DATA ANALYSIS PORTFOLIO

Prepared by

DHRUV SINGH
ds075156@gmail.com

PROFESSIONAL BACKGROUND

I am a dedicated and detail-oriented data analytics professional with a strong foundation in tools and techniques essential for extracting insights from data. My academic background in computer science has provided me with a solid understanding of programming, problem-solving, and data structures.

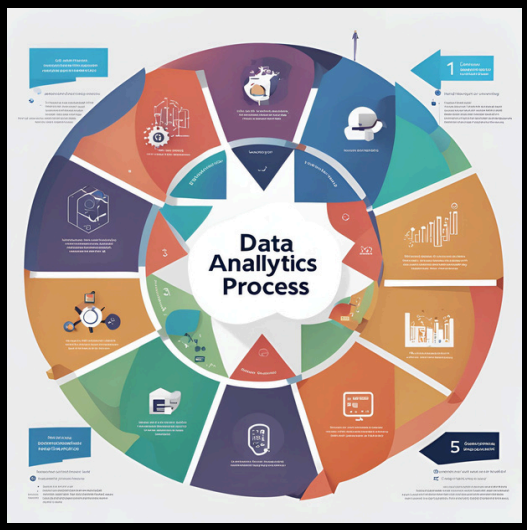
During my learning journey, I have mastered tools and technologies such as Excel, SQL, Python, Tableau, and Power BI, along with statistical methods to analyze and interpret data effectively. My experience includes working on a variety of projects, such as Bank Loan Case Studies, IMDB Movie Analysis, Hiring Process Analytics, Analyzing the Impact of features of car on Price and Profitability Analysis, and ABC Call Volume Trend Analysis, which have honed my abilities in data cleaning, exploratory data analysis (EDA), visualization, and generating actionable insights.

In addition to technical skills, I have cultivated a systematic approach to solving real-world business problems through the data analytics process: planning, preparing, processing, analyzing, sharing, and acting. My hands-on experience with large datasets, SQL queries, and dashboard creation ensures I can deliver meaningful and actionable outcomes for stakeholders.

I am passionate about leveraging data-driven insights to drive decision-making and create impactful solutions. My career aspirations include becoming a data scientist, where I can apply advanced analytics techniques, including machine learning and deep learning, to solve complex challenges and contribute to organizational success.

TABLE OF CONTENTS

DATA ANALYTICS PROCESS



The data analytics process underpins all my projects. It involves six critical stages:

1. Plan: Define objectives, criteria, and timelines for achieving goals.
2. Prepare: Identify and collect relevant data, and choose tools for data management.
3. Process: Clean, organize, and structure data for analysis.
4. Analyze: Explore data to uncover trends, correlations, and actionable insights.
5. Share: Present findings through reports and visualizations.
6. Act: Implement recommendations and monitor outcomes to refine strategies.

trainity

Data Analytics Process: Real World Application

Module 1 : Phone Upgrade Savings Analysis

Shopping & Use of 6 Step Data Analytics Process

Objective:
- To save for a phone upgrade with a budget of ₹40,000 to ₹50,000 within 3-5 months.

Plan:
- Set a target budget and timeline.

Prepare:
- Gather data from:
  1.) Bank statements.
  2.) Expense receipts.
- Use tools like:
  1.) Text editor to track goals.
  2.) Spreadsheet software to monitor expenses and savings.

Process:
- Categorize expenses into utility, food, and petrol.
- Compute monthly savings and expenses.

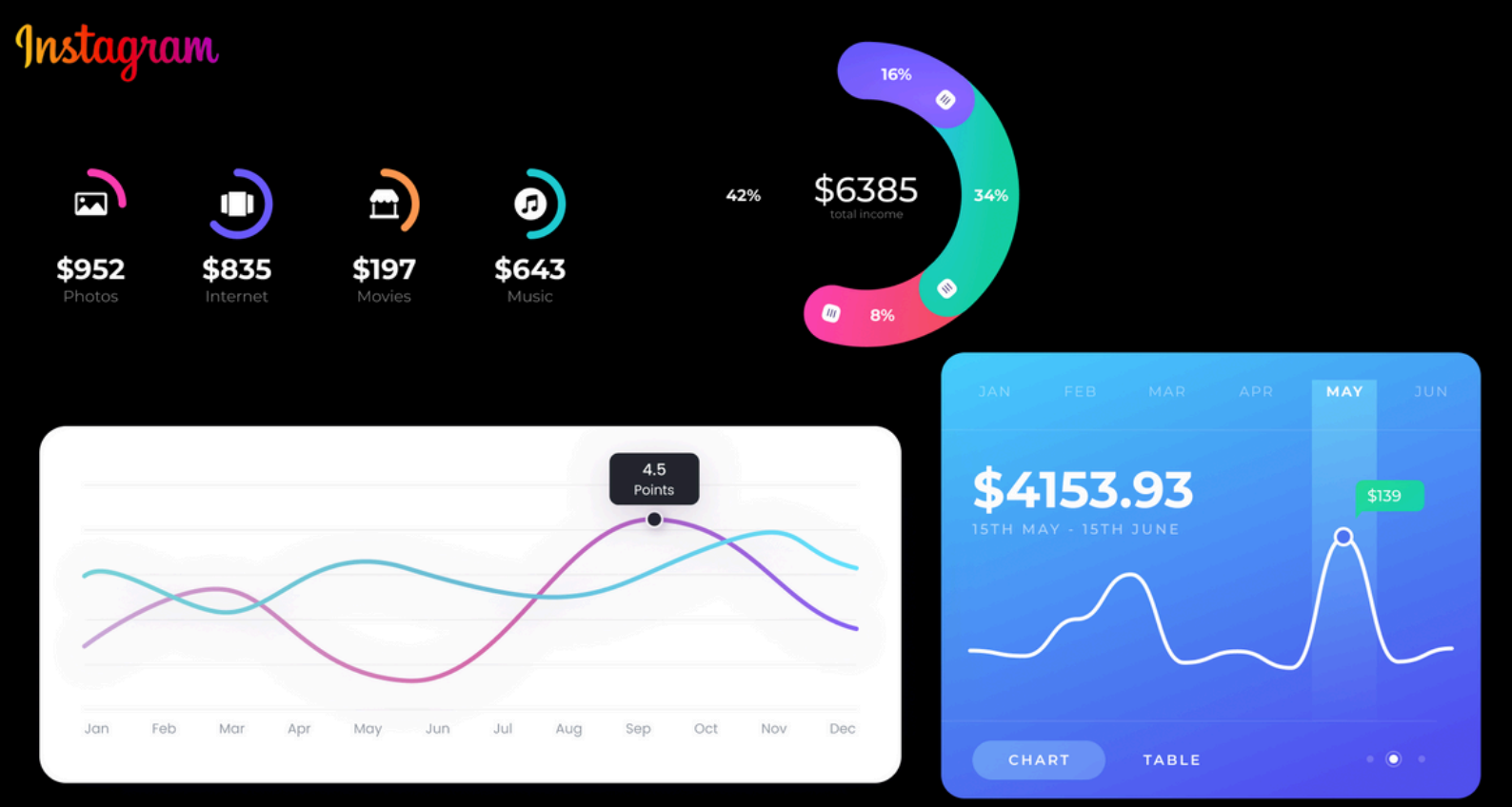
Analyze:
- Compare savings and expenses to identify spending patterns.
- Determine overspending categories and control measures.
- Highlight unusual expenses for review.

Share:
- Create a report summarizing monthly savings and expenses.
- Discuss insights with a guardian/parent for feedback.

Act:
- Implement budgeting strategies.
- Evaluate progress and adjust goals as necessary.
- Research and decide on the best phone model based on needs and budget.

Key Tools and Skills Used:
- Data organization in spreadsheets.
- Expense tracking and categorization.
- Basic financial analysis and reporting.

Outcome:
- Successfully saved the targeted amount within the desired timeline.
- Selected a new phone model meeting requirements and budget.



Module 2 : Instagram User Analytics

Objective:

- To analyze user interactions and engagement on Instagram to provide insights for the product, marketing, and investor teams.

Plan:

- Define key metrics: user activity, engagement rates, and interaction patterns.
- Set a timeline for data collection and analysis.

Prepare:

- Data Sources:
    1.) Instagram user logs.
    2.) Interaction data (likes, comments, shares, follows).

Tools Used:

- SQL for querying the dataset.
- Excel for initial data cleaning.

Analysis and SQL Queries

A) Marketing Analysis

1.) Loyal User Reward : Identify the five longest-standing users on Instagram.

Findings :

```
1 •   select * from users
2     order by created_at
3     Limit 5;
```

Sql Query

id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26
NULL	NULL	NULL

Result

2.) Inactive User Engagement : Identify users who have not posted a single photo.

Findings :

```
1 •   select username from users
2     WHERE id NOT IN (SELECT DISTINCT user_id FROM photos);
```

Sql Query

Result

username
Aniya_Hackett
Kasandra_Homenick
Jaclyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39
Franco_Keebler64
Nia_Haag
Hulda.Macejkovic
Leslie67
Janelle.Nikolaus81
Darby_Herzog
Esther.Zulauf61
Bartholome.Bernhard
Jessyca_West
Esmeralda.Mraz57
Bethany20

3.) Contest Winner Declaration : Determine the user with the most likes on a single photo.

Findings :

```sql
select username, photos.id,photos.image_url, count(likes.user_id) as Count
from photos
Inner Join likes
ON likes.photo_id=photos.id
Inner Join users
on photos.user_id = users.id
Group By photos.id
Order by Count Desc
Limit 1;
```

Sql Query

Result

	username	id	image_url	Count
▶	Zack_Kemmer93	145	https://jarret.name	48

4.) Hashtag Research : Find the top five most commonly used hashtags on Instagram.

Findings :

Result

Sql Query

```sql
Select tags.tag_name, Count(*) as Count
from photo_tags
Join tags on photo_tags.tag_id=tags.id
Group by tags.id
Order by Count Desc
Limit 5;
```

	tag_name	Count
▶	smile	59
	beach	42
	party	39
	fun	38
	concert	24

5.) Ad Campaign Launch : Determine the best day of the week to launch ads based on user registration patterns.

Findings :

```sql
Select dayname(created_at) as Day , count(*) as Count
from users
Group by Day
Order by Count Desc
Limit 1;
```

Sql Query

Result

	Day	Count
▶	Thursday	16

B) Investor Analysis

1.) User Engagement : Calculate the average number of posts per user and the total number of photos divided by total users.

Findings :

Sql Query

```
Select ( Select count(*) from photos )/ (Select count(*) from users ) as Averag
```

Result

Average
2.5700

2.) Bots & Fake Accounts : Identify users who liked every photo, which might indicate bot-like behavior.

Findings :

Sql Query

```
1  Select user_id, Count(*) as Count
2  from likes
3  Group By user_id
4  Having Count = (Select Count(*) from photos);
5
6  Select u.username,count(*) as Count
7  from users u
8  Join likes l on u.id=l.user_id
9  Group by u.id
10 Having Count = (Select count(*) from photos);
```

Result

user_id	Count
5	257
14	257
21	257
24	257
36	257
41	257
54	257
57	257
66	257
71	257
75	257
76	257
91	257

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

Insights

- **Loyal Users:** Identified long-standing users who could be engaged for special loyalty campaigns.
- **Inactive Users:** Recognized users who could be targeted with re-engagement strategies.
- **Popular Hashtags:** Top-performing hashtags were found, guiding brand collaborations and content reach.
- **Optimal Ad Day:** Identified a peak registration day to guide ad scheduling, which could maximize new user engagement.
- **User Engagement:** Calculated an engagement metric useful for understanding overall user interaction.
- **Bots & Fake Accounts:** Detected potential bot accounts, helping to maintain Instagram's authenticity and reliability.

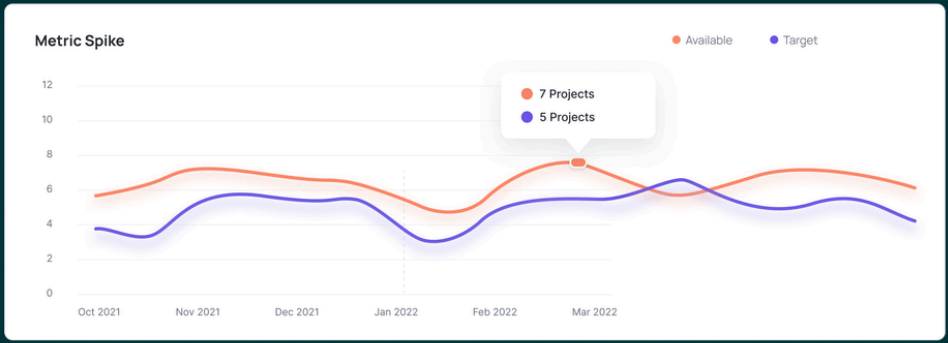
Results and Impact

Through SQL analysis, we provided actionable insights to improve user retention, engagement, and growth. By identifying key user segments and behavioral patterns, this project aids Instagram's teams in enhancing platform value and optimizing user experiences.

Module 3

trainity

Operation Analytics & Investagating metric spike case study



Objective:

- To analyze operational data to identify and investigate metric spikes, providing actionable insights for decision-making.

Plan:

- Define investigation metrics: spike patterns, operational inefficiencies.
- Establish a timeline and goals for analysis and reporting.

Prepare:

- Data Sources:
  1.) Operational logs.
  2.) Historical performance data.
  3.) Relevant metadata files.

Tools Used:

- SQL Workbench for querying.
- Excel for data preparation and cleaning.

Process:

- Import large datasets using SQL commands (e.g., LOAD DATA INFILE).
- Clean data by removing outliers and inconsistencies.
- Segment data into relevant categories (e.g., jobs, metrics, timestamps).
- Query data using SQL to identify trends and anomalies.

Case Studies

Case Study 1: Job Data Analysis

A) Jobs Reviewed Over Time:

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

QUERY:

```
#Calculate the number of jobs reviewed per hour for each day in November 2020.
Select ds as Date,
    Count(job_id) As Jobid,
    Round((Sum(time_spent) / 3600 ), 2) As Total_Time_Spent,
    Round(( Count(job_id) / (Sum(time_spent) / 3600 )),2) As Job_Review
From job_data
Where
    ds Between '2020-11-01' And '2020-11-30'
Group By ds
Order By ds;
```

Solution :

	Date	Jobid	Total_Time_Spent	Job_Review
▶	11/25/2020	1	0.01	80.00
	11/26/2020	1	0.02	64.29
	11/27/2020	1	0.03	34.62
	11/28/2020	2	0.01	218.18
	11/29/2020	1	0.01	180.00
	11/30/2020	2	0.01	180.00

B) **Throughput Analysis:**

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

Query : For Weekly Average

```
Select
    Round(Count(event) / Sum(time_spent), 2) As Weekly_Average
From
    job_data;
```

For Daily Average

```
Select ds As Dates,
Round(Count(event) / Sum(time_spent), 2) AS Daily_Average
From
    job_data
Group By ds
Order By ds;
```

Solution:

	Dates	Daily_Average
▶	11/25/2020	0.02
	11/26/2020	0.02
	11/27/2020	0.01
	11/28/2020	0.06
	11/29/2020	0.05
	11/30/2020	0.05

	Weekly_Average
▶	0.03

C) **Language Share Analysis:**

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

Query :

```
Select
    language,
    Round(100 * Count(*) / total, 2) As Percentage,
    jd.total
From
    job_data
        Cross Join
        (Select
            Count(*) As total
        From
            job_data) As jd
Group by language , jd.total;
```

Solution:

	language	Percentage	total
▶	English	12.50	8
	Arabic	12.50	8
	Persian	37.50	8
	Hindi	12.50	8
	French	12.50	8
	Italian	12.50	8

D) Duplicate Rows Detection:

- Objective: Identify duplicate rows in the data.

QUERY :

```
#Write an SQL query to displ
Select
    actor_id,
    Count(*) As Duplicates
From
    job_data
Group By
    actor_id
Having
    Count(*) > 1;
```

-

Solution :

	actor_id	Duplicates
▶	1003	2

Case Study 2: Investigating Metric Spike

A) Weekly User Engagement:

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

Query :

```
Select
        Extract(Week From occured_at) As week_num,
        Count(Distinct user_id) As active_users
From
    events
Where
    event_type = 'engagement'
Group By week_num
order by week_num;
```

Solution :

week_num	active_users
17	663
18	1068
19	1113
20	1154
21	1121
22	1186
23	1232
24	1275
25	1264
26	1302
27	1372
28	1365
29	1376
30	1467
31	1299
32	1225
33	1225
34	1204
35	104

B) User Growth Analysis:

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

Query :

```
With weekly_active_users AS(
    Select
        Extract(Year from created_at) As year,
        Extract(Week from created_at) As week_number,
        Count(Distinct user_id) As num_of_users
    From users
    Group By year,week_number)

    Select
        year,
        week_number,
        num_of_users,
        Sum(num_of_users) Over (Order By year,Week_number) As cumulative_users
    From weekly_active_users
    Order By year,week_number;
```

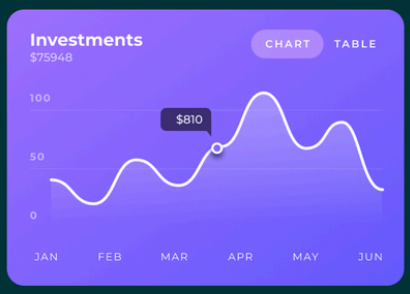
Solution :

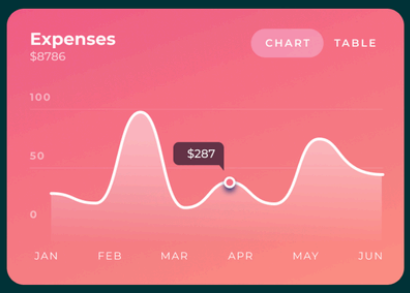
year	week_number	num_of_users	cumulative_users
2013	0	23	23
2013	1	30	53
2013	2	48	101
2013	3	36	137
2013	4	30	167
2013	5	48	215
2013	6	38	253
2013	7	42	295
2013	8	34	329
2013	9	43	372
2013	10	32	404
2013	11	31	435
2013	12	33	468
2013	13	39	507
2013	14	35	542
2013	15	43	585
2013	16	46	631
2013	17	49	680
2013	18	44	724
2013	19	57	781
2013	20	39	820

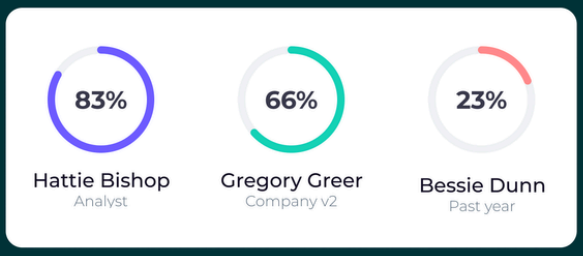
Module 4 : Hiring Process Analytics

trainity

Company Statistics







Objective:

- To analyze a company's hiring process, identify inefficiencies, and provide actionable insights for improvement.

Plan:

- Define the scope of analysis: focus on metrics such as application-to-hire ratio, time-to-hire, and candidate drop-off rates.
- Establish objectives for optimizing the hiring process.

Prepare:

- Data Sources:
    1.) HR records.
    2.) Recruitment platform data.
    3.) Interview feedback.

