



Data Analysis Portfolio

Created by Vivek G

Contents

PROFESSIONAL BACKGROUND	2
INSTAGRAM DATA ANALYTICS.....	3
Project description.....	3
Findings	3
Conclusions	6
Operation Analytics and Investigating Metric Spike.....	7
Project description.....	7
Insights Driven From The case study	8
Hiring process analytics	10
Project description.....	10
Insights	10
Conclusions:	13
IMDB Movie Analysis	14
Project description.....	14
Insights	14
Conclusion.....	16
Bank loan case Study	17
Introduction	17
Business Objectives:.....	17
Data Analytics Tasks:.....	17
Conclusions	28
Analyzing the Impact of Car Features on Price and Profitability	29
Project Overview:.....	29
Business Objective:	29
Insight Required:.....	29
Interactive dashboard	32
ABC Call Volume Trend Analysis	35
Description:.....	35
Data Analytics:	36
Conclusions:	40
LEARNINGS :	40

PROFESSIONAL BACKGROUND

Detail oriented engineering graduate, My goal is to kickstart a career as a data analyst /Business analyst . After exploring entrepreneurship for a while, I've come to realize that my true passion lies in the world of data. Now, I'm actively seeking an internship or entry-level position. In my skillset, you'll find proficiency in SQL, Excel and python, as well as hands-on experience with Tableau. I'm genuinely enthusiastic about diving into data, whether it's analysing numbers or crafting compelling visuals. During my tenure as a Store Manager Products, I gained valuable insights into business operations and improved my communication skills. Yet, what really sets me apart is my Course work and projects. I hold a degree in Electronics and Communication Engineering, providing a strong technical foundation for my data endeavors. I'm eager to further immerse myself in the realm of data analysis. Let's connect and explore how I can contribute my data-driven skills to your team.

INSTAGRAM DATA ANALYTICS

Instagram User Analytics project using MySQL workbench to draw useful insights

Project description

The goal of this project is to analysing user interactions and engagement with the instagram app to provide valuable insights that can help the business grow.

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

In this project, i'll be using SQL and mySQL workbench as the tool to analyze instagram user data and answer questions posed by the management team. My insights will help the product manager and the rest of the team make informed decisions about the future direction of the instagram app.

Findings

A) Marketing Analysis:

1. **Loyal User Reward:** Identify the five oldest users on Instagram from the provided database.

	username	created_at
▶	Darby_Herzog	2016-05-06 00:14:21
	Emilio_Bernier52	2016-05-06 13:04:30
	Elenor88	2016-05-08 01:30:41
	Nicole71	2016-05-09 17:30:22
	Jordyn.Jacobson2	2016-05-14 07:56:26

2. **Inactive User Engagement:** Identify users who have never posted a single photo on Instagram.

	id	username	created_at	id	image_url	user_id	created_dat
▶	5	Aniya_Hackett	2016-12-07 01:04:39	NULL	NULL	NULL	NULL
	7	Kasandra_Homenick	2016-12-12 06:50:08	NULL	NULL	NULL	NULL
	14	Jadlyn81	2017-02-06 23:29:16	NULL	NULL	NULL	NULL
	21	Rocio33	2017-01-23 11:51:15	NULL	NULL	NULL	NULL
	24	Maxwell.Halvorson	2017-04-18 02:32:44	NULL	NULL	NULL	NULL
	25	Tierra.Trantow	2016-10-03 12:49:21	NULL	NULL	NULL	NULL
	34	Pearl7	2016-07-08 21:42:01	NULL	NULL	NULL	NULL
	36	Ollie_Ledner37	2016-08-04 15:42:20	NULL	NULL	NULL	NULL
	41	Mckenna17	2016-07-17 17:25:45	NULL	NULL	NULL	NULL
	45	David.Osinski47	2017-02-05 21:23:37	NULL	NULL	NULL	NULL
	49	Morgan.Kassulke	2016-10-30 12:42:31	NULL	NULL	NULL	NULL
	53	Linnea59	2017-02-07 07:49:34	NULL	NULL	NULL	NULL
	54	Duane60	2016-12-21 04:43:38	NULL	NULL	NULL	NULL
	57	Julien_Schmidt	2017-02-02 23:12:48	NULL	NULL	NULL	NULL
	66	Mike.Auer39	2016-07-01 17:36:15	NULL	NULL	NULL	NULL
	68	Franco_Keebler64	2016-11-13 20:09:27	NULL	NULL	NULL	NULL
	71	Nia_Haag	2016-05-14 15:38:50	NULL	NULL	NULL	NULL
	74	Hulda.Macejkovic	2017-01-25 17:17:28	NULL	NULL	NULL	NULL
	75	Leslie67	2016-09-21 05:14:01	NULL	NULL	NULL	NULL
	76	Janelle.Nikolaus81	2016-07-21 09:26:09	NULL	NULL	NULL	NULL
	80	Darby_Herzog	2016-05-06 00:14:21	NULL	NULL	NULL	NULL

3. **Contest Winner Declaration:** Determine the winner of the contest and provide their details to the team.

	photo_id	username	noflikes
▶	145	Zack_Kemmer93	48
	127	Malinda_Streich	43
	182	Adelle96	43
	123	Seth46	42
	30	Presley_McClure	41
	52	Annalise.Mckenzie16	41
	61	Delpha.Kihn	41
	147	Meggie_Doyle	41
	174	Elenor88	41
	192	Kathryn80	41
	256	Javonte83	41
	13	Harley_Lind18	40
	97	Irwin.Larson	40
	153	Aurelie71	40

4. **Hashtag Research:** Identify and suggest the top five most commonly used hashtags on the platform.

tag_name	ht
smile	59
beach	42
party	39
fun	38
concert	24

5. **Ad Campaign Launch:** Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

wofday	count(username)
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

B) Investor Metrics:

1. **User Engagement:** Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

	totalphotos	totalusers	peruser
▶	257	100	2.5700

2. **Bots & Fake Accounts:** Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

	username	likes
▶	Aniya_Hackett	257
	Bethany20	257
	Duane60	257
	Jadyn81	257
	Janelle.Nikolaus81	257
	Julien_Schmidt	257
	Leslie67	257
	Maxwell.Halvorson	257
	Mckenna17	257
	Mike.Auer39	257
	Nia_Haag	257
	Ollie_Ledner37	257
	Rocio33	257

Conclusions

- Marketing team can reward most loyal customers to increase user activity
- Send promotional emails use popular hashtags and active days for brand promotions
- Can increase insights from average photos per user to account for activity
- Fake bots can be removed from the platform to increase user experience

Operation Analytics and Investigating Metric Spike

Advanced SQL

Project description

Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. Working closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect. One of the key aspects of Operational Analytics is investigating metric spikes. This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales.

Tasks:

A. Jobs Reviewed Over Time:

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

B. Throughput Analysis:

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

C. Language Share Analysis:

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

D. Duplicate Rows Detection:

- Objective: Identify duplicate rows in the data.

Case Study 2: Investigating Metric Spike

- **users**: Contains one row per user, with descriptive information about that user's account.
- **events**: Contains one row per event, where an event is an action that a user has taken (e.g., login, messaging, search).
- **email_events**: Contains events specific to the sending of emails.

Tasks:

A. Weekly User Engagement:

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

B. User Growth Analysis:

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

C. Weekly Retention Analysis:

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

D. Weekly Engagement Per Device:

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

E. Email Engagement Analysis:

- Objective: Analyze how users are engaging with the email service.

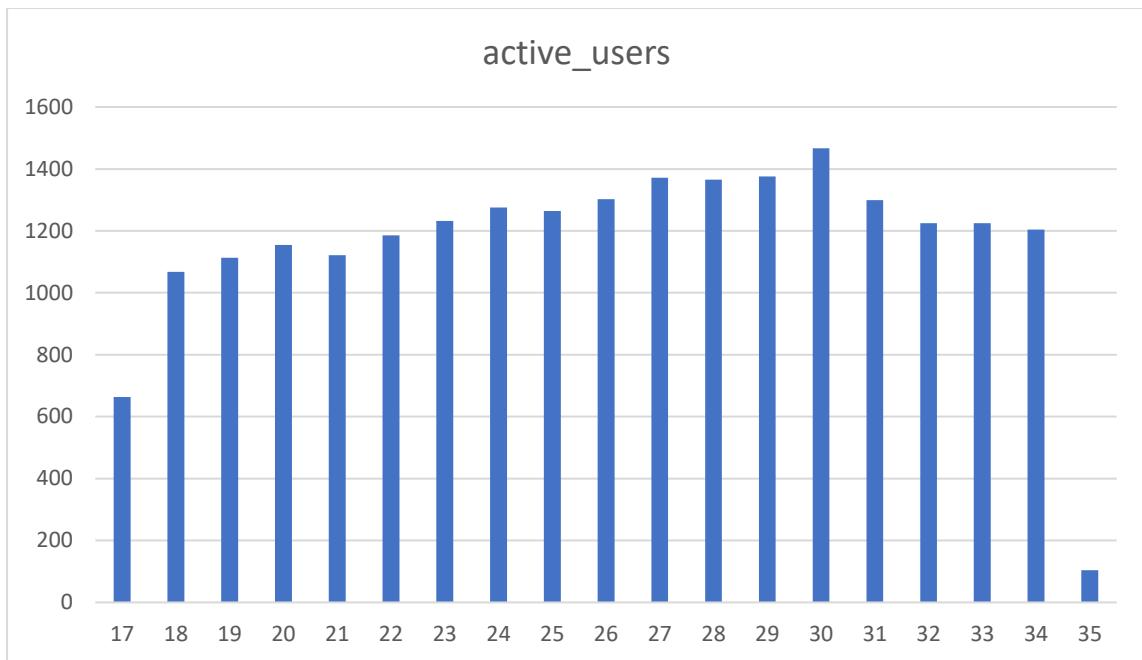
Insights Driven From The case study

Case Study 1 (Job Data):

