PORTFOLIO PROJECT

Here is where my presentation begins

HELLO !

I AM PREETY MOHANTA

I am here to give the Trainity portfolio project

A student of MBA who is a smart worker by nature with an interest in Marketing, Data Analytics, Advertising and Social media. Passionate in excel, sql, Visual Merchandising, Human relations and eager to learn and adapt to different environment. Love to solve problems. Been driven by my curiosity to find answers to the most pressing questions.

I have earned my MBA from Lovely Professional University, with a graduate certificate in Marketing.

Specialties in Microsoft Office (Excel, Word, PowerPoint), sql, M-Commerce and WordPress.

Enjoy travelling and eating my way around the world and talking about the future

In addition, I had the opportunity to work as a data analyst trainee at Trainity, where I gained hands-on experience in data analysis, database management, and creating data visualizations. This experience has helped me to develop a better understanding of the data analysis process and its applications in the industry and also made business strategy decisions .

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.

I am excited to continue developing my skills and knowledge as a data analyst or business analyst and look forward to contributing to meaningful projects in this field.

TABLE OF CONTENTS

CONTENTS OF THIS PORTFOLIO

APPORACH

HERE ARE THE STEPS WE WILL FOLLOW FOR EXECUTING OUR PROJECT:

- Download the data set
- Then clean the data by removing unwanted information, duplicate, handle the missing data.
- After cleaning the data convert the data types and make it ready to use.
- Now understand the data and use the excel and perform the tasks
- With the help of excel we will Create a combo chart, regression analysis, correlation coefficient, formula's, pivot tables, etc And at the end we get insight from it.

Tech Stack Used

Excel

Powerpoint

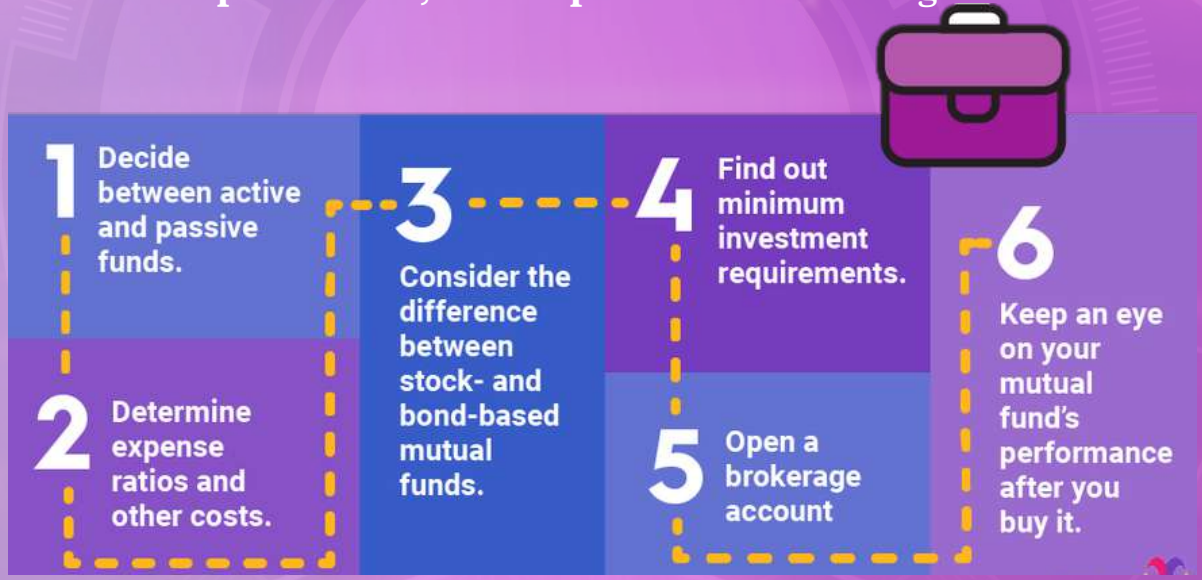
Loom

01

Data Analytics Process

Application in Real Life Scenario

TASK: make a real life scenario of Data analytics process of collecting, Cleansing, Transforming, and Organizing data in order to draw conclusions, make predictions, and help in decision making.



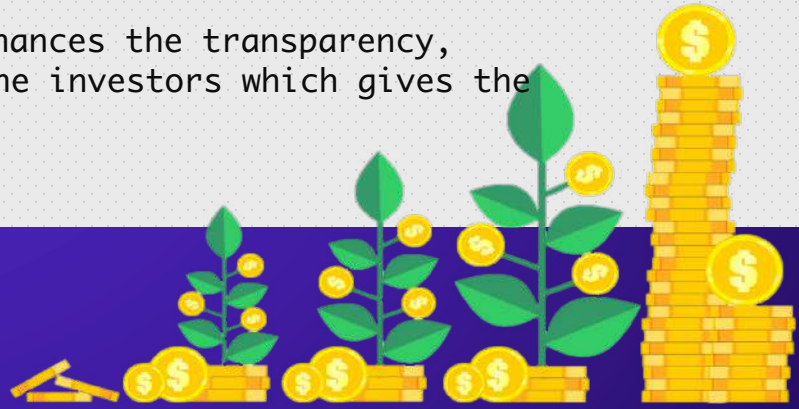
CONCLUSION....

Leveraging data analytics for purchasing mutual funds provides numerous intake in today's dynamic investment market.Its very important to know details related to the mutual funds, not having proper knowledge about our investment we cam make mistake and risk our money.

Data analytics helps as to identify the trends and the patterns in the market by studying the detailed past history which leads to investment portfolio that minimizes overall risk and maximizes returns and optimize portfolio performance.

And when ever we need our money in emergency, we can immediately sell mutual fund shares without incurring losses or reduced liquidity or trading overrun.

Data analytics in the mutual fund buying process enhances the transparency, efficiency, and decision-making effectiveness for the investors which gives the insights of its overall investment portfolio.



02

Instagram User Analytics

SQL Fundamentals

TASKS

01 Find the 5 oldest users of the Instagram from the database provided.

02 Identify users who have never posted a single photo on Instagram

03 Determine the winner of the contest and provide their details to the team

04 Identify and suggest the top five most commonly used hashtags on the platform.

05 Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

A) MARKETING ANALYSIS

Loyal User Reward

Stack used: SQLQUERY

```
SELECT *
FROM users
ORDER BY created_at
LIMIT 5
```

The five oldest users on Instagram from the provided database

Id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson 2	2016-05-14 07:56:26

CONCLUSION: Users 80, 67, 63, 95, 38 are the 5 oldest users on the platform.

A) MARKETING ANALYSIS

Inactive User Engagement

Stack used: SQLQUERY

```
SELECT username
FROM users
LEFT JOIN photos
ON users.id = photos.user_id
WHERE photos.id IS NULL;
```

Remind Inactive Users to Start Posting

05	Aniya_Hackett	74	Hulda.Macejkovic	41	Mckenna17
83	Bartholome.Bernhard	14	Jaclyn81	66	Mike.Auer39
91	Bethany20	76	Janelle.Nikolaus81	49	Morgan.Kassulke
80	Darby_Herzog	89	Jessyca_West	71	Nia_Haag
45	David.Osinski47	57	Julien_Schmid	36	Ollie_Ledner37
54	Duane60	07	Kasandra Homenick	34	Pearl7
90	Esmeralda.Mraz57	75	Leslie67	21	Rocio33
81	Esther.Zulauf61	53	Linnea59	25	Tierra.Trantow
68	Franco_Keebler64	24	Maxwell.Halvorson		

A) MARKETING ANALYSIS

Contest Winner Declaration

Stack used: SQLQUERY

```
SELECT id,
username
FROM users
WHERE id = (SELECT user_id
FROM photos
WHERE id = (SELECT photo_id
FROM likes
GROUP BY photo_id
ORDER BY Count(photo_id) DESC
LIMIT 1));
```

Declaring Contest Winner

Id	username
52	Zack_Kemmer93

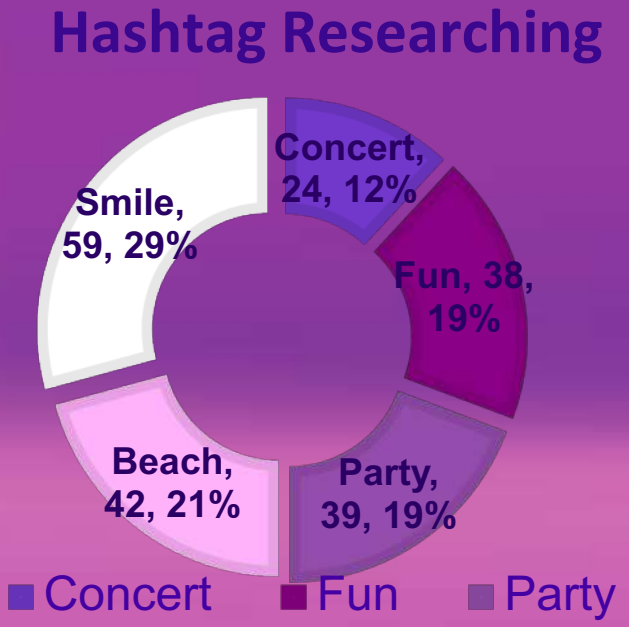
CONCLUSION: The winner of the contest ‘the most likes on a single photo’ is user 52 with a total of 48 likes on post 145.

A) MARKETING ANALYSIS

Hashtag Research

Stack used: SQLQUERY

```
SELECT tags.tag_name,
COUNT(*) AS total
FROM photo_tags
JOIN tags
ON photo_tags.tag_id = tags.id
GROUP BY tags.id
ORDER BY total DESC
LIMIT 5
```



CONCLUSION: This gives a clear recommendation that these hashtag is use to reach most people on platform

A) MARKETING ANALYSIS

Ad Campaign Launch

Stack used: SQLQUERY

```
SELECT Dayname(created_at) "day
of week",
Count(Dayname(created_at)) "count
of users registered"
FROM users
GROUP BY Dayname(created_at)
ORDER BY
Count(Dayname(created_at)) DESC
LIMIT 2;
```

CONCLUSION: According to the data most of the user register on **Sunday and Thursday**

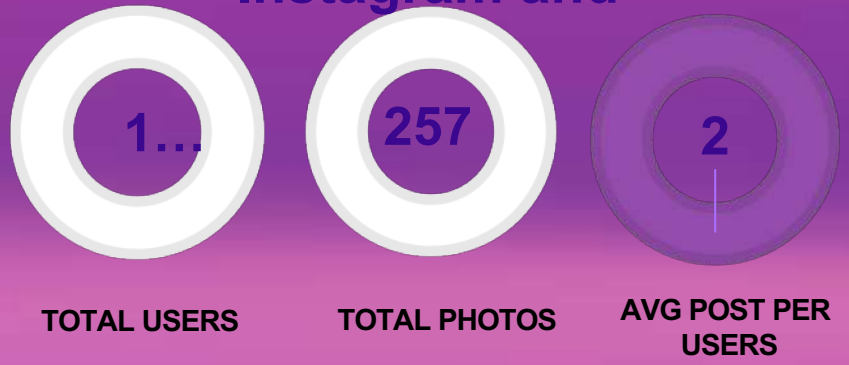
A) INVESTOR METRICS

User Engagement

Stack used: SQLQUERY

```
SELECT (SELECT Count(id)
FROM photos) / (SELECT
Count(DISTINCT user_id)
FROM photos) AS
Average_posts_per_User,
(SELECT Count(id)
FROM photos) / (SELECT Count(id)
FROM users) AS
Ratio_of_Total_Posts_to_Total_User
s;
```

Total number of users in Instagram The ratio between number of photos posted in Instagram and



CONCLUSION: A user posts 2.57 posts on an average. There are 257 photos in total on Instagram. There are100 users in total on Instagram

A) INVESTOR METRICS

Bots & Fake Accounts

Stack used: SQLQUERY

```
SELECT id,
username
FROM users
WHERE id IN (SELECT user_id
FROM likes
GROUP BY user_id
HAVING Count(user_id) = (SELECT
Count(id)
FROM photos));
```

Who have liked every single photo on the site

ID	USERNAME	LIKES	ID	USERNAME	LIKES
05	Aniya_Hackett	257	24	Maxwell.Halvorson	257
91	Bethany20	257	41	Mckenna17	257
54	Duane60	257	66	Mike.Auer39	257
14	Jaclyn81	257	71	Nia_Haag	257
76	Janelle.Nikolaus81	257	36	Ollie_Ledner37	257
57	Julien_Schmidt	257	21	Rocio33	257
75	Leslie67	257			

CONCLUSION: There are 13 bots in total on the platform who have liked all the photos.

