

ANALYZING THE IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY

- DATA ANALYTICS PROJECT

DONE BY: ISWARIYA S

PROFESSIONAL BACKGROUND:

I am currently in my second year of pursuing a B.Tech in Computer Science and Engineering, maintaining a commendable CGPA of 8.35 (till 3rd semester). My skill set encompasses proficiency in MS Excel, data analysis, Python, and SQL. Despite being in the early stages of my academic journey, I have developed a strong foundation in these areas through coursework and self-study. I am eager to further enhance my skills and knowledge through practical experiences and challenges that the professional world has to offer. I am enthusiastic about applying my theoretical knowledge to real-world scenarios and am confident in my ability to adapt and learn quickly.

I have successfully executed 8 projects in SQL, Excel, and Machine Learning, focusing on data analytics. These projects stand as a testament to my proficiency in data analysis and interpretation. Armed with strong analytical skills and expertise in SQL, Excel, and Machine Learning, I offer valuable insights to the realm of data analytics.

TABLE OF CONTENTS

S.NO.	TITLE	PAGENO
1	Professional Background	
2	Table of contents	
3	Instagram User Analytics Description	
4	The Problem	
5	Findings	
6	Analysis	
7	Conclusion	
8	Hiring Process Analytics Description	
9	Design	
10	Findings	
11	Data Analysis	
12	Conclusion	
13	IMDB Movie Analysis Description	
14	The Problem	
15	Findings	
16	Conclusion	
17	Bank loan case study	
18	The Problem	
19	Design	
20	Findings	

21	Analysis	
22	Analyzing the Impact of Car Features on Price and Profitability	
23	Design	
24	Findings	
25	Conclusion	
26	ABC Call volume trend analysis	
27	Description	
28	Design	
29	Findings	
30	Conclusion	

INSTAGRAM USER ANALYTIC

DESCRIPTION:

This project aims to conduct a comprehensive analysis of Instagram users, delving into their behaviors, preferences, and interactions within the platform. Leveraging advanced data analytics and machine learning techniques, we will scrutinize user engagement patterns, content preferences, and the impact of various factors such as demographics and posting frequency on user activity. By examining the relationships between users and their content, we aspire to uncover meaningful insights that can inform marketing strategies, content creation, and platform optimization. Our analysis will not only provide a nuanced understanding of individual user profiles but also contribute to a broader comprehension of the dynamic social ecosystem on Instagram, empowering businesses and influencers to make informed decisions and enhance their digital presence.

PROBLEMS:

You are collaborating with Instagram's product team, and the product manager has tasked you with providing insights into inquiries posed by the management team.

DESIGN:

Steps implemented for data cleaning:

- Initially importing the provided datasets and merging them.
- Subsequently, eliminating duplicate entries and empty cells.
- Enhancing column headers with appropriate labels.
- Identifying and rectifying erroneous data entries.

FINDINGS:

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

Conclusion: These are the oldest users of Instagram.

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.Jacobson2 2016-05-14 07:56:26

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

Conclusion: These users were inactive after their first post.

5 Aniya_Hackett

7 Kasandra_Homenick

14 Jaclyn81

21 Rocio33

24 Maxwell.Halvorson

25 Tierra.Trantow

34 Pearl7

36 Ollie_Ledner37

41 Mckenna17
45 David.Osinski47
49 Morgan.Kassulke
53 Linnea59
54 Duane60
57 Julien_Schmidt
66 Mike.Auer39
68 Franco_Keebler64
71 Nia_Haag
74 Hulda.Macejkovic
75 Leslie67
76 Janelle.Nikolaus81
80 Darby_Herzog
81 Esther.Zulauf61
83 Bartholome.Bernhard
89 Jessyca_West
90 Esmeralda.Mraz57
91 Bethany20

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

Conclusion: He has the most likes in his post.

Zack_Kemmer93 145 <https://jarret.name> 48

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

Conclusion: These are some trending hashtags that a partner brand can use.

smile	59
beach	42
party	39
fun	38
food	24
lol	24
concert	24
hair	23
happy	22
beauty	20
dreamy	20
sunset	19
fashion	19

drunk	19
sunrise	17
landscape	17
style	17
photography	16
stunning	16
delicious	15
foodie	11

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

Conclusion: these days would be best for the AD campaign.

Thursday 16

Sunday 16

B) Investor Metrics

User Engagement: Are users still as active and post on Instagram or they are making fewer posts

Conclusion: A user's average post is more than 2.

2.5700

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

Conclusion: These are some users who can be bot and fake accounts.

Aniya_Hackett	257
Bethany20	257
Duane60	257
Jaclyn81	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

ANALYSIS:

The analysis of the Instagram User Analytic project produced several significant findings:

- User Demographics: The analysis uncovered insights into the age groups, gender distribution, and geographical locations of Instagram users. This information aids in understanding the target audience and enables targeted marketing campaigns.

- Engagement Metrics: The project examined engagement metrics like likes, comments, and shares, identifying the most engaging content types and factors driving user engagement. These insights can inform content creation strategies to increase user interaction.
- Content Performance: The analysis evaluated the performance of different content formats (photos, videos, stories), identifying those that generate the highest engagement. This can guide content creation for maximum impact.
- Follower Growth: The project analyzed follower growth trends, providing insights into acquisition and retention patterns. This helps develop strategies to increase followers and maintain audience loyalty.
- Influencer Impact: The analysis studied the impact of influencer collaborations on engagement and follower growth, offering insights into effective influencer partnerships for expanding brand reach.

Overall, the analysis of the Instagram User Analytics project provided valuable insights into user behavior, content performance, and follower growth. These findings can optimize marketing strategies, create engaging content, and target the desired audience effectively on Instagram.

CONCLUSION:

In conclusion, the Instagram User Analytics project involved analyzing and interpreting data related to Instagram users. Through various analytical techniques, valuable insights into user demographics, engagement patterns, and content performance were gained. This project provided a comprehensive understanding of Instagram user behavior and identified strategies to optimize engagement and reach on the platform. The findings and recommendations derived from this analysis can enhance marketing strategies, target specific audience segments, and drive meaningful user interactions on Instagram.

OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

DESCRIPTION:

The Operation Analytics and Investigating Metric Spike Data Analytics project is a strategic initiative designed to empower organizations with enhanced insights into their operational data. Focused on real-time monitoring and analysis, the project employs advanced data analytics techniques to detect and investigate metric spikes efficiently. By leveraging cutting-edge technologies and algorithms, the system aims to provide actionable intelligence for prompt decision-making. The project encompasses the development of a comprehensive analytics platform that integrates seamlessly with existing systems, ensuring a streamlined approach to data collection, processing, and interpretation. With a primary goal of optimizing operational performance, this project is poised to deliver valuable outcomes for businesses seeking to proactively address issues, improve efficiency, and make informed strategic decisions based on a thorough understanding of their metrics and data patterns.

DESIGN:

- Established the database structure and tables: Developed a database and subsequently constructed tables based on the provided structure and interconnections.
- Conducted an in-depth analysis: Employed SQL to thoroughly analyze the dataset, addressing the posed questions. After reviewing the provided dataset, tables were created to compute diverse queries.
- Integrated Data for Informed Decision-Making: Combined data fragments, organized the table structure for extracting business insights, retrieved the necessary outcomes, and thereby generated valuable insights for the company, enabling informed and strategic decision-making.

ANALYSIS:

CASE STUDY 1: JOB DATA ANALYSIS

1. Jobs Reviewed Over Time:

```
#TASK-1(Jobs Reviewed Over Time)
• select * from job_data;
• select avg(t) as 'avg jobs reviewed per day per hour',
  avg(p) as 'avg jobs reviewed per day per second'
  from
  (select
  ds,
  ((count(job_id)*3600)/sum(time_spent)) as t,
  ((count(job_id))/sum(time_spent)) as p
  from
  job_data
  where
  month(ds)=11
  group by ds) a;
```

Average jobs reviewed per day per hour - 126.18048333

Average jobs reviewed per day per second - 0.03505000

2. Throughput Analysis:

```
SELECT ROUND(COUNT(event)/SUM(time_spent), 2) AS "Weekly Throughput" FROM job_data;

SELECT ds AS Dates, ROUND(COUNT(event)/SUM(time_spent), 2) AS "Daily Throughput" FROM job_data
GROUP BY ds ORDER BY ds;
```

For Daily Throughput :

Dates	Daily Throughput
2020-11-25	0.02
2020-11-26	0.02
2020-11-27	0.01
2020-11-28	0.06
2020-11-29	0.05
2020-11-30	0.05

For Weekly Throughput :

Weekly Throughput
► 0.03

3. Language Share Analysis:

```
SELECT language AS Languages, ROUND(100 * COUNT(*)/TOTAL, 2) AS Percentage, sub.total
FROM job_data
CROSS JOIN (SELECT COUNT(*) AS total FROM job_data) AS sub
GROUP BY language, sub.total;
```

	Languages	Percentage	total
►	English	12.50	16
	Arabic	12.50	16
	Persian	37.50	16
	Hindi	12.50	16
	French	12.50	16
	Italian	12.50	16

4. Duplicate Rows Detection:

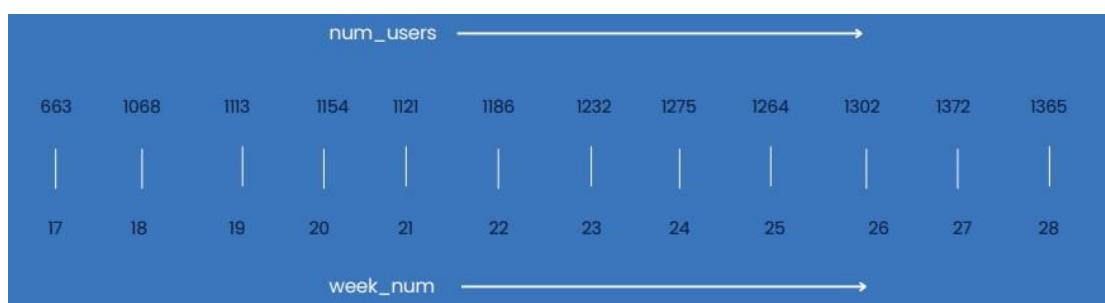
```
SELECT actor_id, COUNT(*) AS Duplicates FROM job_data  
GROUP BY actor_id HAVING COUNT(*) > 1;
```

	actor_id	Duplicates
▶	1001	2
	1006	2
	1003	4
	1005	2
	1002	2
	1007	2
	1004	2

CASE STUDY 2: INVESTIGATING METRIC SPIKE

1. Weekly User Engagement:

```
select extract(week from occurred_at) as week_number,  
count(distinct user_id) as active_user  
from events_tb1  
where event_type='engagement'  
group by week_number  
order by week_number
```



2. User Growth Analysis:

```
select year, week_num, num_users, sum(num_users)
over (order by year, week_num) as cum_users
from (
select extract(year from created_at) as year, extract(week from created_at) as week_num, count(distinct user_id) as num_users
from users_tbl
where state='active'
group by year, week_num
order by year, week_num)sub
```

year	week_num	num_users	cum_users
2013	0	23	23
2013	1	30	53
2013	2	48	101
2013	3	36	137
2013	4	30	167
2013	5	48	215
2013	6	38	253
2013	7	42	295
2013	8	34	329
2013	9	43	372
2013	10	32	404
2013	11	31	435
2013	12	33	468

