Data Analysis Portfolio

PREPARED BY –SAKSHI RAHIWAL

Professional Background

Greetings! I am Sakshi Rahiwal, a dedicated and passionate professional with a strong foundation in Data Analyst. currently pursuing a Master of Computer Applications (MCA) from Uttaranchal University. With a strong foundational background, Sakshi holds a Bachelor of Science (B.Sc) in Mathematics, which has honed her analytical and problem-solving skills.

My academic journey has sparked a deep passion for data science. my mathematical expertise provides a solid base for understanding complex data structures and algorithms. I am particularly enthusiastic about harnessing the power of data to uncover insights, drive decision-making, and solve real-world problems.

I have good knowledge about Python , MS excel , Power BI , Tableau , Mysql to bring out valuable insight from data so organization can take business profitable decision.

I successfully completed my Trainity internship and grab the certifications also I am done pw skill one year data science course.

Thank you for taking the time to review my work.

Table Of Contents

Data Analytics Process

In the realm of grocery shopping, decision-making plays a pivotal role in ensuring efficiency and satisfaction. The importance of data analytics in this context cannot be overstated. By systematically analyzing past consumption patterns, dietary preferences, and budget constraints, individuals can make informed choices on essential and non-essential items. This data-driven approach optimizes the planning process, enhances resource allocation, and ultimately streamlines the overall grocery shopping experience. In a world inundated with choices, leveraging data analytics empowers individuals to make decisions that align with their needs, preferences, and financial goals, creating a more effective and personalized shopping journey.

The Problem

Your task is to give the example(s) of such a real-life situation where we use Data Analytics and link it with the data analytics process. You can prepare a PPT/PDF on a real-life scenario explaining it with the above process (Plan, Prepare, Process, Analyze, Share, Act) and submit it as part of this task.

Methodology

Exploring how data analytics is unconsciously applied in our everyday life, focusing on the process of making informed decisions while shopping for clothing.

1. Planning Phase
2. Preparation Phase
3. Processing Phase
4. Analyzing Phase
5. Sharing Phase
6. Acting Phase

Results

1. Planning Phase: We set the fashion agenda, strategically deciding what our wardrobe needs are, often without consciously acknowledging the analytical process at play.
2. Preparation Phase: By budgeting for style, we financially prepare ourselves for the upcoming fashion endeavors, aligning our preferences with available resources.
3. Processing Phase: The meticulous process of deciphering fashion data unfolds as we choose specific items, considering factors such as comfort, style, and current trends.
4. Analyzing Phase: Crafting the perfect ensemble requires a keen eye for analyzing color combinations, styles, and trends, ensuring our wardrobe is not just functional but also aesthetically pleasing.
5. Sharing Phase: Communication with the shopkeeper becomes crucial, as we seek assistance in finding the most suitable and trendy options, transforming our personal preferences into actionable choices.
6. Acting Phase: Finally, we bring our data-driven fashion choices to life by making purchases that align with our analyzed preferences and current fashion trends.

Conclusion

In essence, what may seem like a routine shopping experience is, in fact, a testament to the seamless integration of data analytics into our everyday lives. By embracing this process, we navigate the fashion landscape with a balance of personal style and the ever-evolving trends surrounding us. The data-driven fashion choices we make are a testament to our ability to adapt, analyze, and curate our wardrobes with both precision and flair.

As we continue to evolve in the digital age, let us appreciate and embrace the unspoken alliance between our fashion choices and the underlying data analytics that guide them. Fashion, once a realm driven purely by personal taste, is now a harmonious blend of individual expression and the subtle influence of data-driven insights. In this fusion, we discover the power of informed decision-making, making our fashion journey not just stylish but inherently intelligent.

"Unveiling the data-driven fashion choices that happen seamlessly in our everyday lives."

Instagram User Analytics

The objective of this project is to leverage SQL and MySQL Workbench for analyzing Instagram user data. The analysis aims to provide valuable insights to the product team at Instagram, enabling them to make informed decisions regarding user engagement, marketing strategies, and investor metrics.

Methodology

Steps for Designing this project are as follow –

1) Database Setup:
    - Executed provided commands to set up the necessary database for the project.
    - Ensured the integrity of the database structure and relationships.

2) Marketing Analysis:
    - Identified the five oldest users on Instagram using appropriate SQL queries and ORDER BY clauses.
    - Isolated users who have never posted a photo to target for promotional emails.
    - Determined the winner of the contest based on the most likes on a single photo.
    - Identified the top five most commonly used hashtags for the partner brand.

3) Ad Campaign Launch:

    - Analyzed user registration data to determine the best day of the week to launch ads.
    - Formulated queries to extract relevant information regarding user registration patterns.

4)Investor Metrics:

- Calculated the average number of posts per user to gauge user engagement.
- Identified potential bots by isolating users who liked every single photo on the site.

MySQL Workbench :

Chose MySQL Workbench for its user-friendly interface, robust query execution, and compatibility with MySQL databases.

Finding – 1

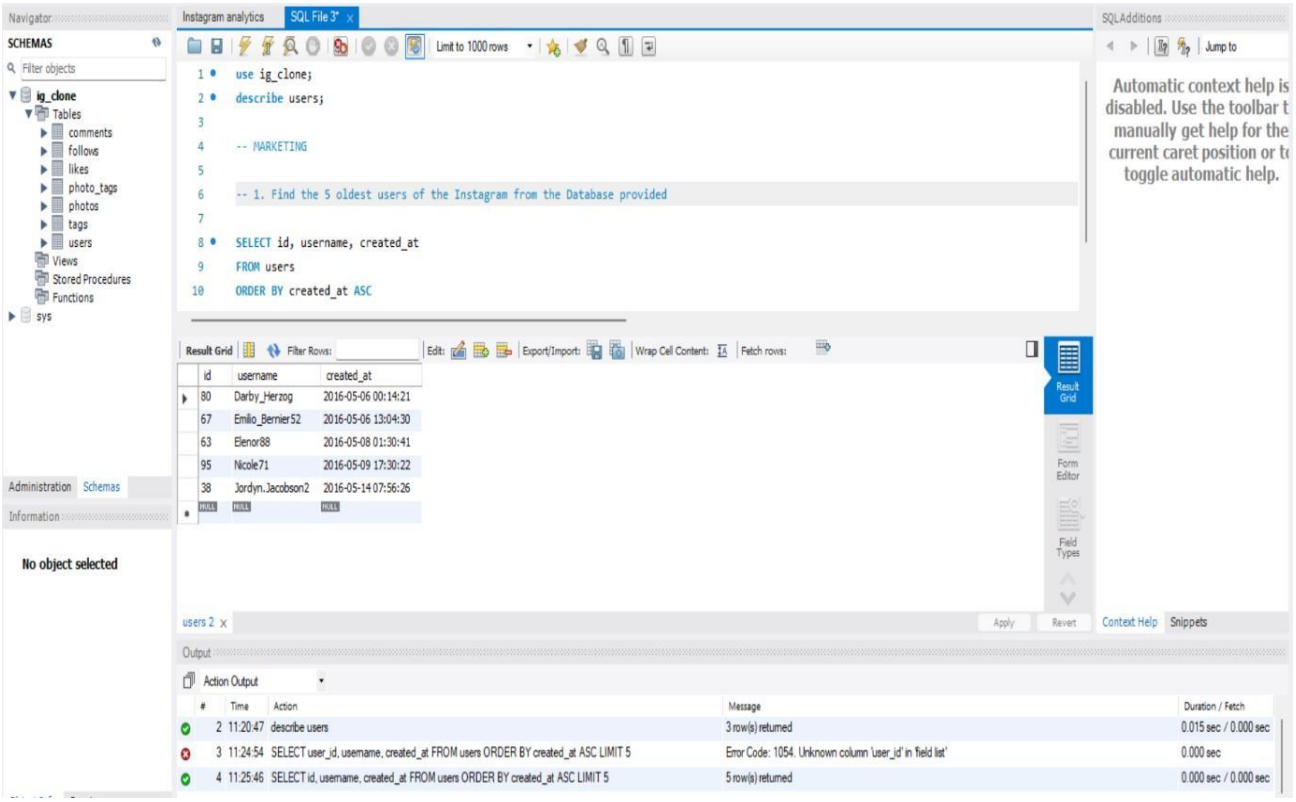


Fig. 1 – Loyal User

Loyal User Reward:

Users Darby_Herzog and Emmo_Bernier52 have been on the platform the longest, providing insights into user loyalty.

Finding -2

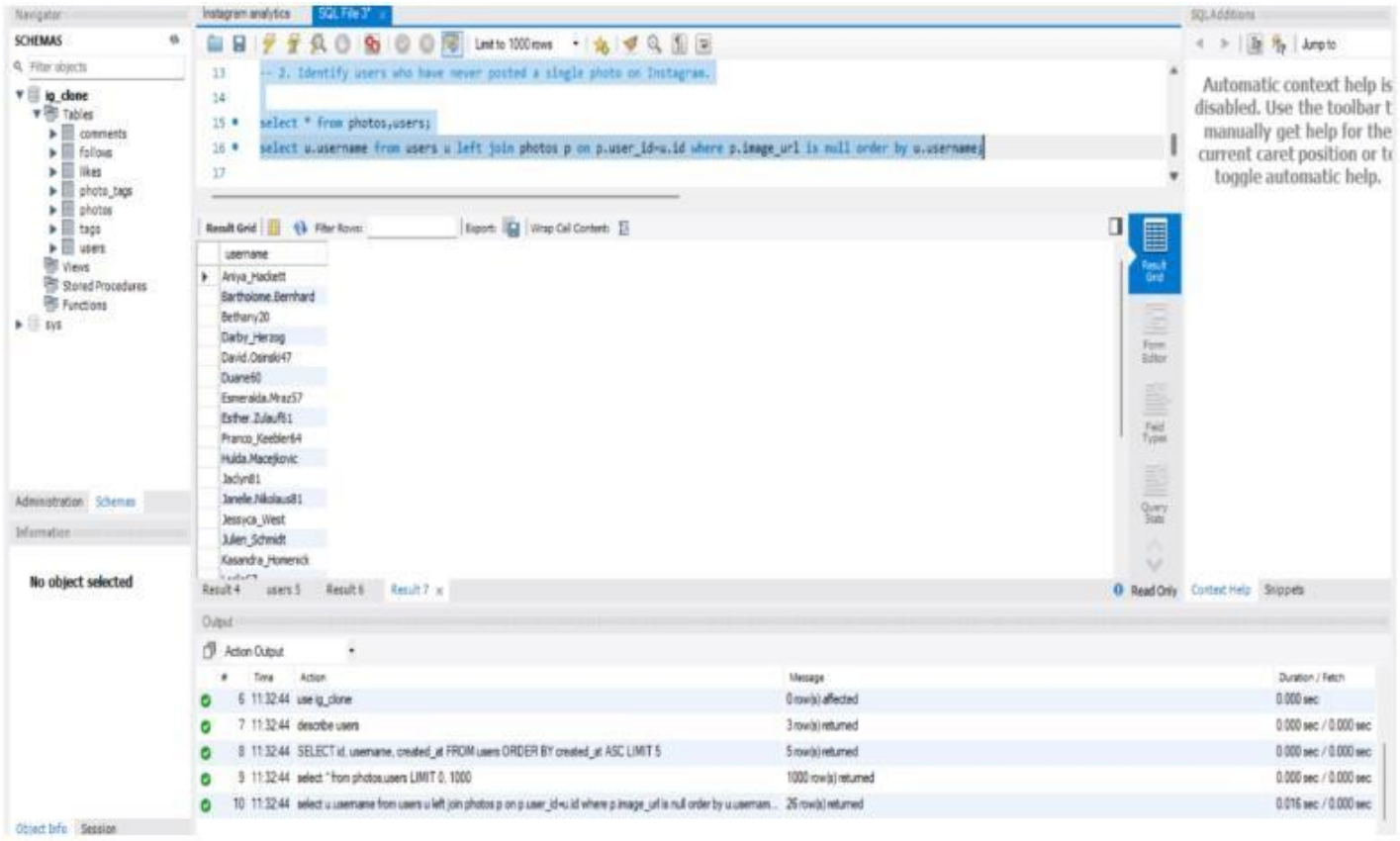


Fig. 2 - Inactive User

Inactive User Engagement:

These users have never posted, enabling targeted efforts to re-engage them.

