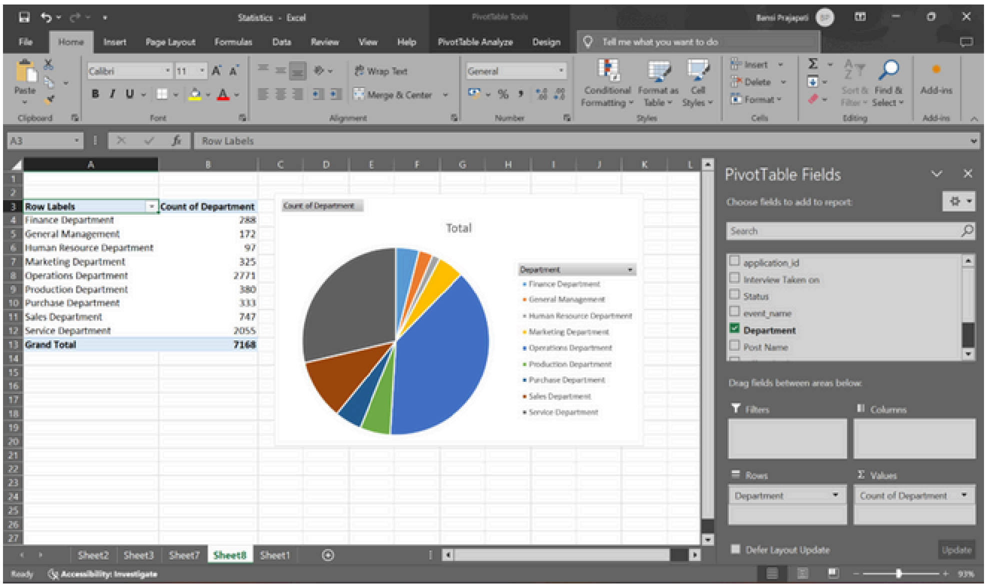



C.Salary Distribution
Visualized how salaries are distributed across different intervals. Here most people have salaries between 40,000-50,000. Which also tells that it is average salary among the company.



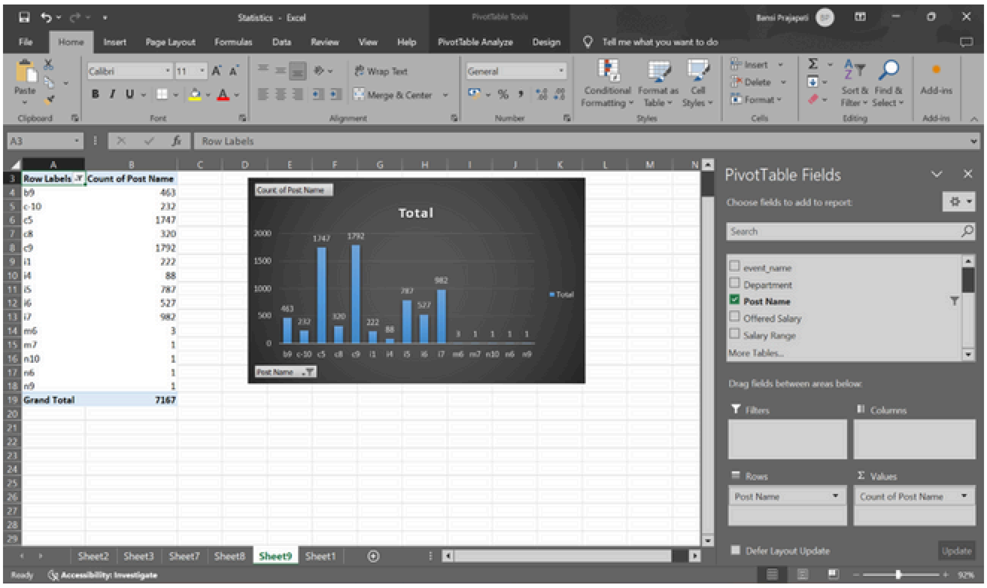

D.Departmental Analysis

Identified that operation department has most employees and human resource department has least number of employees.




E.Position Tier Analysis

Examined distribution of positions among the various tiers and it shows that c9 has most positions in company’s hierarchical structure.




Result
In this project we were able to derive key trends and patterns in salary structure, gender distribution and departmental composition. This insights help HR and management team to take an informed decisions regarding the salary adjustments, hiring practices and departmental resource allocation. Hence project helps to provide comprehensive understanding of hiring process within organization.

Drive link of Excel Sheet
Refer sheet 2, 3, 7, 8, 9 for the answers.
https://docs.google.com/spreadsheets/d/17NBk8-FEd5TLBjCX0QkAX5PN4gKKA-jf/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true

# Project – 5
# IMDB Movie Analysis

Project Description

This project aims to investigate "What factors influence the success of a movie on IMDB?" Here in this project we will analyze movie genre, duration, language, director and budget analysis defined by high IMDB ratings. This analysis is crucial for movie producers, directors, and investors to make informed decisions regarding future projects.

Approach

·Dataset was downloaded and imported into Microsoft excel

·Ensured dataset was clean and free of missing values

·Then we started with analysis using pivot tables and its function.

·For genre analysis we calculated no_of_movies, mean_imdb, max_imdb, min_imdb, StdDev_imdb and var_imdb from pivot table function and we derived the median_imdb, mode_imdb and range_imdb using function in excel sheet.

·Then we generated graph of no_of_movies v/s genre to understand the distribution and a line graph of imdb_score v/s genre to see the trend between there descriptive statistical function.

·For movie duration analysis we analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score using the scatter plot and by observing the trend line.

·For language analysis we have determined the most common languages used in movies and analyzed their impact on the IMDB score using descriptive statistics.

·For director analysis we have calculated the average IMDB score for each director. Used excel's percentile function to identify the directors with the highest scores. Then we compared the scores of these directors to the overall distribution of scores.

·For budget analysis we calculated the correlation coefficient between movie budgets and gross earnings using excel's correl function. Then we calculated the profit margin for each movie and identified the movies with the highest profit margin using excel's max function.

Tech Stack Used

Microsoft® Word 2019 MSO (Version 2501 Build 16.0.18429.20132) 64-bit for data analysis, pivot table creation and chart generation.

Insights

A. Movie Genre Analysis:

· Most common movie genre includes comedy with 1036 no. of movies and then action, drama and adventure with 9770, 700 and 378 no. of movies respectively.

· Least common genres include musical, romance, thriller and western.

· Highest mean imdb score is of biography, documentary and crime.

· #N/A appears when no single rating appers more than once in that genre.

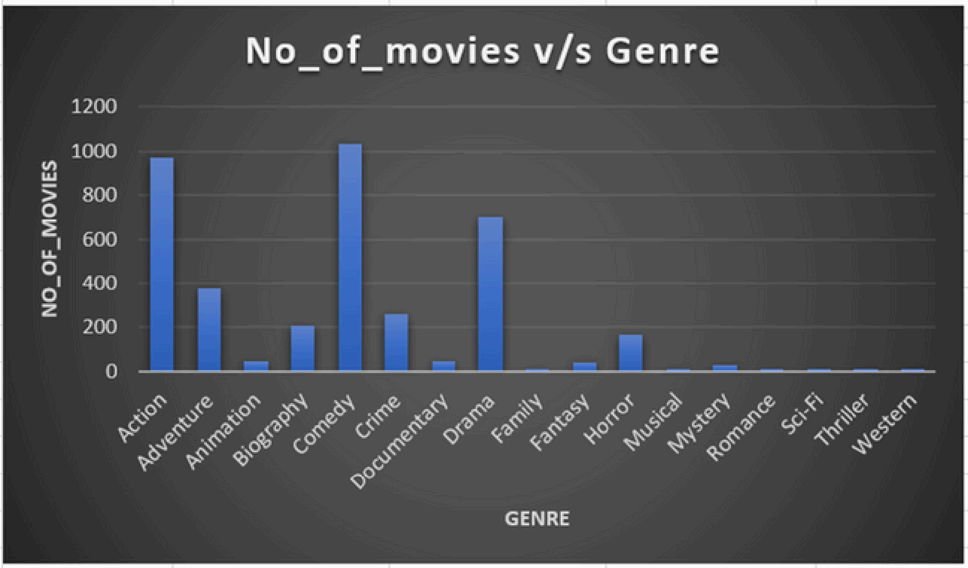
· Documentary is having most rating of 7.5 which is highest and most repeated.

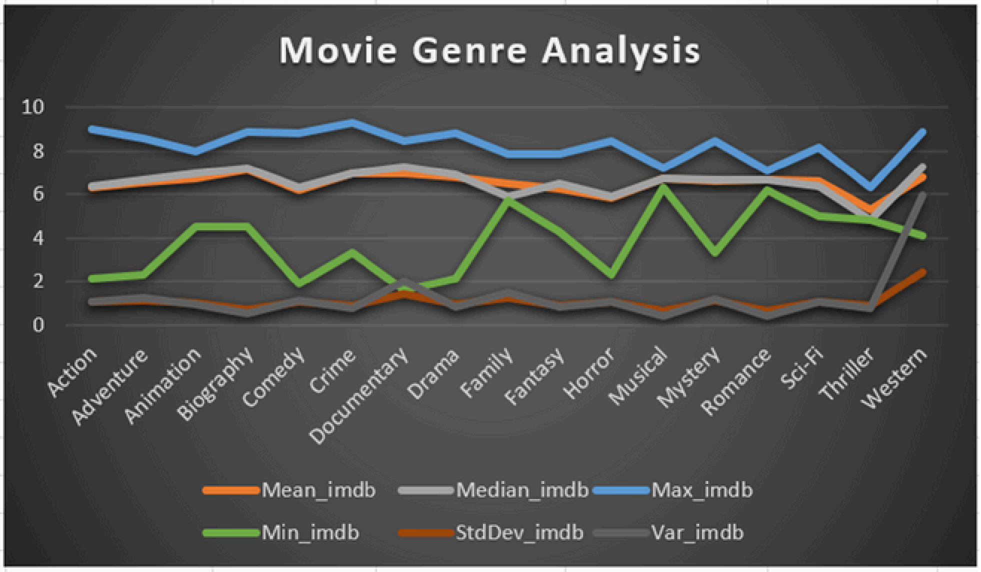
· The highest rated movie is having crime genre with 9.3 imdb score and lowest rated movie is genre having action and horror with 2.1 imdb score.

· Higher standard deviation means more unpredictable in audience reception.

· Western movies have most inconsistent ratings while animation and musical movies tend to have stable ratings.

| Genre | No_of_movies | Mean_imdb | Median_imdb | Mode_imdb | Max_imdb | Min_imdb | Range_imdb | StdDev_imdb | Var_imdb |
|---|---|---|---|---|---|---|---|---|---|
| Action | 970 | 6.290618557 | 6.35 | 6.6 | 9 | 2.1 | 6.9 | 1.038020576 | 1.07748672 |
| Adventure | 378 | 6.555291005 | 6.7 | 7.3 | 8.6 | 2.3 | 6.3 | 1.114308219 | 1.24168281 |
| Animation | 46 | 6.763043478 | 7 | 7.1 | 8 | 4.5 | 3.5 | 0.972593028 | 0.9459372 |
| Biography | 208 | 7.153846154 | 7.2 | 7 | 8.9 | 4.5 | 4.4 | 0.696564805 | 0.48520253 |
| Comedy | 1036 | 6.166409266 | 6.3 | 6.4 | 8.8 | 1.9 | 6.9 | 1.040657096 | 1.08296719 |
| Crime | 259 | 6.944787645 | 7 | 7.4 | 9.3 | 3.3 | 6 | 0.864548238 | 0.74744366 |
| Documentary | 43 | 6.951162791 | 7.3 | 7.5 | 8.5 | 1.6 | 6.9 | 1.41612686 | 2.00541528 |
| Drama | 700 | 6.814428571 | 6.9 | 6.7 | 8.8 | 2.1 | 6.7 | 0.905146841 | 0.8192908 |
| Family | 3 | 6.5 | 5.9 | #N/A | 7.9 | 5.7 | 2.2 | 1.216552506 | 1.48 |
| Fantasy | 37 | 6.281081081 | 6.5 | 6.8 | 7.9 | 4.3 | 3.6 | 0.894066191 | 0.79935435 |
| Horror | 165 | 5.850909091 | 5.9 | 5.9 | 8.5 | 2.3 | 6.2 | 1.032083979 | 1.06519734 |
| Musical | 2 | 6.75 | 6.75 | #N/A | 7.2 | 6.3 | 0.9 | 0.636396103 | 0.405 |
| Mystery | 24 | 6.608333333 | 6.7 | 7.1 | 8.5 | 3.3 | 5.2 | 1.0898411 | 1.18775362 |
| Romance | 2 | 6.65 | 6.65 | #N/A | 7.1 | 6.2 | 0.9 | 0.636396103 | 0.405 |
| Sci-Fi | 8 | 6.5875 | 6.35 | #N/A | 8.2 | 5 | 3.2 | 1.031555691 | 1.06410714 |
| Thriller | 3 | 5.3 | 4.8 | 4.8 | 6.3 | 4.8 | 1.5 | 0.866025404 | 0.75 |
| Western | 3 | 6.766666667 | 7.3 | #N/A | 8.9 | 4.1 | 4.8 | 2.444040371 | 5.97333333 |






B.Movie Duration Analysis:

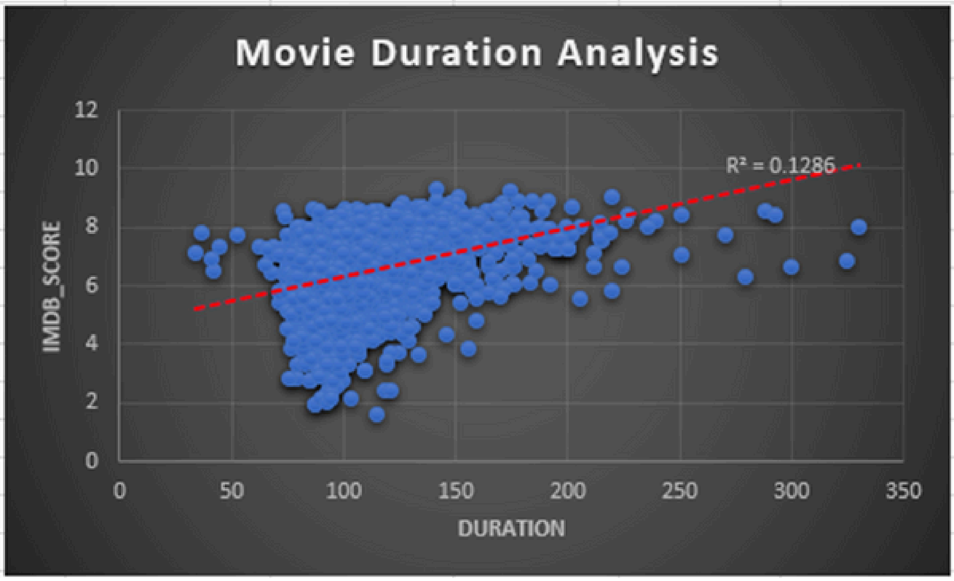
· The scatter plot shows positive correlation between movie duration and IMDB score.

· The $R^2$ value shows that the impact of duration of a movie has less impact on IMDB score and it is not a strong predictor.

· Most movies have duration around 100-110 minutes.

· Longer movies might tend to have slightly higher ratings, but duration only cannot determine IMDB score.

| Operations | Values |
|---|---|
| Mean | 109.901981 |
| Median | 106 |
| Mode | 101 |
| Standard Devat | 22.7125612 |
| Variance | 515.8604361 |

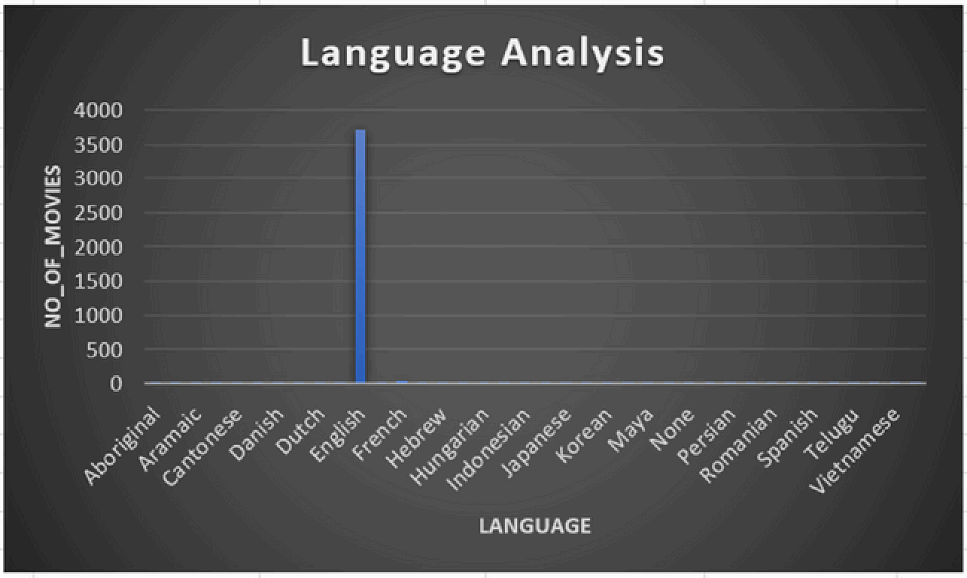



C. Language Analysis:

· With 3706 movies English is the dominating language.

