Data Analytics Portfolio

Prepared By:- Vishal Patil

Professional Background

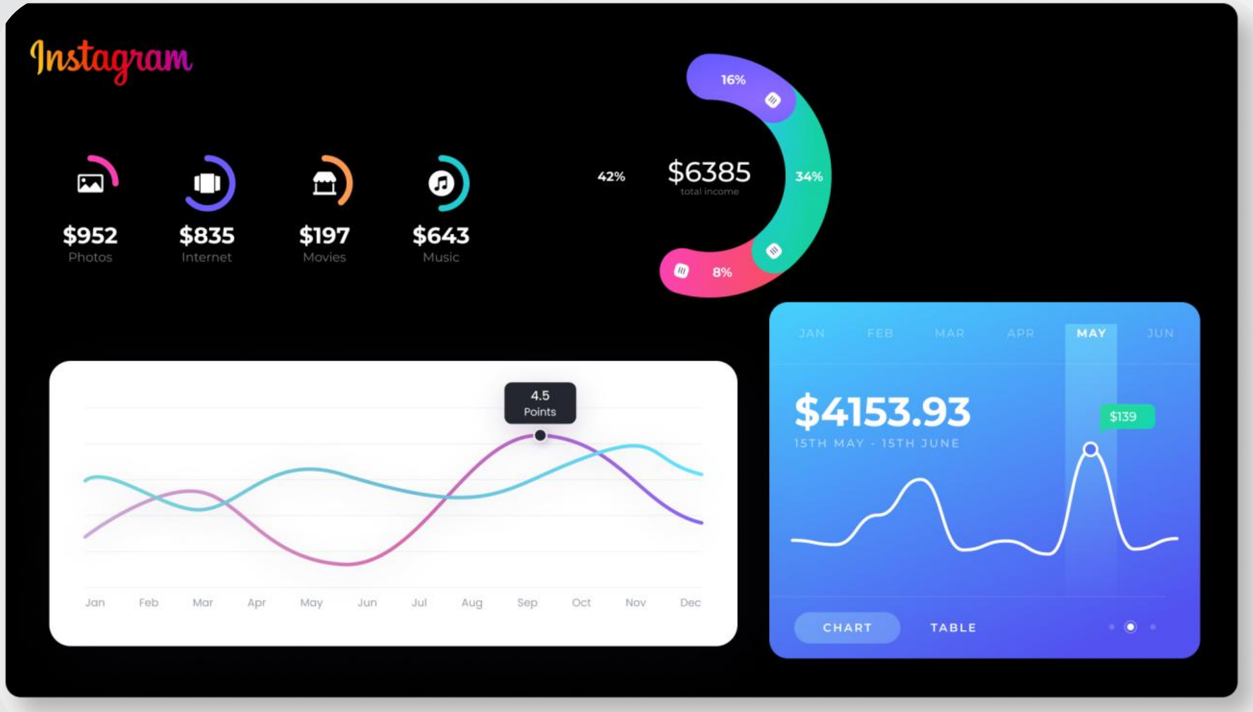
I am currently pursuing my B.Tech Degree(final year) of Computer Science and Engineering in Tontadarya College of engineering, Gadag.I have secured 8.1 CGPA(till 6th sem).

I have attained several skills like Data Analysis, Python, Java and I have worked with different projects by considering above skills.

As a fresher I am willing to experience real world challenges in the cooperate world and as a fresher I am very flexible and adaptive to the culture. I am ready to learn new things which are suitable for the company. I am waiting for the opportunities to work where I can showcase my skills and give my full efforts to the company.

Table Of Contents

Project : Instagram User Analytics



Project Description

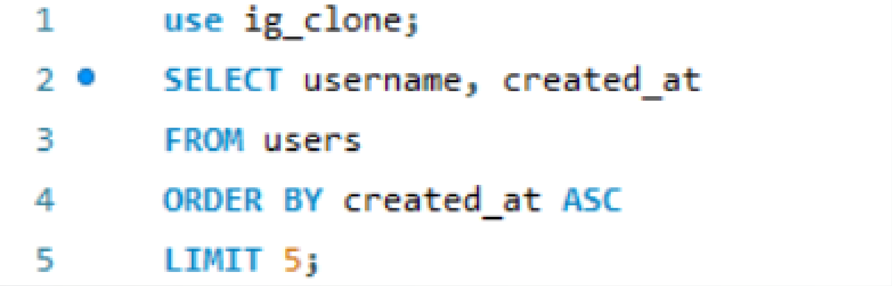
The project involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow. User analysis involves tracking how users engage with a digital product, such as a software application or a mobile app. The goal of this project is to use your SQL skills to extract meaningful insights from the data.

Findings

A) Marketing Analysis:

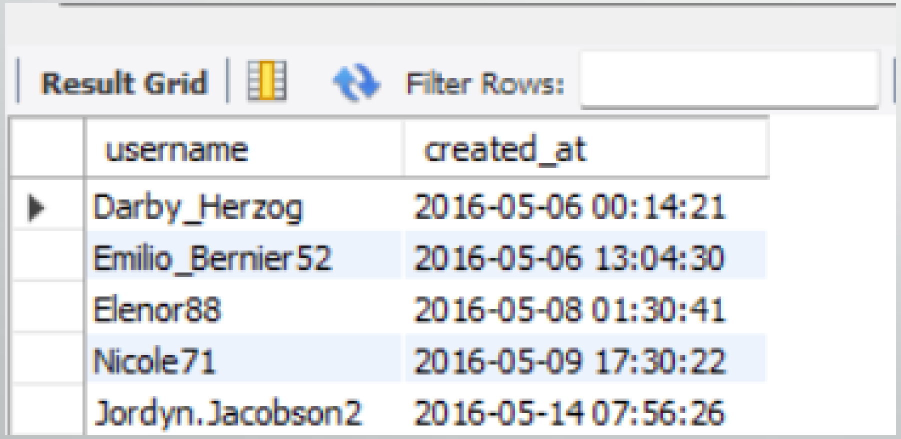
1.Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
Task: Identify the five oldest users on Instagram from the provided database.

Query:



```
use ig_clone;
SELECT username, created_at
FROM users
ORDER BY created_at ASC
LIMIT 5;
```

Output:

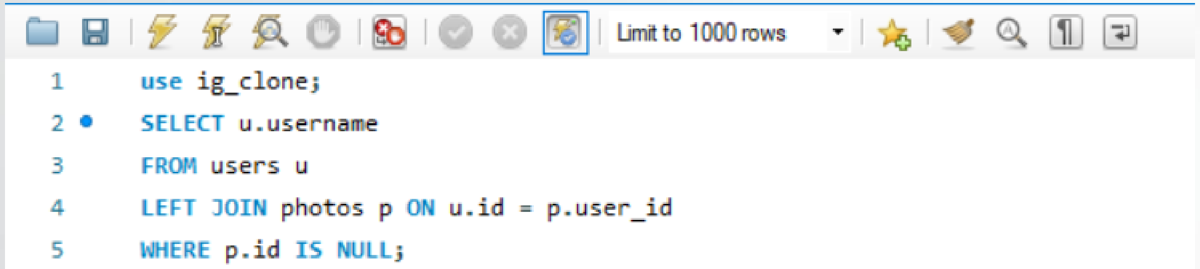


username	created_at
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41
Nicole71	2016-05-09 17:30:22
Jordyn.Jacobson2	2016-05-14 07:56:26

2. Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

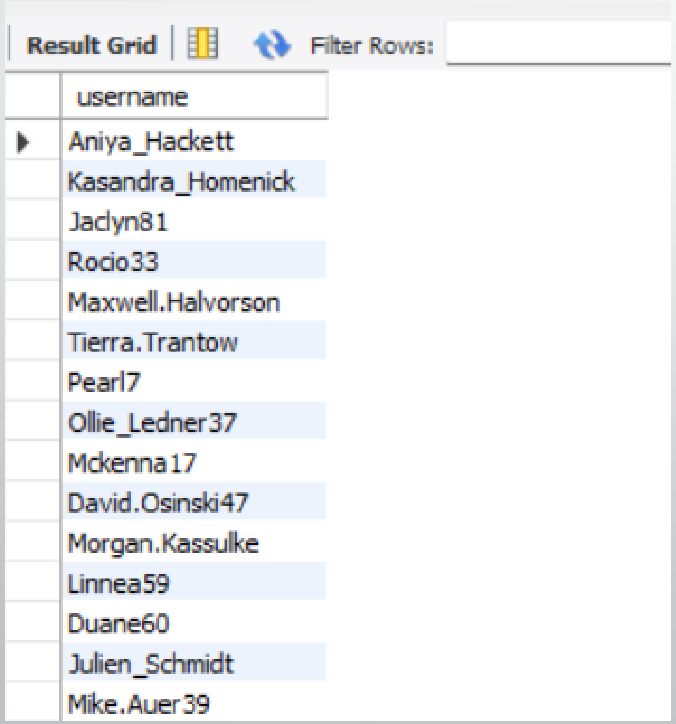
Task: Identify users who have never posted a single photo on Instagram.

Query:



```sql
use ig_clone;
SELECT u.username
FROM users u
LEFT JOIN photos p ON u.id = p.user_id
WHERE p.id IS NULL;
```

Output:

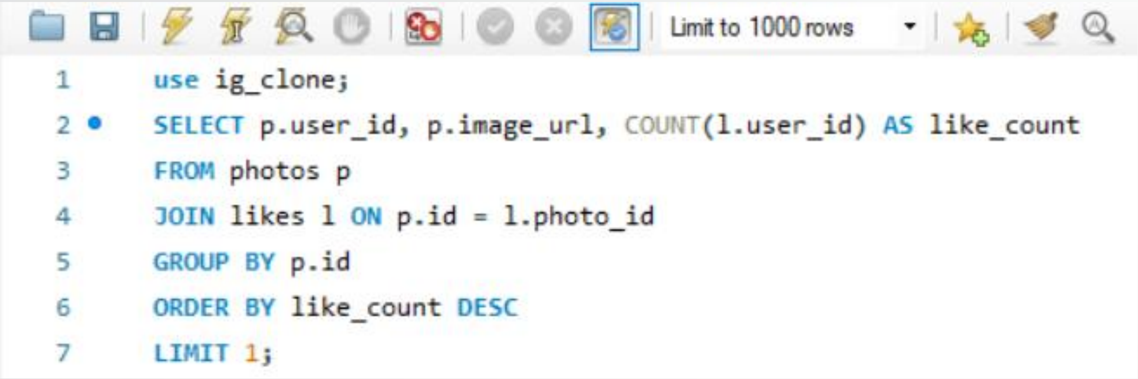


username
Aniya_Hackett
Kasandra_Homenick
Jaclyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39

3. Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.

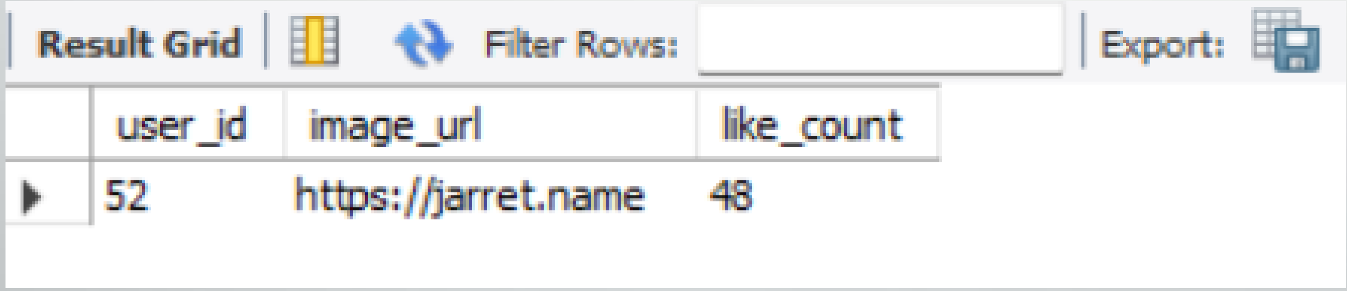
Task: Determine the winner of the contest and provide their details to the team.

Query:



```sql
use ig_clone;
SELECT p.user_id, p.image_url, COUNT(l.user_id) AS like_count
FROM photos p
JOIN likes l ON p.id = l.photo_id
GROUP BY p.id
ORDER BY like_count DESC
LIMIT 1;
```

Output:



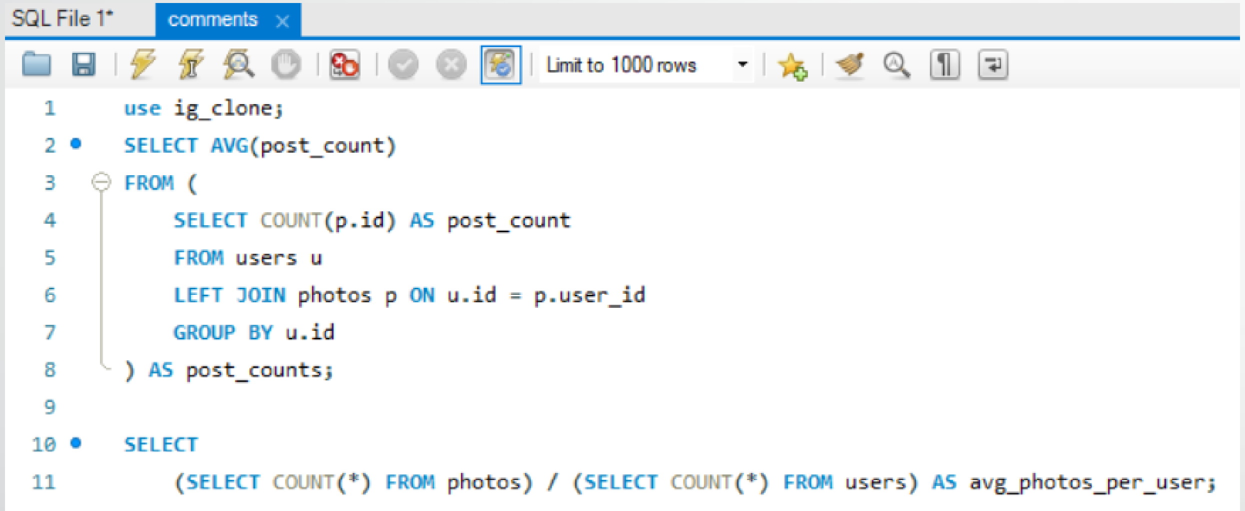
user_id	image_url	like_count
52	https://jarret.name	48

B) Investor Metrics:

1.User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

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

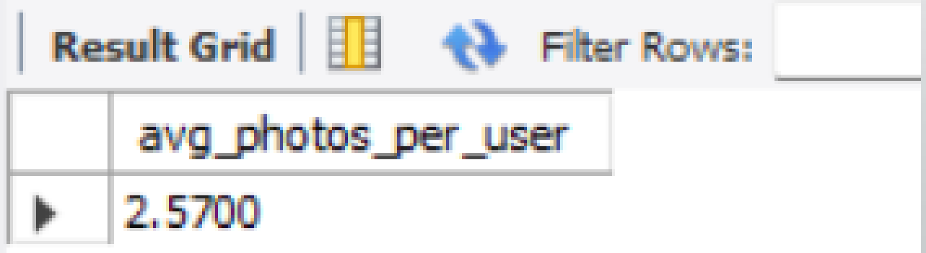
Query:



```sql
use ig_clone;
SELECT AVG(post_count)
FROM (
    SELECT COUNT(p.id) AS post_count
    FROM users u
    LEFT JOIN photos p ON u.id = p.user_id
    GROUP BY u.id
) AS post_counts;

SELECT
    (SELECT COUNT(*) FROM photos) / (SELECT COUNT(*) FROM users) AS avg_photos_per_user;
```

Output:

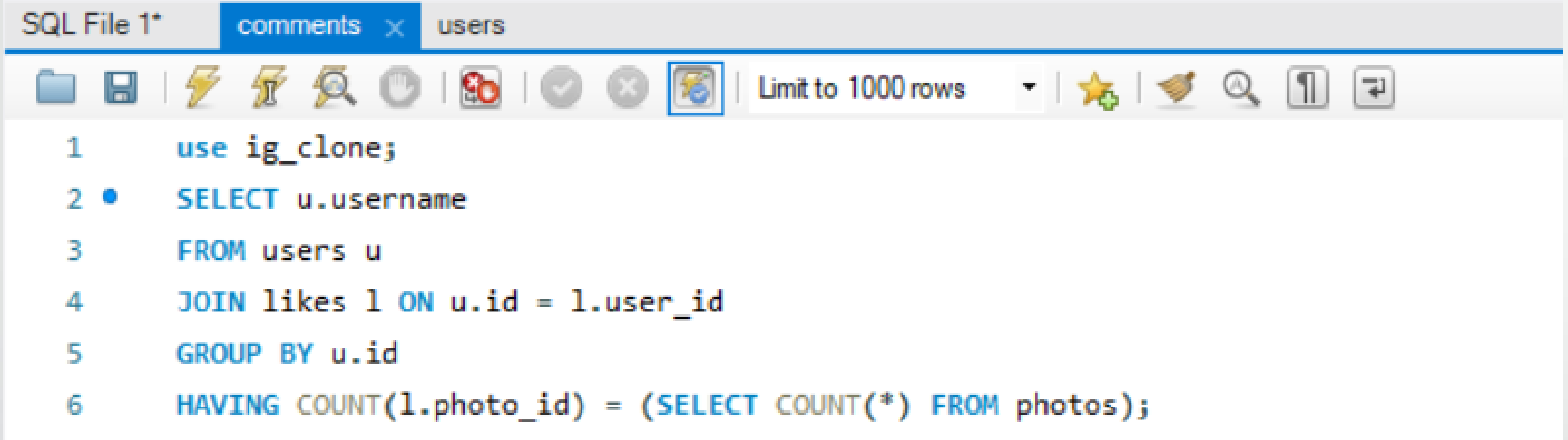


avg_photos_per_user
2.5700

2. Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

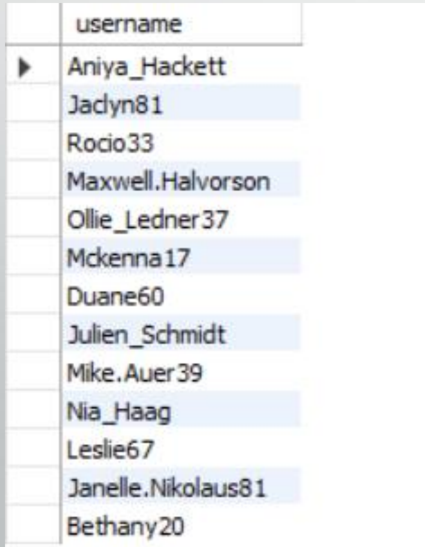
Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

Query:



```sql
use ig_clone;
SELECT u.username
FROM users u
JOIN likes l ON u.id = l.user_id
GROUP BY u.id
HAVING COUNT(l.photo_id) = (SELECT COUNT(*) FROM photos);
```

Output:



username
Aniya_Hackett
Jaclyn81
Rocio33
Maxwell.Halvorson
Ollie_Ledner37
Mckenna17
Duane60
Julien_Schmidt
Mike.Auer39
Nia_Haag
Leslie67
Janelle.Nikolaus81
Bethany20

Insights:

When all the data was synthesized, there were several significant findings that stood out:

- Users with the highest age have been located and most loyal users are also described using queries.
- A good proportion of the users never posted a photo; thus there is a possibility for engagement campaigns.
- The contest winner who got the highest number of likes thereby showing whom the audience liked most was located and five of the most frequently used tags were identified, a great asset for marketing purposes.

Results:

Through this project we are able make decisions and think in a way that would lead to describe a particular outcomes and Can easily find out solution to the problems by writing efficient queries.

Project : Operation Analytics and Investigating Metric Spike

trainity

Operation Analytics & Investagating metric spike case study

Description:

One of the key aspects of Operational Analytics is investigating metric spikes. This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales. The goal is to use your advanced SQL skills to analyze the data and provide valuable insights that can help improve the company's operations and understand sudden changes in key metrics.

Findings

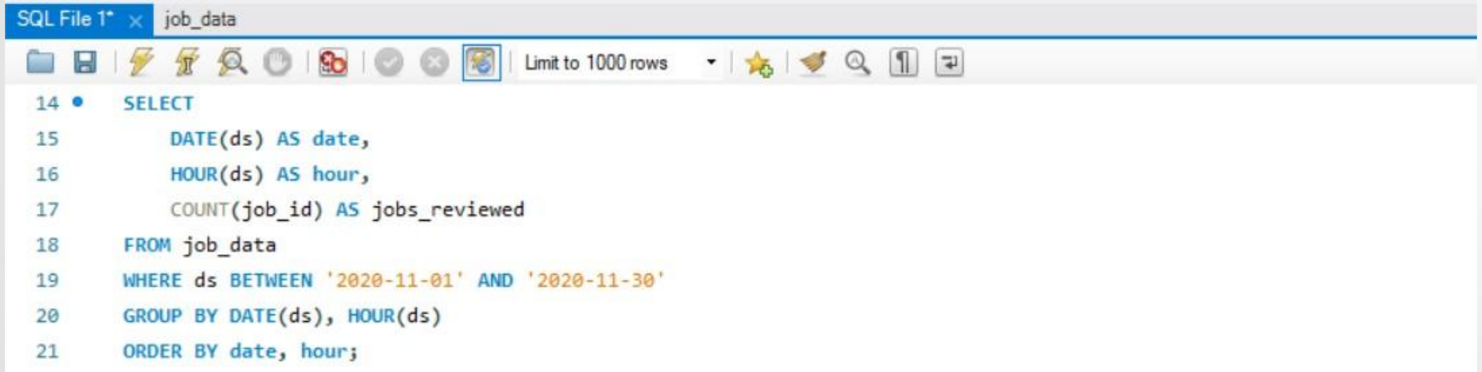
Case Study 1: Job Data Analysis

A.Jobs Reviewed Over Time:

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

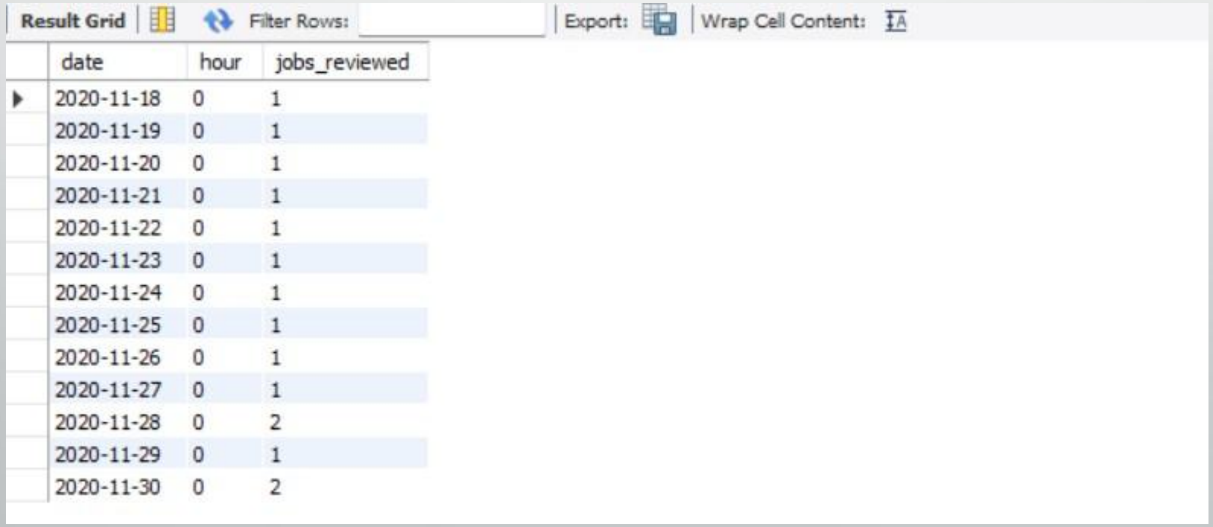
Task: Write an SQL query to calculate the number of jobs reviewed per hour for each day in November 2020.