CONCLUSIONS….

- This project helped me to understand MYSQL works and the importance of data analysis for the organization and from these insight be can make better decision.
- This project helps as to answer all the questions and give insight of the business.
- With this analysis user engagement will be helpful for the company growth.
- Teams can removes all the bots or fake account to get accuracy in the data.

03

Operation & Metric Analytics

Advanced SQL

INSIGHTS TASKS

Jobs Reviewed Over Time

Objective: Calculate the number of jobs reviewed per hour for each day in November 2020.

Throughput Analysis

Objective: Calculate the 7-day rolling average of throughput (number of events per second).

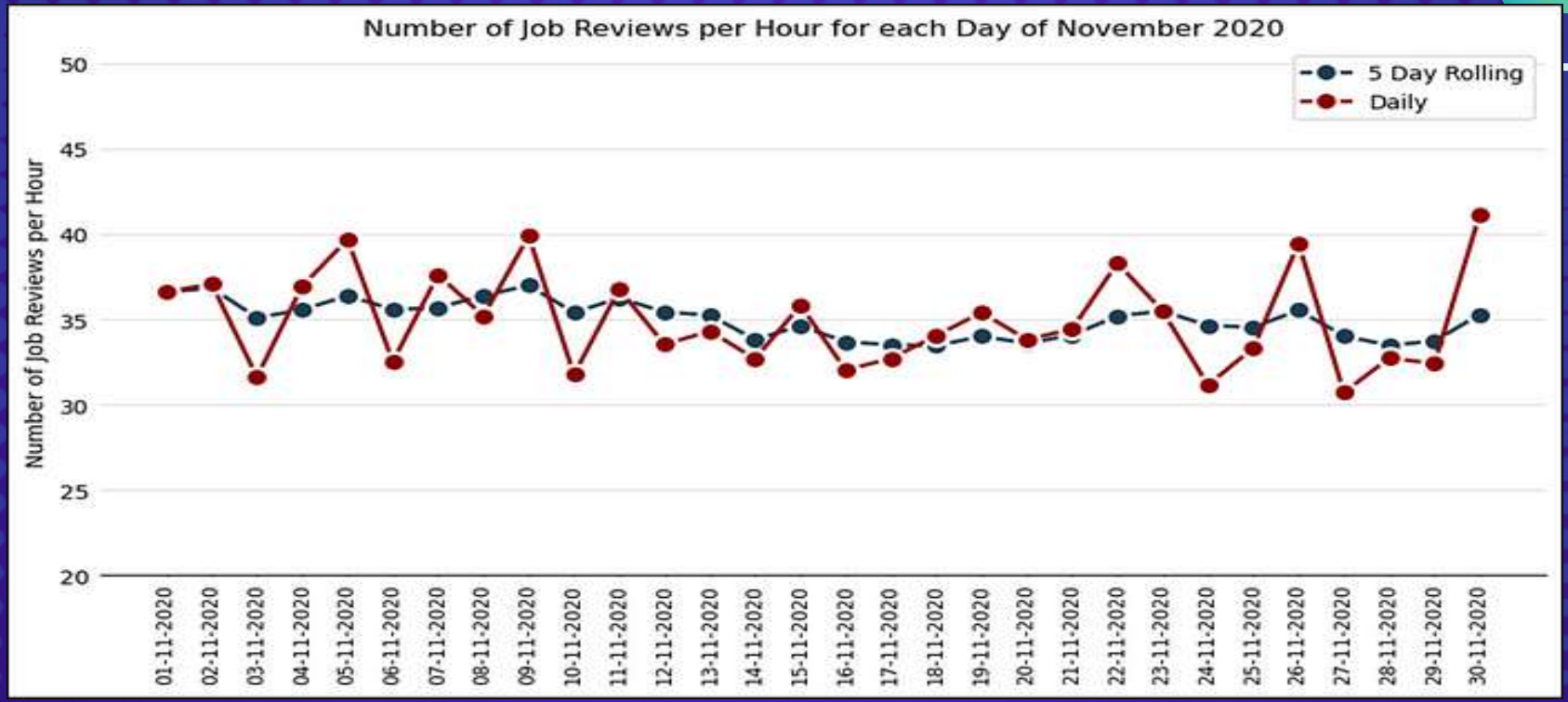
Language Share Analysis

Objective: Calculate the percentage share of each language in the last 30 days.

Duplicate Rows Detection

Objective: Identify duplicate rows in the data.

A. NUMBER OF JOBS REVIEWED



From this table, we can see that the number of **job reviews** done in between 30 and 40 for **most** days of November 2020.

B. THROUGHOUT ANALYSIS

OUTPUT:

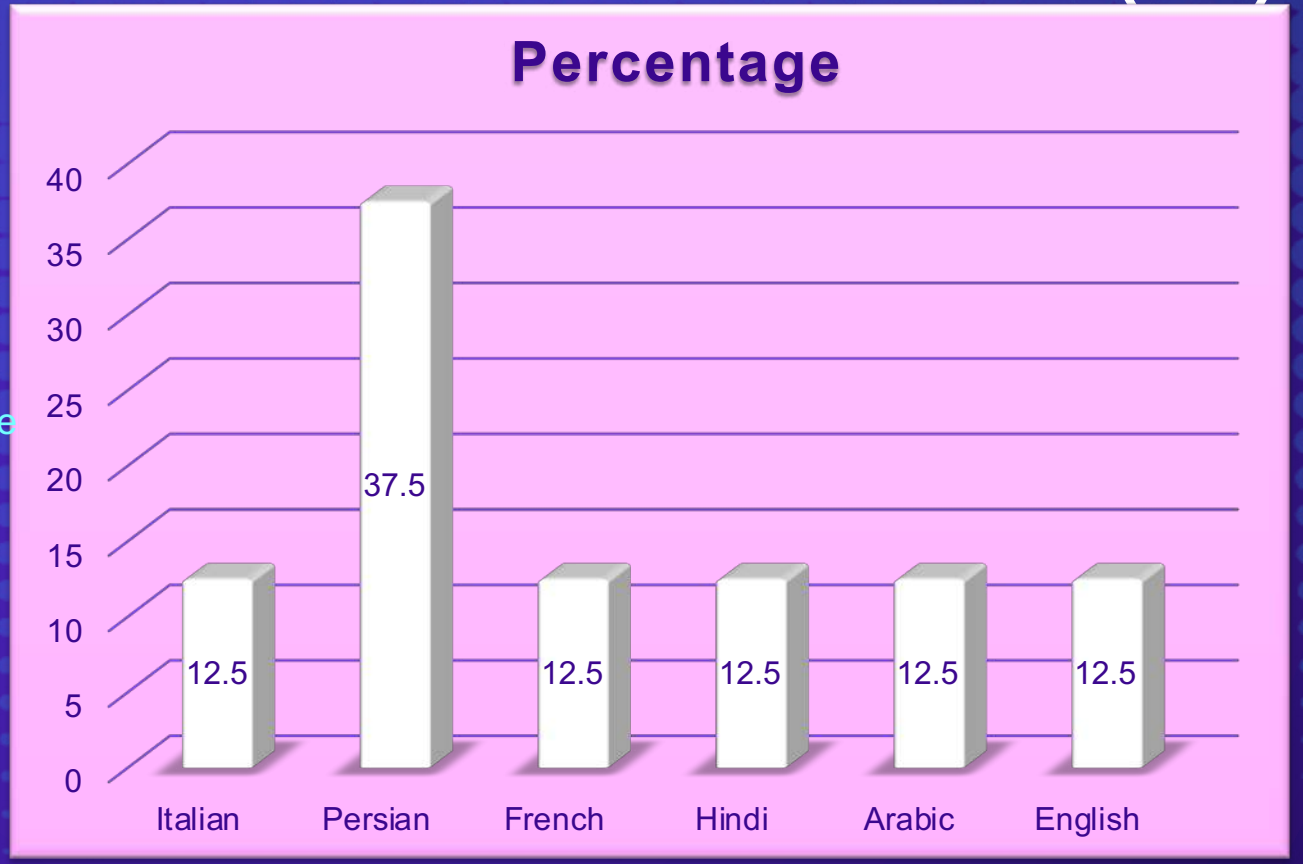
Insight: I would prefer the 7-Day Rolling Average over daily metric for throughput. The reason being daily metrics can go up or down on a daily basis for factors that cannot be controlled by the organizations like seasonality, major events etc. Continue using the rolling average to observe trends without being influenced by daily fluctuations.

Date_of_record	No_events_per_day	Avg_7_day_rolling
25-11-2020	1.3333	1
26-11-2020	1.3333	1
27-11-2020	1.3333	1
28-11-2020	1.3333	1.25
29-11-2020	1.3333	1.2
30-11-2020	1.3333	1.3333

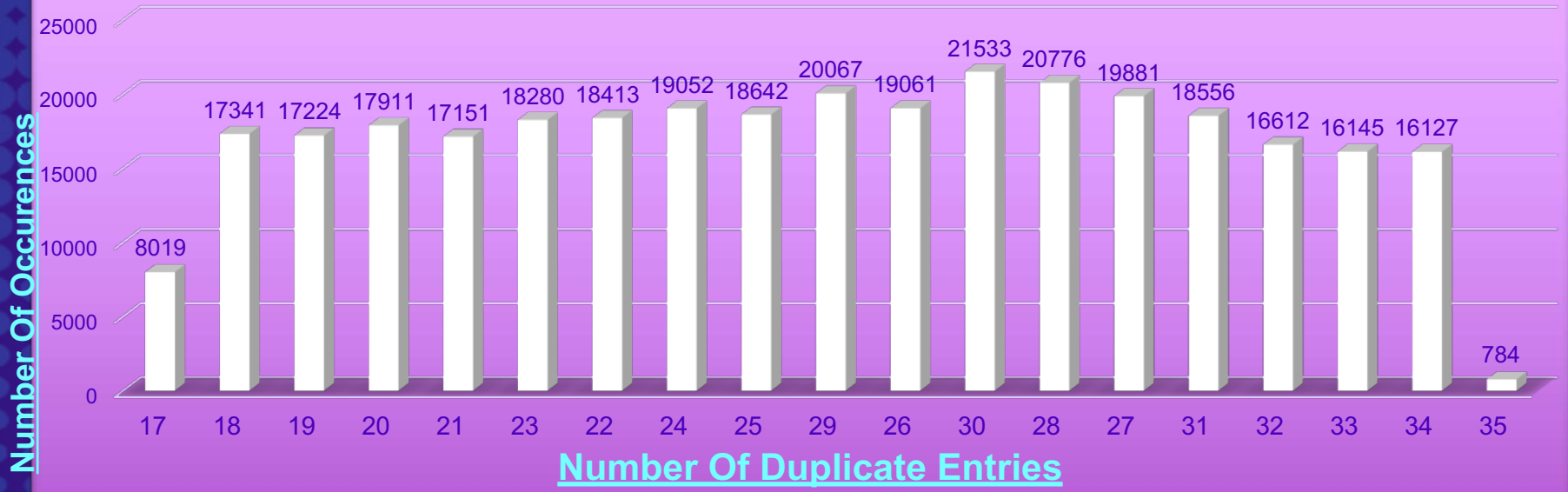
C. LANGUAGE SHARE ANALYSIS

OUTPUT:

***Insight*:** Language distribution is relatively balanced and it rectifies that Persian Language get the highest percentage share



D. DUPLICATE ROWS



OUTPUT:

Insight: There are 89 number of rows with duplicate values of job_id.

INSIGHTS TASKS

Weekly User Engagement

Objective: Measure the activeness of users on a weekly basis.

User Growth Analysis

Objective: Analyze the growth of users over time for a product.

Weekly Retention Analysis

Objective: Analyze the retention of users on a weekly basis after signing up for a product.

Weekly Engagement Per Device

Objective: Measure the activeness of users on a weekly basis per device.