Finding – 3

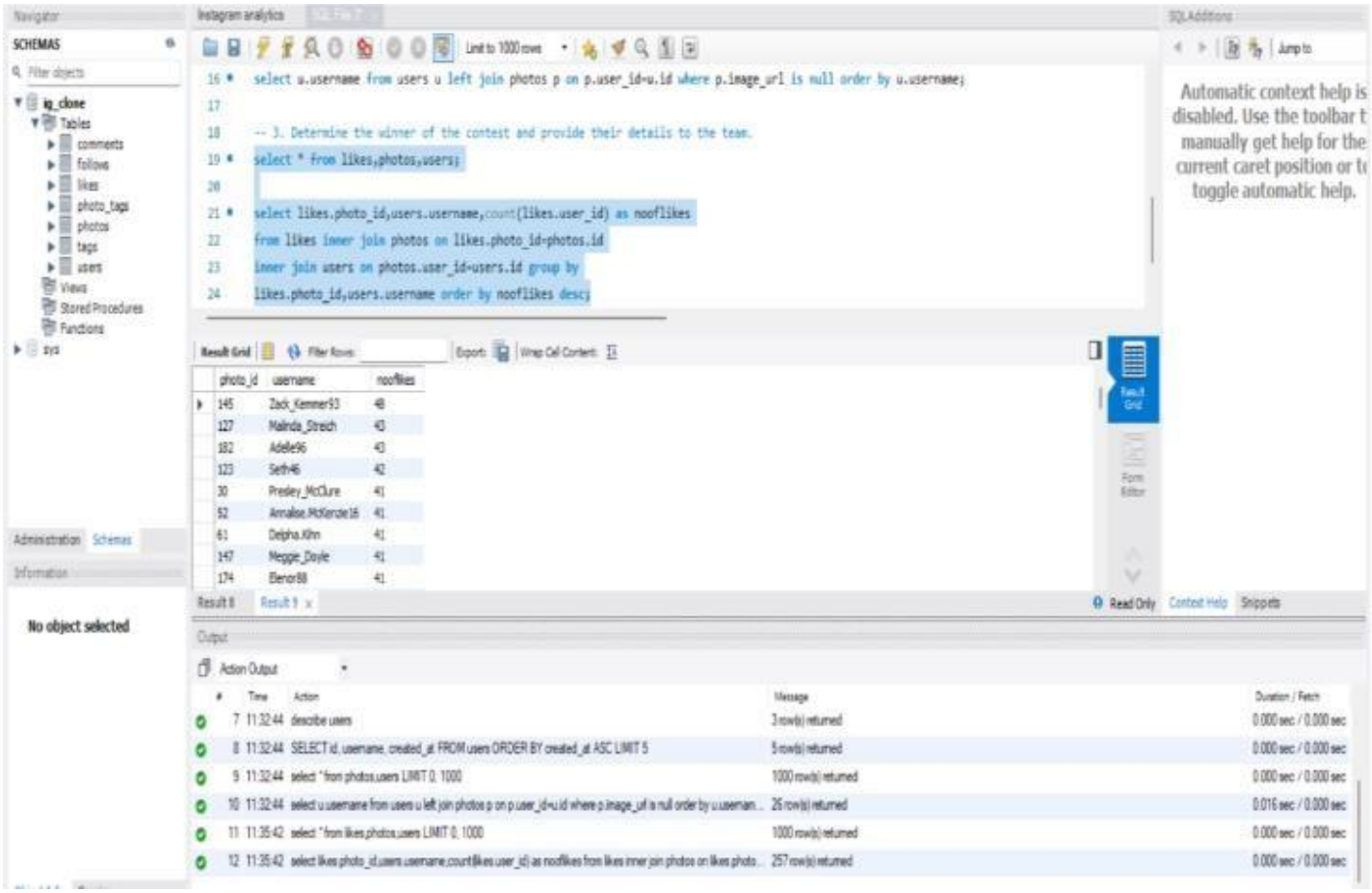


Fig. – 3 Contest Winner

Contest Winner Declaration:

They are winner of the contest based on likes, assisting in the reward process.

Finding – 4

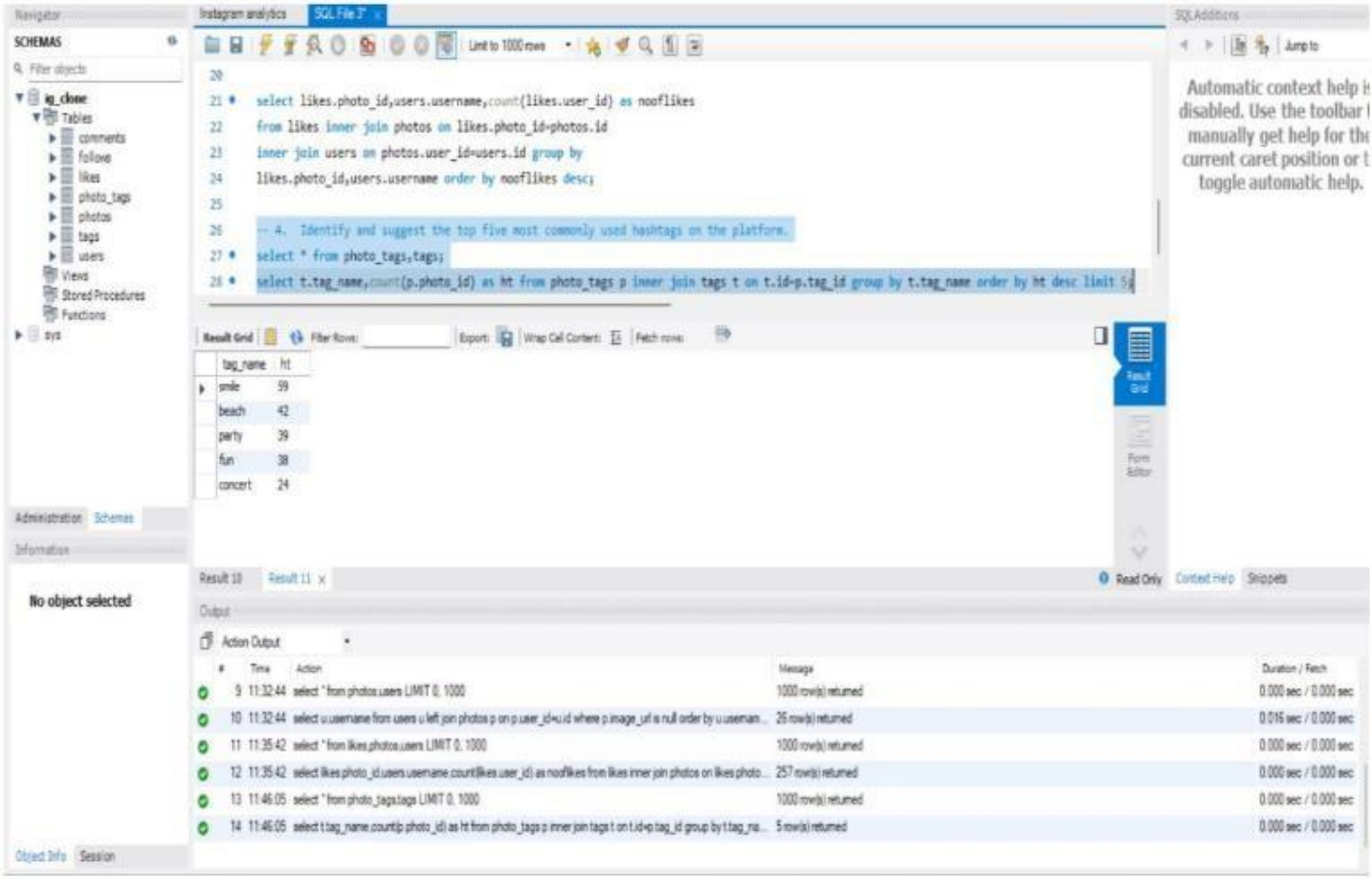


Fig 4 - Hashtag Research

Hashtag Research:

These insights have popular hashtags, aiding the partner brand in reaching a wider audience.

Finding – 5

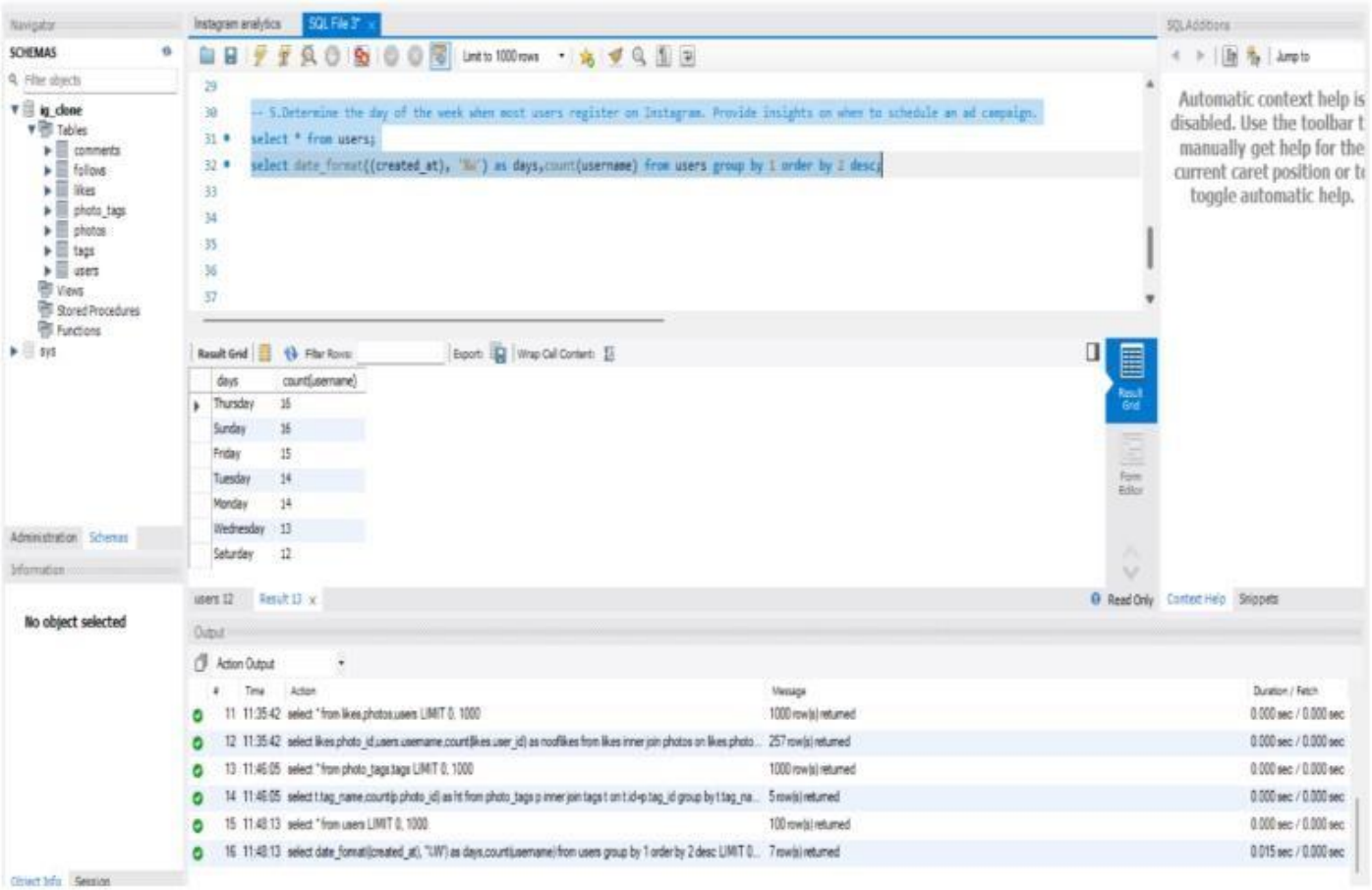


Fig 5 - Ad Campaign Launch

Ad Campaign Launch:

Thursday and Sunday is the best day for ad campaigns based on user registration trends.