Query:



```sql
SELECT
    DATE(ds) AS date,
    HOUR(ds) AS hour,
    COUNT(job_id) AS jobs_reviewed
FROM job_data
WHERE ds BETWEEN '2020-11-01' AND '2020-11-30'
GROUP BY DATE(ds), HOUR(ds)
ORDER BY date, hour;
```

Output:



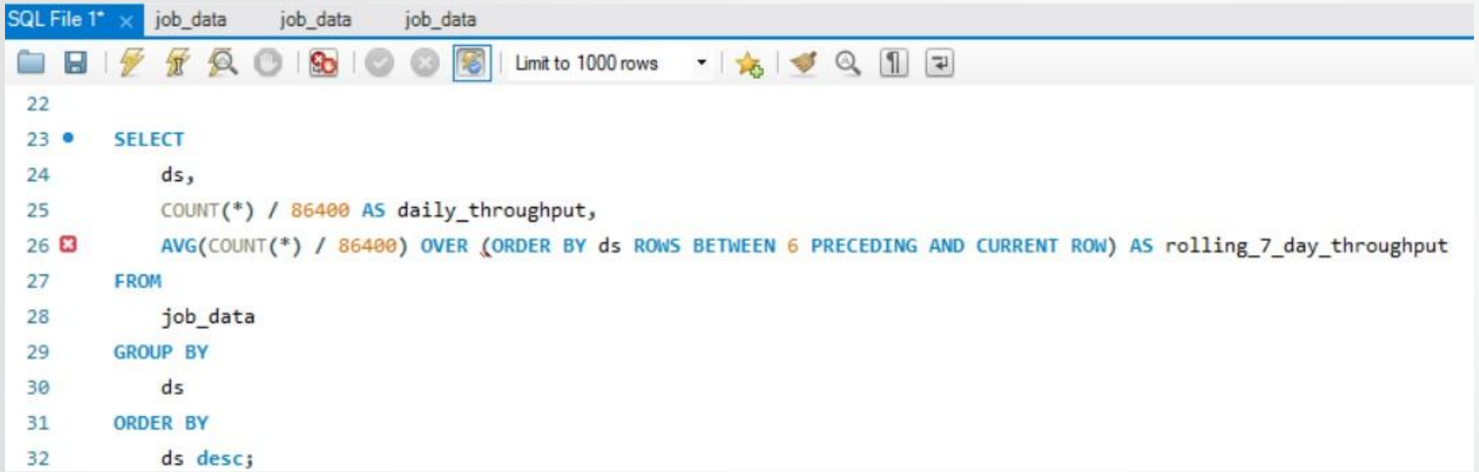
date	hour	jobs_reviewed
2020-11-18	0	1
2020-11-19	0	1
2020-11-20	0	1
2020-11-21	0	1
2020-11-22	0	1
2020-11-23	0	1
2020-11-24	0	1
2020-11-25	0	1
2020-11-26	0	1
2020-11-27	0	1
2020-11-28	0	2
2020-11-29	0	1
2020-11-30	0	2

B. Throughput Analysis:

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

Your Task: Write an SQL query to calculate the 7-day rolling average of throughput. Additionally, explain whether you prefer using the daily metric or the 7-day rolling average for throughput, and why.

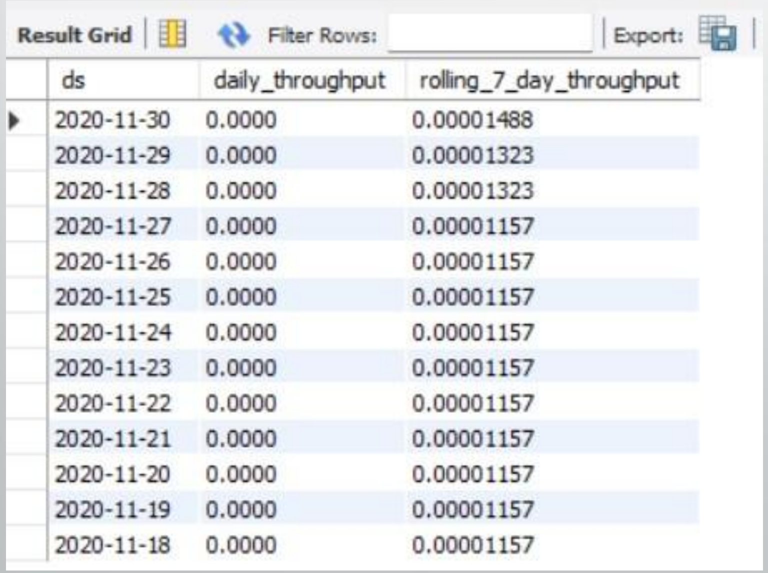
Query:



```sql

SELECT
    ds,
    COUNT(*) / 86400 AS daily_throughput,
    AVG(COUNT(*) / 86400) OVER (ORDER BY ds ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) AS rolling_7_day_throughput
FROM
    job_data
GROUP BY
    ds
ORDER BY
    ds desc;
```

Output:



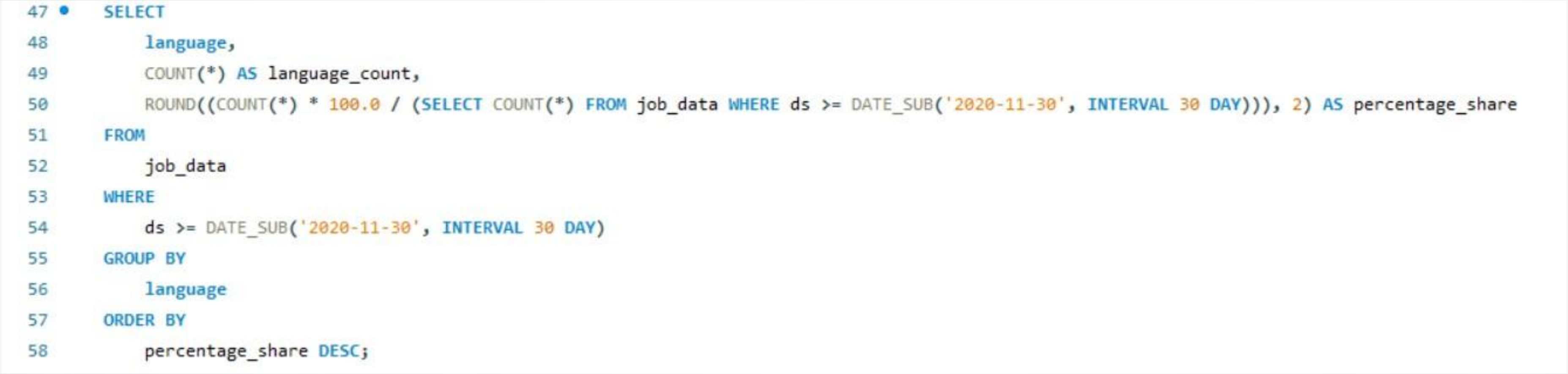
ds	daily_throughput	rolling_7_day_throughput
2020-11-30	0.0000	0.00001488
2020-11-29	0.0000	0.00001323
2020-11-28	0.0000	0.00001323
2020-11-27	0.0000	0.00001157
2020-11-26	0.0000	0.00001157
2020-11-25	0.0000	0.00001157
2020-11-24	0.0000	0.00001157
2020-11-23	0.0000	0.00001157
2020-11-22	0.0000	0.00001157
2020-11-21	0.0000	0.00001157
2020-11-20	0.0000	0.00001157
2020-11-19	0.0000	0.00001157
2020-11-18	0.0000	0.00001157

C. Language Share Analysis:

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

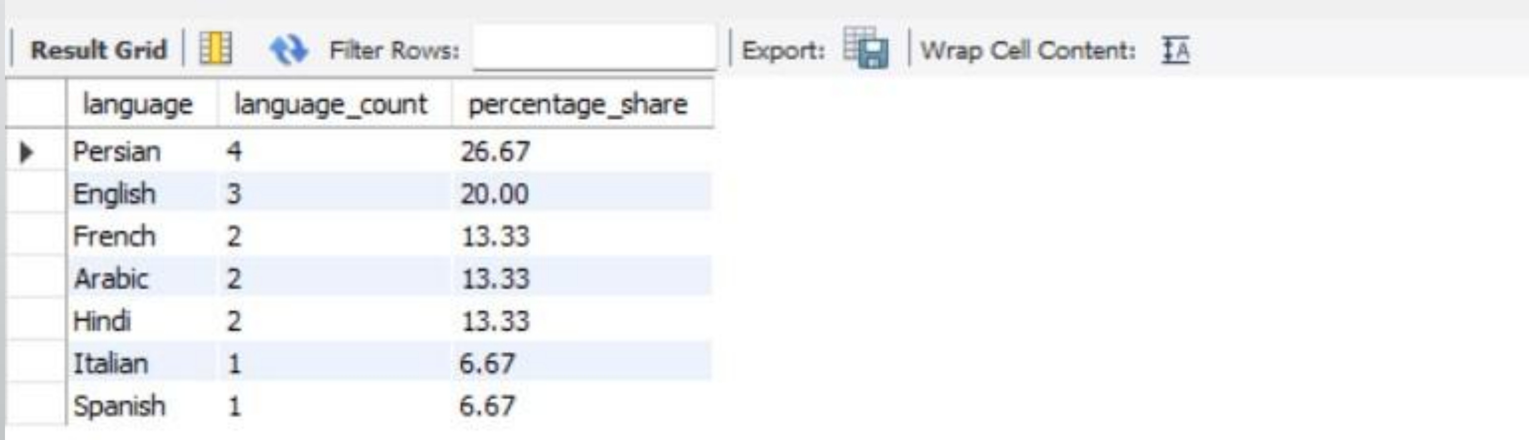
Your Task: Write an SQL query to calculate the percentage share of each language over the last 30 days.

Query:



```
SELECT
    language,
    COUNT(*) AS language_count,
    ROUND((COUNT(*) * 100.0 / (SELECT COUNT(*) FROM job_data WHERE ds >= DATE_SUB('2020-11-30', INTERVAL 30 DAY))), 2) AS percentage_share
FROM
    job_data
WHERE
    ds >= DATE_SUB('2020-11-30', INTERVAL 30 DAY)
GROUP BY
    language
ORDER BY
    percentage_share DESC;
```

Output:



language	language_count	percentage_share
Persian	4	26.67
English	3	20.00
French	2	13.33
Arabic	2	13.33
Hindi	2	13.33
Italian	1	6.67
Spanish	1	6.67

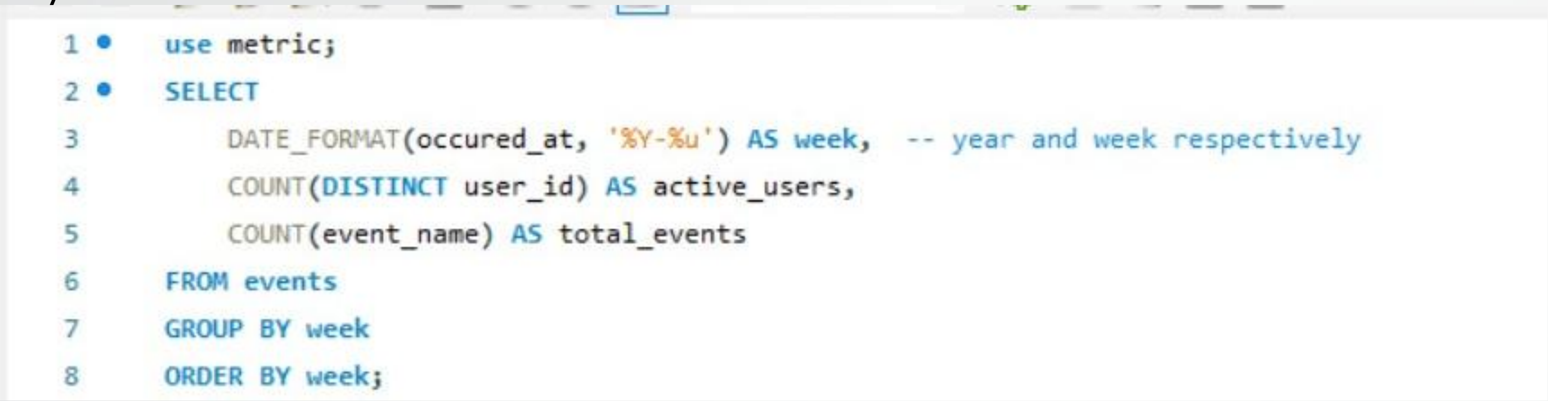
Case Study 2: Investigating Metric Spike

A.Weekly User Engagement:

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

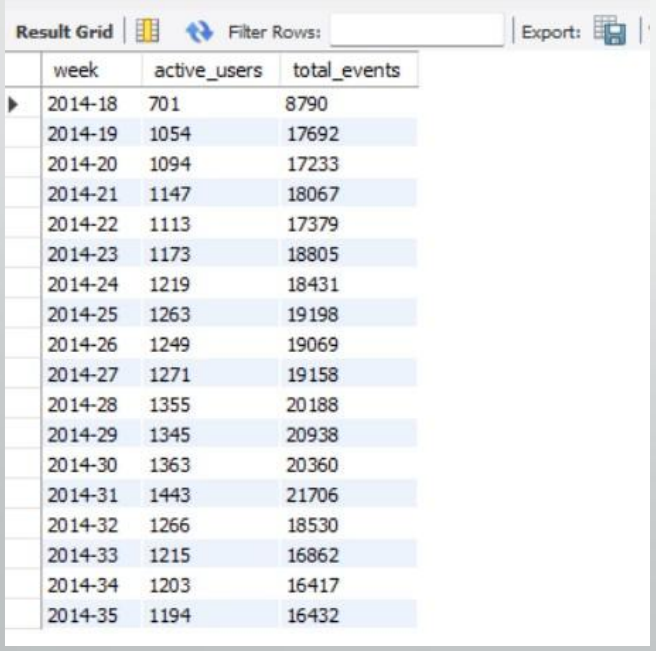
B. Your Task: Write an SQL query to calculate the weekly user engagement.

Query:



```sql
use metric;
SELECT
    DATE_FORMAT(occured_at, '%Y-%u') AS week,  -- year and week respectively
    COUNT(DISTINCT user_id) AS active_users,
    COUNT(event_name) AS total_events
FROM events
GROUP BY week
ORDER BY week;
```

Output:

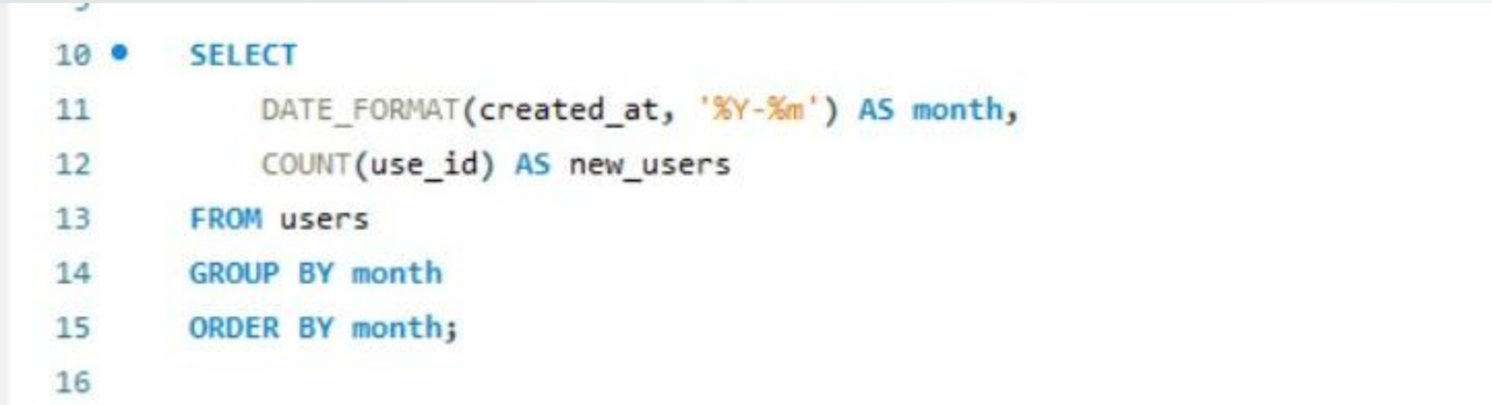


week	active_users	total_events
2014-18	701	8790
2014-19	1054	17692
2014-20	1094	17233
2014-21	1147	18067
2014-22	1113	17379
2014-23	1173	18805
2014-24	1219	18431
2014-25	1263	19198
2014-26	1249	19069
2014-27	1271	19158
2014-28	1355	20188
2014-29	1345	20938
2014-30	1363	20360
2014-31	1443	21706
2014-32	1266	18530
2014-33	1215	16862
2014-34	1203	16417
2014-35	1194	16432

B. User Growth Analysis:

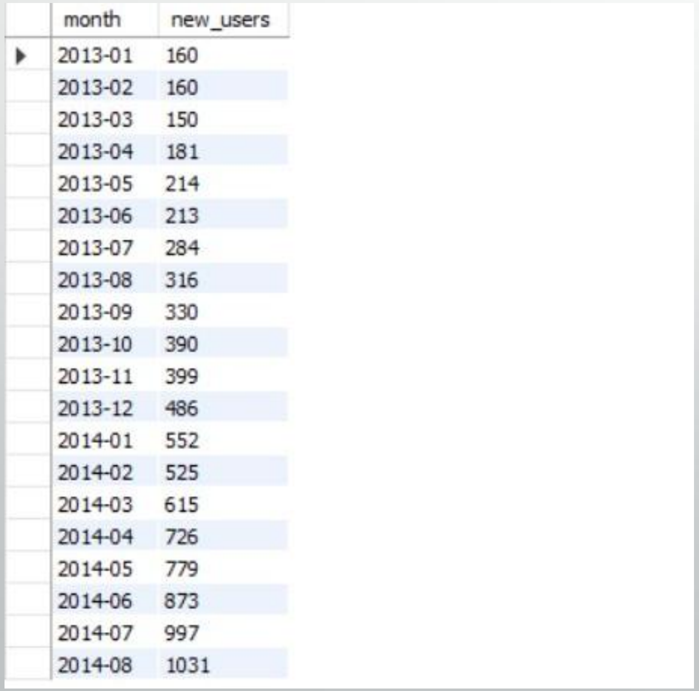
A. Objective: Analyze the growth of users over time for a product.
B. Your Task: Write an SQL query to calculate the user growth for the product.

Query:



```
SELECT
    DATE_FORMAT(created_at, '%Y-%m') AS month,
    COUNT(use_id) AS new_users
FROM users
GROUP BY month
ORDER BY month;
```

Output:



month	new_users
2013-01	160
2013-02	160
2013-03	150
2013-04	181
2013-05	214
2013-06	213
2013-07	284
2013-08	316
2013-09	330
2013-10	390
2013-11	399
2013-12	486
2014-01	552
2014-02	525
2014-03	615
2014-04	726
2014-05	779
2014-06	873
2014-07	997
2014-08	1031

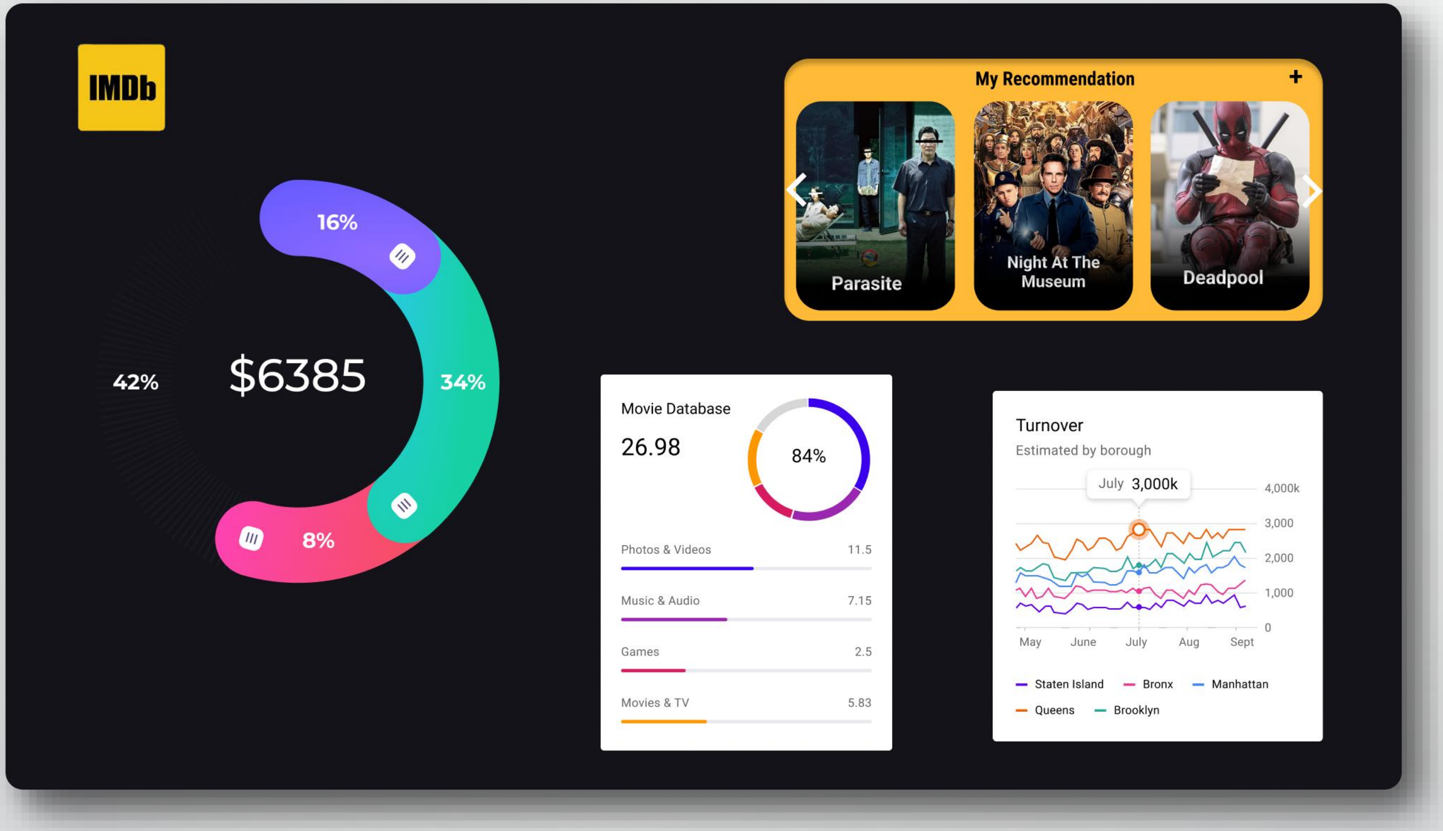
Insights:

- The project generally focuses on **Operational Analytics** and **Investigating Metric Spikes**, using user data to analyze trends in engagement and marketing metrics. Key activities included **job data analysis**, detecting throughput trends, and language usage.
- Tasks involves initially database creation, import the CSV files, and building efficient SQL queries to solve problem statements. Analysis were performed for metrics like weekly user engagement, retention, and device preferences.
- Finding highlighted trends such as peak job review times, throughput averages, language usage patterns, and user engagement by device. These insights help in optimizing resources, understanding growth trends, and improving retention strategies.