3. Weekly Retention Analysis:

```
with cte1 as (
select distinct user_id,
Extract (week from occurred at) as signup_week
from events_tb1
where event_type = 'signup_flow'
and event_name = 'complete_signup' and extract (week from occurred at) = 18 ),
cte2 as (select distinct user_id,
Extract (week from occurred at) as engagement_week
from events_tb1
where event_type = 'engagement')
select count(user_id) total_engaged_users,
sum(case when retention_week > 8 then 1 else end) as retained_users
from (select a.user_id, a.signup_week,
b.engagement_week, b.engagement_week-a.signup_week as retention_week
from cte1 a
LEFT JOIN cte2 b
on a.user_id = b.user_id
order by a.user_id) sub
```

total_engaged_users

317

retained_users

236

4. Weekly Engagement Per Device:

```
with cte as (select extract(year from occurred_at)||'-'||extract(week from occurred_at) as weeknum
device, count(distinct user_id) as usercnt
from events_tb1
where event_type = 'engagement'
group by weeknum, device
order by weeknum)
select weeknum, device, usercnt
from cte
```

weeknum	device	usercnt
2014-18	acer aspire desktop	10
2014-18	acer aspire notebook	21
2014-18	amazon fire phone	4
2014-18	asus chromebook	23
2014-18	dell inspiron desktop	21

5. Email Engagement Analysis:

```
select
100 * sum(case when email_cat = 'email_open' then 1 else 0 end)/
sum(case when email_cat = 'email sent' then 1 else 0 end) as email_open_rate,
100 * sum(case when email_cat = 'email clicked' then 1 else 0 end)/
sum(case when email_cat = 'email sent' then 1 else 0 end) as email_click_rate
from (select *,
case
when action in ('sent_weekly_digest', 'sent_reengagement_email') then 'email_sent'
when action in ('email_open') then 'email_open'
when action in ('email_clickthrough') then 'email_clicked'
end as email_cat
from email_events) sub
```

31.1921

email_open_rate

10.4745

email_click_rate

CONCLUSION:

In this project, I learned advanced SQL techniques like Windows Functions, improved my SQL skills, and discovered how to ask the right questions for business growth. I gained insights into finding valuable information, understanding areas for improvement in companies, and investigating changes in metrics.

HIRING PROCESS ANALYTICS

DESCRIPTION:

In the "Hiring Process Analytics" project, the objective is to refine and optimize hiring procedures within a multinational setting. As a data analyst, insights will be extracted from historical hiring data, focusing on rejection rates, interviews, job types, and vacancies. The analysis aims to provide actionable knowledge to enhance overall recruitment strategies, ensuring the acquisition of top-tier talent for sustained organizational growth and innovation.

DESIGN:

The data cleaning process involved several steps:

- Initially, the provided datasets were imported and merged.
- Duplicate entries and blank cells were then removed.
- The column headers were enhanced with appropriate labels.
- Erroneous data entries were identified and corrected.

FINDINGS:

1. How many males and females are Hired?

event_name	Count of event_name
-	15
Don't want to say	393
Female	2675

Male 4085

(blank)

Grand Total 7168

2. What is the average salary offered in this company?

Analysis	Result	Formula
----------	--------	---------

Offered Salary Mean	49983.02902	Average
---------------------	-------------	---------

Offered Salary Median	49625	Median
-----------------------	-------	--------

Minimum Salary	100	Minimum Salary
----------------	-----	----------------

Maximum Salary	400000	Maximum Salary
----------------	--------	----------------

3. Draw the class intervals for salary in the company.

Offered Salary	Count of Offered Salary
----------------	-------------------------

(blank)

100-50099	3614
-----------	------

50100-100099	3550
--------------	------

150100-200099	1
---------------	---

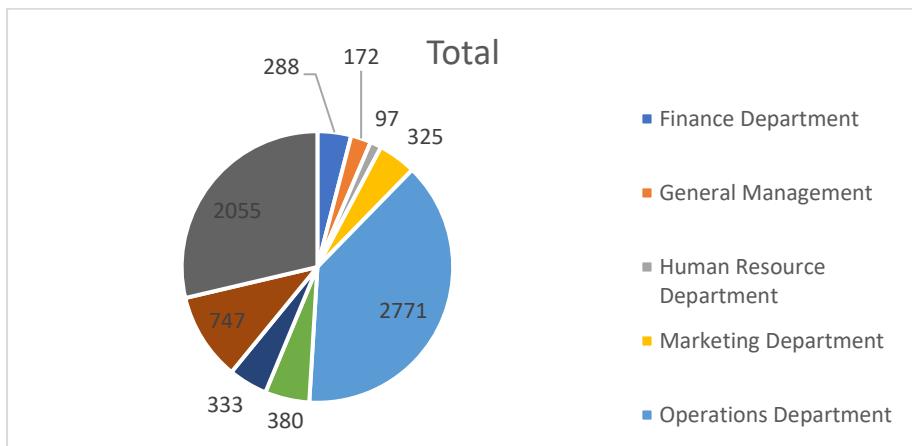
250100-300099	1
---------------	---

350100-400000	1
---------------	---

Grand Total	7167

4. Draw a Pie Chart / Bar Graph (or any other graph) to show the proportion of people working in different departments.

Department	Count of Department
Finance Department	288
General Management	172
Human Resource Department	97
Marketing Department	325
Operations Department	2771
Production Department	380
Purchase Department	333
Sales Department	747
Service Department	2055
Grand Total	7168



5. Represent different post tiers using a chart/graph.

Post Name	Count of Post Name
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
Grand Total	7168

ANALYSIS:

The analysis of Hiring Process Analytics involved a comprehensive examination of the various stages and components of the hiring process within an organization. Key findings included:

- Time-to-Hire: The analysis determined the average duration of the hiring process, from job posting to final offer. It identified bottlenecks and delays, such as extended candidate screening or interview scheduling, which affected the overall time-to-hire.
- Sourcing Channels: The analysis evaluated the effectiveness of different sourcing channels, including job boards, employee referrals, and social media platforms. It identified the most successful channels in attracting qualified candidates, helping to optimize recruitment strategies and allocate resources more effectively.
- Candidate Evaluation: The analysis assessed the methods and criteria used to evaluate candidates' qualifications, skills, and cultural fit. It identified areas for improvement in the evaluation process, such as the need for standardized interview questions or additional assessment tools.
- Candidate Experience: The analysis examined the experience of candidates throughout the hiring process, from initial application to final decision. It identified pain points or areas where candidates may have had a negative experience, such as poor communication or lengthy response times, and recommended measures to enhance the candidate experience.
- Cost Analysis: The analysis included an assessment of the costs associated with the hiring process, including advertising, recruitment agencies, and internal resources. It identified opportunities to optimize costs without compromising the quality of hires.

CONCLUSION:

The examination of Hiring Process Analytics yielded valuable insights into the efficiency and effectiveness of the organization's hiring process. By pinpointing areas for enhancement and applying suitable actions, the organization can simplify the process, decrease time-to-hire, and improve the overall candidate experience. The analysis findings empower the organization to make data-backed decisions in optimizing sourcing channels, refining candidate evaluation methods, and enhancing communication with candidates. This, in turn, aids in attracting and retaining top talent, ensuring a robust pool of qualified candidates for future roles.

IMDB MOVIE ANALYSIS

DESCRIPTION:

The dataset pertains to IMDB Movies, with a focus on understanding the factors influencing a movie's success, defined by high IMDB ratings. This research is valuable for movie industry stakeholders to make informed decisions. The project involves analyzing a dataset containing 28 columns and 5044 rows of IMDB Movies, including movie names, release year, actors, budget, gross, and genre. The process includes defining the problem, cleaning the data, exploring insights, and using root cause analysis. Visual representations like charts and graphs are employed to answer specific questions and create reports for data-driven decision-making.

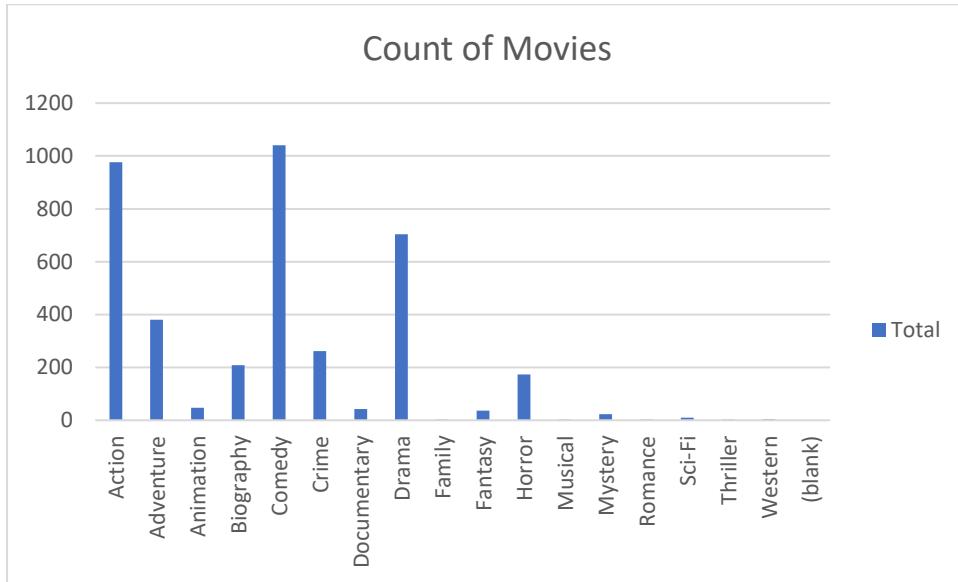
THE PROBLEM:

The IMDB Movie Analysis project aimed to tackle the issue of limited insights into movie genres, ratings, box office performance, and the influence of directors and actors. Its objective was to analyze data from the Internet Movie Database (IMDB) to offer valuable information and meet the demand for data-driven decision-making in the film industry.

FINDINGS:

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

1. Movie Genre Analysis:

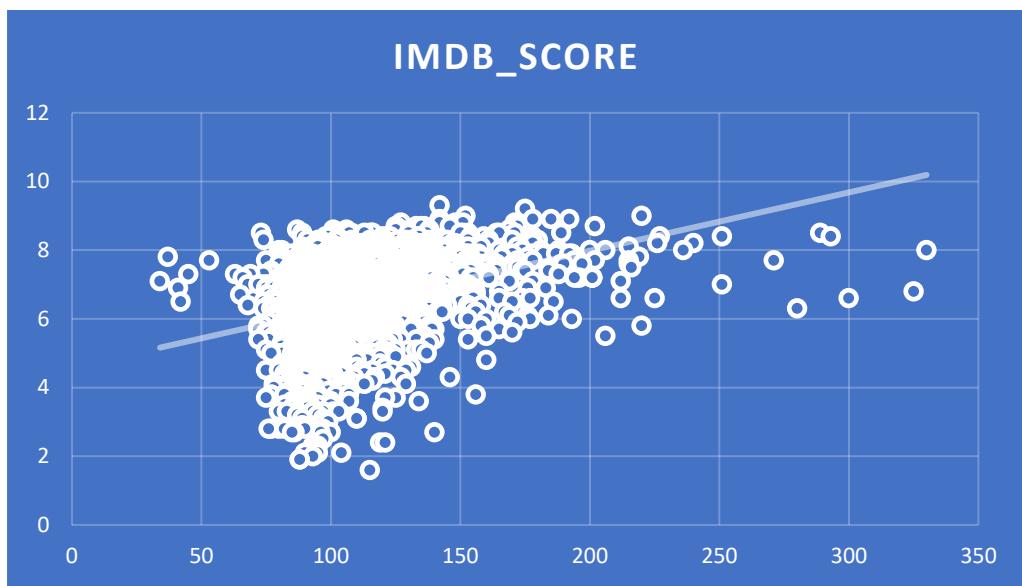


The most common movie genre is Comedy – 1041

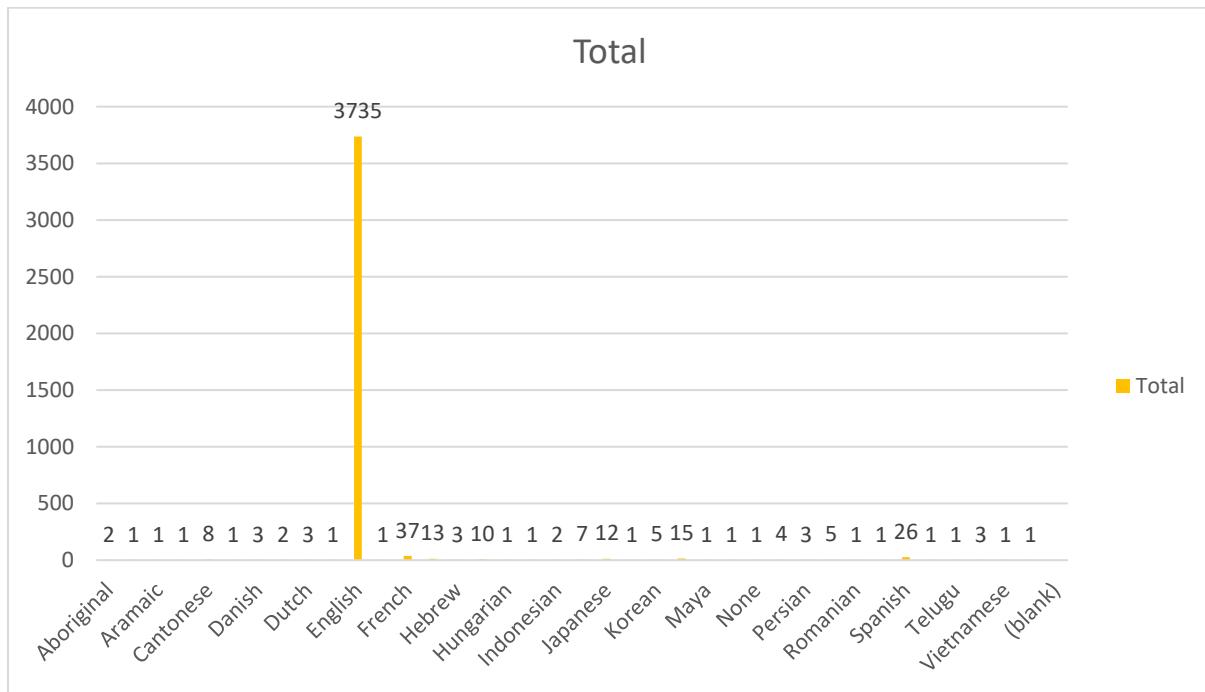
2. Movie Duration Analysis:

Descriptive Analysis:

mean	6.449859658
Standard Error	0.017050391
Median	6.6
Mode	6.7
Standard Deviation	1.067387146
Sample Variance	1.13931532
Range	7.7
Minimum	1.6
Maximum	9.3
Sum	25277
Count	3919

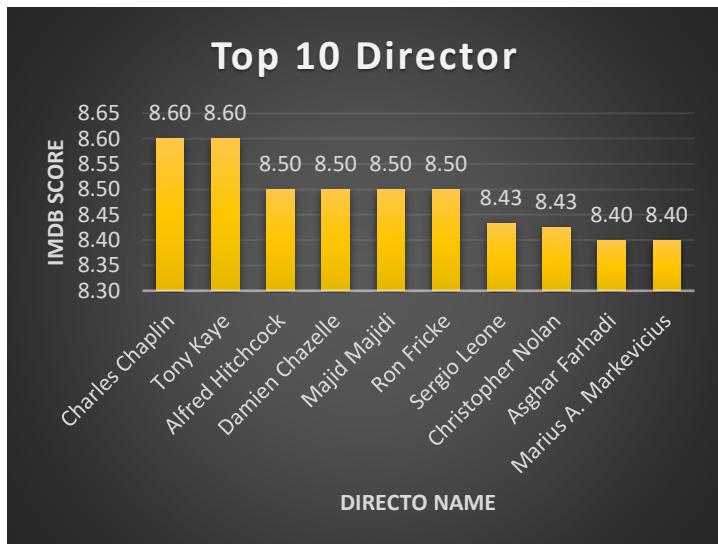


3. Language Analysis:



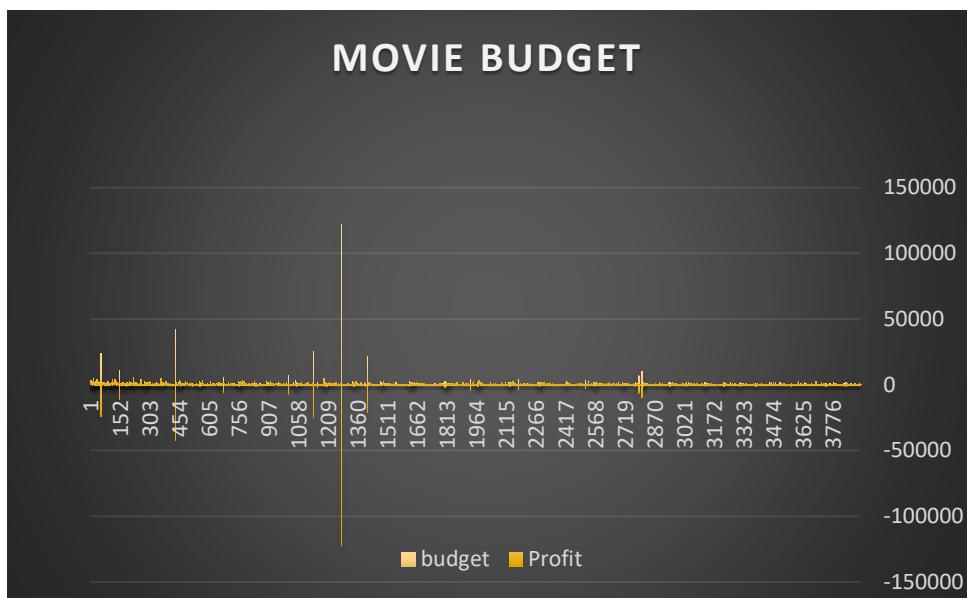
4. Director Analysis:

Top 10 Directors	Average of imdb_score
Charles Chaplin	8.60
Tony Kaye	8.60
Alfred Hitchcock	8.50
Damien Chazelle	8.50
Majid Majidi	8.50
Ron Fricke	8.50
Sergio Leone	8.43
Christopher Nolan	8.43
Asghar Farhadi	8.40
Marius A. Markevicius	8.40



5. Budget Analysis:

Hight profit margin movie	HIGHT PROFIT MARGIN
The Hobbit: The Desolation of Smaug	523505847



CORRELATION - 0.102177361

ANALYSIS:

- The dataset highlighted comedy as the most common genre (1041 instances) and English as the predominant language (3735 instances).
- Descriptive statistics revealed variations in IMDB scores across genres.
- Analysis showed a mild positive correlation between movie durations and IMDB scores, indicating slightly higher ratings for longer movies.
- Top directors with consistently high IMDB scores significantly influenced movie success compared to the overall distribution.
- Correlation analysis indicated a moderate positive relationship (0.102177361) between movie budgets and gross earnings.
- Movies with the highest profit margins totaled 523,505,847.

CONCLUSION:

In conclusion, this project has offered a profound insight into the factors impacting the success of movies on IMDB. From the initial data cleaning to the 'Five Whys' analysis and individual data analytics tasks, the acquired insights can empower stakeholders in the movie industry to make informed, data-driven decisions, elevate the quality of movies, and increase audience satisfaction. The comprehensive report generated by the project encapsulates these valuable insights, offering a convenient reference for decision-making purposes.

BANK LOAN CASE STUDY

DESCRIPTION:

This project immerses participants in the practical application of Exploratory Data Analysis (EDA) within the banking sector, offering a hands-on exploration of risk analytics. It goes beyond conventional EDA methods to illuminate the intricate factors influencing lending decisions. By examining variables such as credit scores and income levels, participants gain insights into how data analysis reduces the risk of financial losses in consumer lending. The case study not only hones EDA skills but also imparts a foundational understanding of risk analytics in financial services. It serves as a practical bridge between theory and application, shedding light on the pivotal role data plays in shaping strategic decisions within the dynamic landscape of banking.

APPROACH:

This case study involves two extensive datasets: one for the current application and another for the previous application. Both datasets contained numerous unnecessary columns and a significant amount of blank data, which were irrelevant for risk assessments. To begin the analysis, I started by cleaning the data. To evaluate these large datasets, I initially cleaned the data, identified and removed outliers, and then proceeded with univariate and bivariate analysis using pivot tables and charts. The following technology stack was utilized: MySQL Workbench 8.0 CE, and Microsoft Excel 2010.

FINDINGS:

FILE 1: APPLICATION DATA

1. Identify Missing Data and Deal with it Appropriately:

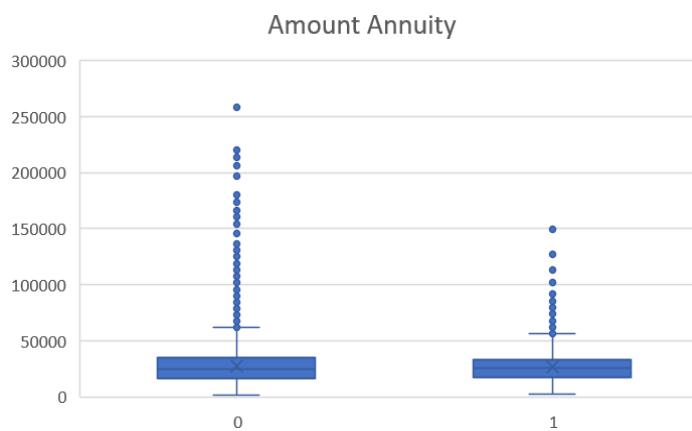
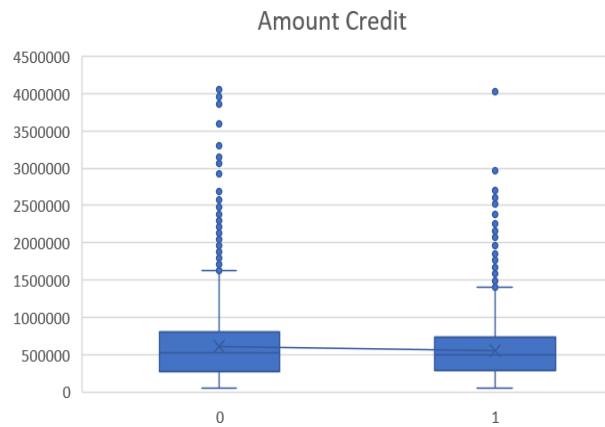
COLUMNS TO DROP :