Email Engagement Analysis

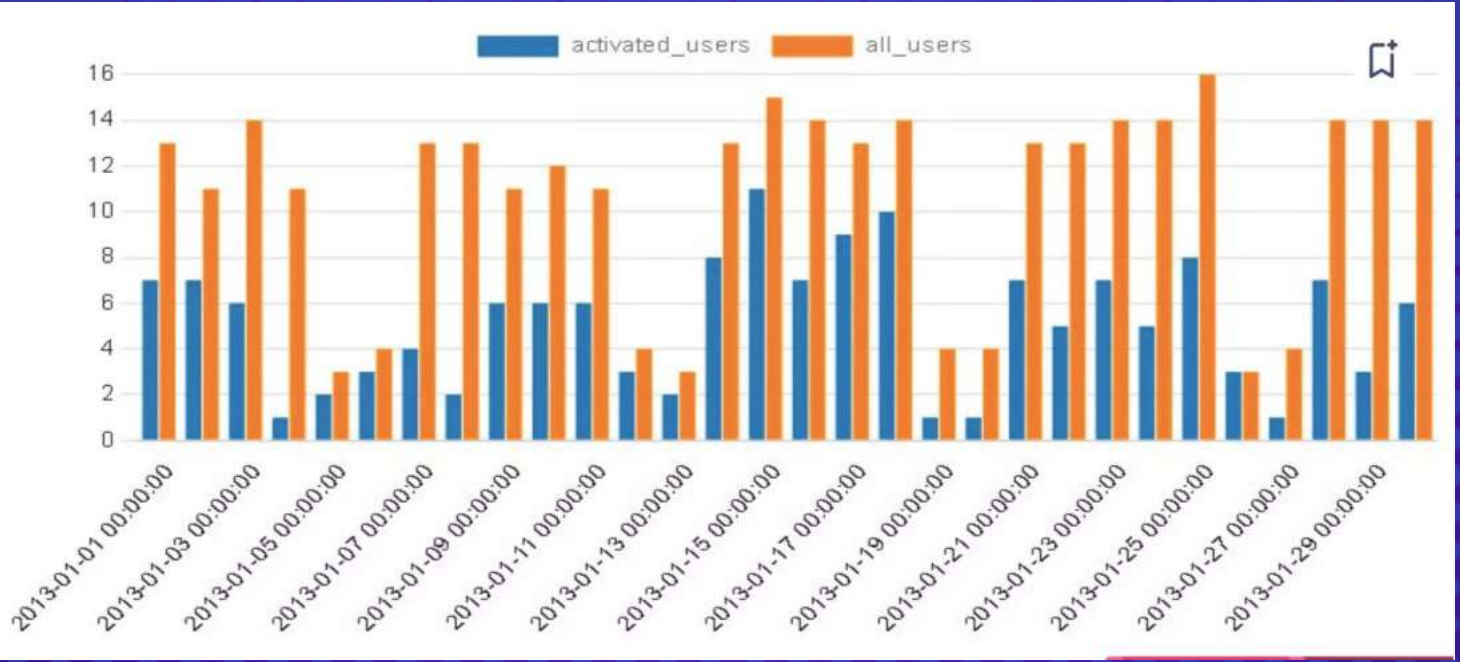
Objective: Analyze how users are engaging with the email service

A. WEEKLY USER ENGAGEMENT

OUTPUT:

***Insight*:** Language distribution is relatively balanced and it rectifies that Persian Language get the highest percentage share

B. USER GROWTH ANALYSIS



OUTPUT:

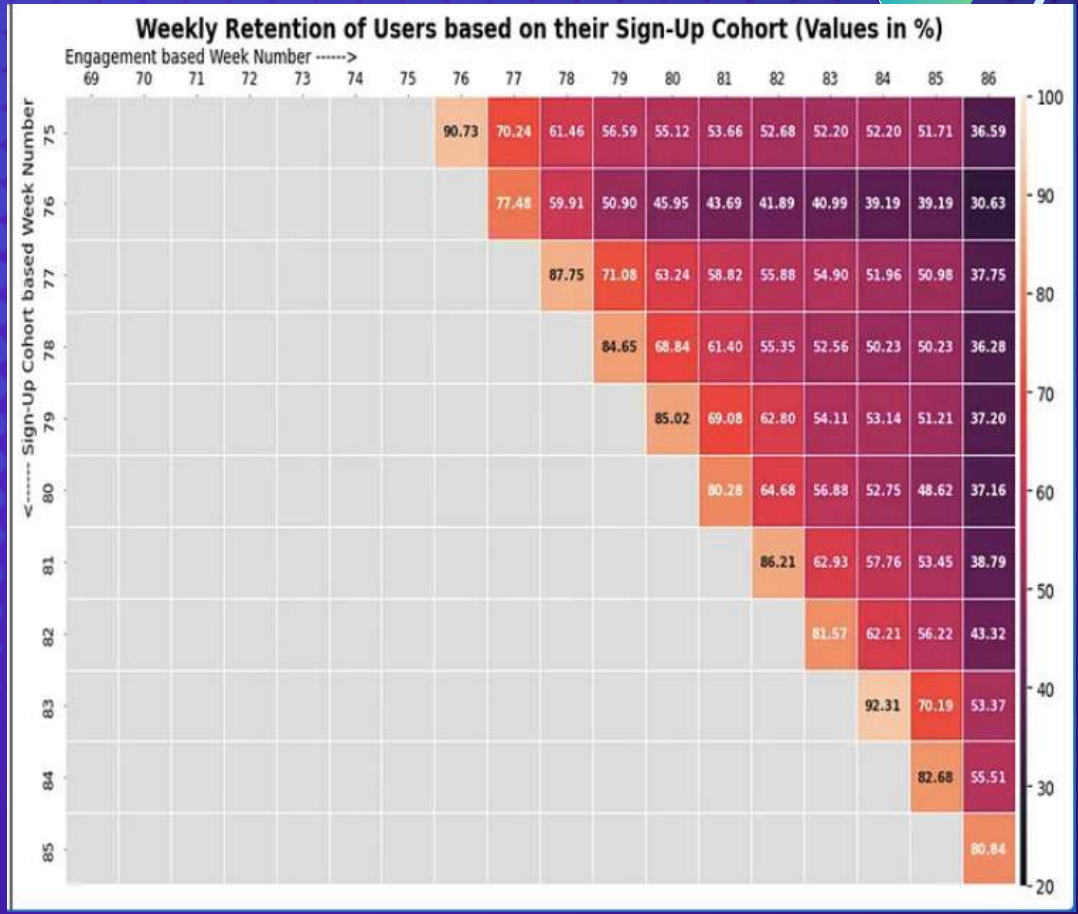
***Insight*:** User growth has been positive over time, with some fluctuations.

From 1st day to Last day in dataset of users there is 9381 users grow

C. WEEKLY RETENTION ANALYSIS

OUTPUT:

***Insight*:** We can observe that the retention rate is remaining same for most of the middle event weeks. On further investigation, we found that event “sent_weekly_digest” forms majority of the events for most of the event weeks from 69 to 85 and the number of occurrence of this event is remaining constant for most of the middle overs for a given sign-up cohort week.



D. Weekly Engagement Per Device

OUTPUT:

***Insight*:** Given is average weekly engagement per device

The weekly data per device was very large (960 rows) hence calculated the weekly data

MacBook pro is used the most
Samsung galaxy table is used least
All three devices are laptops. It is understandable as these are formal events mostly used in corporate environment

Device_name	Avg_weekly_users	Avg_times_used_weekly
Acer Aspire Desktop	26.00	32.95
Acer Aspire Notebook	43.16	56.84
Amazon Fire Phone	10.56	13.78
Asus Chromebook	43.53	58.89
Dell Inspiron Desktop	46.63	62.74
Dell Inspiron Notebook	91.11	123.47
Hp Pavilion Desktop	42.11	55.84
Htc One	21.84	27.68
Ipad Air	51.44	61.72
Ipad Mini	30.00	34.74
Iphone 4s	46.63	60.58
Iphone 5	123.16	161.21
Iphone 5s	73.32	96.79
Kindle Fire	21.16	25.53
Lenovo Thinkpad	172.95	232.58
Mac Mini	20.47	27.37
Macbook Air	123.16	164.89
Macbook Pro	260.16	358.16
Nexus 10	27.05	31.84
Nexus 5	76.37	99.63
Nexus 7	36.37	43.26
Nokia Lumia 635	28.16	36.26
Samsumg Galaxy Tablet	10.28	12.11
Samsung Galaxy Note	13.47	17.58
Samsung Galaxy S4	91.58	118.74
Windows Surface	18.21	21.53

E. EMAIL ENGAGEMENT ANALYSIS

OUTPUT:

***Insight*:** Email engagement metrics include an open rate of approximately 33.58% and a click rate of about 14.79%

From the above table, we can observe that most email activity is related to sent_weekly_digest

action	month	number_of_mails
email_clickthrough	5	2023
email_clickthrough	6	2274
email_clickthrough	7	2721
email_clickthrough	8	1992
email_open	5	4212
email_open	6	4658
email_open	7	5611
email_open	8	5978
sent_reengagement_email	5	758
sent_reengagement_email	6	889
sent_reengagement_email	7	933
sent_reengagement_email	8	1073
sent_weekly_digest	5	11730
sent_weekly_digest	6	13155
sent_weekly_digest	7	15902
sent_weekly_digest	8	16480

05

Conclusion

This project has been highly beneficial as it allowed me to apply my SQL skills and gain hands-on experience in data analysis.
In this project of Operation Analytics and Investigating Metric Spike, I have achieved various Analytics and logical skills as well as technical skills to efficiently use MySQL. I learn how to understand dataset. What kind of questions we have to ask to get proper insights from data.
Whenever utilized correctly, operational analytics can achieve a significant positive effect

04

Hiring Process Analytics

Statistics

Insight task

Step 1

Hiring Analysis

Step 2

Salary Analysis

Step 3

Salary Distribution

Step 4

Departmental Analysis

Step 5

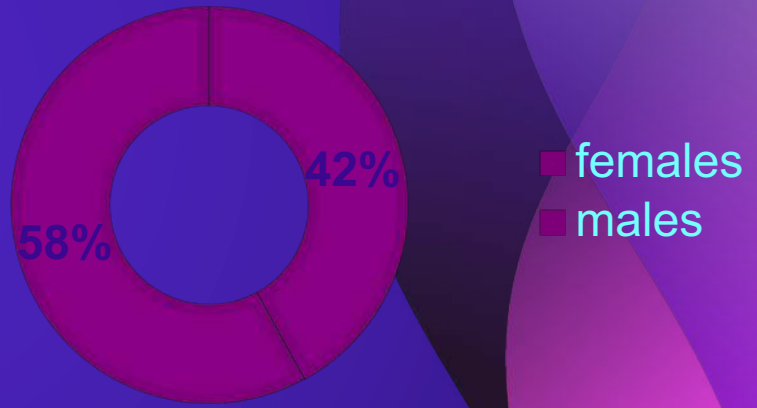
Position Tier Analysis

Hiring Analysis

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

Gender	Number of person
males	2563
females	1856

Row Labels	Sum of Number of person
females	1856
males	2563
Grand Total	**4419**



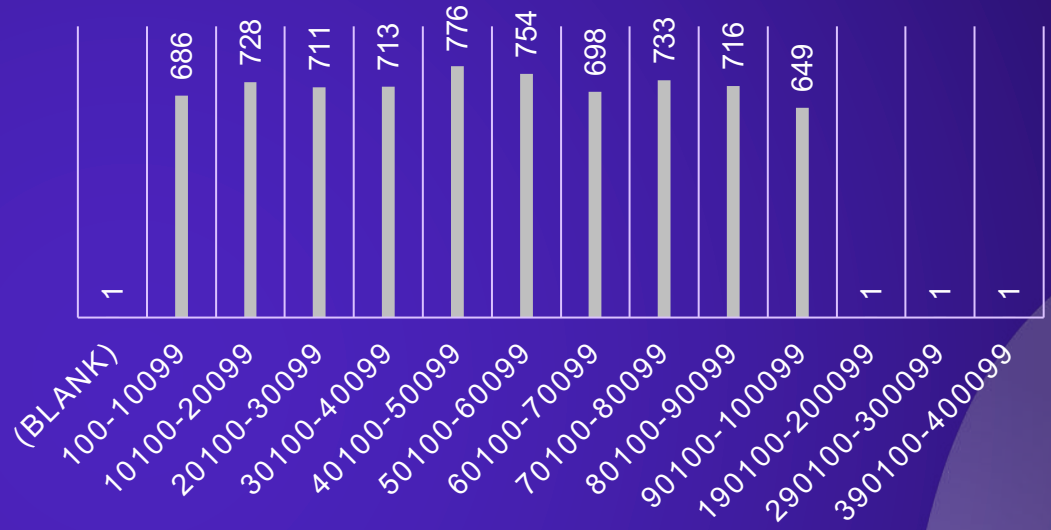
Formula used:

- Male: =COUNTIFS(D:D,D2,C:C,C2)
- Female: =COUNTIFS(D:D,D3,C:C,C3)

Salary Distribution

Min	100	=MIN(G:G)
Max	400000	=MAX(G:G)

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



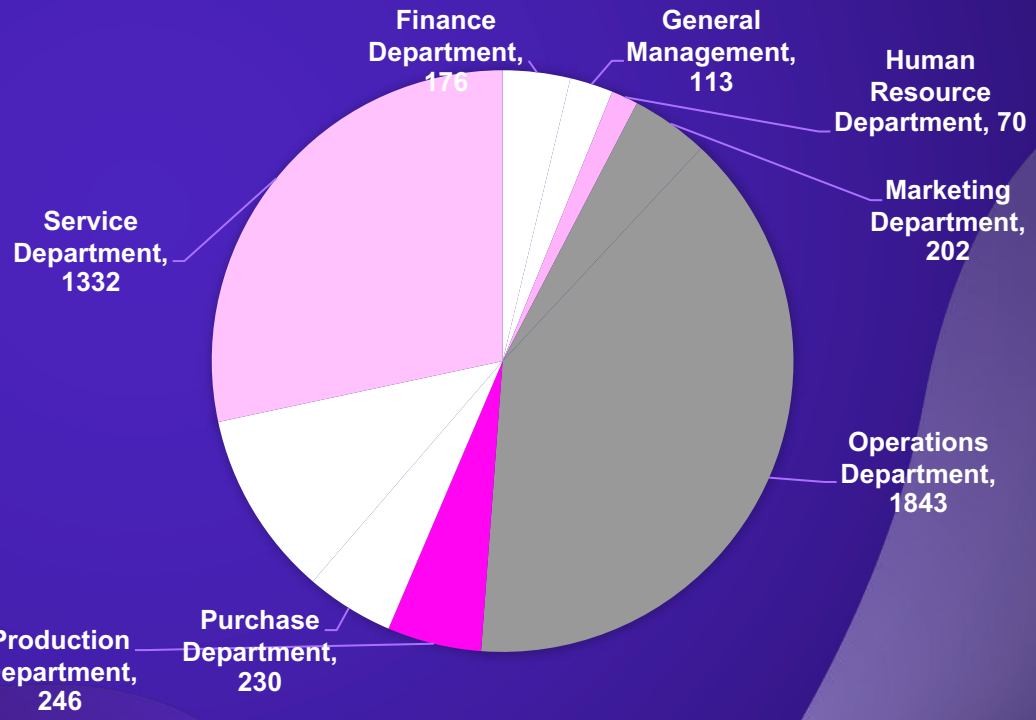
Task: Create class intervals for the salaries in the company.

It is using the PIVOT TABLE and then grouping the salary and finding the range of salary

Departmental Analysis

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

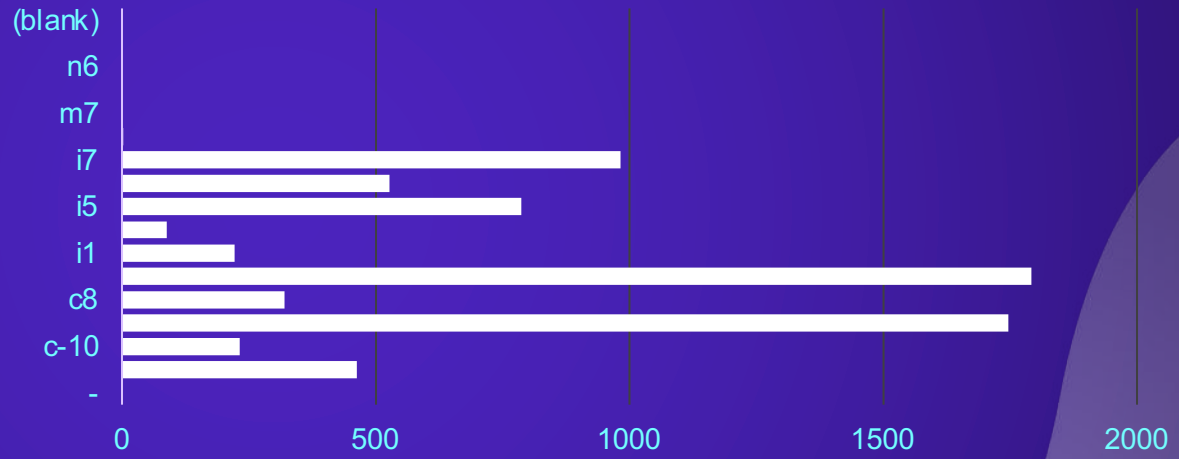
Status	Hired
Row Labels	Count of application_id
Finance Department	176
General Management	113
Human Resource Department	70
Marketing Department	202
Operations Department	1843
Production Department	246
Purchase Department	230
Sales Department	485
Service Department	1332
Grand Total	4697



Position Tier Analysis

Task: Use a chart or graph to represent the different position tiers within the company.

Row Labels	Count of application_id
-	1
b9	463
c-10	232
c5	1747
c8	320
c9	1792
i1	222
i4	88
i5	787
i6	527
i7	982
m6	3
m7	1
n10	1
n6	1
n9	1
(blank)	
Grand Total	7168



RESULT

Hiring Analysis	Salary Analysis	Salary Distribution	Departmental Analysis	Position Tier Analysis
More than half of the hired candidate are male than female	The average salary of the people are 49983.02	We can observe that maximum offered salary is in the interval of 40100-50099 while minimum offered salary are in 190100-200099 290100-300099 390100-400099	From the above pie chart, we can observe that most candidates are hired in **Operations Department** with 39% of hired people	Here, we can observe that the organization has hired most candidates for post tier **c9** 25% of total number of people applied.

05

IMDB Movie Analysis

Final Project-1

TASKS….

You are required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

1st Task	2nd Task	3rd Task	4th Task	5th Task
Movie Genre Analysis	**Movie Duration Analysis**	**Language Analysis**	**Director Analysis**	**Budget Analysis**
Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics of the IMDB scores.	Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.	Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.	Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.	Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

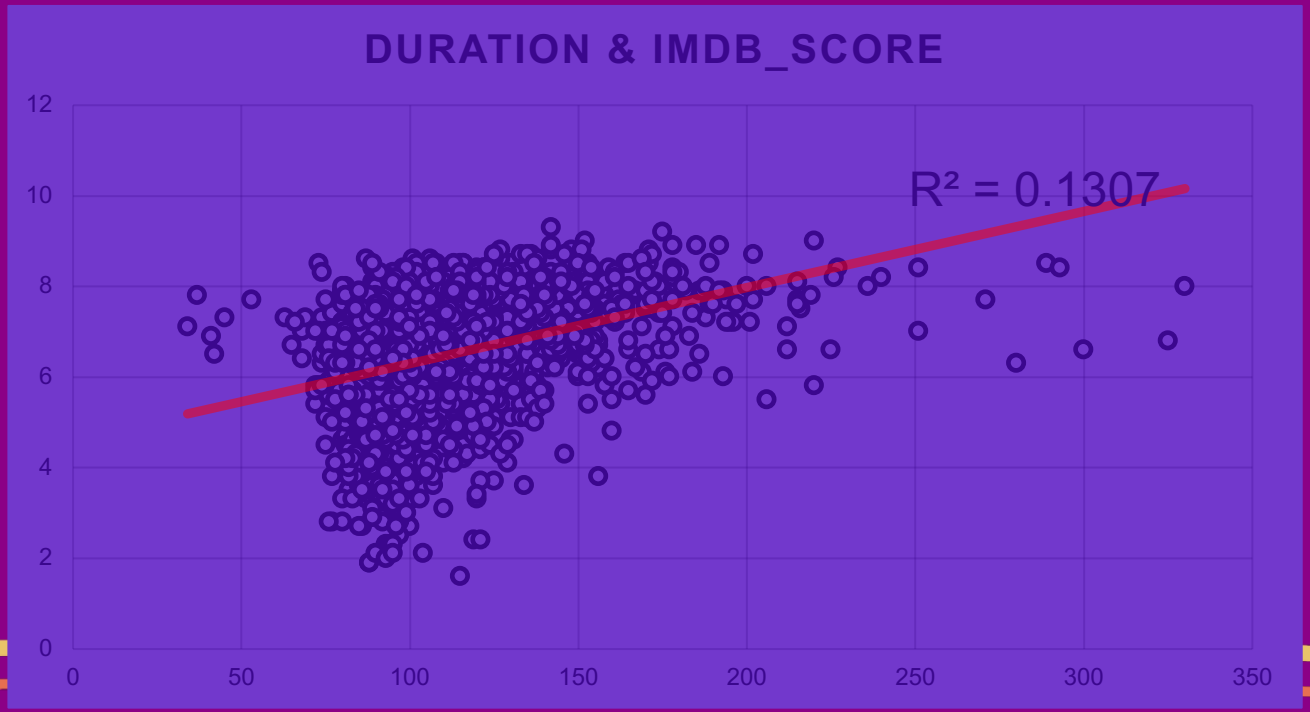
Movie Genre Analysis

Row Labels	Count of Unique Genres	Average of Imdb_Score	Max. of Imdb_Score	Min. of Imdb_Score	StdDevp of Imdb_Score	Varp of Imdb_Score
Action	935	6.286	9	2.1	1.038	1.077
Adventure	367	6.561	8.6	2.3	1.123	1.261
Animation	46	6.763	8	4.5	0.962	0.925
Biography	206	7.152	8.9	4.5	0.698	0.487
Comedy	1026	6.164	8.8	1.9	1.036	1.074
Crime	252	6.945	9.3	3.3	0.867	0.752
Documentary	43	6.951	8.5	1.6	1.400	1.959
Drama	676	6.822	8.8	2.1	0.905	0.820
Family	3	6.500	7.9	5.7	0.993	0.987
Fantasy	35	6.234	7.9	4.3	0.881	0.777
Horror	156	5.813	8.5	2.3	1.005	1.009
Musical	2	6.750	7.2	6.3	0.450	0.203
Mystery	23	6.587	8.5	3.3	1.085	1.177
Romance	2	6.650	7.1	6.2	0.450	0.202
Sci-Fi	8	6.588	8.2	5	0.965	0.931
Thriller	3	5.300	6.3	4.8	0.707	0.500
Western	3	6.767	8.9	4.1	1.996	3.982
(blank)						
Grand Total	3786	6.462	9.3	1.6	1.057	1.118

Movie Duration Analysis

Column1	Column2
Duration	
Mean	109.808505
Standard Error	0.369949997
Median	105
Mode	101
Standard Deviation	22.763201
Sample Variance	518.16332
Kurtosis	12.40512587
Skewness	2.347508256
Range	296
Minimum	34
Maximum	330
Sum	415735
Count	3786
Largest(1)	330
Smallest(1)	34
Confidence Level(95.0%)	0.725320612

	Duration	Imdb_Score
Duration	1	
Imdb_Score	0.361506958	1



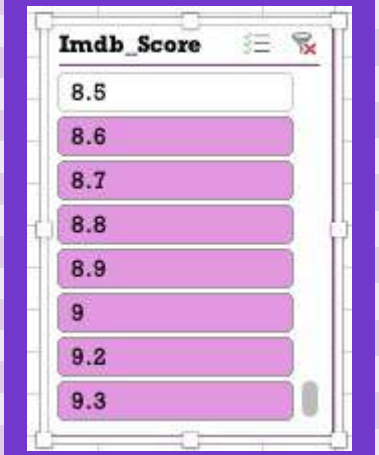
Language Analysis

Row Labels	Count of Language	Average of Imdb_Score	StdDevp of Imdb_Score2	Varp of Imdb_Score
Aboriginal	2.000	6.950	0.550	0.303
Arabic	1.000	7.200	0.000	0.000
Aramaic	1.000	7.100	0.000	0.000
Bosnian	1.000	4.300	0.000	0.000
Cantonese	8.000	7.238	0.412	0.170
Czech	1.000	7.400	0.000	0.000
Danish	3.000	7.900	0.432	0.187
Dari	2.000	7.500	0.100	0.010
Dutch	3.000	7.567	0.330	0.109
Dzongkha	1.000	7.500	0.000	0.000
English	3606.000	6.421	1.052	1.107
Filipino	1.000	6.700	0.000	0.000
French	37.000	7.286	0.554	0.307
German	13.000	7.692	0.616	0.379
Hebrew	3.000	7.500	0.356	0.127
Hindi	10.000	6.760	1.055	1.112
Hungarian	1.000	7.100	0.000	0.000
Icelandic	1.000	6.900	0.000	0.000
Indonesian	2.000	7.900	0.300	0.090
Italian	7.000	7.186	1.070	1.144
Japanese	12.000	7.625	0.861	0.742
Kazakh	1.000	6.000	0.000	0.000
Korean	5.000	7.700	0.510	0.260
Mandarin	14.000	7.021	0.738	0.545
Maya	1.000	7.800	0.000	0.000
Mongolian	1.000	7.300	0.000	0.000
None	1.000	8.500	0.000	0.000
Norwegian	4.000	7.150	0.497	0.247
Persian	3.000	8.133	0.450	0.202
Portuguese	5.000	7.760	0.875	0.766
Romanian	1.000	7.900	0.000	0.000
Russian	1.000	6.500	0.000	0.000
Spanish	26.000	7.050	0.810	0.656
Swedish	1.000	7.600	0.000	0.000
Telugu	1.000	8.400	0.000	0.000
Thai	3.000	6.633	0.368	0.136
Vietnamese	1.000	7.400	0.000	0.000
Zulu	1.000	7.300	0.000	0.000
(blank)				
Grand Total	3786.000	6.462	1.057	1.118