Tools Used:

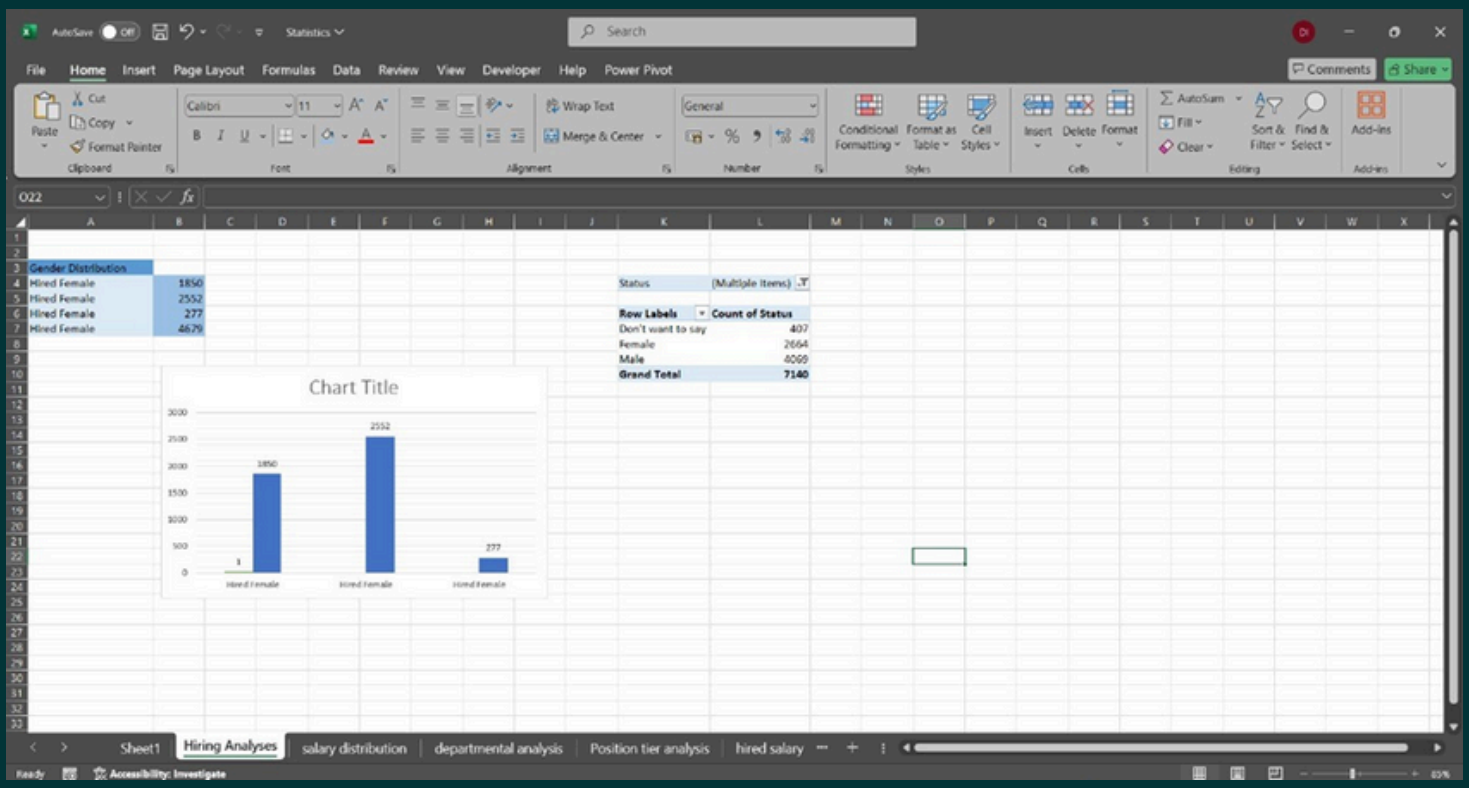
- Excel for data preparation and cleaning.
- Statistical analysis for identifying trends.

Process:

- Clean and preprocess the data by removing duplicates and irrelevant entries.
- Segment data based on job roles, application stages, and timelines.
- Calculate key metrics such as:
    1.) Average time to hire.
    2.) Candidate conversion rates.
    3.) Drop-off rates at various stages.

A. Gender Distribution in Hiring.

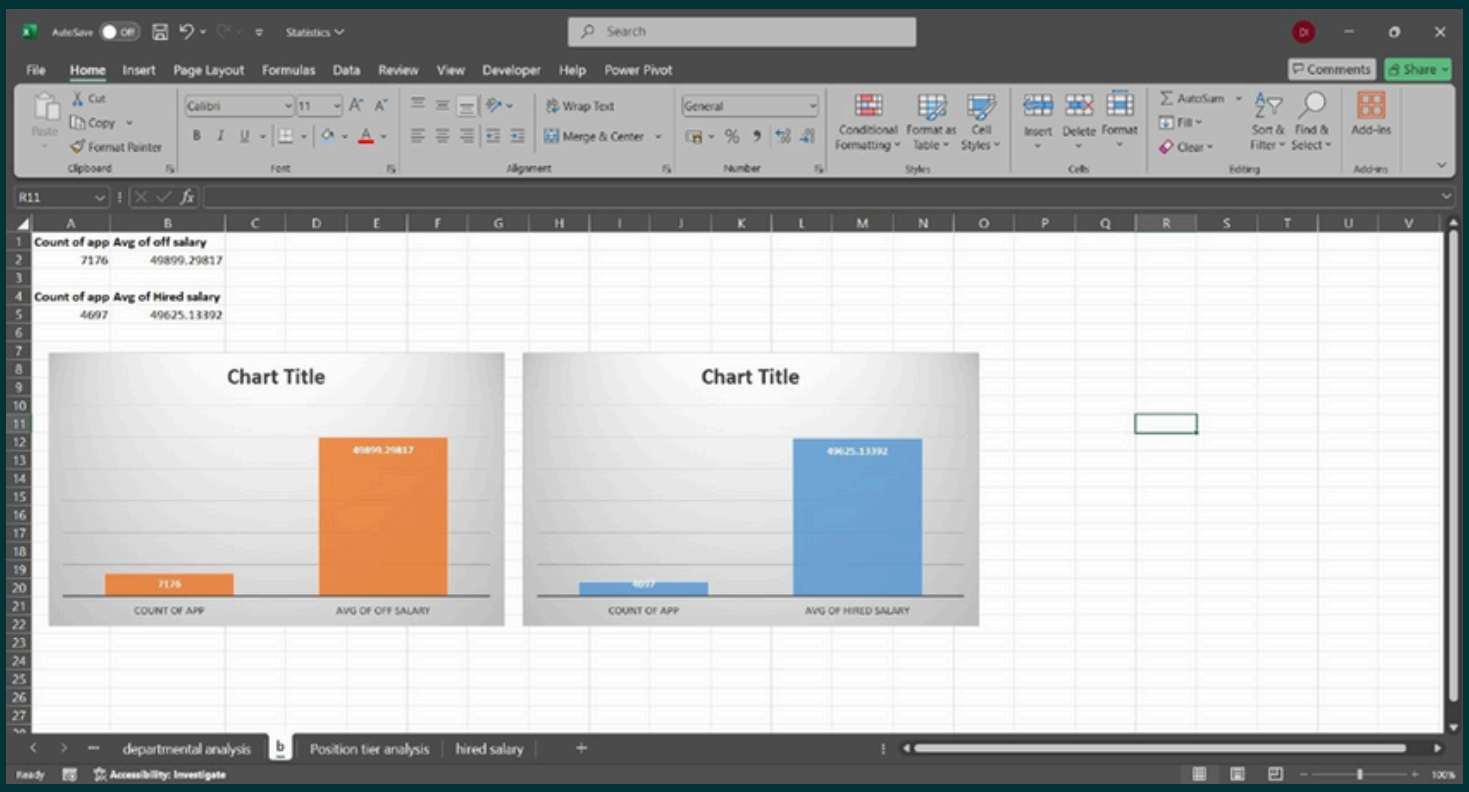
- Analysis: Determined the count of male and female hires using Excel's COUNTIF function



- Insight: The hiring process showed a balanced/unbalanced gender ratio (e.g., 60% male, 40% female).

B. Salary Analysis

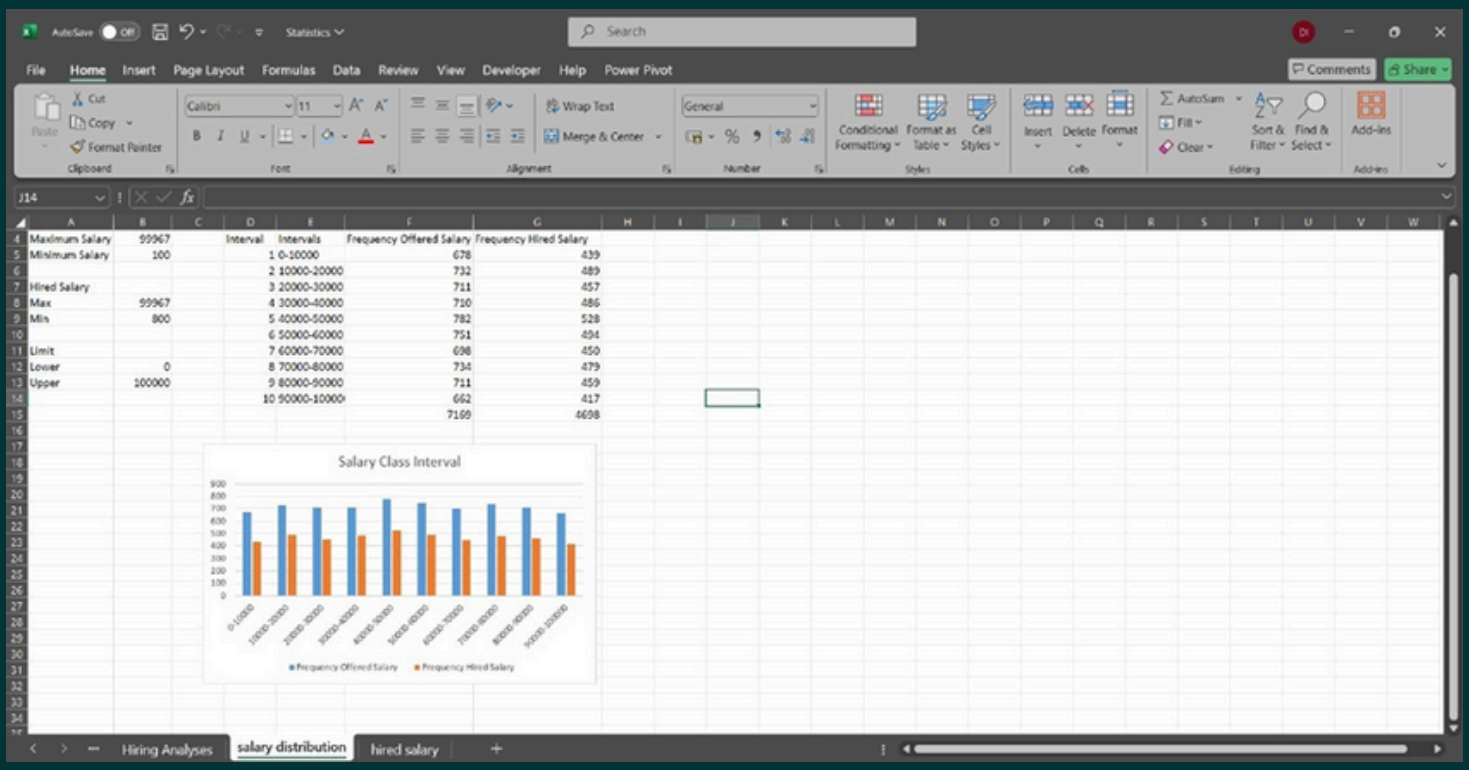
- Average Salary: The mean salary offered by the company was calculated using Excel's AVERAGE function.



- Insight: The average salary was $49899.29817. High-tier positions contributed significantly to the average.

C. Salary Distribution

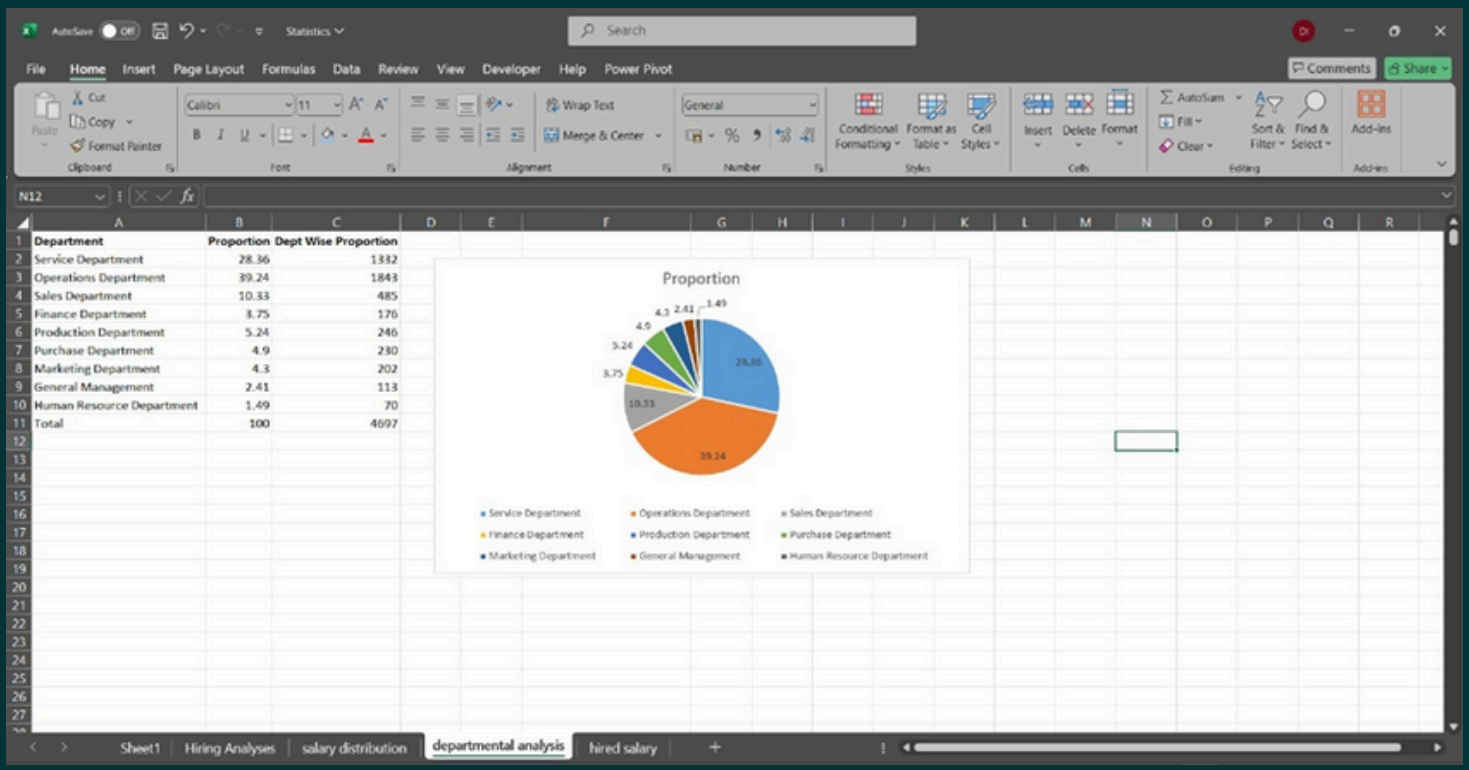
- **Analysis: Salary data was grouped into intervals (e.g., $20,000–$30,000, $30,001–$40,000).**



- **Insight: Most salaries fell within the $400 - $500 range, highlighting the company's pay structure trends.**

D. Departmental Analysis

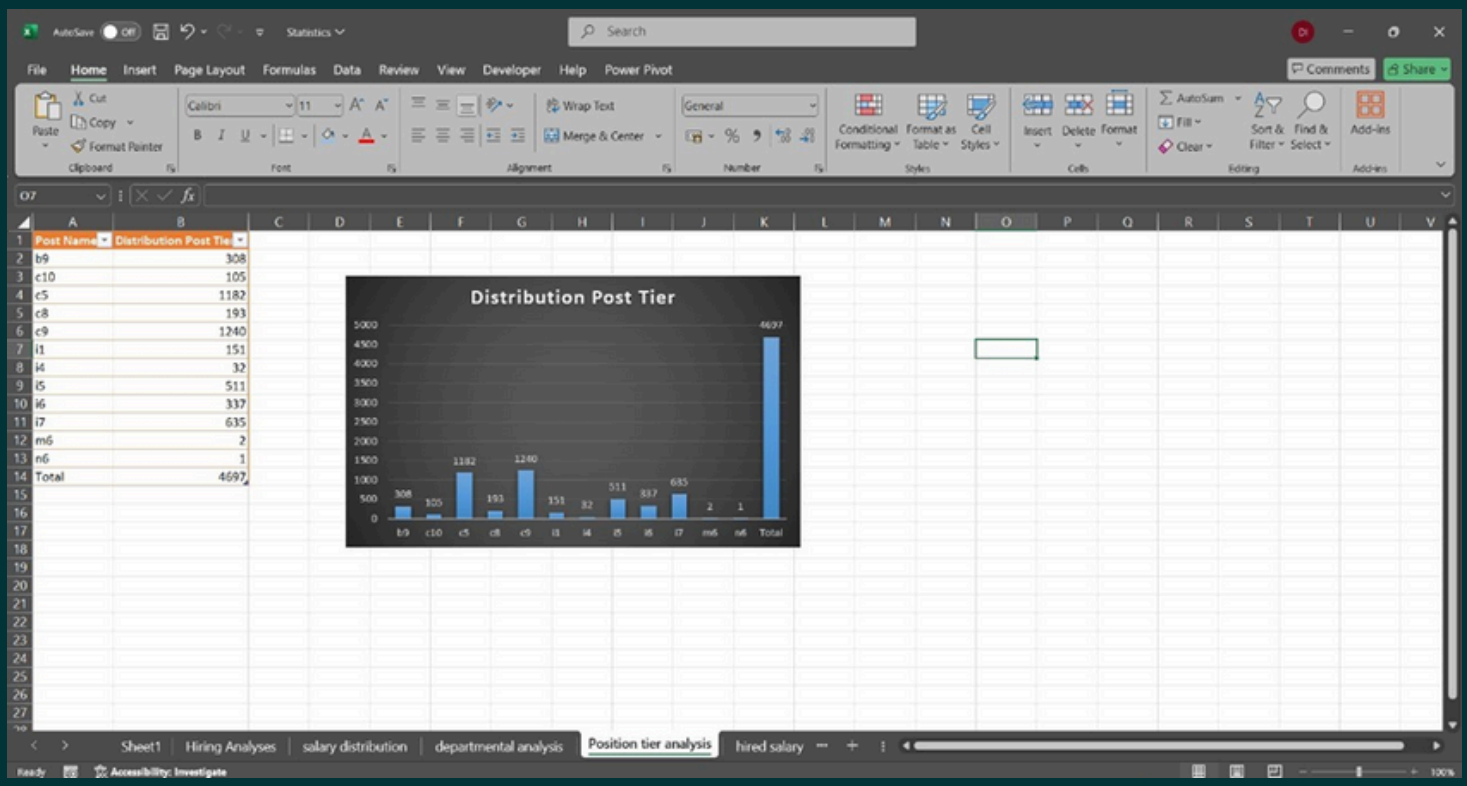
- **Visualization: A pie chart showed the proportion of hires across departments.**



- **Insight: Most hires were concentrated in the Operations/Service department, indicating higher demand in those areas.**

E. Position Tier Analysis

- **Visualization: A bar graph represented the distribution of hires across position tiers.**



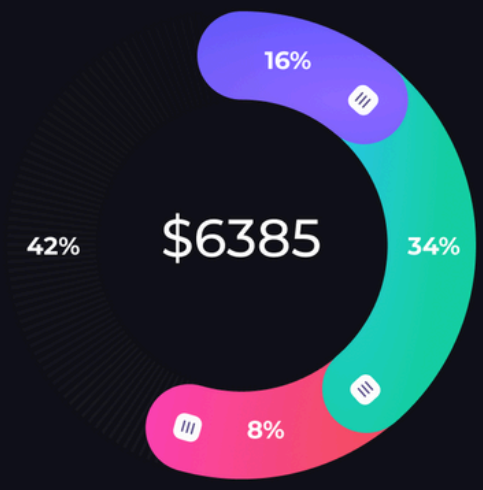
- **Insight: Entry-level positions had the highest hiring rates, followed by managerial roles.**

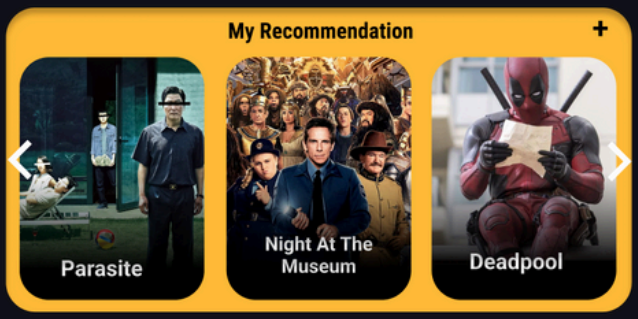
Result

The analysis provided actionable insights into the hiring trends:

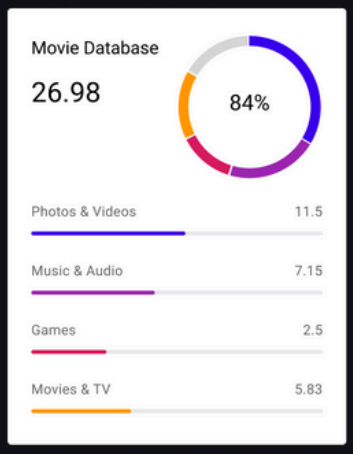
- Gender diversity needs improvement in Operations / Production departments.
- Salary distributions misalign with market standards.
- Departments like Operations/Service show hiring peaks, reflecting strategic growth.