· English movies have mean of 6.42 of imdb score which is lower it might be due larger dataset that contains both high and low rated movies.

· In conclusion we can observed that smaller language dataset might be biased towards high quality films, while English movies would cover wider spectrum including lower rated films.



| Language | No_of_Movies | Average_imdb | Median_imdb | Var_imdb | StdDev_imdb |
|---|---|---|---|---|---|
| Aboriginal | 2 | 6.95 | 6.95 | 0.605 | 0.777817459 |
| Arabic | 1 | 7.2 | 7.2 | #DIV/0! | #DIV/0! |
| Aramaic | 1 | 7.1 | 7.1 | #DIV/0! | #DIV/0! |
| Bosnian | 1 | 4.3 | 4.3 | #DIV/0! | #DIV/0! |
| Cantonese | 8 | 7.2375 | 7.3 | 0.194107143 | 0.440575922 |
| Czech | 1 | 7.4 | 7.4 | #DIV/0! | #DIV/0! |
| Danish | 3 | 7.9 | 8.1 | 0.28 | 0.529150262 |
| Dari | 2 | 7.5 | 7.4 | 0.542205882 | 0.736346306 |
| Dutch | 3 | 7.566666667 | 7.8 | 0.163333333 | 0.404145188 |
| Dzongkha | 1 | 7.5 | 7.5 | #DIV/0! | #DIV/0! |
| English | 3706 | 6.42172619 | 6.5 | 1.104471754 | 1.050938511 |
| Filipino | 1 | 6.7 | 6.7 | #DIV/0! | #DIV/0! |
| French | 37 | 7.286486486 | 7.2 | 0.31509009 | 0.561328861 |
| German | 13 | 7.692307692 | 7.7 | 0.410769231 | 0.640912811 |
| Hebrew | 3 | 7.5 | 7.3 | 0.19 | 0.435889894 |
| Hindi | 10 | 6.76 | 7.05 | 1.236 | 1.111755369 |
| Hungarian | 1 | 7.1 | 7.1 | #DIV/0! | #DIV/0! |
| Icelandic | 1 | 6.9 | 6.9 | #DIV/0! | #DIV/0! |
| Indonesian | 2 | 7.9 | 7.9 | 0.18 | 0.424264069 |
| Italian | 7 | 7.185714286 | 7 | 1.334761905 | 1.155318962 |
| Japanese | 12 | 7.625 | 7.8 | 0.809318182 | 0.899621132 |
| Kazakh | 1 | 6 | 6 | #DIV/0! | #DIV/0! |
| Korean | 5 | 7.7 | 7.7 | 0.325 | 0.570087713 |
| Mandarin | 15 | 7.08 | 7.4 | 0.596 | 0.772010363 |
| Maya | 1 | 7.8 | 7.8 | #DIV/0! | #DIV/0! |
| Mongolian | 1 | 7.3 | 7.3 | #DIV/0! | #DIV/0! |
| None | 1 | 8.5 | 8.5 | #DIV/0! | #DIV/0! |
| Norwegian | 4 | 7.15 | 7.3 | 0.33 | 0.574456265 |
| Persian | 3 | 8.133333333 | 8.4 | 0.303333333 | 0.550757055 |
| Portuguese | 5 | 7.76 | 8 | 0.958 | 0.978774744 |
| Romanian | 1 | 7.9 | 7.9 | #DIV/0! | #DIV/0! |
| Russian | 1 | 6.5 | 6.5 | #DIV/0! | #DIV/0! |
| Spanish | 26 | 7.05 | 7.15 | 0.6826 | 0.826196103 |
| Swedish | 1 | 7.6 | 7.6 | #DIV/0! | #DIV/0! |
| Telugu | 1 | 8.4 | 8.4 | #DIV/0! | #DIV/0! |
| Thai | 3 | 6.633333333 | 6.6 | 0.203333333 | 0.450924975 |
| Vietnamese | 1 | 7.4 | 7.4 | #DIV/0! | #DIV/0! |
| Zulu | 1 | 7.3 | 7.3 | #DIV/0! | #DIV/0! |

D.Director Analysis:

- The average IMDB score of highest rated directors is 7.5 percentile.
- By filtering the data we can identify top directors and which help to understand contributors to higher movie ratings.

| Percentile |
|---|
| 7.5 |

| | |
|---|---|
| r Haigh | 7.7 |
| r Stanton | 7.733333333 |
| tuyleert | 7.9 |
| y Powell | 7.6 |
| nan | 8 |
| Farhadi | 8.4 |
| V. Blaustein | 7.6 |
| Tuck | 7.65 |
| do Bertolucci | 7.8 |
| ilder | 8.3 |
| rd | 7.58 |
| enson | 7.7 |
| ercival | 7.6 |
| e Link | 7.7 |
| rook Park | 7.633333333 |
| Chaplin | 8.6 |
| Ferguson | 7.65 |
| ang Liu | 7.6 |
| uck | 7.6 |
| aine | 7.7 |
| in Carion | 7.8 |
| pho Barratier | 7.9 |
| pher Nolan | 8.425 |
| Lu | 7.7 |
| Mungiu | 7.9 |
| Chazelle | 8.5 |
| nny | 7.9 |
| incher | 7.75 |
| oen | 8 |
| ynch | 7.533333333 |
| ington | 8.1 |
| eBlois | 7.766666667 |
| Villeneuve | 7.966666667 |
| Arcand | 7.7 |
| ill | 7.9 |
| igel | 7.6 |
| Jones | 7.566666667 |
| Wright | 7.6 |
| ain | 8.2 |
| ris | 7.7 |
| Overbey | 7.6 |
| Bielinsky | 7.9 |
| do LeÃ³n de Aranoa | 7.7 |
| do Meirelles | 7.6 |
| ois Girard | 7.7 |
| Ford Coppola | 7.655555556 |
| arabont | 7.975 |
| ng | 8.3 |
| Evans | 7.6 |
| Cukor | 7.9 |
| Roy Hill | 8.2 |
| pe Tornatore | 7.8 |
| Miyazaki | 8.225 |
| Alex Rubin | 7.8 |
| Loiret | 7.6 |
| yona | 7.55 |
| Perrin | 7.9 |
| Nigar | 7.55 |
| Cameron | 7.914285714 |
| Marsh | 7.75 |
| Gang | 8.1 |
| Robbins | 7.6 |
| uhams | 7.8 |
| arney | 7.65 |
| an Dayton | 7.9 |
| amÃ¡rez SuÃ¡rez | 7.7 |
| one | 7.8 |
| Oppenheimer | 8.2 |
| Tickell | 7.6 |
| hedon | 7.925 |
| nÃ© Campanella | 8.2 |
| lak | 7.6 |
| ordan | 7.6 |
| Macdonald | 7.7 |
| erich | 8.3 |
| Abrahamson | 8.3 |
| acon | 7.7 |
| quet | 7.6 |
| Majidi | 8.5 |
| A. Markevicius | 8.4 |
| orman | 7.8 |
| sborne | 7.7 |
| andrich | 7.8 |
| McDonagh | 7.6 |
| Scorsese | 7.675 |
| arron | 7.6 |
| w Vaughn | 7.65 |
| ooks | 7.633333333 |
| nson | 7.766666667 |
| l McGowan | 7.6 |
| l Moore | 7.75 |
| l Roemer | 8.1 |
| l Wadleigh | 8.1 |
| Hazanavicius | 8 |
| an Diem | 7.8 |
| orman | 8.133333333 |
| Tyldum | 7.85 |
| Greno | 7.8 |
| Welles | 7.7 |
| lin | 7.6 |
| orrentino | 7.7 |
| erngrass | 7.585714286 |
| iggis | 7.733333333 |
| omas Anderson | 7.516666667 |
| cter | 8.233333333 |
| ackson | 7.675 |
| Weir | 7.725 |
| NÃ¡ss | 7.6 |
| Kaufman | 7.65 |
| avelle | 7.7 |
| Coffin | 7.6 |
| n Tarantino | 8.2 |
| Siman | 7.8 |
| oore | 7.8 |
| l Curtis | 7.75 |
| l Marquand | 8.4 |
| ern | 7.7 |
| Batra | 7.8 |
| Stevenson | 7.8 |
| cke | 8.5 |
| oogler | 7.6 |
| amouli | 8.4 |
| Leone | 8.433333333 |
| Meadows | 7.7 |
| oeze | 7.575 |
| hane Aubier | 7.9 |
| Kramer | 7.6 |
| Kubrick | 7.8 |
| n Chbosky | 8 |
| n Chow | 7.55 |
| ames | 7.55 |
| McQueen | 7.7 |
| Spielberg | 7.544 |
| Chomet | 7.8 |
| mme | 7.6 |
| eorge | 8.1 |
| harrock | 7.6 |
| s Vinterberg | 7.666666667 |
| Carlies | 7.6 |
| ller | 8.1 |
| ald | 7.55 |
| olenski | 7.65 |
| rd | 7.6 |
| ooper | 7.733333333 |
| cCarthy | 7.9 |
| Moore | 7.7 |

E.Budget Analysis:

·Avatar is having the highest progit margin of $523,505,847 which was calculated by gross earnings – budget.

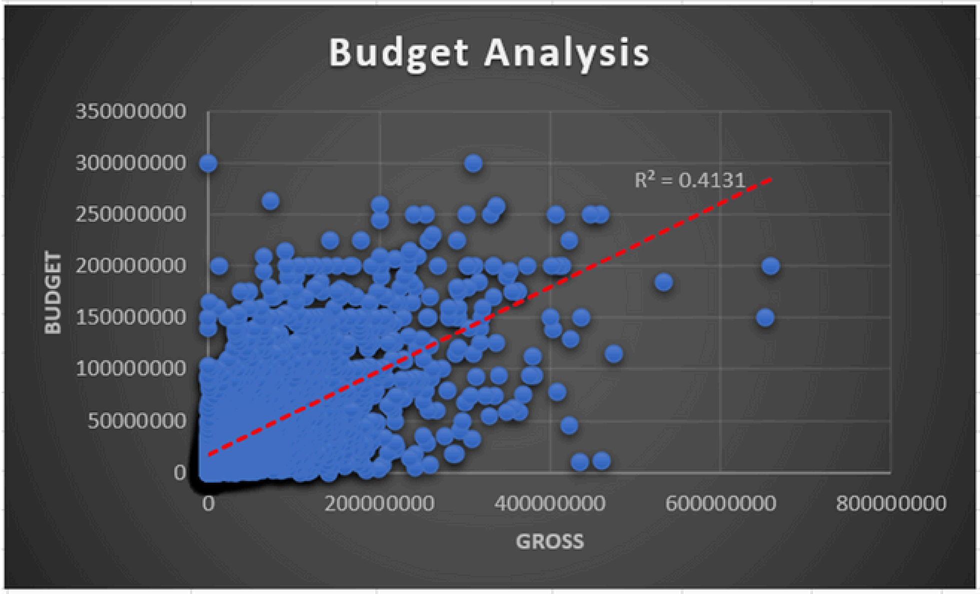
·The correlation coefficient is 0.0408 indicating very weak relationship between budget and gross earnings. This means higher budget does not guarantee higher revenue always.

·We have plot scatter plot by filtering some outliers to observe the values visually and $R^2$ suggest that 42% of variance in gross earnings can be explained by budget.

·In short lower budget movies may achieve exceptional financial success, while some expensive production may sometime under perform.

| movie_title | gross | budget | Profit_margin |
|---|---|---|---|
| AvatarÂ | 760505847 | 237000000 | 523505847 |
| Jurassic WorldÂ | 652177271 | 150000000 | 502177271 |
| TitanicÂ | 658672302 | 200000000 | 458672302 |
| Star Wars: Episode IV - A New HopeÂ | 460935665 | 11000000 | 449935665 |
| E.T. the Extra-TerrestrialÂ | 434949459 | 10500000 | 424449459 |
| The AvengersÂ | 623279547 | 220000000 | 403279547 |
| The AvengersÂ | 623279547 | 220000000 | 403279547 |
| The Lion KingÂ | 422783777 | 45000000 | 377783777 |
| Star Wars: Episode I - The Phantom MenaceÂ | 474544677 | 115000000 | 359544677 |
| The Dark KnightÂ | 533316061 | 185000000 | 348316061 |
| The Hunger GamesÂ | 407999255 | 78000000 | 329999255 |
| DeadpoolÂ | 363024263 | 58000000 | 305024263 |
| The Hunger Games: Catching FireÂ | 424645577 | 130000000 | 294645577 |
| Jurassic ParkÂ | 356784000 | 63000000 | 293784000 |
| Despicable Me 2Â | 368049635 | 76000000 | 292049635 |
| American SniperÂ | 350123553 | 58800000 | 291323553 |
| Finding NemoÂ | 380838870 | 94000000 | 286838870 |
| Shrek 2Â | 436471036 | 150000000 | 286471036 |
| The Lord of the Rings: The Return of the KingÂ | 377019252 | 94000000 | 283019252 |
| Star Wars: Episode VI - Return of the JediÂ | 309125409 | 32500000 | 276625409 |
| Forrest GumpÂ | 329691196 | 55000000 | 274691196 |
| Star Wars: Episode V - The Empire Strikes BackÂ | 290158751 | 18000000 | 272158751 |
| Home AloneÂ | 285761243 | 18000000 | 267761243 |
| Star Wars: Episode III - Revenge of the SithÂ | 380262555 | 113000000 | 267262555 |
| Spider-ManÂ | 403706375 | 139000000 | 264706375 |
| MinionsÂ | 336029560 | 74000000 | 262029560 |
| The Sixth SenseÂ | 293501675 | 40000000 | 253501675 |
| JawsÂ | 260000000 | 8000000 | 252000000 |
| FrozenÂ | 400736600 | 150000000 | 250736600 |
| The Secret Life of PetsÂ | 323505540 | 75000000 | 248505540 |

| Movie_title | Max_Profit |
|---|---|
| AvatarÂ | 523505847 |
| | |
| | |
| Correl | 0.040837208 |




Result
In this project we were able to derive key trends and patterns in factors that influence the success of a movie on IMDB. This insights help movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects. Hence project helps to provide comprehensive understanding of factors that influence success of a movie.

Drive link of Excel Sheet
https://docs.google.com/spreadsheets/d/1nG343gLUPXV4S1TqFZwNE9gRRVbkvg3j/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true

Loom video link
https://www.loom.com/share/234580d50bcc4375a2d3c936e650908b?sid=ef1238f6-d24e-4e0a-8233-ca887d15aba2

# Project – 6
# Bank Loan Case Study

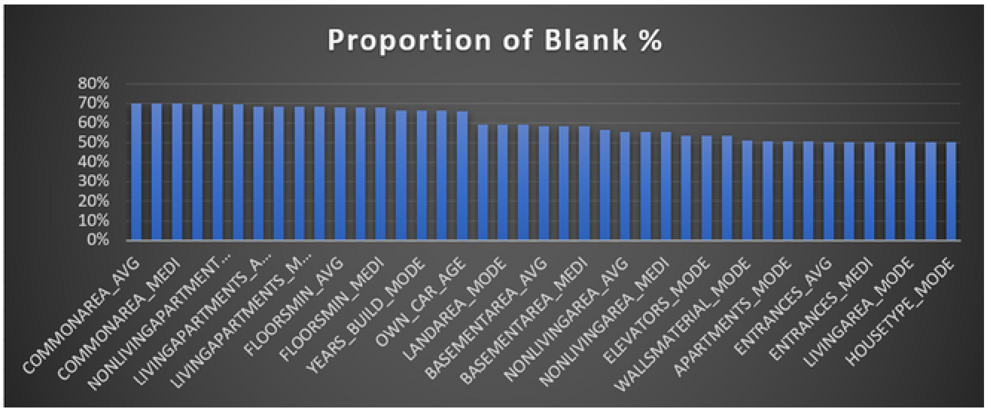
Project Description
The main aim of this project is to identify patterns that indicate if a customer will have difficulty paying their installments. This information can be used to make decisions such as denying the loan, reducing the amount of loan, or lending at a higher interest rate to risky applicants. The company wants to understand the key factors behind loan default so it can make better decisions about loan approval.

