

TRAINITY

# DATA ANALYSIS PORTFOLIO

MOGULAGANI  
PRASHANTH



# PROFESSIONAL BACKGROUND

*I recently obtained my BTECH in Computer Sciences and Engineering (2019-2023) from Ace Engineering College with a CGPA of 7.6. During my internship as a data analyst trainee, I used Excel and MySQL to accomplish 8 real-world projects. I also learned technical skills in Python, MySQL, Tableau, statistics, and Excel that I used to analyse data.*

*I took on the challenge of implementing analytics to more than ten real-world case studies in order to advance my practical experience. I'm ready to contribute my knowledge and excitement to your team because I'm passionate about turning data into insightful knowledge.*

## TABLE OF CONTENTS

PROFESSIONAL BACKGROUND	1
TABLE OF CONTENTS	2-3
INSTAGRAM USER ANALYTICS DESCRIPTION	4
PROJECT DESIGN	5
FINDINGS	6-9
KEY INSIGHTS	10
CONCLUSION	11
HIRING PROCESS ANALYTICS DESCRIPTION AND DESIGN	12
FINDINGS	13-16
KEY INSIGHTS AND CONCLUSION	17

IMDB MOVIE ANALYSIS DESCRIPTION AND DESIGN ----- 18

FINDINGS-----19 - 23

KEY INSIGHTS AND CONCLUSION ----- 24

BANK LOAN CASE STUDY DESCRIPTION AND DESIGN ----- 25

FINDINGS ----- 26 - 31

KEY INSIGHTS AND CONCLUSION ----- 32 - 33

IMPACT OF CAR FEATURES DESCRIPTION AND DESIGN ----- 34

FINDINGS ----- 35 - 39

KEY INSIGHTS AND CONCLUSION ----- 40

ABC CALL VOLUME ANALYSIS DESCRIPTION AND DESIGN ----- 41

FINDINGS ----- 42 - 45

KEY INSIGHTS AND CONCLUSION ----- 46

APPENDIX ----- 47 - 3.

48

# INSTAGRAM USER ANALYTICS

## DESCRIPTION:

*“With the help of our project, Instagram users will receive in-depth information and insights that will help them perform better on the site. In order to help people, companies, and marketers efficiently optimise their Instagram strategies, we analyse engagement, audience demographics, and post effectiveness. We also provide trend analysis to give users a competitive edge by helping them remain on top of market changes and competition benchmarks.”*



# PROJECT DESIGN



## Phase 1: Database Setup

Import dataset into MySQL using DDL and DML SQL commands.  
Create a structured database and relevant tables.

## Phase 2: Data Analysis

Use MySQL Workbench 8.0 CE to execute SQL queries.  
Generate insights and statistics from database tables.  
Leverage MySQL Workbench for data visualization and interpretation.

## Additional Details:

- MySQL Workbench 8.0 CE is chosen due to familiarity.
- MySQL is the selected database system for its open-source nature and SQL compatibility.

Objective: Create a MySQL database from provided data, perform data analysis with SQL, and extract valuable insights for informed decision-making.

# FINDINGS

⇒ Rewarding Most Loyal Users: People who have been using the platform for the longest time.

⇒ Result : 5 oldest users of Instagram from the database provided.

	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



◆ **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.

◆ **Result:** users who have never posted a single photo on Instagram.

username
▶ Aniya_Hackett
Kasandra_Homenick
Jadyn81
Rodio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59



◆ **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner

◆ **Result:** The winner of the contest is.

	photo_id	username	total
▶	145	Zack_Kemmer93	48



⚡ **Hashtag Researching:** A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.

⚡ **Result:** The top 5 most commonly used hashtags on the platform.

	tag_name	tag_id	count(photo_id)
▶	smile	21	59
	beach	20	42
	party	17	39
	fun	13	38
	concert	18	24

⚡ **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs

⚡ **Result:** what day of the week do most users register on

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



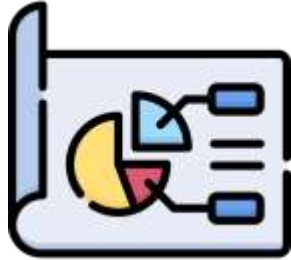
⚡ **User Engagement:** Are users still active and posting on Instagram or they are making fewer posts?

⚡ **Result:** Average user posts on Instagram.

⚡ **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts

⚡ **Result:** Number of Bots and fake accounts present On the platform.

	total_photos	total_users	avg
▶	257	100	2.57000000



username	total_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



## KEY INSIGHTS



Optimizing Instagram Strategies: The project's goal is to boost Instagram tactics, providing insights into performance, engagement, and audience demographics.

🏆 Rewarding Loyalty: Task 1 identifies and rewards the platform's longest-standing users, fostering a stronger community.

🔊 Reengaging Inactives: Task 2 encourages inactive users to post via promotional emails, boosting user participation.

⚖️ Contest Fairness: Task 3 ensures a fair contest by declaring the winner based on user engagement.

🔍 Effective Hashtags: Task 4 helps brands select impactful hashtags for wider audience reach.

📅 Ad Timing: Task 5 identifies the best day for ad campaigns, aligning with peak user activity.

📊 User Engagement: Task 6 monitors user posting trends, offering insights into activity levels.

🛡️ Bot Detection: Task 7 quantifies bots and fake accounts, maintaining platform authenticity.

## CONCLUSION

*In conclusion, the goal of this initiative is to give Instagram users and marketers access to critical insights and analysis. It equips stakeholders to make wise judgements and improve their Instagram strategies by addressing numerous areas of user behaviour and platform authenticity. This initiative enhances the Instagram experience for all users and companies, from recognising devoted users to optimising ad campaign timing.*



# HIRING PROCESS ANALYTICS



## DESCRIPTION

This project aims to enhance an organization's recruitment process, recognizing its pivotal role in long-term company growth and individual development. Careful consideration of various factors during hiring is essential for mutual growth and success.

## DESIGN

Objective: Enhance the organization's recruitment process to facilitate long-term growth for both the company and individuals.

Approach: Utilize Microsoft Excel for data analysis and problem-solving.

Data Collection: Gather relevant recruitment data.  
Data Entry: Organize data in Excel spreadsheets.  
Analysis: Use Excel formulas to assess the process.  
Improvement: Identify areas for enhancement.  
Solutions: Implement process improvements.  
Monitoring: Continuously evaluate and adjust.  
Documentation: Maintain records for reference.

Tools: Microsoft Excel, HR records, candidate profiles.