Director Analysis

Row Labels	Count of Director_Name	Average of Imdb_Score
Akira Kurosawa	1	8.7
Bryan Singer	1	8.6
Charles Chaplin	1	8.6
Christopher Nolan	3	8.8
David Fincher	2	8.7
Fernando Meirelles	1	8.7
Francis Ford Coppola	2	9.1
Frank Darabont	1	9.3
George Lucas	1	8.7
Hayao Miyazaki	1	8.6
Irvin Kershner	1	8.8
Jonathan Demme	1	8.6
Lana Wachowski	1	8.7
Martin Scorsese	1	8.7
Milos Forman	1	8.7
Peter Jackson	3	8.8
Quentin Tarantino	1	8.9
Robert Zemeckis	1	8.8
Sergio Leone	1	8.9
Steven Spielberg	2	8.75
Tony Kaye	1	8.6
(blank)		
Grand Total	28	8.778571429

Percentile	9.3



Point	Column1	Rank	Percent
8	9.3	1	100.00%
7	9.1	2	95.00%
17	8.9	3	85.00%
19	8.9	3	85.00%
11	8.8	5	70.00%
16	8.8	5	70.00%
18	8.8	5	70.00%
4	8.8	8	65.00%
20	8.75	9	60.00%
1	8.7	10	25.00%
5	8.7	10	25.00%
6	8.7	10	25.00%
9	8.7	10	25.00%
13	8.7	10	25.00%
14	8.7	10	25.00%
15	8.7	10	25.00%
2	8.6	17	0.00%
3	8.6	17	0.00%
10	8.6	17	0.00%
12	8.6	17	0.00%
21	8.6	17	0.00%

Budget Analysis

	CORRELATION BETWEEN MOVIE BUDGETS AND GROSS EARNINGS						Gross	Budget
						Gross	1	
						Budget	0.09656892	1
	MOVIES WITH THE HIGHEST PROFIT MARGIN				523505847			

06

Bank Loan Case Study

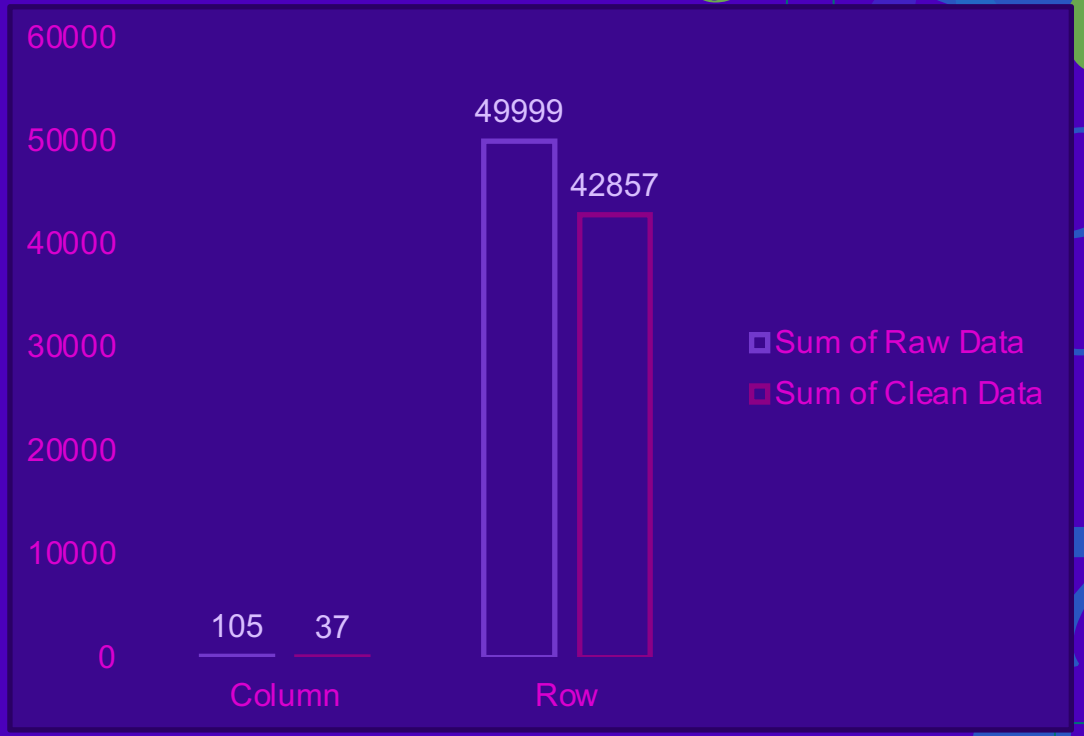
Final Project-2

DATA ANALYTICS TASKS:

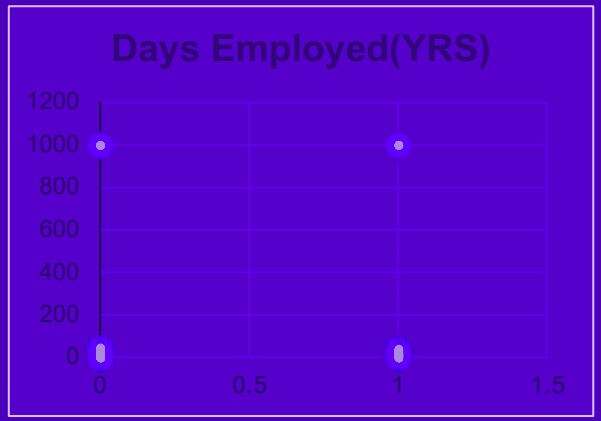
1st Task	2nd Task	3rd Task	4th Task	5th Task
• Identify Missing Data and Deal with it Appropriately	• Identify Outliers in the Dataset	• Analyze Data Imbalance	• Perform Univariate, Segmented Univariate, and Bivariate Analysis	• Identify Top Correlations for Different Scenarios

Identify Missing Data and Deal with it Appropriately

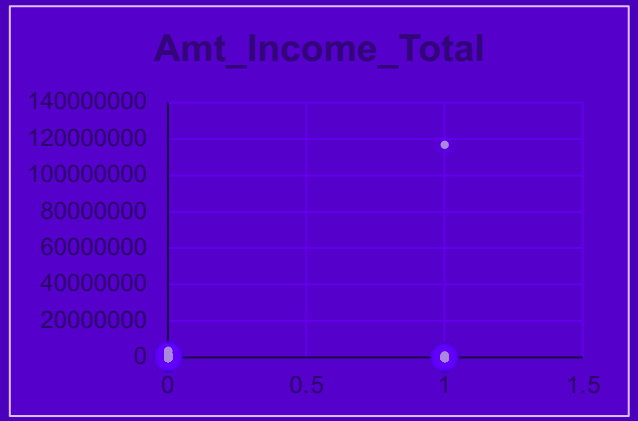
	Sum of Raw Data	Sum of Clean Data
Column	105	37
Row	49999	42857
Grand Total	50104	42894



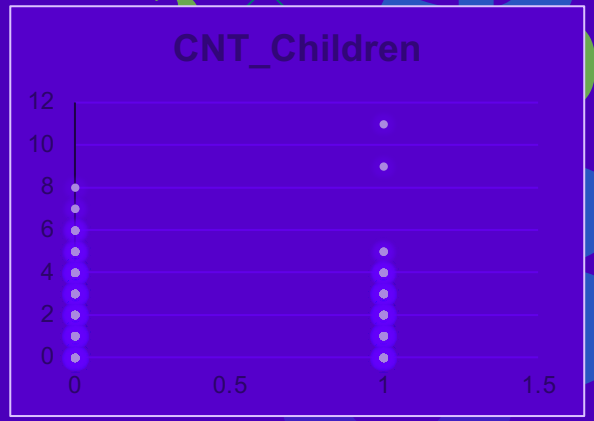
Identify Outliers in the Dataset



Quartile - 1
2.652054795
Quartile - 3
202500
Inter Quartile Range
202497.3479
UPPER LIMIT
506246.0219
Lower Limit
-303743.3699



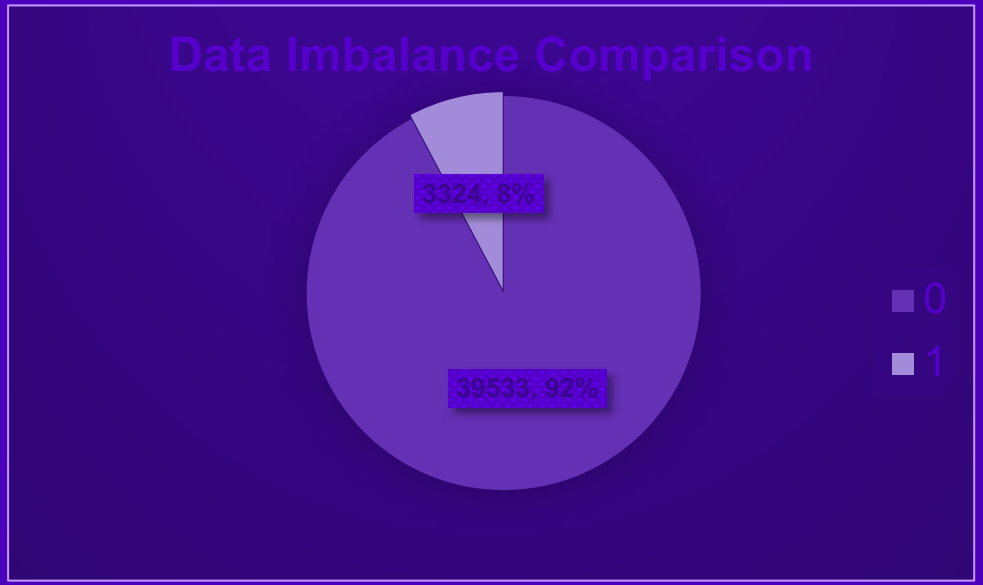
Quartile - 1
112500
Quartile - 3
202500
Inter Quartile Range
90000
UPPER LIMIT
337500
Lower Limit
-22500



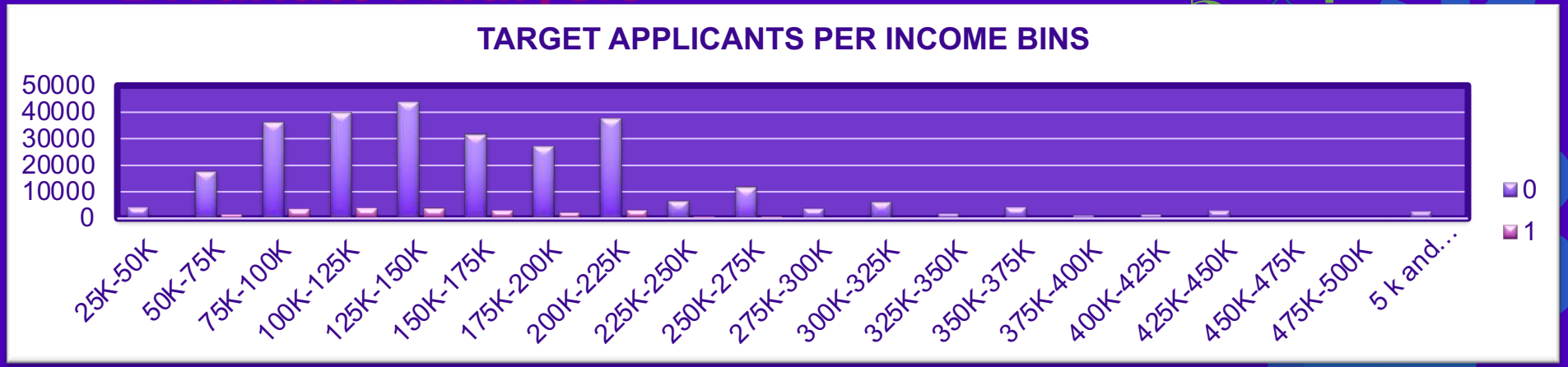
Quartile - 1
0
Quartile - 3
1
Inter Quartile Range
1
UPPER LIMIT
2.5
Lower Limit
-1.5