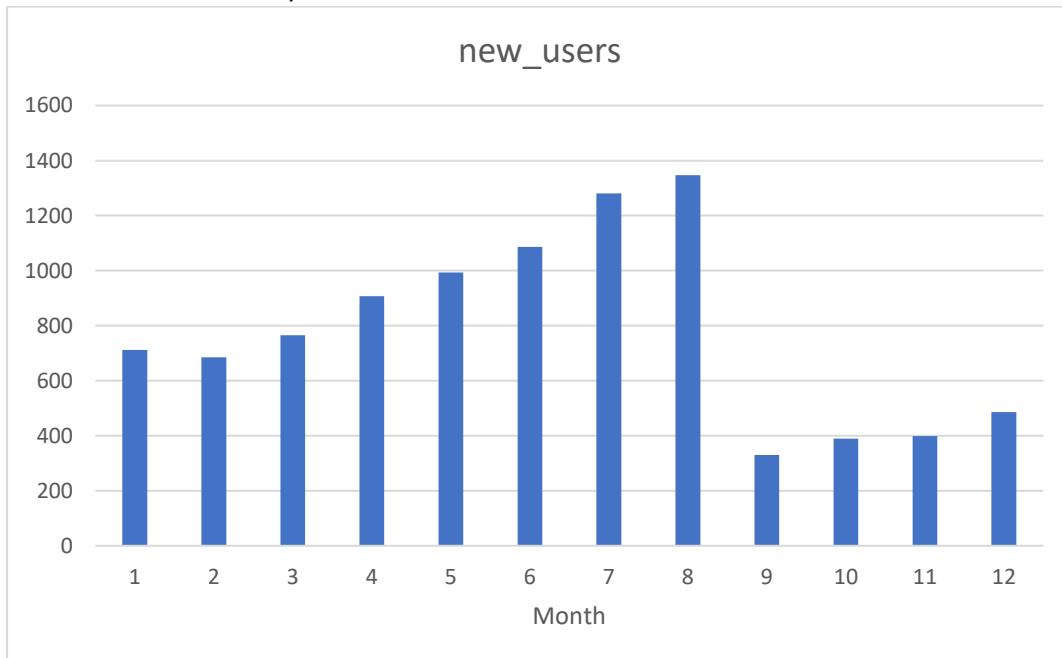
- The number of distinct jobs reviewed per hour per day for November 2020 is 83%.
- We used the 7-day rolling average of throughput as it gives the average for all the days from day 1 to day 7 whereas, daily metric gives the average for only that particular day its
- The percentage share of Persian language is the most (37.5%).
- There are two duplicate rows if we partition the data by job_id. But if we look the overall columns, all the rows are unique.

.Case Study 2 (Investigating metric spike):

- The weekly user engagement increased from week 18th to week 31st and then start declining from then onwards. This means that some of the users do not find much quality the product/service in the last of the weeks.



- The number of users increases towards the 8 th month then fall down and keep going constant towards the end of the year



Hiring process analytics

Project description

The main objective of this project is to analyze the company's hiring process data and draw meaningful insights from it. The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department.

Insights

A. Hiring Analysis:

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

Males hired	2563
Females Hired	1856
Don't want to say	268

B. Salary Analysis:

Average salary offered by this company

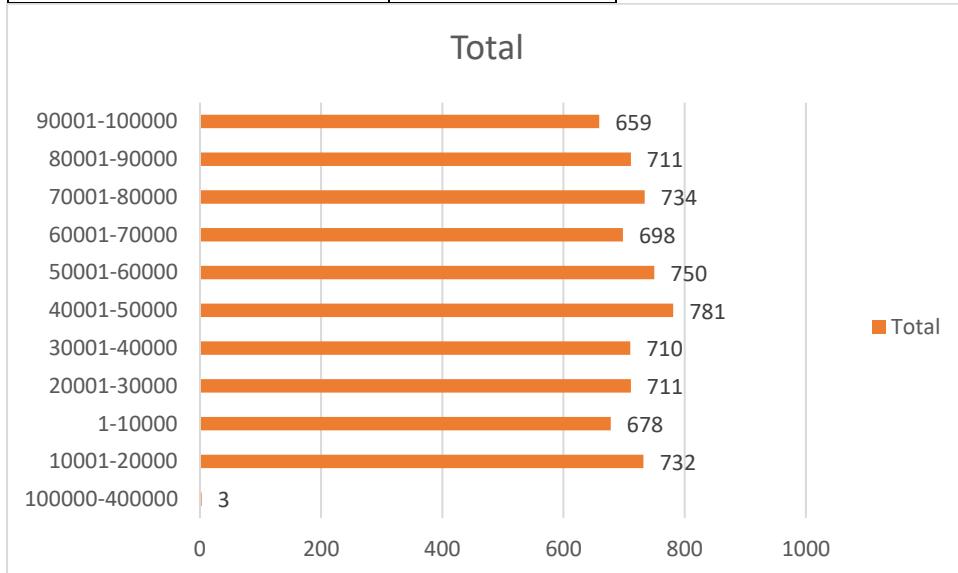
2	average salary offered	49983.03
---	------------------------	----------

C. Salary Distribution:

Create class intervals for the salaries in the company. To help understand the salary distribution.

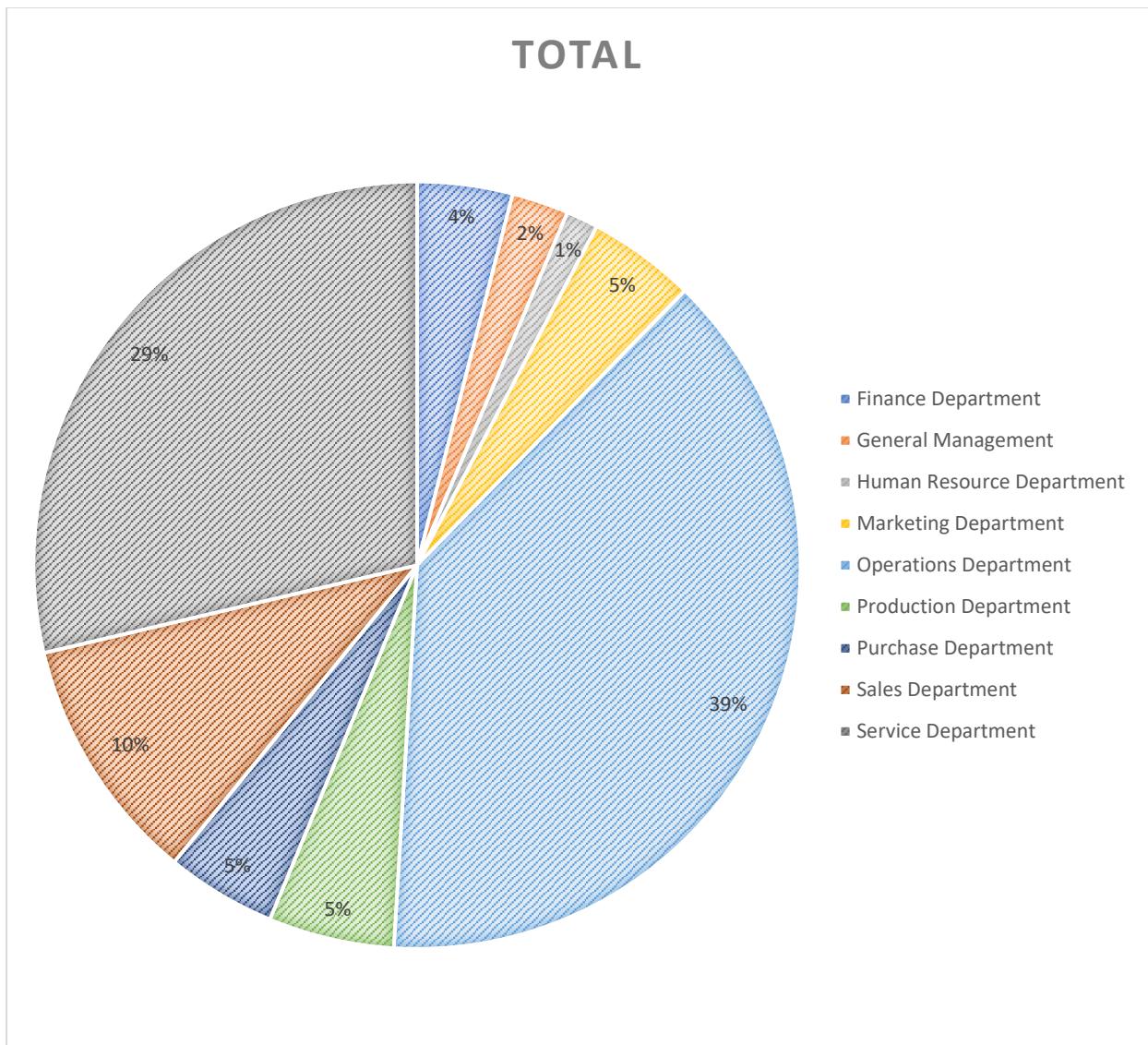
Salary Range	Number of employees in Salary Range
1-10000	678
10001-20000	732
20001-30000	711
30001-40000	710
40001-50000	781

50001-60000	750
60001-70000	698
70001-80000	734
80001-90000	711
90001-100000	659
100000-400000	3



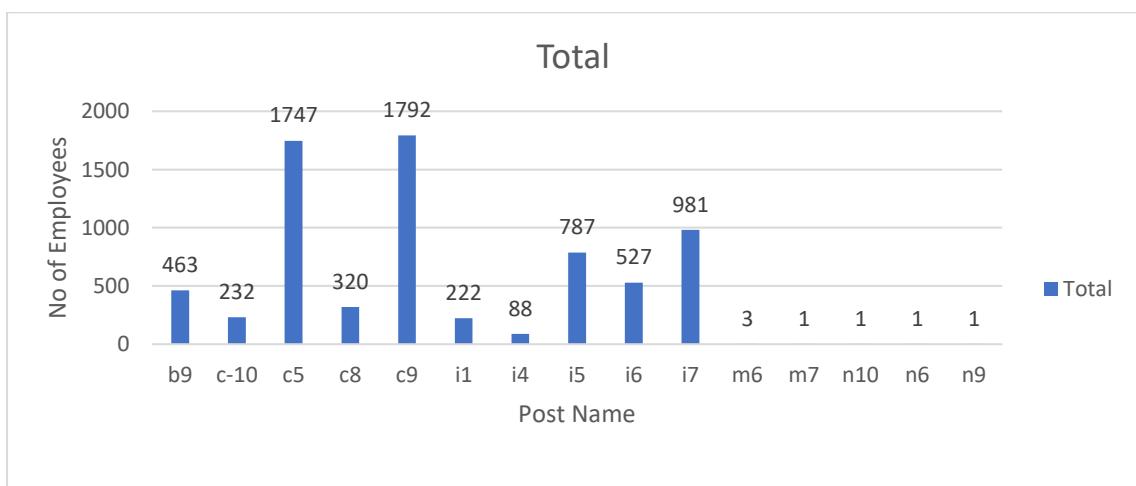
D. Departmental Analysis:

Visualization to show the proportion of people working in different departments.