## FINDINGS

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

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



## OUTPUT

NO OF MALES AND FEMALES HIRED ARE :

Status	event_name	hiring	Column
hired	FEMALE	1856	
hired	MALE	2563	

## OUTPUTS

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

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

AVERAGE SALARY	49983.03
----------------	----------

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

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

Max Salary	400000
Min Salary	100
Diff (Range)	399900
Bins	5
Range/Bins	79980
Class Intervals for Salary	
100-80080	
80081-160060	
160061-240040	
240041-320020	
320021-400000	



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

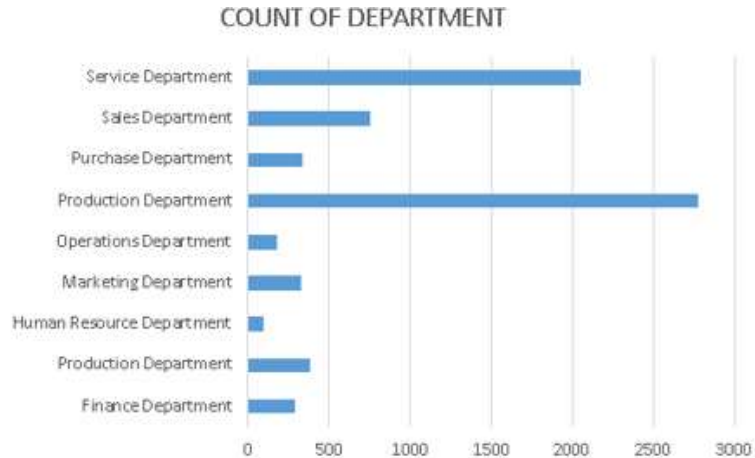
◆ Create class intervals for the salaries in the company. This will help you understand the salary distribution.

## OUTPUT

◆ PRODUCT DEPARTMENT HAS THE HIGHEST COUNT OF SALARIES



DEPARTMENT	COUNT
Finance Department	288
Production Department	380
Human Resource Department	97
Marketing Department	325
Operations Department	176
Production Department	2771
Purchase Department	333
Sales Department	747
Service Department	2055





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

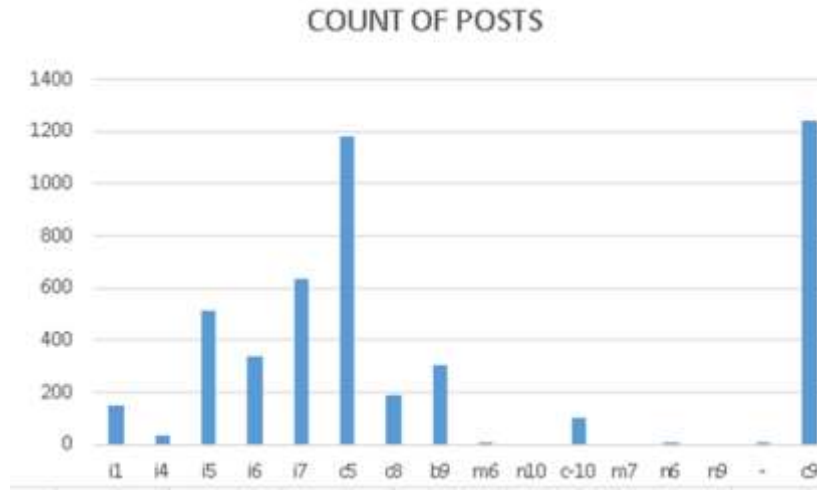
◆ **Your Task:** Use a chart or graph to represent the different position tiers within the company. This will help you understand the distribution of positions across different tiers.

**OUTPUT**

C9 POST HAVE HIGHEST NUMBER OF POSTS



POSTS	NO OF POSTS
i1	151
i4	32
i5	511
i6	337
i7	635
c5	1182
c8	193
b9	308
m6	2
n10	0
c-10	105
m7	0
n6	1
n9	0
-	1
c9	1239



## KEY INSIGHTS


- Task 1: Gender Diversity: The company exhibits gender diversity in hiring with 2,563 males and 1,856 females. 👤
- Task 2: Competitive Compensation: The company offers an average salary of \$49,983, ensuring competitiveness in the job market. 💰
- Task 3 & 4: Salary Distribution: Visualizing salary distribution highlights the Product department's prominence in terms of employee count. 📊
- Task 5: Position Tiers: The C9 position tier holds the highest number of positions within the company, revealing the organizational hierarchy. 🏢


## CONCLUSION



These insights provide a comprehensive understanding of the organization's hiring and compensation structures. Gender diversity and competitive salaries contribute to a positive workplace culture. The prominence of the Product department underscores its significance. Position tier analysis aids in recognizing the organizational hierarchy. These findings empower informed decision-making and strategic planning for future growth and development. 🚀

# IMDB MOVIE ANALYSIS

 **Project Description:** This project centers on exploring IMDB movie data to answer a pivotal question: "What factors contribute to a movie's success on IMDB?" Success is quantified by high IMDB ratings. These insights are indispensable for filmmakers, directors, and investors as they navigate their choices for future film projects.

 **Tools and Resources:** Microsoft Excel serves as the primary analytical tool, harnessing its robust formula capabilities.

 **Key Steps:**

Data Import: Gather and import the IMDB dataset.

Data Cleaning: Ensure data integrity through cleansing.

Data Exploration: Delve into dataset attributes.

Data Analysis: Leverage Excel formulas for deep insights.


Visual Representation: Create impactful visualizations.

Identify Influential Factors: Uncover key drivers of IMDB ratings.

Recommendations: Provide actionable insights based on analysis.

Documentation: Maintain comprehensive records for future reference.



 **Outcome:** This project strives to deliver concise, data-backed insights for stakeholders in the film industry, enabling well-informed decisions and strategic planning for upcoming projects.