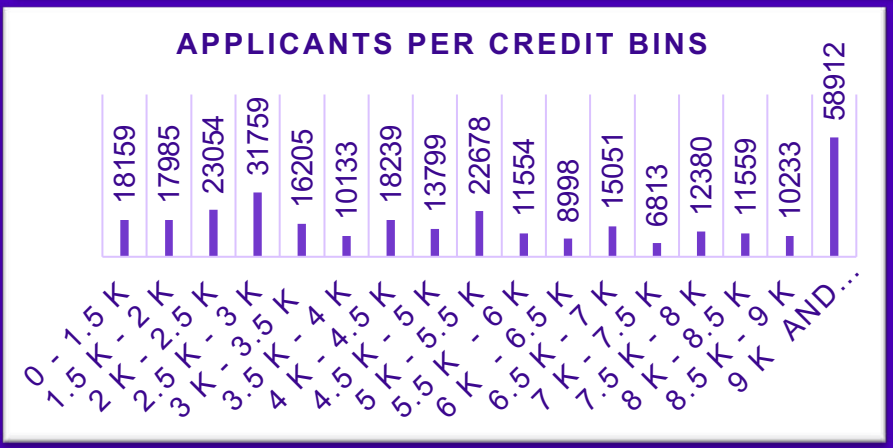
Analyze Data Imbalance

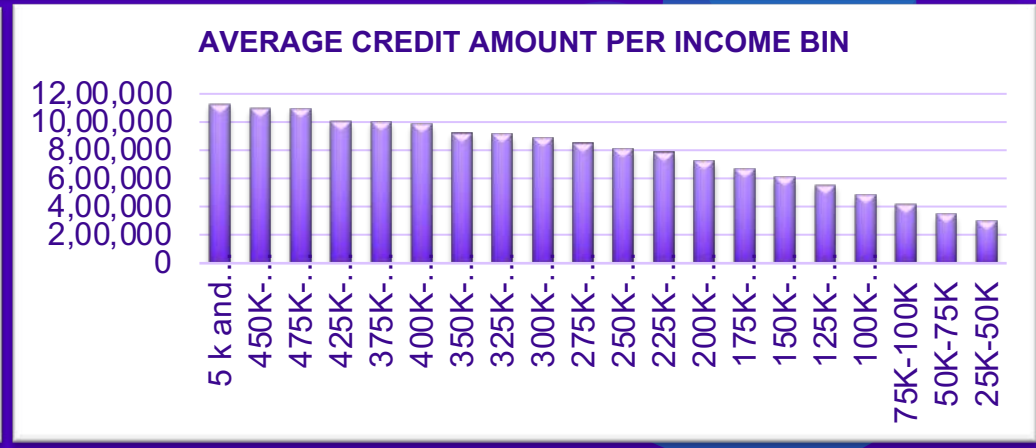
Row Labels	Count of TARGET	RATIO	CONTRIBUTION
0	39533	11.89320096	92.24
1	3324		7.76
Grand Total	42857		100.00



Perform Univariate, Segmented Univariate, and Bivariate Analysis







Identify Top Correlations for Different Scenarios

TARGET 0

CORRELATION FOR TIMELY PAYMENTS

CNT Of Children	1	0.009300239	0.004117726	-0.025043214	-9.7283E-07	0.002368016	0.003949962	0.026733198
AMT_Income_Total	0.009300239	1	0.063511987	0.026543706	-0.000502928	-0.003429977	0.003573112	-0.035127419
AMT_Credit	0.004117726	0.063511987	1	0.098803701	-0.004437401	-0.000222036	-0.004349401	-0.102897278
Region Population Relative	-0.025043214	0.026543706	0.098803701	1	-0.003303866	-0.006023773	0.000844092	-0.527253308
Days_Birth(Yrs)	-9.7283E-07	-0.000502928	-0.004437401	-0.003303866	1	0.523388736	0.50850239	0.003170646
Days_Employed(YRS)	0.002368016	-0.003429977	-0.000222036	-0.006023773	0.523388736	1	0.296604311	0.001543287
Days_ID_Publish(YRS)	0.003949962	0.003573112	-0.004349401	0.000844092	0.50850239	0.296604311	1	0.003088812
Region_Rating_Client	0.026733198	-0.035127419	-0.102897278	-0.527253308	0.003170646	0.001543287	0.003088812	1
	CNT Of Chindren	AMT_Income_Total	AMT_Credit	Region_Population_Relative	Days_Birth(YRS)	Days_Employed(YRS)	Days_ID_Publish(YRS)	Region_Rating_Client

TARGET 1

Correlation For Timely Payments

CNT Of Children	1	0.009300239	0.004117726	-0.025043214	-9.7283E-07	0.002368016	0.003949962	0.026733198
AMT_Income_Total	0.009300239	1	0.063511987	0.026543706	-0.000502928	-0.003429977	0.003573112	-0.035127419
AMT_Credit	0.004117726	0.063511987	1	0.098803701	-0.004437401	-0.000222036	-0.004349401	-0.102897278
Region Population Relative	-0.025043214	0.026543706	0.098803701	1	-0.003303866	-0.006023773	0.000844092	-0.527253308
Days_Birth(Yrs)	-9.7283E-07	-0.000502928	-0.004437401	-0.003303866	1	0.523388736	0.50850239	0.003170646
Days_Employed(YRS)	0.002368016	-0.003429977	-0.000222036	-0.006023773	0.523388736	1	0.296604311	0.001543287
Days_ID_Publish(YRS)	0.003949962	0.003573112	-0.004349401	0.000844092	0.50850239	0.296604311	1	0.003088812
Region_Rating_Client	0.026733198	-0.035127419	-0.102897278	-0.527253308	0.003170646	0.001543287	0.003088812	1
	CNT_Of_Children	AMT_Income_Total	AMT_Credit	Region_Population_Relative	Days_Birth(YRS)	Days_Employed(YRS)	Days_ID_Publish(YRS)	Region_Rating_Client

- As the Age and Years of Experience Increase, the chances of defaulting decrease. So the bank should prioritize older And Experienced Clients.
- Educated Clients tend to default lesser compared to clients with lower education such as Lower Secondary & Secondary education
- Male clients tend to default more than female clients.
- Corporate clients are a safer bet compared to labour class clients.
- People belonging to Region Rating 3 have the highest % of defaulters, bank could formulate a stricter loan policy m the clients from these region. Clients from Region 1 are the safest bet
- We also notice that as the Age increases, the amount taken by the clients is considerably higher and since the default percentage with higher age is lower these should be least risky and highly profitable clients for the bank.

07

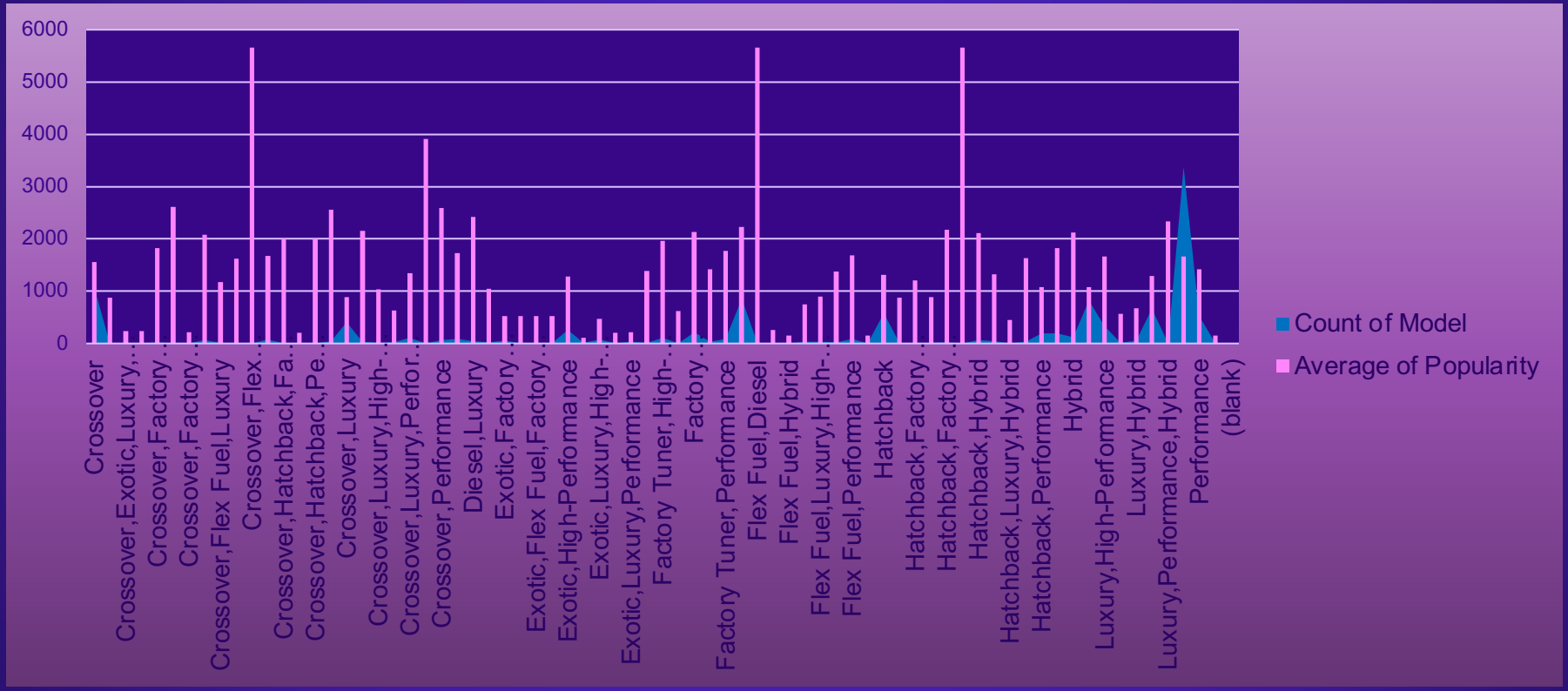
Impact of Car Features

Final Project-3

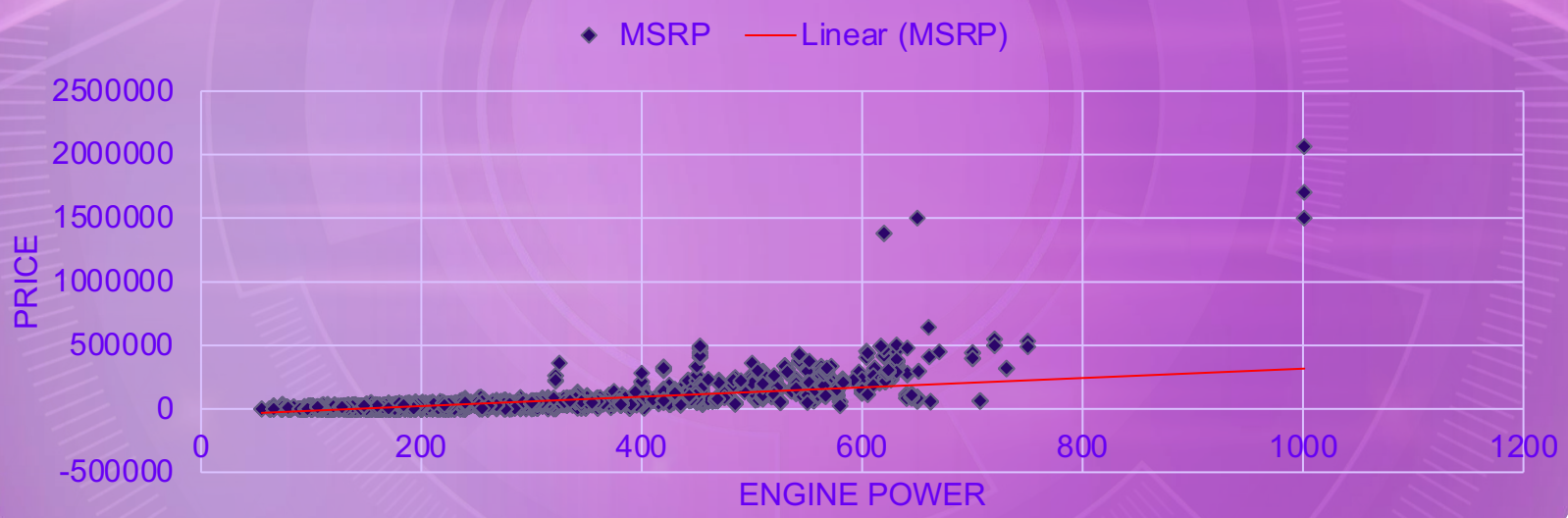
TASKS: ANALYSIS

TASK 1	TASK 2	TASK 3	TASK 4	TASK 5
• How does the popularity of a car model vary across different market categories?	• What is the relationship between a car's engine power and its price?	• Which car features are most important in determining a car's price?	• How does the average price of a car vary across different manufacturers?	• What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