Finding – 6

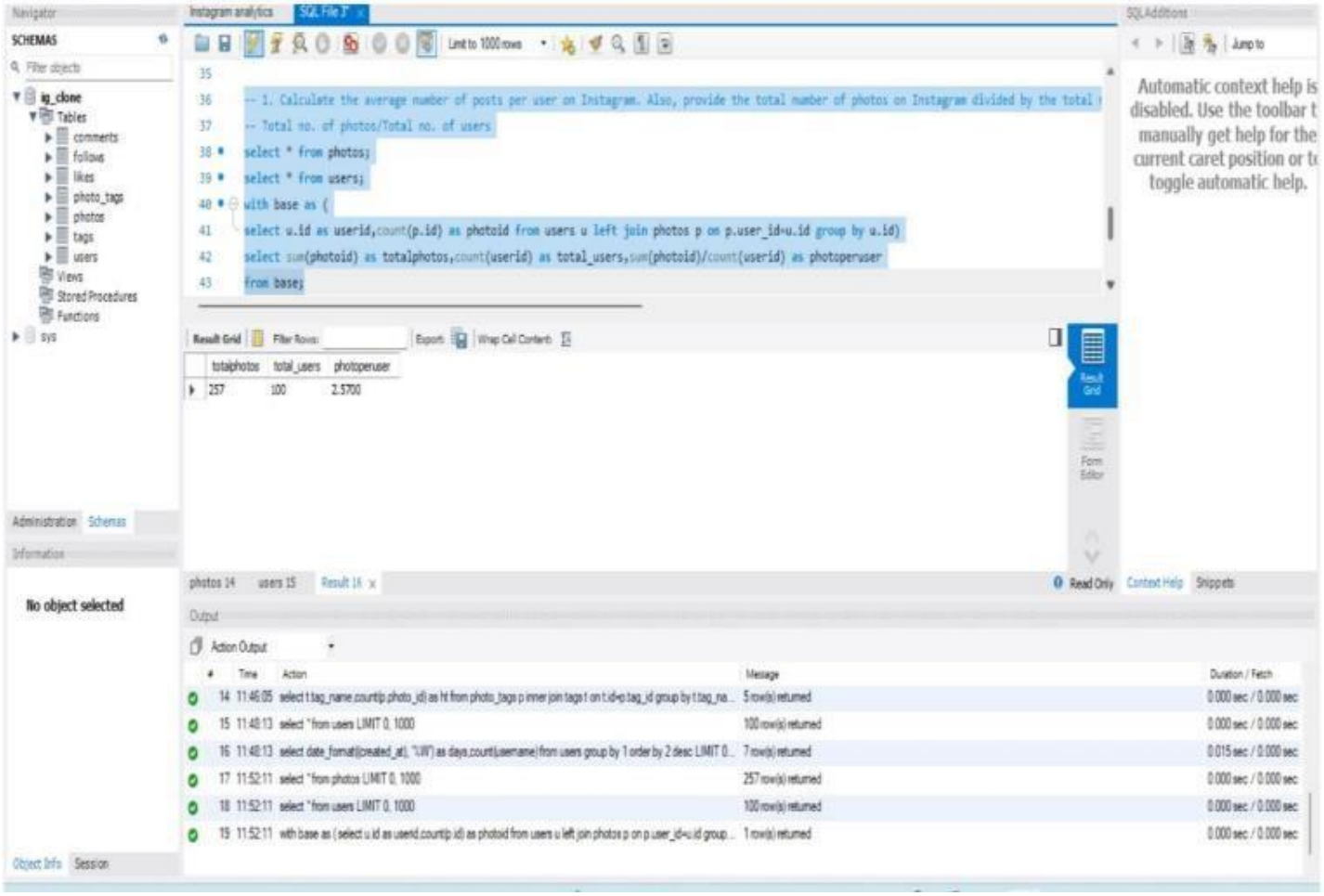


Fig 6 - User Engagement

User Engagement:

2.57 average posts per user, providing a metric for overall user engagement.

Finding – 7

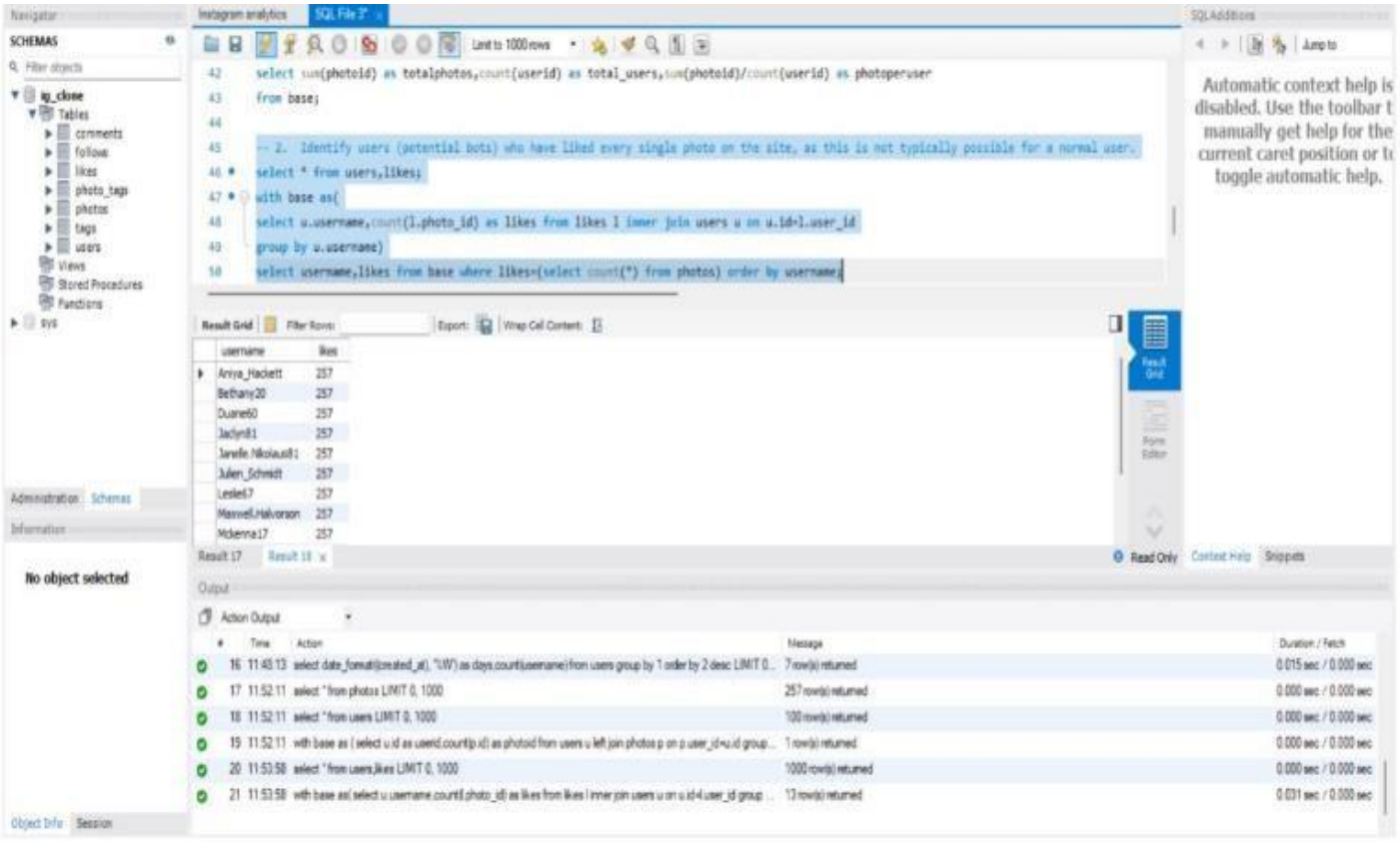


Fig 7 - Bots & Fake Accounts

Bots & Fake Accounts:

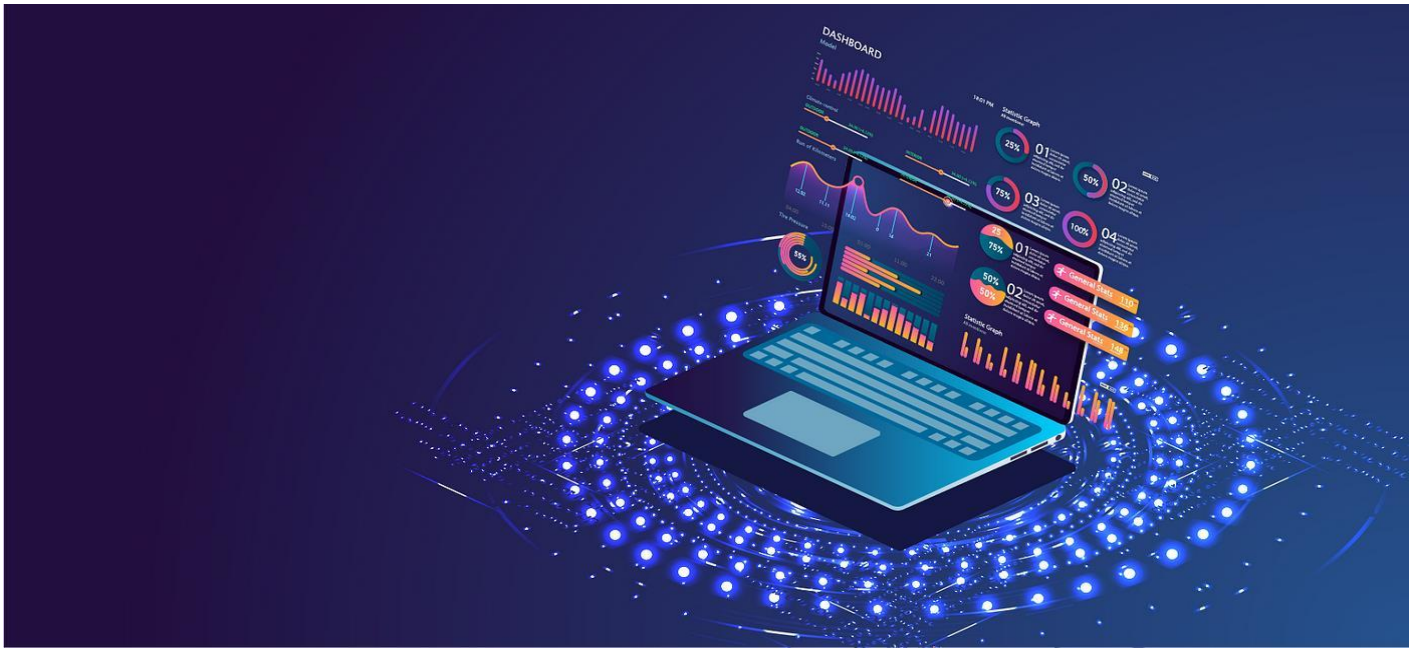
13 User are potential bots, helping investors assess the platform's authenticity.

Results

1. Successfully identified loyal users, potential contest winners, inactive users, and popular hashtags.
2. Provided valuable insights for strategic marketing decisions and optimizing ad campaign launches.
3. Enhanced understanding of user engagement and potential bot activity on the platform.

Conclusion

This analysis equips the Instagram team with actionable insights, fostering data-driven decision-making. The project contributes to the enhancement of user experience, targeted marketing efforts, and transparency for investors regarding platform authenticity. The findings have the potential to positively influence the growth and sustainability of Instagram as one of the leading social media platforms.



Operation Analytics and Investigating Metric Spike

The Operational Analytics project centers on leveraging SQL to analyze a company's comprehensive operations, aligning with the responsibilities of a Lead Data Analyst at a company akin to Microsoft. The primary objectives involve extracting valuable insights from provided datasets to optimize company operations and comprehend abrupt shifts in critical metrics. The Lead Data Analyst engages in tasks such as calculating the number of jobs reviewed per hour, analyzing throughput trends, examining language share dynamics, identifying duplicate data entries, and investigating user engagement and growth patterns. By delving into these facets, the project aims to empower decision-makers with actionable information, fostering a data-driven approach to enhance overall operational efficiency and address fluctuations in key performance indicators.

Methodology

Database Setup: Initiate the project by creating a MySQL database and necessary tables based on the provided structures. Import relevant CSV files into MySQL Workbench to populate the tables.

Analysis Execution: Utilize advanced SQL skills to perform analysis tasks outlined in Case Study 1 and Case Study 2. Understand the table structures and meanings of columns to derive meaningful insights.