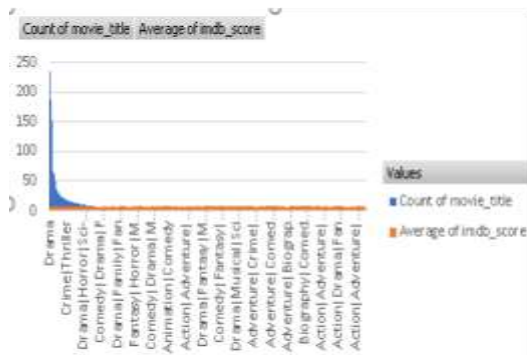
# FINDINGS

Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

DRAMA IS THE MOST COMMON GENRE



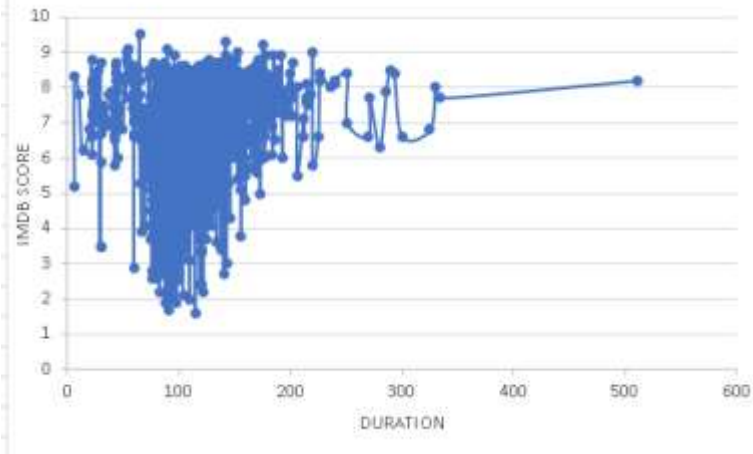
IMDB SCORES	Count of movie_title	Average of imdb_score
Drama	235	6.92
Comedy	205	5.83
Comedy Drama Romance	187	6.53
Comedy Drama	187	6.60
Comedy Romance	156	5.92
Drama Romance	151	6.94
Crime Drama Thriller	99	6.79
Horror	70	5.55
Action Crime Drama Thriller	67	6.45
Drama Thriller	64	6.42



Average of imdb_score	MODE	MEAN	MEDIAN	Max of imdb_score	Min of imdb_score	StdDev of imdb_score	Var of imdb_score
6.92	No Mode	6.922553191	235	9.10	3.30	0.93	0.86
5.83	No Mode	5.834146341	205	9.50	1.90	1.29	1.66
6.53	No Mode	6.526737968	187	8.30	4.30	0.80	0.64
6.60	No Mode	6.6	187	8.80	3.30	0.86	0.74
5.92	No Mode	5.917948718	156	8.40	2.70	0.94	0.89
6.94	No Mode	6.935099338	151	8.10	3.50	0.77	0.60
6.79	No Mode	6.787878788	99	9.00	4.80	0.88	0.77
5.55	No Mode	5.551428571	70	8.00	2.20	1.24	1.55
6.45	No Mode	6.447761194	67	9.00	4.60	0.82	0.67
6.42	No Mode	6.415625	64	8.50	3.90	1.02	1.05

◆ Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score

imdb_score	(All)			
<b>DURATION</b>	<b>FREQUENCY DISTRIBUTION</b>	<b>RELATIVE FREQUENCY DISTRIBUTION</b>	<b>PERCENT DISTRIBUTION</b>	<b>CUMMULATIVE PERCENT</b>
1-100	2161	0.43	43.38%	43.38%
101-200	2784	0.56	55.89%	99.28%
201-300	32	0.01	0.64%	99.92%
301-400	3	0.00	0.06%	99.98%
501-600	1	0.00	0.02%	100.00%
<b>Grand Total</b>	<b>4981</b>	<b>1.00</b>	<b>100.00%</b>	



As we can see the relationship between imdb score and duration is the positive relationship.  
I have used scatter plot to visualize the relation between movie duration and imdb score.

◆ Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics

### TOP 10

language	Average of imdb_score	StdDevp of imdb_score	MEAN	Count of movie	Count of language	MEDIAN
English	6.398426871	1.121948655	30098.2	4704	4704	6.6
French	7.038356164	0.721989287	513.8	73	73	
Spanish	6.9375	0.844300746	277.5	40	40	
Hindi	6.632142857	1.37374711	185.7	28	28	
Mandarin	6.788461538	1.021810958	176.5	26	26	
German	7.342105263	0.928675225	139.5	19	19	
Japanese	7.394444444	0.962907762	133.1	18	18	
NA	6.85	1.199652728	82.2	12	12	
Russian	6.363636364	1.319278541	70	11	11	
Cantonese	6.954545455	0.671989768	76.5	11	11	

◆ Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

THESE ARE THE TOP 10 DIRECTORS BASED ON THE IMDB SCORE.

Used pivot table to find the top directors based on average IMDB scores.

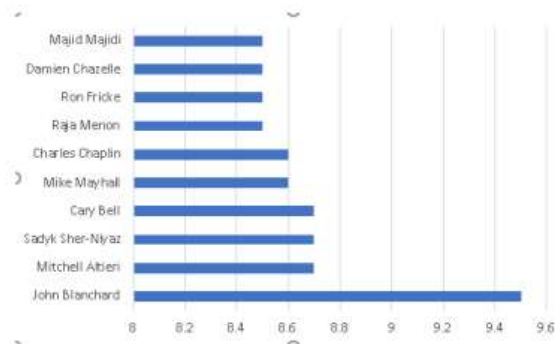
director_name	imdb_score	average imdb scores	percentrank
John Blanchard	9.5	9.5	100.0%
Mitchell Altieri	8.7	8.7	99.9%
Sadyk Sher-Niyaz	8.7	8.7	100.0%
Cary Bell	8.7	8.7	100.0%
Mike Mayhall	8.6	8.6	100.0%
Charles Chaplin	8.6	8.6	100.0%
Raja Menon	8.5	8.5	99.9%
Ron Fricke	8.5	8.5	99.9%
Damien Chazelle	8.5	8.5	100.0%
Majid Majidi	8.5	8.5	100.0%

◆ The below percentile describes the contribution to the success of movies.

90TH % PERCENTILE 7.55



**CLUSTERED BAR:**



Task: :Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

budget	gross	CORRELATION
237000000	760505847	0.102179454
300000000	309404152	
245000000	200074175	
250000000	448130642	
NA	NA	
263700000	73058679	
258000000	336530303	
260000000	200807262	
250000000	458991599	
250000000	301956980	
250000000	330249062	
209000000	200069408	
200000000	168368427	
225000000	423032628	
215000000	89289910	

CORRELATION BETWEEN MOVIE BUDGETS AND GROSS EARNINGS: **0.102179454**

THESE ARE THE TOP 10 MOVIES HAVING HIGHEST PROFIT MARGIN FROM THE GIVEN DATASET

PROFIT	movie_title
523505847	Avatar
9404152	Pirates of the Caribbean: At World's End
198130642	The Dark Knight Rises
78530303	Spider-Man 3
208991599	Avengers: Age of Ultron
51956980	Harry Potter and the Half-Blood Prince
80249062	Batman v Superman: Dawn of Justice
198032628	Pirates of the Caribbean: Dead Man's Chest
66021565	Man of Steel
403279547	The Avengers



## KEY INSIGHTS AND CONCLUSION

### 🔍 Key Insights:

Most Common Genre: "DRAMA" dominates. 🎭

- Descriptive statistics reveal IMDB score patterns within genres. 📊

Duration-IMDB Score Relationship: Positive correlation found. 📈📊

- Scatter plot visualizes this correlation. 📈

Impact of Language: Language affects IMDB scores. 🌐📊

Top Directors: Recognized directors' influence. 🎬🏆

- Percentiles highlight their significance. 📊

Budget vs. Gross Earnings: Budgets correlated with earnings. 💰📈

- High-profit margin movies identified. 🚀💰

### 🚀 Conclusion:

These insights empower decision-makers in the film industry to navigate successful movie projects effectively. Factors like genre, duration, language, top directors, and finances significantly impact IMDB ratings.