Approach
·Dataset was downloaded and imported into Microsoft excel
·Then clean the dataset finding the blanks and missing values ,imputing the missing values with the appropriate method(mean, median ,mode).
·Then I found the outliers in the dataset.
·After all these I used pivot tables and basic charts to visualise the data.
·Then I found top 10 correlation between the columns of repayer and defaulters

Tech-Stack Used
Microsoft® Word 2019 MSO (Version 2501 Build 16.0.18429.20132) 64-bit for data analysis, pivot table creation and chart generation.

Insights
FOR CURRENT APPLICATION
A.Identify Missing Data and Deal with it Appropriately:
·Initially there were 122 columns and 50000 rows including header row in loan application dataset.
·First we used count blank formula to count all the blank cells in column
·Then we converted it into percent to plot the bar graph by transposing it into new sheet.
·Then we filled null values of numerical column with median as data contains outliers therefore mean is inappropriate.
·For non numerical data we deleted whole column
·Columns removed are:

COMMONAREA_AVG
COMMONAREA_MODE
COMMONAREA_MEDI
NONLIVINGAPARTMENTS_AVG
NONLIVINGAPARTMENTS_MODE
NONLIVINGAPARTMENTS_MEDI
LIVINGAPARTMENTS_AVG
LIVINGAPARTMENTS_MODE
LIVINGAPARTMENTS_MEDI
FONDKAPREMONT_MODE
FLOORSMIN_AVG
FLOORSMIN_MODE
FLOORSMIN_MEDI
YEARS_BUILD_AVG
YEARS_BUILD_MODE
YEARS_BUILD_MEDI
OWN_CAR_AGE
LANDAREA_AVG
LANDAREA_MODE
LANDAREA_MEDI
BASEMENTAREA_AVG
BASEMENTAREA_MODE
BASEMENTAREA_MEDI
EXT_SOURCE_1
NONLIVINGAREA_AVG
NONLIVINGAREA_MODE
NONLIVINGAREA_MEDI
ELEVATORS_AVG
ELEVATORS_MODE
ELEVATORS_MEDI
WALLSMATERIAL_MODE
APARTMENTS_AVG
APARTMENTS_MODE
APARTMENTS_MEDI
ENTRANCES_AVG
ENTRANCES_MODE
ENTRANCES_MEDI
LIVINGAREA_AVG
LIVINGAREA_MODE
LIVINGAREA_MEDI
HOUSETYPE_MODE




B.Identify Outliers in the Dataset:

·Outliers are the datapoints which deviate signifiacantly from the rest of the datapoints.

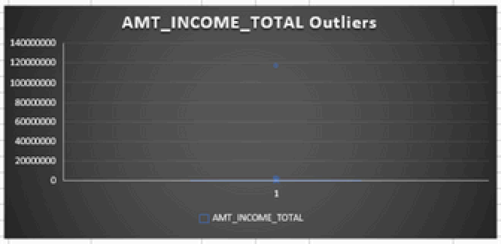
·

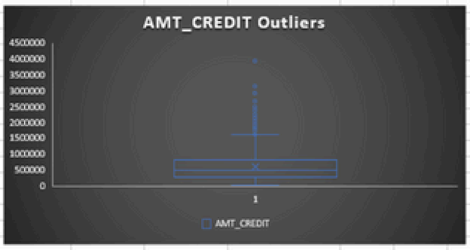
So to analyze the datapoints we have used box plot and scatter plot to find out the outliers.

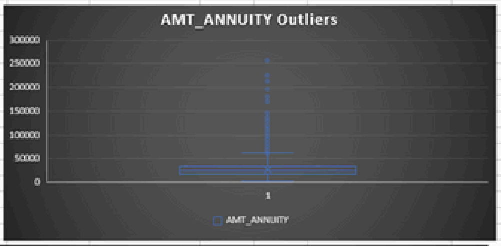
| | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | AMT_GOODS_PRICE |
|---|---|---|---|---|
| Mean | 170763 | 599708 | 27107 | 539002 |
| Median | 145800 | 514778 | 24939 | 450000 |
| Standard Deviation | 531833.55 | 402424.25 | 14563.2 | 369729.26 |
| IQR | 90000 | 538650 | 18140 | 441000 |
| Lower Bound | -22500 | -537975 | -10752.75 | -423000 |
| Upper Bound | 337500 | 1616625 | 61805.25 | 1341000 |

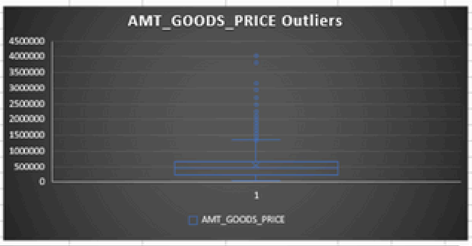
We also used statistical formula to find upper bound manually to highlight the values greater than the upper bound value

| AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | AMT_GOODS_PRICE |
|---|---|---|---|
| 202500 | 406597.5 | 24700.5 | 351000 |
| 270000 | 1293502.5 | 35698.5 | 1129500 |
| 67500 | 135000 | 6750 | 135000 |
| 135000 | 312682.5 | 29686.5 | 297000 |
| 121500 | 513000 | 21865.5 | 513000 |
| 99000 | 490495.5 | 27517.5 | 454500 |
| 171000 | 1560726 | 41301 | 1395000 |
| 360000 | 1530000 | 42075 | 1530000 |
| 112500 | 1019610 | 33826.5 | 913500 |
| 135000 | 405000 | 20250 | 405000 |
| 112500 | 652500 | 21177 | 652500 |
| 38419 | 148365 | 10678.5 | 135000 |
| 67500 | 80865 | 5881.5 | 67500 |
| 225000 | 918468 | 28966.5 | 697500 |
| 189000 | 773680.5 | 32778 | 679500 |
| 157500 | 299772 | 20160 | 247500 |
| 108000 | 509602.5 | 26149.5 | 387000 |
| 81000 | 270000 | 13500 | 270000 |
| 112500 | 157500 | 7875 | 157500 |
| 90000 | 544491 | 17563.5 | 454500 |
| 135000 | 427500 | 21375 | 427500 |
| 202500 | 1132573.5 | 37561.5 | 927000 |
| 450000 | 497520 | 32521.5 | 450000 |
| 83250 | 239850 | 23850 | 225000 |
| 135000 | 247500 | 12703.5 | 247500 |
| 90000 | 225000 | 11074.5 | 225000 |
| 112500 | 979992 | 27076.5 | 702000 |
| 112500 | 327024 | 23827.5 | 270000 |
| 270000 | 790830 | 57676.5 | 675000 |
| 90000 | 180000 | 9000 | 180000 |
| 292500 | 665892 | 24592.5 | 477000 |
| 112500 | 512064 | 25033.5 | 360000 |
| 90000 | 199008 | 20893.5 | 180000 |
| 360000 | 733315.5 | 39069 | 679500 |
| 135000 | 1125000 | 32895 | 1125000 |
| 112500 | 450000 | 44509.5 | 450000 |

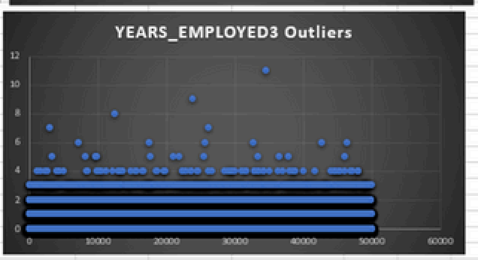


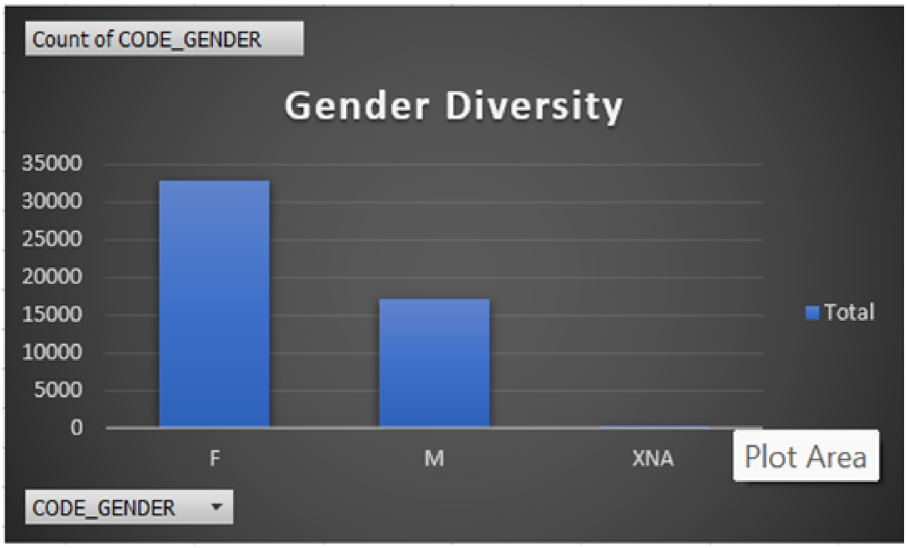





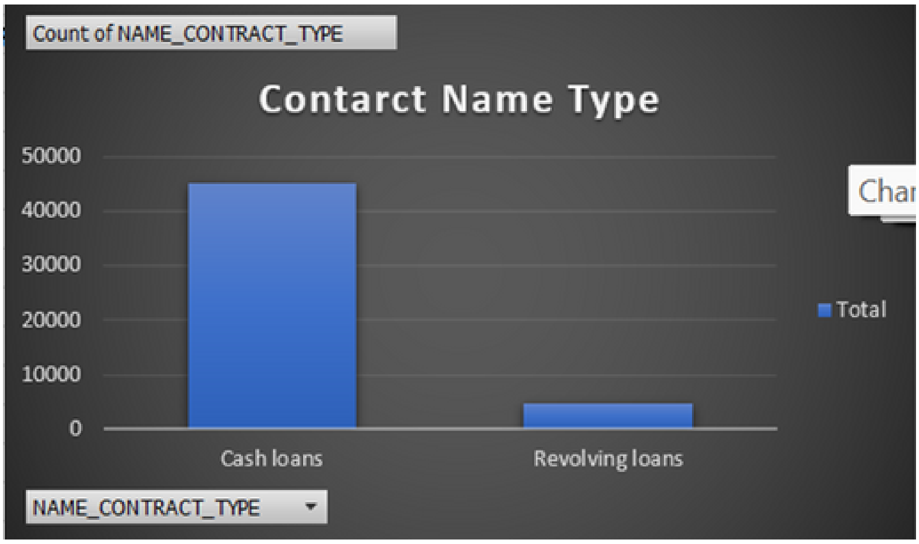


Insights
·AMT_TOTAL_INCOME
By the box plot we can understand that most of the values are below the 40,00,000.
The only value which seems unrealistic is 11,70,00,000. As it is outlier it affects the measure of central tendency.
·CNT_CHILDREN
In few cases there are errors in data entry as some have count upto 11 childrens which seems unrealistic in today's world.
·YEARS_EMPLOYED2
By observing the scatter plot we can analyse that some values are above 1000 years which is clearly not possible and is an outlier.
·AMT_GOOD_PRICES
Some the client are applying for loan with extremely high prices and not realistic. Therefore treated as outliers.
·AMT_ANNUITY
As there is very few chances of having annuity more than 200000 the points observed in box plot above the upper bound are treated as outliers.
·AMT_CREDIT
As there is very few chances that a client has this much credit more than 200000 the points above the upper bound are treated as outliers.

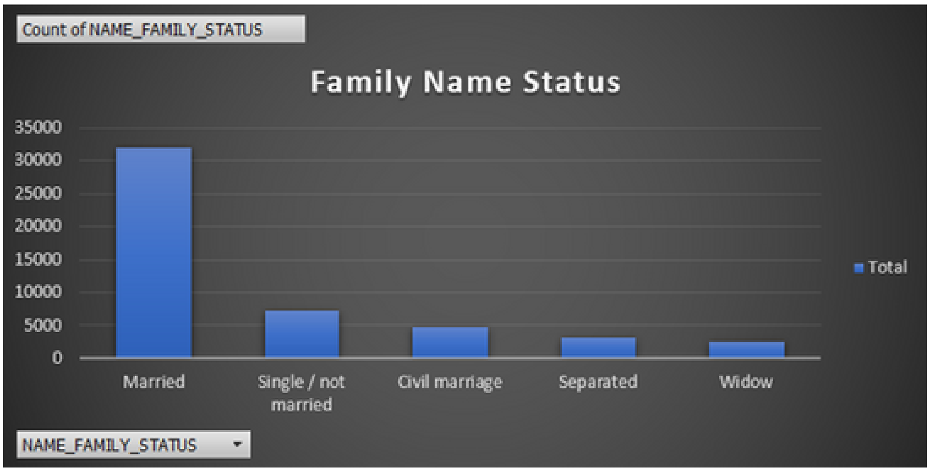
C. Analyze Data Imbalance:



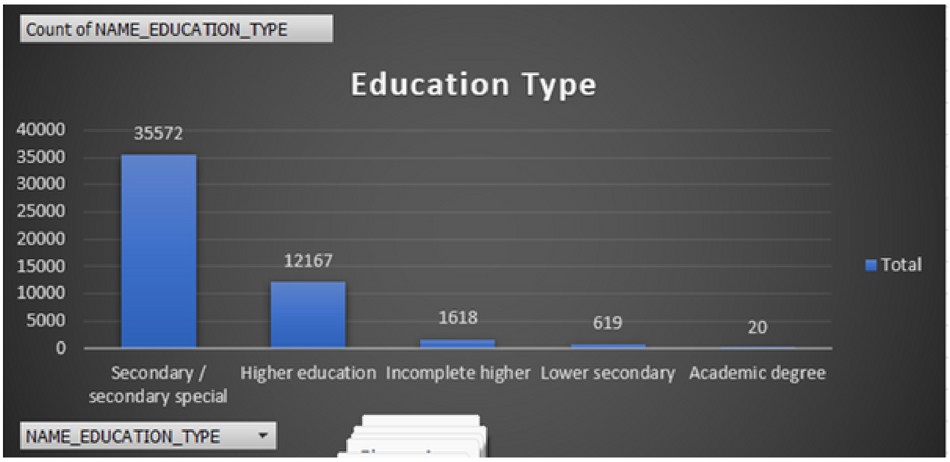

Gender Diversity:
The dataset displays that number of females is much higher than males and also the imbalanced distribution in number of xna.



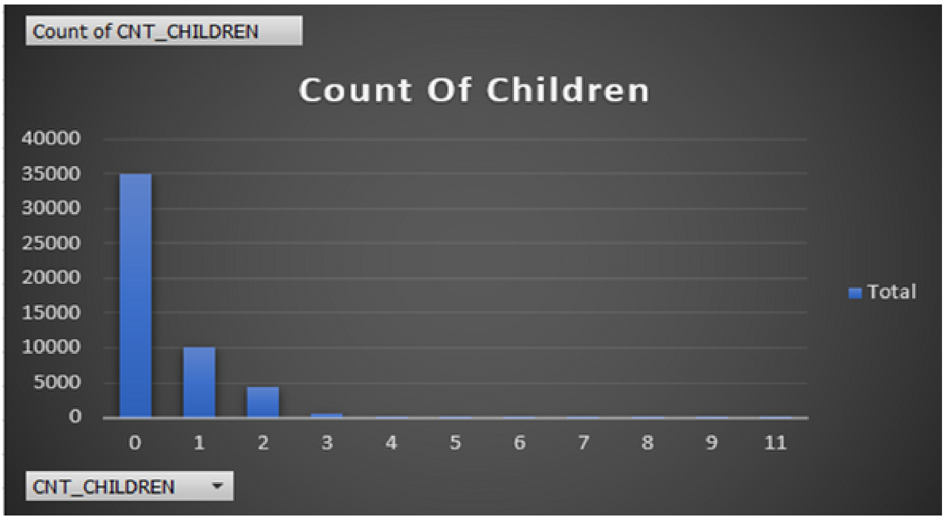

Contract name type:
The frequency of cash loans is much greater than the revolving loans. The revolving loan has small fraction of part out of the total number which is 10.4%.