E. Position Tier Analysis:

Graph to represent the different position tiers within the company. This will help understand the distribution of positions across different tiers.



Link to the Hiring Process Analytics excel

File:https://docs.google.com/spreadsheets/d/1tTITEINyTOCb8iaq7XLJaKKwZ_NdIMZc/edit?usp=sharing&ouid=100819582638973735490&rtpof=true&sd=true

Conclusions:

- ❖ Out of all the 4687 employees hired 54.68% are males and 39.59% are females ,5.7% don't want to say
- ❖ Average salary offered by the company during hiring is 49983.03
- ❖ Only 3 employees have a salary of above 100,000 while other salary ranges have between 600 to 800 employees
- ❖ Operations Department has the most no of employees in the company i.e ,nearly 39% and Service Department comes second with 29% the least no of employees are present in Human Resources Department
- ❖ Most employees are on the c9 post and least no of employees are on m7, n6, n10 and n9

IMDB Movie Analysis

Project description

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

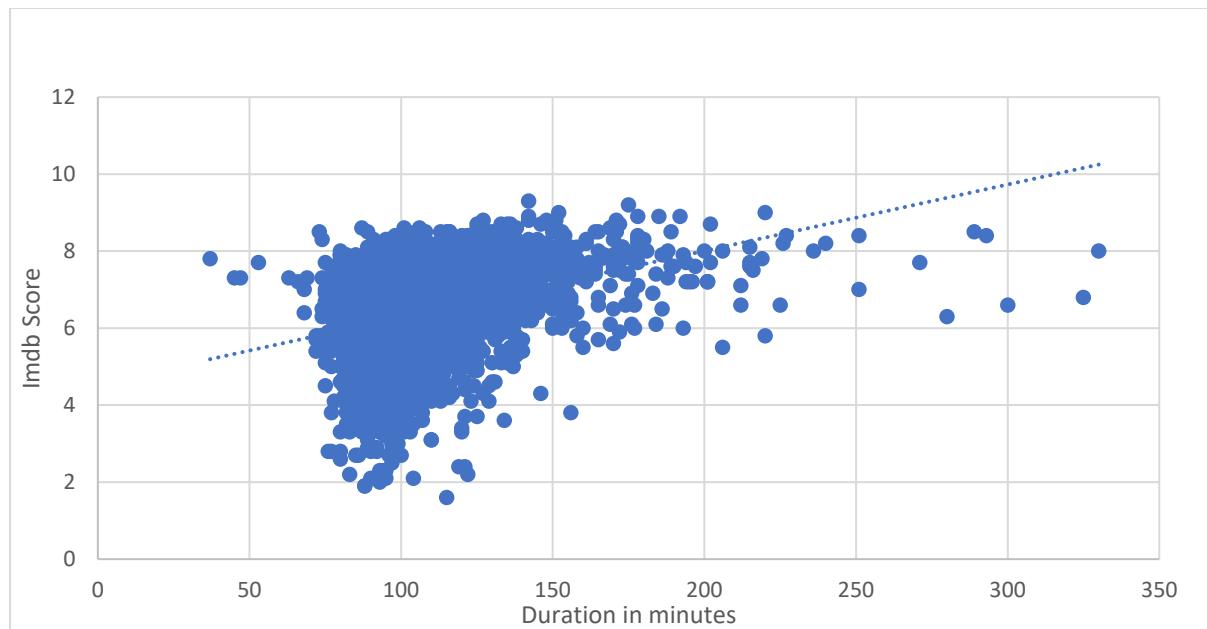
Insights

1. Top 5 movie genres and their statistical functions of Imdb ratings

genres names	No (genre)	Mean	Median	Mode	Variance	Standard Deviation	Max	Min
Comedy	1008	6.153869	6.3	6.4	1.1012549	1.049406914	8.8	1.9
Action	966	6.274948	6.3	6.6	1.1126436	1.054819241	9	2.1
Drama	685	6.828175	6.9	6.7	0.8172062	0.903994557	8.8	2.1
Adventure	372	6.547581	6.7	7.3	1.2503436	1.118187645	8.6	2.3
Crime	259	6.935521	7	7.4	0.7510162	.8611928	9.3	3.3

- Most common movie genre is Comedy
- Crime has lower standard deviation among the top 5 for IMDB ratings meaning the data points are closer to the mean

2. Analyze the distribution of movie durations and its impact on the IMDB score.



Mean	Median	Standard Deviation
109.9906	106	22.609668

3. Examine the distribution of movies based on their language.

Languages	No of movies in each language		Mean	Median	Standard deviation
English	3665	6.412988	6.5	6.5	1.064924055
French	34	7.355882	7.3	7.3	0.51173935
Spanish	24	7.0875	7.2	7.2	0.82427367
Mandarin	15	7.08	7.4	7.4	0.745832868
German	11	7.763636	7.8	7.8	0.644237008
Japanese	10	7.66	8	8	0.939361485

4. Influence of directors on movie ratings.

director_name	Average Imdb Score	Percentile
Akira Kurosawa	8.7	1
Tony Kaye	8.6	0.9988
Charles Chaplin	8.6	0.9988
Alfred Hitchcock	8.5	0.9964
Ron Fricke	8.5	0.9964
Damien Chazelle	8.5	0.9964
Majid Majidi	8.5	0.9964
Sergio Leone	8.433333333	0.9959
Christopher Nolan	8.425	0.9953
Richard Marquand	8.4	0.9941

5. Explore the relationship between movie budgets and their financial success.

movie_title	gross	budget	Profit Margin
Avatar	761M	237M	524M
Jurassic World	652M	150M	502M
Titanic	659M	200M	459M
Star Wars: Episode IV - A New Hope	461M	11M	450M
E.T. the Extra-Terrestrial	435M	11M	424M
The Avengers	623M	220M	403M
The Avengers	623M	220M	403M
The Lion King	423M	45M	378M
Star Wars: Episode I - The Phantom Menace	475M	115M	360M
The Dark Knight	533M	185M	348M

Correlation coefficient between Gross and Budget	0.0999745
Movie with Highest profit	Avatar
Profit Margin	524M

- There is a positive correlation between Gross and Budget of movies
- The correlation is closer to zero means there is weak correlation

Conclusion

- Comedy is the most popular movie genre and action comes second
- The trendline indicates a positive trend and 100 to 150 minutes are ideal for a higher imdb ratings
- English is the most common language due to more people understanding the language
- Akira kurosawa is the most popular director
- There is a positive correlation between the gross and the budget but the correlation is weak

Bank loan case Study

Introduction

Some customers who don't have a sufficient credit history take advantage of this and default on their loans. Using Exploratory Data Analysis (EDA) to analyze patterns in the data and ensure that capable applicants are not rejected.

When a customer applies for a loan, your company faces two risks:

1. If the applicant can repay the loan but is not approved, the company loses business.
2. If the applicant cannot repay the loan and is approved, the company faces a financial loss.

Business Objectives:

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

Data Analytics Tasks:

The dataset contains information about loan applications. It includes two types of scenarios:

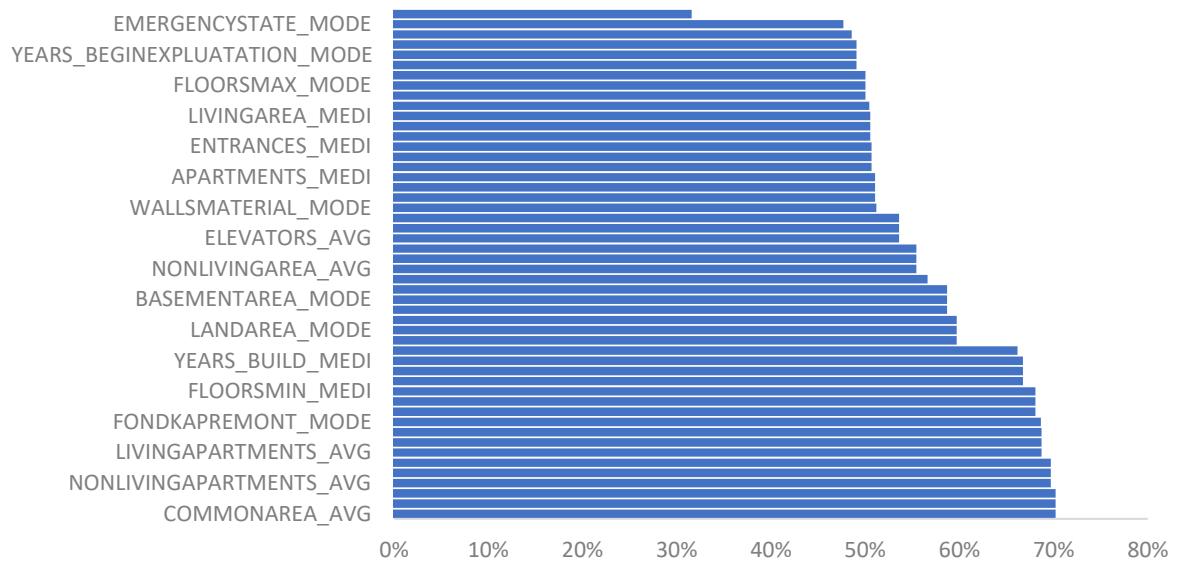
1. Customers with payment difficulties: These are customers who had a late payment of more than X days on at least one of the first Y installments of the loan.
2. All other cases: These are cases where the payment was made on time.

https://drive.google.com/file/d/18BI_bW8NGkzMpsanaE3-JKEc8zxf5AS/view?usp=sharing

Task A:

Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.