## BANK LOAN CASE STUDY

### 🔑 Role Description:

As a data analyst in a financial institution specializing in urban client loans, my role involves addressing loan defaults by clients with limited credit history. Utilizing exploratory data analysis (EDA), I aim to identify data patterns to ensure that deserving candidates are not declined. 🏠💼

### Project Design

📊 Approach: Leverage Microsoft Excel for comprehensive data analysis, applying formulas for solutions.

### 📈 Key Steps:

Data Import: Gather and import loan-related data into Microsoft Excel.

Data Cleaning: Ensure data accuracy by addressing discrepancies and missing values.

Data Exploration: Investigate data patterns to uncover factors contributing to loan defaults.

Data Analysis: Utilize Excel formulas for in-depth analysis of credit history, loan types, and client demographics.

Visual Representation: Create compelling visualizations in Excel to effectively communicate findings.

Recommendations: Formulate strategies to mitigate defaults and ensure deserving clients are not turned away.

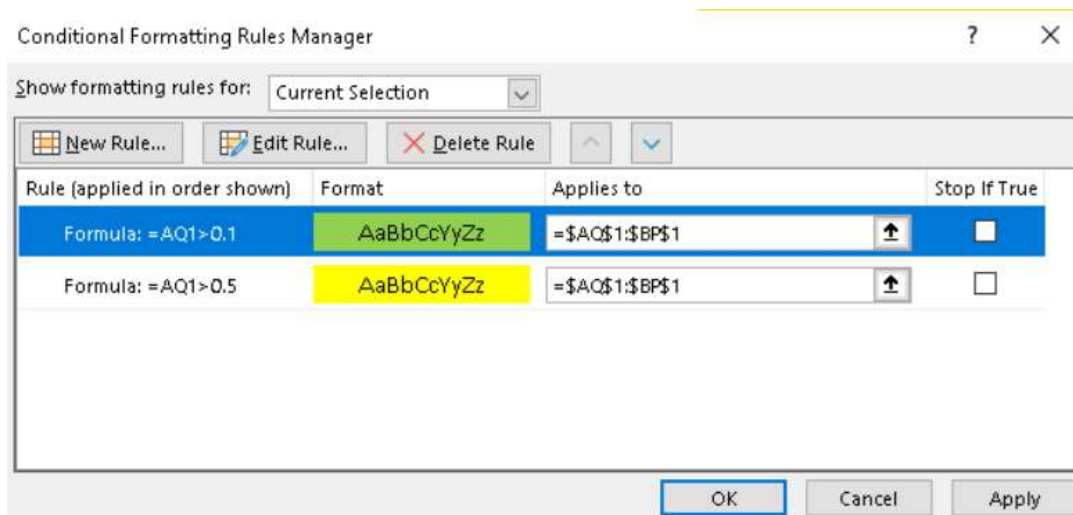
### 🔧 Tools and Resources:

- Microsoft Excel for data analysis and visualization.

## FINDINGS

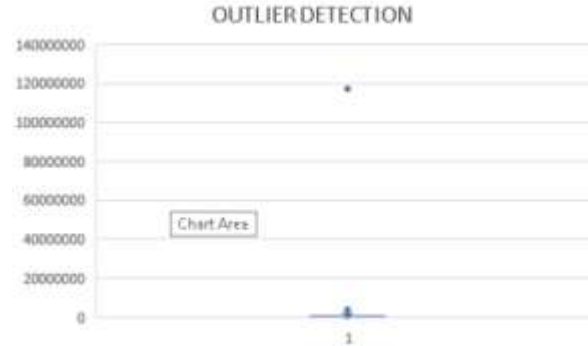
◆ Task: Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.

USED CONDITIONAL FORMATTING TO FIND OUT THE BLANKS MORE THAN 50% AND REMOVE THEM

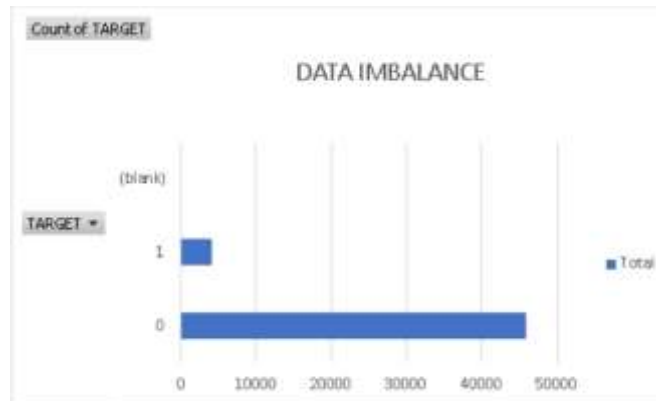


Task: Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.

DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_DOCUMENT	FLAG_MOBILE	FLAG_EMP_PHONE	FLAG_WORK_PHONE
-637	-3646	-3120	1	1	0
-1398	-1196	-291	1	1	0
-225	-4760	-2931	1	1	0
-3039	-3633	-2437	1	1	0
-3039	-4311	-3409	1	1	0
-1509	-8970	-477	1	1	0
-3180	-1213	-619	1	1	0
-445	-4957	-2979	1	1	0
-2171	-7827	-9514	1	1	0
-2019	-14807	-3992	1	1	0
-479	-4427	-798	1	1	0
-2171	-5146	-2512	1	1	0
-2117	-211	-3227	1	1	0
-3028	-643	-4911	1	1	0
-209	-415	-2956	1	1	0
-1857	-3494	-1368	1	1	0
-1817	-6392	-3866	1	1	0
-351	-4149	-2427	1	1	0
-209	-8751	-1295	1	1	0
-2038	-1921	-3964	1	1	0
-4286	-290	-1800	1	1	0



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

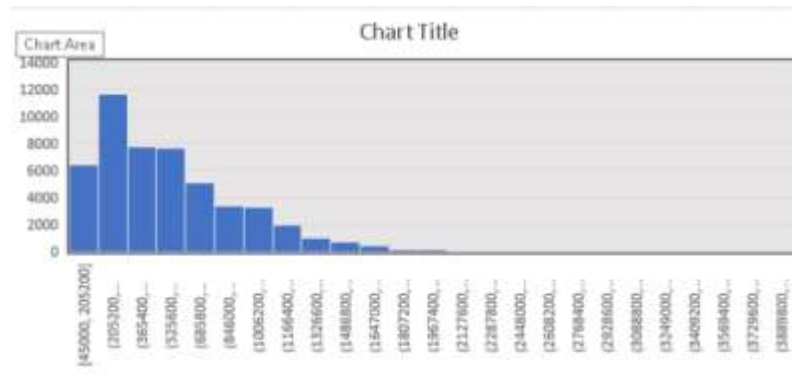


Task: Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.