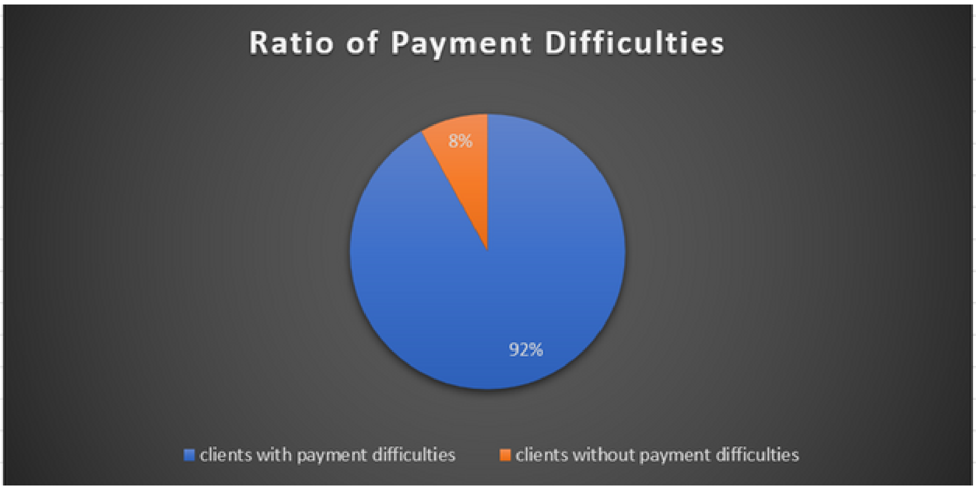

Family Name status:
Married people have taken most of the loans followed by Single/not married and civil marriage.




Education Type:
Secondary/secondary special education is of majority clients then followed by Higher education. Clients having academic degree has very small number.




Count of Children:
Clients with no children have higher frequency while other are less. It shows the imbalanced distribution in number of children feature.

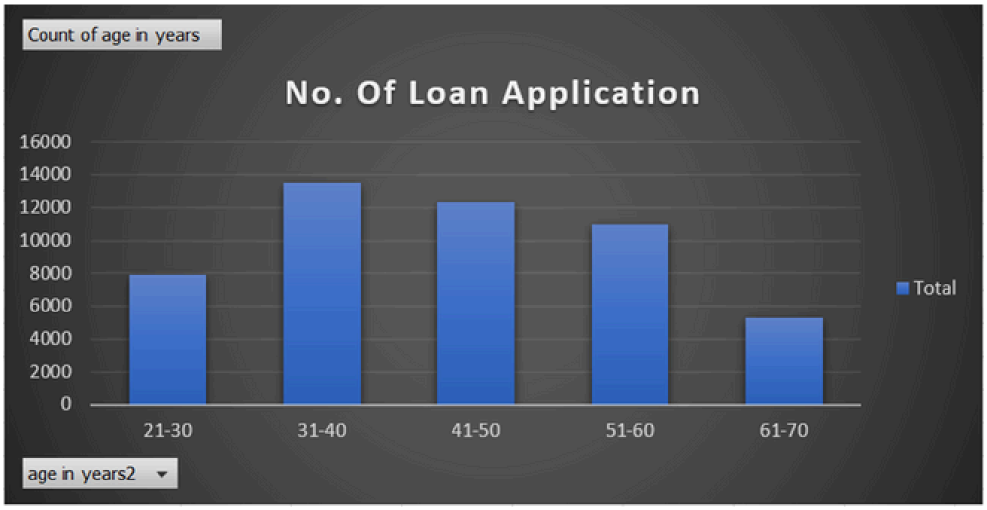



Ratio of Payment Difficulties:
We observe that most of the clients are facing difficulties in payment which 92% and only 8% are able to do payment without any difficulties. This shows that the data is imbalanced as the one class is under represented.

A.Perform Univariate, Segmented Univariate, and Bivariate Analysis:

Univariate:

| Row Labels | Count of age in years |
|---|---|
| 21-30 | 7928 |
| 31-40 | 13505 |
| 41-50 | 12327 |
| 51-60 | 10973 |
| 61-70 | 5263 |
| **Grand Total** | **49996** |




Insight:
Data shows that majority of loan applicants fall within 31-60 age group.

Bivariate:

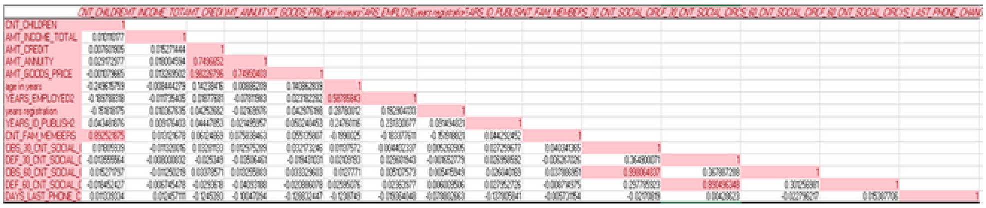
| Sum of TARGET | Column Labels | | | | | |
|---|---|---|---|---|---|---|
| **Row Labels** | **21-30** | **31-40** | **41-50** | **51-60** | **61-70** | **Grand Total** |
| 1 | 897 | 1284 | 924 | 654 | 267 | 4026 |
| **Grand Total** | **897** | **1284** | **924** | **654** | **267** | **4026** |


Insight:
The data shows that after age of 31 the number of clients with payment difficulties decreases as age increases. This gives us a insight that older clients may have more financial stability and experience in managing finances so they have less difficulties in payment of loan.

E.Identify Top Correlations for Different Scenarios:

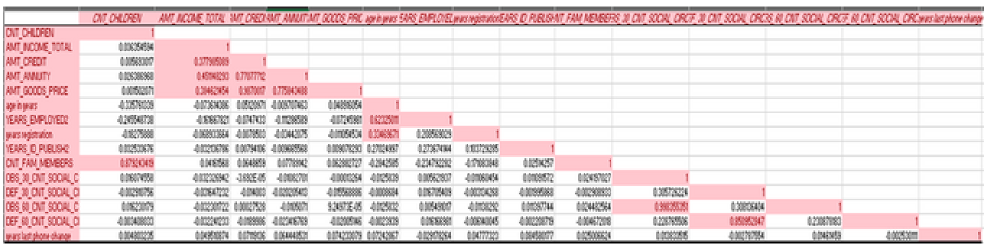
The top 10 correlation between different variables and client who are defaulter is:

| variable 1 | variable 2 | correlation |
|---|---|---|
| OBS_60_CNT_SOCIAL_CIRCLE | OBS_30_CNT_SOCIAL_CIRCLE | 0.998064837 |
| AMT_GOODS_PRICE | AMT_CREDIT | 0.982267963 |
| CNT_FAM_MEMBERS | CNT_CHILDREN | 0.892521875 |
| DEF_60_CNT_SOCIAL_CIRCLE | DEF_30_CNT_SOCIAL_CIRCLE | 0.890496348 |
| AMT_ANNUITY | AMT_CREDIT | 0.749665201 |
| AMT_GOODS_PRICE | AMT_ANNUITY | 0.74950403 |
| YEARS_EMPLOYED2 | age in years | 0.587858433 |
| OBS_60_CNT_SOCIAL_CIRCLE | DEF_30_CNT_SOCIAL_CIRCLE | 0.367887288 |
| DEF_30_CNT_SOCIAL_CIRCLE | OBS_30_CNT_SOCIAL_CIRCLE | 0.364900071 |
| DEF_60_CNT_SOCIAL_CIRCLE | OBS_60_CNT_SOCIAL_CIRCLE | 0.301256981 |



The top 10 correlation between different variable and clients who are repayer

| variable 1 | variable 2 | coreelation |
|---|---|---|
| OBS_60_CNT_SOCIAL_CIRCLE | OBS_30_CNT_SOCIAL_CIRCLE | 0.998355351 |
| AMT_GOODS_PRICE | AMT_INCOME_TOTAL | 0.987001704 |
| CNT_FAM_MEMBERS | CNT_CHILDREN | 0.879243419 |
| DEF_60_CNT_SOCIAL_CIRCLE | DEF_30_CNT_SOCIAL_CIRCLE | 0.850952847 |
| AMT_GOODS_PRICE | AMT_ANNUITY | 0.775843488 |
| AMT_ANNUITY | AMT_INCOME_TOTAL | 0.77077712 |
| YEARS_EMPLOYED2 | age in years | 0.623250115 |
| AMT_ANNUITY | AMT_INCOME_TOTAL | 0.451148293 |
| AMT_GOODS_PRICE | AMT_INCOME_TOTAL | 0.384621454 |
| AMT_CREDIT | AMT_INCOME_TOTAL | 0.377985089 |



Drive link:
https://docs.google.com/spreadsheets/d/14lAwEXxC09Dxz_GIuOaeyzTIn3rLk5YT/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true

FOR PREVIOUS APPLICATION
Similarly performing all the tasks for previous application dataset

Drive Link:
https://docs.google.com/spreadsheets/d/1uSWZS87SSBVZpexxIr2B6UEDpoiA7Iu_/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true

Conclusion
·The dataset was observed to be imbalanced with one class as being underrepresented.
·The major demographic age group for seeking loan were observed to be between 31-60.
·The age group of 31-60 where observed to have higher incomes which suggests the prime stage of earning in life.
·The clients having higher incomes generally did not face difficulties in payment of loan.
·It was observed cash loans were the most common loan type which gave us insight of liquidity needs among the clients.
·Clients with secondary/secondary special education were more likely to take loans which showed us correlation between education level and loan dependency.
·Married individuals are more likely to apply for loan which suggest greater financial obligations or goals.
·Largest occupation group of loan applicants is observed of labourers which tells financial need in this sector.
·There large amount of repeat clients suggesting customers returning more often for additional loans.
·Age group between 31-60 tend to have more credit score, further establishing this age group as financially stable and reliable.

# Project – 7 Analyzing the Impact of Car Features on Price and Profitability

Project Description
This project aims to analyse 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.

Approach
· First I have cleaned the data to ensure accurate results. I observed that columns having null values where Engine HP, Engine fuel type, Engine cylinders, Market Category and Number of doors.
· Engine fuel type had three blank cells so we filled it with 'regular unleaded'.
· Then we removed the null value rows from Engine HP, Engine cylinders and Number of doors.
· Then for better visualization and understanding we converted MSRP into $ currency.
· Afterwards using pivot table, regression analysis and different types of charts such as bar chart, scatter chart, bubble chart, stacked bar chart, line chart and combo chart we have explored the relationships between different features.
· This helped us to build an interactive dashboard which shows the relationship between car features, market categories, popularity, fuel efficiency, manufacturers and engine power.

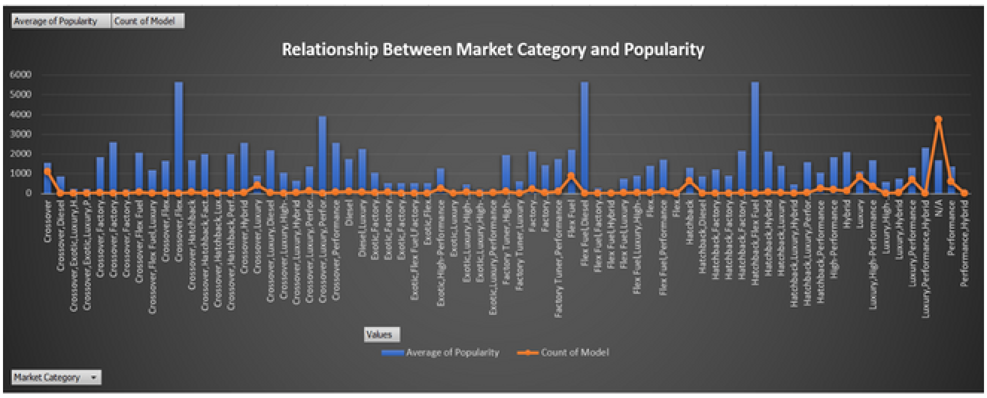
Tech-Stack Used
Microsoft® Word 2019 MSO (Version 2501 Build 16.0.18429.20132) 64-bit for data analysis, pivot table creation and chart generation.

Insights

Tasks: Analysis
Insight Required: How does the popularity of a car model vary across different market categories?
●Task 1.A: Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

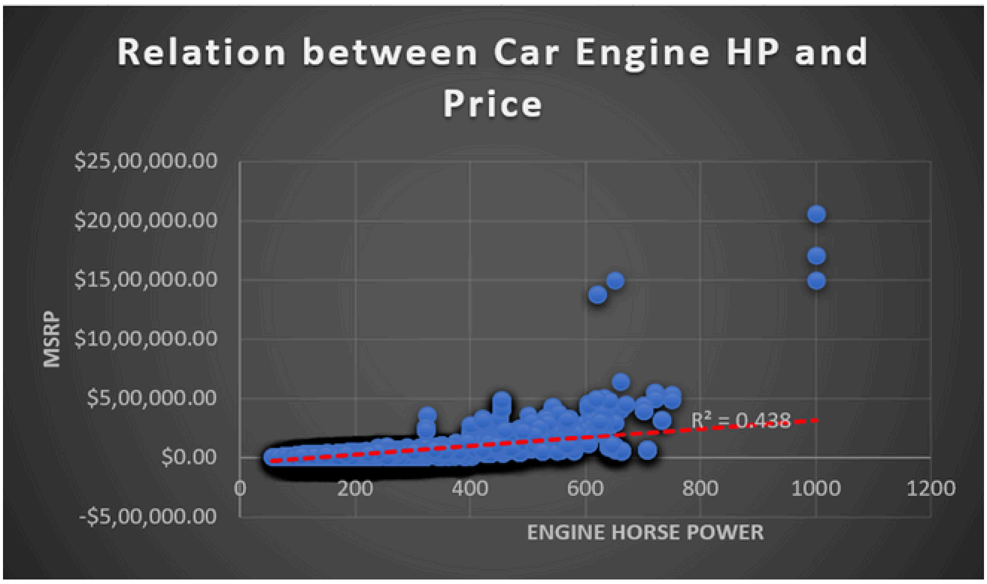
| Popularity of Car Model | | |
|---|---|---|
| Market Category | Average of Popularity | Count of Model |
| Crossover | 1529.030825 | 1103 |
| Crossover,Diesel | 873 | 7 |
| Crossover,Exotic,Luxury,High-Performance | 238 | 1 |
| Crossover,Exotic,Luxury,Performance | 238 | 1 |
| Crossover,Factory Tuner,Luxury,High-Performance | 1823.461538 | 26 |
| Crossover,Factory Tuner,Luxury,Performance | 2607.4 | 5 |
| Crossover,Factory Tuner,Performance | 210 | 4 |
| Crossover,Flex Fuel | 2073.75 | 64 |
| Crossover,Flex Fuel,Luxury | 1173.2 | 10 |
| Crossover,Flex Fuel,Luxury,Performance | 1624 | 6 |
| Crossover,Flex Fuel,Performance | 5657 | 6 |
| Crossover,Hatchback | 1675.694444 | 72 |
| Crossover,Hatchback,Factory Tuner,Performance | 2009 | 6 |
| Crossover,Hatchback,Luxury | 204 | 7 |
| Crossover,Hatchback,Performance | 2009 | 6 |
| Crossover,Hybrid | 2563.380952 | 42 |
| Crossover,Luxury | 884.5487805 | 410 |
| Crossover,Luxury,Diesel | 2195.848485 | 33 |
| Crossover,Luxury,High-Performance | 1037.222222 | 9 |
| Crossover,Luxury,Hybrid | 630.9166667 | 24 |
| Crossover,Luxury,Performance | 1344.849558 | 113 |
| Crossover,Luxury,Performance,Hybrid | 3916 | 2 |
| Crossover,Performance | 2585.956522 | 69 |
| Diesel | 1730.904762 | 84 |
| Diesel,Luxury | 2275 | 51 |
| Exotic,Factory Tuner,High-Performance | 1046.380952 | 21 |
| Exotic,Factory Tuner,Luxury,High-Performance | 517.5384615 | 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 | 1261.571429 | 252 |
| Exotic,Luxury | 112.6666667 | 12 |
| Exotic,Luxury,High-Performance | 467.0759494 | 79 |
| Exotic,Luxury,High-Performance,Hybrid | 204 | 1 |
| Exotic,Luxury,Performance | 217.0277778 | 36 |
| Factory Tuner,High-Performance | 1941.415094 | 106 |
| Factory Tuner,Luxury | 617 | 2 |
| Factory Tuner,Luxury,High-Performance | 2133.367442 | 215 |
| Factory Tuner,Luxury,Performance | 1413.419355 | 31 |
| Factory Tuner,Performance | 1733.101124 | 89 |
| Flex Fuel | 2217.302752 | 872 |
| Flex Fuel,Diesel | 5657 | 16 |
| Flex Fuel,Factory Tuner,Luxury,High-Performance | 258 | 1 |
| Flex Fuel,Hybrid | 155 | 2 |
| Flex Fuel,Luxury | 746.5384615 | 39 |
| Flex Fuel,Luxury,High-Performance | 878.9090909 | 33 |
| Flex Fuel,Luxury,Performance | 1380.071429 | 28 |
| Flex Fuel,Performance | 1702.358025 | 81 |
| Flex Fuel,Performance,Hybrid | 155 | 2 |
| Hatchback | 1292.998371 | 614 |
| Hatchback,Diesel | 873 | 14 |
| Hatchback,Factory Tuner,High-Performance | 1205.153846 | 13 |
| Hatchback,Factory Tuner,Luxury,Performance | 886.8888889 | 9 |
| Hatchback,Factory Tuner,Performance | 2159.045455 | 22 |
| Hatchback,Flex Fuel | 5657 | 7 |
| Hatchback,Hybrid | 2121.25 | 72 |
| Hatchback,Luxury | 1379.5 | 46 |
| Hatchback,Luxury,Hybrid | 454 | 3 |
| Hatchback,Luxury,Performance | 1566.131579 | 38 |
| Hatchback,Performance | 1039.646825 | 252 |
| High-Performance | 1821.447236 | 199 |
| Hybrid | 2105.569106 | 123 |
| Luxury | 1107.553467 | 851 |
| Luxury,High-Performance | 1668.017964 | 334 |
| Luxury,High-Performance,Hybrid | 568.8333333 | 12 |
| Luxury,Hybrid | 724.6875 | 48 |
| Luxury,Performance | 1292.615156 | 673 |
| Luxury,Performance,Hybrid | 2333.181818 | 11 |
| N/A | 1670.430984 | 3731 |
| Performance | 1371.080479 | 584 |
| Performance,Hybrid | 155 | 1 |

Task 1.B: Create a combo chart that visualizes the relationship between market category and popularity.