Query Optimization: Optimize SQL queries for performance, considering the volume of data and the need for real-time analysis.

Documentation: Document the queries, including their purpose and any assumptions made during the analysis.

Tech-Stack Used:

MySQL Workbench: Used for database creation, table management, and executing SQL queries.

CSV Files: Imported into MySQL Workbench for data population.

SQL: Employed for data analysis, including querying, aggregating, and filtering.

Finding – 1

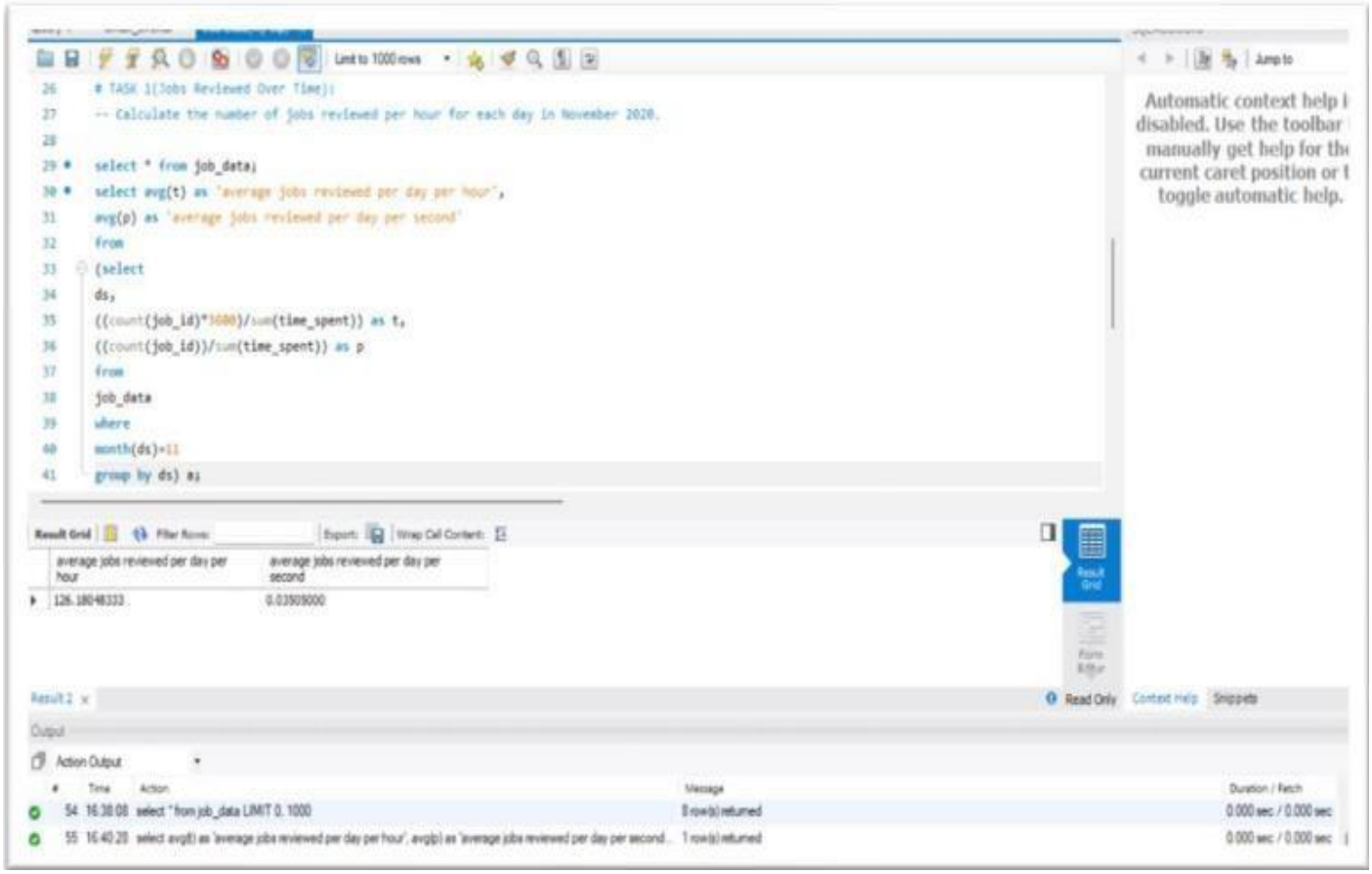
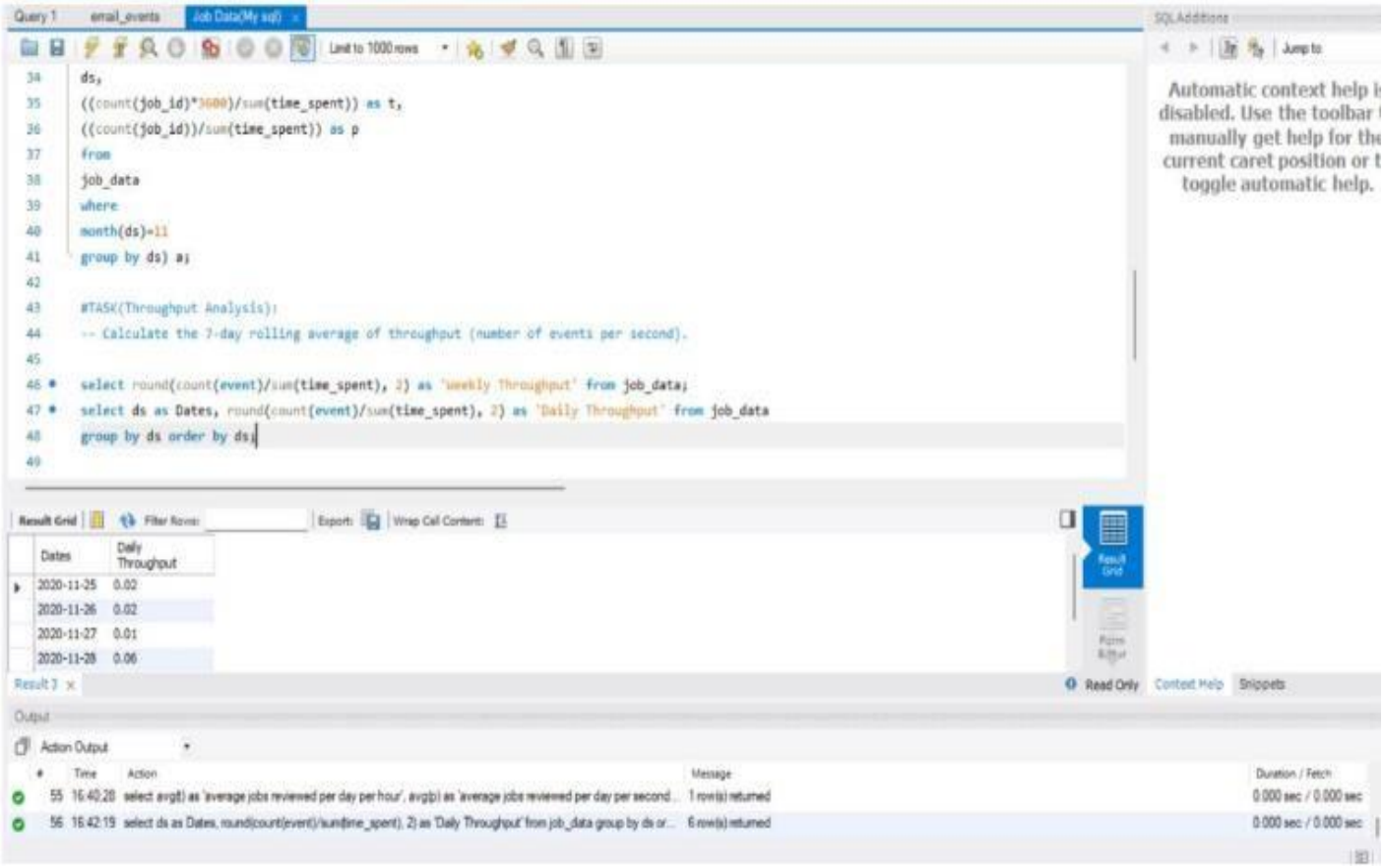


Fig. 1 - Jobs Reviewed Over Time

Jobs Reviewed Over Time:

- 126.18 have peak hours of job reviews each day in November 2020.
- This insight can guide resource allocation for job review processes.

Finding – 2



```
ds,
((count(job_id)*3600)/sum(time_spent)) as t,
((count(job_id))/sum(time_spent)) as p
from
job_data
where
month(ds)=11
group by ds) a;

#TASK(Throughput Analysis):
-- Calculate the 7-day rolling average of throughput (number of events per second).

select round(count(event)/sum(time_spent), 2) as 'weekly Throughput' from job_data;
select ds as Dates, round(count(event)/sum(time_spent), 2) as 'Daily Throughput' from job_data
group by ds order by ds;
```

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

Fig.2 - SQL Query for 7 days rolling average of throughput

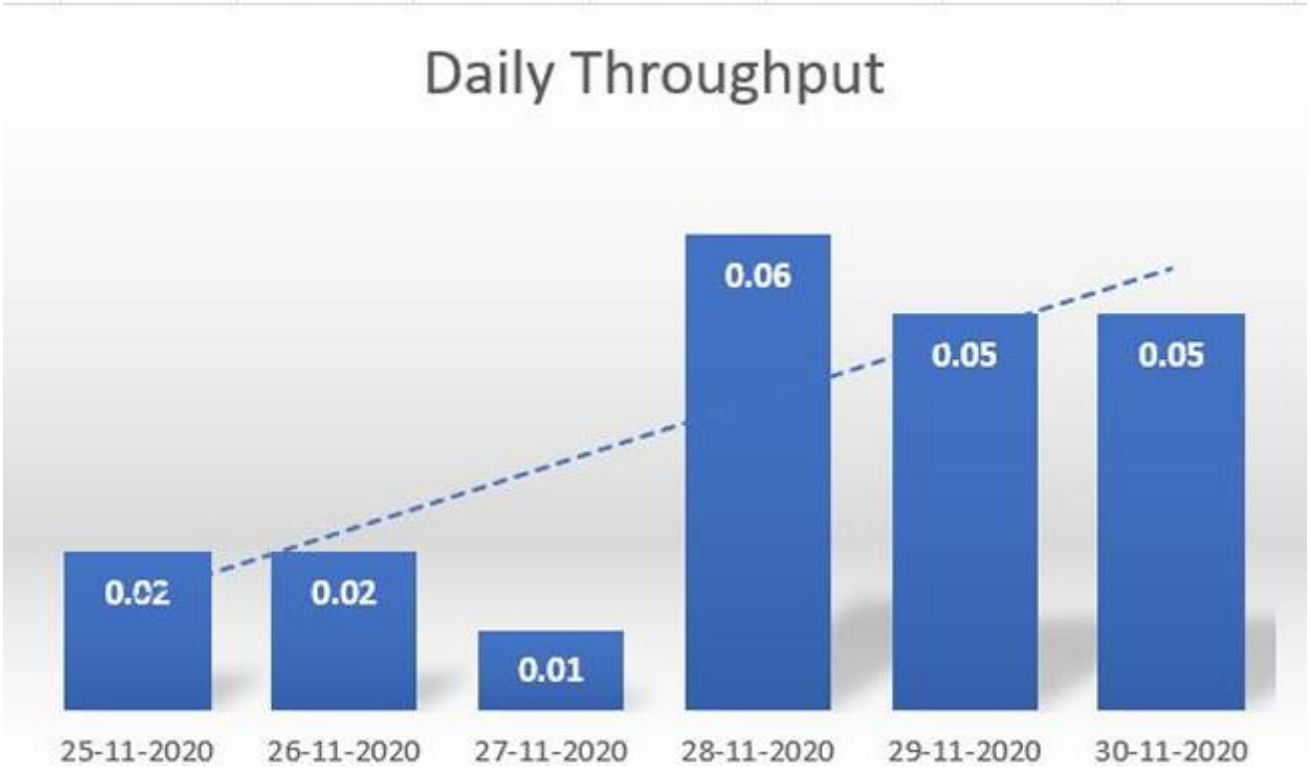


Fig.3 – Daily Throughput

Throughput Analysis:

- Calculated 7-day rolling average of throughput to smooth out daily fluctuations.
- Preferred the rolling average as it provides a more stable representation of trends, reducing the impact of outliers.