## 1.UNIVARIATE ANALYSIS:

### DESCRIPTIVE STATISTICS FOR AMT\_CREDIT VARIABLE:

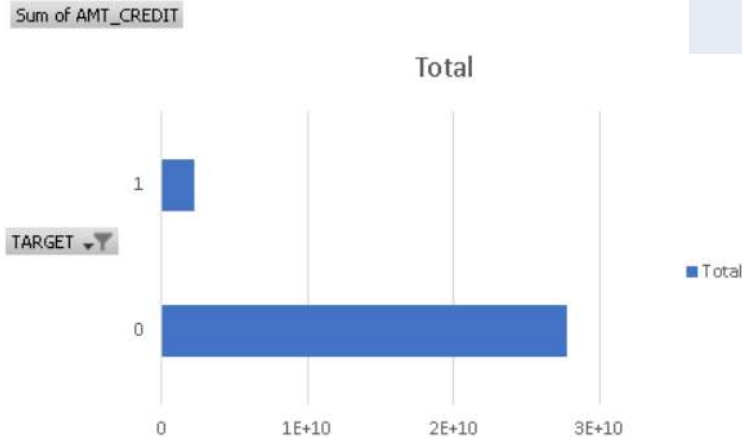
AVERAGE/MEAN	MEDIAN	MODE	MAX	MIN	COUNT
597519.1377	508495.5	450000	2517300	45000	4999



## 2.SEGMENTED UNIVARIATE ANALYSIS:



TARGET	Sum of AMT_CREDIT
0	27747569597
1	2236859780
Grand Total	29984429376



🔗 Targeted Credit Usage: The analysis reveals distinct credit usage patterns among different segments. Specifically, it highlights a notable increase in credit usage among a specific demographic group, indicating a potential market opportunity for targeted financial products or services in this segment.

### 3.BIVARIATE ANALYSIS:

INCOME	Average of AMT_CREDIT
25650-125650	425228.7416
125650-225650	632779.0874
225650-325650	839540.732
325650-425650	935945.9604
425650-525650	1009091.246
525650-625650	1123616.396
625650-725650	1046201.618
725650-825650	1287182.647
825650-925650	1161345.214
925650-1025650	1303200
1025650-1125650	1153857.971
1225650-1325650	1095111
1325650-1425650	914911.2
1525650-1625650	900000
1725650-1825650	1237500
1825650-1925650	781920
1925650-2025650	731068.5
2225650-2325650	1125000
3525650-3625650	953460
3725650-3825650	1241023.5
116925650-117025650	562491
<b>Grand Total</b>	<b>599700.5815</b>



💰 **Income vs. Credit Amount:** Higher income levels correspond to larger credit amounts, indicating a positive correlation. Tailoring credit offerings based on income can optimize credit utilization and repayment.

- Task: 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.

HERE WE ARE SEGMENTING THE TARGET VARIABLE INTO 2 THAT IS : TARGET - 0 , TARGET - 1.

FOR TARGET 1:

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT
CNT_CHILDREN	1	0.00960733	0.005042552
AMT_INCOME_TOTAL	0.00960733	1	0.069319162
AMT_CREDIT	0.005042552	0.069319162	1

- HERE CONDITIONAL FORMATTING IS USED TO DIFFERENTIATE BETWEEN LOWEST , MIDPOINT AND HIGHEST VALUES USING TRI-COLORS



## FOR TARGET 0:

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT
CNT_CHILDREN	1	0.00960733	0.005042552
AMT_INCOME_TOTAL	0.00960733	1	0.069319162
AMT_CREDIT	0.005042552	0.069319162	1

## KEY INSIGHTS AND CONCLUSION

### Task 1: Missing Data Handling

- Conditional formatting identified blanks exceeding 50%, ensuring data completeness. ☒

### Task 2: Outlier Detection

- Conditional formatting and charts exposed outliers, critical for robust analysis. 📈

### Task 3: Data Imbalance Assessment

- Calculated data imbalance ratio, highlighting potential modeling challenges. 📊🔍

### Task 4: Exploratory Data Analysis (EDA)

- Univariate, segmented univariate, and bivariate analysis provided essential insights. 📊

### Task 5: Correlation in Segmented Data

- Segmentation revealed segment-specific correlations, aiding tailored strategies. 📊🔍

🔍 In summary, EDA using Excel uncovered valuable insights, facilitating data-driven decisions for loan approval and risk management. 📊🔍

### 🔍 Conclusion:

Our EDA journey using Excel functions and features has unearthed vital insights. We tackled missing data, spotted outliers, gauged data imbalance, and unraveled variable relationships. These revelations guide our loan approval, risk management, and tailored strategies. EDA in Excel remains our compass for data-driven decisions. 📊🔍



## IMPACT OF CAR FEATURES

### ◆ Description:

In the fast-evolving automotive industry, our focus is on fuel efficiency, sustainability, and innovation. We employ regression analysis and market segmentation to analyze the relationships between car features, market categories, and pricing. This approach helps us strike a balance between consumer demand and profitability, guiding future product development and enhancing our market competitiveness. 🚗📊💡

### ◆ Project Design

◆ Approach: Leverage Microsoft Excel for data analysis with Excel formulas.

Data Collection: Gather car features, market categories, and pricing data.

Data Preparation: Clean and preprocess data for accuracy.

Regression Analysis: Explore car feature impact on pricing.

Market Segmentation: Segment markets based on preferences.

Pricing Optimization: Align pricing with insights.

Innovation Focus: Identify innovation opportunities.

Competitiveness: Enhance market competitiveness.