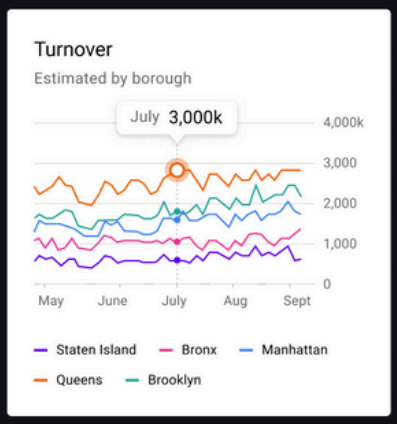
These insights can guide decision-making for improving hiring processes and workforce planning.





Module 5 : IMDB Movie Analysis





Objective:

- To analyze IMDB movie data to identify factors that influence a movie's success, such as genre, duration, language, director, and budget.

Plan:

- Define metrics for analysis: average ratings, box office collections, and trends in genres or languages.
- Establish a timeline for data preparation, analysis, and reporting.

Prepare:

- Data Sources:
  1.) IMDB movie datasets.
  2.) Supplementary datasets (e.g., box office revenues).

Tools Used:

- Excel for data cleaning and analysis.
- Statistical methods for deeper insights.

Process:

- Clean data by handling missing or inconsistent entries.
- Segment movies based on genres, directors, and languages.
- Compute averages and distributions for key metrics.

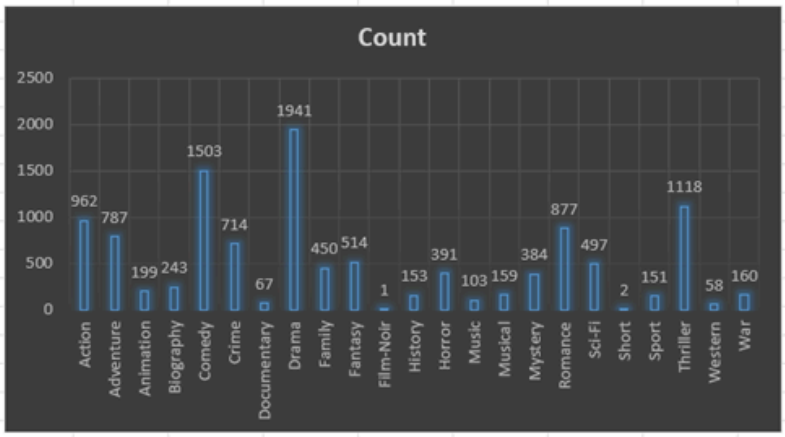
Data Analytics Tasks and Insights

A. Movie Genre Analysis

- **Task**: Analyze the distribution of movie genres and their impact on IMDB scores.

Category	Count
Action	962
Adventure	787
Animation	199
Biography	243
Comedy	1503
Crime	714
Document	67
Drama	1941
Family	450
Fantasy	514
Film-Noir	1
History	153
Horror	391
Music	103
Musical	159
Mystery	384
Romance	877
Sci-Fi	497
Short	2
Sport	151
Thriller	1118
Western	58
War	160
Total	11434

Genres Maniplutaion	
Median	384
Max	1941
Min	1
Mode	#N/A
Var	245400.5
Sd	495.3792



Process:

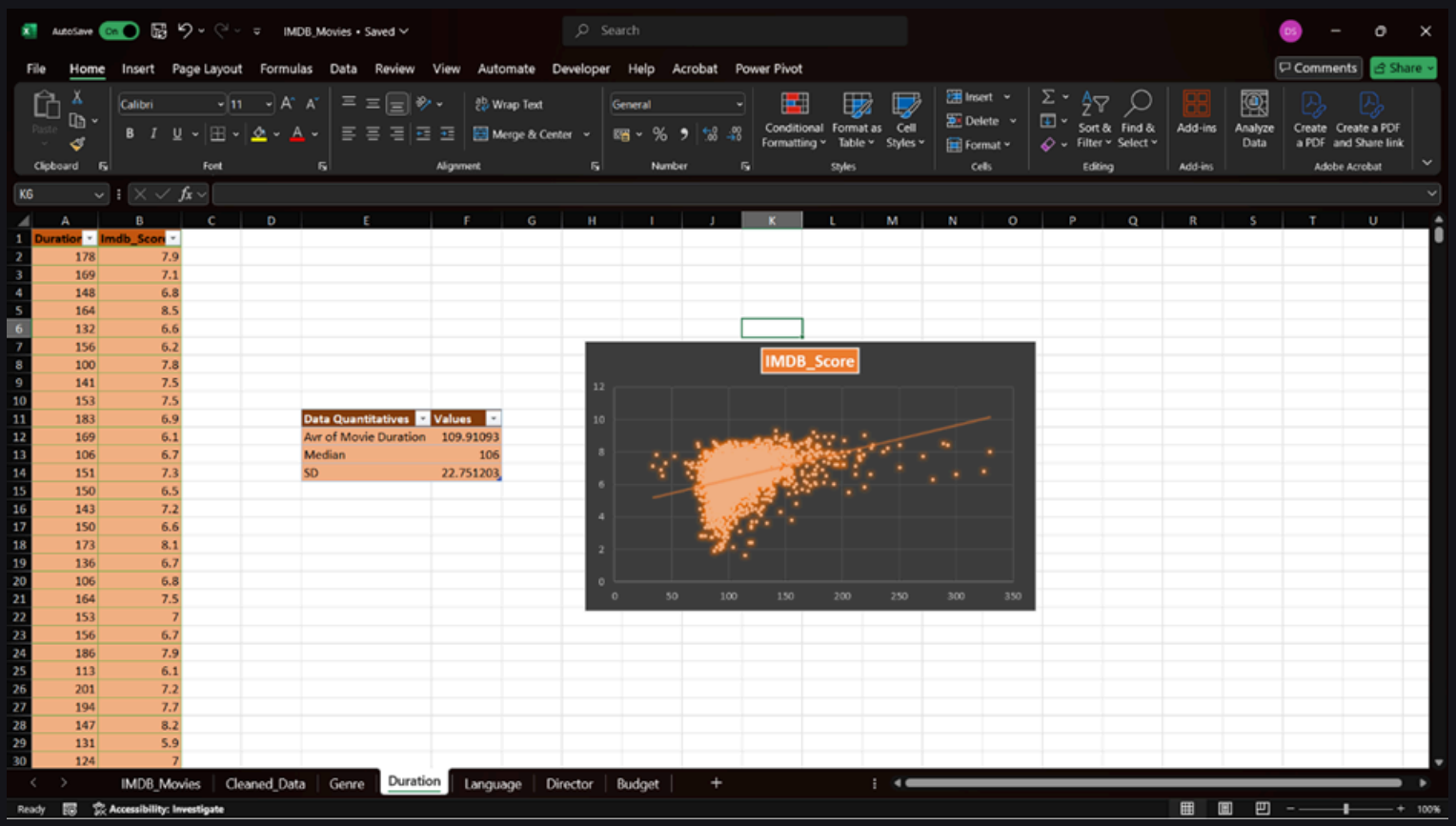
- Split genres using Excel formulas.
- Use COUNTIF to find the most common genres.
- Calculate mean, median, mode, range, variance, and standard deviation of IMDB scores for each genre.

Insights:

- Genre Drama has the highest Count in the movie count, suggesting it is well-received by the audience.
- Variance in scores highlights the consistency of viewer reception.

B. Movie Duration Analysis

- **Task**: Understand the relationship between movie duration and IMDB scores.



Process:

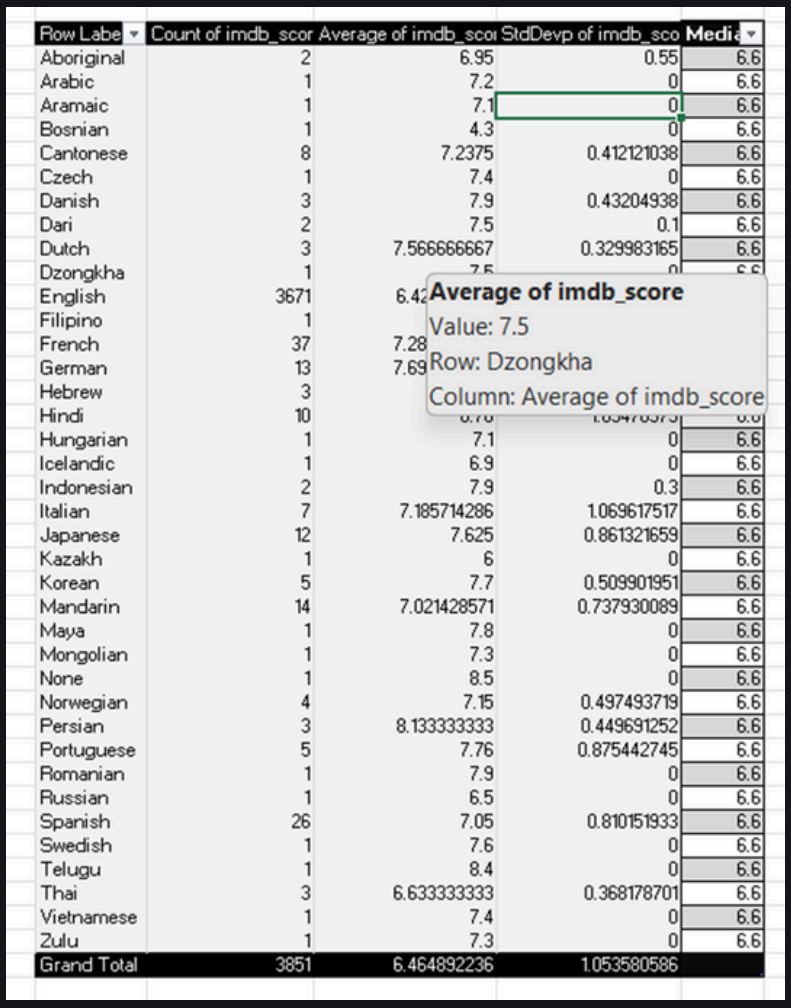
- Compute descriptive statistics for durations.
- Plot duration against IMDB scores and fit a trendline.

Insights:

- Movies with durations between 150 and 180 minutes tend to receive higher ratings.
- Extremely long movies are less favourably rated.

C. Language Analysis

- Task: Investigate the impact of movie language on IMDB ratings.



Row Labe	Count of imdb_scor	Average of imdb_sco	StdDevp of imdb_sco	Media
Aboriginal	2	6.95	0.55	6.6
Arabic	1	7.2	0	6.6
Aramaic	1	7.1	0	6.6
Bosnian	1	4.3	0	6.6
Cantonese	8	7.2375	0.412121038	6.6
Czech	1	7.4	0	6.6
Danish	3	7.9	0.43204938	6.6
Dari	2	7.5	0.1	6.6
Dutch	3	7.566666667	0.329983165	6.6
Dzongkha	1			
English	3671	6.42		
Filipino	1			
French	37	7.28		
German	13	7.69		
Hebrew	3			
Hindi	10			
Hungarian	1	7.1	0	6.6
Icelandic	1	6.9	0	6.6
Indonesian	2	7.9	0.3	6.6
Italian	7	7.185714286	1.069617517	6.6
Japanese	12	7.625	0.861321659	6.6
Kazakh	1	6	0	6.6
Korean	5	7.7	0.509901951	6.6
Mandarin	14	7.021428571	0.737930089	6.6
Maya	1	7.8	0	6.6
Mongolian	1	7.3	0	6.6
None	1	8.5	0	6.6
Norwegian	4	7.15	0.497493719	6.6
Persian	3	8.133333333	0.449691252	6.6
Portuguese	5	7.76	0.875442745	6.6
Romanian	1	7.9	0	6.6
Russian	1	6.5	0	6.6
Spanish	26	7.05	0.810151933	6.6
Swedish	1	7.6	0	6.6
Telugu	1	8.4	0	6.6
Thai	3	6.633333333	0.368178701	6.6
Vietnamese	1	7.4	0	6.6
Zulu	1	7.3	0	6.6
Grand Total	3851	6.464892236	1.053580586	

Process:

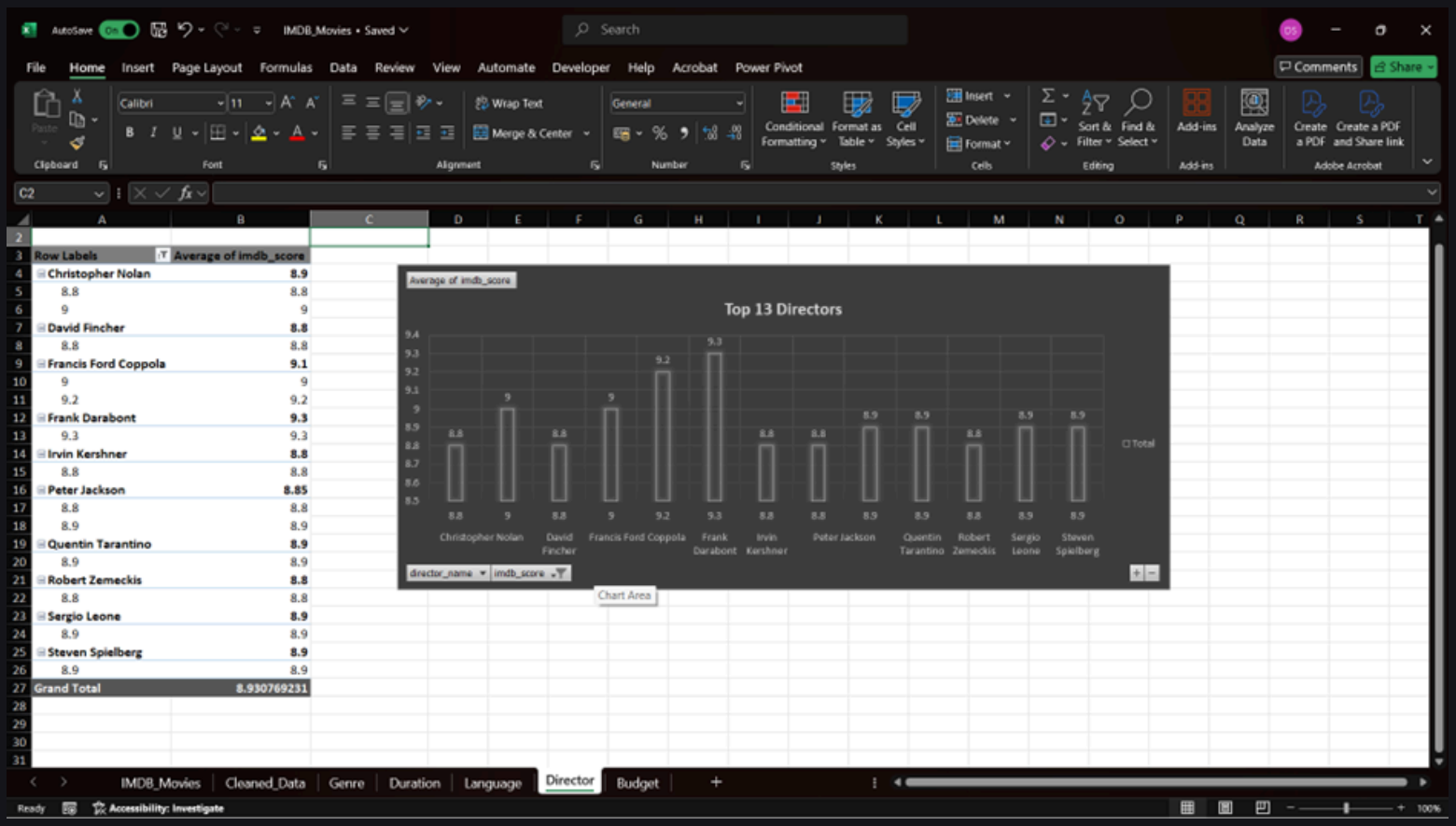
- Identify the most common languages using COUNTIF.
- Compute mean, median, and standard deviation of scores for each language.

Insights:

- Movies in Language English have the highest average scores, possibly due to larger audience reach.
- Lesser-known languages show higher variance, indicating niche audience reception.

D. Director Analysis

- Task:Evaluate the influence of directors on movie success.