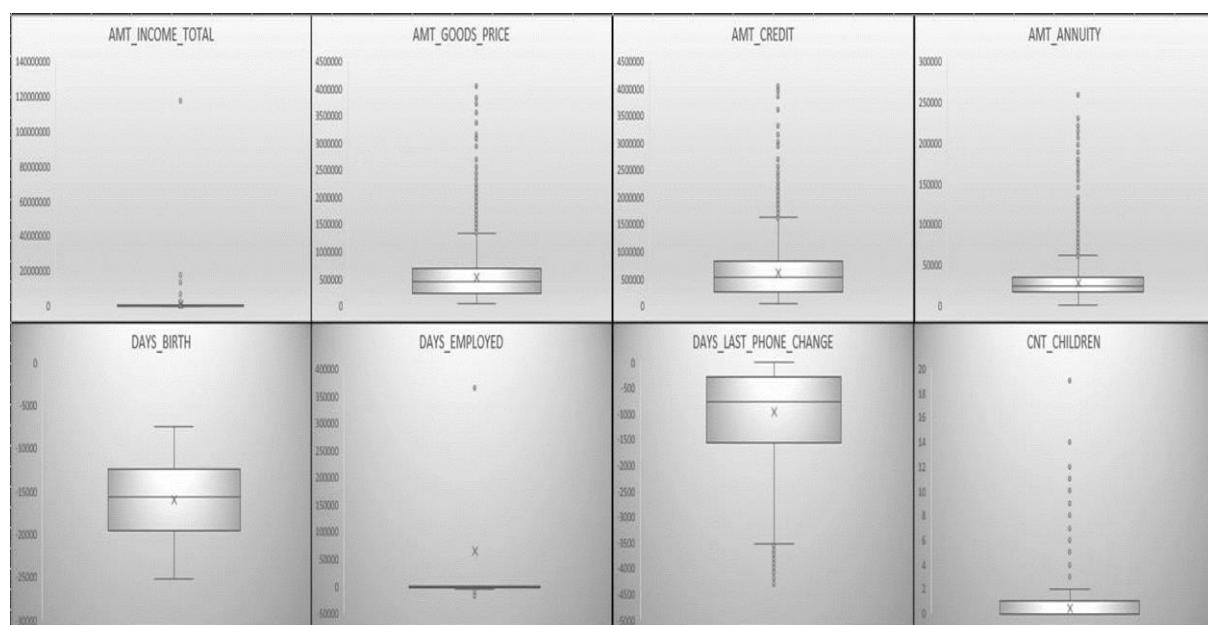
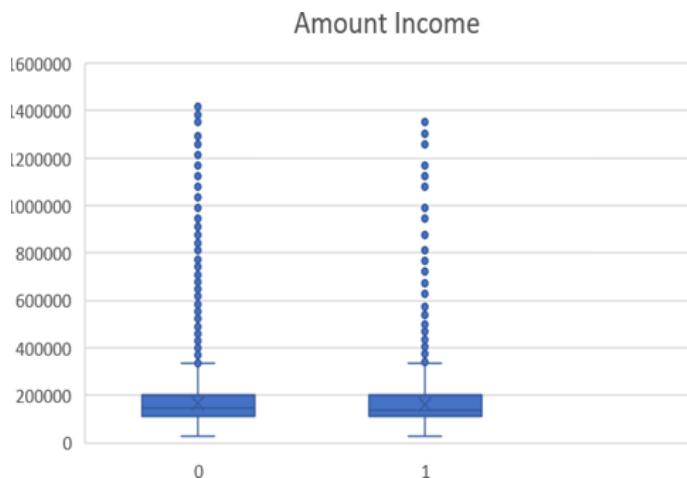
	HOUSETYPE_MODE	WAL LSMATERIAL_MODE	BASEMENTAREA_MEDI	FLOORSMIN_MEDI
LIVINGAREA_AVG	ELEVATORS_AVG	LANDAREA_AVG	LIVINGAPARTMENTS_AVG	
LIVINGAREA_MODE	ELEVATORS_MODE	LANDAREA_MODE	LIVINGAPARTMENTS_MODE	
LIVINGAREA_MEDI	ELEVATORS_MEDI	LANDAREA_MEDI	LIVINGAPARTMENTS_MEDI	
ENTRANCES_AVG	NONLIVINGAREA_AVG	OWN_CAR_AGE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

AMT_ANNUITY	YEARS_BEGINEXPLUATATION_MODE	DEF_60_CNT_SOCIAL_CIRCLE
AMT_GOODS_PRICE	FLOORSMAX_MODE	DAYS_LAST_PHONE_CHANGE
NAME_TYPE_SUITE	YEARS_BEGINEXPLUATATION_MEDI	AMT_REQ_CREDIT_BUREAU_HOUR
OCCUPATION_TYPE	FLOORSMAX_MEDI	AMT_REQ_CREDIT_BUREAU_DAY
CNT_FAM_MEMBERS	TOTALAREA_MODE	AMT_REQ_CREDIT_BUREAU_WEEK
EXT_SOURCE_2	EMERGENCYSTATE_MODE	AMT_REQ_CREDIT_BUREAU_MON
EXT_SOURCE_3	OBS_30_CNT_SOCIAL_CIRCLE	AMT_REQ_CREDIT_BUREAU_QRT
YEARS_BEGINEXPLUATATION_AVG	DEF_30_CNT_SOCIAL_CIRCLE	AMT_REQ_CREDIT_BUREAU_YEAR
FLOORSMAX_AVG	OBS_60_CNT_SOCIAL_CIRCLE	

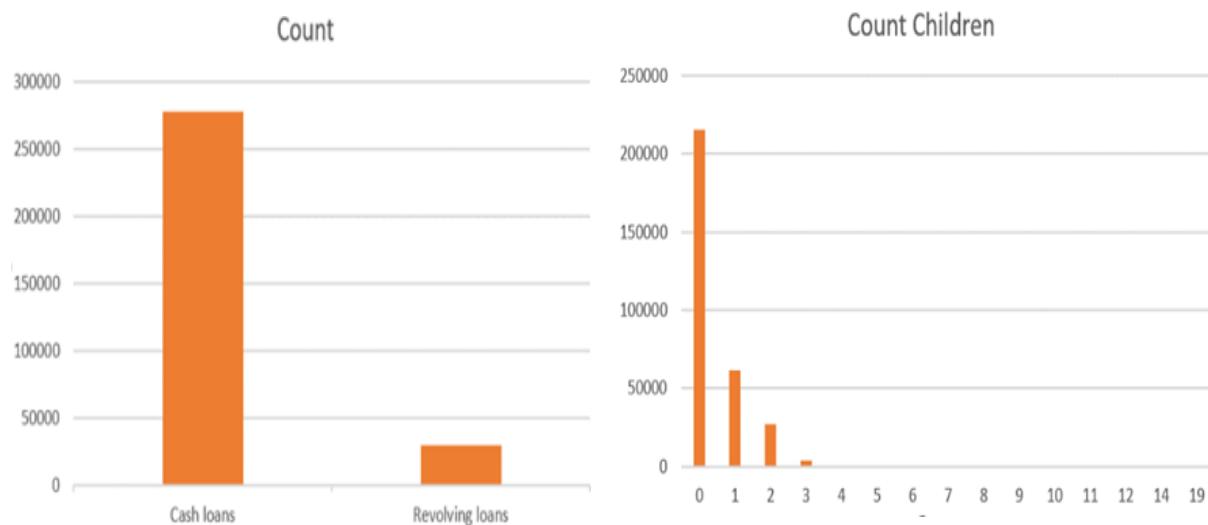
FEATURES TO DROP		FEATURES TO KEEP	
FLOORSMAX_AVG		OCCUPATION_TYPE	OBS_30_CNT_SOCIAL_CIRCLE
FLOORSMAX_MODE		AMT_REQ_CREDIT_BUREAU_HOUR	DEF_30_CNT_SOCIAL_CIRCLE
FLOORSMAX_MEDI		AMT_REQ_CREDIT_BUREAU_DAY	OBS_60_CNT_SOCIAL_CIRCLE
EXT_SOURCE_2		AMT_REQ_CREDIT_BUREAU_WEEK	DEF_60_CNT_SOCIAL_CIRCLE
YEARS_BEGINEXPLUATATION_AVG		AMT_REQ_CREDIT_BUREAU_MON	DAYS_LAST_PHONE_CHANGE
YEARS_BEGINEXPLUATATION_MODE		AMT_REQ_CREDIT_BUREAU_QRT	AMT_GOODS_PRICE
YEARS_BEGINEXPLUATATION_MEDI		AMT_REQ_CREDIT_BUREAU_YEAR	AMT_ANNUITY
TOTALAREA_MODE		NAME_TYPE_SUITE	CNT_FAM_MEMBERS
EXT_SOURCE_3			
EMERGENCYSTATE_MODE			

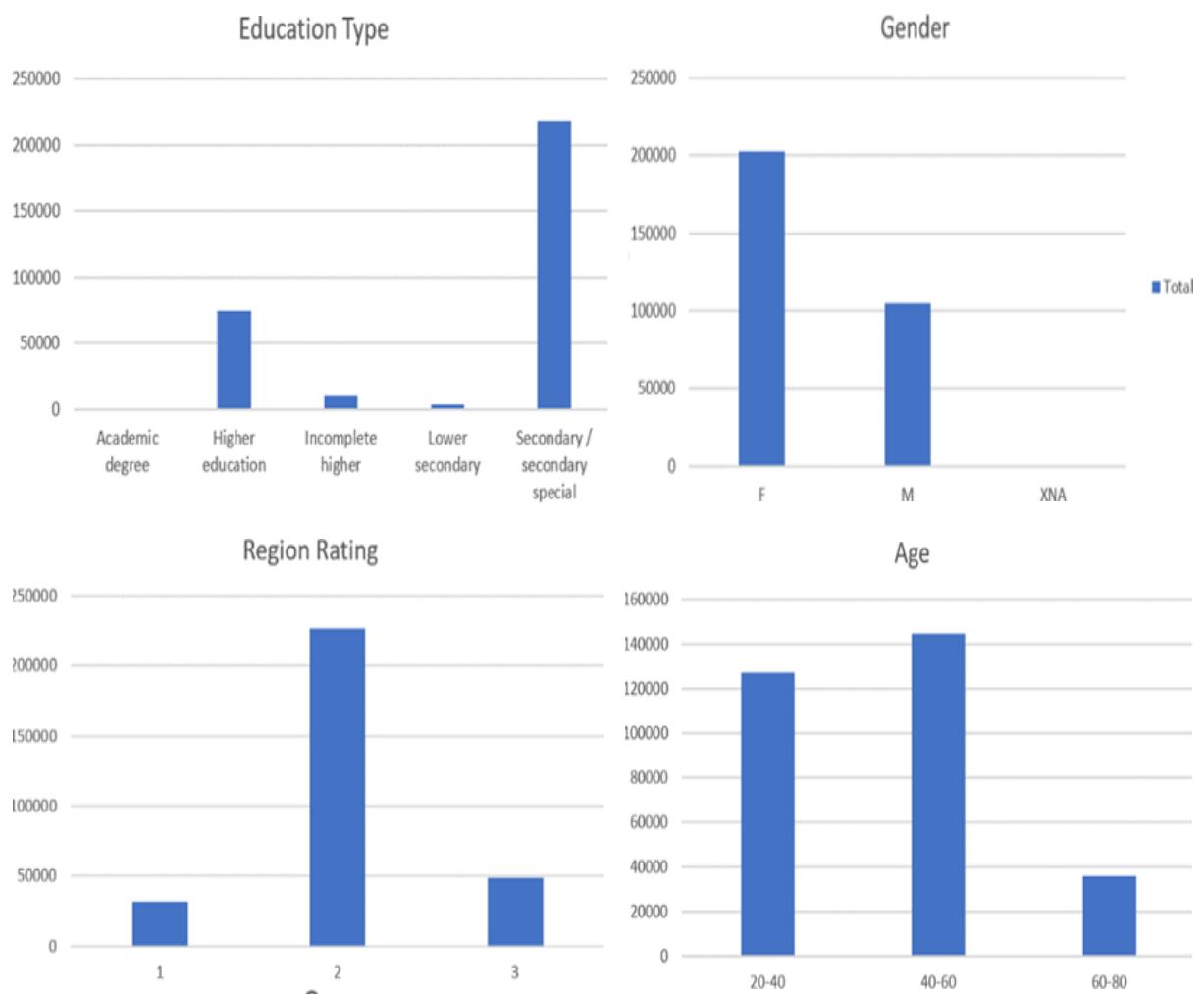
2. Identify Outliers in the Dataset:





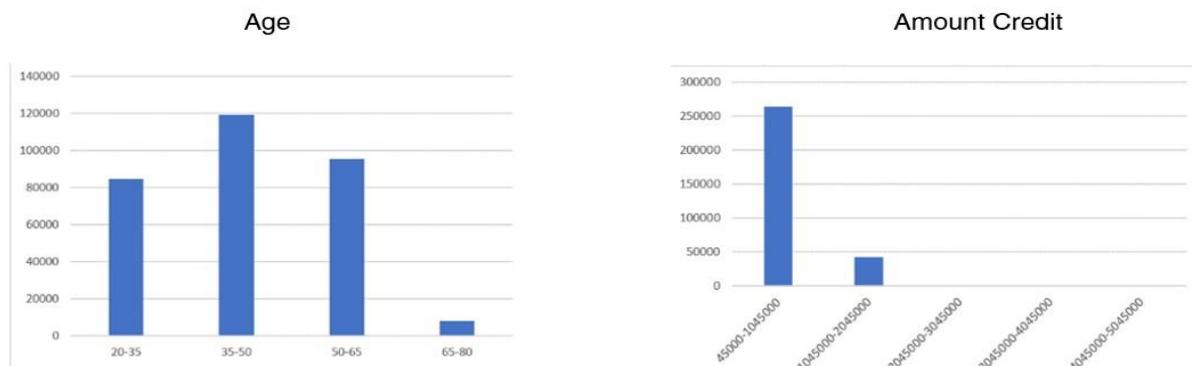
C. Analyze Data Imbalance:

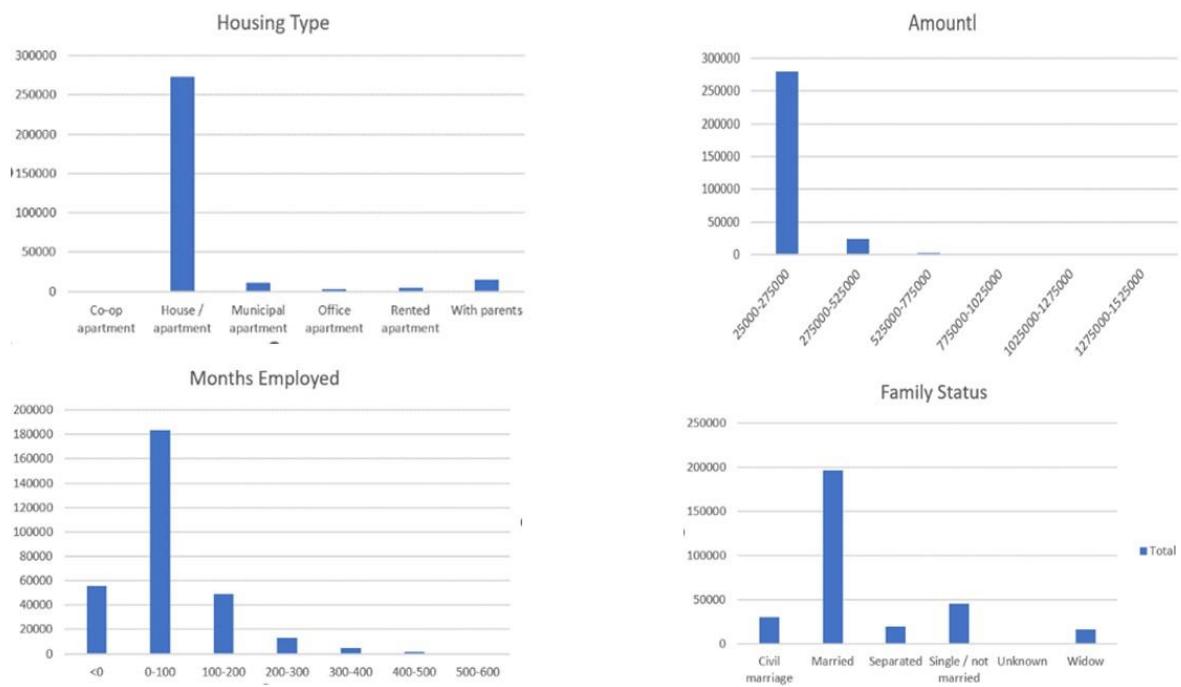




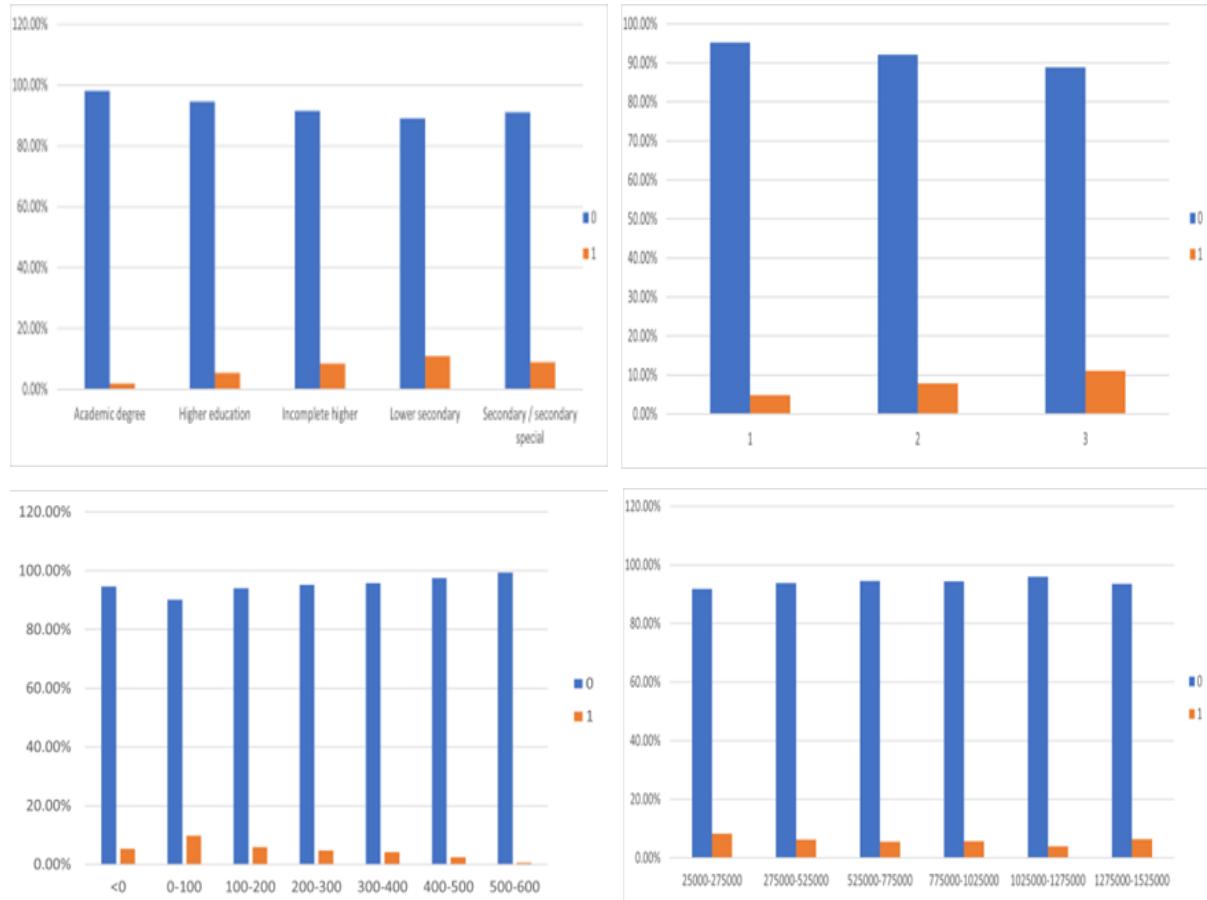
4. Perform Univariate, Segmented Univariate, and Bivariate Analysis:

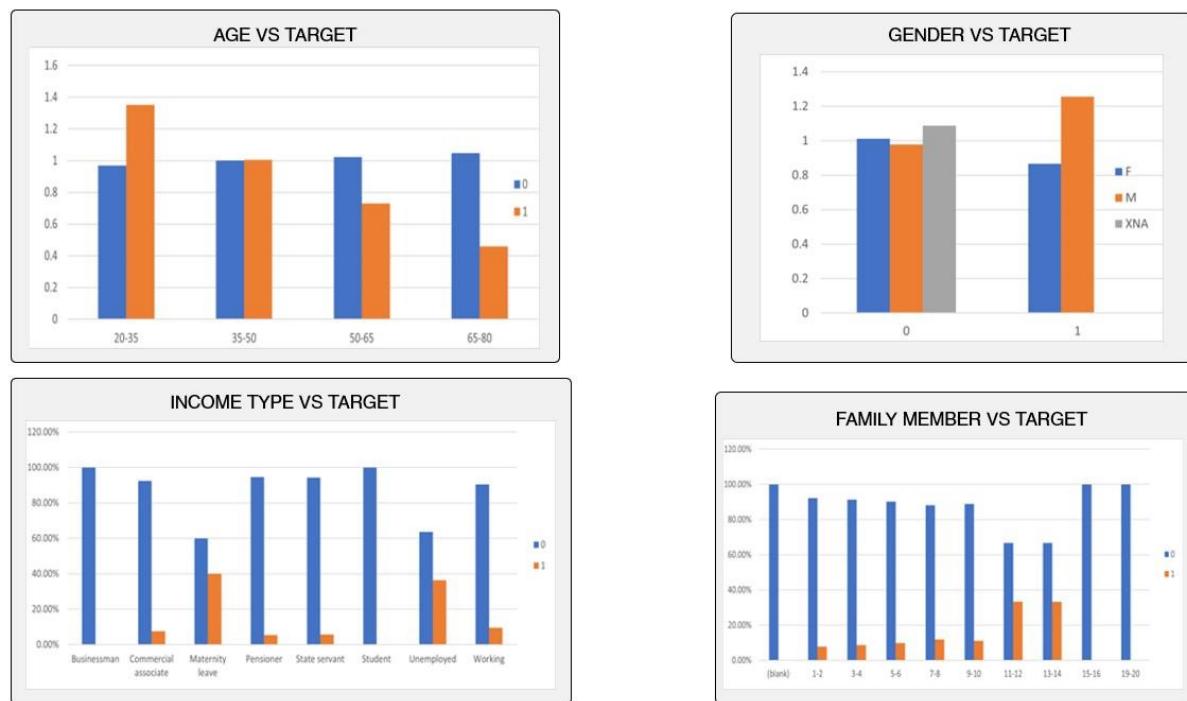
- **Univariate Analysis:**





- Bivariate Analysis:





E. Identify Top Correlations for Different Scenarios:

Top 10 driving factors in current application.csv

1	Income type
2	Count of Family Members
3	Children count
4	External source
5	Region rating of client
6	Age
7	Months Employed
8	Amount credit
9	Amount Goods Price
10	Amount total income

ANALYSIS:

Recommended Loan Groups:

1. Clients with approved previous applications
2. Married individuals
3. Senior citizens
4. Clients with higher educational qualifications
5. Individuals with high-income levels
6. Clients with significant external sources of income
7. Female applicants
8. Customers with substantial work experience

High-Risk Loan Groups:

1. Unemployed clients
2. Young clientele
3. Customers with a history of denied prior applications
4. Low-income individuals
5. Clients lacking sufficient external sources of income
6. Customers with limited work experience
7. Clients on Maternity Leave
8. Individuals with a larger number of family members

CONCLUSION:

Through this project, we aimed to identify patterns predicting customer payment challenges, aiding decisions on loan approval, denial, or adjustment. The insights gained enhance the company's ability to make informed decisions, particularly in understanding key factors influencing loan defaults. The project's success lies in providing actionable insights, contributing to more effective risk assessment and improved decision-making in the Bank Loan Case Study.