Result:

The project analyzed job data and user engagement to uncover trends like peak activity times, language preferences, and user retention. Efficient SQL queries enabled problem identification and actionable insights for resource optimization. Using MySQL Workbench, the results improved decision-making and enhanced strategic planning.

Project : IMDB Movie Analysis



Description:

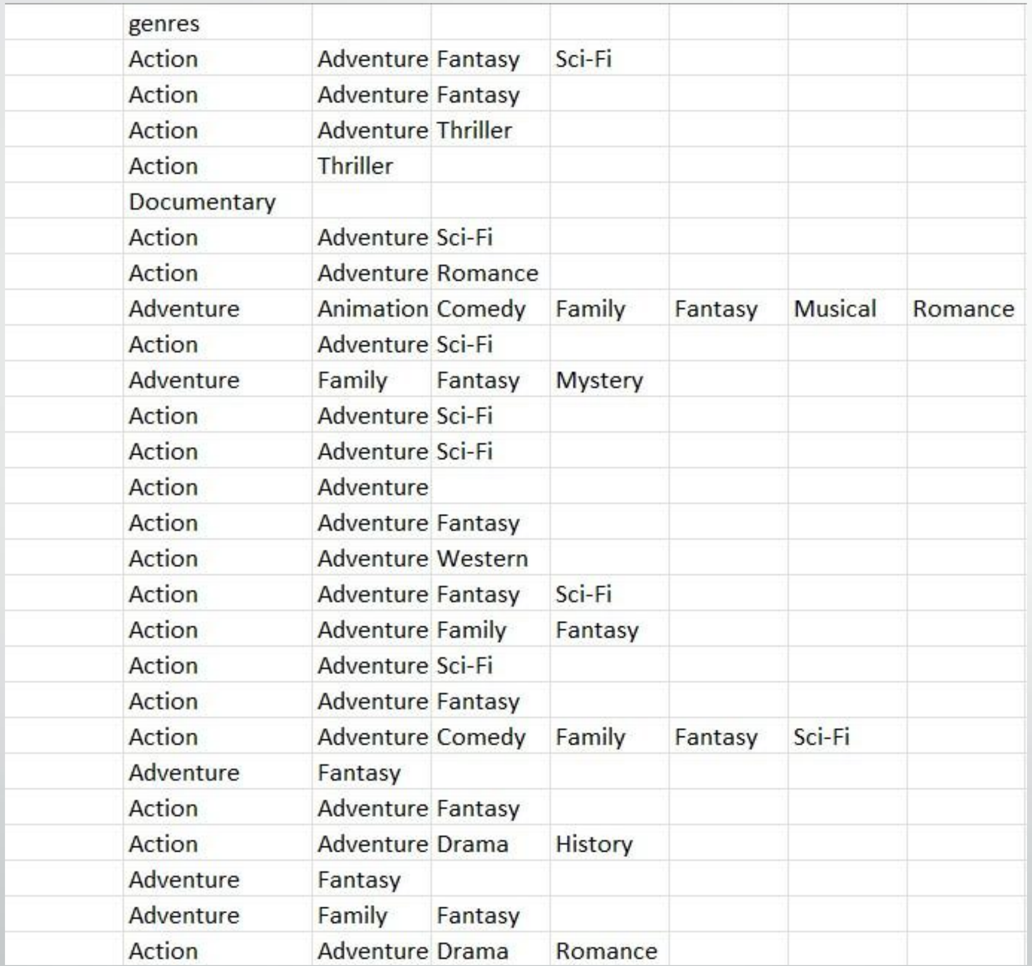
The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. Consider the correlation between movie ratings and other factors like genre, director, budget, etc. You might also want to consider the year of release, the actors involved, and other relevant factors.

Excel sheet analysis : Click Here

Findings

A. **Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.

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

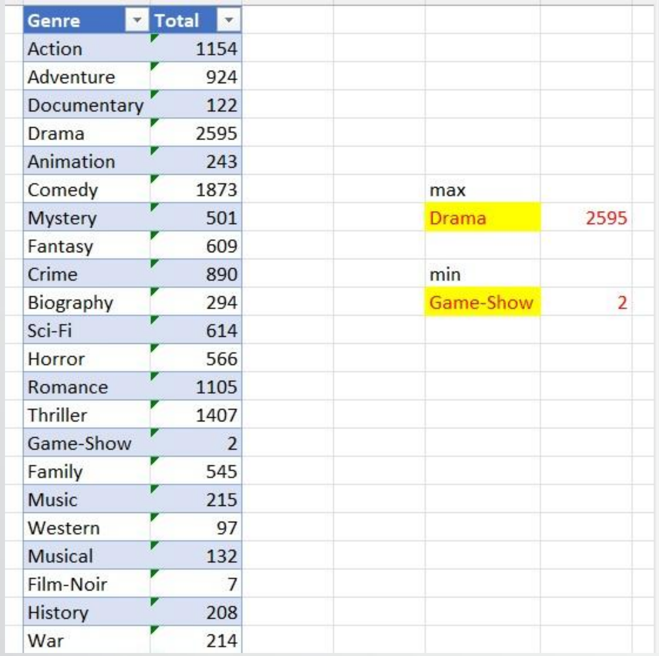


genres						
Action	Adventure	Fantasy	Sci-Fi			
Action	Adventure	Fantasy				
Action	Adventure	Thriller				
Action	Thriller					
Documentary						
Action	Adventure	Sci-Fi				
Action	Adventure	Romance				
Adventure	Animation	Comedy	Family	Fantasy	Musical	Romance
Action	Adventure	Sci-Fi				
Adventure	Family	Fantasy	Mystery			
Action	Adventure	Sci-Fi				
Action	Adventure	Sci-Fi				
Action	Adventure					
Action	Adventure	Fantasy				
Action	Adventure	Western				
Action	Adventure	Fantasy	Sci-Fi			
Action	Adventure	Family	Fantasy			
Action	Adventure	Sci-Fi				
Action	Adventure	Fantasy				
Action	Adventure	Comedy	Family	Fantasy	Sci-Fi	
Adventure	Fantasy					
Action	Adventure	Fantasy				
Action	Adventure	Drama	History			
Adventure	Fantasy					
Adventure	Family	Fantasy				
Action	Adventure	Drama	Romance			

Separation of genres by using Text to Columns > Delimited .

Finding Total number of movies for each genre. Top Genres are:

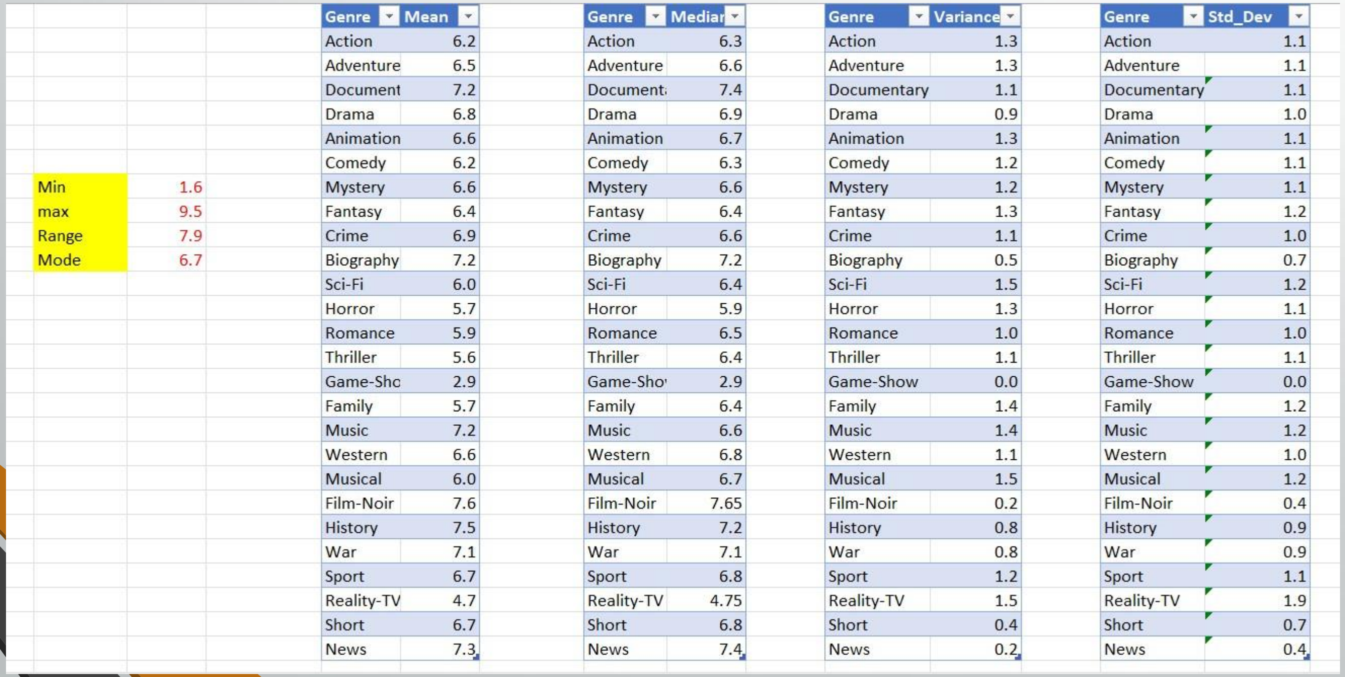
- Drama
- Comedy
- Thriller
- Action
- Romance



Genre	Total
Action	1154
Adventure	924
Documentary	122
Drama	2595
Animation	243
Comedy	1873
Mystery	501
Fantasy	609
Crime	890
Biography	294
Sci-Fi	614
Horror	566
Romance	1105
Thriller	1407
Game-Show	2
Family	545
Music	215
Western	97
Musical	132
Film-Noir	7
History	208
War	214

max	
Drama	2595
min	
Game-Show	2

Calculation of Discrete Statistics based on genres to IMDB Scores. Min, Max, Range, Mode based on Average IMDB Scores.



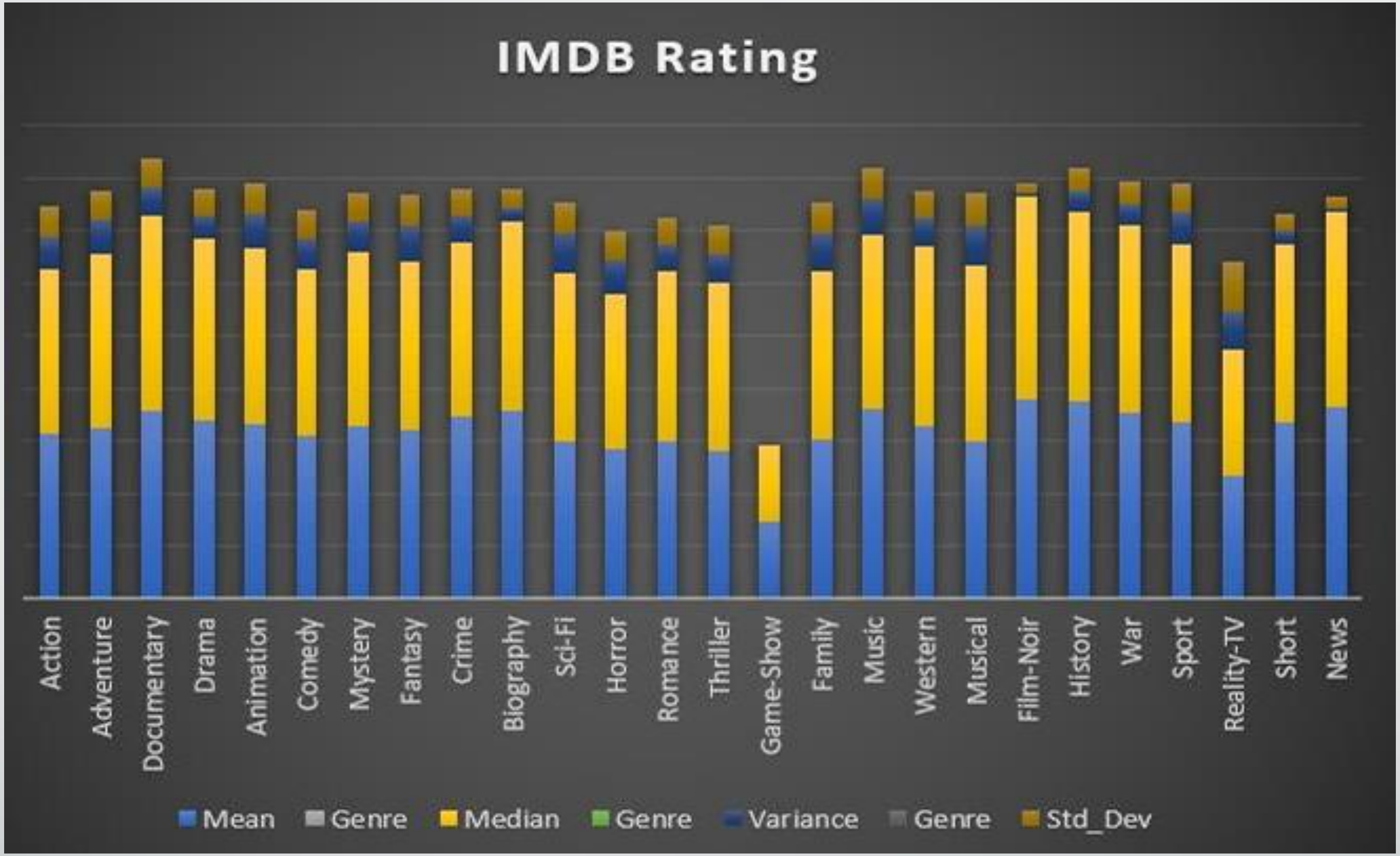
Min	1.6
max	9.5
Range	7.9
Mode	6.7

Genre	Mean
Action	6.2
Adventure	6.5
Document	7.2
Drama	6.8
Animation	6.6
Comedy	6.2
Mystery	6.6
Fantasy	6.4
Crime	6.9
Biography	7.2
Sci-Fi	6.0
Horror	5.7
Romance	5.9
Thriller	5.6
Game-Shc	2.9
Family	5.7
Music	7.2
Western	6.6
Musical	6.0
Film-Noir	7.6
History	7.5
War	7.1
Sport	6.7
Reality-TV	4.7
Short	6.7
News	7.3

Genre	Mediar
Action	6.3
Adventure	6.6
Document	7.4
Drama	6.9
Animation	6.7
Comedy	6.3
Mystery	6.6
Fantasy	6.4
Crime	6.6
Biography	7.2
Sci-Fi	6.4
Horror	5.9
Romance	6.5
Thriller	6.4
Game-Sho	2.9
Family	6.4
Music	6.6
Western	6.8
Musical	6.7
Film-Noir	7.65
History	7.2
War	7.1
Sport	6.8
Reality-TV	4.75
Short	6.8
News	7.4

Genre	Variance
Action	1.3
Adventure	1.3
Documentary	1.1
Drama	0.9
Animation	1.3
Comedy	1.2
Mystery	1.2
Fantasy	1.3
Crime	1.1
Biography	0.5
Sci-Fi	1.5
Horror	1.3
Romance	1.0
Thriller	1.1
Game-Show	0.0
Family	1.4
Music	1.4
Western	1.1
Musical	1.5
Film-Noir	0.2
History	0.8
War	0.8
Sport	1.2
Reality-TV	1.5
Short	0.4
News	0.2

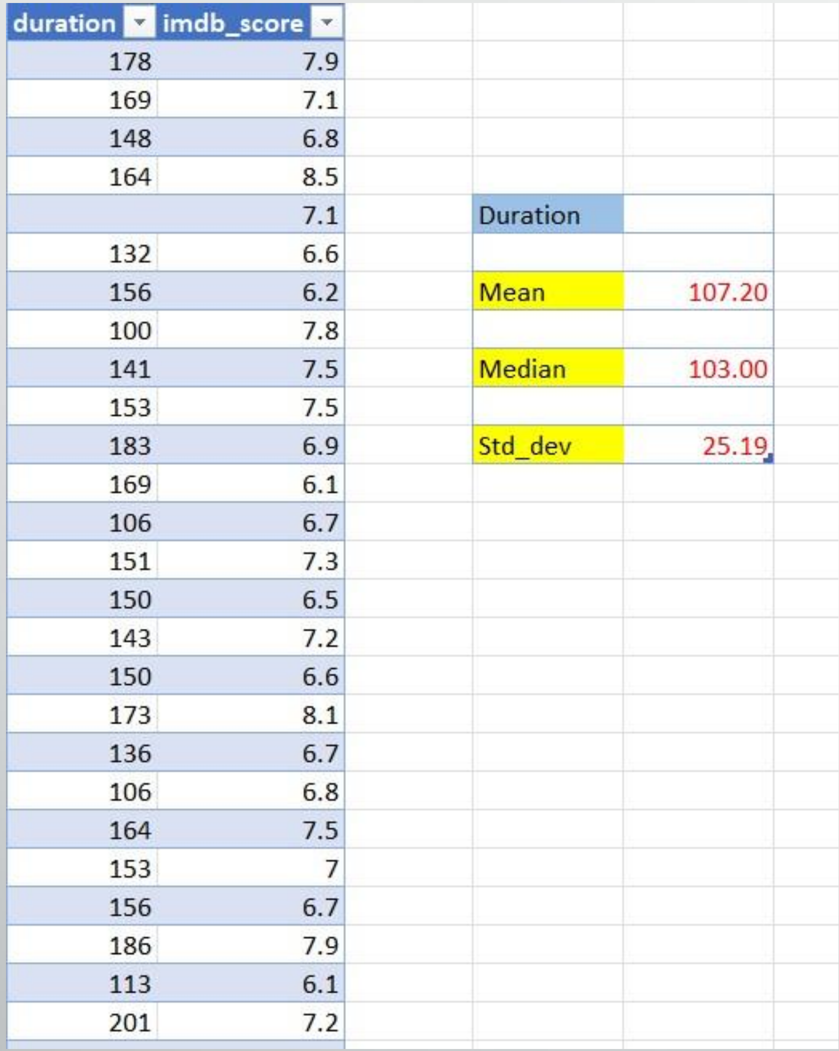
Genre	Std_Dev
Action	1.1
Adventure	1.1
Documentary	1.1
Drama	1.0
Animation	1.1
Comedy	1.1
Mystery	1.1
Fantasy	1.2
Crime	1.0
Biography	0.7
Sci-Fi	1.2
Horror	1.1
Romance	1.0
Thriller	1.1
Game-Show	0.0
Family	1.2
Music	1.2
Western	1.0
Musical	1.2
Film-Noir	0.4
History	0.9
War	0.9
Sport	1.1
Reality-TV	1.9
Short	0.7
News	0.4



Here you can see IMDB Scores of specific genre with discrete statistics like : Mean, Median, Variance and Standard Deviation.

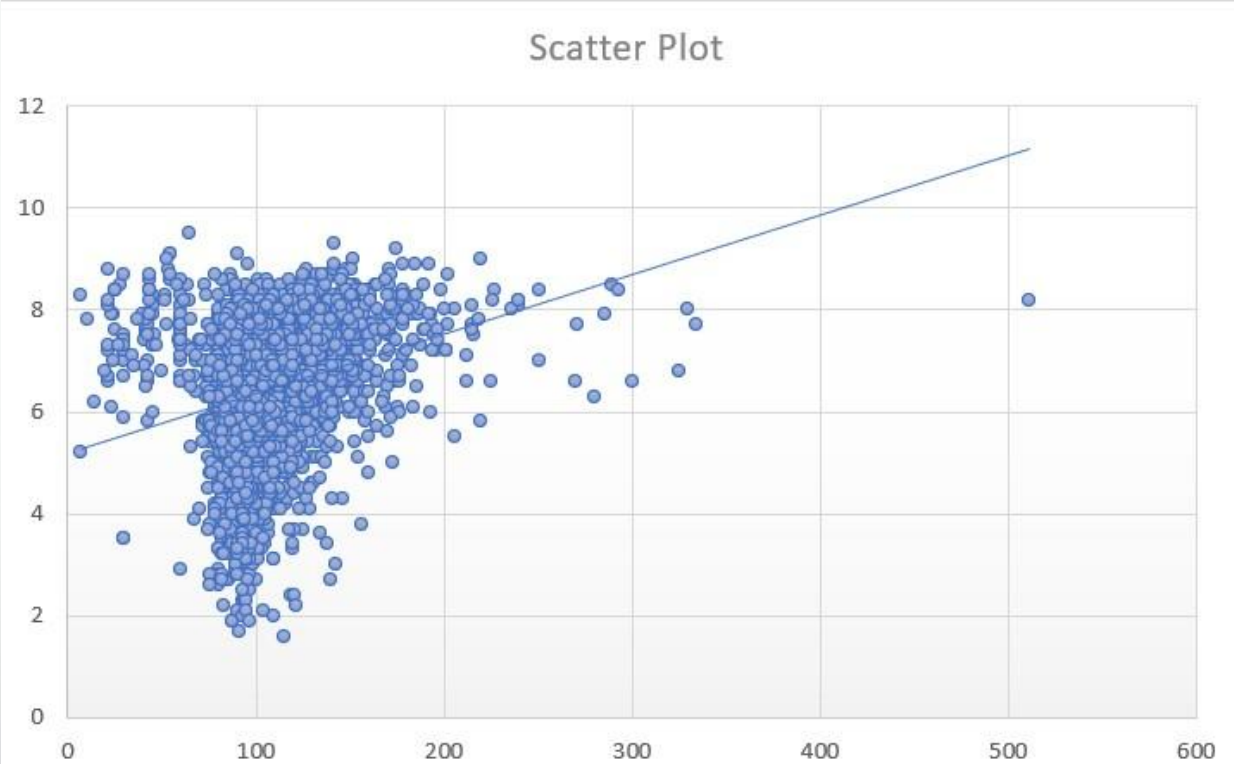
B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

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



duration	imdb_score
178	7.9
169	7.1
148	6.8
164	8.5
	7.1
132	6.6
156	6.2
100	7.8
141	7.5
153	7.5
183	6.9
169	6.1
106	6.7
151	7.3
150	6.5
143	7.2
150	6.6
173	8.1
136	6.7
106	6.8
164	7.5
153	7
156	6.7
186	7.9
113	6.1
201	7.2

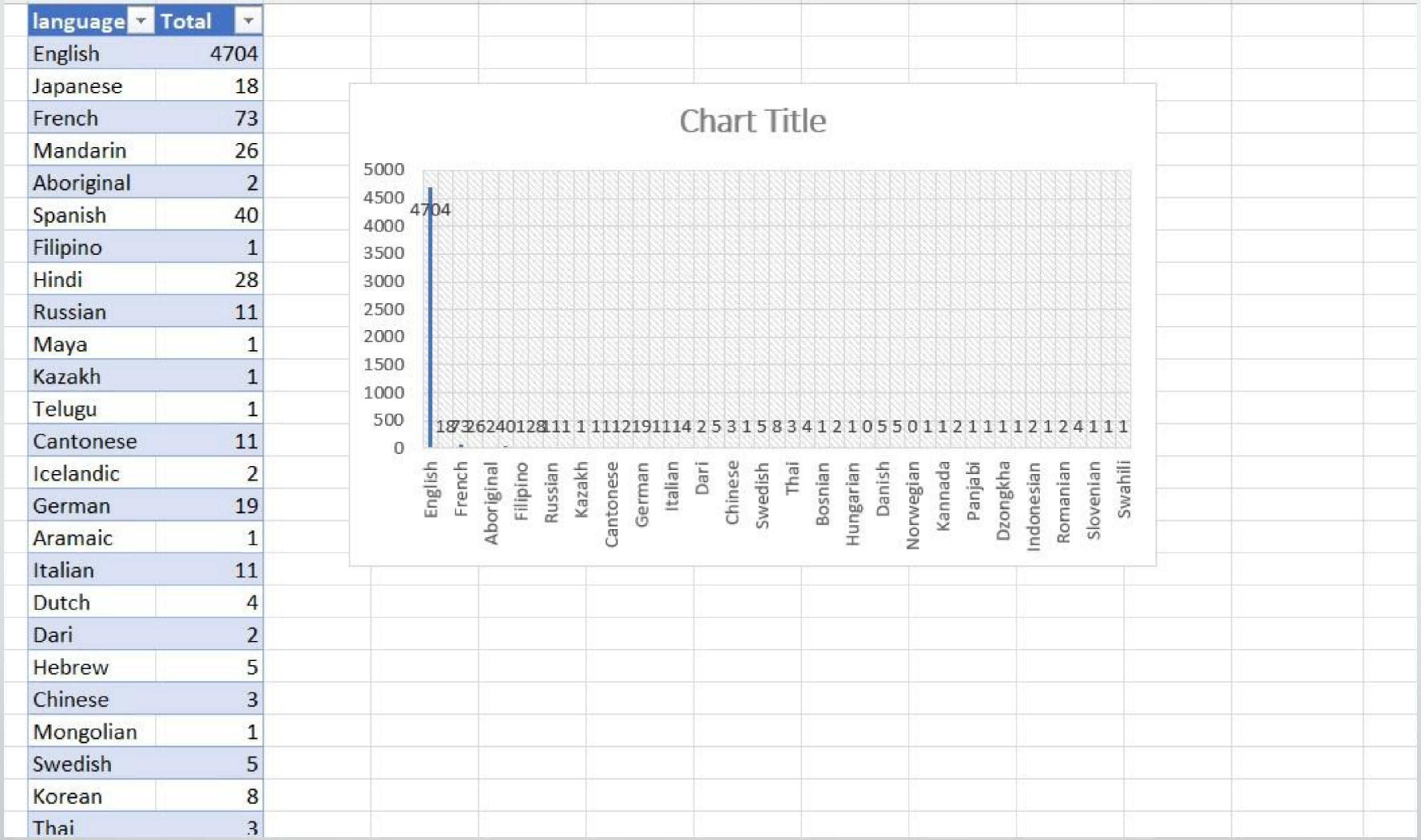
Duration	
Mean	107.20
Median	103.00
Std_dev	25.19



Calculation of Discrete Statistics based on Duration to IMDB Scores. Mean, Median, Standard Deviation based on IMDB Scores.