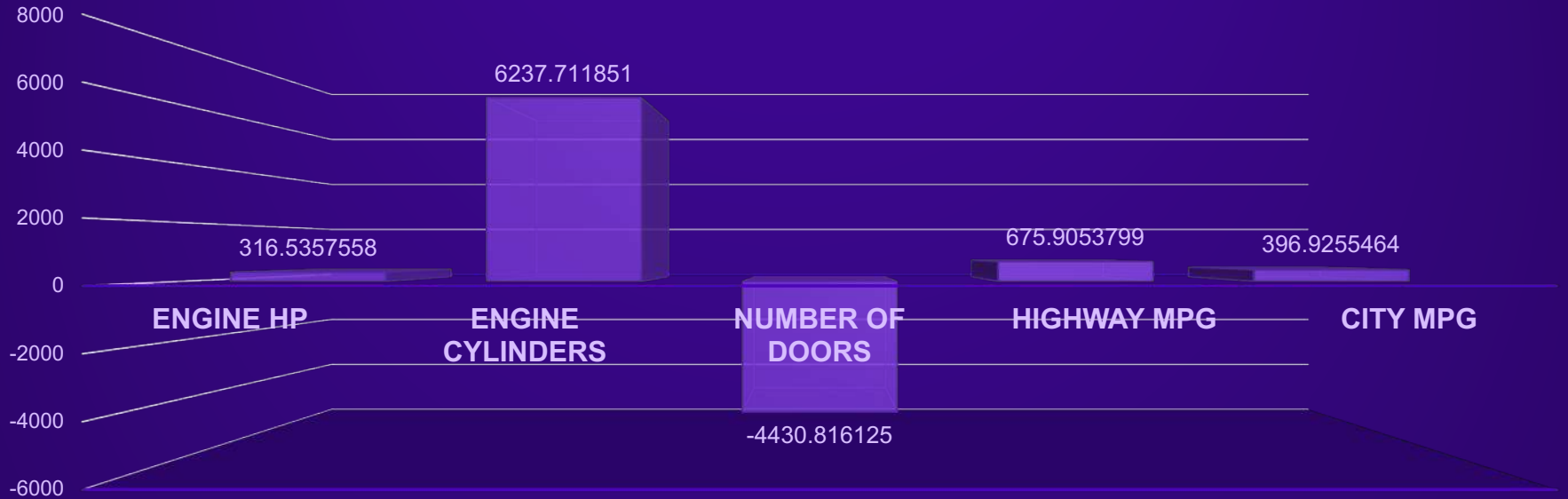
Task 1.B: Create a combo chart that visualizes the relationship between market category and popularity.



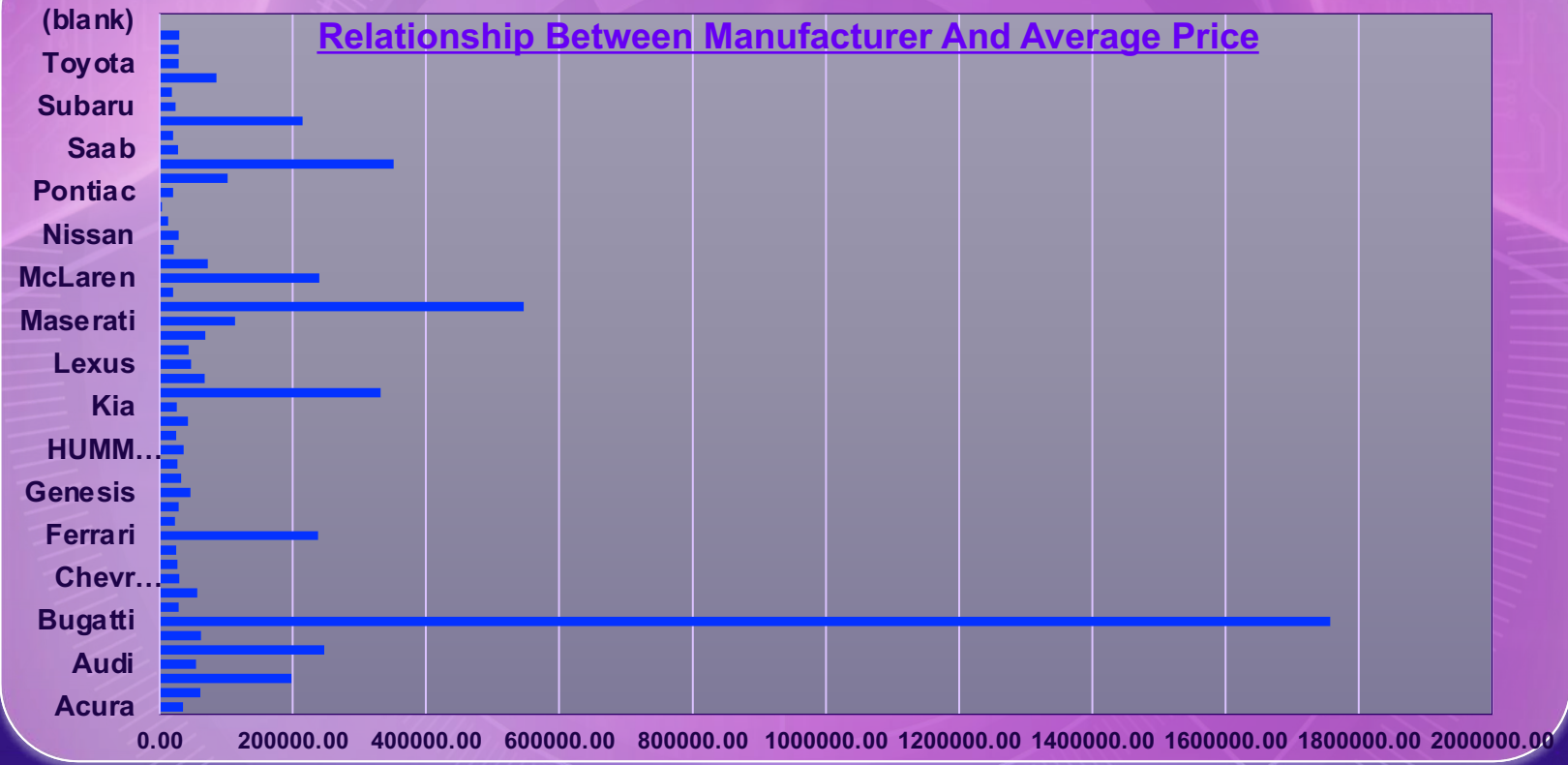
Task 2: Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.



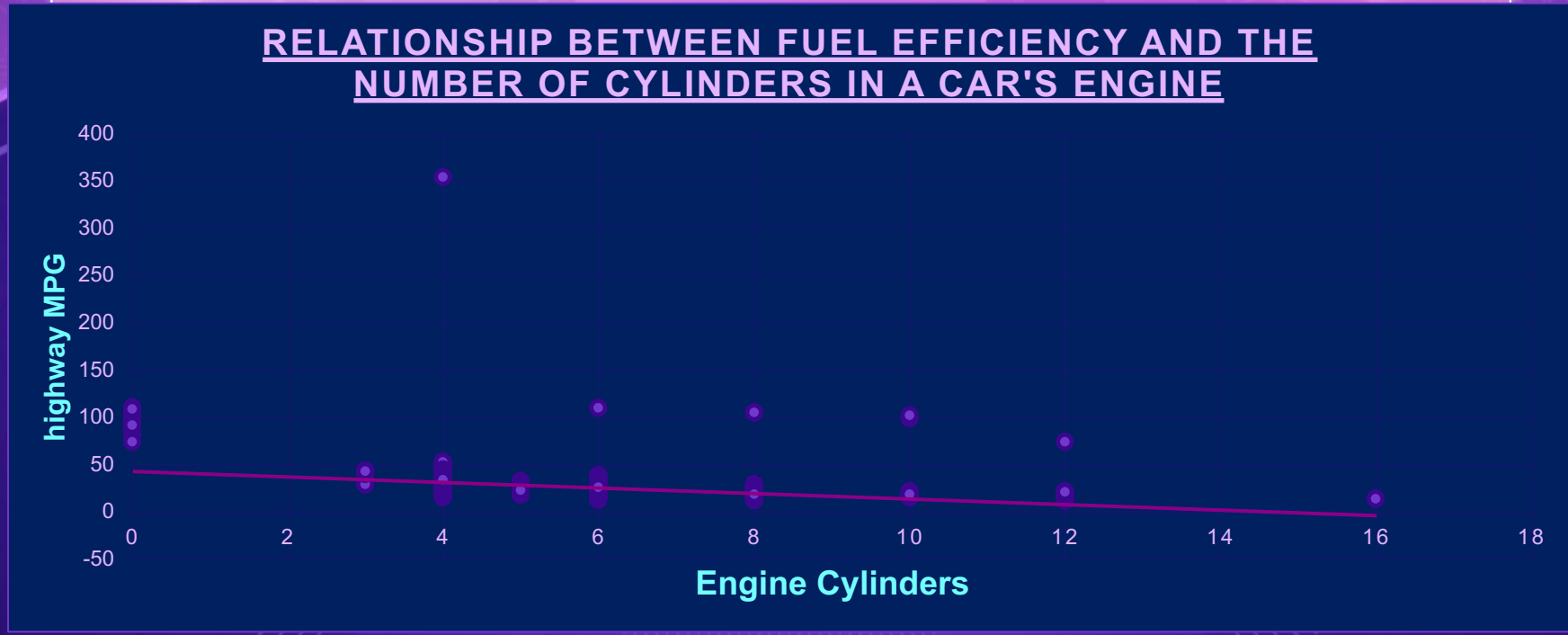
Coefficient Values For Each Variable



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



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.



The "Impact of Car Features" paper examines how various car features affect their price and profitability. Highlights include:

Objective: To understand consumer motivations by analyzing vehicle types, market categories, and prices using data analytics techniques.

Methodology: Data cleaning, regression analysis and chart creation in Excel.

Key findings:

- Popularity: The popularity of cars varies greatly from market to market.
- Engine Power vs. Engine Power Cost: More engine cylinders increase cost, while more doors lower it.
- Important Factors: Engine cylinders, highway MPG, and city MPG have a big impact on car prices.
- Product prices: Luxury brands like Bugatti and Rolls-Royce have the highest prices.
- Fuel Consumption: Increased engine output is generally associated with lower fuel efficiency.

Overall, the study helps manufacturers develop pricing strategies and focus on factors that enhance market competitiveness and profitability.

08

ABC Call Volume Trend

Final Project-4

INSIGHTS TASKS

Average Call Duration

Objective: The Average Duration of all Incoming Calls Received By Agents

Call Volume Analysis

Objective: Create a Chart or Graph That Shows the Number of Calls Received in Each Time Bucket?