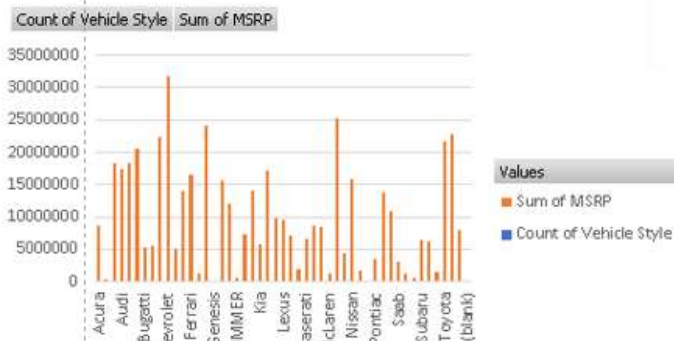
◆ Tools: Microsoft Excel for analysis and visualization. 🚗📊💡



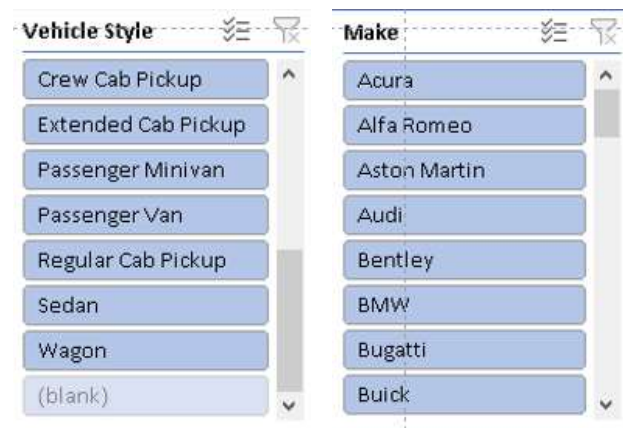
## FINDINGS

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

Sum of MSRP	Column Labels			
BRAND	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV
Acura	480917		357440	2663505
Alfa Romeo				
Aston Martin				
Audi	4000			2674900
Bentley				
BMW	80097		1144950	3160950
Bugatti				
Buick				2141770
Cadillac				7182555
Chevrolet	8000	213310	1287260	6569568
Chrysler	98805			250545
Dodge	48000	44000	18000	2572405
Ferrari				
FIAT	420715			369305
Ford	36000	479873	567615	4482771
Genesis				
GMC		144319		6641919
Honda	413200		2088520	3953209

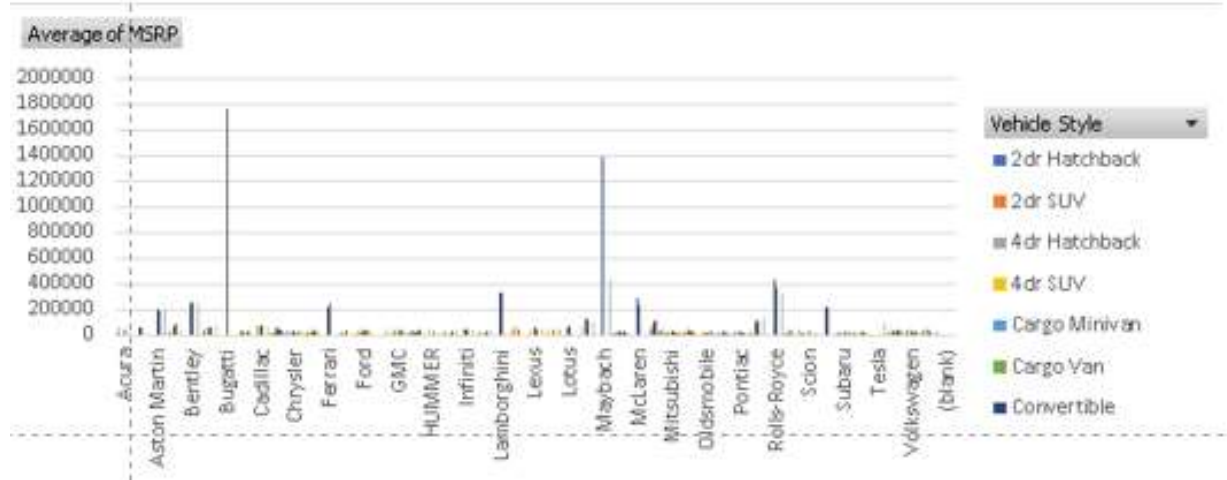


◆ Interactive Pivot Table and slicers provide dynamic visuals, making pricing analysis effortless and insightful for users.



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

Average of MSRP	VEHICLE STYLE		
BRAND	2dr Hatchback	2dr SUV	4dr Hatchback
Acura	17175.60714		51062.85714
Alfa Romeo			
Aston Martin			
Audi	2000		
Bentley			
BMW	26699		54521.42857
Bugatti			
Buick			
Cadillac			
Chevrolet	2000	8887.916667	18930.29412
Chrysler	32935		
Dodge	2000	2000	2000
Ferrari			
FIAT	21035.75		
Ford	2000	13710.65714	19572.93103
Genesis			
GMC		5550.730769	
Honda	17216.66667		26106.5
HUMMER			
Hyundai	18536.60714		17629.33333



\$ Striking price disparities among car brands and body styles.

🚗 Identifying priciest and budget-friendly brands.

👁 Body style's significant influence on car prices.

📊 Dynamic charts for interactive exploration of pricing dynamics. 💰

Make

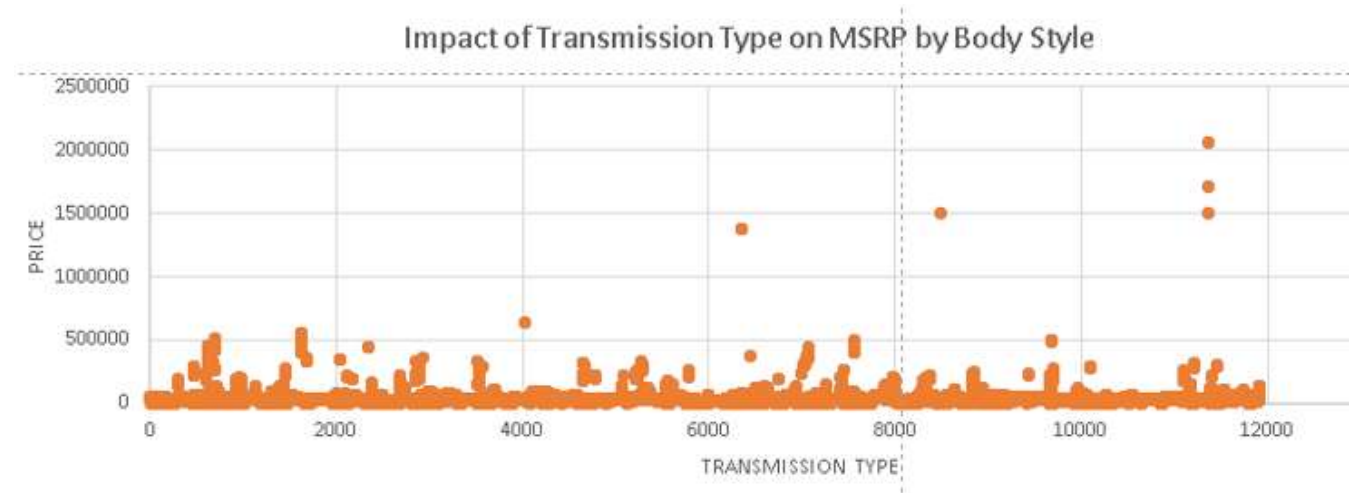
Acura  
Alfa Romeo  
Aston Martin  
Audi  
Bentley  
BMW  
Bugatti  
Buick