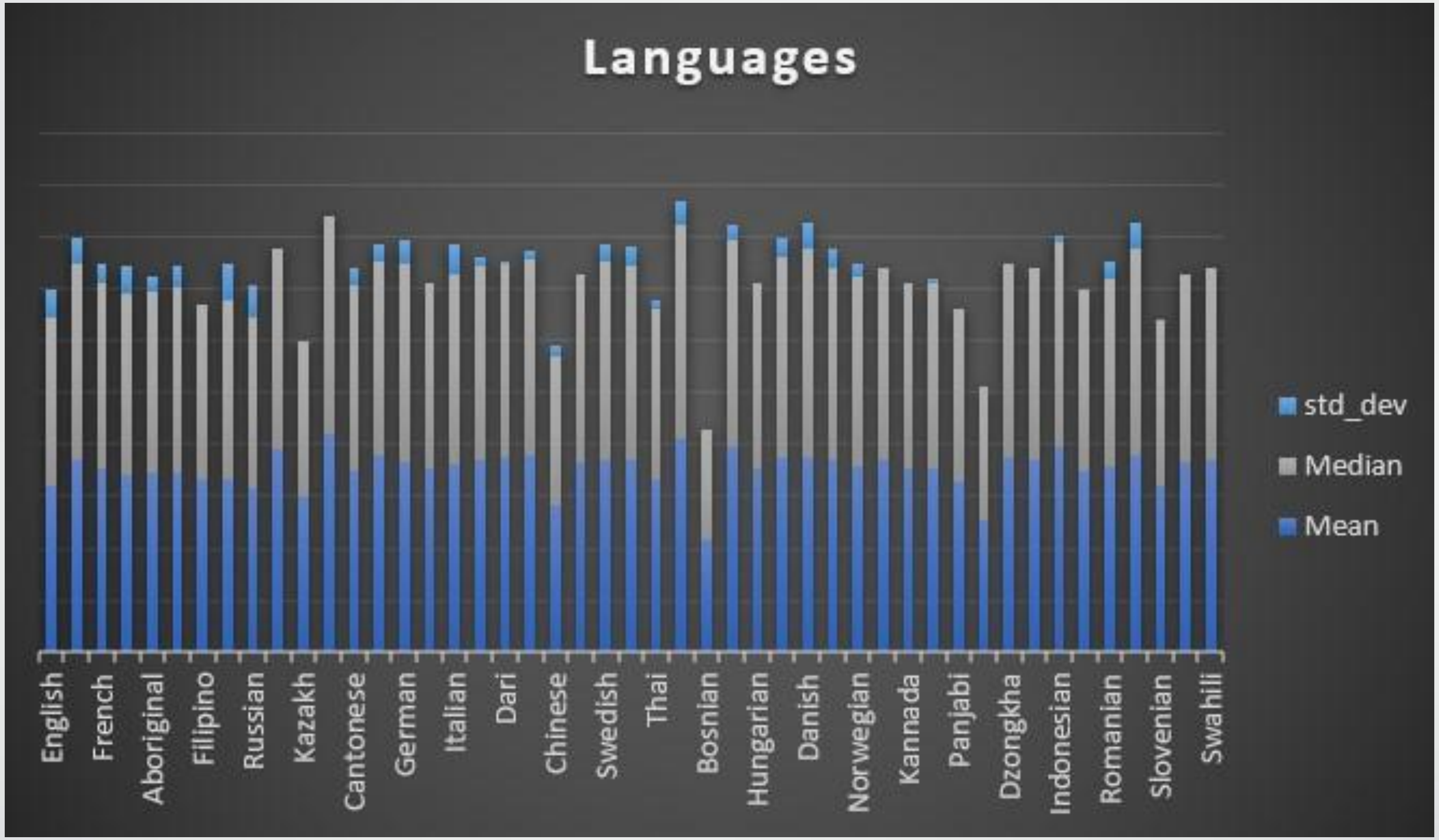
D. Director Analysis: Influence of directors on movie ratings.

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



language	Total
English	4704
Japanese	18
French	73
Mandarin	26
Aboriginal	2
Spanish	40
Filipino	1
Hindi	28
Russian	11
Maya	1
Kazakh	1
Telugu	1
Cantonese	11
Icelandic	2
German	19
Aramaic	1
Italian	11
Dutch	4
Dari	2
Hebrew	5
Chinese	3
Mongolian	1
Swedish	5
Korean	8
Thai	3

Total number of languages are determined by using COUNTIF function in Excel. For Insight the data is visualized by a chart.



Calculation of Discrete Statistics based on Languages to IMDB Scores.
Mean, Median, Standard Deviation based on IMDB Scores.

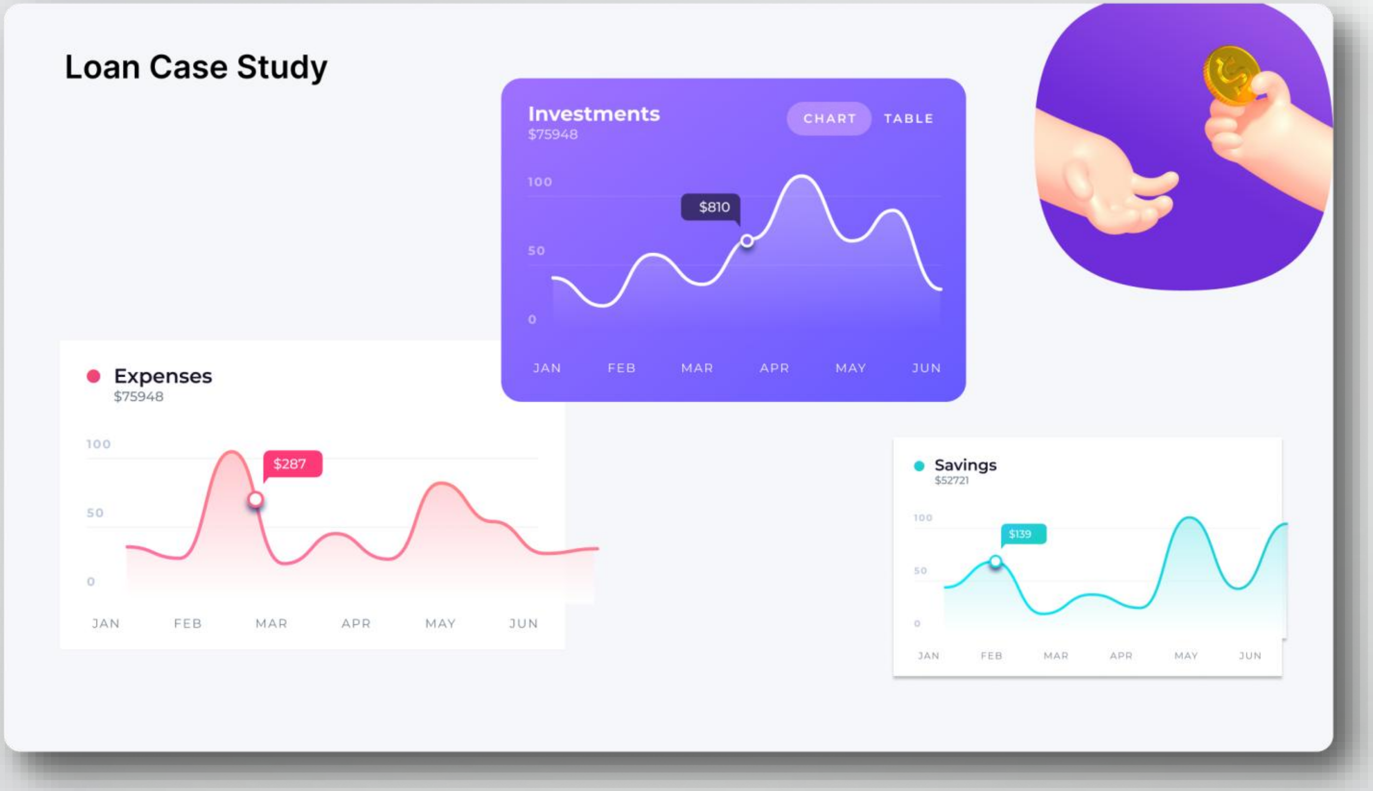
Insights:

Drama, Comedy, Thriller, Action, and Romance emerged as the most popular genres, with significant impacts on IMDB scores observed through statistical analysis. Movie durations showed a noticeable correlation with ratings, as longer films often exhibited more varied scores, visualized using scatter plots. Language analysis highlighted English as the leading language, with statistical insights revealing its dominance and global audience appeal. Directors played a critical role in influencing IMDB scores, with top-performing directors identified through percentile-based analysis. Budget analysis demonstrated that higher budgets often lead to greater profitability, supported by correlation findings and maximum profit evaluations.

Results:

- Key patterns were identified for genres, durations, languages, directors, budgets, and their corresponding IMDB scores.
- The analysis provided actionable insights to make decisions and solve problems, demonstrating the efficacy of Excel functions for statistical analysis and visualization.

Project : Bank Loan Case Study



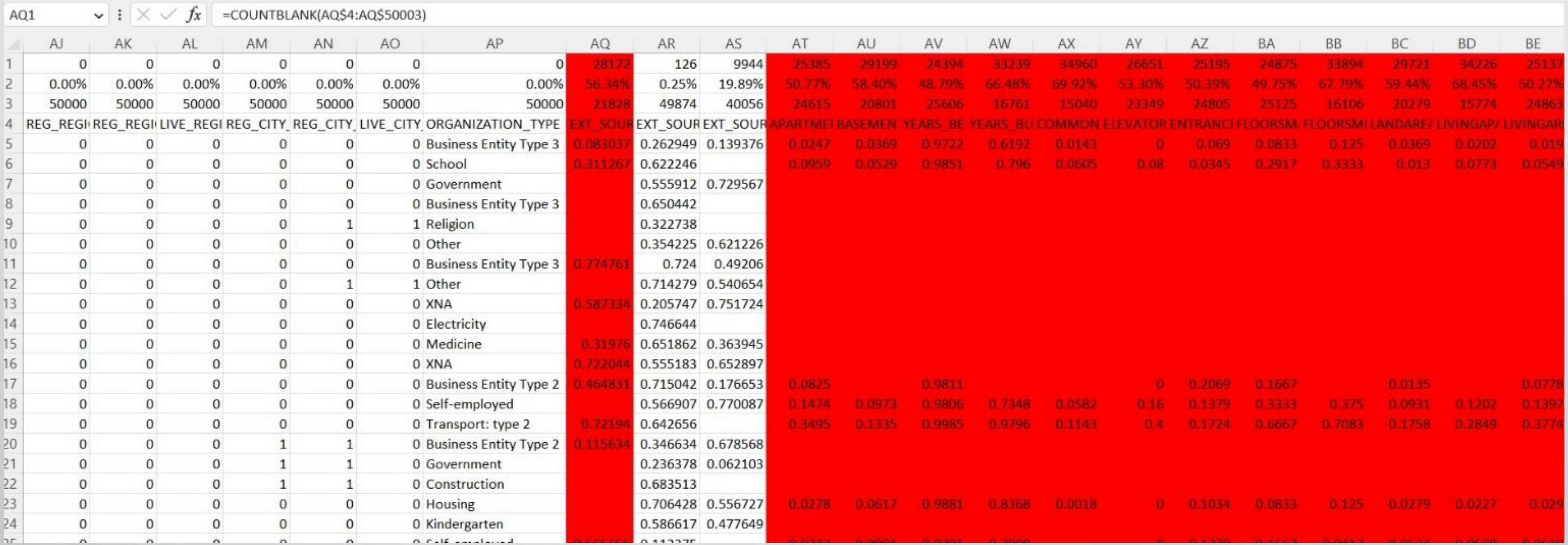
Description:

A finance company that specializes in lending various types of loans to urban customers. Company faces a challenge some customers who don't have a sufficient credit history take advantage of this and default on their loans. The main aim of this project is to identify patterns that indicate if a customer will have difficulty paying their installments. This information can be used to make decisions such as denying the loan, reducing the amount of loan, or lending at a higher interest rate to risky applicants. The task is to use Exploratory Data Analysis (EDA) to analyze patterns in the data and ensure that capable applicants are not rejected.

Excel sheet analysis : Click Here

Findings

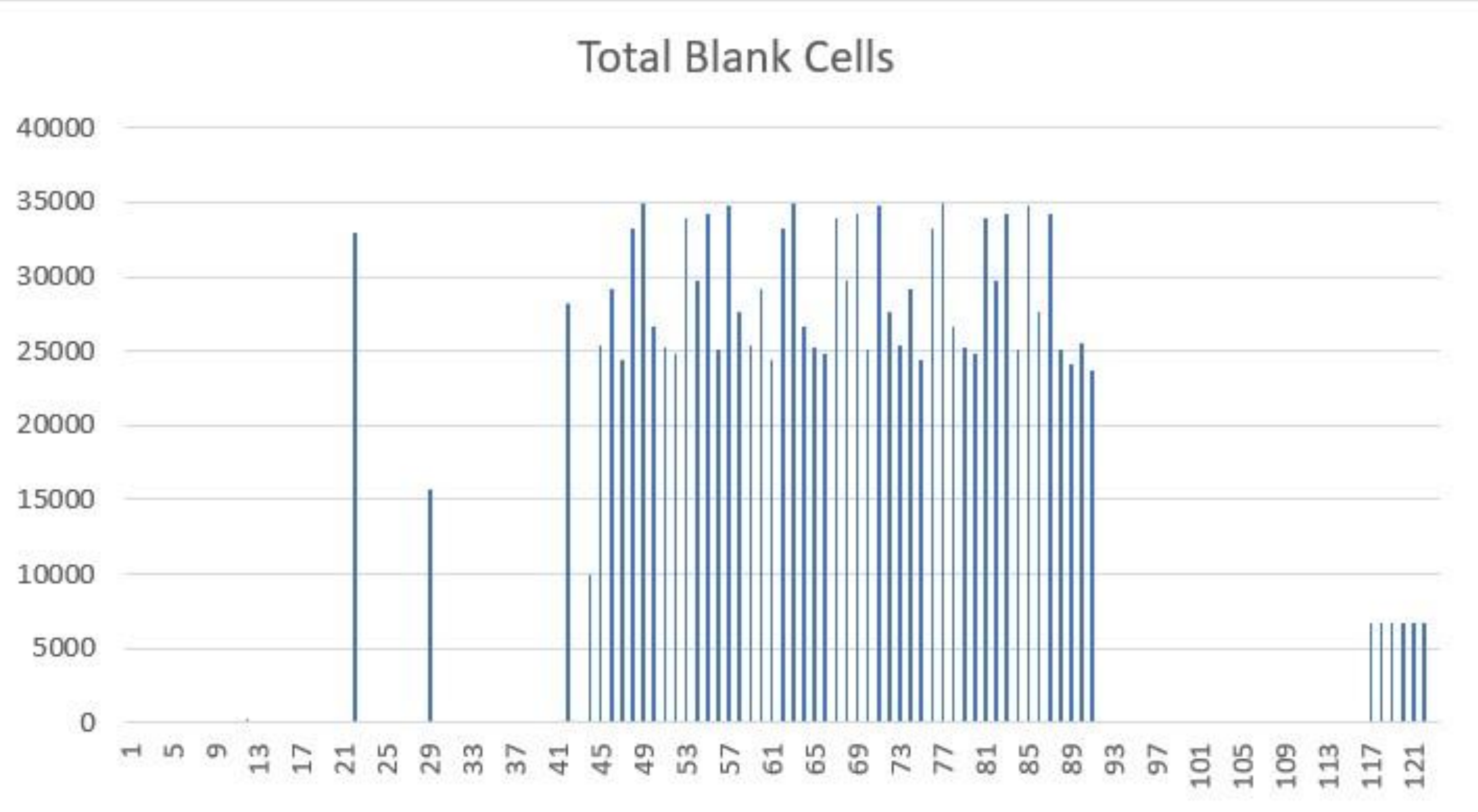
A. **Identify Missing Data and Deal with it Appropriately:** As a data analyst, you come across missing data in the loan application dataset. It is essential to handle missing data effectively to ensure the accuracy of the analysis. **Task:** Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.



AQ1 =COUNTBLANK(AQ$4:AQ$50003)

	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE
1	0	0	0	0	0	0	0	28172	126	9944	25385	29199	24394	33239	34960	26651	25195	24875	33894	29721	34226	25137
2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	56.34%	0.25%	19.89%	50.77%	58.40%	48.79%	66.48%	69.92%	53.30%	50.39%	49.75%	67.79%	59.44%	68.45%	50.27%
3	50000	50000	50000	50000	50000	50000	50000	21828	49874	40056	24615	20801	25606	16761	15040	23349	24805	25125	16106	20279	15774	24863
4	REG_REGI	REG_REGI	LIVE_REGI	REG_CITY_	REG_CITY_	LIVE_CITY	ORGANIZATION_TYPE	EXT_SOUR	EXT_SOUR	EXT_SOUR	APARTMEI	BASEMEN	YEARS_BE	YEARS_BU	COMMON	ELEVATOR	ENTRANCI	FLOORSM.	FLOORSMI	LANDARE/	LIVINGAP/	LIVINGARI
5	0	0	0	0	0	0	Business Entity Type 3	0.083037	0.262949	0.139376	0.0247	0.0369	0.9722	0.6192	0.0143	0	0.069	0.0833	0.125	0.0369	0.0202	0.019
6	0	0	0	0	0	0	School	0.311267	0.622246		0.0959	0.0529	0.9851	0.796	0.0605	0.08	0.0345	0.2917	0.3333	0.013	0.0773	0.0549
7	0	0	0	0	0	0	Government		0.555912	0.729567												
8	0	0	0	0	0	0	Business Entity Type 3		0.650442													
9	0	0	0	0	1	1	Religion		0.322738													
10	0	0	0	0	0	0	Other		0.354225	0.621226												
11	0	0	0	0	0	0	Business Entity Type 3	0.774761	0.724	0.49206												
12	0	0	0	0	1	1	Other		0.714279	0.540654												
13	0	0	0	0	0	0	XNA	0.587334	0.205747	0.751724												
14	0	0	0	0	0	0	Electricity		0.746644													
15	0	0	0	0	0	0	Medicine	0.31976	0.651862	0.363945												
16	0	0	0	0	0	0	XNA	0.722044	0.555183	0.652897												
17	0	0	0	0	0	0	Business Entity Type 2	0.464831	0.715042	0.176653	0.0825		0.9811			0	0.2069	0.1667		0.0135		0.0778
18	0	0	0	0	0	0	Self-employed		0.566907	0.770087	0.1474	0.0973	0.9806	0.7348	0.0582	0.16	0.1379	0.3333	0.375	0.0931	0.1202	0.1397
19	0	0	0	0	0	0	Transport: type 2	0.72194	0.642656		0.3495	0.1335	0.9985	0.9796	0.1143	0.4	0.1724	0.6667	0.7083	0.1758	0.2849	0.3774
20	0	0	0	1	1	0	Business Entity Type 2	0.115634	0.346634	0.678568												
21	0	0	0	1	1	0	Government		0.236378	0.062103												
22	0	0	0	1	1	0	Construction		0.683513													
23	0	0	0	0	0	0	Housing		0.706428	0.556727	0.0278	0.0617	0.9881	0.8368	0.0018	0	0.1034	0.0833	0.125	0.0279	0.0227	0.029
24	0	0	0	0	0	0	Kindergarten		0.586617	0.477649												
25	0	0	0	0	0	0	Self-employed	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

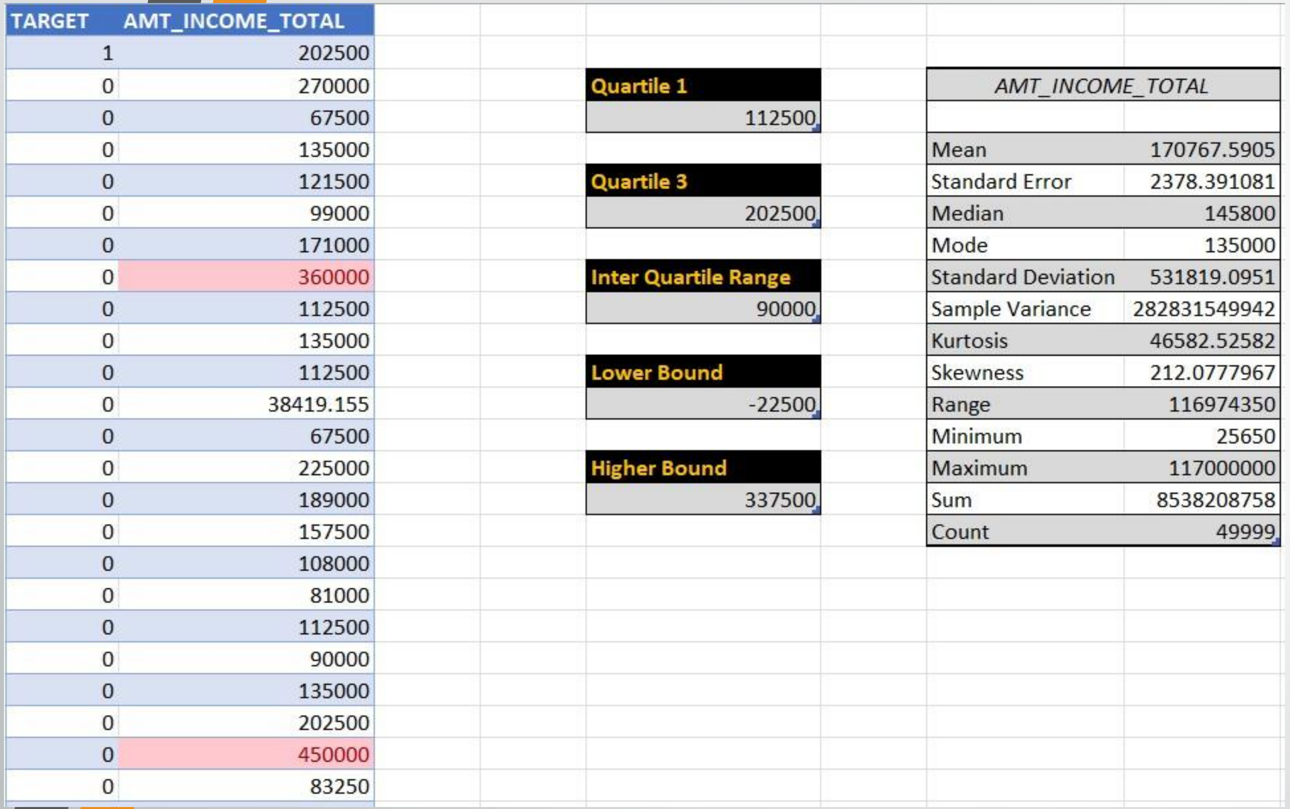
The columns with more than 30% of missing values are dropped and rest of the columns and rest of the columns having minimal missing values are flagged and are filled with average of their values. Non-numerical values the blanks are replaced with “UNKOWN” as we cannot determine accurate values.