Insight:
The popularity of car model varies across different market categories based on various factors. Here the most popular market categories for car models are Crossover, flex fuel, diesel, hatchback, and performance.

Insight Required: What is the relationship between a car's engine power and its price?
●Task 2: Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.



Insight:
We can observe the positive correlation between engine power and price, which means that as engine power increases the price of car tends to increase as well. This is because higher engine power requires more advanced technology and components which increases the cost of production.

Insight Required: Which car features are most important in determining a car's price?
●Task 3: Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

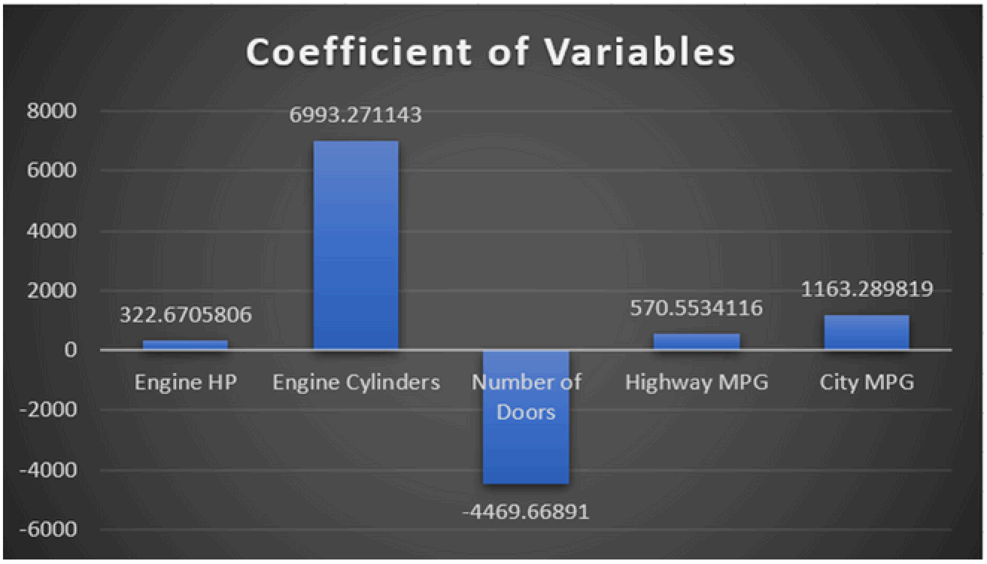
SUMMARY OUTPUT

| Regression Statistics | |
|---|---|
| Multiple R | 0.680713809 |
| R Square | 0.463371289 |
| Adjusted R Square | 0.463144077 |
| Standard Error | 44165.59385 |
| Observations | 11815 |

ANOVA

| | df | SS | MS | F | Significance F |
|---|---|---|---|---|---|
| Regression | 5 | 1.98901E+13 | 3.97802E+12 | 2039.38084 | 0 |
| Residual | 11809 | 2.30346E+13 | 1950599680 | | |
| Total | 11814 | 4.29247E+13 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|---|---|---|---|---|---|---|---|---|
| Intercept | -101612.1812 | 3683.853676 | -27.58312088 | 2.3253E-162 | -108833.1419 | -94391.2206 | -108833.1419 | -94391.2206 |
| Engine HP | 322.6705806 | 6.014842757 | 53.64572169 | 0 | 310.880497 | 334.4606642 | 310.880497 | 334.4606642 |
| Engine Cylinders | 6993.271143 | 439.5089745 | 15.91155482 | 2.02935E-56 | 6131.761081 | 7854.781204 | 6131.761081 | 7854.781204 |
| Number of Doors | -4469.66891 | 465.6339314 | -9.599104809 | 9.68047E-22 | -5382.388195 | -3556.949626 | -5382.388195 | -3556.949626 |
| Highway MPG | 570.5534116 | 105.7686551 | 5.394352522 | 7.0098E-08 | 363.2294072 | 777.8774161 | 363.2294072 | 777.8774161 |
| City MPG | 1163.289819 | 121.9795585 | 9.536760372 | 1.76081E-21 | 924.1897715 | 1402.389867 | 924.1897715 | 1402.389867 |

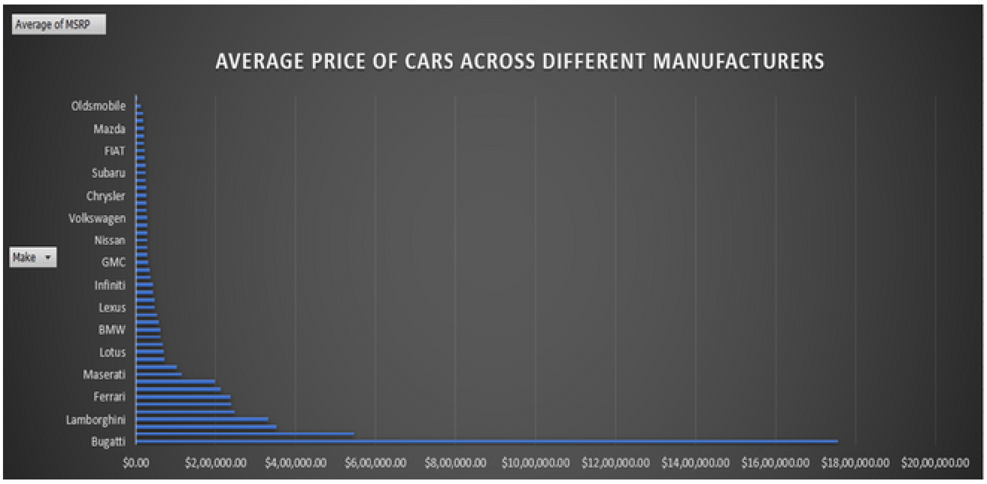


Insight:
We can observe that engine cylinders is the most important feature to determine the price of a car, then followed by City MPG and Highway MPG.

Insight Required: How does the average price of a car vary across different manufacturers?
●Task 4.A: Create a pivot table that shows the average price of cars for each manufacturer.

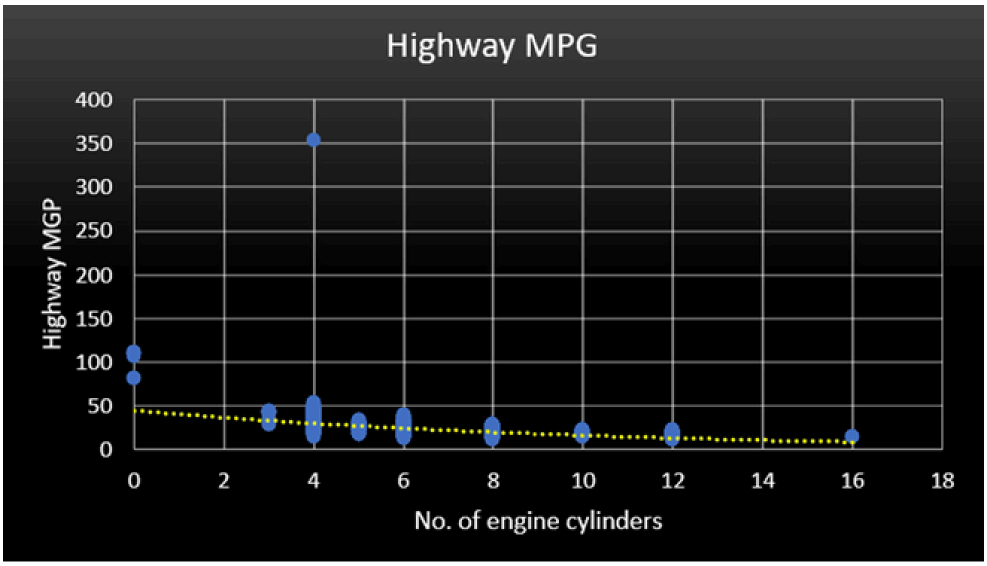
| Row Labels | Average of MSRP |
|---|---|
| Bugatti | $17,57,223.67 |
| Maybach | $5,46,221.88 |
| Rolls-Royce | $3,51,130.65 |
| Lamborghini | $3,31,567.31 |
| Bentley | $2,47,169.32 |
| McLaren | $2,39,805.00 |
| Ferrari | $2,37,383.82 |
| Spyker | $2,13,323.33 |
| Aston Martin | $1,97,910.38 |
| Maserati | $1,14,207.71 |
| Porsche | $1,01,622.40 |
| Mercedes-Benz | $71,537.81 |
| Lotus | $69,188.28 |
| Land Rover | $67,823.22 |
| Alfa Romeo | $61,600.00 |
| BMW | $61,546.76 |
| Cadillac | $56,231.32 |
| Audi | $53,452.11 |
| Lexus | $47,549.07 |
| Genesis | $46,616.67 |
| Lincoln | $42,494.37 |
| Infiniti | $42,394.21 |
| HUMMER | $36,464.41 |
| Acura | $34,887.59 |
| Acura | $34,887.59 |
| GMC | $30,493.30 |
| Toyota | $28,946.15 |
| Volvo | $28,541.16 |
| Nissan | $28,513.37 |
| Chevrolet | $28,273.36 |
| Buick | $28,206.61 |
| Volkswagen | $28,076.20 |
| Saab | $27,413.50 |
| Ford | $27,393.42 |
| Chrysler | $26,722.96 |
| Honda | $26,629.82 |
| Kia | $25,112.39 |
| Subaru | $24,827.50 |
| Hyundai | $24,597.04 |
| Dodge | $22,390.06 |
| FIAT | $22,206.02 |
| Mitsubishi | $21,215.47 |
| Scion | $19,932.50 |
| Mazda | $19,719.06 |
| Pontiac | $19,321.55 |
| Suzuki | $17,907.21 |
| Oldsmobile | $11,542.54 |
| Plymouth | $3,122.90 |

●Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.




Insight:
Average price of a car depends on the manufacturers as they vary significantly, as each manufacturer has its own pricing strategy, market target, product offerings and brand positioning. Which contributes in variations in average prices. Here we can observe that Bugatti has highest average price and Oldsmobile has the lowest average price.

Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?
●Task 5.A: Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.



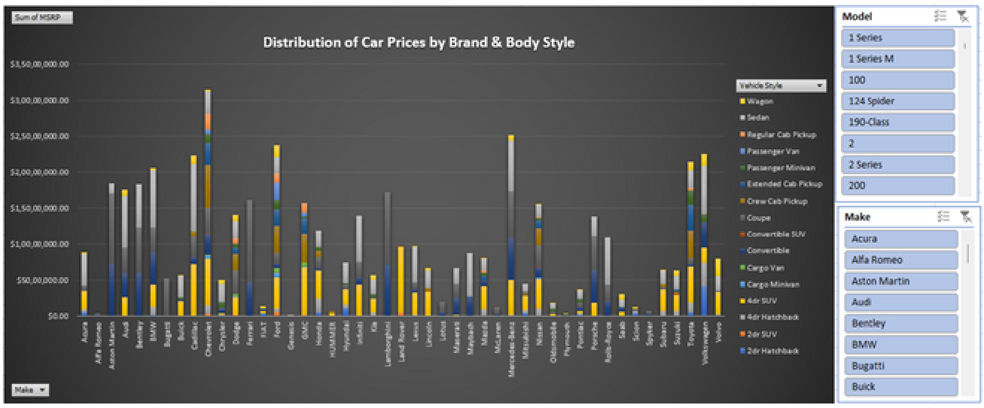
●Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

| Correlation | -0.62032 |
|---|---|

=CORREL(A2:A11816,B2:B11816)

Insight:
It is observed that Number of Engine Cylinders increases and the Highway MPG decreases. So it shows that both of them have negative relationship.
The correlation coefficient between them is -0.62032 which also tells they have negative correlation which means as one will increase the other will decrease and vice versa.

Building the Dashboard:
Task 1: How does the distribution of car prices vary by brand and body style?