MISSING DATA ANALYSIS



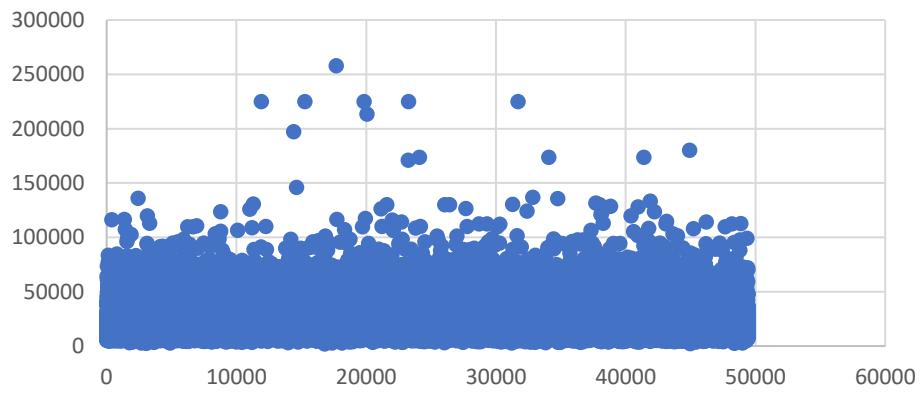
Imputation:

- Columns which contain more than 30% missing values is eliminated from the data
- For numerical data that is missing from columns with less than 20% missing data we use median imputation
- For textual data that is missing we remove the entire row

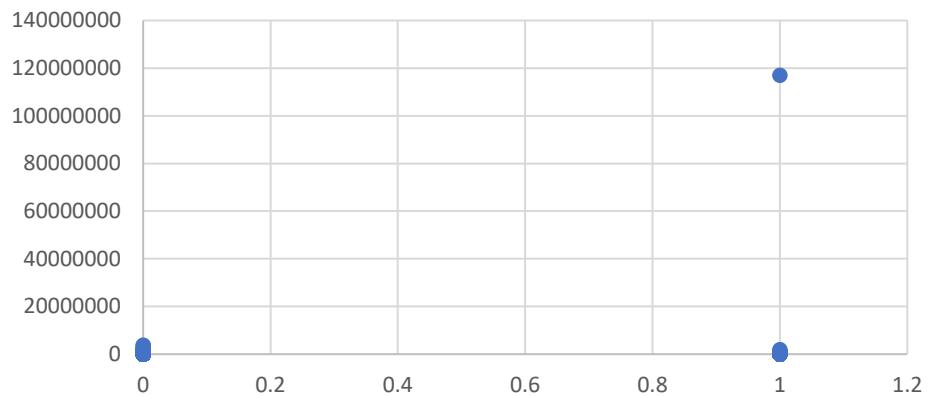
Task B.

Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.

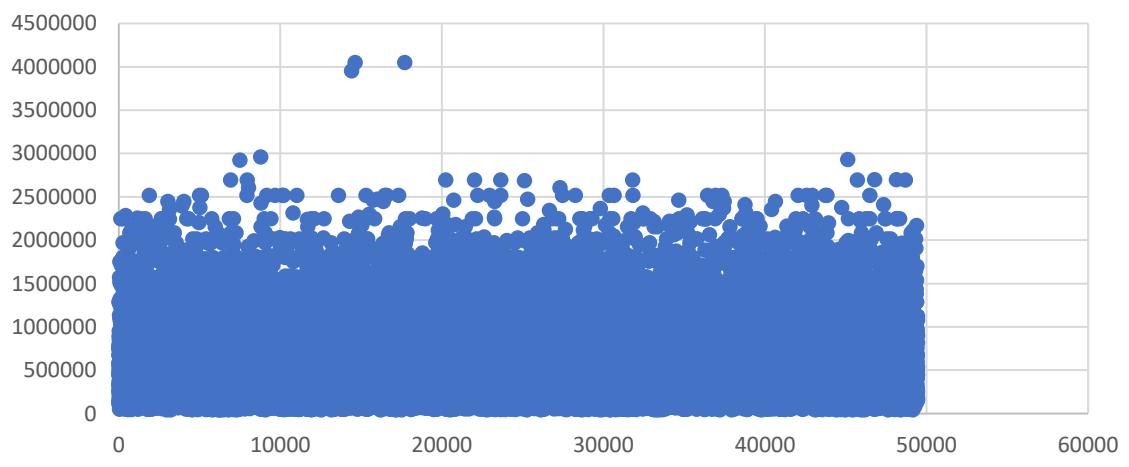
AMT_ANNUITY



TOTAL INCOME OUTLIERS

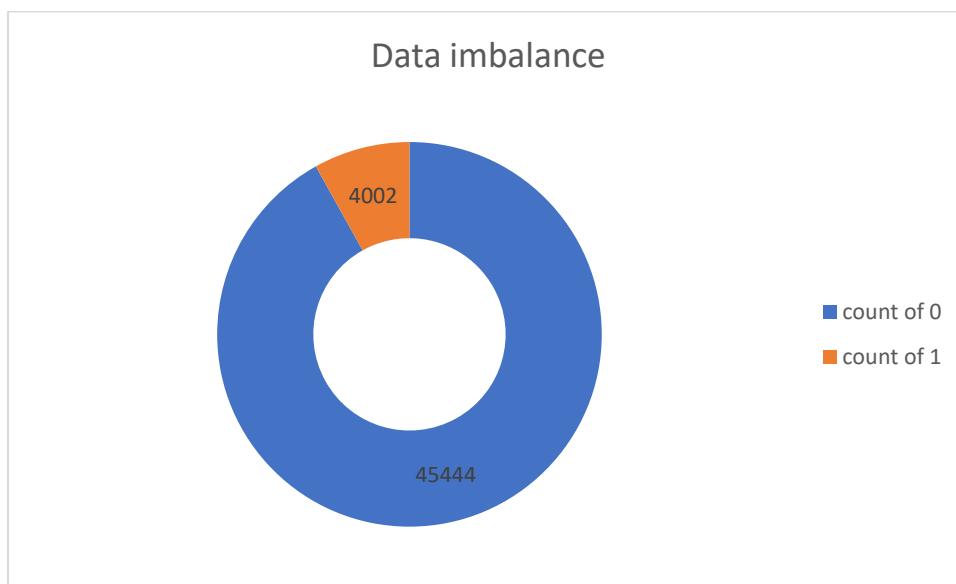


CREDIT AMOUNT OUTLIERS



Task C.

Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.



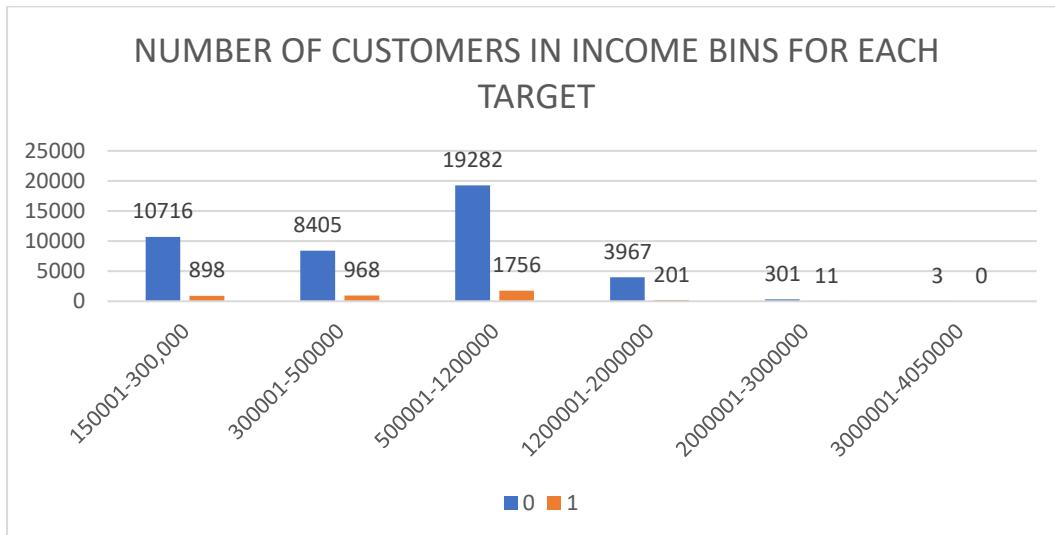
Task D.

Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and feature

1. Segmented univariate analysis of credit bins

Credit Bins	All other cases(0)	Defaulters (1)
0-150000	2770	168
150001-300,000	10716	898
300001-500000	8405	968
500001-1200000	19282	1756
1200001-2000000	3967	201

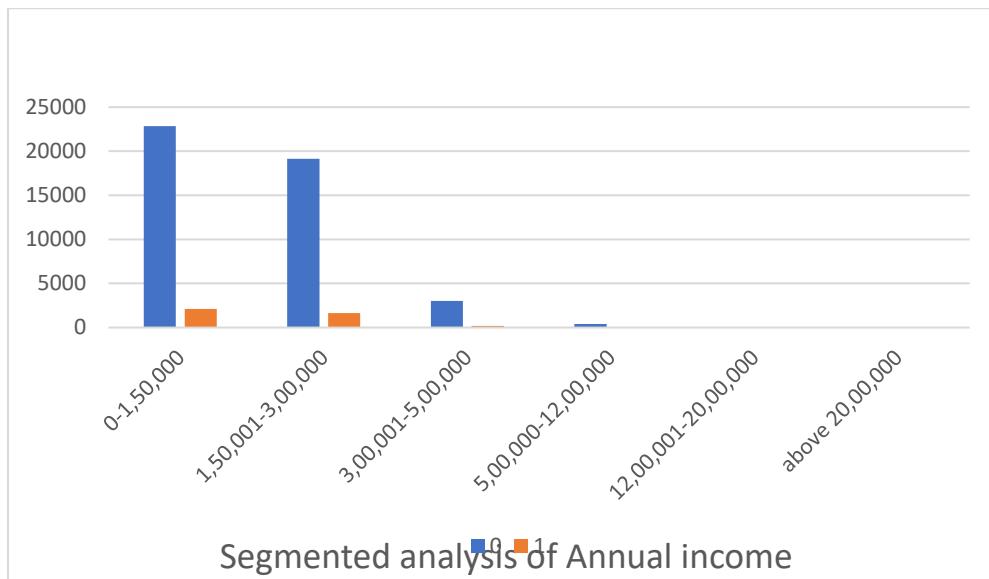
2000001-3000000	301	11
3000001-4050000	3	0