Analyzing the Impact of Car Features on Price and Profitability

PROJECT DESCRIPTION:

The automotive sector has undergone significant changes in recent decades, emphasizing fuel efficiency, eco-friendliness, and technological advancement. With fierce competition and shifting consumer preferences, understanding the drivers of car demand is crucial. One way to tackle this challenge is by examining the correlation between a vehicle's attributes, market segment, and pricing, pinpointing the most sought-after features and segments for consumers, as well as the most profitable for manufacturers. Employing analytical methods like regression analysis and market segmentation can aid in crafting a pricing strategy that balances consumer demand and profitability. Moreover, it can help identify key features for future product development, enhancing the manufacturer's competitive edge and long-term profitability.

PROBLEM:

Studying the correlation between a car's features and its popularity: By analyzing the popularity metric within the dataset, a data analyst can determine which features are most favored by consumers and how they impact a car's popularity. This insight can assist manufacturers in making informed choices regarding product development and marketing strategies. Forecasting a car's price based on its features and market segment: Utilizing the dataset's various feature and market segment variables, a data analyst can construct a model to forecast a car's price. This can aid both manufacturers and consumers in comprehending how different features influence a car's price, enabling informed decisions regarding pricing and purchasing.

In general, this dataset could be a valuable asset for data analysts interested in exploring diverse facets of the automotive industry, potentially offering insights that can guide decisions related to product development, marketing, and pricing. I used the provided dataset of cars. Initially, I familiarized myself with the data and then performed data cleanup. For data cleanup, I removed rows with blank cells and duplicate entries.

DESIGN:

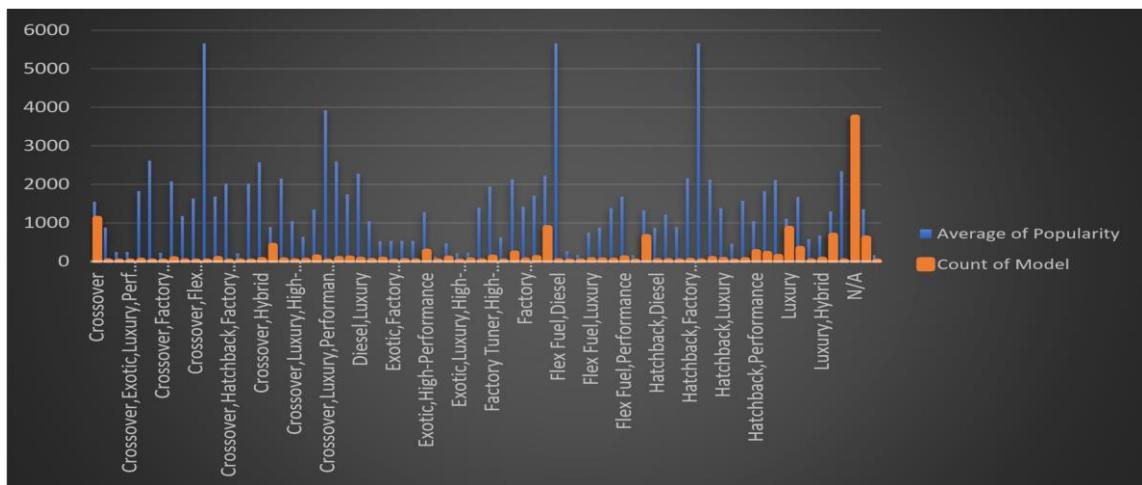
For analytical purposes, I utilized MS Excel to execute various functions such as pivot tables, graphs, charts, and regression analysis. This choice was made due to its user-friendly interface and the ability to create visually appealing dashboards. Tech Stack Used: MS Excel and MS Word.

ANALYSIS:

1. How does the popularity of a car model fluctuate across various market categories?

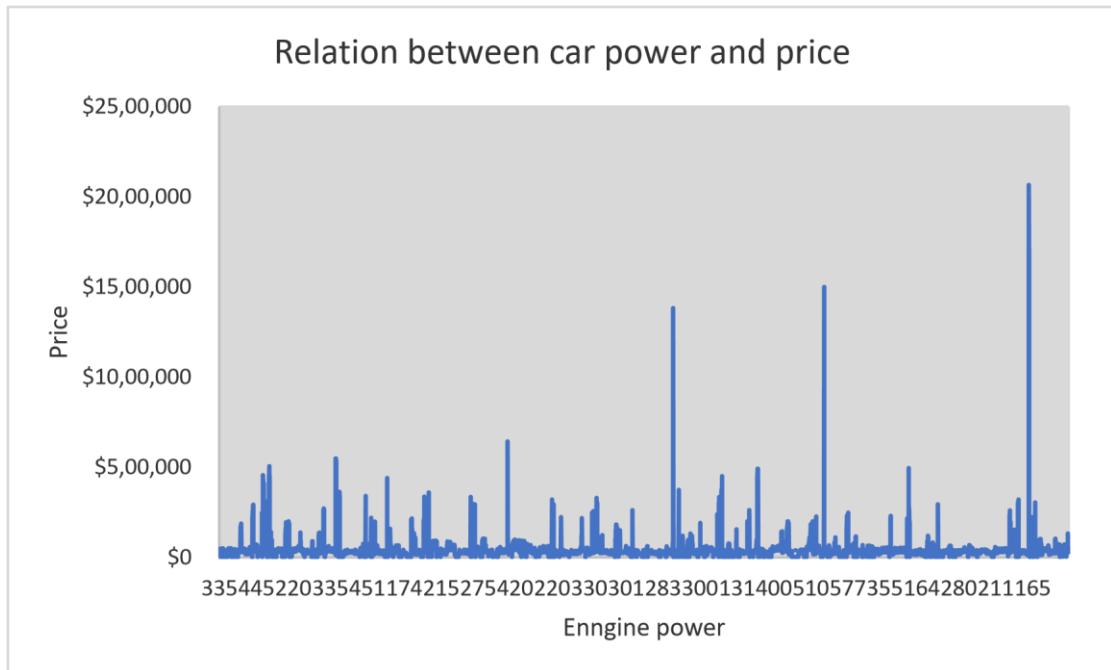
- Task 1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.
- Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.

Market Category	Average of Popularity	Count of Model			
2 Crossover	1545.263063	1110	38 Factory Tuner,High-Performance	1941.415094	106
3 Crossover,Diesel	873	7	39 Factory Tuner,Luxury	617	2
4 Crossover,Exotic,Luxury,High-Performance	238	1	40 Factory Tuner,Luxury,High-Performance	2133.367442	215
5 Crossover,Exotic,Luxury,Performance	238	1	41 Factory Tuner,Luxury,Performance	1413.419355	31
6 Crossover,Factory Tuner,Luxury,High-Performance	1823.461538	26	42 Factory Tuner,Performance	1695.695652	92
7 Crossover,Factory Tuner,Luxury,Performance	2607.4	5	43 Flex Fuel	2217.302752	872
8 Crossover,Factory Tuner,Performance	210	4	44 Flex Fuel,Diesel	5657	16
9 Crossover,Flex Fuel	2073.75	64	45 Flex Fuel,Factory Tuner,Luxury,High-Performance	258	1
10 Crossover,Flex Fuel,Luxury	1173.2	10	46 Flex Fuel,Hybrid	155	2
11 Crossover,Flex Fuel,Luxury,Performance	1624	6	47 Flex Fuel,Luxury	746.5384615	39
12 Crossover,Flex Fuel,Performance	5657	6	48 Flex Fuel,Luxury,High-Performance	878.9090909	33
13 Crossover,Hatchback	1675.694444	72	49 Flex Fuel,Luxury,Performance	1380.071429	28
14 Crossover,Hatchback,Factory Tuner,Performance	2009	6	50 Flex Fuel,Performance	1680.471264	87
15 Crossover,Hatchback,Luxury	204	7	51 Flex Fuel,Performance,Hybrid	155	2
16 Crossover,Hatchback,Performance	2009	6	52 Hatchback	1318.865835	641
17 Crossover,Hybrid	2563.380952	42	53 Hatchback,Diesel	873	14
18 Crossover,Luxury	884.5487805	410	54 Hatchback,Factory Tuner,High-Performance	1205.153846	13
19 Crossover,Luxury,Diesel	2149.411765	34	55 Hatchback,Factory Tuner,Luxury,Performance	886.8888889	9
20 Crossover,Luxury,High-Performance	1037.222222	9	56 Hatchback,Factory Tuner,Performance	2159.045455	22
21 Crossover,Luxury,Hybrid	630.9166667	24	57 Hatchback,Flex Fuel	5657	7
22 Crossover,Luxury,Performance	1344.849558	113	58 Hatchback,Hybrid	2121.25	72
23 Crossover,Luxury,Performance,Hybrid	3916	2	59 Hatchback,Luxury	1379.5	46
24 Crossover,Performance	2585.956522	69	60 Hatchback,Luxury,Hybrid	454	3
25 Diesel	1730.904762	84	61 Hatchback,Luxury,Performance	1566.131579	38
26 Diesel,Luxury	2275	51	62 Hatchback,Performance	1039.646825	252
27 Exotic,Factory Tuner,High-Performance	1046.380952	21	63 High-Performance	1821.447236	199
28 Exotic,Factory Tuner,Luxury,High-Performance	517.5384615	52	64 Hybrid	2105.569106	123
29 Exotic,Factory Tuner,Luxury,Performance	520	3	65 Luxury	1102.65731	855
30 Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	520	13	66 Luxury,High-Performance	1668.017964	334
31 Exotic,Flex Fuel,Luxury,High-Performance	520	11	67 Luxury,High-Performance,Hybrid	568.833333	12
32 Exotic,High-Performance	1271.333333	261	68 Luxury,Hybrid	673.6346154	52
33 Exotic,Luxury	112.6666667	12	69 Luxury,Performance	1292.615156	673
34 Exotic,Luxury,High-Performance	467.0759494	79	70 Luxury,Performance,Hybrid	2333.181818	11
35 Exotic,Luxury,High-Performance,Hybrid	204	1	71 N/A	1676.889364	3742
36 Exotic,Luxury,Performance	217.027778	36	72 Performance	1348.873544	601
37 Exotic,Performance	1391	10	73 Performance,Hybrid	155	1
Grand Total			1554.911197	11914	



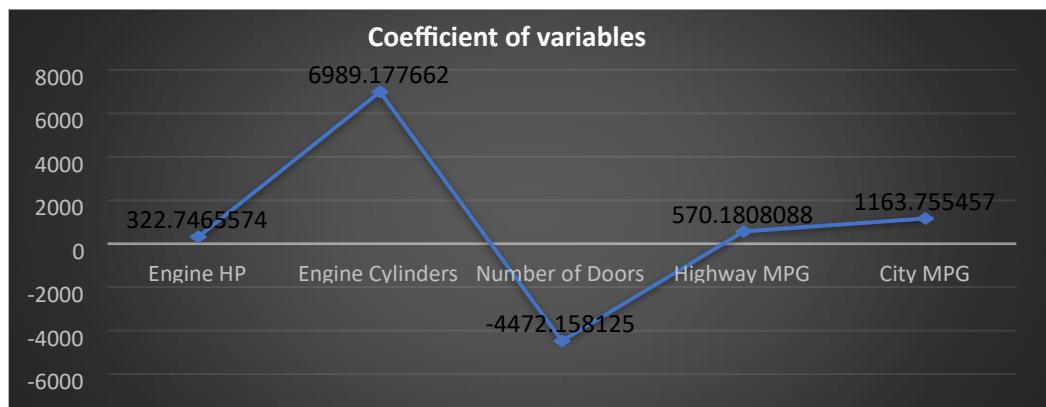
Insight Required: What is the relationship between a car's engine power and its price?