| Sum of MSRP | Vehicle Style | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Make | 2dr Hatchback | 2dr SUV | 4dr Hatchback | 4dr SUV | Cargo Minivan | Cargo Van | Convertible | Convertible SUV | Coupe | Crew Cab Pickup | Extended Cab Pickup | Passenger Minivan | Passenger Van | Regular Cab Pickup | Sedan | Wagon | Grand Total |
| Acura | $4,80,917.00 | | $3,57,440.00 | $26,63,505.00 | | | | | $7,93,748.00 | | | | | | $42,94,702.00 | $2,01,360.00 | $87,91,672.00 |
| Alfa Romeo | | | | | | | $1,29,800.00 | | $1,78,200.00 | | | | | | | | $3,08,000.00 |
| Aston Martin | | | | | | | $73,21,655.00 | | $96,35,275.00 | | | | | | $14,48,735.00 | | $1,84,05,665.00 |
| Audi | $4,000.00 | | | $26,74,900.00 | | | $32,91,405.00 | | $35,56,290.00 | | | | | | $71,58,348.00 | $8,47,350.00 | $1,75,32,293.00 |
| Bentley | | | | | | | $60,12,870.00 | | $63,56,760.00 | | | | | | $59,20,900.00 | | $1,82,90,530.00 |
| BMW | $80,097.00 | | $11,44,950.00 | $31,60,950.00 | | | $45,02,671.00 | | $34,19,051.00 | | | | | | $79,89,300.00 | $2,59,600.00 | $2,05,56,619.00 |
| Bugatti | | | | | | | | | $52,71,671.00 | | | | | | | | $52,71,671.00 |
| Buick | | | | $21,41,770.00 | | | $1,79,325.00 | | $18,534.00 | | | $3,30,065.00 | | | $28,50,590.00 | $8,212.00 | $55,28,496.00 |
| Cadillac | | | | $71,82,555.00 | | | $9,85,607.00 | | $29,53,574.00 | $5,99,150.00 | | | | | $94,18,847.00 | $11,84,100.00 | $2,23,23,833.00 |
| Chevrolet | $8,000.00 | $2,13,310.00 | $12,09,735.00 | $65,69,568.00 | $4,20,150.00 | $78,688.00 | $29,53,245.00 | $1,06,300.00 | $35,04,525.00 | $59,27,617.00 | $31,17,951.00 | $11,78,515.00 | $6,07,670.00 | $22,60,032.00 | $30,68,812.00 | $3,00,675.00 | $3,15,24,793.00 |
| Chrysler | $98,805.00 | | | $2,50,545.00 | | | $6,30,105.00 | | $1,14,510.00 | | | $9,22,295.00 | | | $24,79,859.00 | $5,01,075.00 | $49,97,194.00 |
| Dodge | $48,000.00 | $44,000.00 | $18,000.00 | $25,72,405.00 | $60,520.00 | $3,38,497.00 | $12,000.00 | | $32,64,627.00 | $22,35,775.00 | $8,64,172.00 | $5,57,425.00 | $70,708.00 | $7,19,408.00 | $24,17,585.00 | $7,93,055.00 | $1,40,16,177.00 |
| Ferrari | | | | | | | $47,23,811.00 | | $1,14,18,289.00 | | | | | | | | $1,61,42,100.00 |
| FIAT | $3,25,315.00 | | | $3,69,305.00 | | | $3,27,965.00 | | | | | | | | | $2,87,570.00 | $13,10,155.00 |
| Ford | $36,000.00 | $4,79,873.00 | $4,80,155.00 | $43,70,871.00 | $6,80,770.00 | $5,66,151.00 | $7,30,007.00 | | $13,98,144.00 | $38,12,353.00 | $22,85,584.00 | $12,71,330.00 | $24,31,898.00 | $12,99,240.00 | $22,99,348.00 | $16,35,565.00 | $2,37,77,489.00 |
| Genesis | | | | | | | | | | | | | | | $1,39,850.00 | | $1,39,850.00 |
| GMC | | $1,44,319.00 | | $66,41,919.00 | $1,42,750.00 | $4,68,085.00 | | | | $40,62,482.00 | $21,83,866.00 | $1,50,630.00 | $6,03,670.00 | $13,06,328.00 | | | $1,57,04,049.00 |
| Honda | $4,13,200.00 | | $20,15,270.00 | $39,53,209.00 | | | $2,52,135.00 | | $15,88,705.00 | $7,87,720.00 | | $5,53,185.00 | | | $23,40,105.00 | | $1,19,03,529.00 |
| HUMMER | | | | $3,77,490.00 | | | | | | $2,42,405.00 | | | | | | | $6,19,895.00 |
| Hyundai | $10,38,050.00 | | $5,28,880.00 | $21,28,890.00 | | | | | $7,24,070.00 | | | $1,33,075.00 | | | $28,99,937.00 | | $74,52,902.00 |
| Infiniti | | | | $43,40,200.00 | | | $9,80,050.00 | | $21,75,750.00 | | | | | | $64,94,090.00 | | $1,39,90,090.00 |
| Kia | | | $4,06,960.00 | $20,49,645.00 | | | | | $1,42,630.00 | | | $4,94,650.00 | | | $19,80,360.00 | $6,01,155.00 | $56,75,400.00 |
| Lamborghini | | | | | | | $70,64,450.00 | | $1,01,77,050.00 | | | | | | | | $1,72,41,500.00 |
| Land Rover | | $4,76,394.00 | | $90,76,595.00 | | | | $1,45,731.00 | | | | | | | | | $96,98,720.00 |
| Lexus | | | $94,700.00 | $31,52,974.00 | | | $4,72,065.00 | | $10,16,472.00 | | | | | | $48,37,596.00 | $31,105.00 | $96,04,912.00 |
| Lincoln | | | | $34,22,570.00 | | | | | $25,342.00 | $4,53,260.00 | | | | | $24,58,245.00 | $2,69,705.00 | $66,29,122.00 |
| Lotus | | | | | | | $4,13,260.00 | | $15,93,200.00 | | | | | | | | $20,06,460.00 |
| Maserati | | | | $1,55,000.00 | | | $23,42,963.00 | | $19,72,284.00 | | | | | | $21,53,800.00 | | $66,24,047.00 |
| Maybach | | | | | | | $27,62,750.00 | | | | | | | | $59,76,800.00 | | $87,39,550.00 |
| Mazda | $22,000.00 | $24,000.00 | $8,53,180.00 | $32,22,525.00 | | | $8,70,505.00 | | $14,000.00 | | $5,80,033.00 | $4,43,130.00 | | $2,65,486.00 | $16,18,571.00 | $33,350.00 | $79,46,780.00 |
| McLaren | | | | | | | $2,80,225.00 | | $9,18,800.00 | | | | | | | | $11,99,025.00 |

Result:

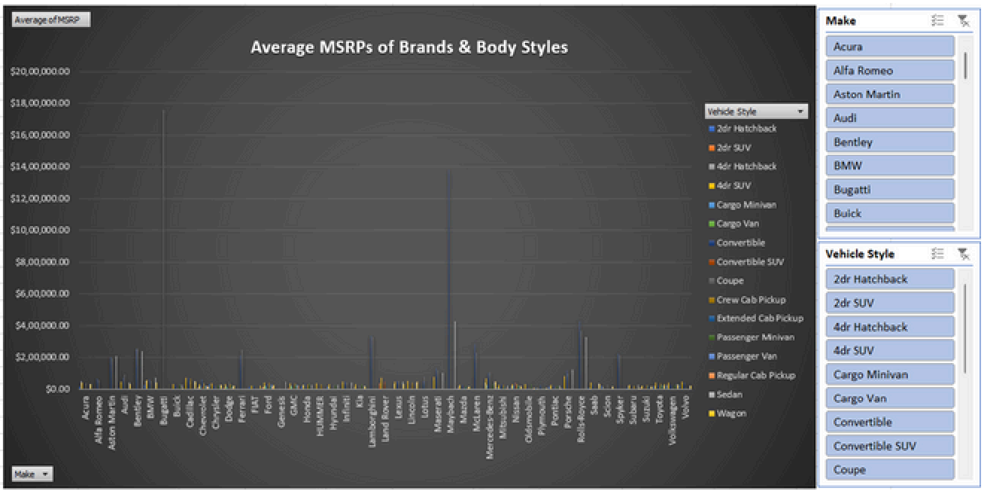
Chevrolet has the highest price distribution by body style. The price distribution of automobile by brand and body type depends on market trends such as:

· Pricing differs on the basis of target audience and brand positioning such as high end brands have premium feature and cutting edge technology which makes it higher at price point while cost effective brands focus on affordability.

· Pricing is also influenced and affected by the branding. The well known brands with reputation has quality products which are dependable due to which cost is higher.

· Body type of vehicle also influence the price point as the increased size and perceived prestige influence the prices at higher end.

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

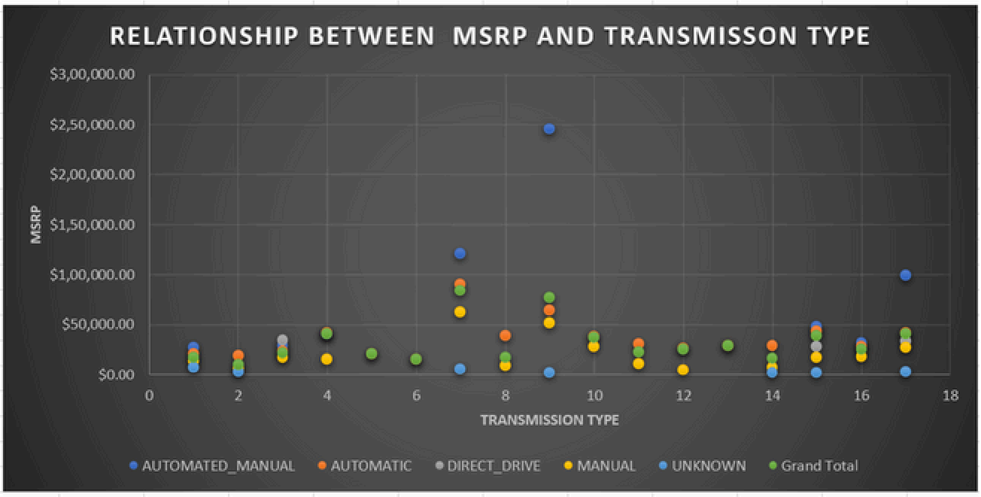



| Average of MSRP / Make | 2dr Hatchback | 2dr SUV | 4dr Hatchback | 4dr SUV | Cargo Minivan | Cargo Van | Convertible | Convertible SUV | Coupe | Crew Cab Pickup | Extended Cab Pickup | Passenger Minivan | Passenger Van | Regular Cab Pickup | Sedan | Wagon | Grand Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Acura | $17,175.61 | | $51,062.86 | $42,959.76 | | | | | $39,687.40 | | | | | | $33,292.26 | $33,560.00 | $34,887.59 |
| Alfa Romeo | | | | | | | $64,900.00 | | $59,400.00 | | | | | | | | $61,600.00 |
| Aston Martin | | | | | | | $2,03,379.31 | | $1,92,705.50 | | | | | | $2,06,962.14 | | $1,97,910.38 |
| Audi | $2,000.00 | | | $48,634.55 | | | $70,029.89 | | $93,586.58 | | | | | | $44,461.79 | $33,894.00 | $53,452.11 |
| Bentley | | | | | | | $2,50,536.25 | | $2,54,270.40 | | | | | | $2,36,836.00 | | $2,47,169.32 |
| BMW | $26,699.00 | | $54,521.43 | $58,536.11 | | | $63,417.90 | | $51,803.80 | | | | | | $70,701.77 | $43,266.67 | $61,546.76 |
| Bugatti | | | | | | | | | $17,57,223.67 | | | | | | | | $17,57,223.67 |
| Buick | | | | $33,996.35 | | | $25,617.86 | | $2,059.33 | | | $30,005.91 | | | $27,946.96 | $2,053.00 | $28,206.61 |
| Cadillac | | | | $72,551.06 | | | $70,400.50 | | $45,439.60 | $66,572.22 | | | | | $50,912.69 | $47,364.00 | $56,231.32 |
| Chevrolet | $2,000.00 | $8,887.92 | $18,329.32 | $32,046.67 | $20,007.14 | $7,153.45 | $62,835.00 | $17,716.67 | $38,939.17 | $39,255.74 | $24,170.16 | $24,552.40 | $24,306.80 | $19,824.84 | $19,798.79 | $15,825.00 | $28,273.36 |
| Chrysler | $32,935.00 | | | $35,792.14 | | | $24,234.81 | | $19,085.00 | | | $29,751.45 | | | $26,103.78 | $26,372.37 | $26,722.96 |
| Dodge | $2,000.00 | $2,000.00 | $2,000.00 | $30,992.83 | $20,173.33 | $12,536.93 | $2,000.00 | | $45,980.66 | $31,052.43 | $13,938.26 | $25,337.50 | $14,141.60 | $9,342.96 | $21,780.05 | $24,782.97 | $22,390.06 |
| Ferrari | | | | | | | $2,14,718.68 | | $2,48,223.67 | | | | | | | | $2,37,383.82 |
| FIAT | $19,136.18 | | | $24,620.33 | | | $23,426.07 | | | | | | | | | $22,120.77 | $22,206.02 |
| Ford | $2,000.00 | $13,710.66 | $18,467.50 | $42,027.61 | $21,274.06 | $17,698.47 | $34,762.24 | | $34,101.07 | $41,438.62 | $23,808.17 | $23,115.09 | $32,425.31 | $17,797.81 | $21,290.26 | $27,259.42 | $27,393.42 |
| Genesis | | | | | | | | | | | | | | | $46,616.67 | | $46,616.67 |
| GMC | | $5,550.73 | | $36,695.69 | $23,791.67 | $18,723.40 | | | | $39,062.33 | $26,632.51 | $25,105.00 | $26,246.52 | $21,069.81 | | | $30,493.30 |
| Honda | $17,216.67 | | $25,836.79 | $28,855.54 | | | $36,019.29 | Chart Area | 3.08 | $34,248.70 | | $36,879.00 | | | $26,001.17 | | $26,629.82 |
| HUMMER | | | | $37,749.00 | | | | | | $34,629.29 | | | | | | | $36,464.41 |
| Hyundai | $18,536.61 | | $17,629.33 | $30,412.71 | | | | | $20,687.71 | | | $26,615.00 | | | $27,102.21 | | $24,597.04 |
| Infiniti | | | | $45,686.32 | | | $46,669.05 | | $40,291.67 | | | | | | $40,588.06 | | $42,394.21 |
| Kia | | | $19,379.05 | $31,533.00 | | | | | $20,375.71 | | | $32,976.67 | | | $23,298.35 | $18,216.82 | $25,112.39 |
| Lamborghini | | | | | | | $3,36,402.38 | | $3,28,291.94 | | | | | | | | $3,31,567.31 |
| Land Rover | | $39,699.50 | | $70,910.90 | | | | $48,577.00 | | | | | | | | | $67,823.22 |
| Lexus | | | $31,566.67 | $45,042.49 | | | $52,451.67 | | $50,823.60 | | | | | | $48,864.61 | $31,105.00 | $47,549.07 |
| Lincoln | | | | $50,331.91 | | | | | $2,111.83 | $41,205.45 | | | | | $41,665.17 | $44,950.83 | $42,494.37 |
| Lotus | | | | | | | $51,657.50 | | $75,866.67 | | | | | | | | $69,188.28 |
| Maserati | | | | $77,500.00 | | | $1,30,164.61 | | $1,16,016.71 | | | | | | $1,02,561.90 | | $1,14,207.71 |
| Maybach | | | | | | | $13,81,375.00 | | | | | | | | $4,26,914.29 | | $5,46,221.88 |
| Mazda | $2,000.00 | $2,000.00 | $20,809.27 | $27,080.04 | | | $28,080.81 | | $2,000.00 | | $11,600.66 | $23,322.63 | | $9,154.69 | $19,738.67 | $16,675.00 | $19,719.06 |
| McLaren | | | | | | | $2,80,225.00 | | $2,29,700.00 | | | | | | | | $2,39,805.00 |
| Mercedes-Benz | | | $40,933.33 | $68,400.14 | $28,950.00 | | $1,04,617.53 | | $1,09,713.68 | | | $32,500.00 | | | $49,168.35 | $44,996.18 | $71,537.81 |
| Mitsubishi | $13,162.27 | | $32,101.79 | $26,158.29 | $2,000.00 | | $29,984.71 | | | $26,690.00 | $19,194.29 | $2,000.00 | | $2,000.00 | $24,058.25 | | $21,215.47 |

Result:

- Bugatti has highest average MSRP which suggests it as luxury brand. This luxury brands are known for its high end product with cutting edge technology.
- On the other hand Plymouth has lowest MSRP which tell it as a budget brand that offers reasonable prices depending on the model and body type.
- The tendencies of brand may be seen as which cars are frequently bought according to their model, material used, engine type and technology used.
- The mainstream companies provide variety of compact and midsize sedans and hatchbacks. This all cars seems to be more economical solution for daily use.