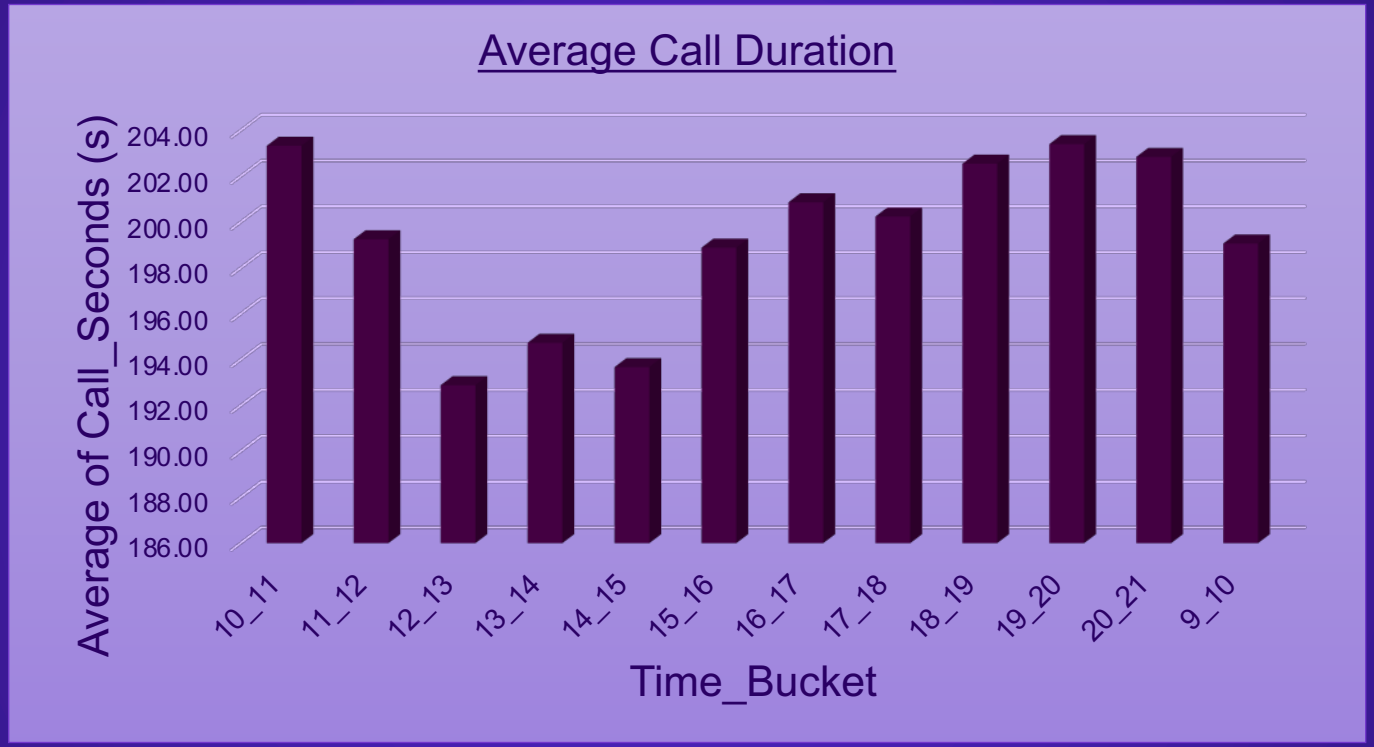
Manpower Planning

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

Duplicate Rows Detection

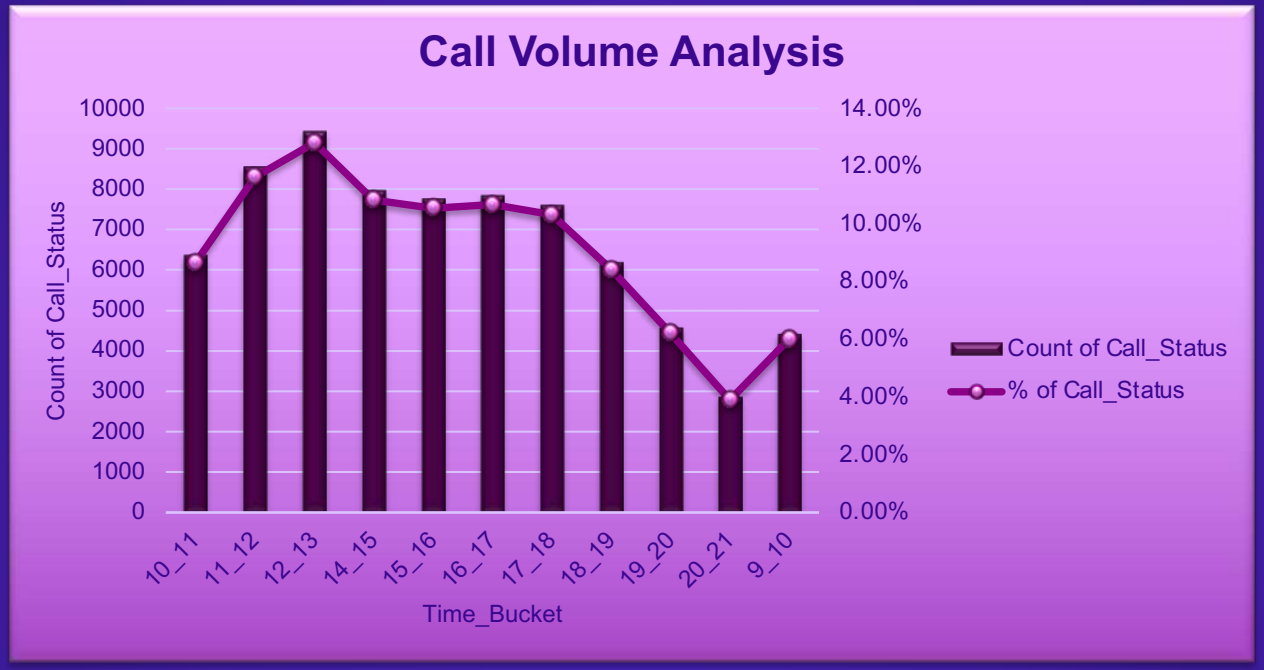
Objective: Propose a manpower plan for each time bucket throughout the day, keeping the maximum abandon rate at 10%

The Average Duration of all Incoming Calls Received By Agents



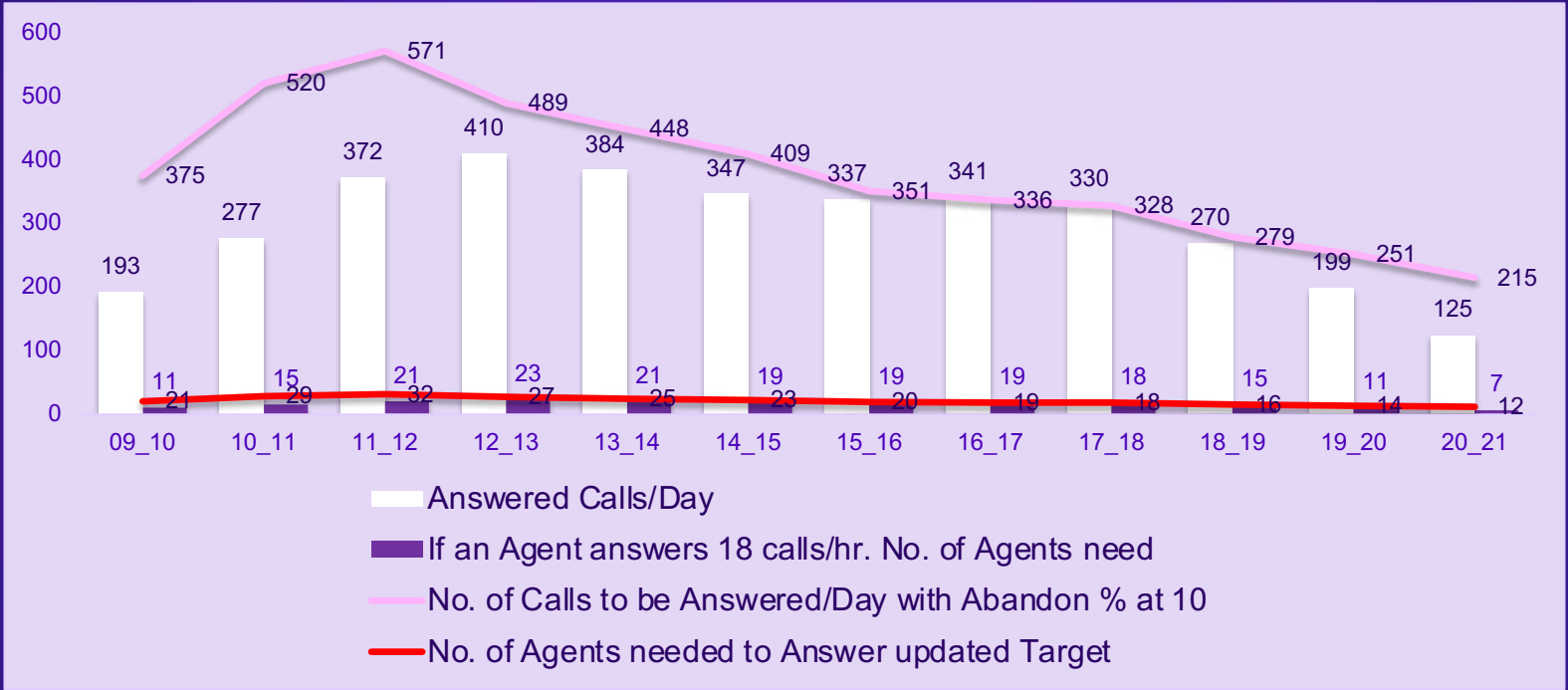
- An Overall Increasing Trend from 9am to 9pm with average duration of 198.62 seconds
- lowest during 12pm to 1pm slot followed by 2pm to 3pm then 1pm to 2pm
- Longest duration during 10am to 11am followed by 8pm to 9pm then 7pm to 8pm
- In morning hours from 9 am to 12 pm and from 6pm to 9pm the call duration is highest

Create a Chart or Graph That Shows the Number of Calls Received in Each Time Bucket?



- The Call volume follows a left skewed bell curve, with the 9588 at 9am to 10am peaking at 11 to 12 with 14626 then continuously declining to 5505 in 8pm to 9pm slot
- Overall decreasing trend is followed
- During the initial number of hours large number of calls are abandoned, and during the last hour large number of calls are abandoned in comparison to the call answered

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



1. **Calls answered/day: Varies from hour to hour, peaking at 410 calls between 12-13 hours.**
2. **Agents Required: Calculated based on agents handling 18 calls per hour, ranging from 7 to 23 agents depending on the time bucket.**
3. **Target calls with a 10% abandonment rate: Changes the number of calls that must be answered to maintain a 10% abandonment rate, which means more targets.**
4. **Update Agents Required: Increases the number of agents required to meet the update call goal, displayed from 12 to 32 agents.**

Propose a manpower plan for each time bucket throughout the day, keeping the maximum abandon rate at 10%

This chart shows the number of employees required to work night shifts in a call center. The data points represent buckets of time at night, indicating the number of laborers required.
This visualization helps to understand staffing needs throughout the night, ensuring that the best staffing is allocated at different times

1. **Peak Hours**: The highest demand for agents occurs between 11:00 and 12:00, requiring 32 agents. Other high-demand periods include 10:00-11:00 (29 agents) and 12:00-13:00 (27 agents).
2. **Off-Peak Hours**: During the night and early morning (01:00-05:00), the required number of agents drops significantly to as low as 3 agents.
3. **Moderate Demand**: Evening hours and early morning (05:00-09:00) show a moderate demand, with the number of agents needed ranging from 8 to 13.
4. **Transition Periods**: The demand gradually decreases after the peak, with a steady need for 8 to 18 agents from 17:00 to 22:00.

This project focuses on analyzing a company's inbound calling team data for customer experience (CX) analysis. The dataset includes information such as agent name and ID, queue time, call time, duration, and status. The CX team uses AI-powered tools like IVR, RPA, Predictive Analytics, and Intelligent Routing to enhance the customer experience. The analysis reveals insights such as the overall increasing trend of calls from 9 am to 9 pm, lowest average duration during 12 pm to 1 pm, and highest call duration during morning hours and evening hours. The visualization helps understand staffing needs and transition periods in the call center.

In conclusion, the challenge presents valuable insights into call extent tendencies, common name period, and staffing necessities for keeping a 10% abandon price. The utilization of excel gear and the thorough information analysis show a comprehensive method to improving the purchaser experience and optimizing the performance of the inbound calling team.

09

APPENDIX

Links for excel

LINK TO PROJECT REPORTS

1. DATA ANALYTICS PROCESS - click here
2. INSTAGRAM USER ANALYTICS - click here
3. OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE - click here
4. HIRING PROCESS ANALYTICS - click here
5. IMDB MOVIE ANALYSIS - click here
6. BANK LOAN CASE STUDY - click here
9. ANALYZING THE IMPACT OF CAR FEATURES - click here
10. ABC CALL VOLUME TREND ANALYSIS - click here

THANK YOU