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



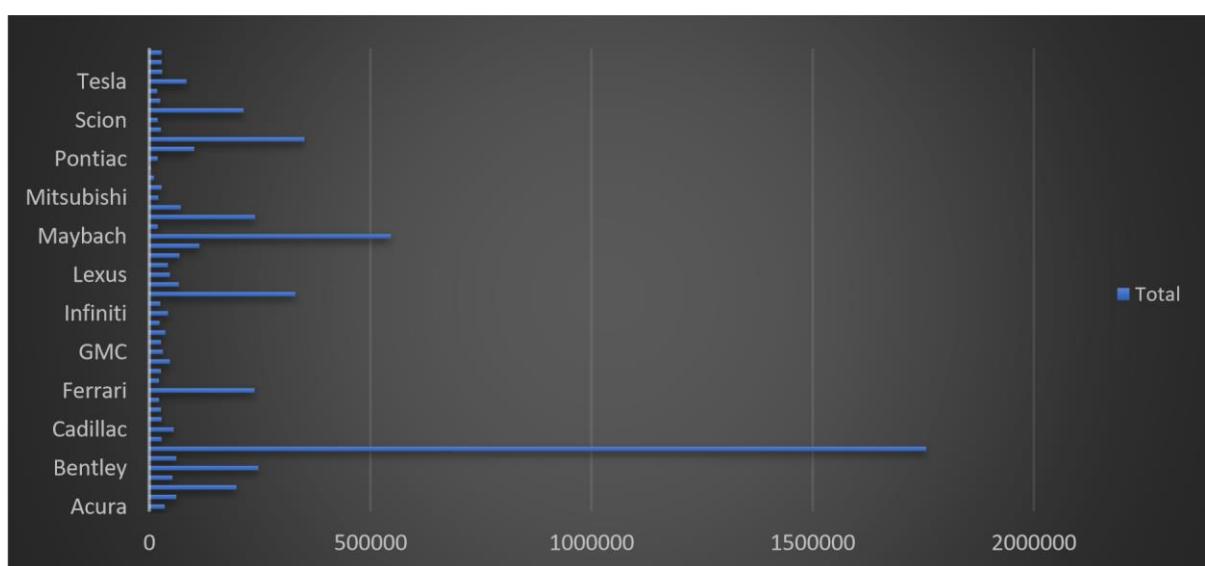
3. Which car features are most important in determining a car's price?

- **Task 3:** Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.



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

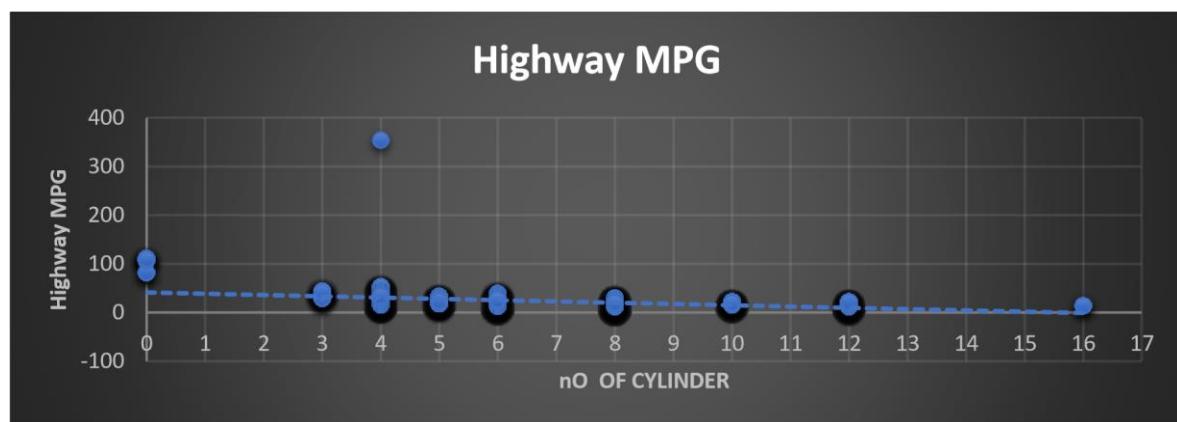
- **Task 4.A:** Create a pivot table that shows the average price of cars for each manufacturer.
- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



Outcome: Bugatti boasts the highest average price.

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

- **Task 5.A:** Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.
- **Task 5.B:** Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.



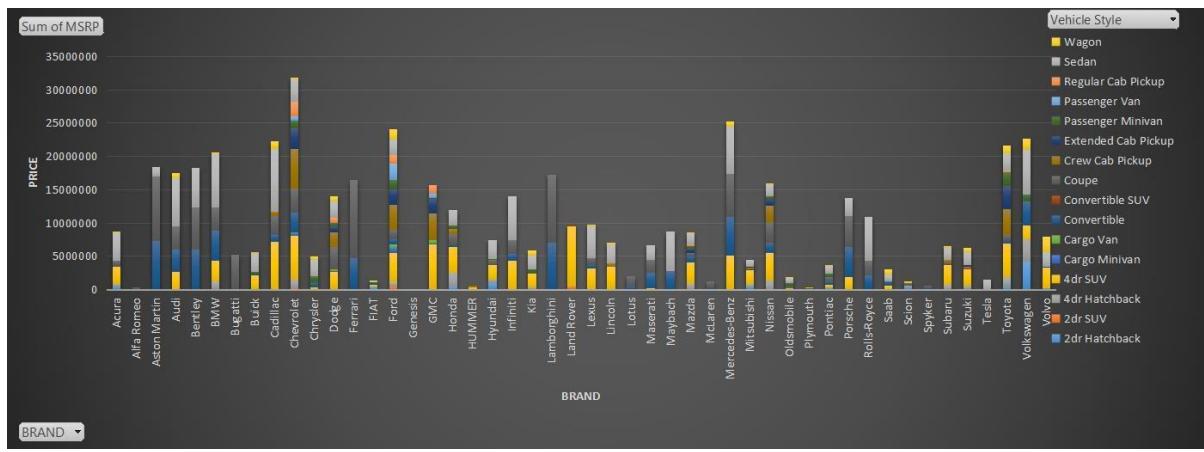
Correlation coefficient = -0.620312551

Outcome: The data demonstrates that as the number of cylinders increases, highway mpg decreases, indicating a negative correlation between the two variables.

BUILDING THE DASHBOARD:

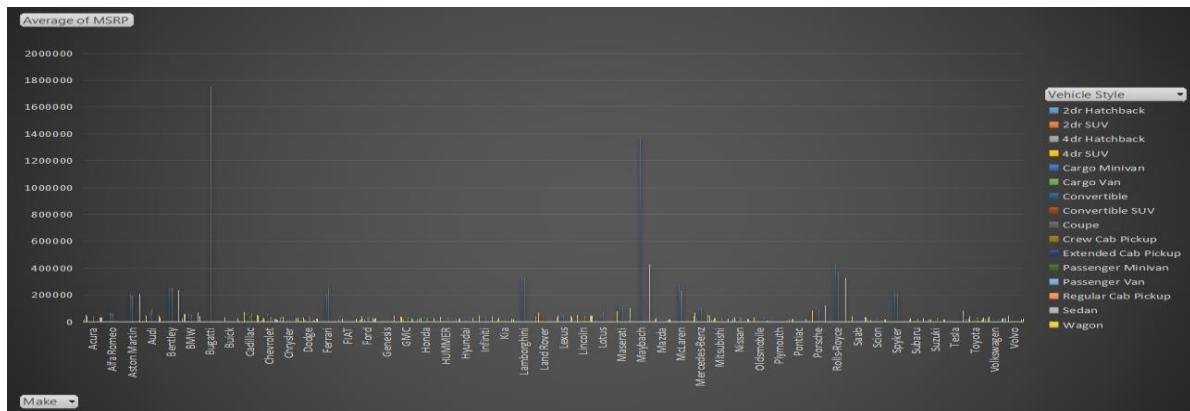
For the next stage of the project, the task is to develop an interactive dashboard using filters and slicers to enhance chart interactivity. The client has provided specific questions to address.

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



Outcome: Chevrolet has the highest price distribution

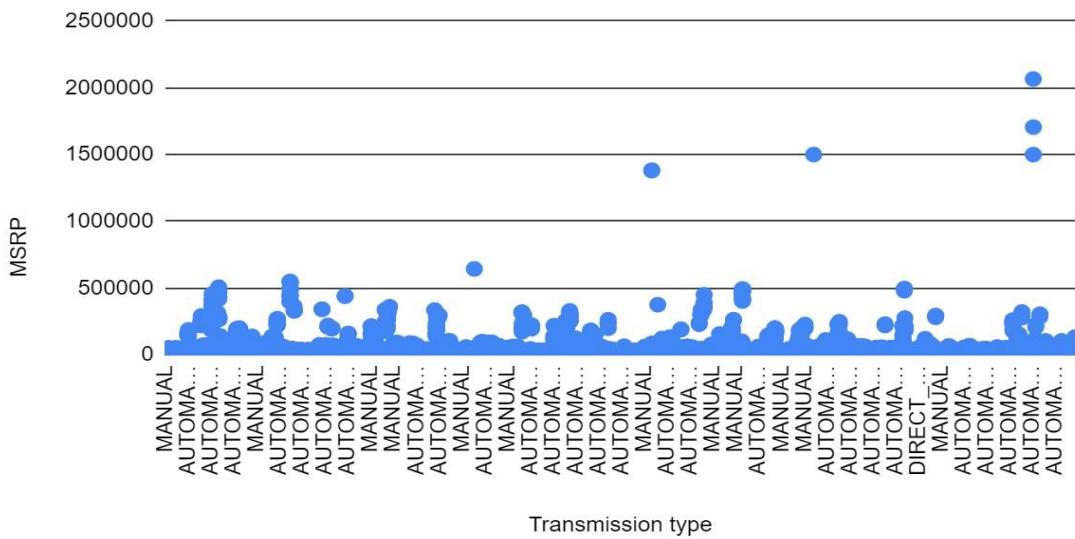
2. Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



Outcome: Bugatti has the highest MSRP and Plymouth has the lowest Average MSRP

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

Relationship between MSRP AND TRANSMISSION TYPE



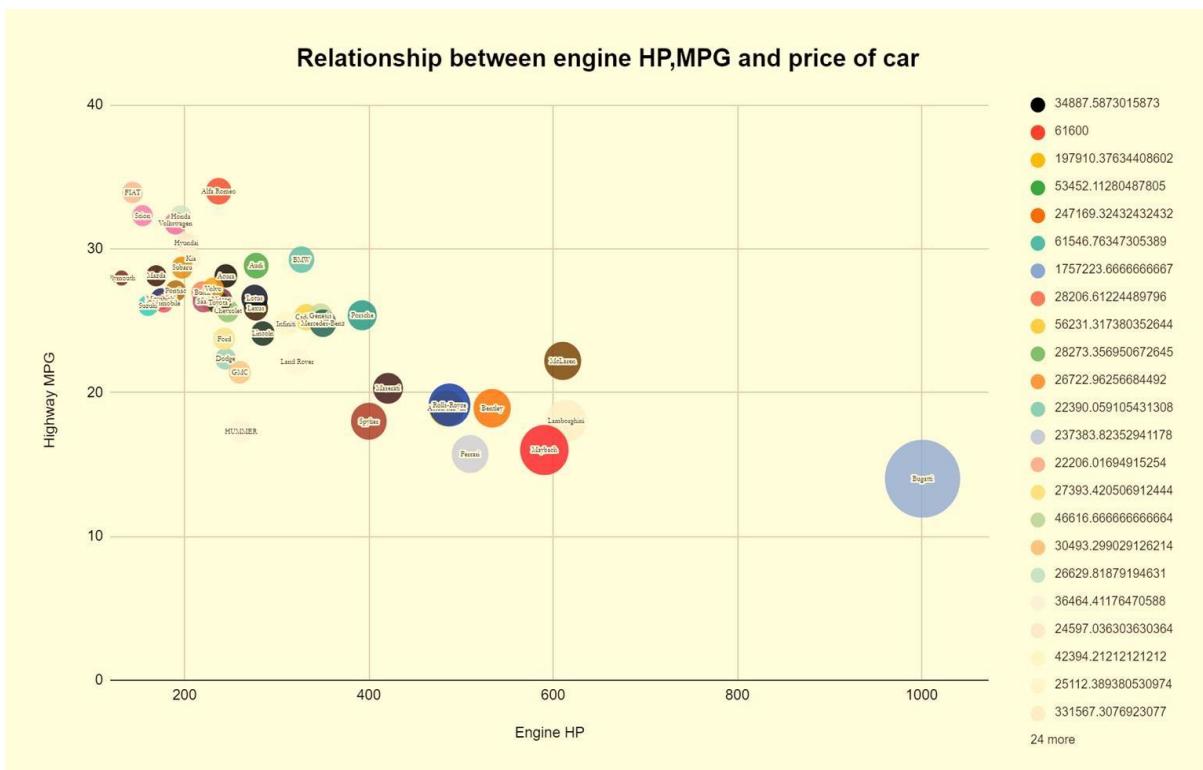
Outcome: The automatic_manual category is the priciest and the most popular as well.