Data Visualization for Total Blank Cells

B. Identify Outliers in the Dataset: Outliers can significantly impact the analysis and distort the results. You need to identify outliers in the loan application dataset.

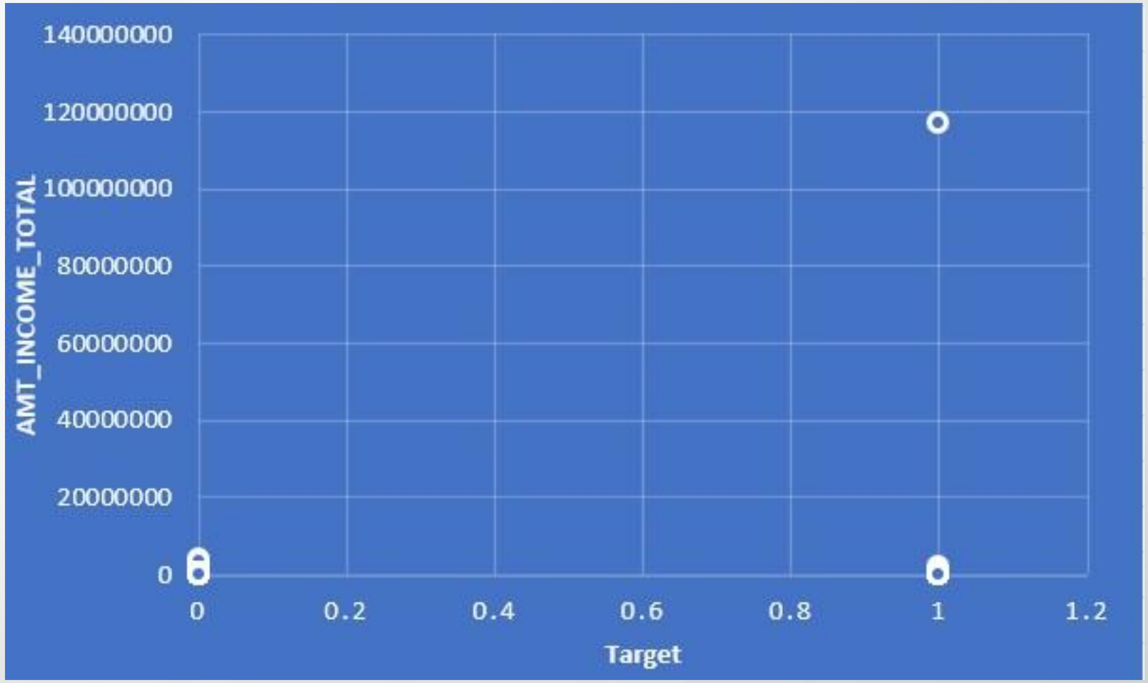
•**Task:** Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.



TARGET	AMT_INCOME_TOTAL
1	202500
0	270000
0	67500
0	135000
0	121500
0	99000
0	171000
0	360000
0	112500
0	135000
0	112500
0	38419.155
0	67500
0	225000
0	189000
0	157500
0	108000
0	81000
0	112500
0	90000
0	135000
0	202500
0	450000
0	83250

Quartile 1	112500
Quartile 3	202500
Inter Quartile Range	90000
Lower Bound	-22500
Higher Bound	337500

AMT_INCOME_TOTAL	
Mean	170767.5905
Standard Error	2378.391081
Median	145800
Mode	135000
Standard Deviation	531819.0951
Sample Variance	282831549942
Kurtosis	46582.52582
Skewness	212.0777967
Range	116974350
Minimum	25650
Maximum	117000000
Sum	8538208758
Count	49999



Using the formula **=QUARTILE(range, 1)** **to calculate Quartile 1 i.e. 25%**

Using the formula **=QUARTILE(range, 3)** **to calculate Quartile 3 i.e. 75%**

Using the formula **=Quartile 3 – Quartile1** **to calculate Inter Quartile Range**

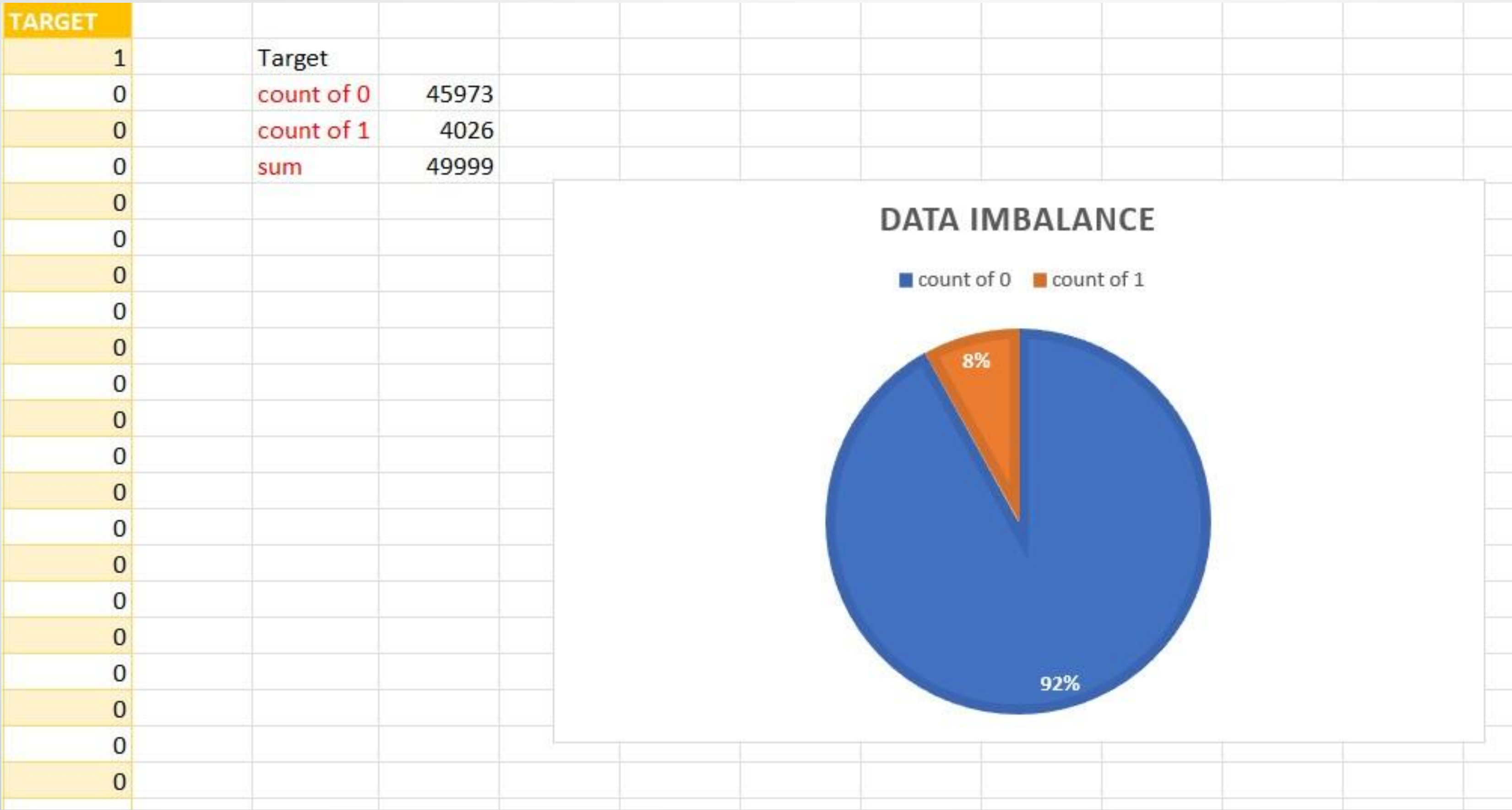
Using the formula **=Q1 - 1.5 * IQR** **to calculate Lower Bound**

Using the formula **= Q3 + 1.5 * IQR** **to calculate Higher Bound**

And determine the Outliers using the above formulas

C. Analyze Data Imbalance: Data imbalance can affect the accuracy of the analysis, especially for binary classification problems. Understanding the data distribution is crucial for building reliable models.

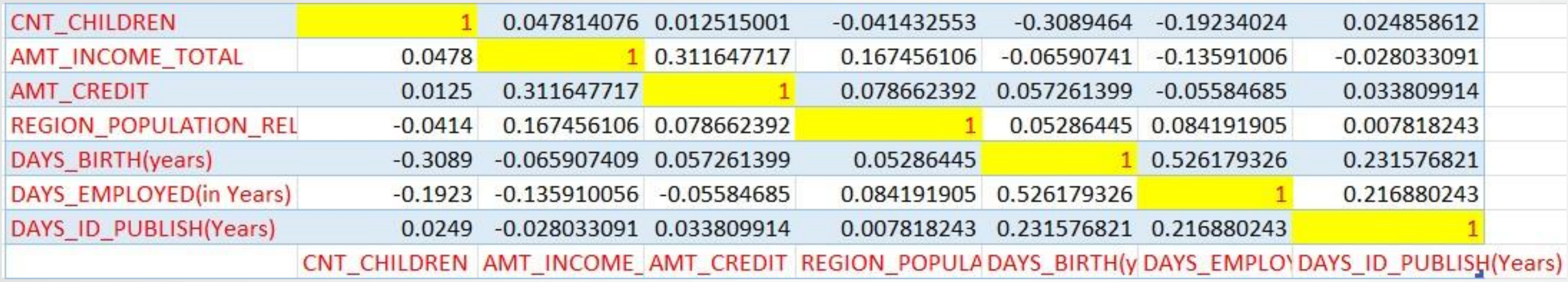
Task: Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.



TARGET			
1		Target	
0		count of 0	45973
0		count of 1	4026
0		sum	49999
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			
0			

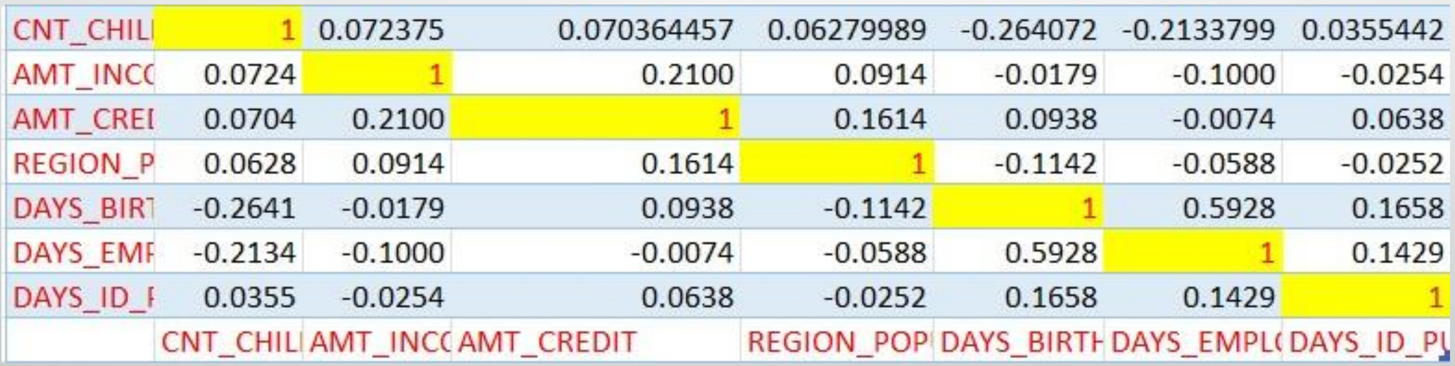
As we can see in the above data and visualization the data is highly imbalance in the column **Target** it consists of 0 and 1 with the total distribution of **92%** of **‘0’** and only **8%** of **‘1’**.

D. **Identify Top Correlations for Different Scenarios:** Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.
Task: Segment the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases) and identify the top correlations for each segmented data using Excel functions.



CNT_CHILDREN	1	0.047814076	0.012515001	-0.041432553	-0.3089464	-0.19234024	0.024858612
AMT_INCOME_TOTAL	0.0478	1	0.311647717	0.167456106	-0.06590741	-0.13591006	-0.028033091
AMT_CREDIT	0.0125	0.311647717	1	0.078662392	0.057261399	-0.05584685	0.033809914
REGION_POPULATION_REL	-0.0414	0.167456106	0.078662392	1	0.05286445	0.084191905	0.007818243
DAYS_BIRTH(years)	-0.3089	-0.065907409	0.057261399	0.05286445	1	0.526179326	0.231576821
DAYS_EMPLOYED(in Years)	-0.1923	-0.135910056	-0.05584685	0.084191905	0.526179326	1	0.216880243
DAYS_ID_PUBLISH(Years)	0.0249	-0.028033091	0.033809914	0.007818243	0.231576821	0.216880243	1
	CNT_CHILDREN	AMT_INCOME_	AMT_CREDIT	REGION_POPULA	DAYS_BIRTH(y	DAYS_EMPLOY	DAYS_ID_PUBLISH(Years)

Univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the **Target '0'** variable using Excel functions and features. Function =CORREL(r1,r2)



CNT_CHIL	1	0.072375	0.070364457	0.06279989	-0.264072	-0.2133799	0.0355442
AMT_INCC	0.0724	1	0.2100	0.0914	-0.0179	-0.1000	-0.0254
AMT_CREI	0.0704	0.2100	1	0.1614	0.0938	-0.0074	0.0638
REGION_P	0.0628	0.0914	0.1614	1	-0.1142	-0.0588	-0.0252
DAYS_BIRT	-0.2641	-0.0179	0.0938	-0.1142	1	0.5928	0.1658
DAYS_EMF	-0.2134	-0.1000	-0.0074	-0.0588	0.5928	1	0.1429
DAYS_ID_I	0.0355	-0.0254	0.0638	-0.0252	0.1658	0.1429	1
	CNT_CHIL	AMT_INCC	AMT_CREDIT	REGION_POP	DAYS_BIRTH	DAYS_EMPL(	DAYS_ID_PL

Univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the **Target '1'** variable using Excel functions and features. Function =CORREL(r1,r2)

Insights:

1.Missing data was handled by dropping highly null columns and imputing averages or "UNKNOWN" for remaining gaps.
2.Data imbalance revealed that 92% of applicants were non-defaulters, highlighting a skewed target distribution.
3.Statistical analysis and visualizations showed correlations between income, loan amounts, and default risks.

Results:

- Identified factors influencing defaults, enabling better loan approval strategies.
- Enhanced decision-making through risk-based recommendations like adjusting loan terms or interest rates.
- Improved data integrity by addressing missing values, outliers, and imbalances systematically.

Analyzing the Impact of Car Features on Price and Profitability

Analyzing the Impact of Car Features on Price and Profitability
77%

Description:

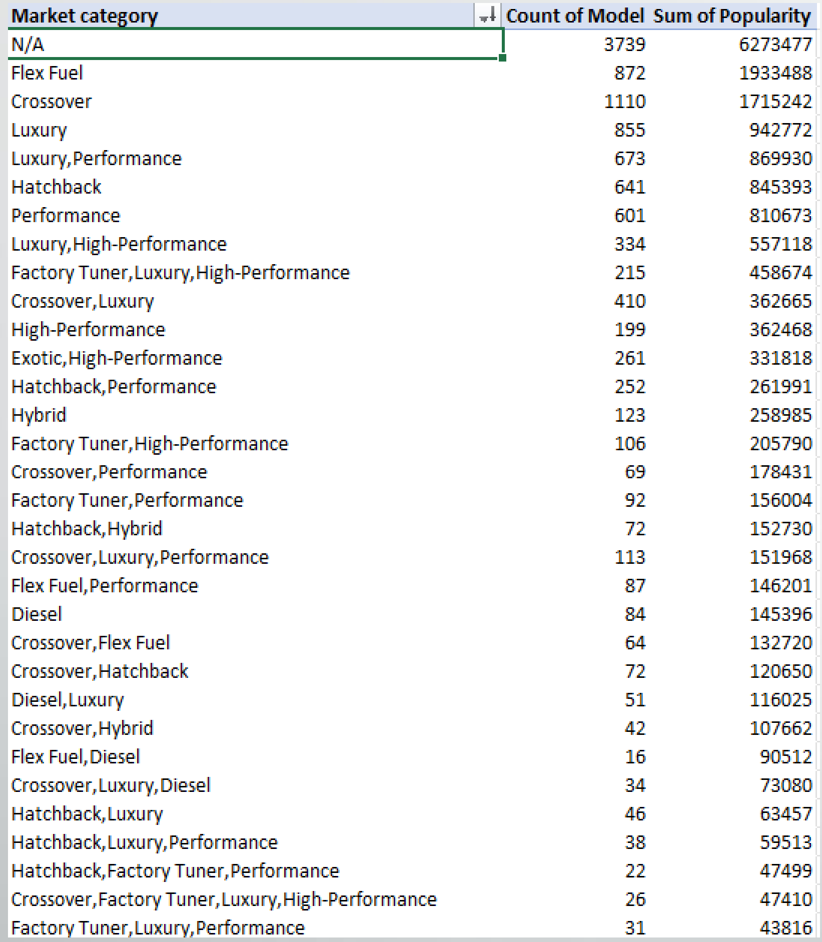
The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances consumer demand with profitability, and identify which product features to focus on in future product development efforts.

Excel sheet analysis : Click Here

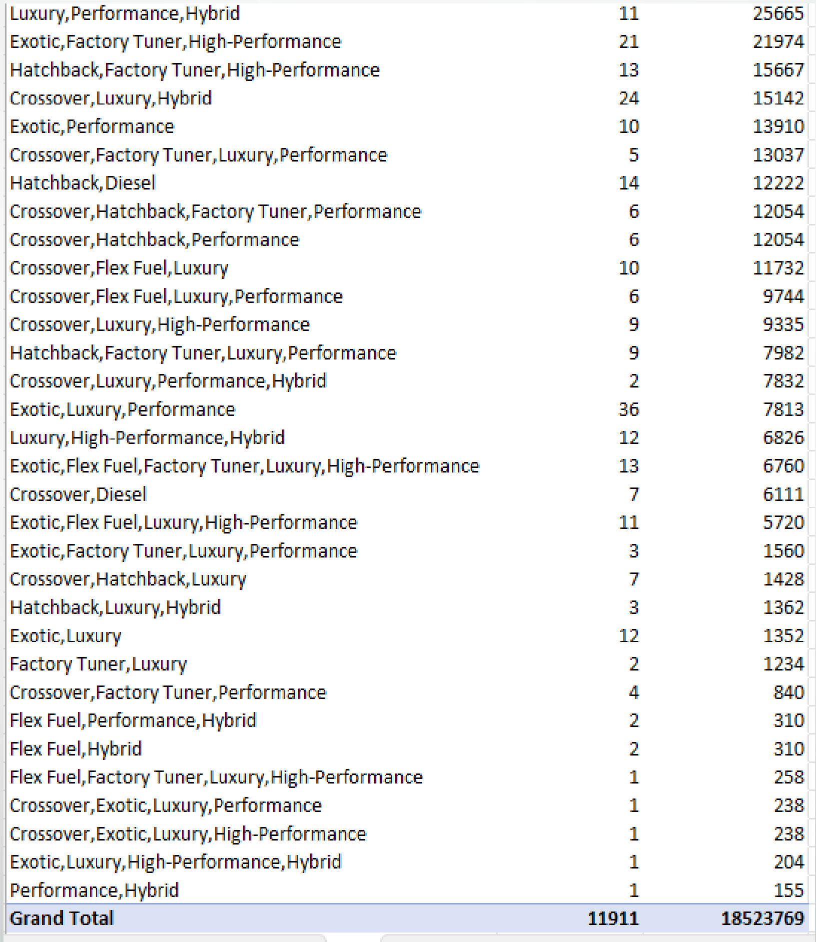
Findings

Insight Required: How does the popularity of a car model vary across different 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.