2. Customers in different income bins

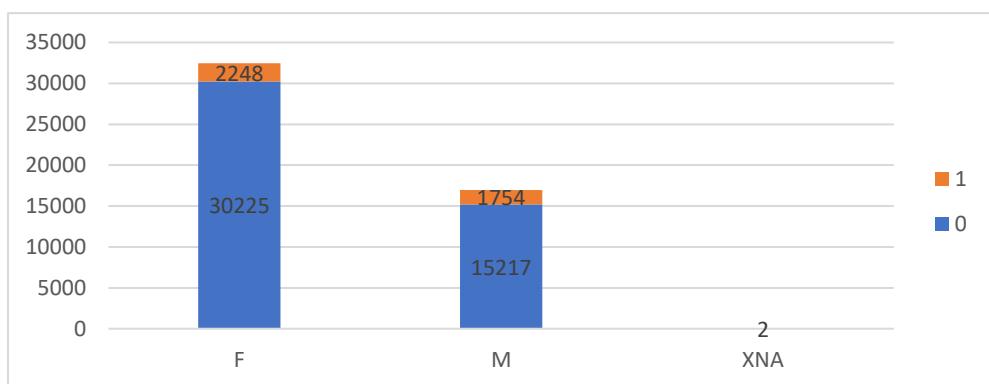
income bins	All other cases (0)	Defaulters (1)
0-1,50,000	22850	2127
1,50,001-3,00,000	19153	1646
3,00,001-5,00,000	3029	198
5,00,000-12,00,000	392	29
12,00,001-20,00,000	14	1

above 20,00,000 6 1



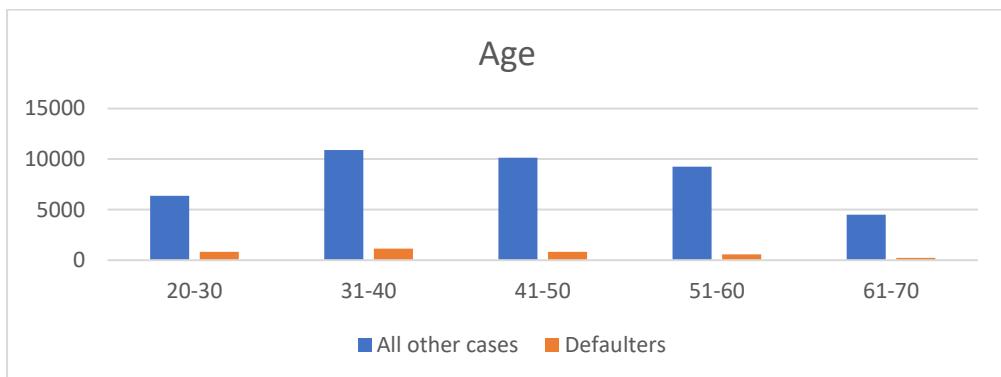
3. Segmented univariate analysis of Gender

Row Labels	All other cases (0)	Defaulters (1)	Grand Total
F	30225	2248	32473
M	15217	1754	16971
XNA	2	0	2
Grand Total	45444	4002	49446



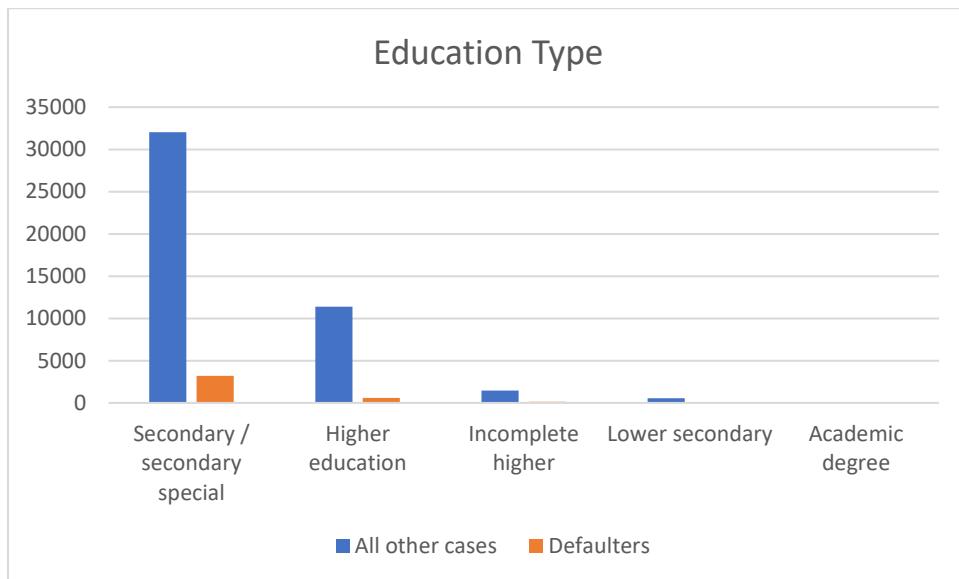
4. Segmented univariate analysis of Age

Age group	All other cases	Defaulters
20-30	6379	814
31-40	10899	1157
41-50	10148	826
51-60	9263	579
61-70	4487	240



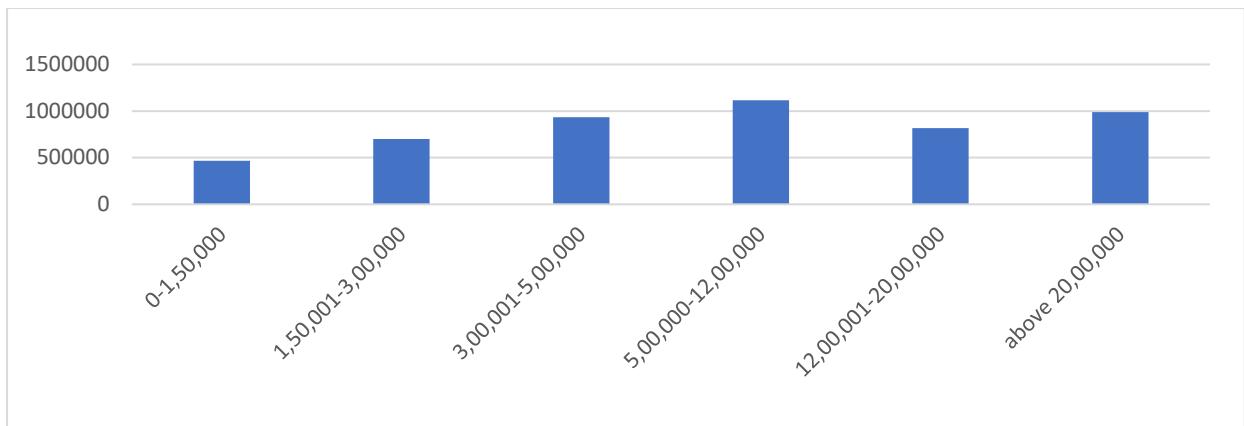
5. Segmented univariate of education type

NAME_EDUCATION_TYPE	All other cases	Defaulters	percentage of defaulters
Secondary / secondary special	32050	3191	9.054794132
Higher education	11369	602	5.028819647
Incomplete higher	1467	136	8.484092327
Lower secondary	538	73	11.94762684
Academic degree	20	0	0



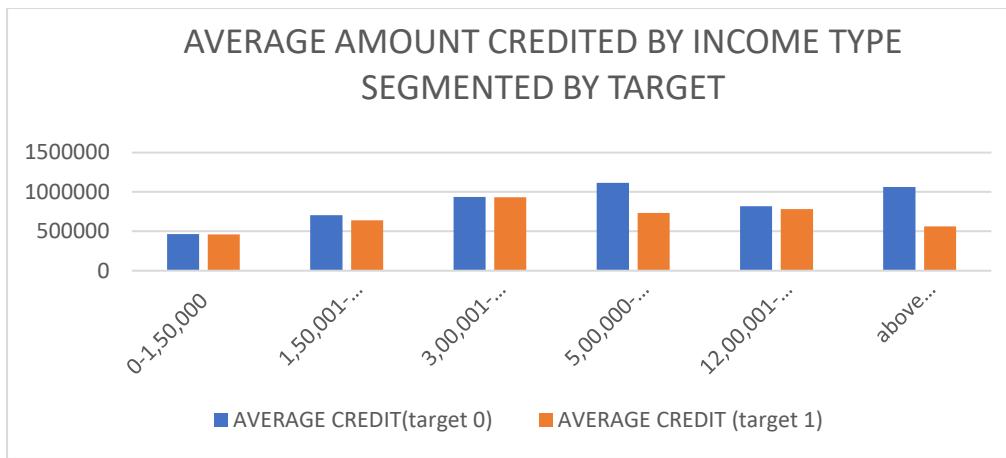
6.Bivariate analysis of income bins and average credit amount

income bins	average credit amount
0-1,50,000	465638.1381
1,50,001-3,00,000	699120.9933
3,00,001-5,00,000	933703.1035
5,00,000-12,00,000	1116821.95
12,00,001-20,00,000	817354.2
above 20,00,000	990090

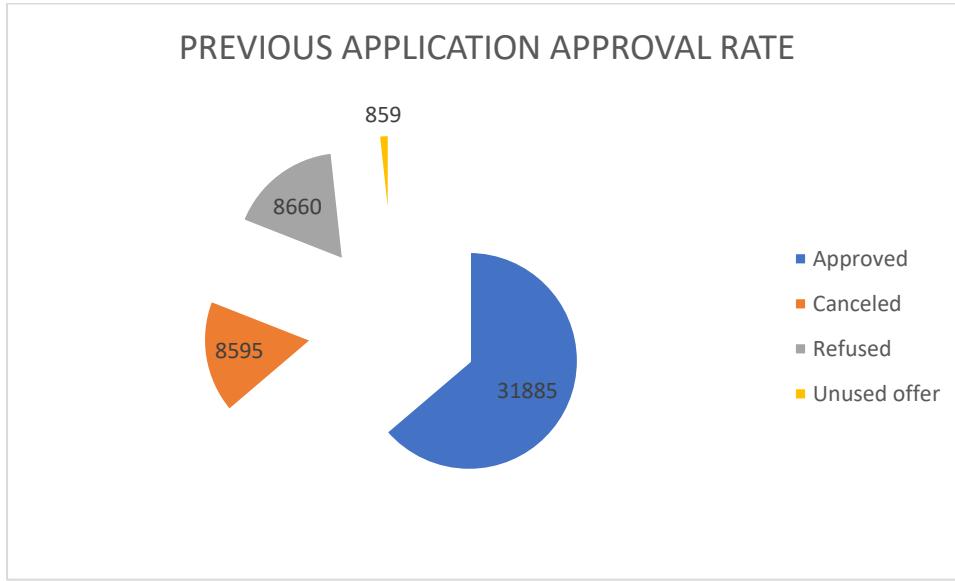


7. Segmented bivariate analysis of income bins and average credit

income bins	AVERAGE CREDIT(target 0)	AVERAGE CREDIT (target 1)
0-1,50,000	465970	462074
1,50,001-3,00,000	704365	638103
3,00,001-5,00,000	938682	933703
5,00,000-12,00,000	1116822	733751
12,00,001-20,00,000	819885	781920
above 20,00,000	1061357	562491



8.Univariate analysis of Loan approval rate in previous application data



Task E.