Process:

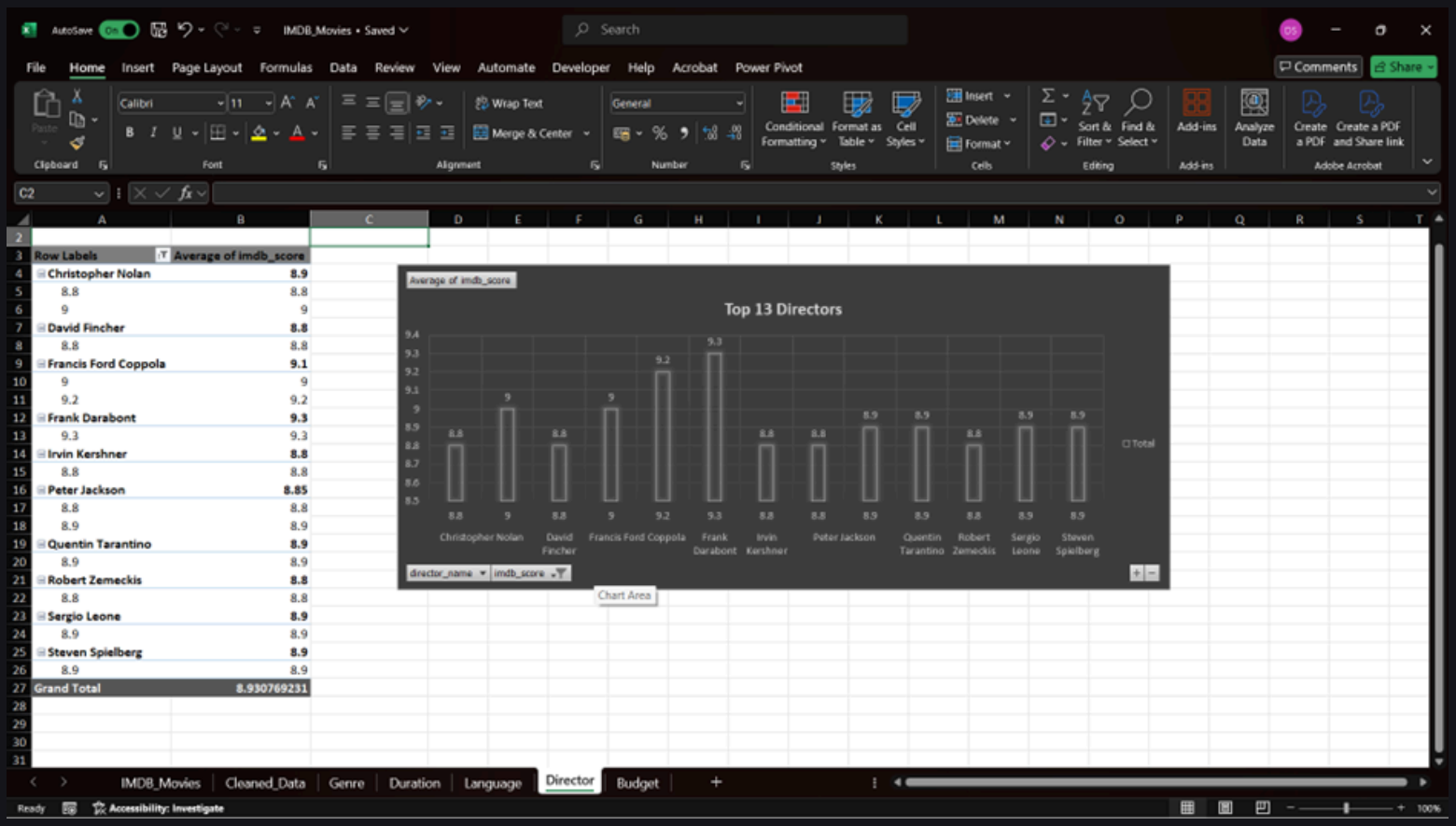
- Compute the average IMDB score for each director.
- Use PERCENTILE to identify top-performing directors.

Insights:

- Directors in the 90th percentile consistently produce highly rated movies.
- Patterns in their filmography reveal common success factors like genre or budget.

D. Director Analysis

- Task:Evaluate the influence of directors on movie success.



Process:

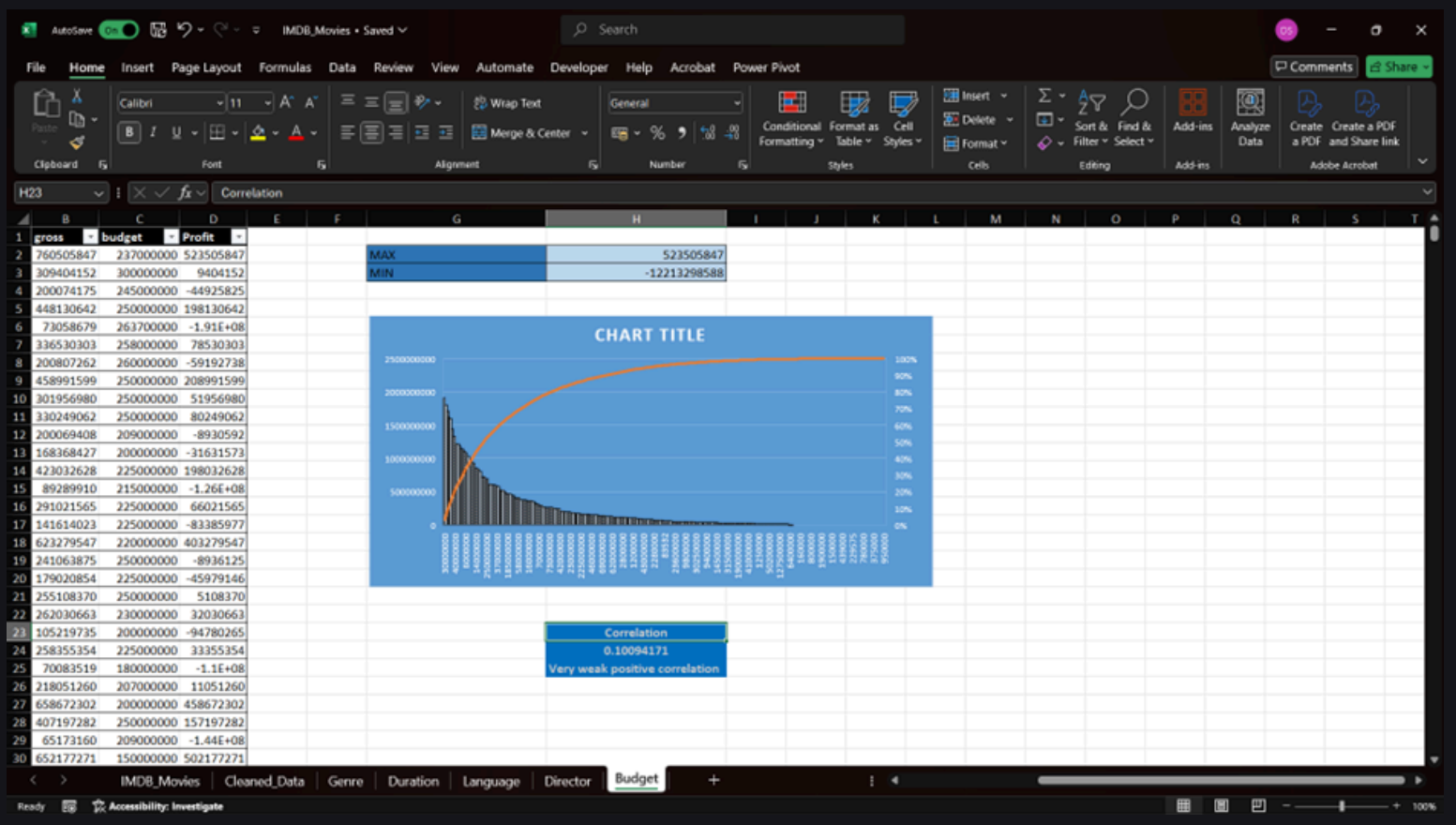
- Compute the average IMDB score for each director.
- Use PERCENTILE to identify top-performing directors.

Insights:

- Directors in the 90th percentile consistently produce highly rated movies.
- Patterns in their filmography reveal common success factors like genre or budget.

E. Budget Analysis

- Task: Explore correlations between budgets and financial success..



Process:

- Calculate correlation between budgets and gross earnings.
- Compute profit margins and identify top-performing movies.

Insights:

- High correlation suggests a strong link between budget and earnings.

Results and Insights

- Factors like genres, durations, and languages are critical in shaping audience reception.
- Directors and budgets significantly influence both ratings and profitability.
- These insights offer a data-driven roadmap for producers and investors to optimize future projects

MODULE 6 : Bank Loan Case Study

Objective:

- To analyze patterns in loan application data and identify key factors influencing loan defaults, aiding better decision-making in loan approvals.

Plan:

- Define the scope of analysis, focusing on:
- Approval rate trends.
- Loan default patterns.
- Demographic and financial characteristics affecting loan outcomes.

Prepare:

- Data Sources:
  1.) Loan application records.
  2.) Applicant demographic and financial data.

Tools Used:

- Excel for data cleaning and exploratory analysis.
- Statistical concepts for insight generation.

Process:

- Data Cleaning and Preparation:
  1.) Handle missing data by appropriate imputation methods.
  2.) Remove irrelevant or duplicate records.

Data Organization:

- Segment data by income, credit score, loan amount, and demographics.
- Identify outliers and assess their impact on loan outcomes.

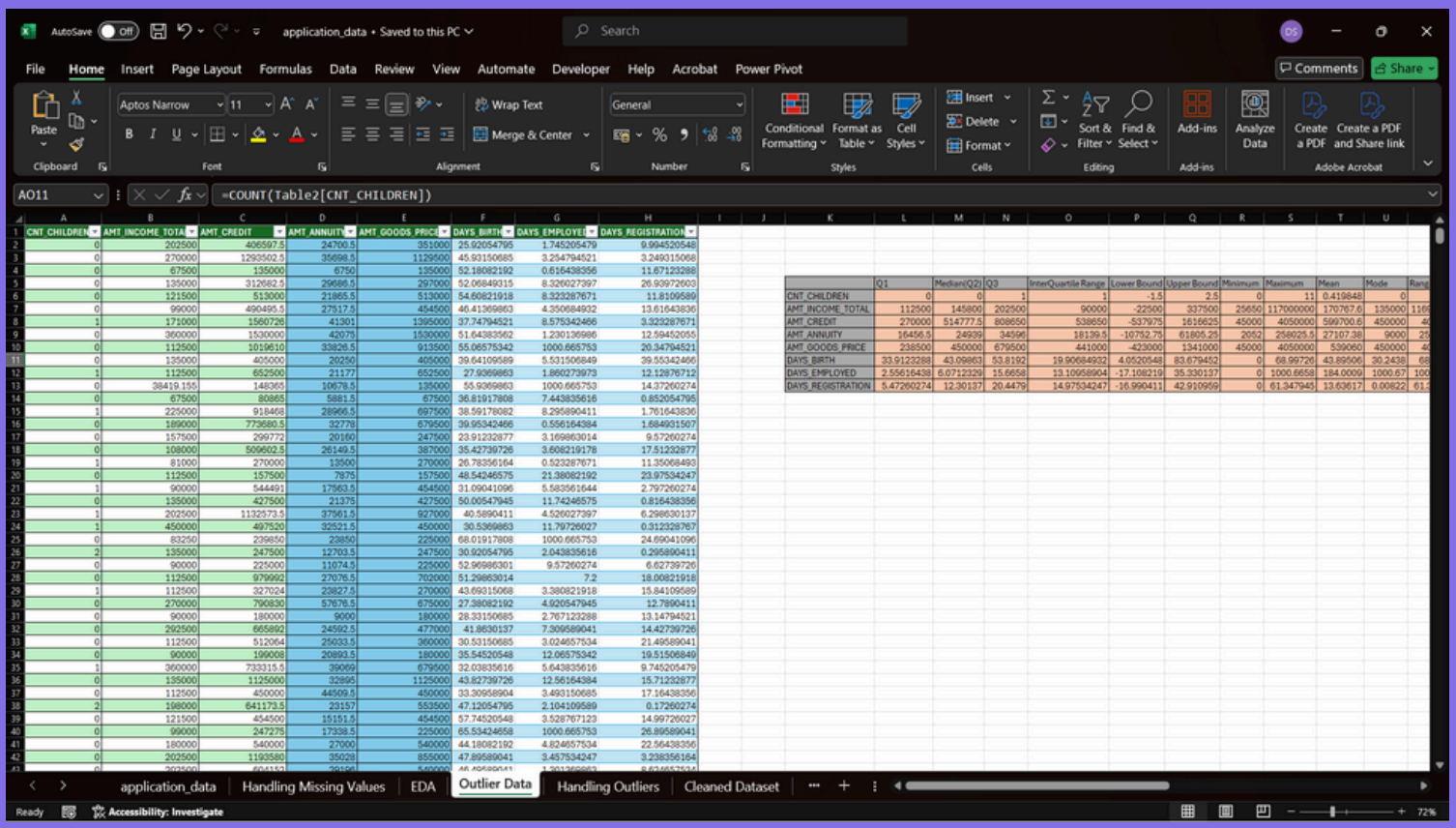
OUTLIER ANALYSIS

Method:

- Identified outliers using Interquartile Range (IQR).
- Calculated Q1, Q3, and IQR.
- Defined lower and upper bounds: Q1 - 1.5*IQR, Q3 + 1.5*IQR.
- Treated outliers by capping/extending within bounds.

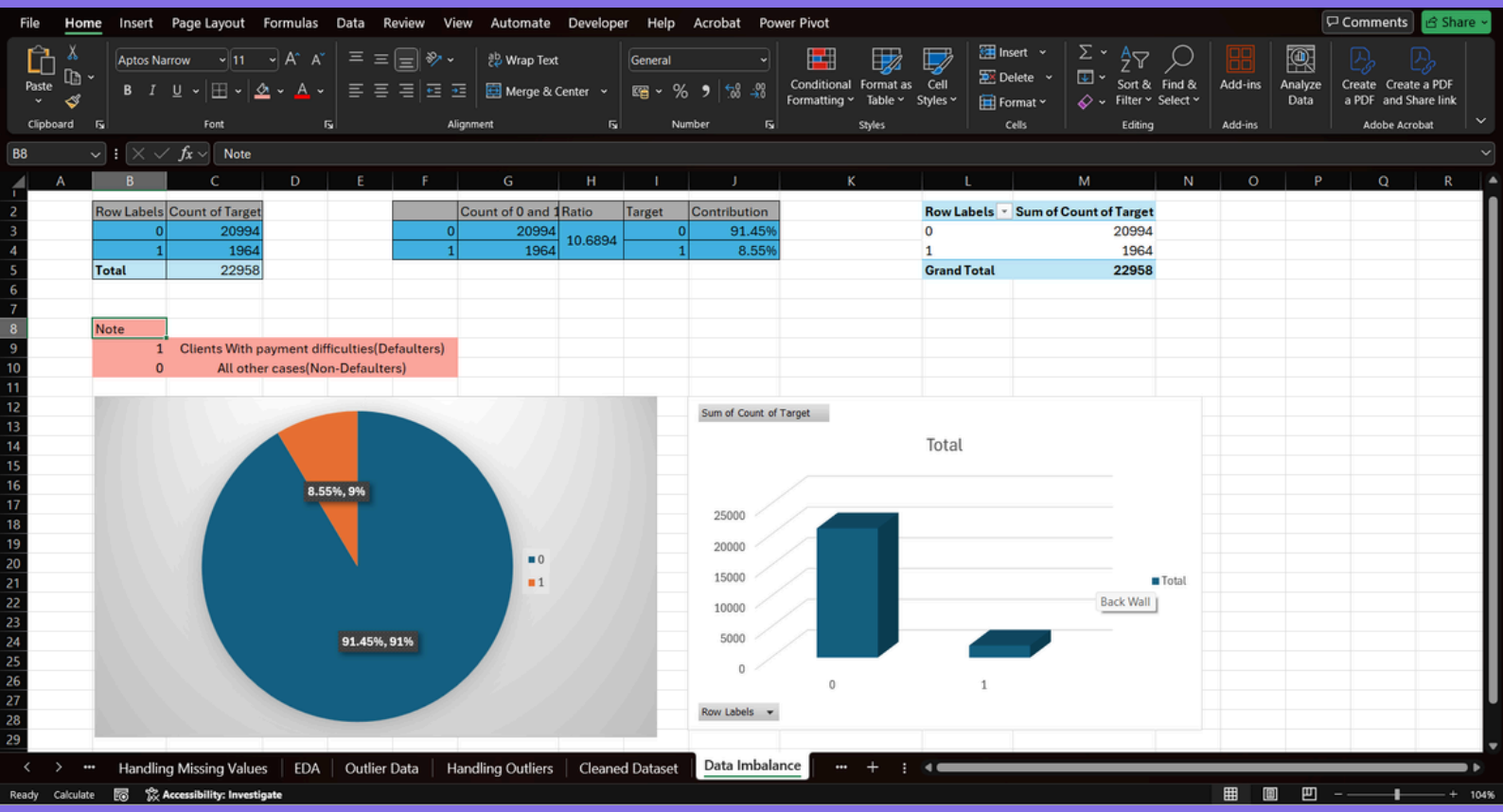
Visualization:

- Box plots for numerical variables (see Excel).



- The above tables depict two different set of data.
- The first are the numerical rows suspected of having outliers found through filtering the data.
- The second contains the quartile values and IQR(Inter Quartile Range) as well as the bounds required for the Box Plots and some other descriptive analysis such as mean, median, mode, standard deviation etc.

DATA IMBALANCE



Analysis:

- Assessed class distribution of target variable using COUNTIF.
- Imbalance found in loan outcomes, i.e., the male to female criteria has a big gap.

Visualization:

- Pie chart representing class distribution (see Excel).

UNIVARIATE AND BIVARIATE ANALYSIS

Univariate Analysis:

- Examined distributions of individual variables using histograms and bar charts.

Segmented Univariate Analysis:

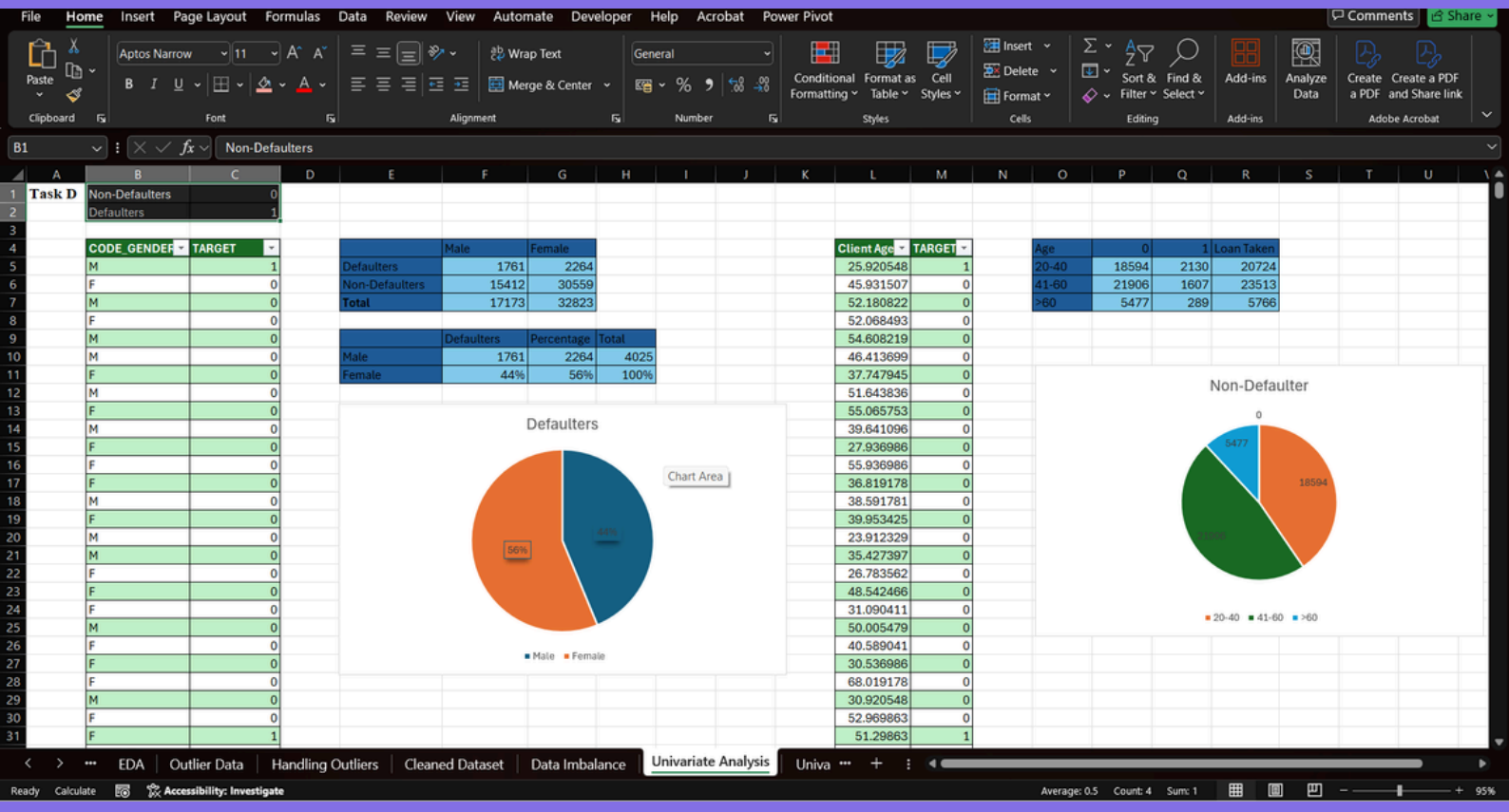
- Compared distributions for scenarios (e.g., defaulters vs non-defaulters).