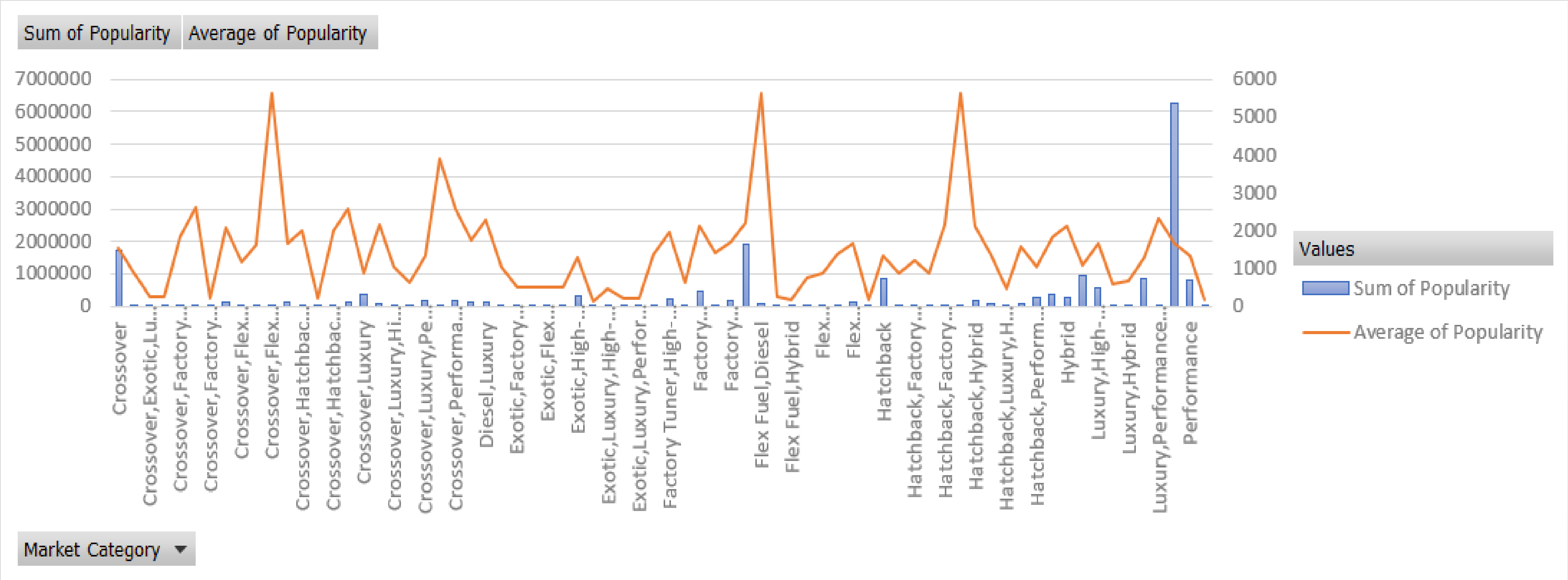
Market category	Count of Model	Sum of Popularity
N/A	3739	6273477
Flex Fuel	872	1933488
Crossover	1110	1715242
Luxury	855	942772
Luxury,Performance	673	869930
Hatchback	641	845393
Performance	601	810673
Luxury,High-Performance	334	557118
Factory Tuner,Luxury,High-Performance	215	458674
Crossover,Luxury	410	362665
High-Performance	199	362468
Exotic,High-Performance	261	331818
Hatchback,Performance	252	261991
Hybrid	123	258985
Factory Tuner,High-Performance	106	205790
Crossover,Performance	69	178431
Factory Tuner,Performance	92	156004
Hatchback,Hybrid	72	152730
Crossover,Luxury,Performance	113	151968
Flex Fuel,Performance	87	146201
Diesel	84	145396
Crossover,Flex Fuel	64	132720
Crossover,Hatchback	72	120650
Diesel,Luxury	51	116025
Crossover,Hybrid	42	107662
Flex Fuel,Diesel	16	90512
Crossover,Luxury,Diesel	34	73080
Hatchback,Luxury	46	63457
Hatchback,Luxury,Performance	38	59513
Hatchback,Factory Tuner,Performance	22	47499
Crossover,Factory Tuner,Luxury,High-Performance	26	47410
Factory Tuner,Luxury,Performance	31	43816



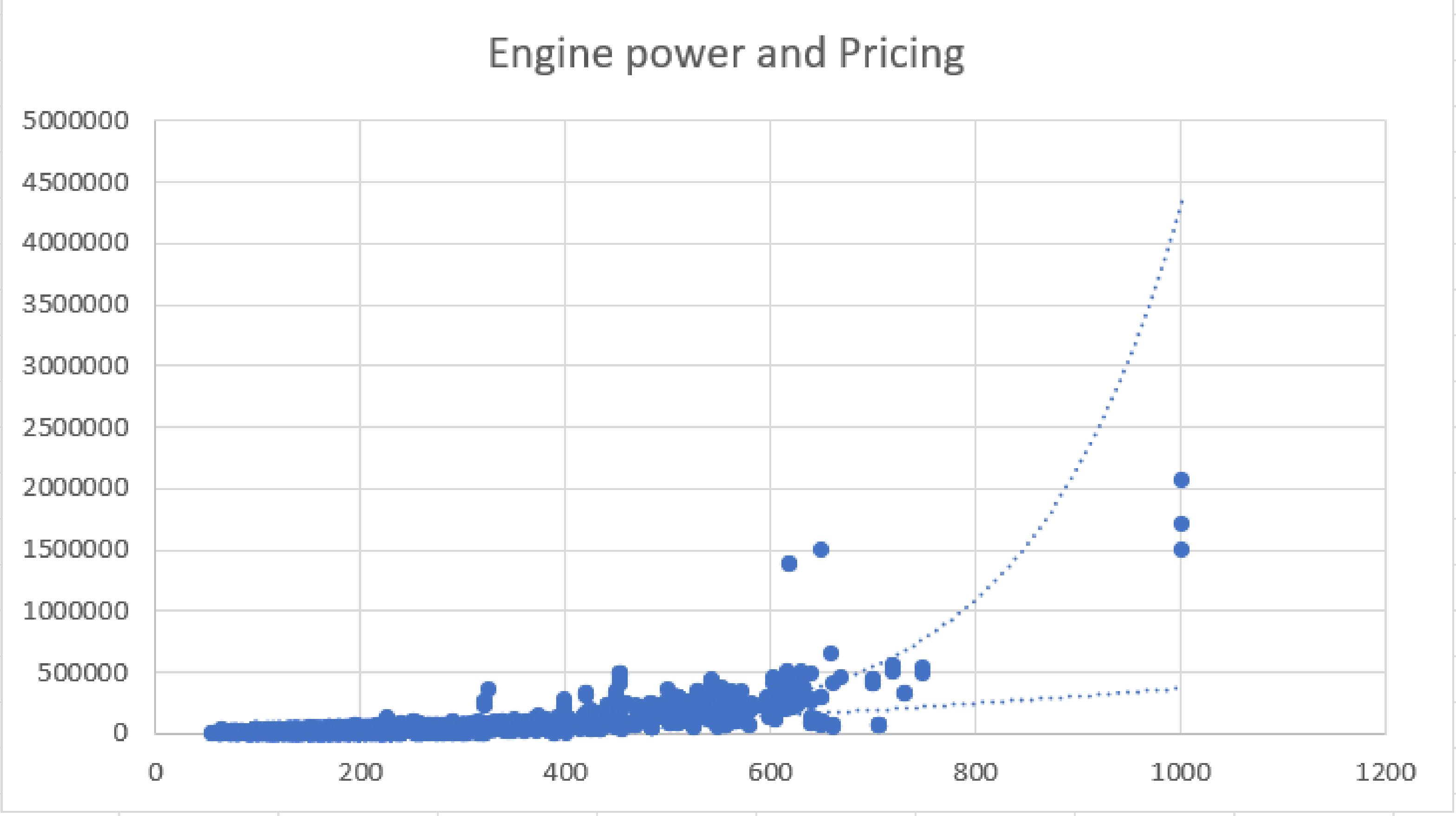
Luxury,Performance,Hybrid	11	25665
Exotic,Factory Tuner,High-Performance	21	21974
Hatchback,Factory Tuner,High-Performance	13	15667
Crossover,Luxury,Hybrid	24	15142
Exotic,Performance	10	13910
Crossover,Factory Tuner,Luxury,Performance	5	13037
Hatchback,Diesel	14	12222
Crossover,Hatchback,Factory Tuner,Performance	6	12054
Crossover,Hatchback,Performance	6	12054
Crossover,Flex Fuel,Luxury	10	11732
Crossover,Flex Fuel,Luxury,Performance	6	9744
Crossover,Luxury,High-Performance	9	9335
Hatchback,Factory Tuner,Luxury,Performance	9	7982
Crossover,Luxury,Performance,Hybrid	2	7832
Exotic,Luxury,Performance	36	7813
Luxury,High-Performance,Hybrid	12	6826
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	13	6760
Crossover,Diesel	7	6111
Exotic,Flex Fuel,Luxury,High-Performance	11	5720
Exotic,Factory Tuner,Luxury,Performance	3	1560
Crossover,Hatchback,Luxury	7	1428
Hatchback,Luxury,Hybrid	3	1362
Exotic,Luxury	12	1352
Factory Tuner,Luxury	2	1234
Crossover,Factory Tuner,Performance	4	840
Flex Fuel,Performance,Hybrid	2	310
Flex Fuel,Hybrid	2	310
Flex Fuel,Factory Tuner,Luxury,High-Performance	1	258
Crossover,Exotic,Luxury,Performance	1	238
Crossover,Exotic,Luxury,High-Performance	1	238
Exotic,Luxury,High-Performance,Hybrid	1	204
Performance,Hybrid	1	155
Grand Total	**11911**	**18523769**

The popularity of a car model with respect to different market categories is represented through pivot table.

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



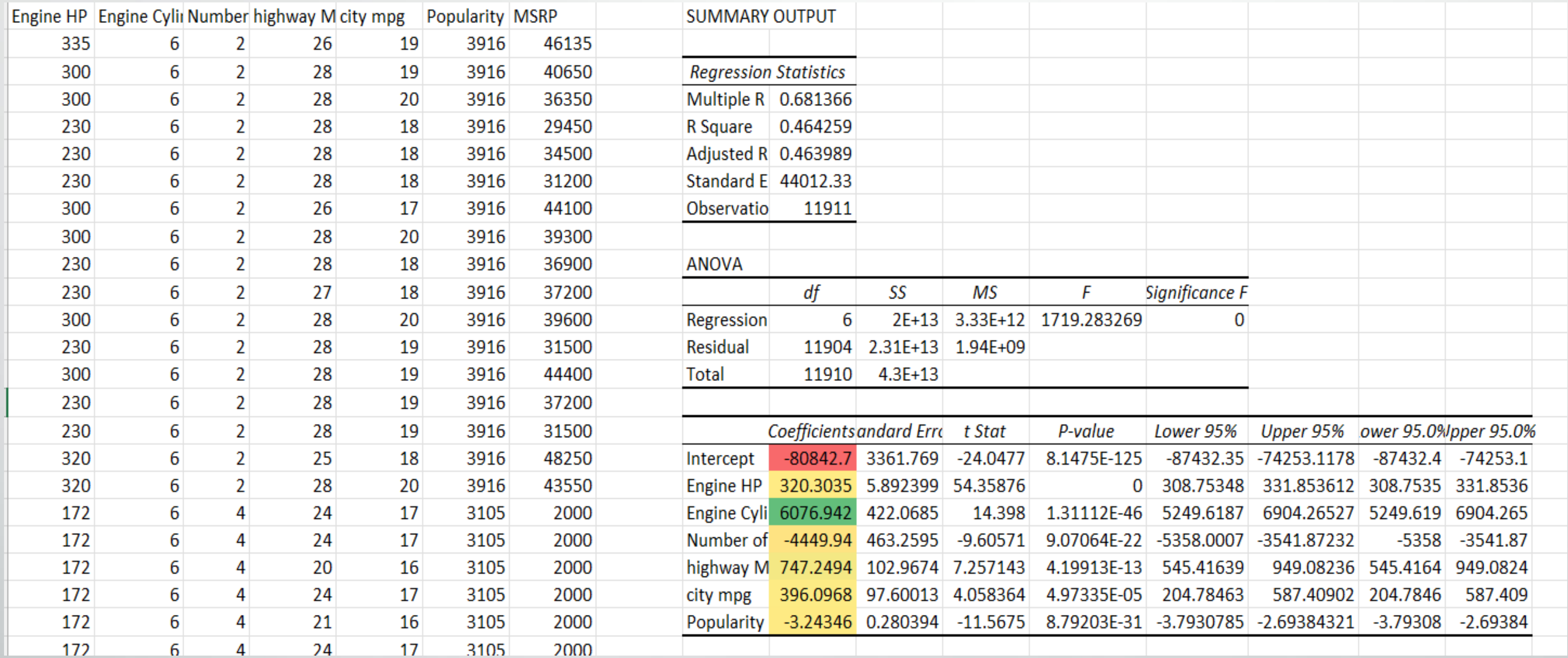
Insight Required: What is the relationship between a car's engine power and its price?
Task 2: Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.



The graph shows a increasing trendline slope, which tells there is a direct correlation between a car's engine power and its price. This implies that vehicles with more powerful engines tend have higher price tags.

Insight Required: 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.



Engine HP	Engine Cyli	Number	highway M	city mpg	Popularity	MSRP
335	6	2	26	19	3916	46135
300	6	2	28	19	3916	40650
300	6	2	28	20	3916	36350
230	6	2	28	18	3916	29450
230	6	2	28	18	3916	34500
230	6	2	28	18	3916	31200
300	6	2	26	17	3916	44100
300	6	2	28	20	3916	39300
230	6	2	28	18	3916	36900
230	6	2	27	18	3916	37200
300	6	2	28	20	3916	39600
230	6	2	28	19	3916	31500
300	6	2	28	19	3916	44400
230	6	2	28	19	3916	37200
230	6	2	28	19	3916	31500
320	6	2	25	18	3916	48250
320	6	2	28	20	3916	43550
172	6	4	24	17	3105	2000
172	6	4	24	17	3105	2000
172	6	4	20	16	3105	2000
172	6	4	24	17	3105	2000
172	6	4	21	16	3105	2000
172	6	4	24	17	3105	2000

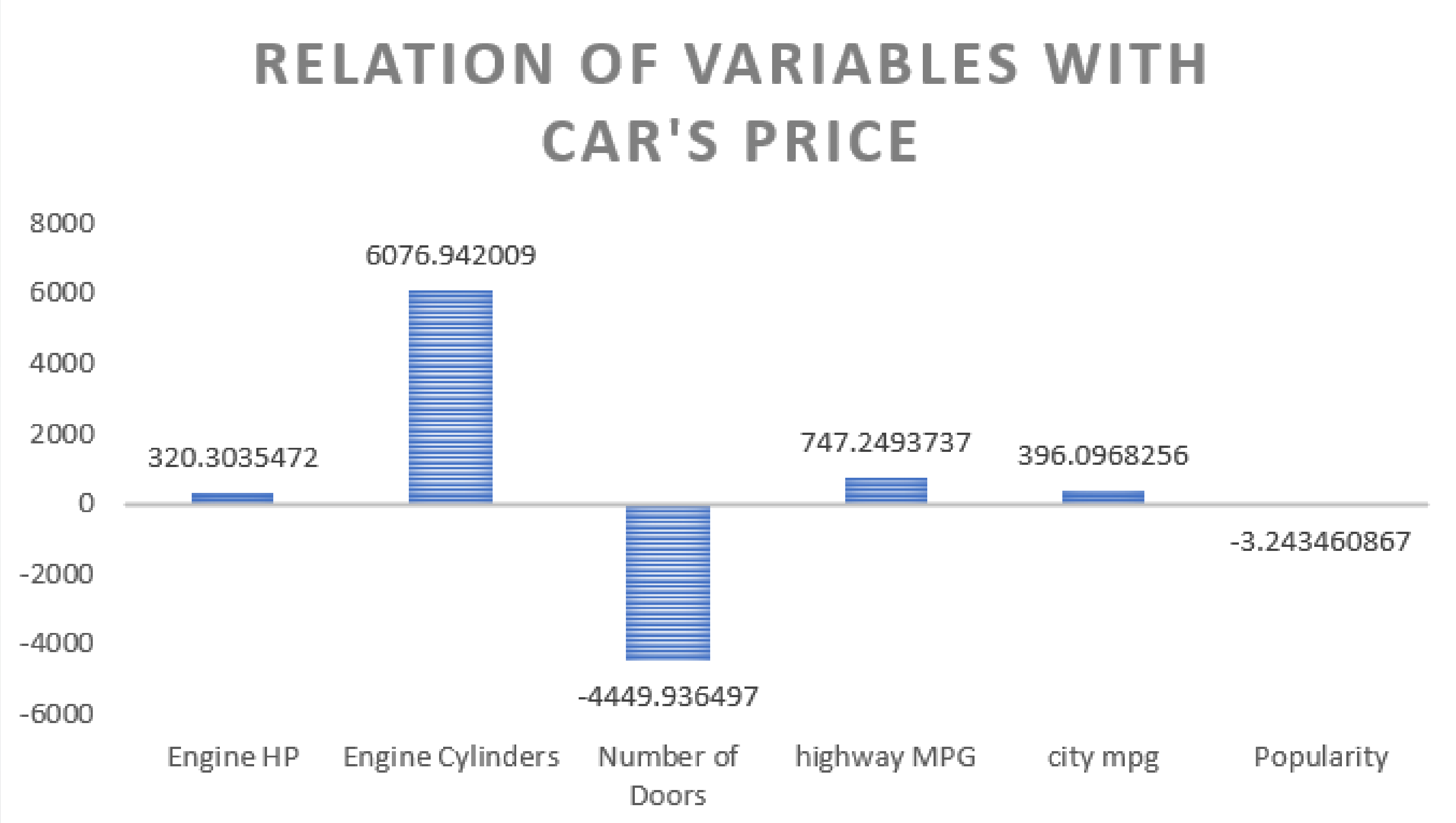
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.681366
R Square	0.464259
Adjusted R	0.463989
Standard E	44012.33
Observatio	11911

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	2E+13	3.33E+12	1719.283269	0
Residual	11904	2.31E+13	1.94E+09		
Total	11910	4.3E+13			

	Coefficients	*andard Errc*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	-80842.7	3361.769	-24.0477	8.1475E-125	-87432.35	-74253.1178	-87432.4	-74253.1
Engine HP	320.3035	5.892399	54.35876	0	308.75348	331.853612	308.7535	331.8536
Engine Cyli	6076.942	422.0685	14.398	1.31112E-46	5249.6187	6904.26527	5249.619	6904.265
Number of	-4449.94	463.2595	-9.60571	9.07064E-22	-5358.0007	-3541.87232	-5358	-3541.87
highway M	747.2494	102.9674	7.257143	4.19913E-13	545.41639	949.08236	545.4164	949.0824
city mpg	396.0968	97.60013	4.058364	4.97335E-05	204.78463	587.40902	204.7846	587.409
Popularity	-3.24346	0.280394	-11.5675	8.79203E-31	-3.7930785	-2.69384321	-3.79308	-2.69384

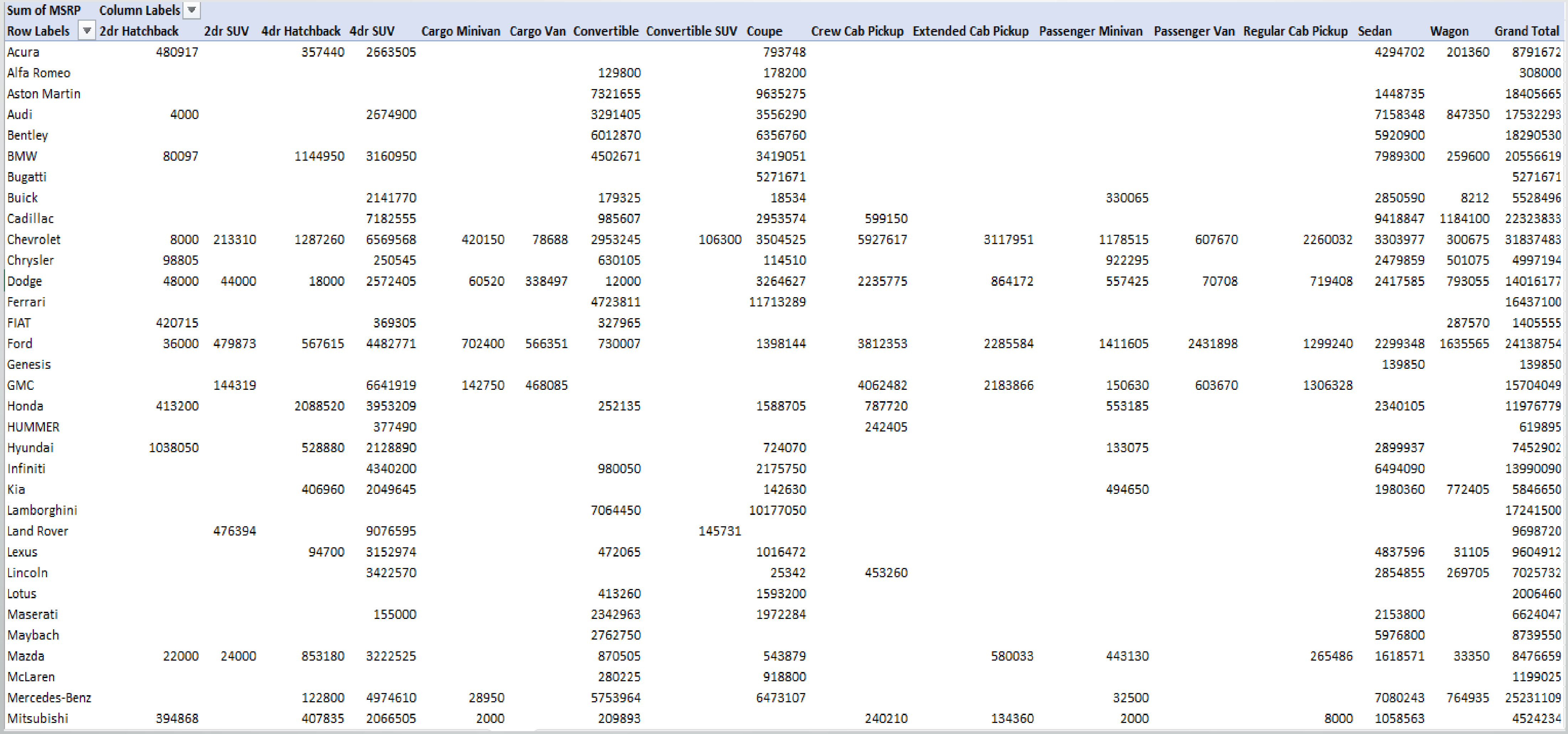


On the basis of regression analysis, the variables of Engine horsepower, Engine cylinders, City MPG, Highway MPG, and MSRP shows the strongest relationship with a car's price.