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



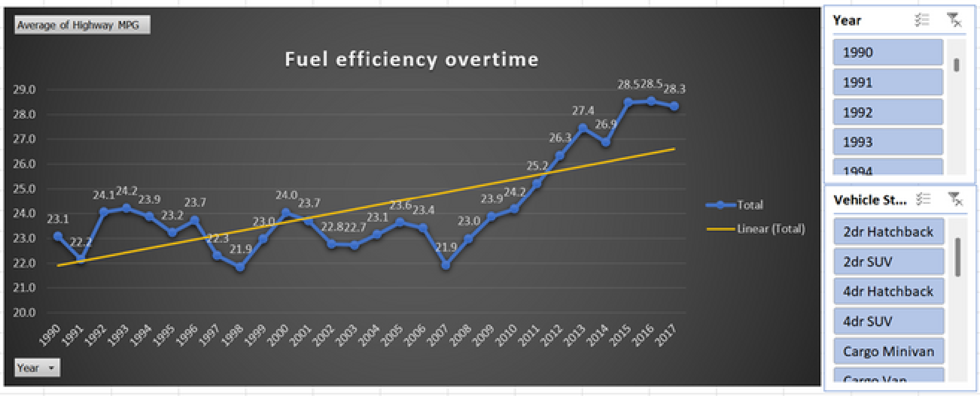

| Average of MSRP | Transmission Type | | | | | |
|---|---|---|---|---|---|---|
| Vehicle Style | AUTOMATED_MANUAL | AUTOMATIC | DIRECT_DRIVE | MANUAL | UNKNOWN | Grand Total |
| 2dr Hatchback | $27,180.96 | $20,926.46 | | $13,353.66 | $7,361.50 | $16,778.65 |
| 2dr SUV | | $18,615.20 | | $6,303.81 | $2,371.00 | $10,115.19 |
| 4dr Hatchback | $29,249.07 | $23,833.68 | $34,511.92 | $17,594.41 | | $22,086.30 |
| 4dr SUV | $40,451.15 | $41,555.19 | | $15,426.46 | | $40,426.82 |
| Cargo Minivan | | $20,910.86 | | | | $20,910.86 |
| Cargo Van | | $15,280.22 | | | | $15,280.22 |
| Convertible | $1,21,256.64 | $90,637.39 | | $62,357.76 | $5,783.50 | $84,224.28 |
| Convertible SUV | | $38,925.50 | | $9,233.14 | | $17,424.14 |
| Coupe | $2,45,588.36 | $63,852.01 | | $51,070.48 | $2,000.00 | $76,900.71 |
| Crew Cab Pickup | | $37,744.07 | | $28,360.53 | | $37,220.47 |
| Extended Cab Pickup | | $30,637.35 | | $10,884.19 | | $22,488.78 |
| Passenger Minivan | | $26,392.00 | | $4,405.33 | | $25,591.51 |
| Passenger Van | | $29,015.20 | | | | $29,015.20 |
| Regular Cab Pickup | | $28,536.82 | | $7,557.77 | $2,000.00 | $15,953.71 |
| Sedan | $47,498.71 | $43,760.61 | $27,822.50 | $17,119.23 | $2,000.00 | $38,969.06 |
| Wagon | $31,985.28 | $27,613.19 | | $17,844.14 | | $25,483.90 |
| **Grand Total** | **$99,195.58** | **$41,129.06** | **$33,620.00** | **$26,671.40** | **$3,040.74** | **$40,554.37** |

Result:

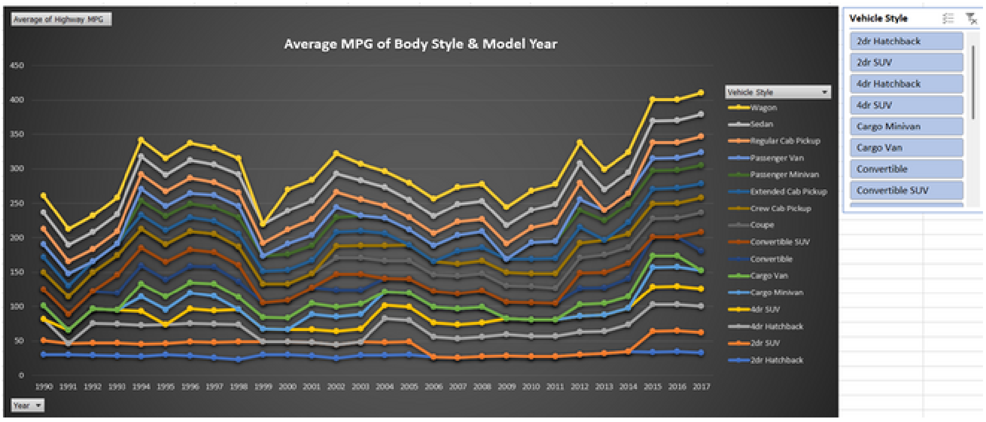
- The most expensive transmission is automated manual and also the automatic is one of the most popular transmission.
- The transmission type of a vehicle can vary also on body type of the vehicle such as for sedan cars they have higher trim levels and automatic transmission levels which generally offers comfortable driving options.
- Which suggests us that sedan having automatic transmission have expensive MSRP than equivalent model with manual transmissions.
- Generally the MSRP have an important feature which is transmission type. It varies according to the body style of a vehicle. As automated transmission provides more comfort and ease to use it has higher price than manual transmission.

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

| Row Labels | Average of Highway MPG |
|---|---|
| 1990 | 23.1 |
| 1991 | 22.2 |
| 1992 | 24.1 |
| 1993 | 24.2 |
| 1994 | 23.9 |
| 1995 | 23.2 |
| 1996 | 23.7 |
| 1997 | 22.3 |
| 1998 | 21.9 |
| 1999 | 23.0 |
| 2000 | 24.0 |
| 2001 | 23.7 |
| 2002 | 22.8 |
| 2003 | 22.7 |
| 2004 | 23.1 |
| 2005 | 23.6 |
| 2006 | 23.4 |
| 2007 | 21.9 |
| 2008 | 23.0 |
| 2009 | 23.9 |
| 2010 | 24.2 |
| 2011 | 25.2 |
| 2012 | 26.3 |
| 2013 | 27.4 |
| 2014 | 26.9 |
| 2015 | 28.5 |
| 2016 | 28.5 |
| 2017 | 28.3 |
| **Grand Total** | **26.3** |




| AVERAGE HIGHWAY MPG FOR EACH COMBNATION OF BODY STYLE CORRESPONDING TO YEAR | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average of Highway MPG | Vehicle Style | | | | | | | | | | | | | | | | |
| Year | 2dr Hatch | 2dr SUV | 4dr Hatch | 4dr SUV | Cargo Min | Cargo Van | Convertibl | Convertibl | Coupe | Crew Cab | Extended | Passenger | Passenger | Regular Ca | Sedan | Wagon | Grand Tota |
| 1990 | 30.4 | 20.0 | 31.0 | | 20.0 | | 23.5 | | 24.5 | | 22.0 | 18.9 | | 22.2 | 24.0 | 24.1 | 23.1 |
| 1991 | 30.1 | 16.3 | | 19.3 | | | 22.6 | | 26.2 | | 15.8 | 18.0 | | 17.0 | 24.2 | 22.6 | 22.2 |
| 1992 | 29.7 | 17.5 | 28.4 | 21.3 | | | 25.5 | | 27.3 | | 15.6 | | | 17.9 | 24.5 | 24.3 | 24.1 |
| 1993 | 28.5 | 18.5 | 27.3 | 21.0 | | | 24.5 | 26.0 | 28.5 | | 16.7 | | | 17.6 | 25.3 | 24.5 | 24.2 |
| 1994 | 27.4 | 18.4 | 27.1 | 20.0 | 21.0 | 19.3 | 26.0 | 26.0 | 27.5 | | 20.3 | 21.0 | 16.4 | 21.7 | 25.2 | 23.8 | 23.9 |
| 1995 | 30.1 | 16.0 | 27.7 | | 21.5 | 19.0 | 24.5 | 26.0 | 25.8 | | 20.0 | 20.1 | 15.0 | 21.2 | 24.1 | 24.1 | 23.2 |
| 1996 | 29.0 | 20.0 | 26.1 | 21.6 | 23.0 | 14.6 | 23.8 | 24.0 | 26.7 | | 20.0 | 20.8 | 15.0 | 22.2 | 25.7 | 24.7 | 23.7 |
| 1997 | 26.1 | 22.0 | 26.5 | 19.7 | 21.0 | 17.1 | 25.3 | 20.7 | 27.2 | | 18.4 | 20.6 | 17.0 | 18.8 | 25.3 | 24.4 | 22.3 |
| 1998 | 23.2 | 26.0 | 24.5 | 22.1 | | 17.2 | 23.7 | 24.0 | 26.3 | | 18.6 | 23.4 | 17.0 | 19.2 | 27.1 | 23.0 | 21.9 |
| 1999 | 30.3 | 18.8 | | 18.3 | | 16.7 | 21.5 | | 27.6 | | 18.4 | 22.3 | | 18.4 | 27.4 | | 23.0 |
| 2000 | 30.4 | 18.8 | | 17.7 | | 16.4 | 25.3 | | 24.2 | | 20.5 | 23.2 | 14.5 | 20.8 | 26.8 | 31.0 | 24.0 |
| 2001 | 29.0 | 18.7 | | 18.7 | 22.0 | 15.8 | 23.4 | | 20.3 | | 19.0 | 21.2 | 15.0 | 23.0 | 27.4 | 30.6 | 23.7 |
| 2002 | 25.3 | 19.0 | | 19.8 | 21.0 | 14.6 | 24.1 | 23.3 | 23.6 | 17.0 | 20.2 | 21.7 | 15.0 | 22.1 | 26.1 | 28.9 | 22.8 |
| 2003 | 29.8 | 18.8 | | 19.2 | 20.7 | 15.0 | 20.2 | 23.4 | 23.9 | 18.0 | 20.8 | 22.3 | | 24.1 | 27.1 | 24.0 | 22.7 |
| 2004 | 29.7 | 18.8 | 34.0 | 19.0 | 19.6 | | 20.1 | | 25.3 | 22.0 | 17.8 | 22.2 | | 18.5 | 26.4 | 22.8 | 23.1 |
| 2005 | 30.3 | 18.7 | 30.6 | 19.3 | 20.7 | | 20.7 | | 26.0 | 23.0 | | 22.0 | | 18.0 | 25.8 | 24.3 | 23.6 |
| 2006 | 27.3 | | 28.8 | 20.4 | 23.0 | | 22.9 | | 24.3 | 19.4 | | 22.5 | | 18.0 | 24.8 | 25.0 | 23.4 |
| 2007 | 25.5 | | 27.5 | 20.5 | 22.7 | | 22.8 | | 25.2 | 18.0 | 18.4 | 22.8 | | 19.6 | 25.3 | 24.8 | 21.9 |
| 2008 | 27.3 | | 28.5 | 20.8 | 23.0 | | 23.5 | | 24.8 | 18.5 | 19.2 | 23.0 | | 18.0 | 26.7 | 24.7 | 23.0 |
| 2009 | 29.0 | | 30.8 | 22.6 | | | 23.8 | | 24.0 | 19.1 | 19.9 | | | 21.9 | 26.6 | 26.8 | 23.9 |
| 2010 | 27.8 | | 29.5 | 23.2 | | | 24.6 | | 23.8 | 18.9 | 20.8 | 24.2 | | 21.0 | 26.1 | 28.5 | 24.2 |
| 2011 | 27.8 | | 28.9 | 23.6 | | | 23.9 | | 22.7 | 21.1 | 21.9 | 25.0 | | 27.0 | 27.0 | 28.7 | 25.2 |
| 2012 | 30.7 | | 32.7 | 22.7 | | 16.7 | 23.6 | 22.0 | 22.1 | 21.4 | 23.1 | 25.0 | 15.3 | 24.1 | 27.9 | 30.6 | 26.3 |
| 2013 | 31.9 | | 32.3 | 23.6 | | 16.7 | 23.2 | 22.0 | 25.5 | 21.3 | | 28.0 | 15.3 | | 29.8 | 29.6 | 27.4 |
| 2014 | 34.7 | | 38.8 | 23.8 | | 16.9 | 26.6 | 22.0 | 23.2 | 18.9 | 17.4 | 26.0 | 16.4 | | 29.9 | 29.4 | 26.9 |
| 2015 | 34.3 | 30.0 | 38.5 | 25.8 | 28.0 | 17.0 | 27.6 | | 26.2 | 22.0 | 21.7 | 25.8 | 18.1 | 22.7 | 32.2 | 30.9 | 28.5 |
| 2016 | 34.5 | 30.0 | 38.8 | 26.2 | 27.7 | 16.0 | 27.9 | | 27.1 | 22.3 | 21.8 | 25.6 | 17.7 | 22.5 | 32.1 | 29.9 | 28.5 |
| 2017 | 33.1 | 29.0 | 38.0 | 25.7 | 26.7 | | 27.8 | 28.0 | 27.7 | 21.8 | 21.0 | 26.1 | 19.0 | 22.5 | 32.6 | 30.9 | 28.3 |
| Grand Total | 30.9 | 19.1 | 35.3 | 24.4 | 24.5 | 16.6 | 25.8 | 23.7 | 25.8 | 21.1 | 20.1 | 23.6 | 17.2 | 20.6 | 29.8 | 27.8 | 26.3 |



Result:

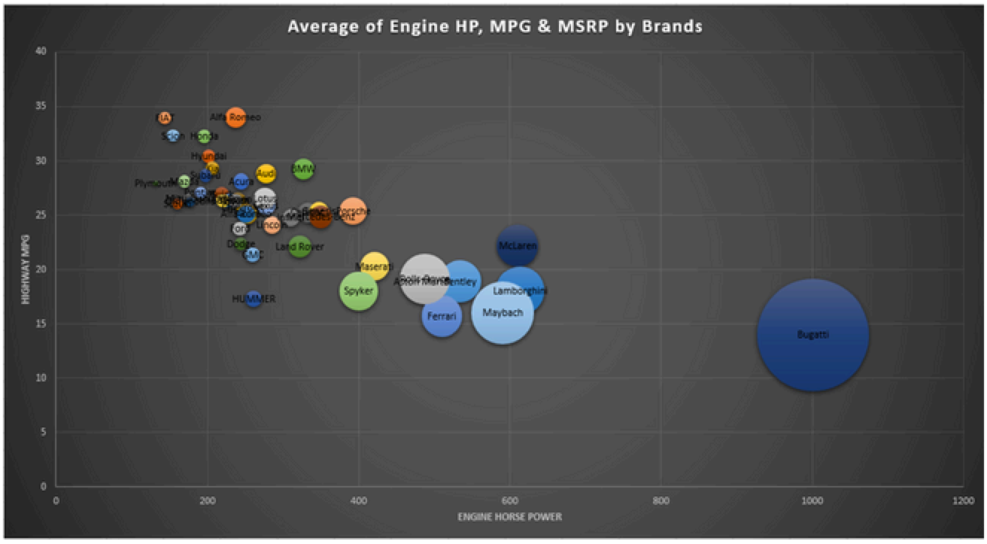
· Overall fuel efficiency has increased after 2007 at slower rate.

· Car’s fuel efficiency range as per the body type and model years. It can be seen that different body style have impact on fuel efficiency.

· As most efficient engines have been developed by automakers which later improves fuel economy.

· As the year passes the manufacturer improves their design and use of technology so it is observed that there is seen the increase in fuel economy after 2007.

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



| Row Labels | Average of Engine HP | Average of Highway MPG | Average of MSRP |
|---|---|---|---|
| Acura | 245 | 28 | $34,888 |
| Alfa Romeo | 237 | 34 | $61,600 |
| Aston Martin | 484 | 19 | $1,97,910 |
| Audi | 278 | 29 | $53,452 |
| Bentley | 534 | 19 | $2,47,169 |
| BMW | 327 | 29 | $61,547 |
| Bugatti | 1001 | 14 | $17,57,224 |
| Buick | 219 | 27 | $28,207 |
| Cadillac | 332 | 25 | $56,231 |
| Chevrolet | 247 | 26 | $28,273 |
| Chrysler | 229 | 26 | $26,723 |
| Dodge | 244 | 22 | $22,390 |
| Ferrari | 510 | 16 | $2,37,384 |
| FIAT | 144 | 34 | $22,206 |
| Ford | 243 | 24 | $27,393 |
| Genesis | 347 | 25 | $46,617 |
| GMC | 260 | 21 | $30,493 |
| Honda | 196 | 32 | $26,630 |
| HUMMER | 261 | 17 | $36,464 |
| Hyundai | 202 | 30 | $24,597 |
| Infiniti | 310 | 25 | $42,394 |
| Kia | 207 | 29 | $25,112 |
| Lamborghini | 614 | 18 | $3,31,567 |
| Land Rover | 322 | 22 | $67,823 |
| Lexus | 277 | 26 | $47,549 |
| Lincoln | 285 | 24 | $42,494 |
| Lotus | 276 | 27 | $69,188 |
| Maserati | 421 | 20 | $1,14,208 |
| Maybach | 591 | 16 | $5,46,222 |
| Maybach | 591 | 16 | $5,46,22 |
| Mazda | 169 | 28 | $19,71 |
| McLaren | 610 | 22 | $2,39,80 |
| Mercedes-Benz | 350 | 25 | $71,53 |
| Mitsubishi | 174 | 27 | $21,21 |
| Nissan | 240 | 26 | $28,51 |
| Oldsmobile | 177 | 26 | $11,54 |
| Plymouth | 132 | 28 | $3,12 |
| Pontiac | 190 | 27 | $19,32 |
| Porsche | 393 | 25 | $1,01,62 |
| Rolls-Royce | 488 | 19 | $3,51,13 |
| Saab | 221 | 26 | $27,41 |
| Scion | 154 | 32 | $19,93 |
| Spyker | 400 | 18 | $2,13,32 |
| Subaru | 197 | 29 | $24,82 |
| Suzuki | 160 | 26 | $17,90 |
| Toyota | 236 | 26 | $28,94 |
| Volkswagen | 190 | 32 | $28,07 |
| Volvo | 231 | 27 | $28,54 |
| **Grand Total** | **249** | **26** | **$40,55** |

Result:

· It is observed as the engine horse power increases the highway MPG decreases and the price goes up.

· It helps us to see how different manufacturers design and price their vehicles.