Bivariate Analysis:

- Analyzed relationships between variables using scatter plots and pivot tables.

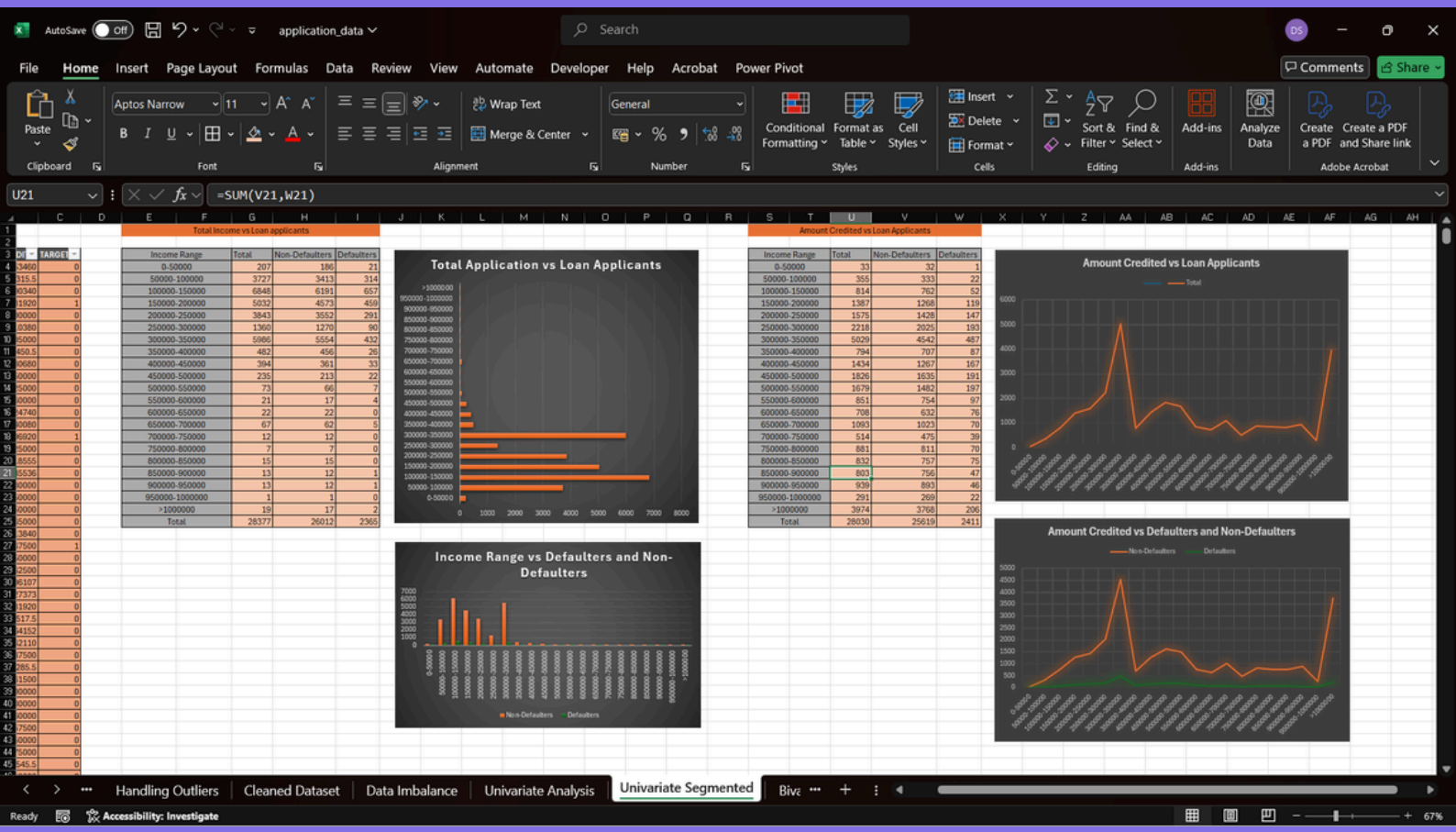
Visualizations:

- Various charts (see Excel).

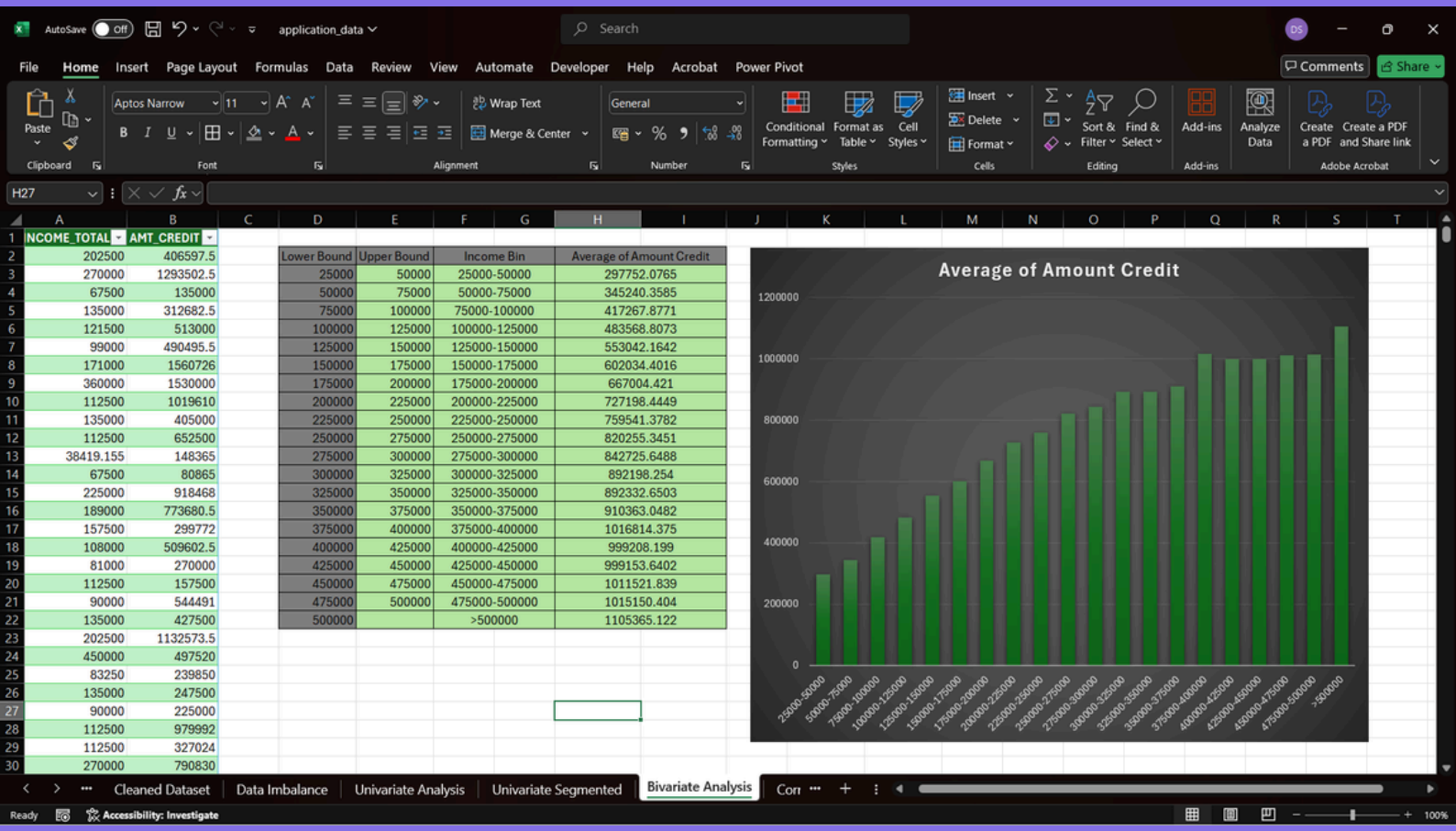


Univariate Analysis

Univariate Segmented Analysis



Bivariate Analysis



INCOME_TOTAL	AMT_CREDIT
202500	406597.5
270000	1293502.5
67500	135000
135000	312682.5
121500	513000
99000	490495.5
171000	1560726
360000	1530000
112500	1019610
135000	405000
112500	652500
38419.155	148365
67500	80865
225000	918468
189000	773680.5
157500	299772
108000	509602.5
81000	270000
112500	157500
90000	544491
135000	427500
202500	1132573.5
450000	497520
83250	239850
135000	247500
90000	225000
112500	979992
112500	327024
270000	790830

Lower Bound	Upper Bound	Income Bin	Average of Amount Credit
25000	50000	25000-50000	297752.0765
50000	75000	50000-75000	345240.3585
75000	100000	75000-100000	417267.8771
100000	125000	100000-125000	483568.8073
125000	150000	125000-150000	553042.1642
150000	175000	150000-175000	602034.4016
175000	200000	175000-200000	667004.421
200000	225000	200000-225000	727198.4449
225000	250000	225000-250000	759541.3782
250000	275000	250000-275000	820255.3451
275000	300000	275000-300000	842725.6488
300000	325000	300000-325000	892198.254
325000	350000	325000-350000	892332.6503
350000	375000	350000-375000	910363.0482
375000	400000	375000-400000	1016814.375
400000	425000	400000-425000	999208.199
425000	450000	425000-450000	999153.6402
450000	475000	450000-475000	1011521.839
475000	500000	475000-500000	1015150.404
500000		>500000	1105365.122

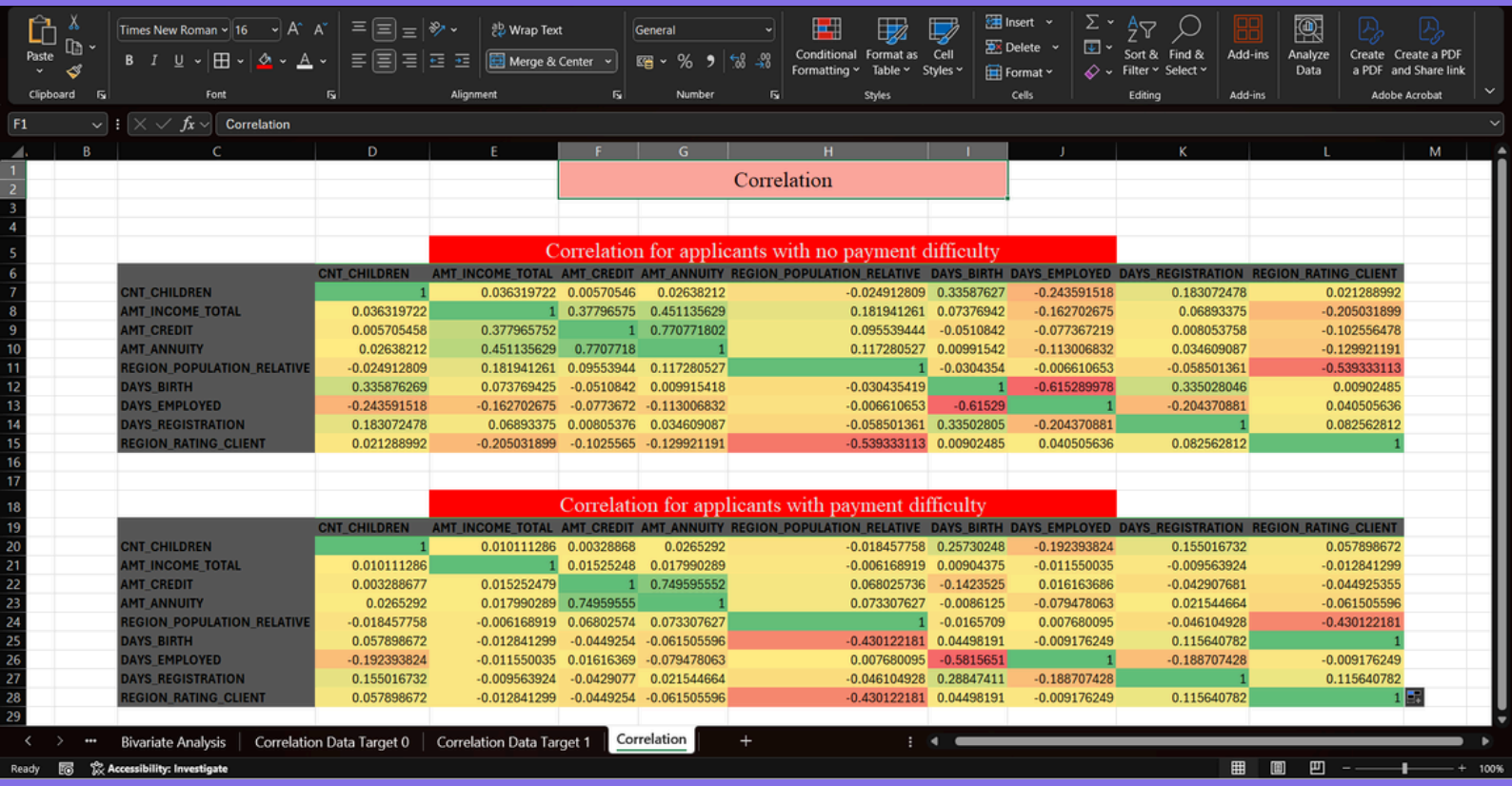
CORRELATION ANALYSIS

Segmented Correlation:

- Identified correlations within segmented data (e.g., defaulters).
- Used CORREL function to calculate correlation coefficients.
- Highlighted top predictors influencing loan default.

Visualization:

- Correlation heatmaps (see Excel).



Correlation

Correlation for applicants with no payment difficulty

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	REGION_RATING_CLIENT
CNT_CHILDREN	1	0.036319722	0.00570546	0.02638212	-0.024912809	0.33587627	-0.243591518	0.183072478	0.021288992
AMT_INCOME_TOTAL	0.036319722	1	0.37796575	0.451135629	0.181941261	0.07376942	-0.162702675	0.06893375	-0.205031899
AMT_CREDIT	0.005705458	0.377965752	1	0.770771802	0.095539444	-0.0510842	-0.077367219	0.008053758	-0.102556478
AMT_ANNUITY	0.02638212	0.451135629	0.7707718	1	0.117280527	0.00991542	-0.113006832	0.034609087	-0.129921191
REGION_POPULATION_RELATIVE	-0.024912809	0.181941261	0.09553944	0.117280527	1	-0.0304354	-0.006610653	-0.058501361	-0.539333113
DAYS_BIRTH	0.335876269	0.073769425	-0.0510842	0.009915418	-0.030435419	1	-0.615289978	0.335028046	0.00902485
DAYS_EMPLOYED	-0.243591518	-0.162702675	-0.0773672	-0.113006832	-0.006610653	-0.61529	1	-0.204370881	0.040505636
DAYS_REGISTRATION	0.183072478	0.06893375	0.00805376	0.034609087	-0.058501361	0.33502805	-0.204370881	1	0.082562812
REGION_RATING_CLIENT	0.021288992	-0.205031899	-0.1025565	-0.129921191	-0.539333113	0.00902485	0.040505636	0.082562812	1

Correlation for applicants with payment difficulty

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	REGION_RATING_CLIENT
CNT_CHILDREN	1	0.010111286	0.00328868	0.0265292	-0.018457758	0.25730248	-0.192393824	0.155016732	0.057898672
AMT_INCOME_TOTAL	0.010111286	1	0.01525248	0.017990289	-0.006168919	0.00904375	-0.011550035	-0.009563924	-0.012841299
AMT_CREDIT	0.003288677	0.015252479	1	0.749595552	0.068025736	-0.1423525	0.016163686	-0.042907681	-0.044925355
AMT_ANNUITY	0.0265292	0.017990289	0.74959555	1	0.073307627	-0.0086125	-0.079478063	0.021544664	-0.061505596
REGION_POPULATION_RELATIVE	-0.018457758	-0.006168919	0.06802574	0.073307627	1	-0.0165709	0.007680095	-0.046104928	-0.430122181
DAYS_BIRTH	0.057898672	-0.012841299	-0.0449254	-0.061505596	-0.430122181	0.04498191	-0.009176249	0.115640782	1
DAYS_EMPLOYED	-0.192393824	-0.011550035	0.01616369	-0.079478063	0.007680095	-0.5815651	1	-0.188707428	-0.009176249
DAYS_REGISTRATION	0.155016732	-0.009563924	-0.0429077	0.021544664	-0.046104928	0.28847411	-0.188707428	1	0.115640782
REGION_RATING_CLIENT	0.057898672	-0.012841299	-0.0449254	-0.061505596	-0.430122181	0.04498191	-0.009176249	0.115640782	1

Bivariate Analysis | Correlation Data Target 0 | Correlation Data Target 1 | Correlation

Correlation Analysis

INSIGHTS AND RECOMMENDATIONS

Key Insights:

- High-income customers less likely to default.
- Loan amounts exceeding a threshold increase default risk.
- Data imbalance in loan approvals impacts analysis.

Recommendations:

- Implement stricter checks for high-risk applicants.
- Offer loans with adjusted terms to risky applicants (e.g., higher interest).
- Improve data collection to reduce missing values.

Module 7 : Analyzing Impact of Features on Car Prices & Profitability

Objective:

- To analyze the influence of various car features on customer preferences and purchasing decisions, providing insights for the automobile industry.

Plan:

- Define the scope of analysis:
- Focus on factors such as pricing, safety features, fuel efficiency, and design.
- Examine customer demographics and preferences.

Prepare:

- Data Sources:
  1.) Car sales data.
  2.) Customer feedback and survey data.

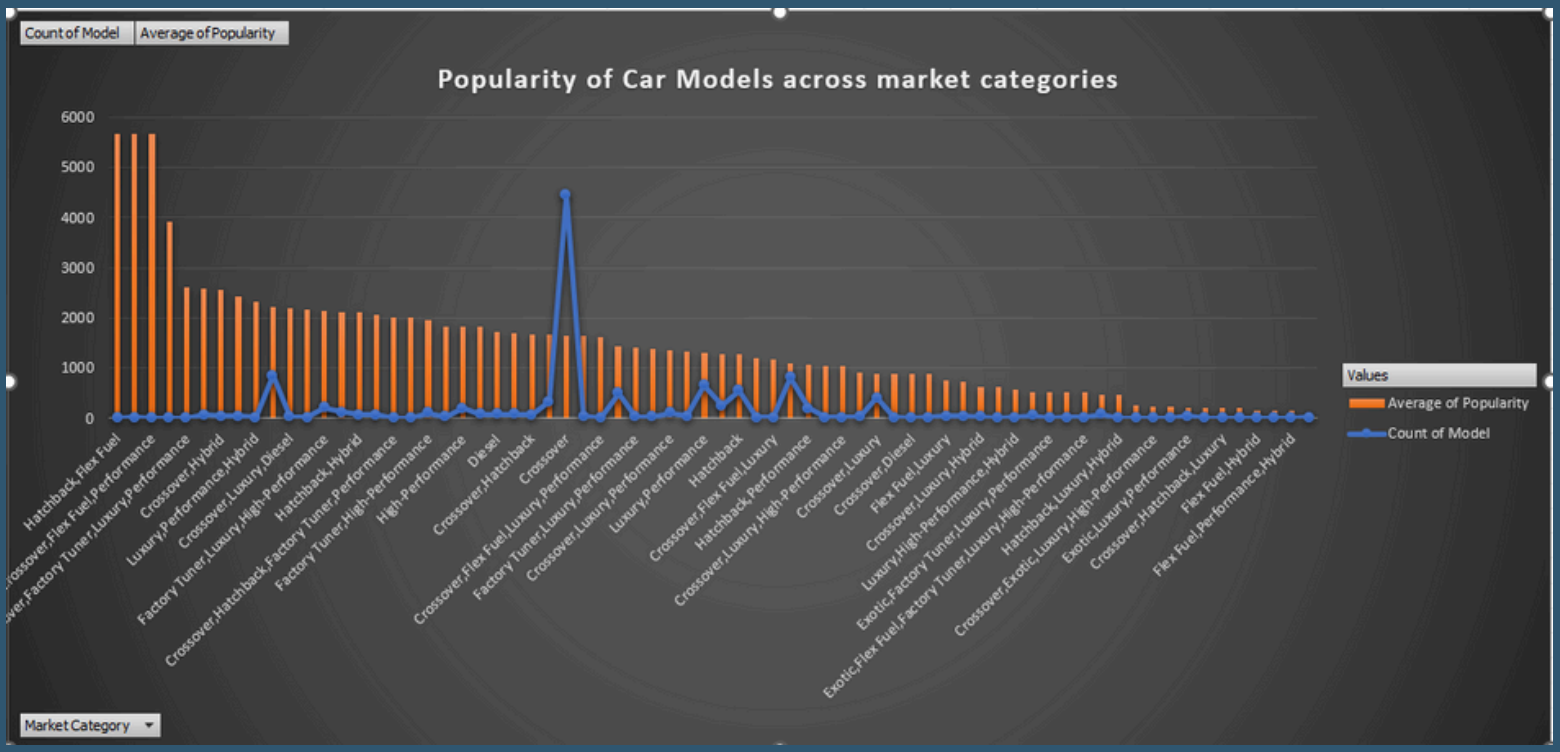
Tools Used:

- Excel for data cleaning and analysis.
- Statistical tools for identifying trends and correlations.