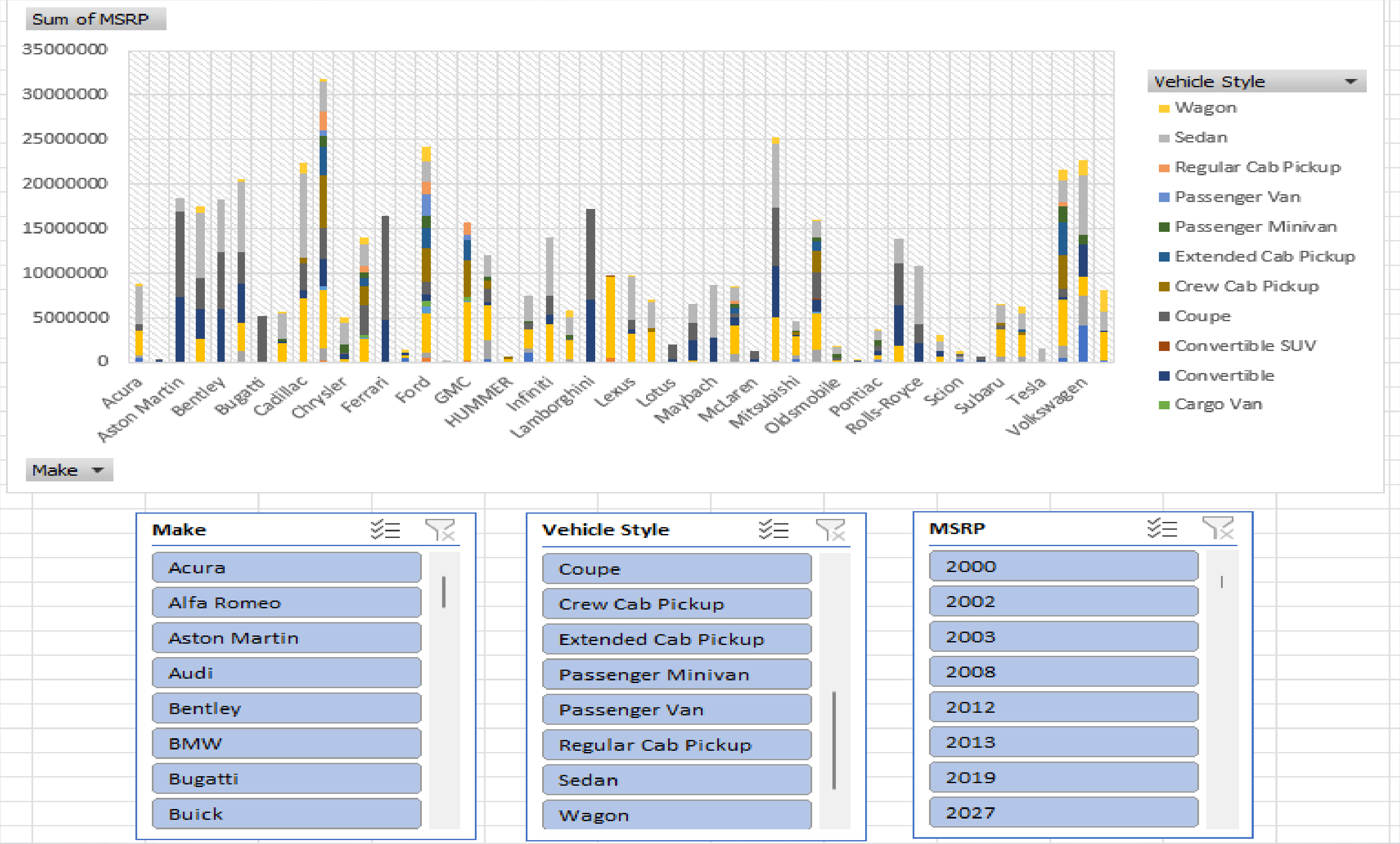
From the chart we can say that the strongest relationship with price is of Engine Cylinders and the negative relationship is with Number of Doors, which means it is inversely related to the car's price.

Building the Dashboard:

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



Sum of MSRP	Column Labels ▼															
Row Labels ▼	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Passenger Minivan	Passenger Van	Regular Cab Pickup	Sedan	Wagon	Grand Total
Acura	480917		357440	2663505					793748						4294702	201360	8791672
Alfa Romeo							129800		178200								308000
Aston Martin							7321655		9635275						1448735		18405665
Audi	4000			2674900			3291405		3556290						7158348	847350	17532293
Bentley							6012870		6356760						5920900		18290530
BMW	80097		1144950	3160950			4502671		3419051						7989300	259600	20556619
Bugatti									5271671								5271671
Buick				2141770			179325		18534			330065			2850590	8212	5528496
Cadillac				7182555			985607		2953574	599150					9418847	1184100	22323833
Chevrolet	8000	213310	1287260	6569568	420150	78688	2953245	106300	3504525	5927617	3117951	1178515	607670	2260032	3303977	300675	31837483
Chrysler	98805			250545			630105		114510			922295			2479859	501075	4997194
Dodge	48000	44000	18000	2572405	60520	338497	12000		3264627	2235775	864172	557425	70708	719408	2417585	793055	14016177
Ferrari							4723811		11713289								16437100
FIAT	420715			369305			327965									287570	1405555
Ford	36000	479873	567615	4482771	702400	566351	730007		1398144	3812353	2285584	1411605	2431898	1299240	2299348	1635565	24138754
Genesis															139850		139850
GMC		144319		6641919	142750	468085				4062482	2183866	150630	603670	1306328			15704049
Honda	413200		2088520	3953209			252135		1588705	787720		553185			2340105		11976779
HUMMER				377490						242405							619895
Hyundai	1038050		528880	2128890					724070			133075			2899937		7452902
Infiniti				4340200			980050		2175750						6494090		13990090
Kia			406960	2049645					142630			494650			1980360	772405	5846650
Lamborghini							7064450		10177050								17241500
Land Rover		476394		9076595				145731									9698720
Lexus			94700	3152974			472065		1016472						4837596	31105	9604912
Lincoln				3422570					25342	453260					2854855	269705	7025732
Lotus							413260		1593200								2006460
Maserati				155000			2342963		1972284						2153800		6624047
Maybach							2762750								5976800		8739550
Mazda	22000	24000	853180	3222525			870505		543879		580033	443130		265486	1618571	33350	8476659
McLaren							280225		918800								1199025
Mercedes-Benz			122800	4974610	28950		5753964		6473107			32500			7080243	764935	25231109
Mitsubishi	394868		407835	2066505	2000		209893			240210	134360	2000		8000	1058563		4524234



Make

- Acura
- Alfa Romeo
- Aston Martin
- Audi
- Bentley
- BMW
- Bugatti
- Buick

Vehicle Style

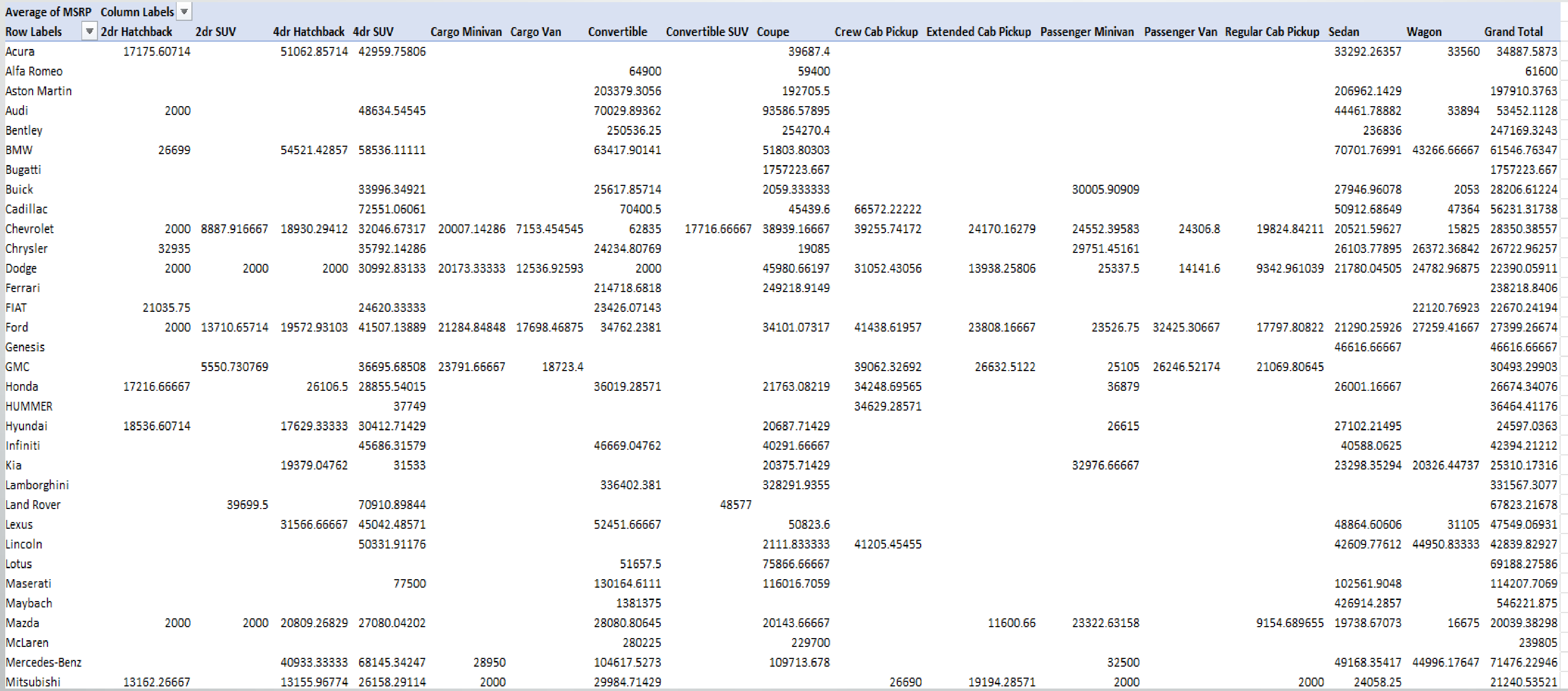
- Coupe
- Crew Cab Pickup
- Extended Cab Pickup
- Passenger Minivan
- Passenger Van
- Regular Cab Pickup
- Sedan
- Wagon

MSRP

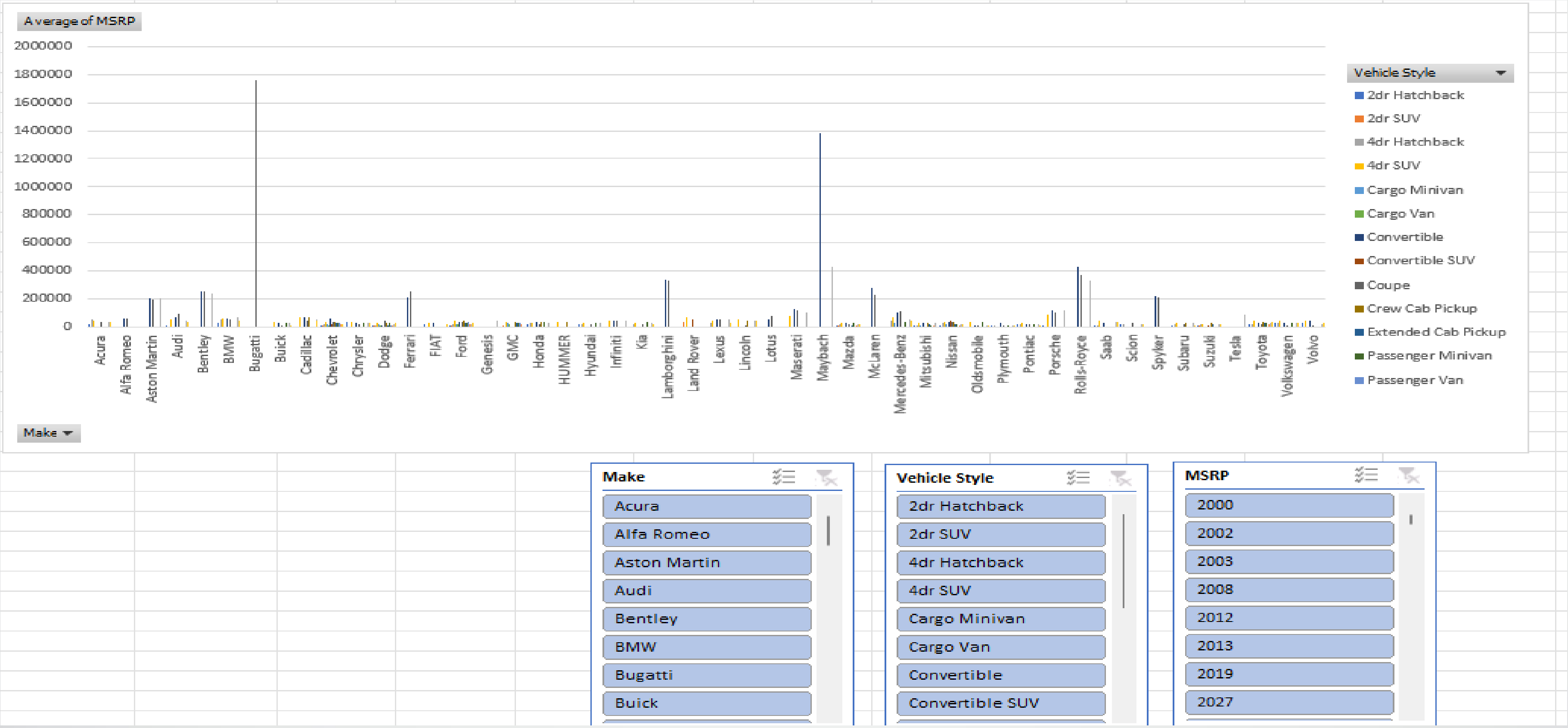
- 2000
- 2002
- 2003
- 2008
- 2012
- 2013
- 2019
- 2027

This analysis gives valuable insights into the variations in car prices based on brand and body style. Such insights can help for manufacturers in optimizing their pricing strategies and enhancing profitability. With addition to the utilization of slicers enables a deeper exploration of the data, allowing for a more detailed examination of specific details and patterns.

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



Average of MSRP	Column Labels																
Row Labels	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Passenger Minivan	Passenger Van	Regular Cab Pickup	Sedan	Wagon	Grand Total
Acura	17175.60714		51062.85714	42959.75806					39687.4						33292.26357	33560	34887.5873
Alfa Romeo							64900		59400								61600
Aston Martin							203379.3056		192705.5						206962.1429		197910.3763
Audi	2000			48634.54545			70029.89362		93586.57895						44461.78882	33894	53452.1128
Bentley							250536.25		254270.4						236836		247169.3243
BMW	26699		54521.42857	58536.11111			63417.90141		51803.80303						70701.76991	43266.66667	61546.76347
Bugatti									1757223.667								1757223.667
Buick				33996.34921			25617.85714		2059.333333			30005.90909			27946.96078	2053	28206.61224
Cadillac				72551.06061			70400.5		45439.6	66572.22222					50912.68649	47364	56231.31738
Chevrolet	2000	8887.916667	18930.29412	32046.67317	20007.14286	7153.454545	62835	17716.66667	38939.16667	39255.74172	24170.16279	24552.39583	24306.8	19824.84211	20521.59627	15825	28350.38557
Chrysler	32935			35792.14286			24234.80769		19085			29751.45161			26103.77895	26372.36842	26722.96257
Dodge	2000	2000	2000	30992.83133	20173.33333	12536.92593	2000		45980.66197	31052.43056	13938.25806	25337.5	14141.6	9342.961039	21780.04505	24782.96875	22390.05911
Ferrari							214718.6818		249218.9149								238218.8406
FIAT	21035.75			24620.33333			23426.07143									22120.76923	22670.24194
Ford	2000	13710.65714	19572.93103	41507.13889	21284.84848	17698.46875	34762.2381		34101.07317	41438.61957	23808.16667	23526.75	32425.30667	17797.80822	21290.25926	27259.41667	27399.26674
Genesis															46616.66667		46616.66667
GMC		5550.730769		36695.68508	23791.66667	18723.4				39062.32692	26632.5122	25105	26246.52174	21069.80645			30493.29903
Honda	17216.66667		26106.5	28855.54015			36019.28571		21763.08219	34248.69565		36879			26001.16667		26674.34076
HUMMER				37749						34629.28571							36464.41176
Hyundai	18536.60714		17629.33333	30412.71429					20687.71429			26615			27102.21495		24597.0363
Infiniti				45686.31579			46669.04762		40291.66667						40588.0625		42394.21212
Kia			19379.04762	31533					20375.71429			32976.66667			23298.35294	20326.44737	25310.17316
Lamborghini							336402.381		328291.9355								331567.3077
Land Rover		39699.5		70910.89844				48577									67823.21678
Lexus			31566.66667	45042.48571			52451.66667		50823.6						48864.60606	31105	47549.06931
Lincoln				50331.91176					2111.833333	41205.45455					42609.77612	44950.83333	42839.82927
Lotus							51657.5		75866.66667								69188.27586
Maserati				77500			130164.6111		116016.7059						102561.9048		114207.7069
Maybach							1381375								426914.2857		546221.875
Mazda	2000	2000	20809.26829	27080.04202			28080.80645		20143.66667		11600.66	23322.63158		9154.689655	19738.67073	16675	20039.38298
McLaren							280225		229700								239805
Mercedes-Benz			40933.33333	68145.34247	28950		104617.5273		109713.678			32500			49168.35417	44996.17647	71476.22946
Mitsubishi	13162.26667		13155.96774	26158.29114	2000		29984.71429			26690	19194.28571	2000		2000	24058.25		21240.53521



During the analysis, it was observed that certain brands exhibit significantly higher or lower average Manufacturer's Suggested Retail Price (MSRP) compared to others. Luxury brands such as Bugatti, Maybach, and Rolls Royce generally have higher average MSRP values compared to brands like BMW, Toyota, and Audi. Additionally, specific vehicle styles, including Sedan, 4Dr SUV, and Coupe, tend to have higher average MSRP values compared to other styles. This underscores the price variations based on brand and vehicle style, highlighting the diverse pricing landscape within the automotive market. Furthermore, the results of this analysis illustrate the progression of fuel efficiency over time across different body styles.

Insights:

1.Engine power correlates positively with car price, vehicles with higher horsepower generally have higher prices.

2.Market categories, such as luxury and performance, influence a car's popularity and pricing significantly.

3.Fuel efficiency is inversely related to engine cylinders; more cylinders result in lower highway MPG, emphasizing a trade-off between performance and efficiency.

Results:

•Identified key factors like engine specifications and fuel efficiency affecting car pricing and demand.

•Established manufacturer-wise and body style-based price variations, aiding pricing strategy optimization.

•Highlighted market trends, such as growing emphasis on fuel-efficient body styles and their impact on profitability.

ABC Call Volume Trend Analysis

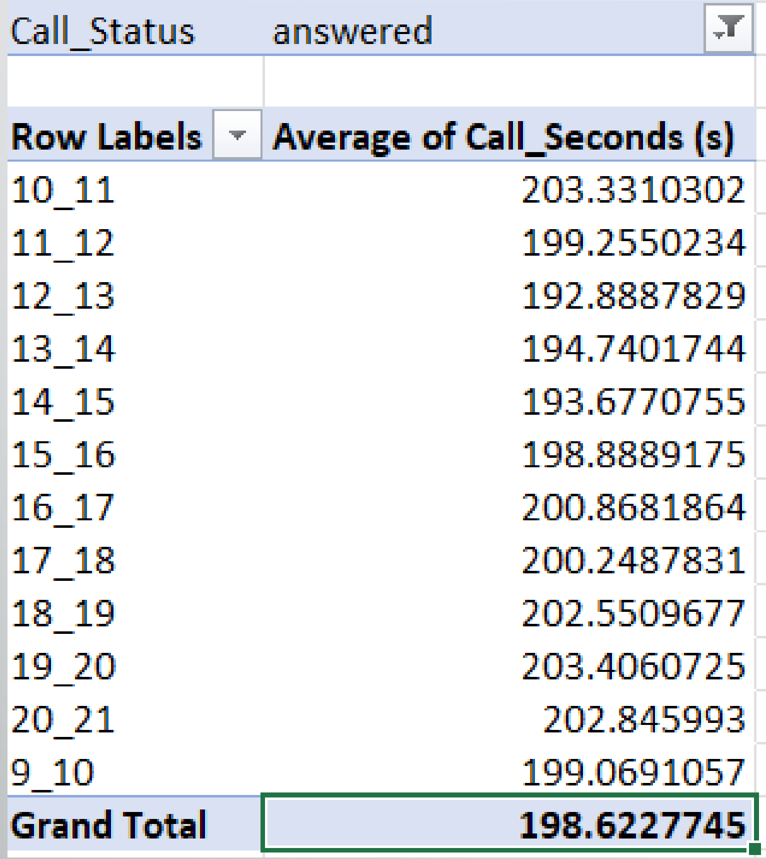
Description:

In this project, you'll be diving into the world of Customer Experience (CX) analytics, specifically focusing on the inbound calling team of a company. You'll be provided with a dataset that spans 23 days and includes various details such as the agent's name and ID, the queue time (how long a customer had to wait before connecting with an agent), the time of the call, the duration of the call, and the call status (whether it was abandoned, answered, or transferred). Inbound customer support, which is the focus of this project, involves handling incoming calls from existing or prospective customers. The goal is to attract, engage, and delight customers, turning them into loyal advocates for the business.

Findings

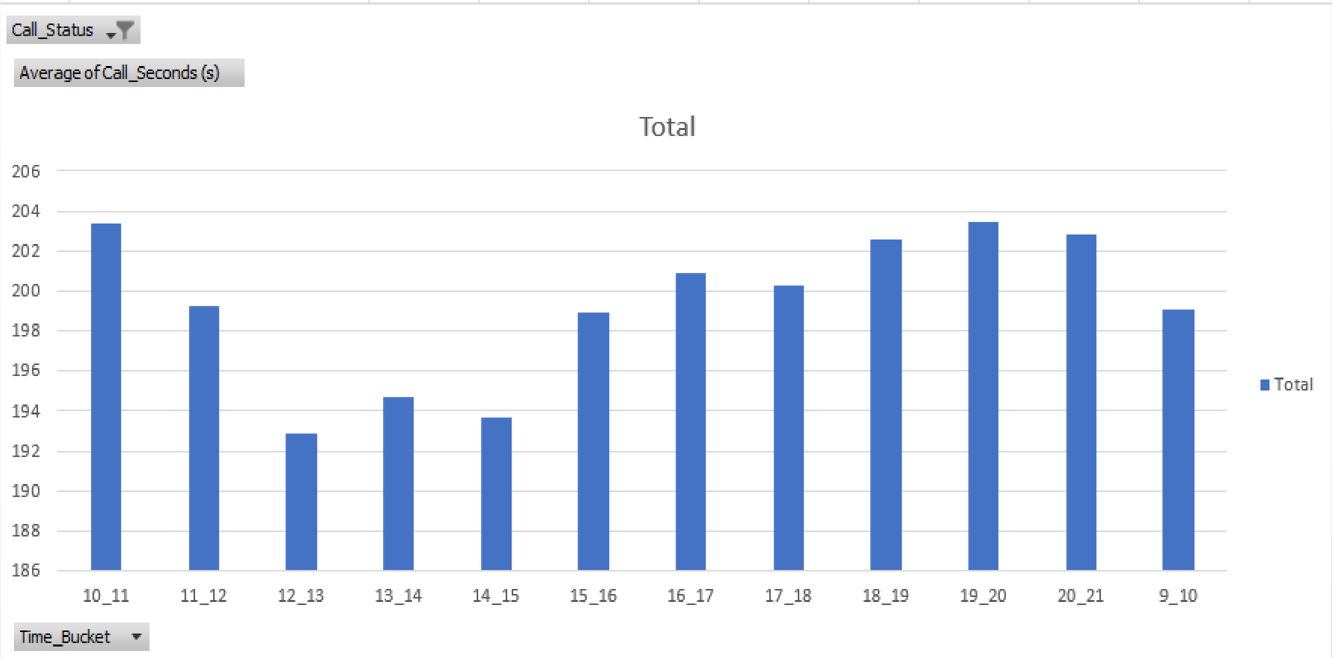
1.Average Call Duration: Determine the average duration of all incoming calls received by agents. This should be calculated for each time bucket.