Finding – 3

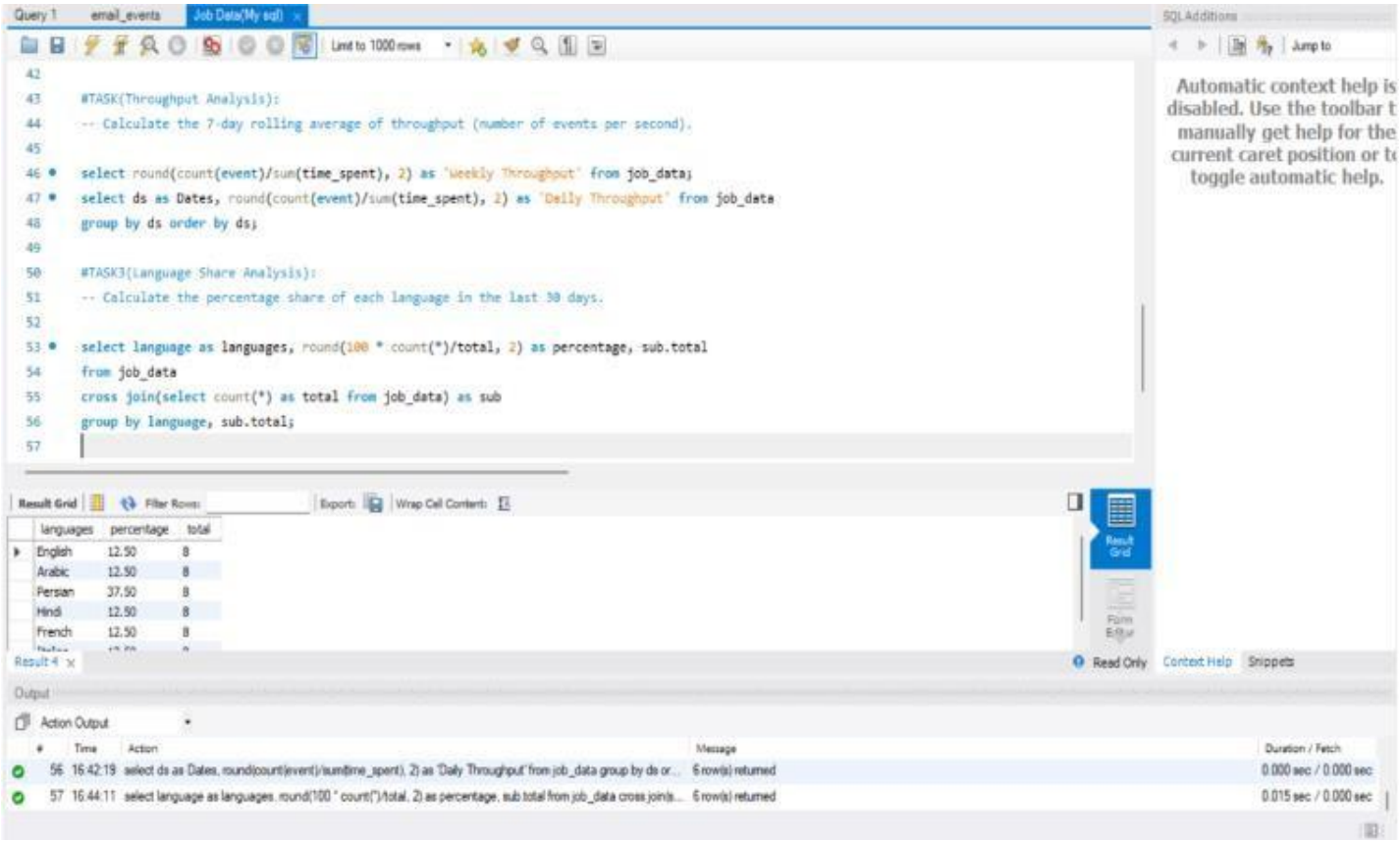


Fig.4 – SQL Query for percentage share of each language in last 30 days

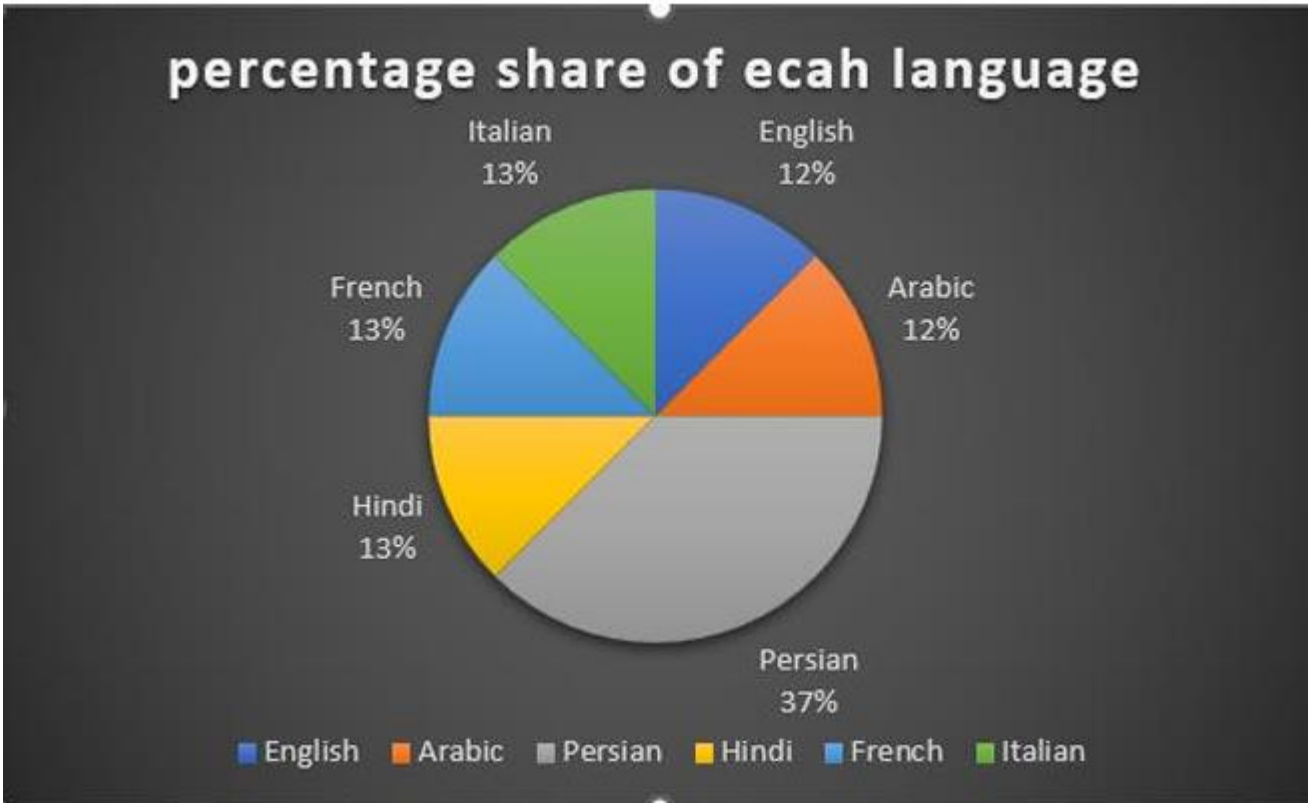


Fig. 5 – Insight for percentage share of each language

Language Share Analysis:

- In figure 4 there are percentage share of each language over the last 30 days.
- Persian is the language preferences and potentially adapting content strategy.

Finding – 4

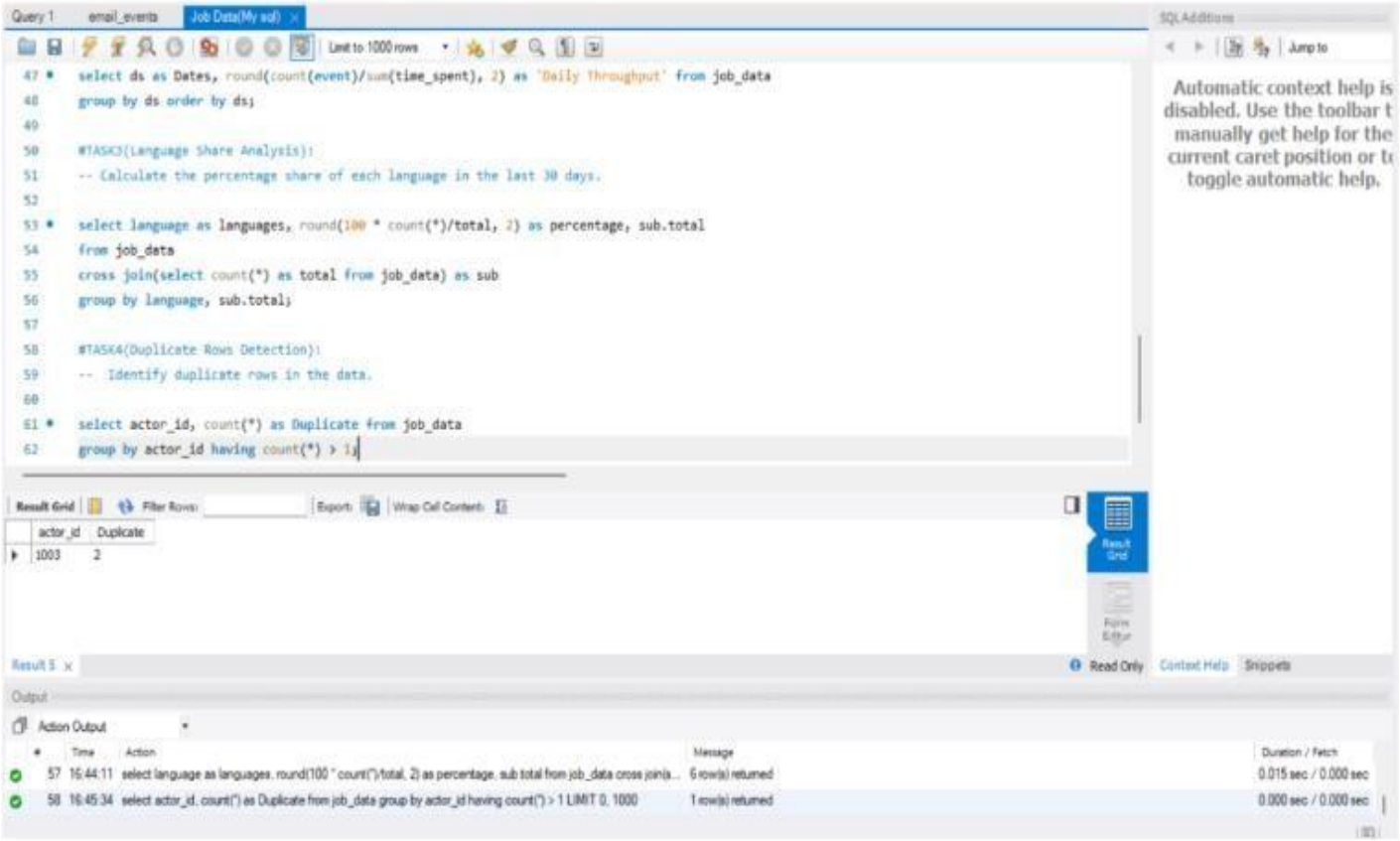
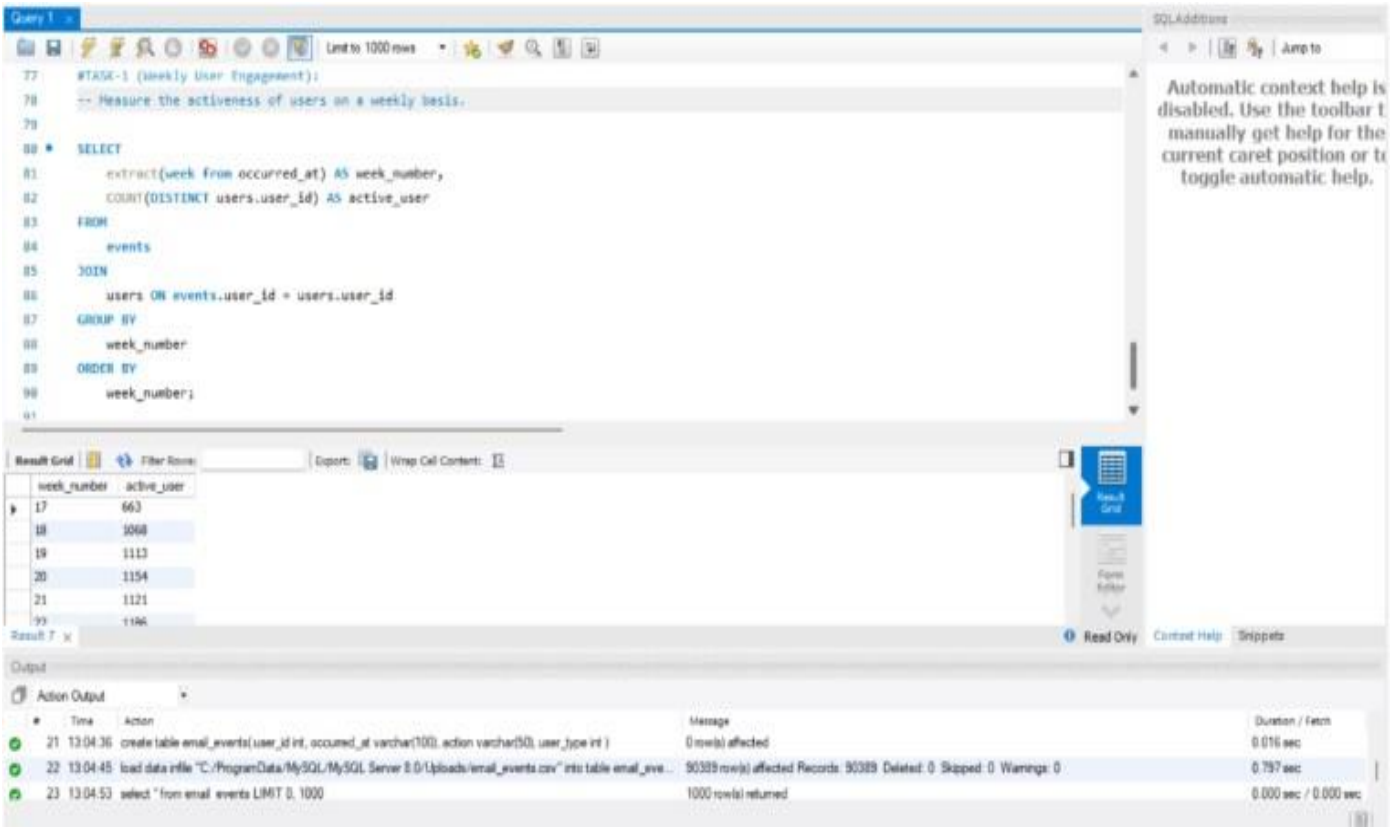