Process:

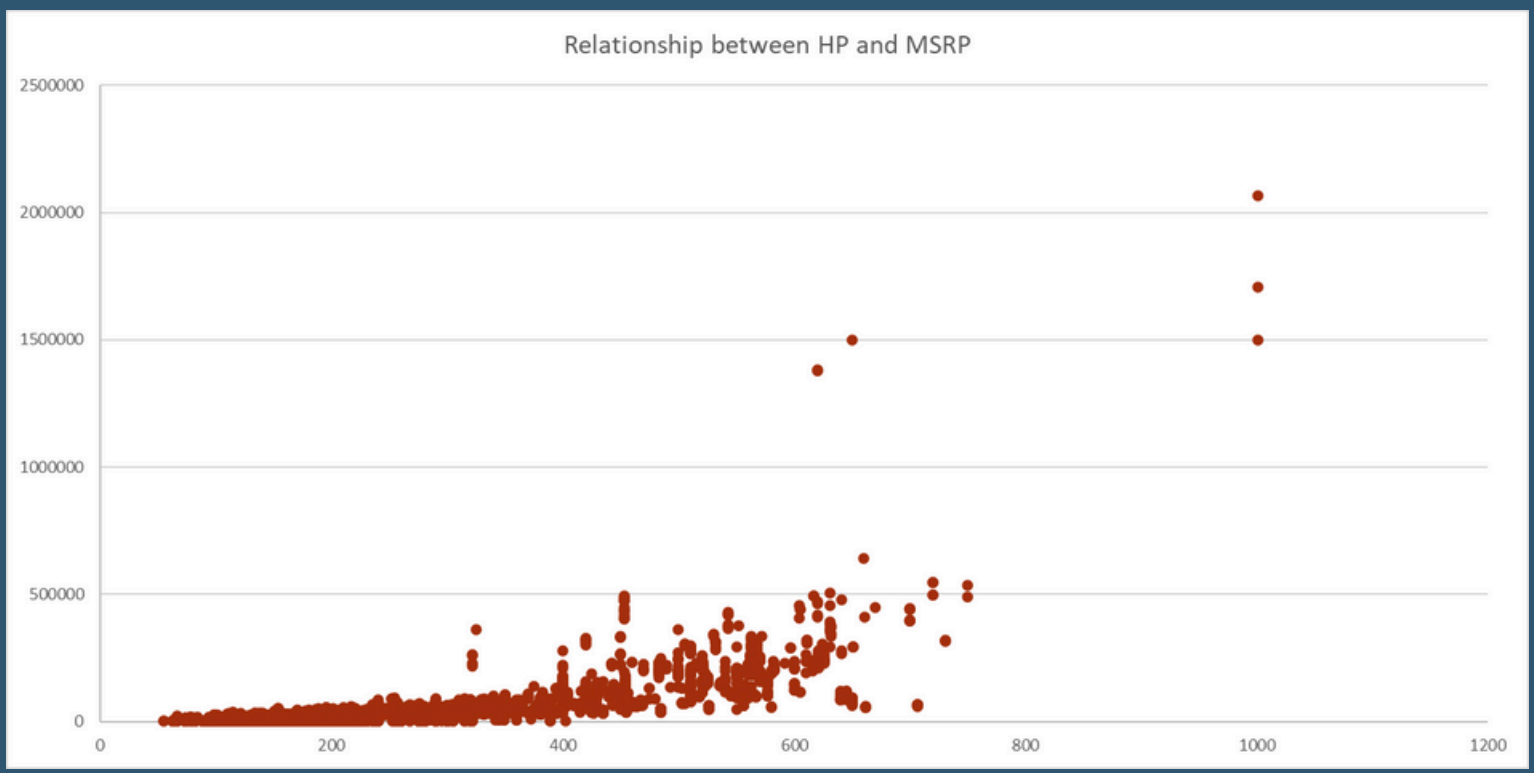
- Data Cleaning and Preparation:
  1.) Remove incomplete or redundant data entries.
  2.) Categorize features such as safety ratings, mileage, and luxury components.
- Data Analysis Setup:
  1.) Organize data into segments by price range, brand, and feature type.
  2.) Apply statistical measures to assess feature impact.

Task 1: Popularity of car models across market categories.

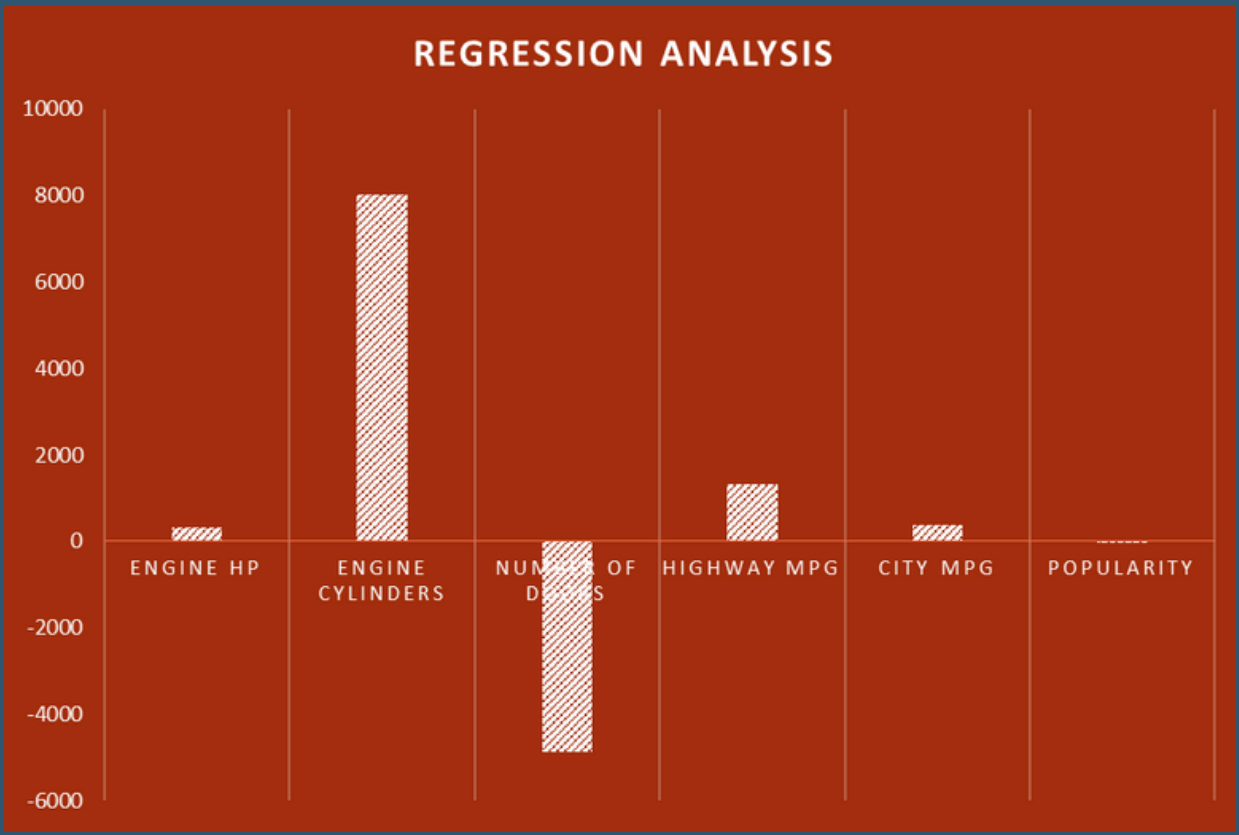


INSIGHTS : Crossover has the most number of models with the value of 4440

Task 2: Relationship between engine power (HP) and price (MSRP)



Task 3: Features affecting price.

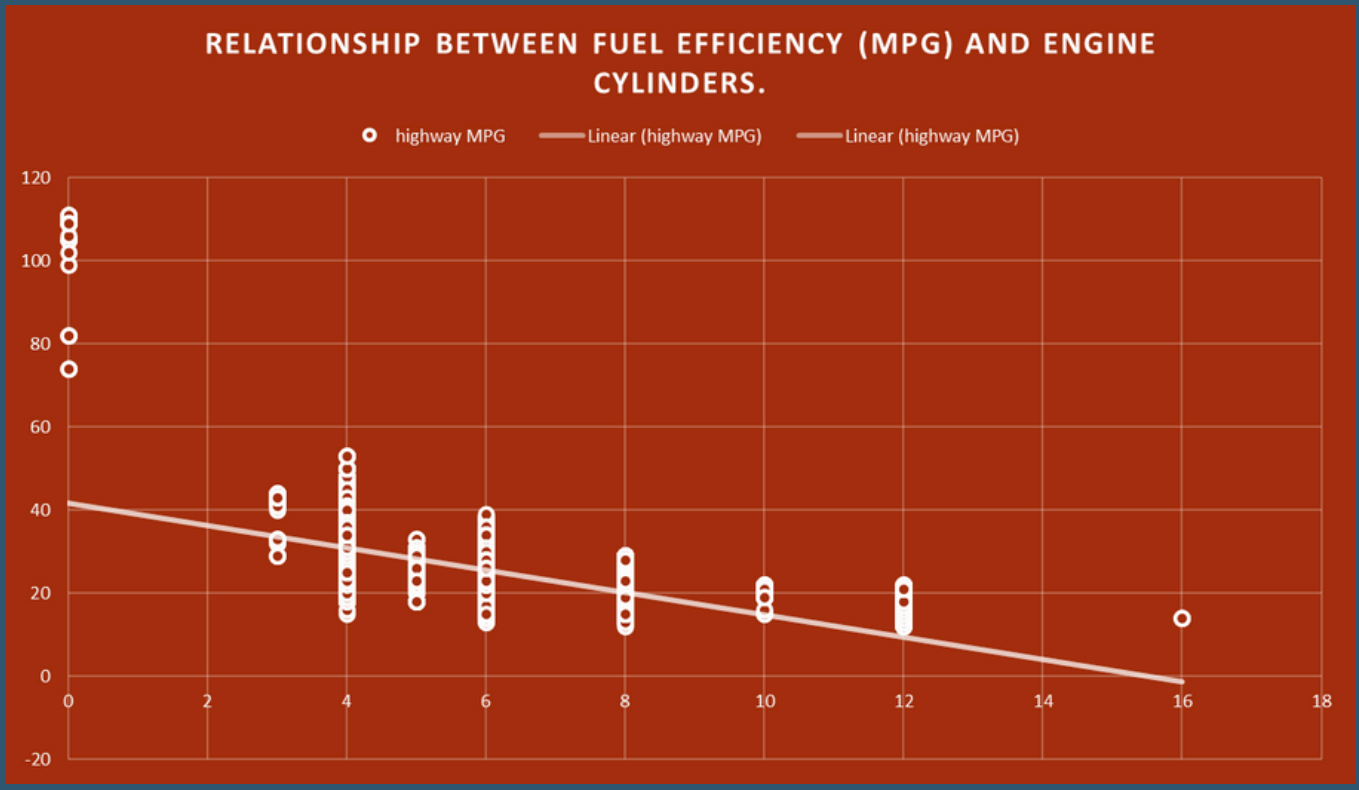


Insights : Engine Cylinders with the highest while Number of doors with the lowest regression value.

Task 4: Average price variation by manufacturer.

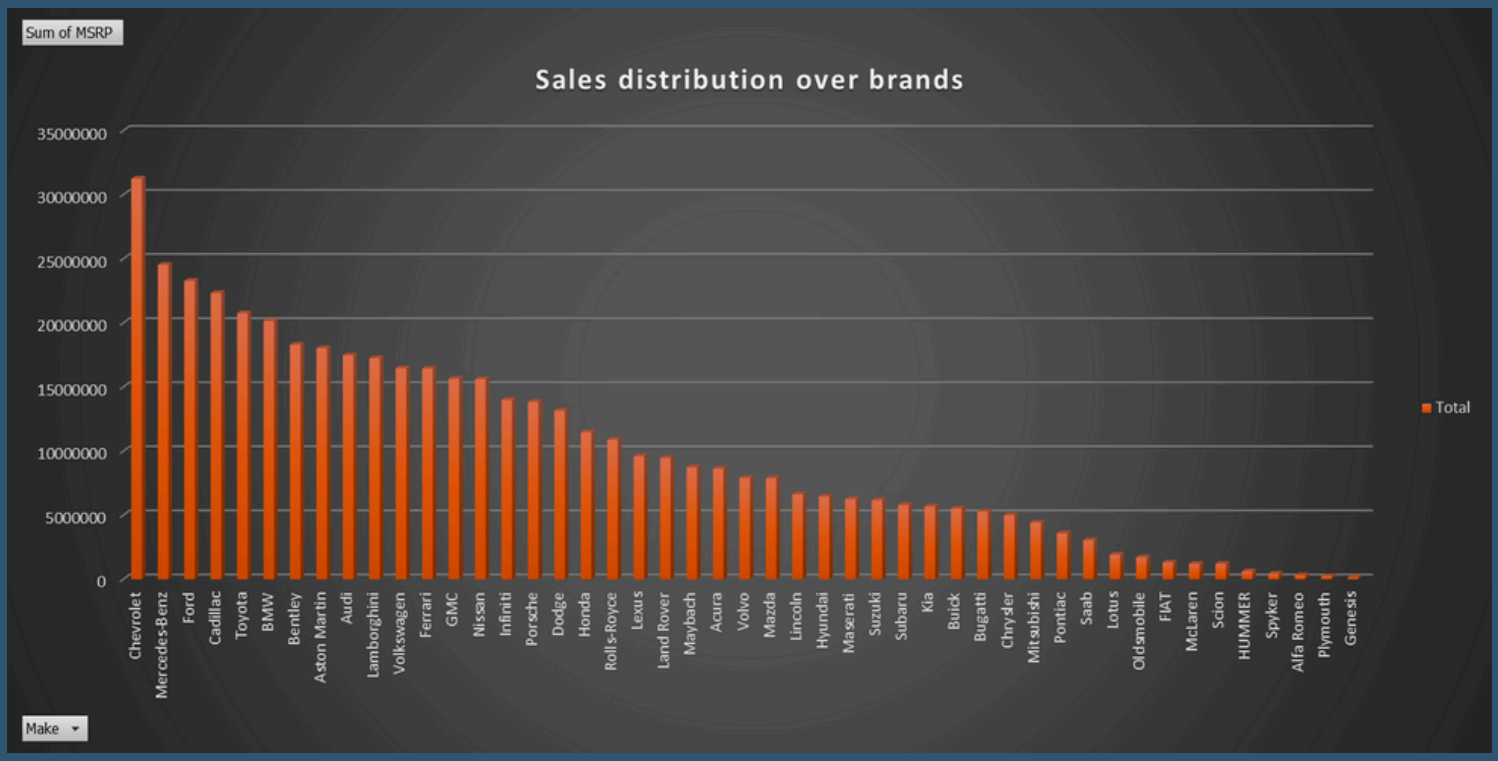
Insights : Bugatti and Rolls-Royce lead the average pricing in the market.

Task 5: Relationship between fuel efficiency (MPG) and engine cylinders.



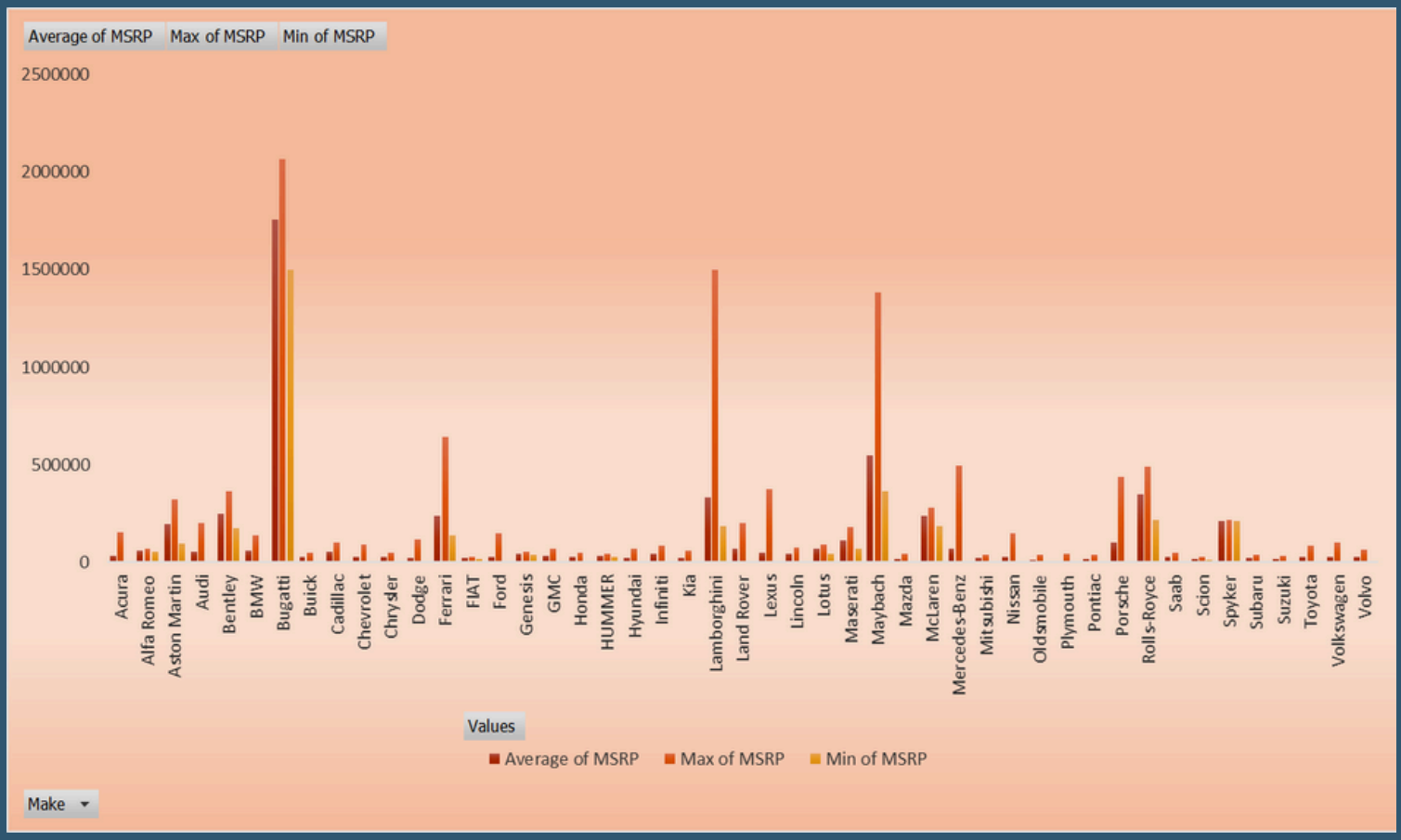
DASHBOARD TASKS

Task 1: Distribution of car prices by brand and body style.