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



Call_Status answered

Row Labels	Average of Call_Seconds (s)
10_11	203.3310302
11_12	199.2550234
12_13	192.8887829
13_14	194.7401744
14_15	193.6770755
15_16	198.8889175
16_17	200.8681864
17_18	200.2487831
18_19	202.5509677
19_20	203.4060725
20_21	202.845993
9_10	199.0691057
Grand Total	**198.6227745**

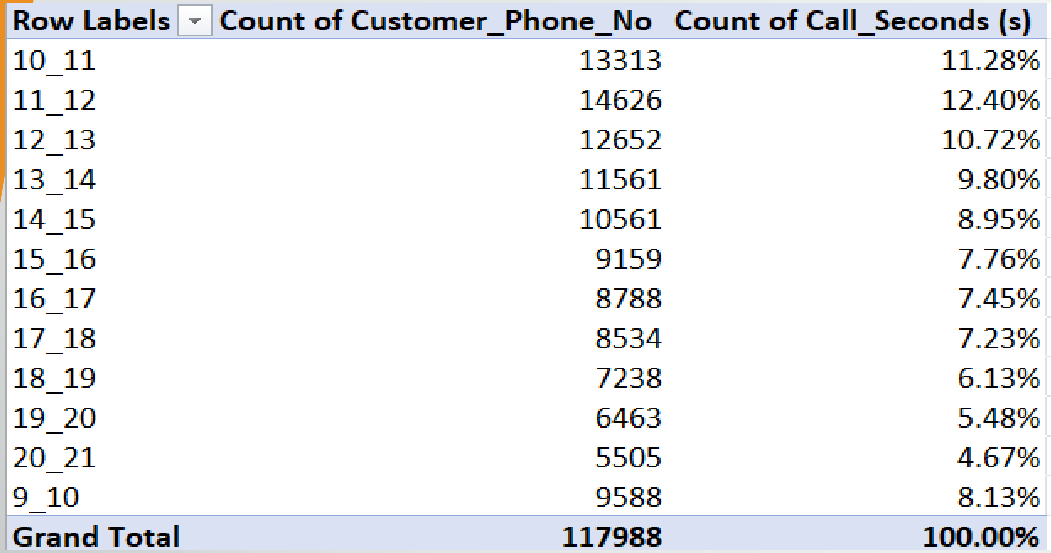


The average of Call seconds which are answered is of 198.6 seconds in total.

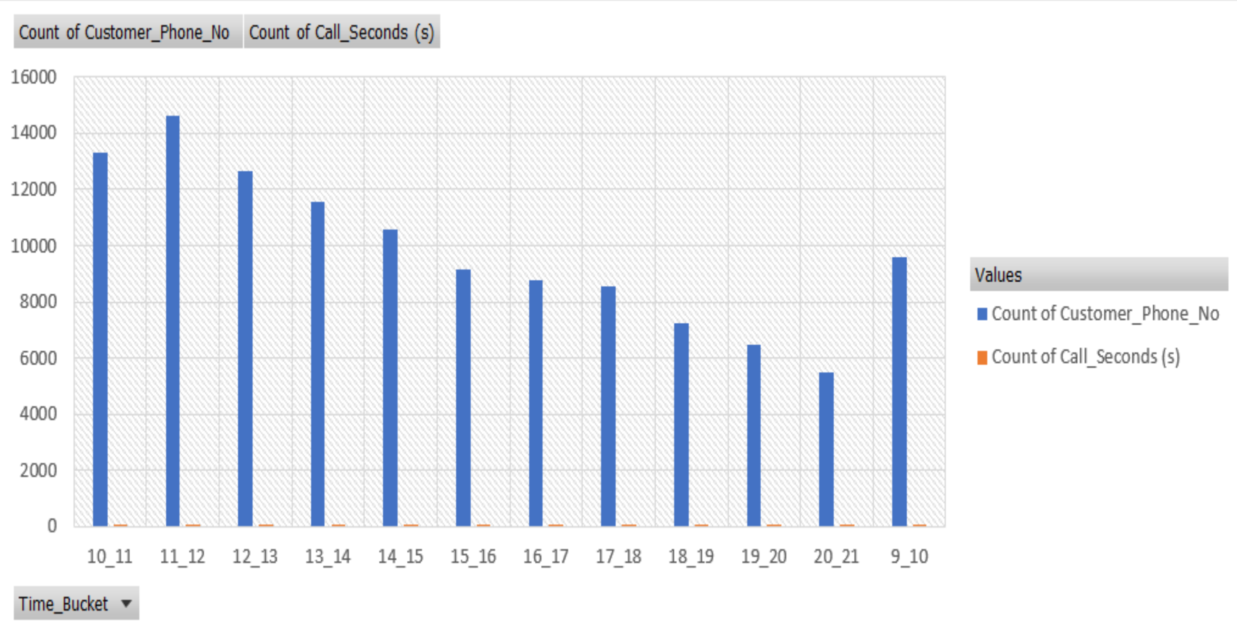
- The analysis tells us that the average call time duration for incoming calls received by agents is highest between 10 am to 11 am and from 7 pm to 8 pm.

2.Call Volume Analysis: Visualize the total number of calls received. This should be represented as a graph or chart showing the number of calls against time. Time should be represented in buckets (e.g., 1-2, 2-3, etc.).

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



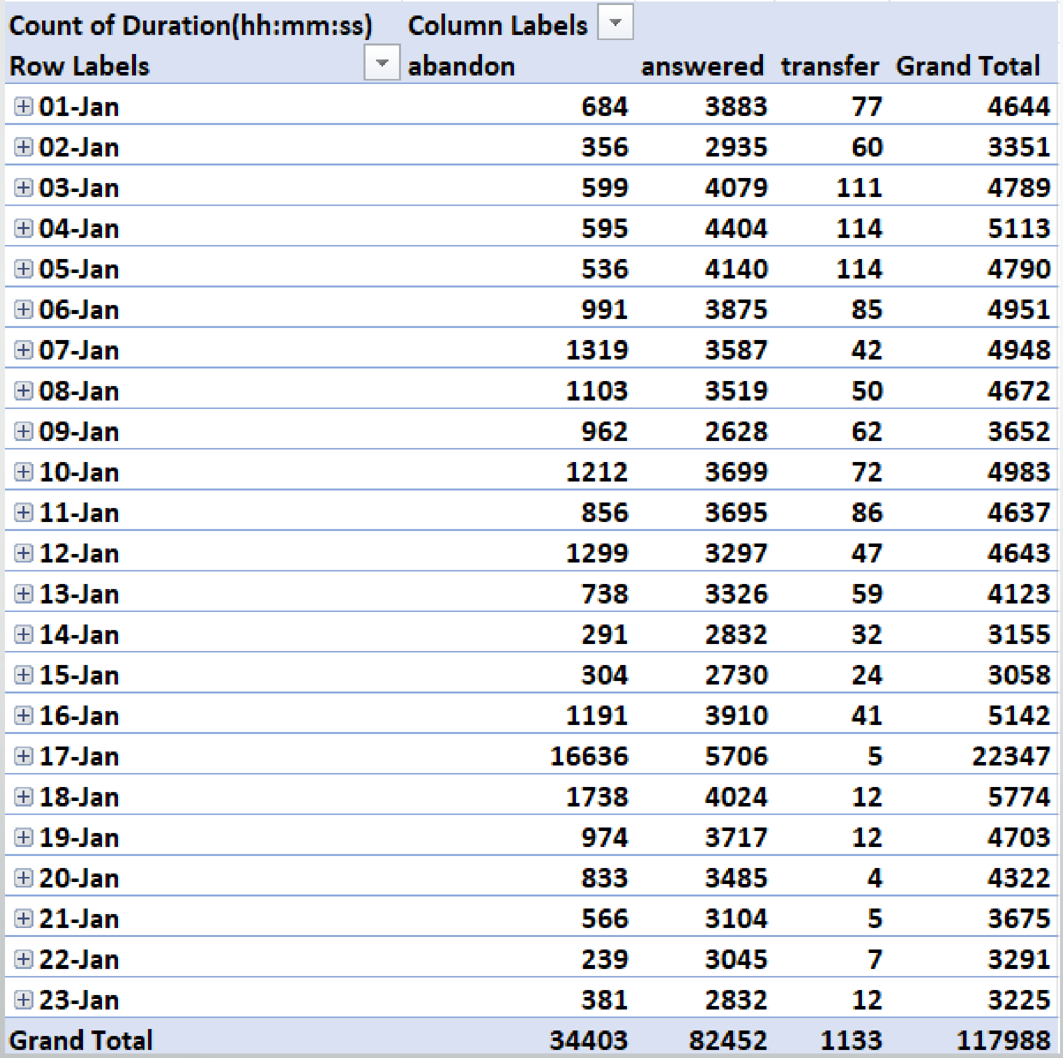
Row Labels ▾	Count of Customer_Phone_No	Count of Call_Seconds (s)
10_11	13313	11.28%
11_12	14626	12.40%
12_13	12652	10.72%
13_14	11561	9.80%
14_15	10561	8.95%
15_16	9159	7.76%
16_17	8788	7.45%
17_18	8534	7.23%
18_19	7238	6.13%
19_20	6463	5.48%
20_21	5505	4.67%
9_10	9588	8.13%
Grand Total	**117988**	**100.00%**



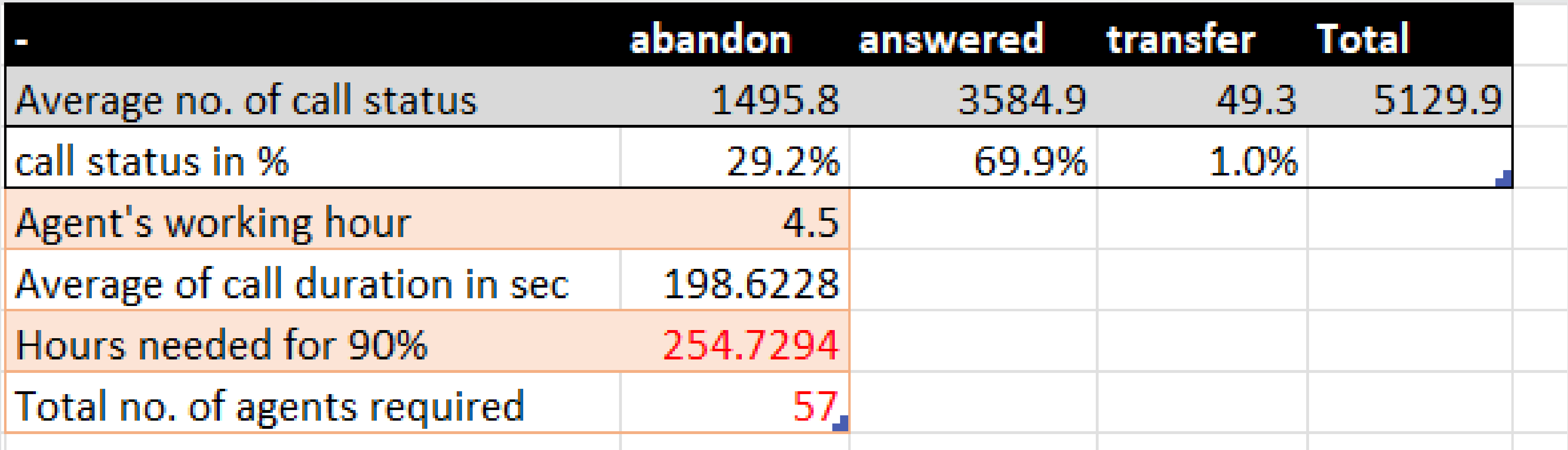
- Considering the analysis, it was observed that customers make the maximum number of calls between 11 am to 12 noon.
- The analysis also suggests that customers make the minimum number of calls between 8 pm to 9 pm.
- This analysis suggests the user to know at what time range the customers will more likely to answer the call.

3.Manpower Planning: The current rate of abandoned calls is approximately 30%. Propose a plan for manpower allocation during each time bucket (from 9 am to 9 pm) to reduce the abandon rate to 10%. In other words, you need to calculate the minimum number of agents required in each time bucket to ensure that at least 90 out of 100 calls are answered.

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



Count of Duration(hh:mm:ss)	Column Labels			
Row Labels	abandon	answered	transfer	Grand Total
01-Jan	684	3883	77	4644
02-Jan	356	2935	60	3351
03-Jan	599	4079	111	4789
04-Jan	595	4404	114	5113
05-Jan	536	4140	114	4790
06-Jan	991	3875	85	4951
07-Jan	1319	3587	42	4948
08-Jan	1103	3519	50	4672
09-Jan	962	2628	62	3652
10-Jan	1212	3699	72	4983
11-Jan	856	3695	86	4637
12-Jan	1299	3297	47	4643
13-Jan	738	3326	59	4123
14-Jan	291	2832	32	3155
15-Jan	304	2730	24	3058
16-Jan	1191	3910	41	5142
17-Jan	16636	5706	5	22347
18-Jan	1738	4024	12	5774
19-Jan	974	3717	12	4703
20-Jan	833	3485	4	4322
21-Jan	566	3104	5	3675
22-Jan	239	3045	7	3291
23-Jan	381	2832	12	3225
Grand Total	34403	82452	1133	117988



-	abandon	answered	transfer	Total
Average no. of call status	1495.8	3584.9	49.3	5129.9
call status in %	29.2%	69.9%	1.0%	
Agent's working hour	4.5			
Average of call duration in sec	198.6228			
Hours needed for 90%	254.7294			
Total no. of agents required	57			

By assuming that a person works 7.5 hrs a day, 6 days in a week with 60% in engaging to answered calls.

By using the formula =(60/100)*7.5, with this we can determine the worker being in call i.e. 4.5 hrs a day

We have already calculated average Call Duration in sec in Task 1.

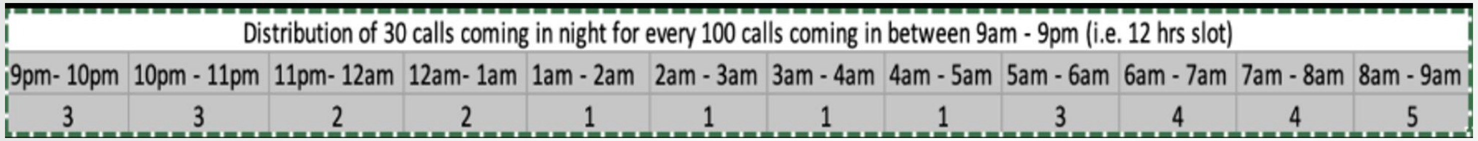
We can calculate Hours needed for 90%, (5129.9*198.6*0.9)/3600 .

We can calculate Total no. of agents required, (254.7/4.5) .

This analysis helps to know the total man power required/used and to distribute the work among the employees.

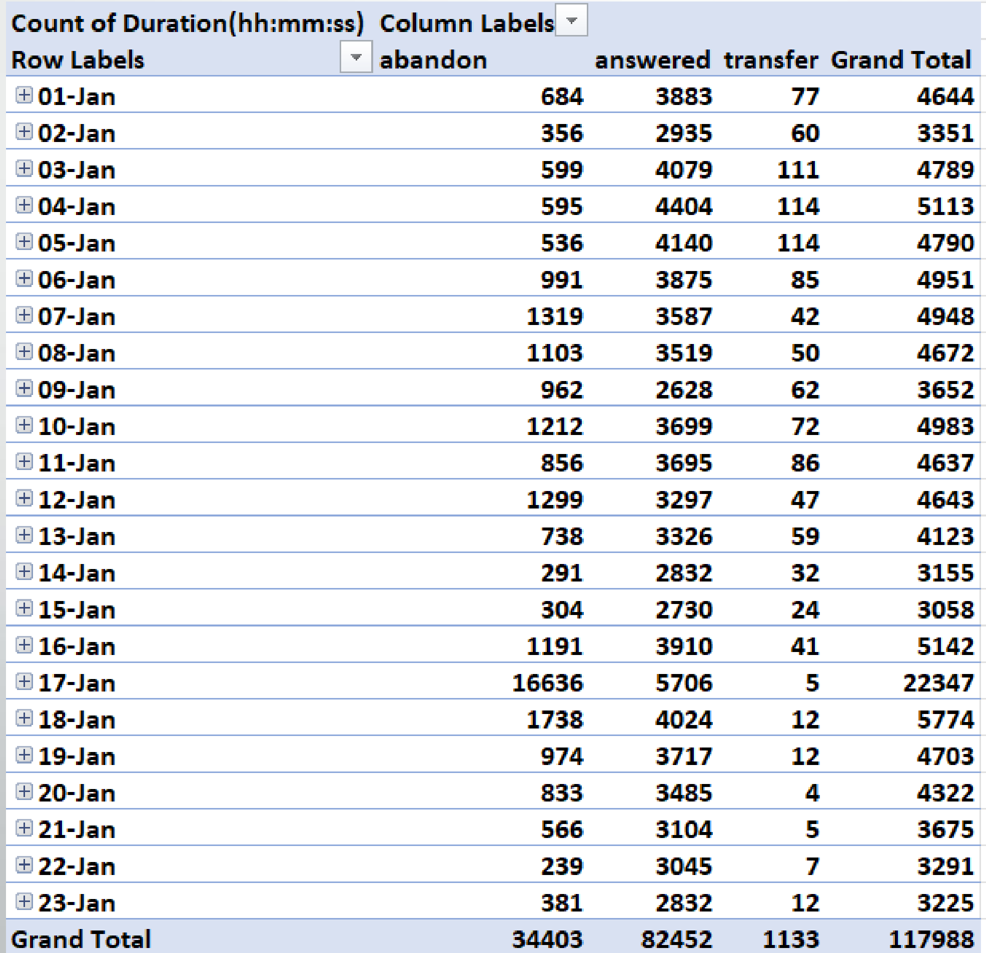
4.Night Shift Manpower Planning: Customers also call ABC Insurance Company at night but don't get an answer because there are no agents available. This creates a poor customer experience. Assume that for every 100 calls that customers make between 9 am and 9 pm, they also make 30 calls at night between 9 pm and 9 am. The distribution of these 30 calls is as follows:

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

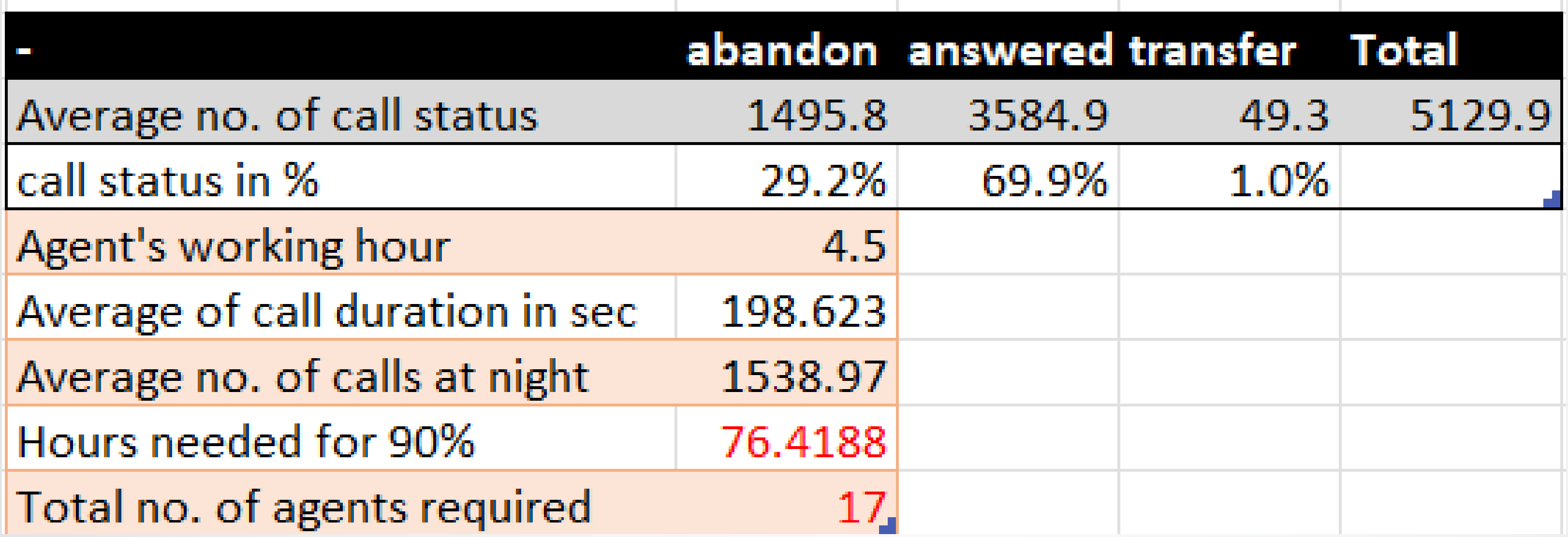


Distribution of 30 calls coming in night for every 100 calls coming in between 9am - 9pm (i.e. 12 hrs slot)

9pm- 10pm	10pm - 11pm	11pm- 12am	12am- 1am	1am - 2am	2am - 3am	3am - 4am	4am - 5am	5am - 6am	6am - 7am	7am - 8am	8am - 9am
3	3	2	2	1	1	1	1	3	4	4	5



Count of Duration(hh:mm:ss)	Column Labels			
Row Labels	abandon	answered	transfer	Grand Total
⊞ 01-Jan	684	3883	77	4644
⊞ 02-Jan	356	2935	60	3351
⊞ 03-Jan	599	4079	111	4789
⊞ 04-Jan	595	4404	114	5113
⊞ 05-Jan	536	4140	114	4790
⊞ 06-Jan	991	3875	85	4951
⊞ 07-Jan	1319	3587	42	4948
⊞ 08-Jan	1103	3519	50	4672
⊞ 09-Jan	962	2628	62	3652
⊞ 10-Jan	1212	3699	72	4983
⊞ 11-Jan	856	3695	86	4637
⊞ 12-Jan	1299	3297	47	4643
⊞ 13-Jan	738	3326	59	4123
⊞ 14-Jan	291	2832	32	3155
⊞ 15-Jan	304	2730	24	3058
⊞ 16-Jan	1191	3910	41	5142
⊞ 17-Jan	16636	5706	5	22347
⊞ 18-Jan	1738	4024	12	5774
⊞ 19-Jan	974	3717	12	4703
⊞ 20-Jan	833	3485	4	4322
⊞ 21-Jan	566	3104	5	3675
⊞ 22-Jan	239	3045	7	3291
⊞ 23-Jan	381	2832	12	3225
Grand Total	34403	82452	1133	117988



-	abandon	answered	transfer	Total
Average no. of call status	1495.8	3584.9	49.3	5129.9
call status in %	29.2%	69.9%	1.0%	
Agent's working hour	4.5			
Average of call duration in sec	198.623			
Average no. of calls at night	1538.97			
Hours needed for 90%	76.4188			
Total no. of agents required	17			

By assuming that a person works 7.5 hrs a day, 6 days in a week with 60% in engaging to answered calls.

By using the formula =(60/100)*7.5, with this we can determine the worker being in call i.e. 4.5 hrs a day

We have already calculated **average Call Duration in sec in Task 1.**

To calculate average no. of calls at night, using formula =(0.3*5129.9)

We can calculate Hours needed for 90%, (198.6*1538.9*0.9)/3600 .

We can calculate Total no. of agents required, (76.4/4.5) .

Insights:

1.Peak call volumes occur between 11 a.m. and 12 p.m., while the least activity is observed from 8 p.m. to 9 p.m.
2.Average call duration is highest between 10 a.m. and 11 a.m. and from 7 p.m. to 8 p.m., with the lowest durations recorded between 12 p.m. and 1 p.m.
3.A significant 30% of calls are abandoned during the day due to insufficient manpower, necessitating optimized staffing plans.

Results:

1.Identified optimal time slots for scheduling agents to reduce abandoned calls and improve service quality.
2.Proposed manpower allocation strategies to achieve a 90% call-answer rate during both day and night shifts.
3.Enhanced understanding of call patterns, enabling better customer experience and resource management.