Fig.6 - Duplicate Rows Detection

Duplicate Rows Detection:

- Detected and displayed duplicate rows in the job_data table.
- This insight aids in data quality assurance and potential process improvements.

Finding – 5



```
#TASK-1 (Weekly User Engagement):
-- Measure the activeness of users on a weekly basis.

SELECT
    extract(week from occurred_at) AS week_number,
    COUNT(DISTINCT users.user_id) AS active_user
FROM
    events
JOIN
    users ON events.user_id = users.user_id
GROUP BY
    week_number
ORDER BY
    week_number;
```

week_number	active_user
17	663
18	1068
19	1113
20	1154
21	1121

Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or t toggle automatic help.

Fig.7 - Weekly User Engagement

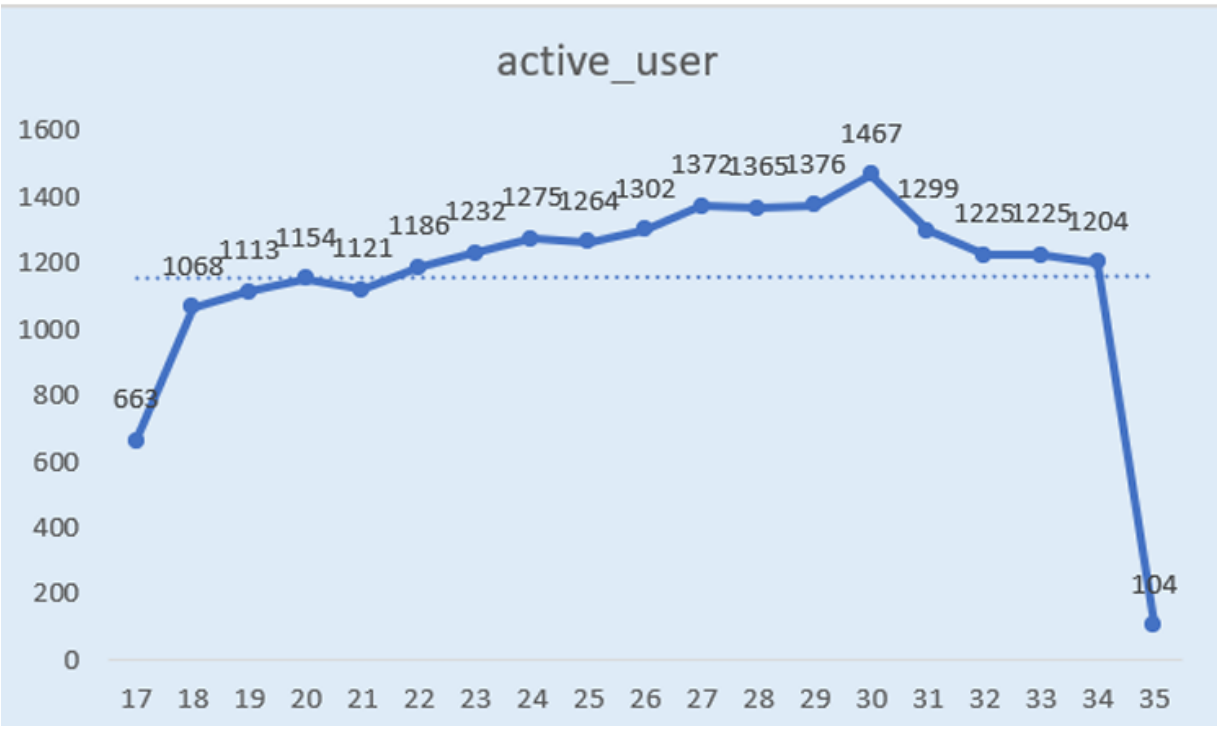


Fig.8 – Insightful user activity patterns and identifying potential trends or dips

Weekly User Engagement:

- Week no. 30 have highest measured user engagement on a weekly basis.
- Insightful for understanding user activity patterns and identifying potential trends or dips.