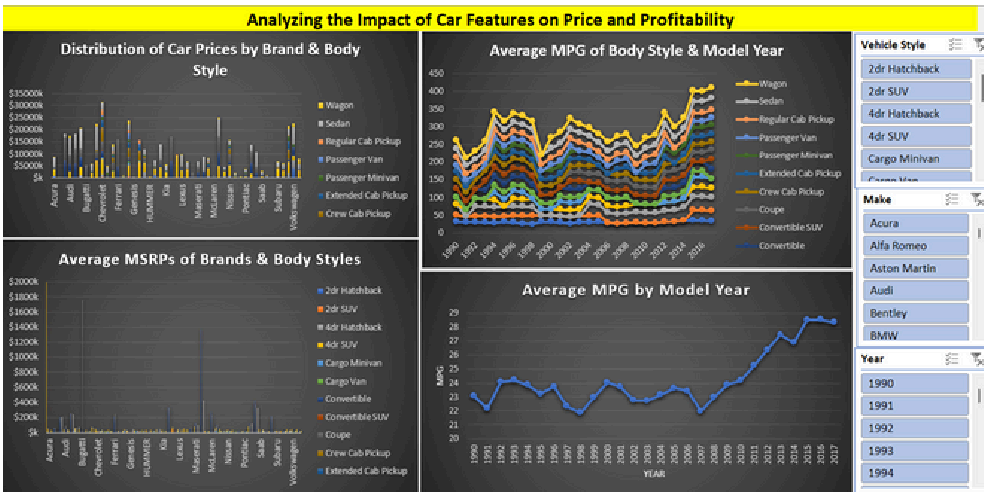
· Such as Bugatti, Maybach, Lamborghini have high performance automobiles. This brands manufacturers mainly focuses on acceleration, speed and track performance. It provides varied horsepower numbers.

· On the other hand large amount of manufacturers focuses on mass market such as Toyota, Honda, and ford as they mainly focuses on utility and fuel economy.

· There are two types of automakers one is luxury which are like Rolls Royce, Bently, Bugatti and many more this focuses on creating expensive automobiles that have cutting edge technology.

· And the another one is mainstream brands which includes Toyota, Honda, Ford and many more this all provides affordable options and also offers higher end priced models with advanced features which offers varied price range.

Dashboard



Result:
This interactive dashboard helps stakeholders to explore different aspects of dataset. It helps to visualize distribution of car prices by brand and body style, average MSRPs of brands and body styles, average MPG by model year and average MPG of body style and model year. This helps to understand the relationship between horsepower, MPG and price across different car brands. This information helps the manufacturers to make informed decisions of pricing, product development, marketing and competitiveness of market. Manufacturers by optimizing the factors as discussed can optimize maximum profitability while keeping in mind the needs of consumer demand.

Drive Link
https://docs.google.com/spreadsheets/d/184tjYBAI2fvyYnGquVBQgMvfo8xwjjnK/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true

Loom Link
https://www.loom.com/share/d6cf481072664eb09872f0d7243e3600

# Project – 8
# ABC Call Volume Trend Analysis

Project Description

In this project we have done analysis of inbound call data for ABC Insurance Company, which focuses on optimizing call handling efficiency and improving customer service. The analysis is done by performing certain analysis which will give us useful insights. This begins by performing average call duration analysis which provides us insights into efficiency of call handling process throughout the day. Then using visualization call volume analysis is presented to show the trends in call over time. Then we have proposed a detailed manpower strategy which aims to reduce abandon rate to 10% this is done by calculating minimum number of agents required in each time bucket. Then at last we have address the issue of unanswered calls during night shifts, where also we have proposed manpower plan to make sure customer satisfactory experience. Overall this project highlights the assumptions, methodologies and recommendations, that provides actionable insights for ABC Insurance Company's management team to optimize call centre operations and increase the customer satisfaction.

Approach

·We begin by importing and preprocessing the dataset.

·We then ensured that data is cleaned to have accurate outcomes and results.

·Then using descriptive statistics and visualization we have calculated the average call duration for each time bucket.

·For call volume analysis, using visualization through graphs we have plot the bar chart which shows the temporal distribution of call volume.

·Then after for we have plot the graph to display the phone number counts, ratio distribution and average call durations. This is done to present the manpower allocation plan for each time bucket which ensures adequate coverage to meet service level targets.

·Then we have analyzed the distribution of calls during night shift period. This helps to determine manpower requirements for calls to handle which maintains the abandon rate. This helps to propose manpower plan for night shift according to the availability and scheduling of the agents.

Tech-Stack Used

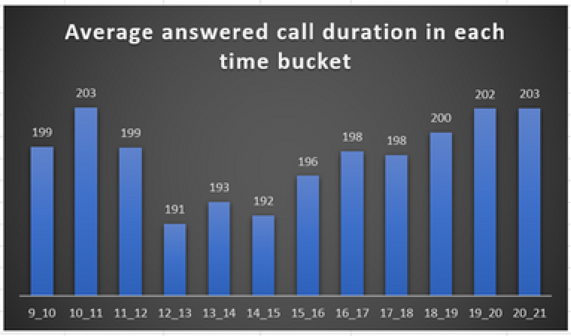
Microsoft® Word 2019 MSO (Version 2501 Build 16.0.18429.20132) 64-bit for data analysis, pivot table creation and chart generation.

Insights

1. Average Call Duration: Determine the average duration of all incoming calls received by agents. This should be calculated for each time bucket.

Your Task: What is the average duration of calls for each time bucket?

| Time Bucket | Average Call Duration |
|---|---|
| 9_10 | 199 |
| 10_11 | 203 |
| 11_12 | 199 |
| 12_13 | 191 |
| 13_14 | 193 |
| 14_15 | 192 |
| 15_16 | 196 |
| 16_17 | 198 |
| 17_18 | 198 |
| 18_19 | 200 |
| 19_20 | 202 |
| 20_21 | 203 |

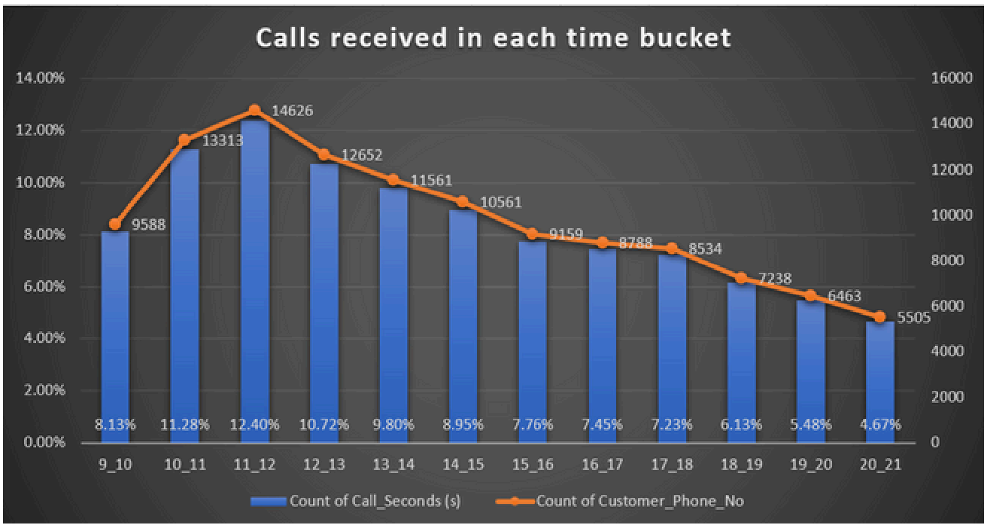



Highest Average Call Duration: 10_11

Lowest Average Call Duration: 12_13

1. Call Volume Analysis: Visualize the total number of calls received. This should be represented as a graph or chart showing the number of calls against time. Time should be represented in buckets (e.g., 1-2, 2-3, etc.).

Your Task: Can you create a chart or graph that shows the number of calls received in each time bucket?

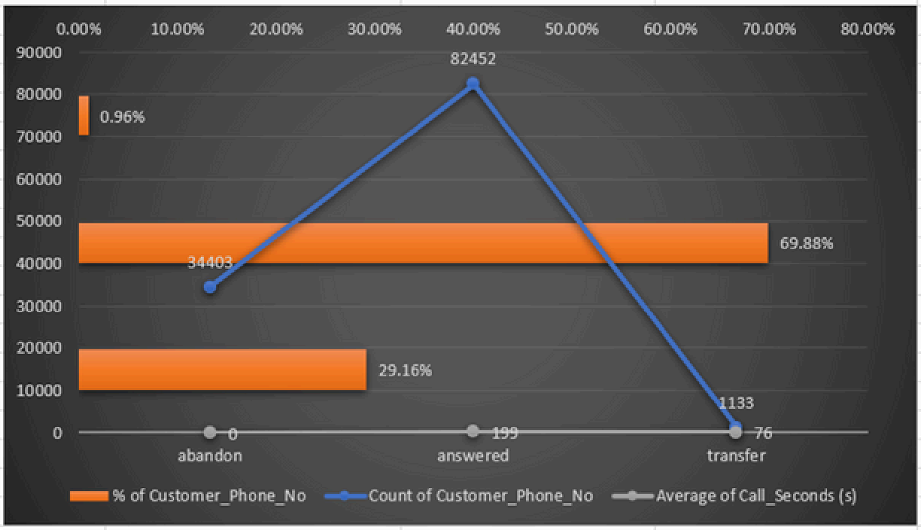


Max number of calls answered in duration: 11_12
Min number of calls answered in duration: 20_21

1. Manpower Planning: The current rate of abandoned calls is approximately 30%. Propose a plan for manpower allocation during each time bucket (from 9 am to 9 pm) to reduce the abandon rate to 10%. In other words, you need to calculate the minimum number of agents required in each time bucket to ensure that at least 90 out of 100 calls are answered.

Your Task: What is the minimum number of agents required in each time bucket to reduce the abandon rate to 10%?

| Row Labels | Count of Customer_Phone_No | % of Customer_Phone_No | Average of Call_Seconds (s) |
|---|---|---|---|
| abandon | 34403 | 29.16% | 0 |
| answered | 82452 | 69.88% | 199 |
| transfer | 1133 | 0.96% | 76 |
| **Grand Total** | **117988** | **100.00%** | **140** |



| | |
|---|---|
| For January the total sum of call duration = 676664s | |
| Sum of call records(sec) | 676664 |
| Sum of call records(hr) | 187.9622222 |
| Abandon rate = 30% | |
| Efficiency rate = 60% as it is mentionend agent works for 6 hours per day | |
| Total number of agents working in a day | 31.32703704 |
| This means 31 agents work in a day | |
| Efficiency should be 90% in order to reduce abandon rate to 10% | |
| Total number of agents required for above condition | 46.5 |

| Row Labels | Count of Call_Seconds (s) |
|---|---|
| 10_11 | 11.28% |
| 11_12 | 12.40% |
| 12_13 | 10.72% |
| 13_14 | 9.80% |
| 14_15 | 8.95% |
| 15_16 | 7.76% |
| 16_17 | 7.45% |
| 17_18 | 7.23% |
| 18_19 | 6.13% |
| 19_20 | 5.48% |
| 20_21 | 4.67% |
| 9_10 | 8.13% |
| **Grand Total** | **100.00%** |

| Row Labels | % Count of Call_Seconds (s) | Decimal Count of Call_Seconds (s) | Agents Required |
|---|---|---|---|
| 9_10 | 8.13% | 0.08 | 4 |
| 10_11 | 11.28% | 0.11 | 5 |
| 11_12 | 12.40% | 0.12 | 6 |
| 12_13 | 10.72% | 0.11 | 5 |
| 13_14 | 9.80% | 0.10 | 5 |
| 14_15 | 8.95% | 0.09 | 4 |
| 15_16 | 7.76% | 0.08 | 4 |
| 16_17 | 7.45% | 0.07 | 3 |
| 17_18 | 7.23% | 0.07 | 3 |
| 18_19 | 6.13% | 0.06 | 3 |
| 19_20 | 5.48% | 0.05 | 3 |
| 20_21 | 4.67% | 0.05 | 2 |

Total agents required to reduce the abandon rate to 10% is 47.
The distribution according to each hour is given in the table (agent required) column.

1. Night Shift Manpower Planning: Customers also call ABC Insurance Company at night but don't get an answer because there are no agents available. 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. The distribution of these 30 calls is as follows:

Your Task: Propose a manpower plan for each time bucket throughout the day, keeping the maximum abandon rate at 10%.
Assumptions: An agent works for 6 days a week; On average, each agent takes 4 unplanned leaves per month; An agent's total working hours are 9 hours, out of which 1.5 hours are spent on lunch and snacks in the office. On average, an agent spends 60% of their total actual working hours (i.e., 60% of 7.5 hours) on calls with customers/users. The total number of days in a month is 30.

| 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 |

| Count of Call_Status | Column Labels | | | |
|---|---|---|---|---|
| Row Labels | abandon | answered | transfer | Grand Total |
| 01-01-2022 | 4117 | 526 | 1 | 4644 |
| 02-01-2022 | 869 | 2478 | 4 | 3351 |
| 03-01-2022 | 935 | 3791 | 63 | 4789 |
| 04-01-2022 | 1494 | 3567 | 52 | 5113 |
| 05-01-2022 | 559 | 4166 | 65 | 4790 |
| 06-01-2022 | 1113 | 3755 | 83 | 4951 |
| 07-01-2022 | 1209 | 3676 | 63 | 4948 |
| 08-01-2022 | 2099 | 2543 | 30 | 4672 |
| 09-01-2022 | 897 | 2746 | 9 | 3652 |
| 10-01-2022 | 965 | 3937 | 81 | 4983 |
| 11-01-2022 | 3524 | 1113 | | 4637 |
| 12-01-2022 | 3807 | 836 | | 4643 |
| 13-01-2022 | 1685 | 2433 | 5 | 4123 |
| 14-01-2022 | 264 | 2885 | 6 | 3155 |
| 15-01-2022 | 339 | 2707 | 12 | 3058 |
| 16-01-2022 | 1413 | 3685 | 44 | 5142 |
| 17-01-2022 | 2737 | 19140 | 470 | 22347 |
| 18-01-2022 | 1232 | 4451 | 91 | 5774 |
| 19-01-2022 | 682 | 3987 | 34 | 4703 |
| 20-01-2022 | 2586 | 1735 | 1 | 4322 |
| 21-01-2022 | 945 | 2721 | 9 | 3675 |
| 22-01-2022 | 619 | 2668 | 4 | 3291 |
| 23-01-2022 | 313 | 2906 | 6 | 3225 |
| Grand Total | 34403 | 82452 | 1133 | 117988 |
| | | | | 5129.913 |

| Time_Buckets | Distribution of calls | Distribution of time | Agents required |
|---|---|---|---|
| 9_10 | 3 | 10 | 1 |
| 10_11 | 3 | 10 | 1 |
| 11_12 | 2 | 15 | 1 |
| 12_13 | 2 | 15 | 1 |
| 13_14 | 1 | 30 | 0 |
| 14_15 | 1 | 30 | 0 |
| 15_16 | 1 | 30 | 0 |
| 16_17 | 1 | 30 | 0 |
| 17_18 | 3 | 10 | 1 |
| 18_19 | 4 | 7.5 | 2 |
| 19_20 | 4 | 7.5 | 2 |
| 20_21 | 5 | 6 | 2 |
| Total | 30 | | 11 |

The average of incoming calls from 01-01-2022 to 23-01-2022 is 5129.91304
In 30 days,average of 5130 calls are coming in a day and as we know a day's 30% calls are coming at nights,

Average number of calls coming at night is 1539

From Task 1,We have already got the average duration of all calls(in s) that is 199.
And we will have to answer 90% calls(means upto 10% call can be abandoned),

Additional Time Required is(in seconds): 275635
Additional Time Required is(in hours): 77

Since an agent works for 6 hours,
For additional 77 hours of work,number of additional agents required will be: 13

This is Total number of **estimated** additional agent required is **13**.
Total number of additional agents required in **practical** is **11**.

Total number of additional agents required is 11. Discrete number of agents required for each time interval from 9PM to 9AM is given in the table (agent required) column.

Result
In analyzing the inbound call data for ABC Insurance Company, we found that the average call duration varies across different time buckets throughout the day. Call volume follows a distinct pattern, with peaks and troughs corresponding to different hours. To reduce the abandon rate from 30% to 10%, we determined the minimum number of agents required during each time bucket, ensuring that at least 90 out of 100 calls are answered. Additionally, to address nighttime calls and provide a better customer experience, we proposed a manpower plan for each time bucket, maintaining the abandon rate at or below 10%. These solutions are based on specific calculations and assumptions regarding agent availability and call distribution patterns.

Drive Link for Excel
https://docs.google.com/spreadsheets/d/1WxiyxMxCuUH4O93kNWwY8RGBwE5aSdfU/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true
Drive Link for Loom Video
https://www.loom.com/share/b0048433361c4d148c21f1aef65ae9b6?sid=0dd2e282-f75f-40ff-8ebc-851bb75b75f2