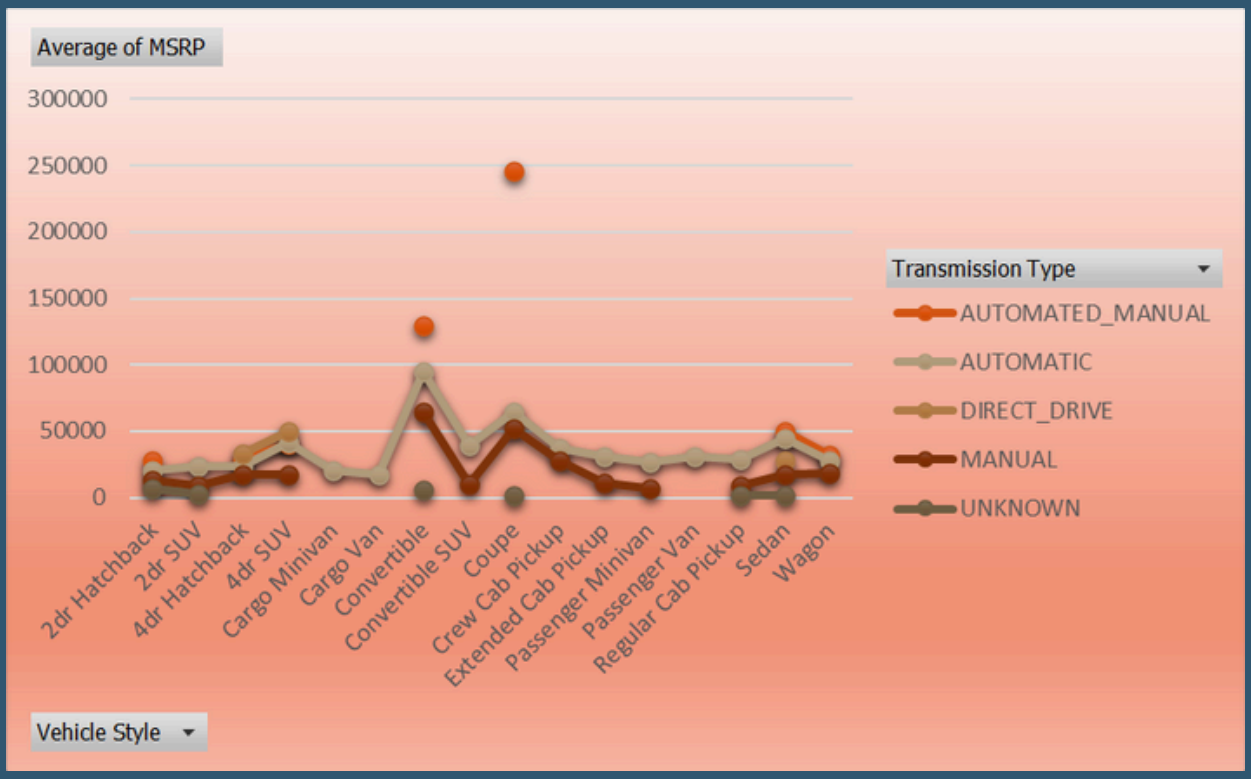


Insights : Chevrolet has the highest Sales distribution of 31252763

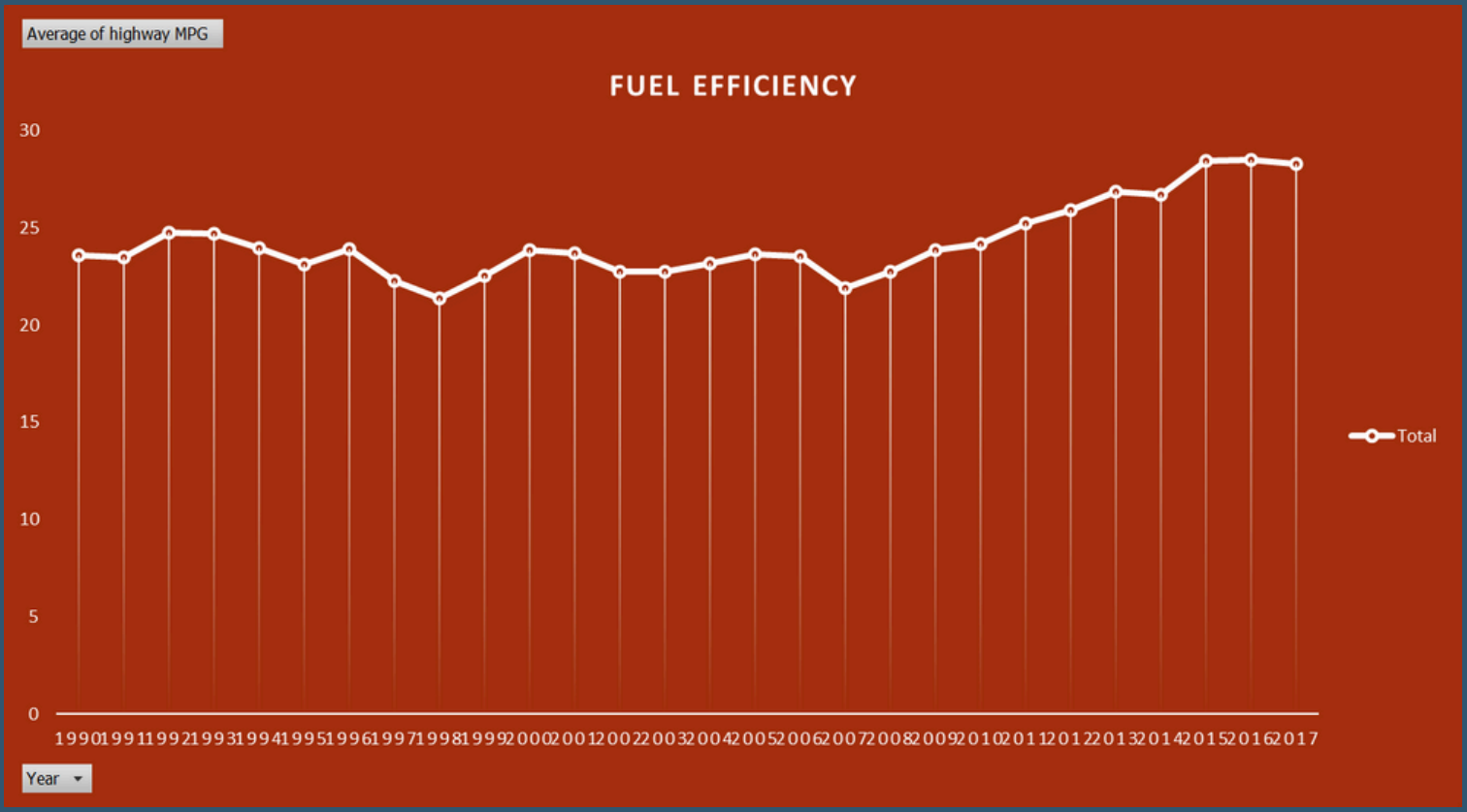
Task 2: Brands with the highest and lowest average MSRPs by body style.



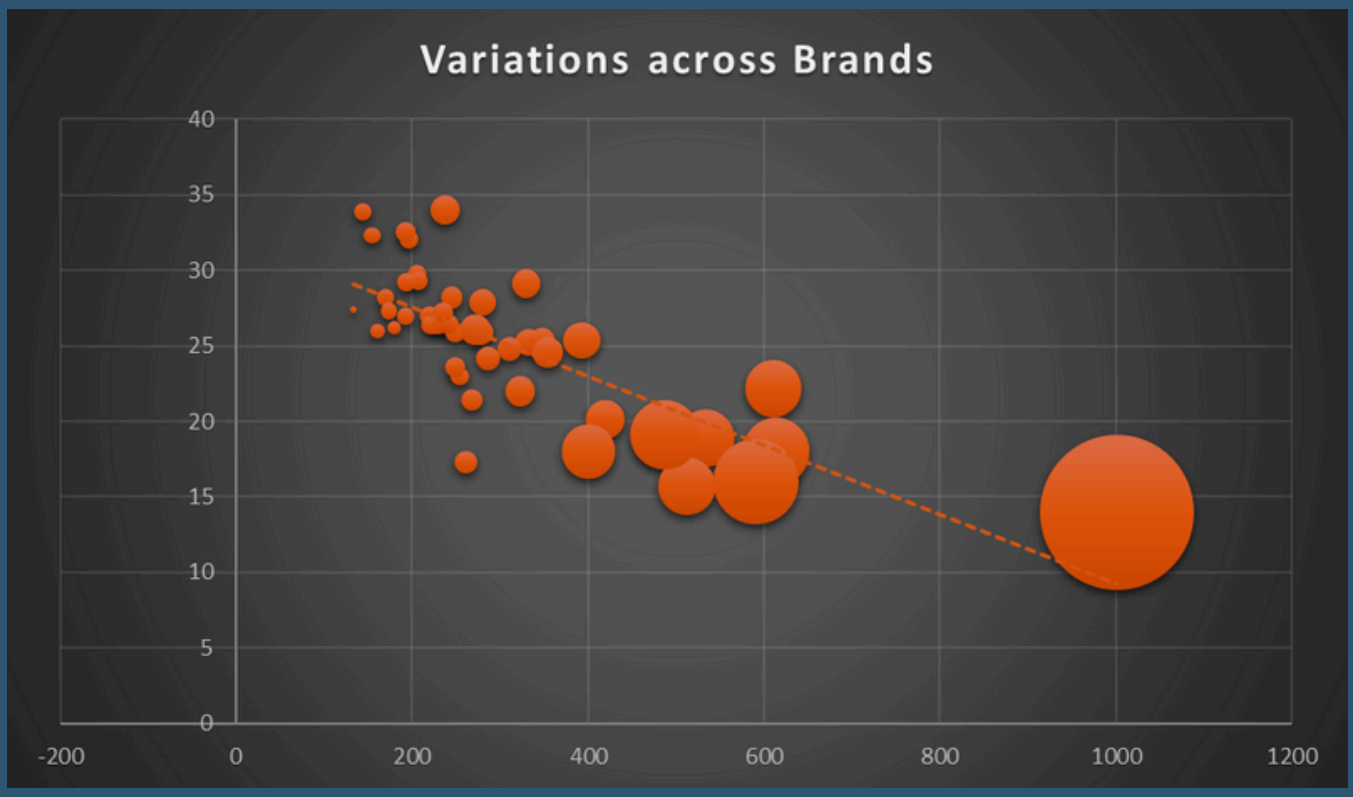
Task 3: Effect of transmission type on MSRP by body style.



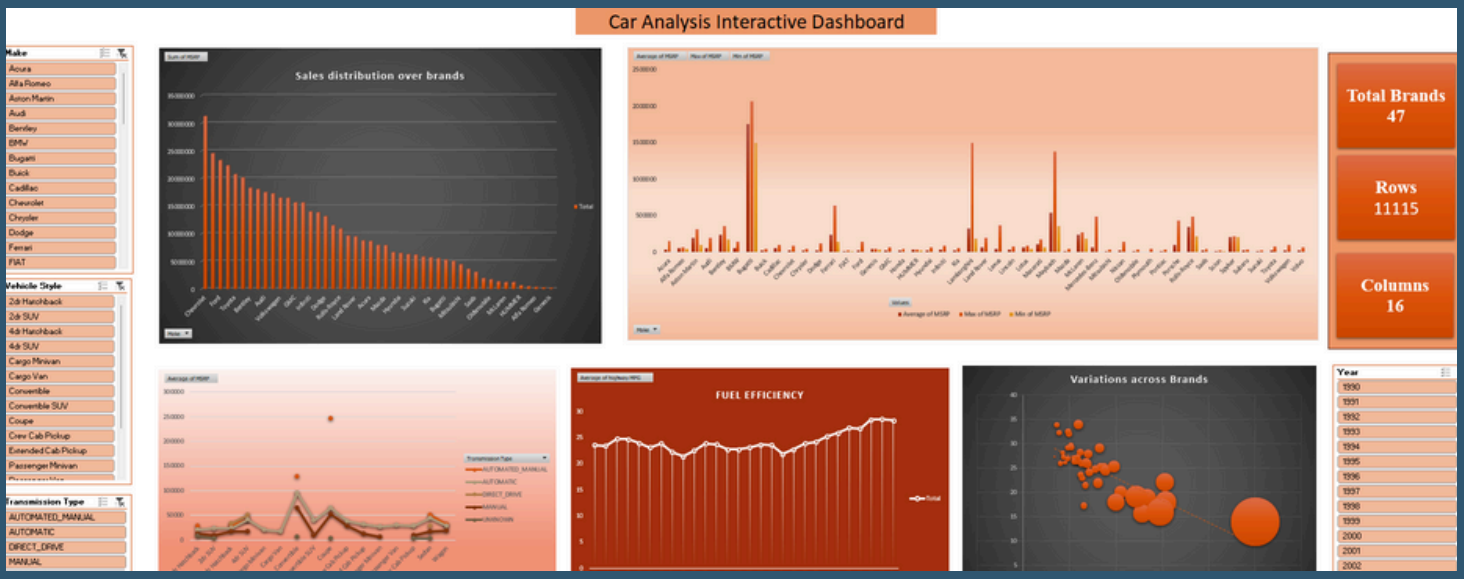
Task 4: Fuel efficiency trends by body styles and model years.



Task 5: Variations in horsepower, MPG, and price across brands.



Dashboard



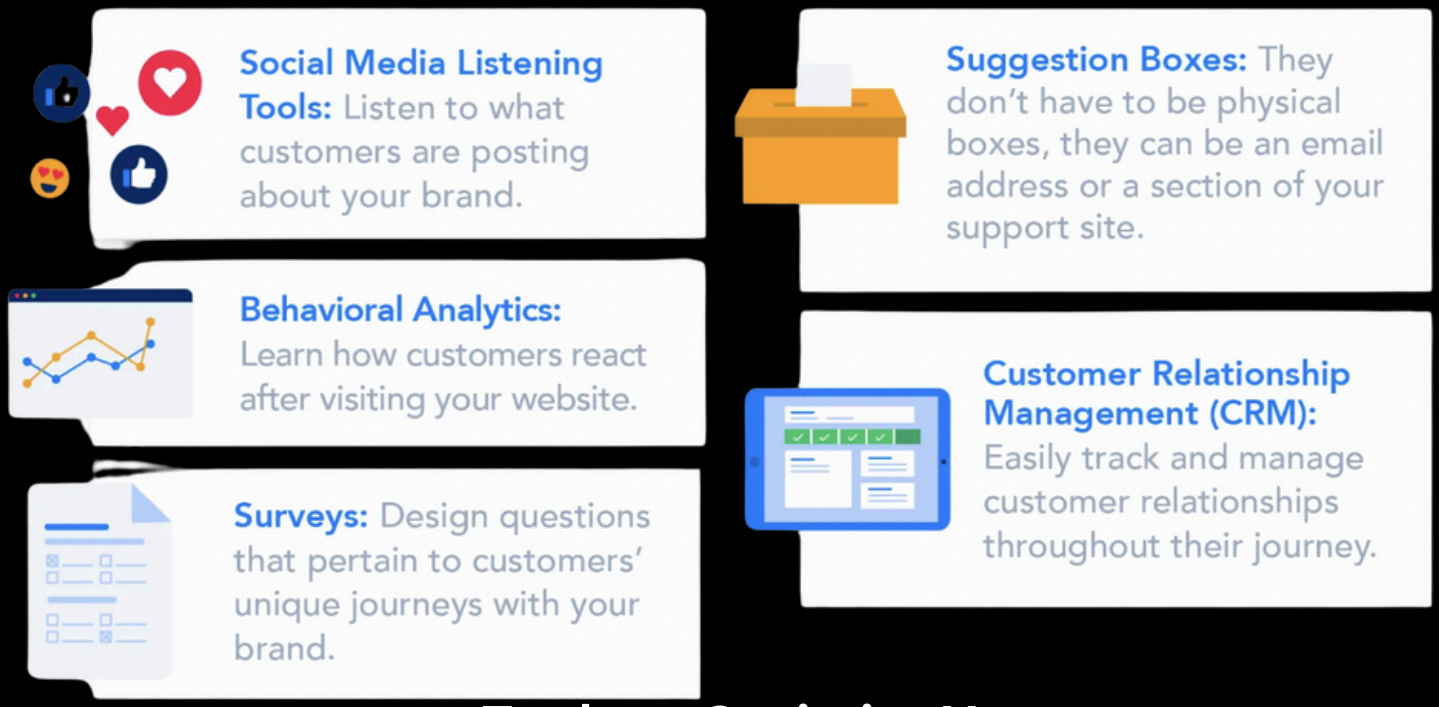
INSIGHTS SUMMARY

1. Popularity Insights: Luxury and SUVs are consistently popular market categories.
2. Horsepower and Price: Strong positive correlation; higher HP vehicles demand premium pricing.
3. Price Drivers: Engine HP, fuel type, and luxury category significantly influence MSRP.
4. Manufacturer Pricing: Premium brands like BMW and Mercedes dominate high average MSRP segments.
5. Fuel Efficiency: Higher cylinder counts correlate with lower MPG.

RECOMMENDATIONS

- Invest in high-HP, fuel-efficient cars for premium pricing.
- Focus R&D on improving MPG for SUVs and high-cylinder cars.
- Target marketing efforts on popular segments like luxury SUVs.
- Expand offerings in low-MSRP segments for emerging markets.

Module 8 : ABC Call Volume Trend Analysis



Tools to Optimize Your Customer Experience

Objective:
- To analyze call volume trends for ABC, identifying peak periods and underlying patterns to improve resource allocation and customer support efficiency.

Plan:
- Define key metrics for analysis:
- Call volume by time and day.
- Average call handling time.
- Trends in call types and resolutions.
- Establish the timeline and objectives for analysis and reporting.

Prepare:
- Data Sources:
  1.) Call logs from ABC's customer service platform.
  2.) Historical records of call handling times and resolutions.