Finding – 6

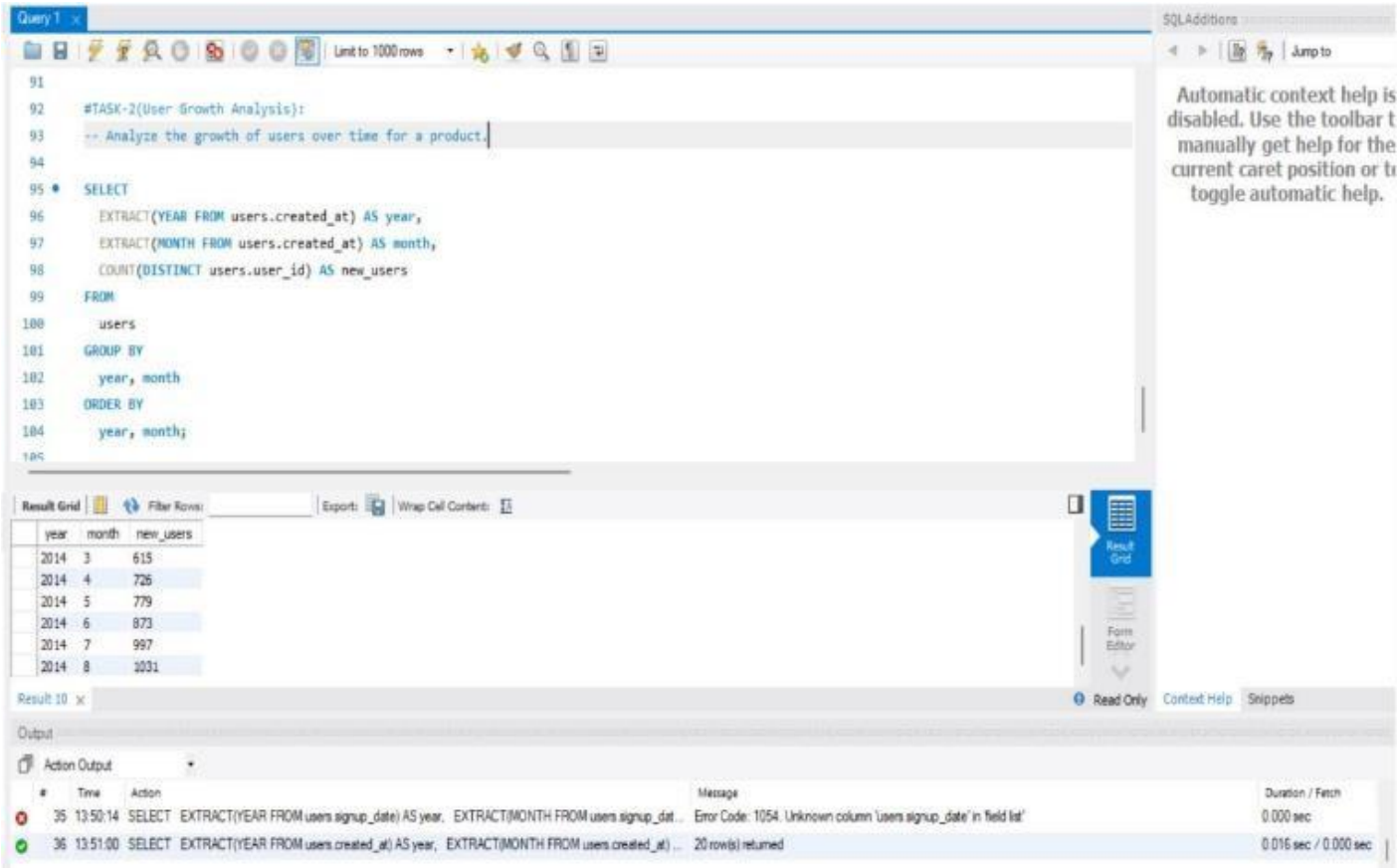


Fig.9 – SQL Query for User Growth Analysis

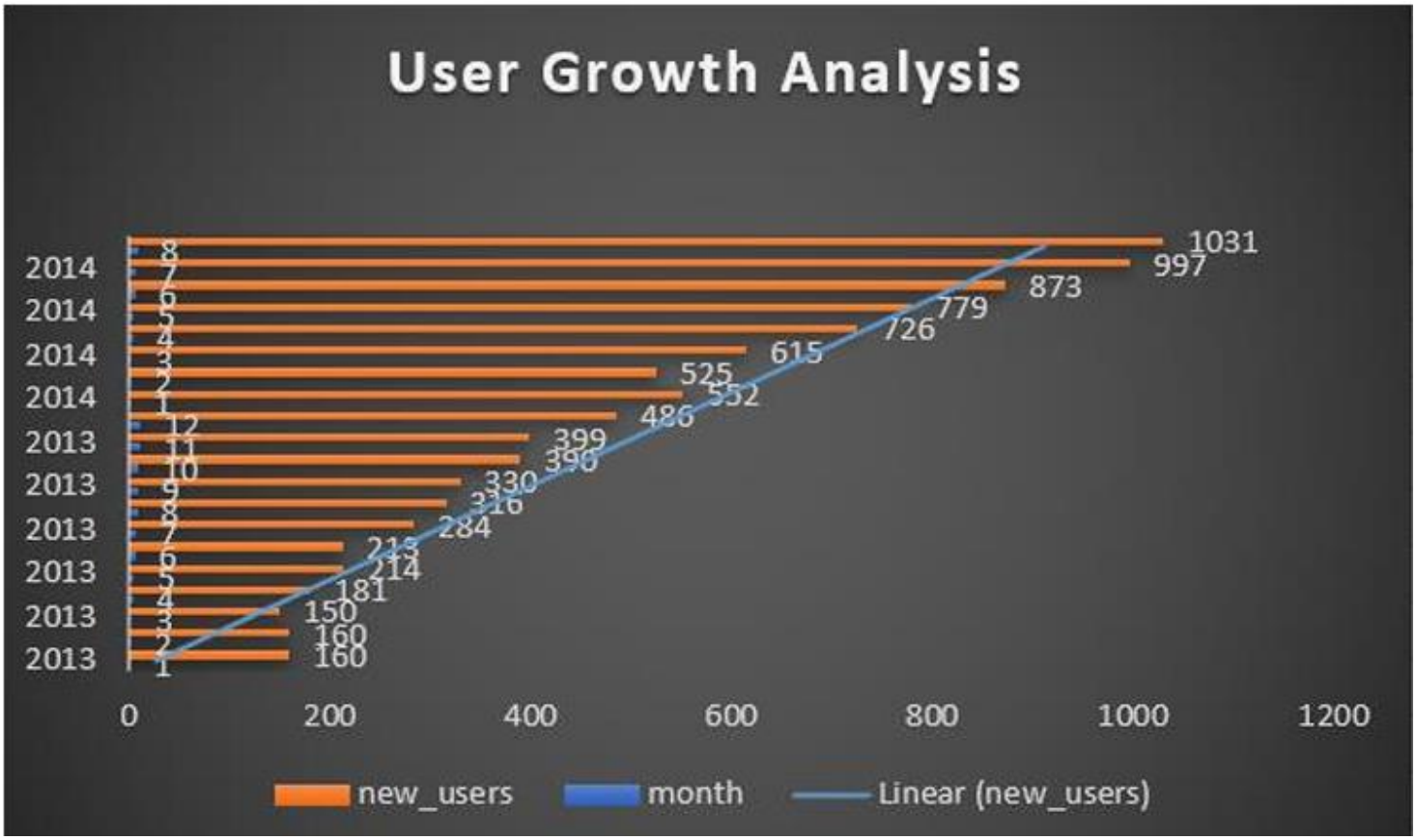
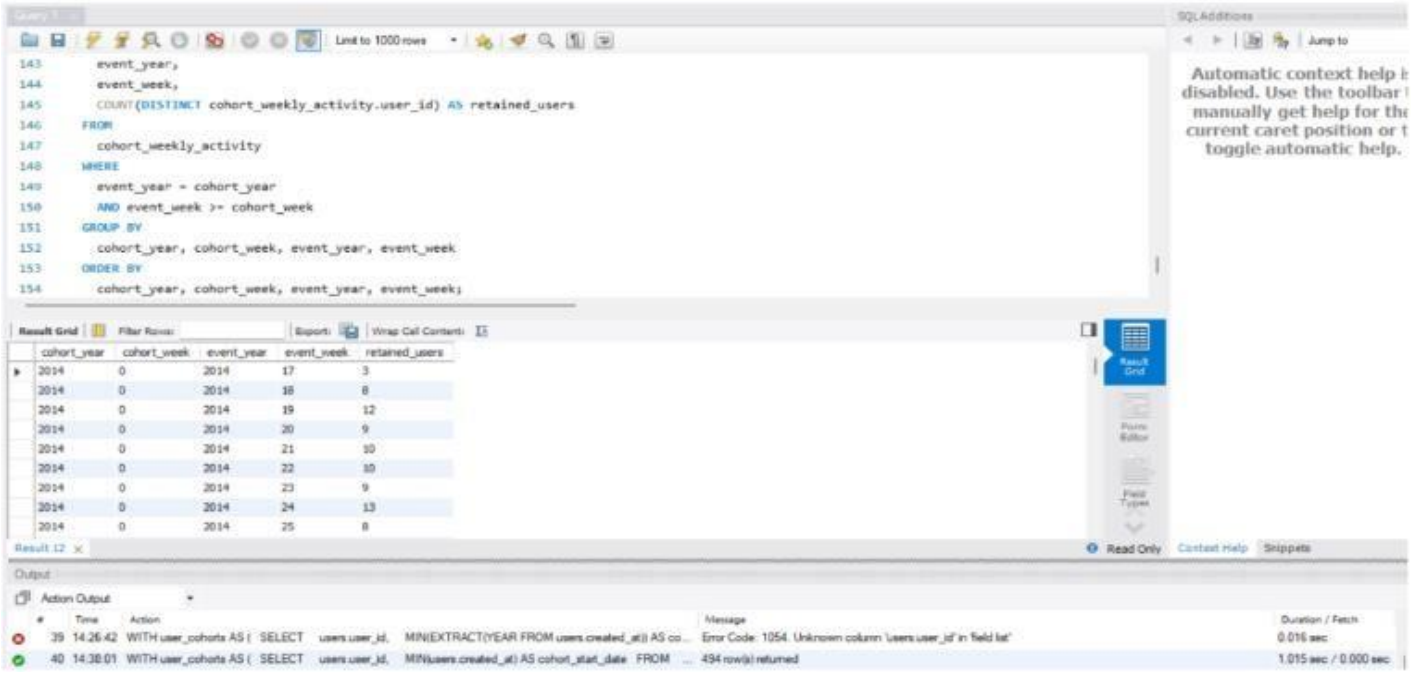


Fig.10 – Insight b/w month and new_user

User Growth Analysis:

- Analyzed the growth of users over time.
- Helpful for strategic planning and resource allocation based on user adoption trends.

Finding – 7



```
143     event_year,
144     event_week,
145     COUNT(DISTINCT cohort_weekly_activity.user_id) AS retained_users
146   FROM
147     cohort_weekly_activity
148   WHERE
149     event_year = cohort_year
150     AND event_week >= cohort_week
151   GROUP BY
152     cohort_year, cohort_week, event_year, event_week
153   ORDER BY
154     cohort_year, cohort_week, event_year, event_week;
```

cohort_year	cohort_week	event_year	event_week	retained_users
2014	0	2014	17	3
2014	0	2014	18	8
2014	0	2014	19	12
2014	0	2014	20	9
2014	0	2014	21	10
2014	0	2014	22	10
2014	0	2014	23	9
2014	0	2014	24	13
2014	0	2014	25	8

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

#	Time	Action	Message	Duration / Fetch
39	14:26:42	WITH user_cohorts AS (SELECT users.user_id, MIN(EXTRACT(YEAR FROM users.created_at)) AS co...	Error Code: 1054. Unknown column 'users.user_id' in 'field list'	0.016 sec
40	14:38:01	WITH user_cohorts AS (SELECT users.user_id, MIN(users.created_at) AS cohort_start_date FROM ...	494 row(s) returned	1.015 sec / 0.000 sec

Fig.11 - SQL Query for Weekly Retention Analysis

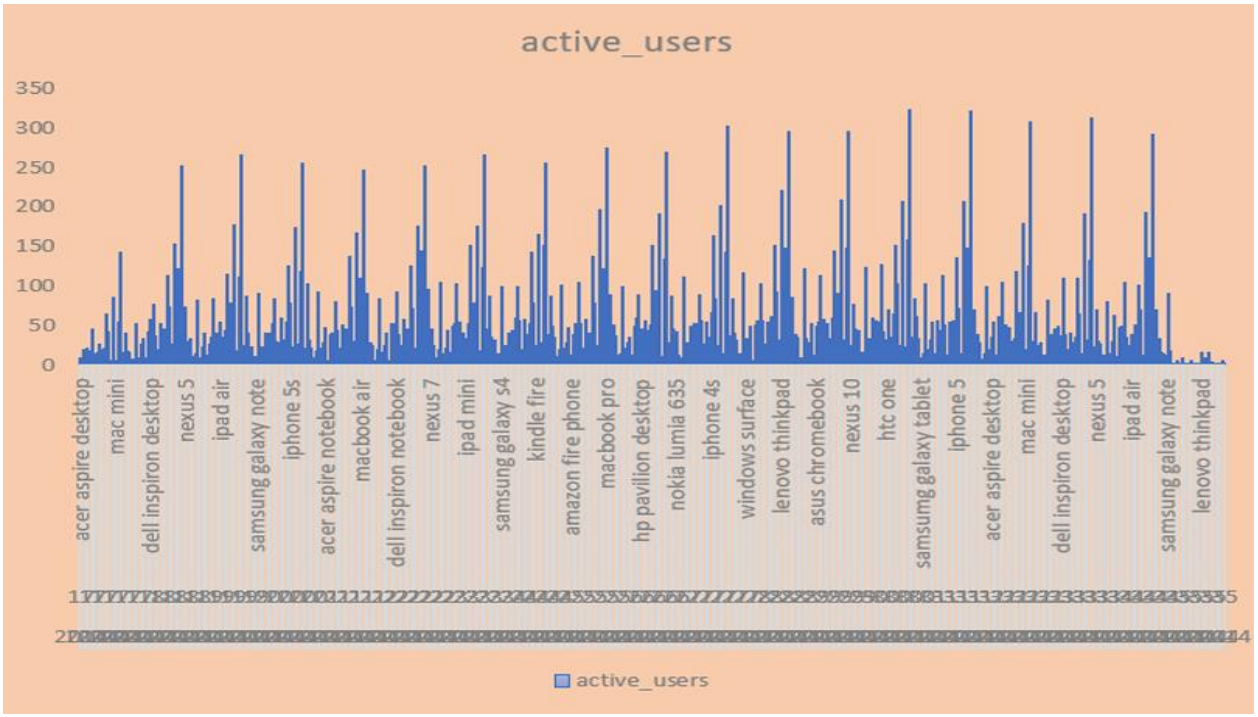


Fig.12 – Insight for Weekly Retention Analysis

Weekly Retention Analysis:

- Calculated weekly retention rates based on user sign-up cohorts.
- Provides insights into user loyalty and product stickiness.