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



Outcome: Over the year fuel efficiency is increasing at a slow speed.

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



Outcome: An increase in engine horsepower leads to a decrease in highway mpg and an increase in price.

CONCLUSION:

The analysis of the automotive dataset has provided valuable insights into various aspects of the industry, including consumer preferences, pricing strategies, and technological trends. Key findings include:

1. Consumer Preferences: Certain car features, such as engine power, transmission type, and body style, significantly influence consumer demand and pricing. Cars with higher engine power tend to command higher prices, indicating a preference for performance among consumers.

2. Pricing Strategies: Manufacturers can use insights from the analysis to develop effective pricing strategies. Understanding the relationship between car features and prices allows manufacturers to tailor their product offerings to meet consumer preferences while ensuring profitability.
3. Market Dynamics: Market segmentation plays a crucial role in understanding consumer behaviour. Categorizing cars based on market segments and analyzing their popularity and pricing dynamics helps manufacturers better target their products to specific consumer segments.
4. Technology Trends: Technological advancements, such as fuel efficiency and Eco friendliness, are increasingly important factors driving consumer demand and influencing pricing decisions.

Overall, the analysis provides a comprehensive understanding of the factors driving the automotive industry's evolution. Manufacturers can use these insights to make informed decisions regarding product development, marketing strategies, and pricing, ultimately enhancing their competitiveness in the market.

ABC Call Volume Trend Analysis

DESCRIPTION:

The dataset provided contains valuable insights into the inbound call volume of an ABC company's Customer Experience (CX) Inbound calling team, specifically focusing on the insurance category over a span of 23 days. This data encompasses crucial metrics such as Agent_Name, Agent_ID, Queue_Time (the duration customers wait before connecting to an agent), Time (the time of day when the call was made), Time_Bucket (a simplified time categorization), Duration (the length of the call), Call_Seconds (call duration in seconds), and call status (Abandon, answered, transferred). This analysis aims to uncover trends and patterns in call volume to enhance customer service and operational efficiency.

DESIGN:

Steps taken to clean the data:

First importing the datasets provided and merging them. Then removing the duplicates and the blank cells. Improving the headers of each column with proper values. Finding and replacing the data with correct values.

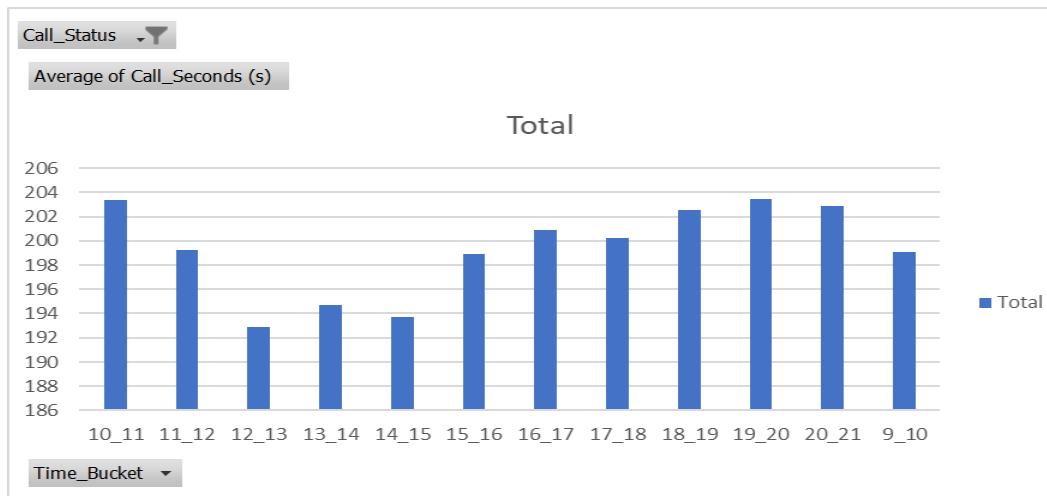
Tools used for visualization: Excel

Tech Stack Used

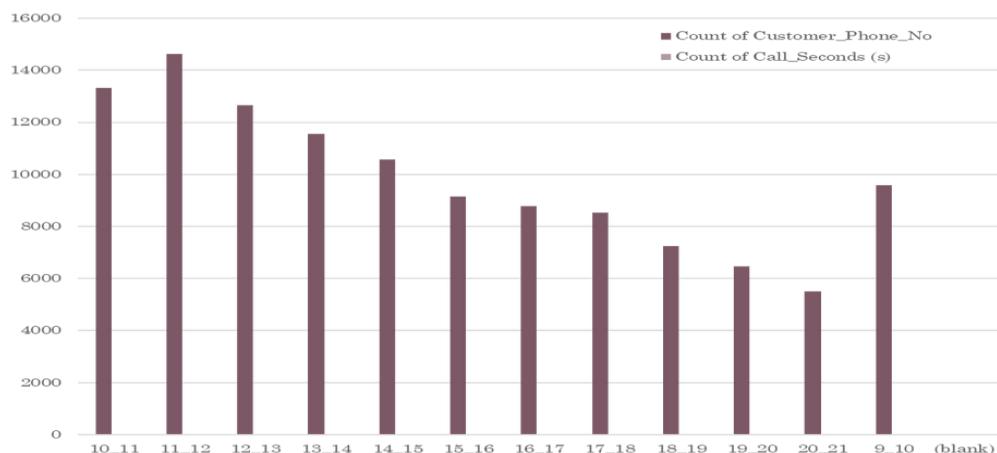
- Microsoft Excel

FINDINGS:

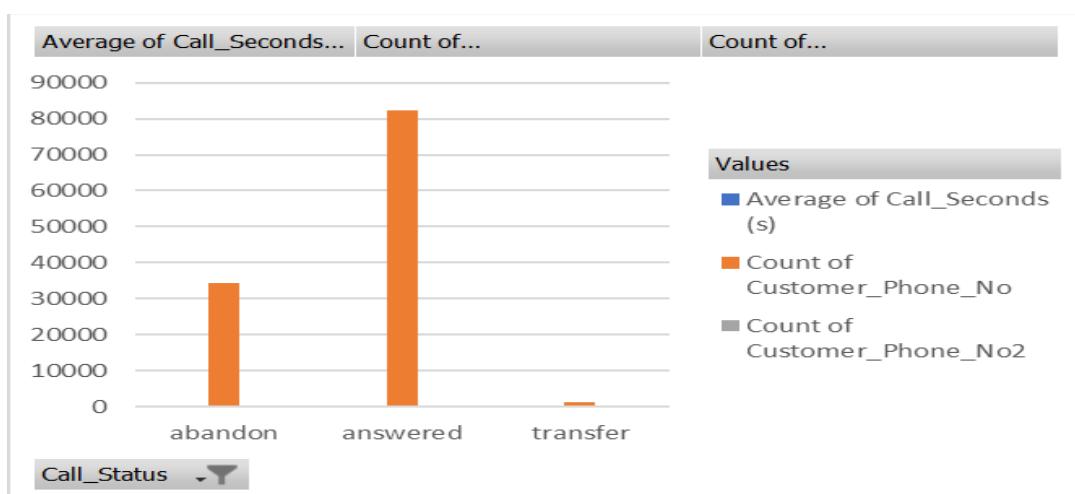
1. Average Call Duration:



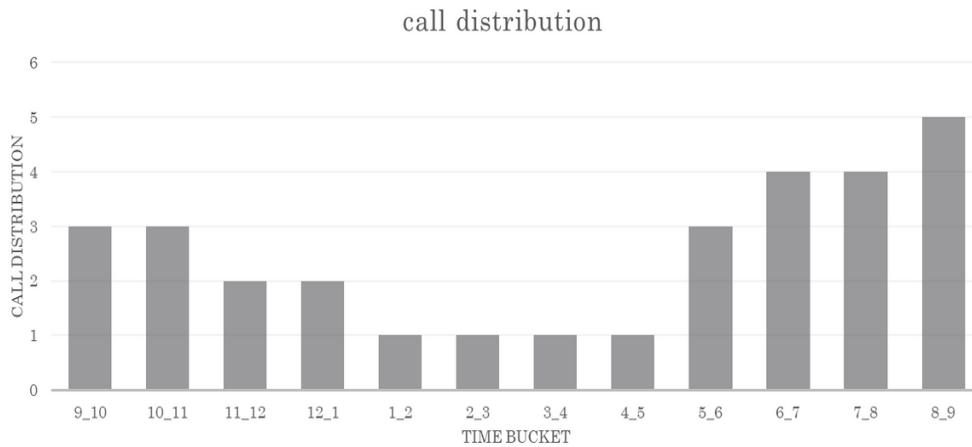
2. Call Volume Analysis:



3. Manpower Planning:



4. Night Shift Manpower Planning:



CONCLUSION:

- Throughout this project, I have gained valuable insights into the role of an analyst in the customer service department and its impact on ensuring maximum customer satisfaction. The company's commitment to effective customer handling strategies is evident.
- An important tool utilized is the Interactive Voice Response (IVR) system, which uses AI technology to address customer queries by identifying their specific concerns and routing calls to the appropriate agents for resolution.
- The analysis of the provided data was made more efficient by the pre-calculated time buckets and call durations converted into seconds, which saved time and effort in calculations.
- Additionally, I have explored behavioral analytics, which involves studying customer behavior patterns to identify trends, preferences, and opportunities for enhancing the overall customer experience.
- Overall, this project has provided me with valuable knowledge and insights into the dynamics of customer service and the significant role of an analyst in optimizing customer satisfaction.

APPENDIX

1. Instagram User Analytics Description:

https://drive.google.com/file/d/1bOYnbq_DpnWnNgyEq77eLQe71ucnMm/view?usp=sharing

2. Operation Analytics And Investigating Metric Spike:

<https://drive.google.com/file/d/1DETvXGbYJGw6z8UuslrZJRi0iMDiP6A4/view?usp=sharing>

3. Hiring process:

https://drive.google.com/file/d/11YlnzGZnHtO9ndrL_MHFjyAzqf3qH4cI/view?usp=sharing

4. IMDB Movie Analysis:

<https://drive.google.com/file/d/1fy3Yd5sbMsDqjJbz-uaxIWcYFLGekKIV/view?usp=sharing>

5. Analysing the impact of car features on price and profit:

<https://drive.google.com/file/d/1mLRUfggFdUXVRyXjCzUCKc7L2HwoNpkR/view?usp=sharing>

6. Bank loan case study:

<https://drive.google.com/file/d/1Zls1Q1OyLHh3zLnsOu53QlbkC4q8CuC8/view?usp=sharing>

7. ABC call volume trend analysis:

https://drive.google.com/file/d/1_Bs6lzah8R-f0tmRuwZoV4MACmh-PWiI/view?usp=drive_link