Segment the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases) and identify the top correlations for each segmented data using Excel functions.

All other cases

Top 10 Correlations

AMT_GOODS_PRICE	AMT_CREDIT	0.99
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.95
CNT_FAM_MEMBERS	CNT_CHILDREN	0.88
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.86
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.83
AMT_GOODS_PRICE	AMT_ANNUITY	0.78
AMT_ANNUITY	AMT_CREDIT	0.77
DAYS_EMPLOYED	client age	0.62
REG_REGION_NOT_WORK_REGION	REG_REGION_NOT_LIVE_REGION	0.45
AMT_ANNUITY	AMT_INCOME_TOTAL	0.45

Defaulters

	CNT_CHILIAMT_INCCAMT_CREIAMT_ANNAMT_GOCREGION_POPUL	client age	DAYS_EMICNT_FAM	REGION_RREGION_RHOUR_APIREG_REGIREG_RELIGIVE_REGIREG_CITY_REG_CITY_I
AMT_INCOME_TOTAL	0.01			
AMT_CREDIT	0.01	0.02		
AMT_ANNUITY	0.03	0.02	0.75	
AMT_GOODS_PRICE	0.00	-0.01	0.98	0.75
REGION_POPULATION_RELATIVE	-0.02	-0.01	0.07	0.07
client age	-0.25	-0.01	0.14	0.01
DAYS_EMPLOYED	-0.19	-0.01	0.02	-0.08
CNT_FAM_MEMBERS	0.89	0.01	0.06	0.08
REGION_RATING_CLIENT	0.06	-0.01	-0.04	-0.06
REGION_RATING_CLIENT_W_CITY	0.06	-0.01	-0.05	-0.08
HOUR_APPR_PROCESS_START	-0.01	0.01	0.05	0.04
REG_REGION_NOT_LIVE_REGION	-0.01	0.00	0.01	0.03
REG_REGION_NOT_WORK_REGION	0.00	0.00	0.02	0.07
LIVE_REGION_NOT_WORK_REGION	0.00	0.00	0.04	0.08
REG_CITY_NOT_LIVE_CITY	0.00	-0.01	-0.05	-0.02
REG_CITY_NOT_WORK_CITY	0.05	-0.01	-0.04	0.00
LIVE_CITY_NOT_WORK_CITY	0.06	-0.01	-0.01	0.01

Top 10 Correlations

AMT_GOODS_PRICE	CNT_CHILDREN	0.98
REGION_RATING_CLIENT_W_CITY	AMT_CREDIT	0.95
CNT_FAM_MEMBERS	AMT_ANNUITY	0.89
LIVE_REGION_NOT_WORK_REGION	client age	0.81
LIVE_CITY_NOT_WORK_CITY	REGION_RATING_CLIENT	0.78
AMT_GOODS_PRICE	REG_REGION_NOT_LIVE_REGION	
AMT_ANNUITY	ION	0.75
DAYS_EMPLOYED	AMT_CREDIT	0.75
REG_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	
REG_CITY_NOT_LIVE_CITY	EGION	0.58
REG_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.52
	REG_CITY_NOT_WORK_CITY	0.47

Link to Excel file:

<https://docs.google.com/spreadsheets/d/1ESUEI0QWRuW2MeG9hQpuNZidXOjvvNol/edit?usp=sharing&ouid=100819582638973735490&rtpof=true&sd=true>

Conclusions

- For class imbalance, proportion of majority class is 91.9% and proportion of minority class is 8.09%
- Majority of customers want Loan amount from 5 to 12 lakhs these are also the credit bin where most defaulters are
- 6.92% of female customers are defaulters and 10.33% male customers are defaulters so priority should be given to female customers while approving loan
- Older customers are most likely to not default payments so priority must be given to them
- People with income from 5 lakhs to 12 lakhs should be focused more while lending
- Customers coming from higher education background tend to default payments the least

Analyzing the Impact of Car Features on Price and Profitability

Project Overview:

The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars. For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand? This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer.

Business Objective:

To find car features that influence its pricing so that a manufacturer can optimize pricing and take product development decisions to maximize profitability while meeting consumer demand

Dataset:

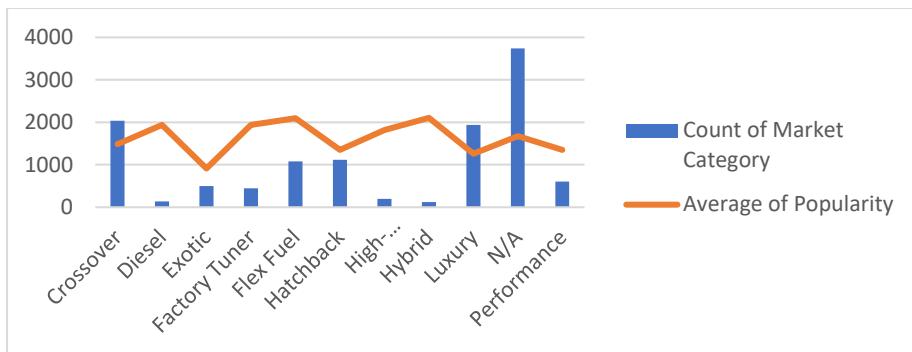
The dataset contains information on over 11,000 car models and their specifications, including details on the car's make, model, year, fuel type, engine power, transmission, wheels, number of doors, market category, size, style, estimated miles per gallon, popularity, and manufacturer's suggested retail price (MSRP).

Raw dataset link:

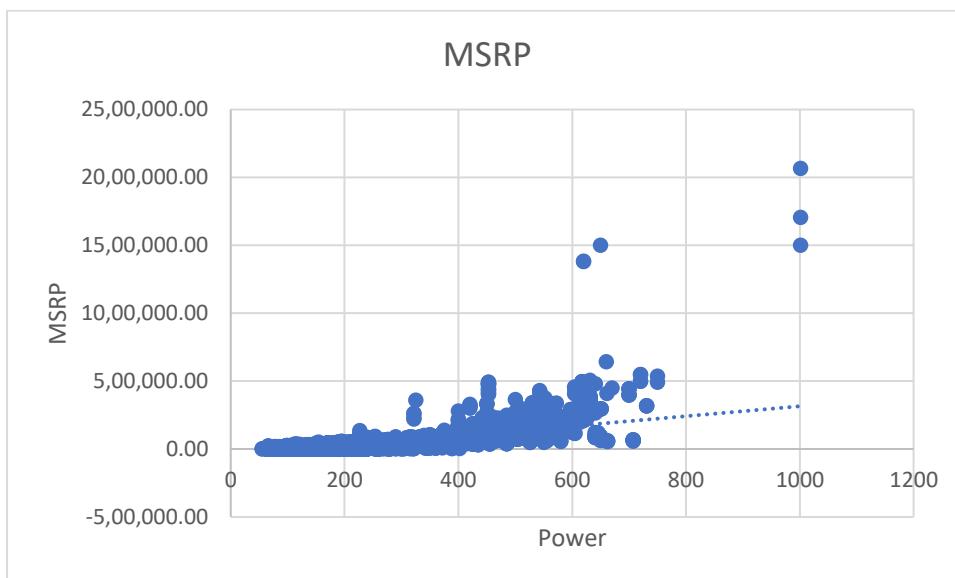
<https://drive.google.com/file/d/10OJxyi3TexbRHb7JrSYaWs2oc-WyvuB/view?usp=sharing>

Insight Required:

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



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



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

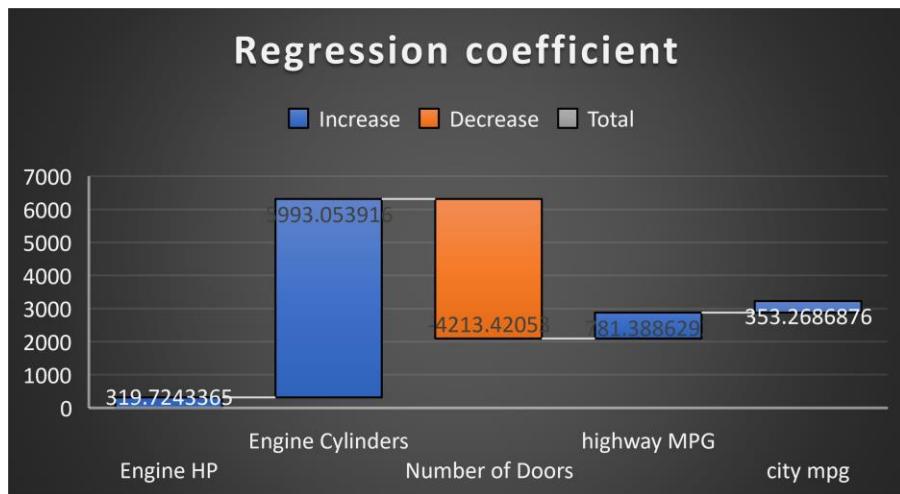
Regression Analysis

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%	
Intercept	-86147.2115	3348.871021	-25.72425482	4.5686E-142	-92711.54529	-79582.9	-92711.5	-79582.9	
Engine HP	319.7243365	5.924961033	53.96226823		0	308.1104458	331.3382	308.1104	331.3382
Engine Cylinders	5993.053916	424.3535762	14.12278405	6.33284E-45	5161.251646	6824.856	5161.252	6824.856	
Number of Doors	-4213.42053	465.3823111	-9.05367572	1.59623E-19	-5125.645818	-3301.2	-5125.65	-3301.2	
highway MPG	781.3886292	103.4976347	7.549821131	4.67637E-14	578.5163728	984.2609	578.5164	984.2609	
city mpg	353.2686876	98.07237372	3.602122333	0.000316915	161.0308282	545.5065	161.0308	545.5065	