Vehicle Style

2dr Hatchback  
2dr SUV  
4dr Hatchback  
4dr SUV  
Cargo Minivan  
Cargo Van  
Convertible  
Convertible SUV

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



Average of MSRP	Column Labels						
TRANSMISSION TYPE	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible
AUTOMATED_MANUAL	27180.96491		29249.07407	40451.15385			121256.6444
AUTOMATIC	20926.464	18615.20455	23833.67898	41535.60646	20920.98592	15280.22105	90637.3869
DIRECT_DRIVE	31800		32799.72973	49800			
MANUAL	13353.65831	6303.811111	17594.41313	15426.46226			62357.75625
UNKNOWN	7361.5	2371					5783.5
(blank)							
Grand Total	16867.71344	10115.18841	22420.8661	40421.87178	20920.98592	15280.22105	84224.28499



Transmission Type

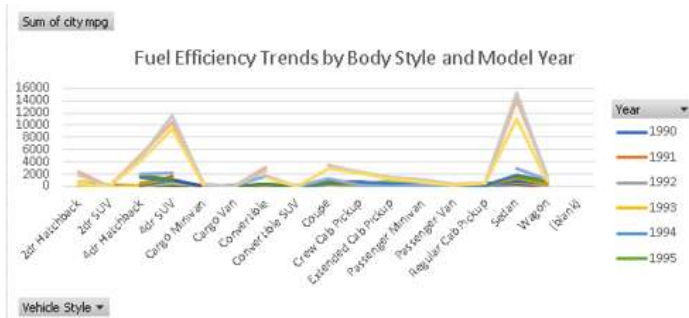
- AUTOMATED\_MANUAL
- AUTOMATIC
- DIRECT\_DRIVE
- MANUAL
- UNKNOWN
- (blank)

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

Sum of city mpg	YEAR 											
BODY STYLE 	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	
2dr Hatchback	118	335	748	652	412	162	129	174	86	216	277	
2dr SUV	214	150	224	258	194	48	81	56	66	56	57	
4dr Hatchback	44		174	214	149	66	149	188	36			
4dr SUV		86	47	93	90		94	160	164	133	204	
Cargo Minivan	18				17	33	33	15				
Cargo Van					87	73	110	104	128	40	63	
Convertible	62	127	223	229	71	67	87	129	51	30	123	
Convertible SUV				48	24	24	42	54	63			
Coupe	238	349	267	550	479	567	647	570	285	176	103	
Crew Cab Pickup												
Extended Cab Pickup	108	148	114	183	112	175	128	390	449	368	127	
Passenger Minivan	224	64			17	183	139	137	167	98	98	

Vehicle Style	Year
2dr Hatchback	1990
2dr SUV	1991
4dr Hatchback	1992
4dr SUV	1993
Cargo Minivan	1994
Cargo Van	1995
Convertible	1996
Convertible SUV	1997

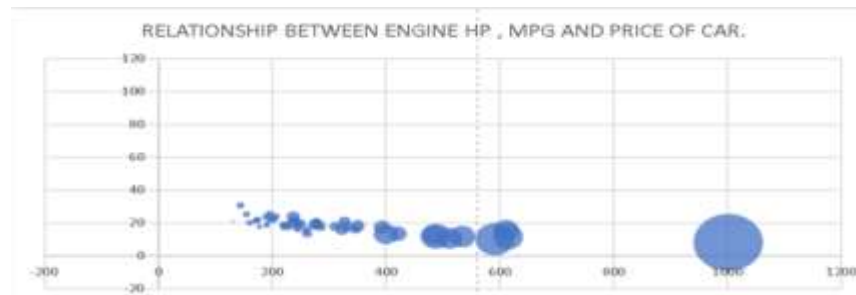
⇒ Fuel efficiency variations across body styles and model years elegantly visualized in dynamic charts. ☒ Evolution of fuel efficiency trends empowers consumers and manufacturers for savvy choices and design decisions. 🚗📊





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

BRAND	Average of Engine HP	Average of city mpg	Average of MSRP
Acura	244.797619	19.94047619	34887.5873
Alfa Romeo	237	24	61600
Aston Martin	484.3225806	12.52688172	197910.3763
Audi	277.695122	19.58536585	53452.1128
Bentley	533.8513514	11.55405405	247169.3243
BMW	326.9071856	20.73952096	61546.76347
Bugatti	1001	8	1757223.667
Buick	219.244898	18.70408163	28206.61224
Cadillac	332.3098237	17.35516373	56231.31738
Chevrolet	246.9722471	19.02137133	28350.38557
Chrysler	229.1390374	17.75935829	26722.96257
Dodge	244.4153355	16.06549521	22390.05911
Ferrari	511.9565217	10.56521739	238218.8406
FIAT	143.559322	30.64516129	22670.24194



Make
Acura
Alfa Romeo
Aston Martin
Audi
Bentley
BMW
Bugatti
Buick

Engaging Bubble Chart reveals trade-offs between performance, efficiency, and price, offering insights into brand positioning. 🚗💡



## ◆ Key Insights:

Price Diversity: Dynamic visuals via Pivot Tables and slicers streamline pricing analysis by brand and body style.

Brand MSRPs: Uncover brands with high and low average MSRPs, highlighting pricing variations across body styles.

Feature Influence: Explore how factors like transmission type affect MSRP, with body style nuances.

Fuel Efficiency Trends: Visualize fuel efficiency shifts over time and across body styles for informed choices.

Performance vs. Efficiency: Engaging Bubble Chart reveals trade-offs between performance, efficiency, and price, illuminating brand positioning

## ◆ Conclusion:

This analysis equips users with interactive tools for data exploration, enabling data-driven decisions in the dynamic automotive market. 🚗📊👉💰📈



# ABC CALL VOLUME ANALYSIS

## DESCRIPTION

"In this project, I enhanced customer experiences for a company's inbound calling team. We employed AI tools like Interactive Voice Response, Predictive Analytics, and Intelligent Routing. My role focused on analyzing agent performance in handling incoming calls, impacting customer satisfaction." 📞 📊 📈

🔹 Project Design:

🔹 Approach: Utilize Microsoft Excel for comprehensive data analysis and solution evaluation with Excel formulas.