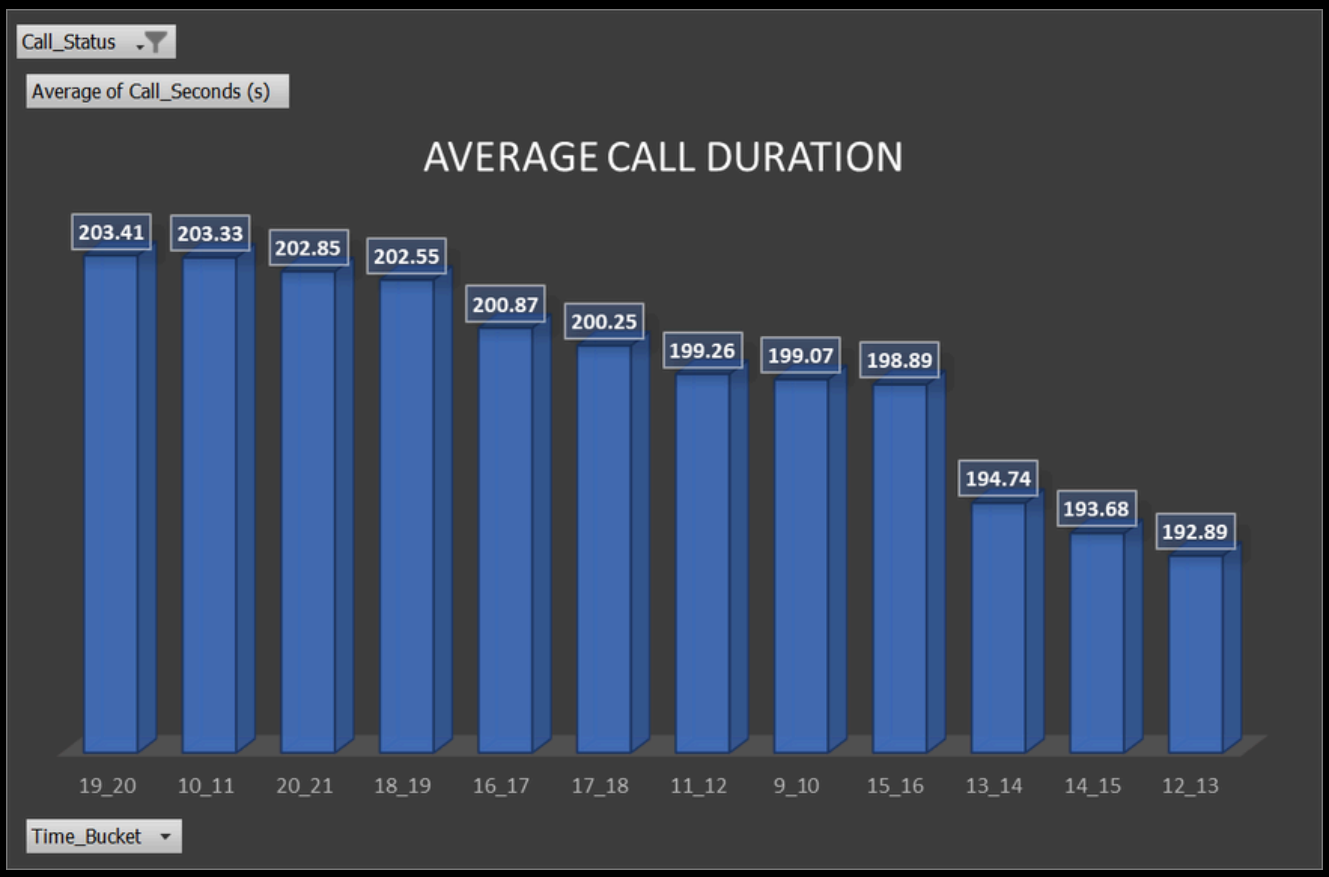
Tools Used:
- SQL for data extraction and querying.
- Excel for data preparation and initial analysis.
- Tableau for visualizing trends.

Process:
- Data Cleaning and Preparation:
  1.) Remove invalid or incomplete call entries.
  2.) Organize data by date, time, and call type.

Data Segmentation:
- Categorize calls by type (e.g., inquiries, complaints, technical support).
- Segment data based on call timing (peak vs. off-peak hours).

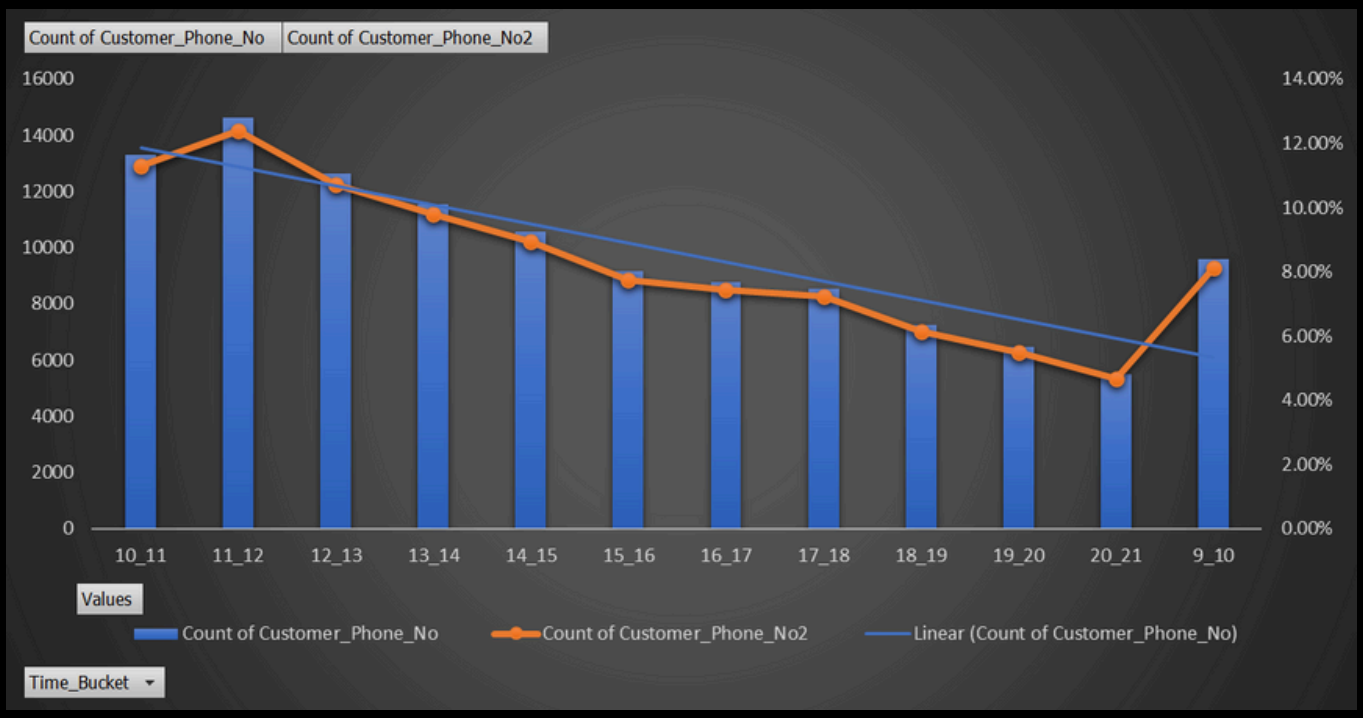
Data Analytic Tasks

Task 1 : Average Duration of calls



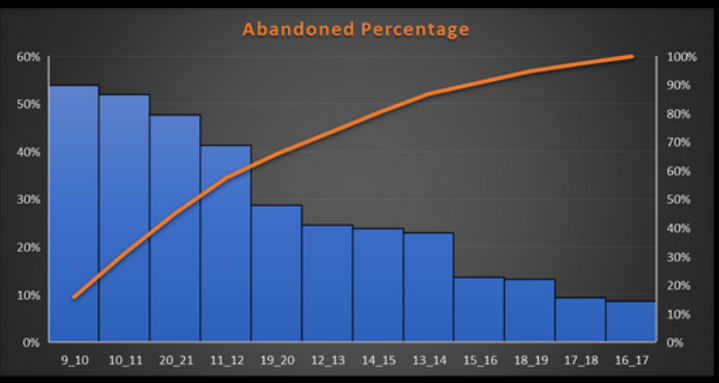
Insights : Time Bucket 7-8 evening and 10 -11 morning are the busiest with the highest average duration of call time.

Task 2 : Call Volume Analysis



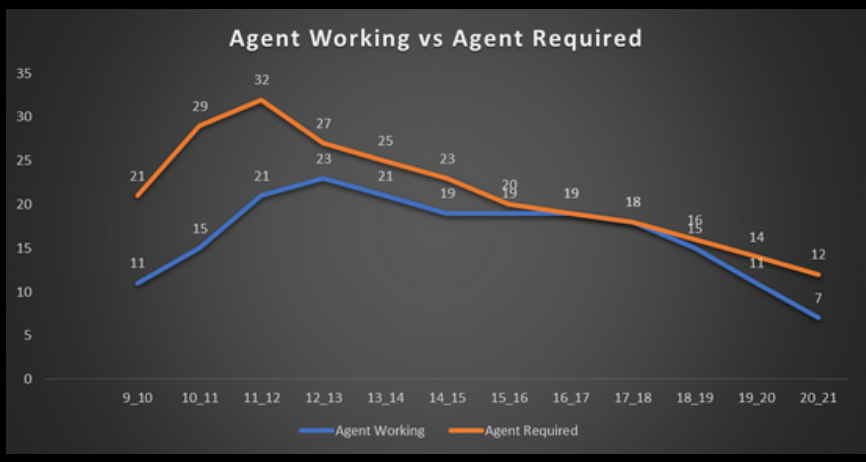
Insights : Similar to above , Time bucket of morning shift and late night hold the most volume of calls.

Task 3 : Manpower Planning



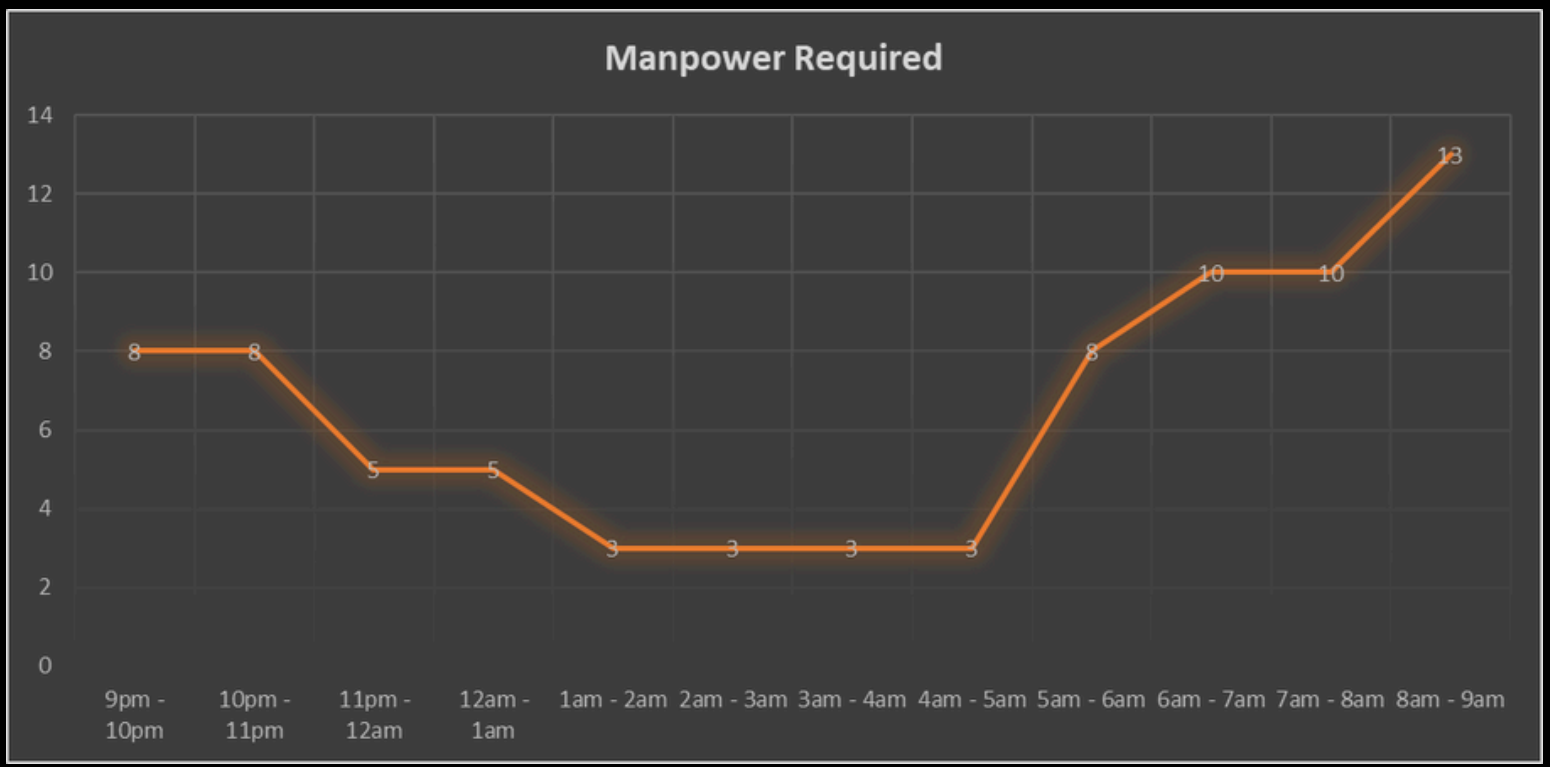
Abandoned calls Percentage

Agents Required for Proper functioning



Insights : Early morning calls are the most abandoned getting less with increasing time and agents are less in number than required.

Task 4 : Manpower Planning for Night-Shift



Insights : Early morning receives the most interest for the night time shift with late night just behind

INSIGHTS

1. Average Call Duration:
   - The average call duration varied across time buckets, with peak durations occurring in the late morning (10 AM to 12 PM).
2. Call Volume Trends:
   - Call volumes peaked during mid-morning (10 AM to 12 PM) and late afternoon (3 PM to 5 PM).
   - A consistent decline in call volumes was observed after 7 PM.
3. Day Shift Manpower Planning:
   - To reduce the abandonment rate to 10%, the required agents per time bucket were calculated. Peak hours required up to 12 agents, while non-peak hours needed fewer agents.
4. Night Shift Manpower Planning:
   - For every 100 calls during the day, an additional 30 calls occurred at night. Manpower allocation was proposed for night shifts to maintain a 10% abandonment rate.

RESULTS

1. **Day Shift Manpower Requirements:**
   - A detailed manpower allocation plan was created for each time bucket from 9 AM to 9 PM.
   - Unplanned leaves and agent productivity were considered to ensure adequate staffing.
2. **Night Shift Plan:**
   - Staffing was recommended for time buckets from 9 PM to 9 AM based on a proportional distribution of night calls.
   - This included 3-5 agents per time bucket during night hours.
3. **Customer Experience Improvement:**
   - The proposed plan is expected to significantly reduce the call abandonment rate, improving customer satisfaction and loyalty.

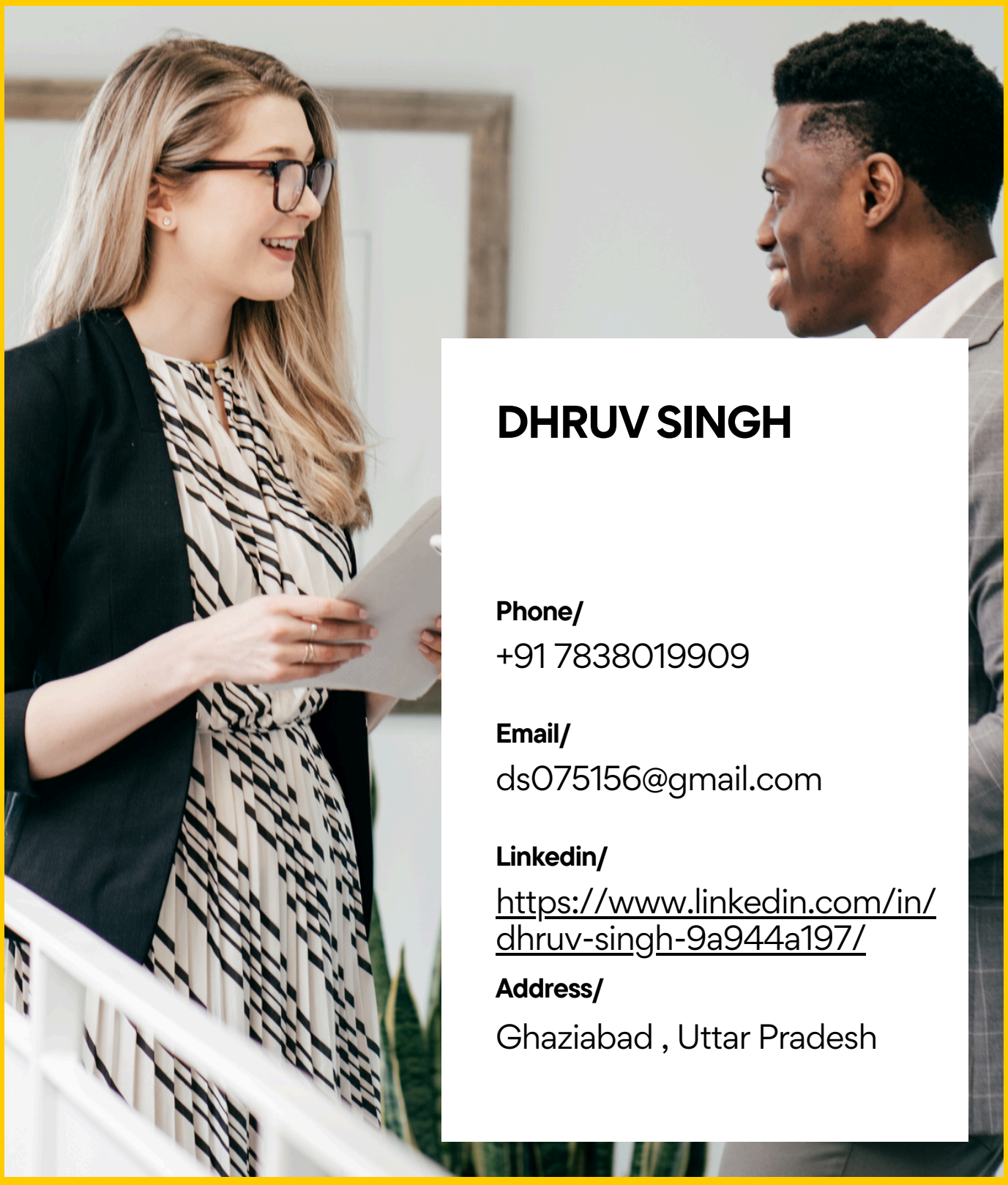
LEARNINGS AND REFLECTIONS

These projects have equipped me with the skills and confidence to handle diverse datasets, extract meaningful insights, and deliver impactful recommendations. They have honed my ability to use analytical tools effectively and communicate results clearly. Moving forward, I aim to apply these learnings in real-world scenarios to drive innovation and success in data-driven organizations.

APPENDIX

1. **Module 1(Phone Upgrade Savings Analysis) :** https://drive.google.com/drive/folders/18Eo2iS9tBG5i5tI9zU9rlMNjDL5pgF41?usp=drive_link
2. **Module 2(Instagram User Analytics) :** https://drive.google.com/drive/folders/1bkDat6vuLS-fjZihxZJx60_4Byp6t3hM?usp=drive_link
3. **Module 3(Operation & Metric Analytics) :** https://drive.google.com/drive/folders/12-zfj041T3C4uwkTlci55-7nEItLlYjw?usp=drive_link
4. **Module 4(Hiring Process Analytics) :** https://drive.google.com/drive/folders/19JN6FOHOlAcoAGq9YuZ-5H3UJL4-t-Hg?usp=drive_link
5. **Module 5(IMDB Movie Analysis) :** https://drive.google.com/drive/folders/1uBEmyn_q2dQf5ak37GGjiSh1pQHuaFFq?usp=drive_link
6. **Module 6(Bank Loan Case Study) :** https://drive.google.com/drive/folders/1aqDBXx-RHSiZ63So1l9Pk7CDI5sVfwcZ?usp=drive_link
7. **Module 7(Impact of Car Features) :** https://drive.google.com/drive/folders/1VLxPrAnevAY371zeDRyAQM5Gu-cWLjHw?usp=drive_link
8. **Module 8(ABC Call Volume Trend) :** https://drive.google.com/drive/folders/1G5qvYXTLPcatGHAt3-Q5AkiTYxqJKUNF?usp=drive_link

CONTACT ME



DHRUV SINGH

Phone/
+91 7838019909

Email/
ds075156@gmail.com

Linkedin/
https://www.linkedin.com/in/dhruv-singh-9a944a197/

Address/
Ghaziabad , Uttar Pradesh