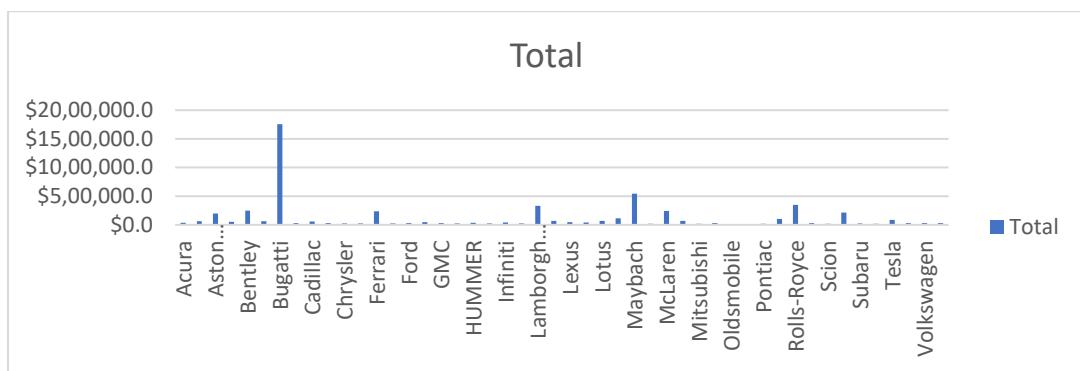
SUMMARY OUTPUT

Regression Statistics

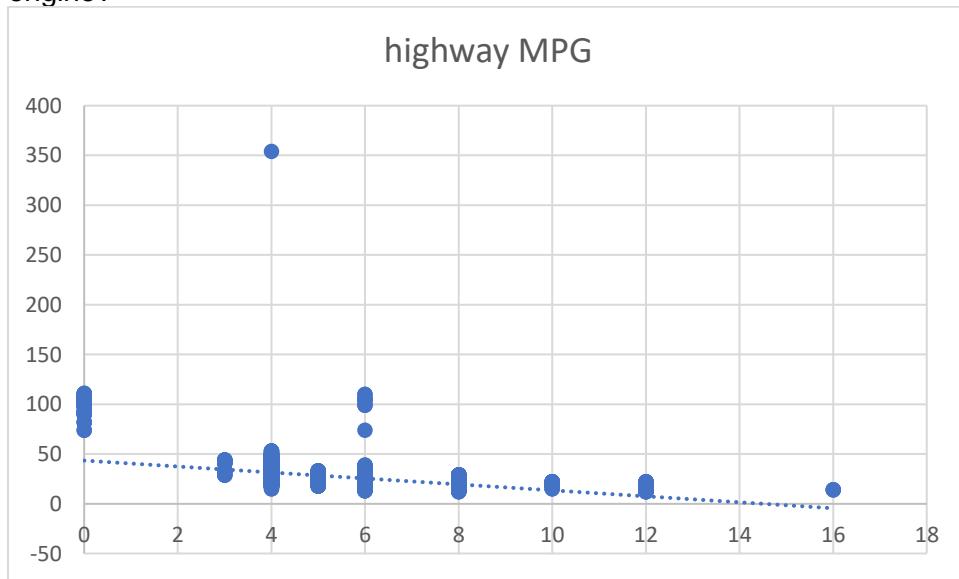
Multiple R	:0.67693217
R Square	:0.458237163
Adjusted R Square	:0.458009627
Standard Error	:44257.14587
Observations	:11911