🔹 Key Steps:

Data Gathering: 📊 Collect and organize data related to incoming customer calls and agent performance.

Data Preprocessing: 🧹 Clean and prepare the data for accurate analysis.

Analysis Phase: 📈 Employ Excel functions and formulas to analyze agent performance metrics.

AI Integration: 🤖💻 Collaborate with AI tools like Interactive Voice Response and Predictive Analytics for data-driven insights.

Performance Evaluation: 📊 Assess the impact of agent performance on customer satisfaction and loyalty.

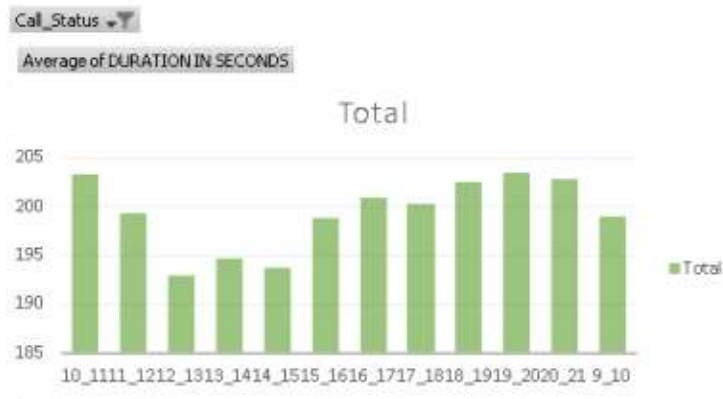
🔹 Tools and Resources:

- Microsoft Excel for data analysis and solution assessment.

## FINDINGS

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

Consistent call durations (190-205 seconds) enhance service experience, optimize staffing, set performance benchmarks, and are visually validated by a clustered column chart. 🕒📊



Call_Status	answered
TIME BUCKET	Average of DURATION IN SECONDS
10_11	203.3310302
11_12	199.2550234
12_13	192.8887829
13_14	194.7401744
14_15	193.6770755
15_16	198.8889175
16_17	200.8681864
17_18	200.2487831
18_19	202.5509677
19_20	203.4060725
20_21	202.845993
9_10	199.0691057
<b>Grand Total</b>	<b>198.6227745</b>

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

Row Labels	Count of DURATION IN SECONDS
10_11	13313
11_12	14626
12_13	12652
13_14	11561
14_15	10561
15_16	9159
16_17	8788
17_18	8534
18_19	7238
19_20	6463
20_21	5505
9_10	9588
<b>Grand Total</b>	<b>117988</b>



Time-based call volume patterns reveal a consistent peak between 11 AM and 12 PM, as visually confirmed by a clustered column chart. This indicates the need for extra resources during this period to manage the higher call volume effectively. 📊

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

ASSUMPTIONS	
WORKING HOURS :	9
BREAK TIME :	1.5
ACTUAL WORKING HOURS :	7.5
OCCUPANCY :	60%
WORKING IN SECONDS :	16200
AVERAGE CALL TIME / AGENT	199
CALL CAPACITY OF AGENT/DAY	81
CALL CAPACITY OF AGENT/HOUR :	16



- 📊 Peak call times demand optimized staffing; efficient resource allocation is vital.
- 📊 Consistent call durations ensure predictability; agent shortage requires adjustments.
- 🕒 Precise staffing during peaks reduces abandon rates;
- 💡 assumption-driven strategies aim for a 10% abandon rate with 90% call answer rate.

TIME_BUCKET	CURRENT NO OF AGENTS	REQUIRED NO OF AGENTS
10_11	17	32
11_12	23	36
12_13	26	31
13_14	24	28
14_15	22	26
15_16	21	22
16_17	21	21
17_18	21	21
18_19	17	17
19_20	12	16
20_21	8	13
9_10	12	23

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

TIME SLOT	NO OF AGENTS NEEDED
9pm - 10pm	9
10pm - 11pm	9
11pm - 12am	6
12am - 1am	6
1am - 2am	3
2am - 3am	3
3am - 4am	3
4am - 5am	3
5am - 6am	9
6am - 7am	11
7am - 8am	11
8am - 9am	14



🌙 Manpower planning insights reveal peak staffing needs: 8 AM - 9 AM (14 agents) and 9 PM - 12 AM. Night shift optimization is crucial for late-night calls, emphasizing the importance of morning boosts and nighttime optimization for improved service quality and call management. 📞🌙

## 🚀 Project Insights and Conclusions:

### 📊 Call Volume Patterns:

- Peak hours at 11 AM to 12 PM highlight the need for optimized staffing.

### 📈 Consistent Call Durations:

- Call durations within 190 to 205 seconds offer predictability.
- Efficient resource allocation can be maintained throughout the day.

### 👥 Agent Shortages:

- Addressing shortages during critical time periods is essential.

### 💡 Manpower Planning Strategies:

- Assumption-driven plans aim to reduce the abandon rate to 10%.
- Night shift optimization is vital for late-night call management.

These insights guide strategies to enhance customer experiences, optimize agent staffing, and address operational challenges within the call center.



## **APPENDIX**

### INSTAGRAM USER ANALYTICS:

<https://drive.google.com/file/d/1b2pIGkxIUvy8JTlmtaxB9kjYAk2Dhg7u/view?usp=sharing>

### HIRING PROCESS ANALYTICS:

[https://drive.google.com/file/d/1SgamNN81JCp1AjWtmkJTA7Uo2\\_821IEb/view?usp=sharing](https://drive.google.com/file/d/1SgamNN81JCp1AjWtmkJTA7Uo2_821IEb/view?usp=sharing)

### IMDB MOVIE ANALYSIS:

[https://drive.google.com/file/d/1VJ6Ou0yTvoMqSE8DhsFF7VShlsW\\_2Cr5/view?usp=sharing](https://drive.google.com/file/d/1VJ6Ou0yTvoMqSE8DhsFF7VShlsW_2Cr5/view?usp=sharing)



## BANK LOAN CASE STUDY:

<https://drive.google.com/file/d/1KbdV1NvFBy5W6mOP-oJvc8pVzlJkM7k0/view?usp=sharing>

## IMPACT OF CAR FEATURES :

<https://docs.google.com/presentation/d/1zs56UqYT4hl9RCtrxpA9SzkD9wpvx2X/edit?usp=sharing&oid=108547673521600619650&rtpof=true&sd=true>

## ABC CALL VOLUME ANALYSIS:

<https://drive.google.com/file/d/164rbsb6QMyQ0IK29IVe0cHmdtcTH1opU/view?usp=sharing>