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



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

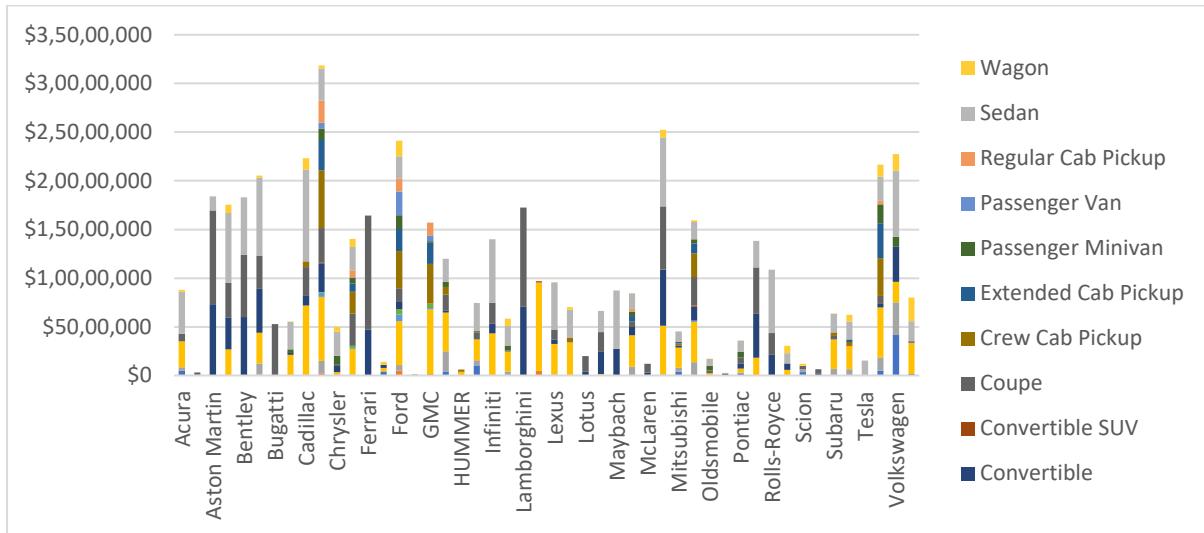


Correlation coefficient

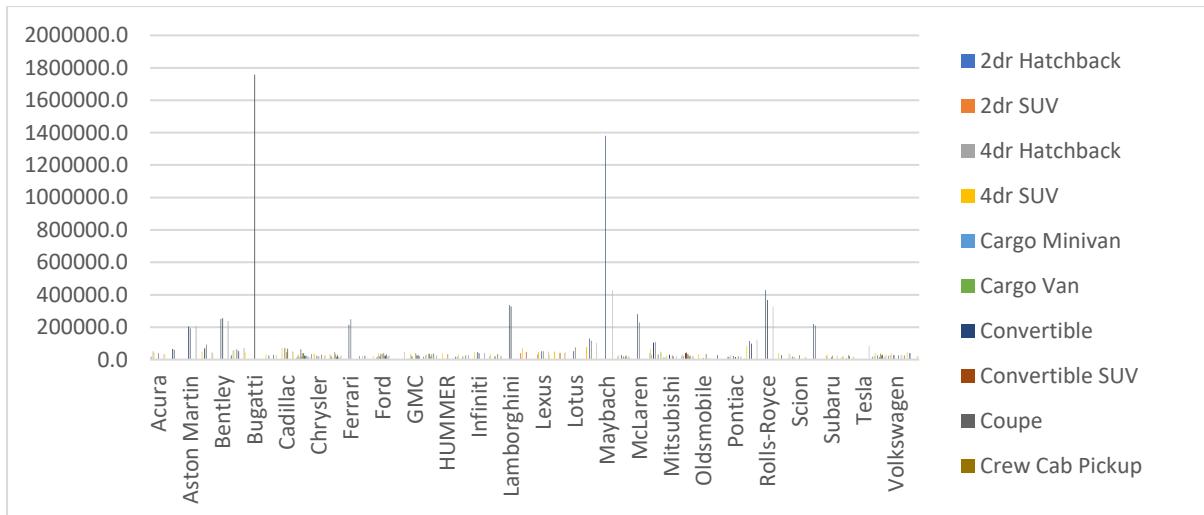
-0.60094

Interactive dashboard

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

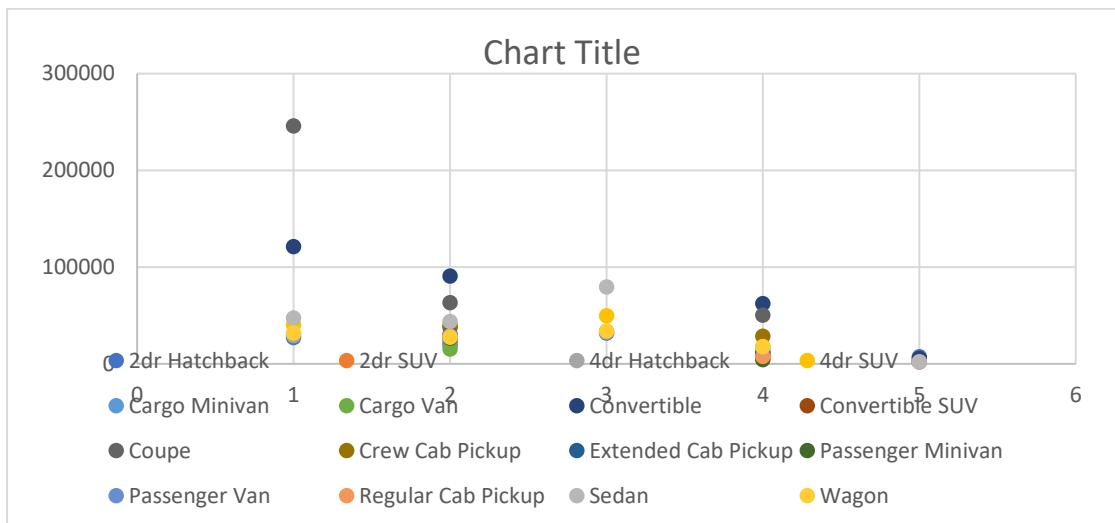


Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

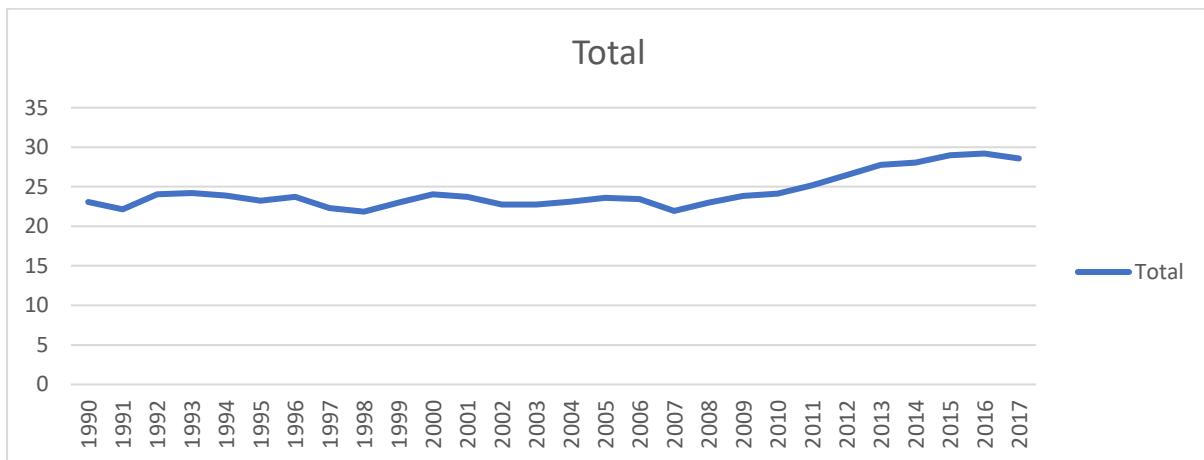


MSRP	Average	Brand
Highest	1757223.7	Bugatti
Lowest	2000.0	Suzuki

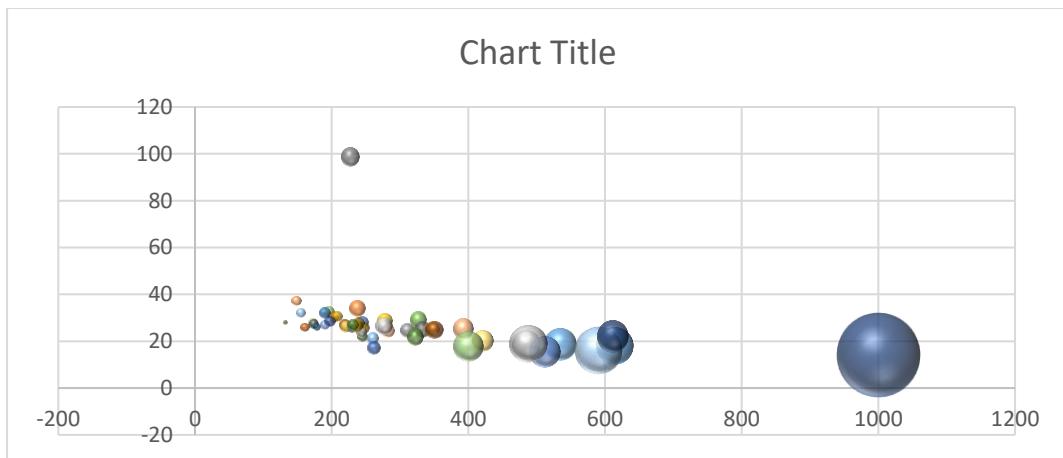
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?



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



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



[Link to dashboard:](#)

https://docs.google.com/spreadsheets/d/1LL8vqU_uxT8sEZrY3qCLjwNums3BQwrs/edit?usp=sharing&ouid=100819582638973735490&rtpof=true&sd=true

Conclusion:

- Bugatti has the highest price where Plymouth has lowest average price
- The number of cylinders in a car has a negative impact on the fuel efficiency
- Engine HP is a determining factor on the car's MSRP , The regression analysis gives highest coefficient and zero p value and low standard error
- Luxury cars have the highest popularity
- There is an overall increase in fuel efficiency over the years
- Bugatti has the highest horsepower , highest MSRP and lowest fuel efficiency

ABC Call Volume Trend Analysis

Description:

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

Inbound customer support, which is the focus of this project, involves handling incoming calls from existing or prospective customers. The goal is to attract, engage, and delight customers, turning them into loyal advocates for the business.

Business Understanding:

Advertising is a crucial aspect of any business. It helps increase sales and makes the audience aware of the company's products or services. The first impressions of a business are often formed through its advertising efforts.

The target audience for businesses can be local, regional, national, or international. Various types of advertising are used to reach these audiences, including online directories, trade and technical press, radio, cinema, outdoor advertising, and national papers, magazines, and TV.

The advertising business is highly competitive, with many players bidding large amounts of money to target the same audience segment. This is where the company's analytical skills come into play. The goal is to identify those media platforms that can convert audiences into customers at a low cost.

In this project, we'll be using your analytical skills to understand the trends in the call volume of the CX team and derive valuable insights from it.

Dataset:

Dataset contains information about the inbound calls received by a company named ABC, which operates in the insurance sector. Your task is to use this data to answer the following questions:

<https://docs.google.com/spreadsheets/d/1RmMKasMWTU7x2EEc-LHAjG70yzMAHpUf/edit?usp=sharing&ouid=100819582638973735490&rtpof=true&sd=true>

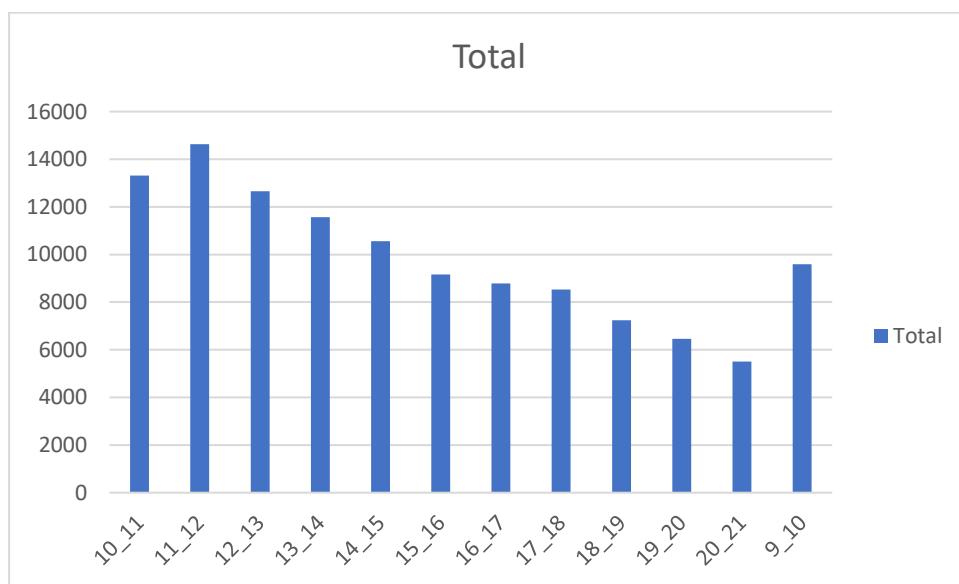
Assumptions: An agent works for 6 days a week; On average, each agent takes 4 unplanned leaves per month; An agent's total working hours are 9 hours, out of which 1.5 hours are spent on lunch and snacks in the office. On average, an agent spends 60% of their total actual working hours (i.e., 60% of 7.5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

Data Analytics:

1. Average duration of calls for each time bucket

Time Bucket	Average Call Duration (s)
9_10	92
11_12	117
12_13	145
13_14	150
14_15	147
15_16	170
16_17	181
17_18	180
18_19	174
19_20	145
20_21	106

2. Number of calls received in each time bucket



Total working hours	187.9622
no of agents working for 4.5 hours with 30% abandment rate	42
Percentage of working hours	70
No of agents required to reduce 90 % abandment rate	54

3. Manpower Planning: Minimum number of agents required in each time bucket to reduce the abandon rate to 10%
- The current rate of abandoned calls is approximately 30%. Plan for manpower allocation during each time bucket (from 9 am to 9 pm) to reduce the abandon rate to 10%.

Row Labels	Average of Call_Seconds (s)	Count of Customer_Phone_No2
abandon	0	29.16%
answered	198.6227745	69.88%
transfer	76.14651368	0.96%
Grand Total	139.5321473	100.00%

Assumptions:

Total Working Hours by the company	9 Hrs
Break	1.5 Hrs
IT downtime	0.5 Hrs
Meetings	1 Hrs
Actual working hour by the agent	5 Hrs

Row Labels	Sum of Call_Seconds (s)	Hours
01-Jan	676664	187.9622

Time Bucket	Count of Call_Seconds (s)	Count of Call_Seconds (s)2	Additional agent required for each time bucket
9_10	0.081262501	9588	4
10_11	0.112833508	13313	6
11_12	0.123961759	14626	7
12_13	0.107231244	12652	6
13_14	0.097984541	11561	5
14_15	0.089509103	10561	5
15_16	0.077626538	9159	4
16_17	0.074482151	8788	4
17_18	0.072329389	8534	4
18_19	0.061345222	7238	3
19_20	0.054776757	6463	3
20_21	0.046657287	5505	3
Grand Total	1	117988	54

4. 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:

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	1	3	4	4			5

Manpower plan for each time bucket throughout the day, keeping the maximum abandon rate at 10%.

average incoming calls	5130
calls at night	1539
Average call hours at night	85
hours required for 90% call answer rate	77
Agents required with 4.5 hours working	17

Time bucket	Distribution of calls	Time distribution	Agent Distribution
9_10	3	0.10	2
10_11	3	0.10	2
11_12	2	0.07	1
12_13	2	0.07	1
13_14	1	0.03	1
14_15	1	0.03	1
15_16	1	0.03	1
16_17	1	0.03	1
17_18	3	0.10	2
18_19	4	0.13	2
19_20	4	0.13	2
20_21	5	0.17	3
Grand Total	30	1.00	17

Analysis Excel file link:

<https://docs.google.com/spreadsheets/d/1ct9WCGqNfss66YxAzNYLmKOdyjvPlYzW/edit?usp=sharing&ouid=100819582638973735490&rtpof=true&sd=true>

Conclusions:

1. Maximum no of calls are in morning time bucket and decreases towards the evenings whereas the Longer call duration is between 3pm and 7pm which could be the reason for lesser number of calls during this time
2. In order to increase the call answering rate to 90% additional 54 agents are required with more agents in the morning time bucket due to more calls
3. If 30% of calls from the total calls in the morning are received in the night 17 more agents have to be assigned

LEARNINGS :

Through these projects I have learned different domains in data analytics such as User engagement analytics , Risk Analytics , Customer experience analytics ,Hiring process analytics

Tools Learned:

- Advanced excel
- MySQL
- Tableau