Finding – 8

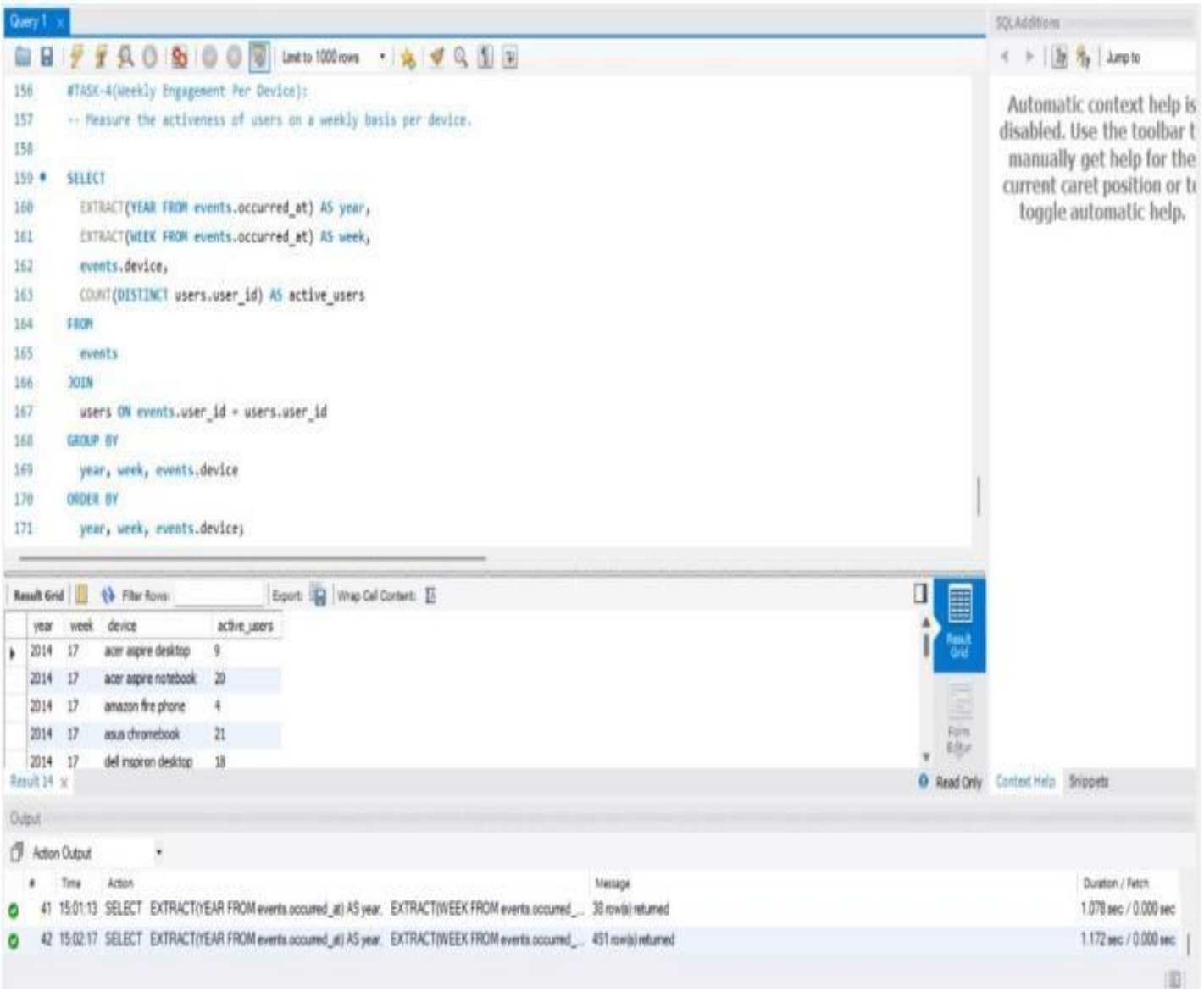


Fig.13 – SQL Query Weekly Engagement Per Device

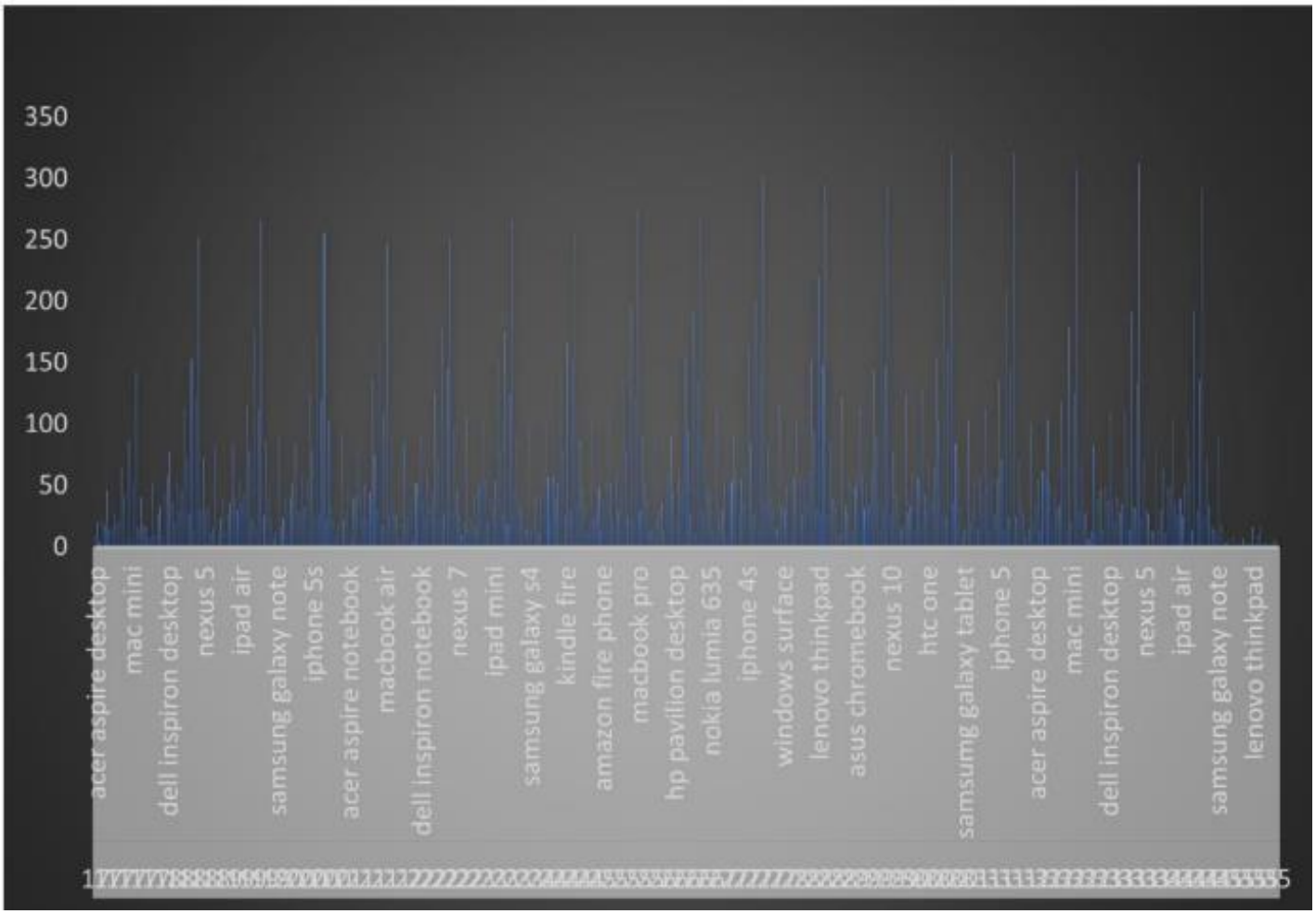


Fig.14 – Insight for Weekly Engagement Per Device

Weekly Engagement Per Device:

- Measured user engagement on a weekly basis per device.
- Useful for tailoring user experiences based on device preferences.

Finding – 9

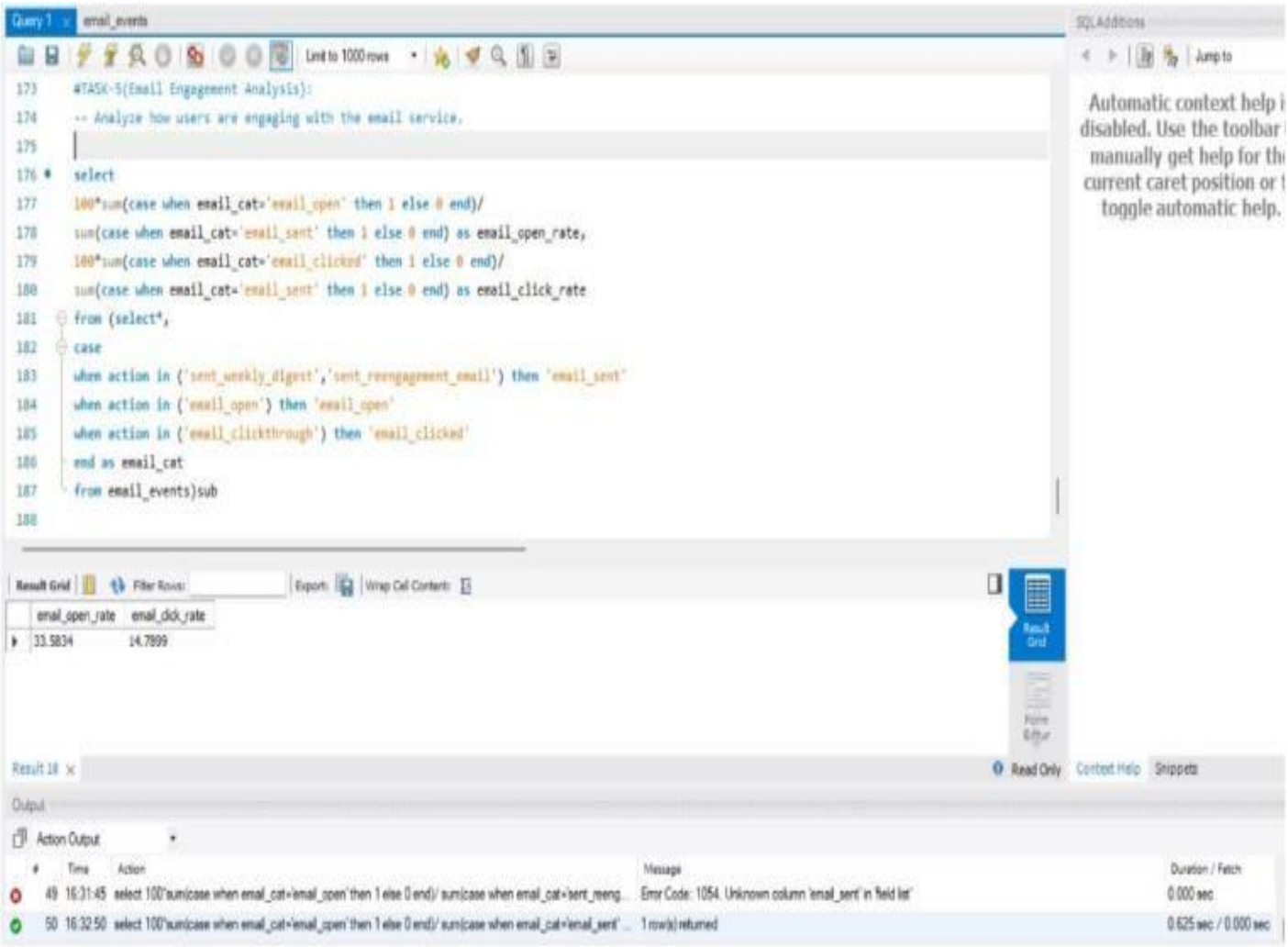


Fig.15 – SQL Query for Email Engagement Analysis

Email Engagement Analysis:

- Explored how users engage with the email service.
- Valuable for optimizing email campaigns and improving user communication.

Result

The successful execution of SQL queries in this project yielded actionable insights crucial for decision-making across various domains within the company. The derived information played a pivotal role in shaping strategic decisions related to operations, marketing, and user engagement. By harnessing SQL analytics, the Lead Data Analyst contributed significantly to the enhancement of the company's overall comprehension of its operations and key metrics. The insights obtained paved the way for informed decision-making processes, allowing the organization to adapt and optimize its strategies in response to dynamic operational scenarios. This outcome underscores the efficacy of leveraging advanced analytical techniques in extracting meaningful intelligence from data, ultimately empowering the company to make well-informed and impactful choices in its pursuit of operational excellence and business success.

Conclusion

In conclusion, this project exemplifies the adept use of SQL for operational analytics and metric investigation. The insights derived from the analysis not only offer immediate value for decision-making across various business functions but also establish a robust foundation for ongoing data-driven strategies. The documented queries and results, serving as a comprehensive reference, position the organization for future analyses and decision support. The successful application of SQL in this context underscores its pivotal role in unravelling valuable insights from complex datasets, ultimately contributing to the holistic improvement of business operations. As the company moves forward, the knowledge gained from this project equips decision-makers with a powerful tool to navigate challenges, optimize processes, and foster a culture of continual improvement grounded in data-driven principles.

Hiring Process Analytics

This project centres on a comprehensive analysis of the hiring process data of a multinational company, akin to Google. The primary goal is to extract actionable insights that can optimize the efficiency of the company's hiring procedures. The dataset, encompassing records of previous hires, serves as the foundation for uncovering valuable information related to gender distribution, salary statistics, departmental composition, and position tier distribution.

- Understand the gender distribution of hires.
- Calculate the average salary offered by the company.
- Create salary distribution class intervals.
- Visualize departmental composition.
- Represent position tiers through charts/graphs.

Methodology

Handling Missing Data:

- Identified missing values in the dataset.
- Imputed missing values using appropriate strategies (mean, median, mode).

Clubbing Columns:

- Combined columns with multiple categories to simplify analysis.
- Ensured consistency in data representation.

Outlier Detection and Removal:

- Identified outliers using statistical methods (z-score, IQR).
- Decided on the strategy to handle outliers based on the context (removal, replacement, or retention).

Data Summary:

- Calculated averages, medians, and other statistical measures.
- Created visualizations using Excel functions and charts.

Software:

Microsoft Excel 2022:

  - Used for data cleaning, analysis, and visualization.
  - Excel functions (e.g., VLOOKUP, IF, AVERAGE) employed for calculations.
  - Charts and graphs (e.g., pie chart, bar graph) utilized for visual representation.

Finding – 1

Fig.1 Hiring Analysis

Gender Distribution:

- Discovered the gender distribution of hires.
- Identified potential gender imbalances, providing a basis for diversity and inclusion strategies.
- The company has hired 2563 males and 1856 females, indicating a gender distribution.

Finding – 2

Average Salary:

- 49983.03 is the average salary offered by the company.
- Insightful for budgeting, salary negotiations, and benchmarking against industry standards.

Finding – 3

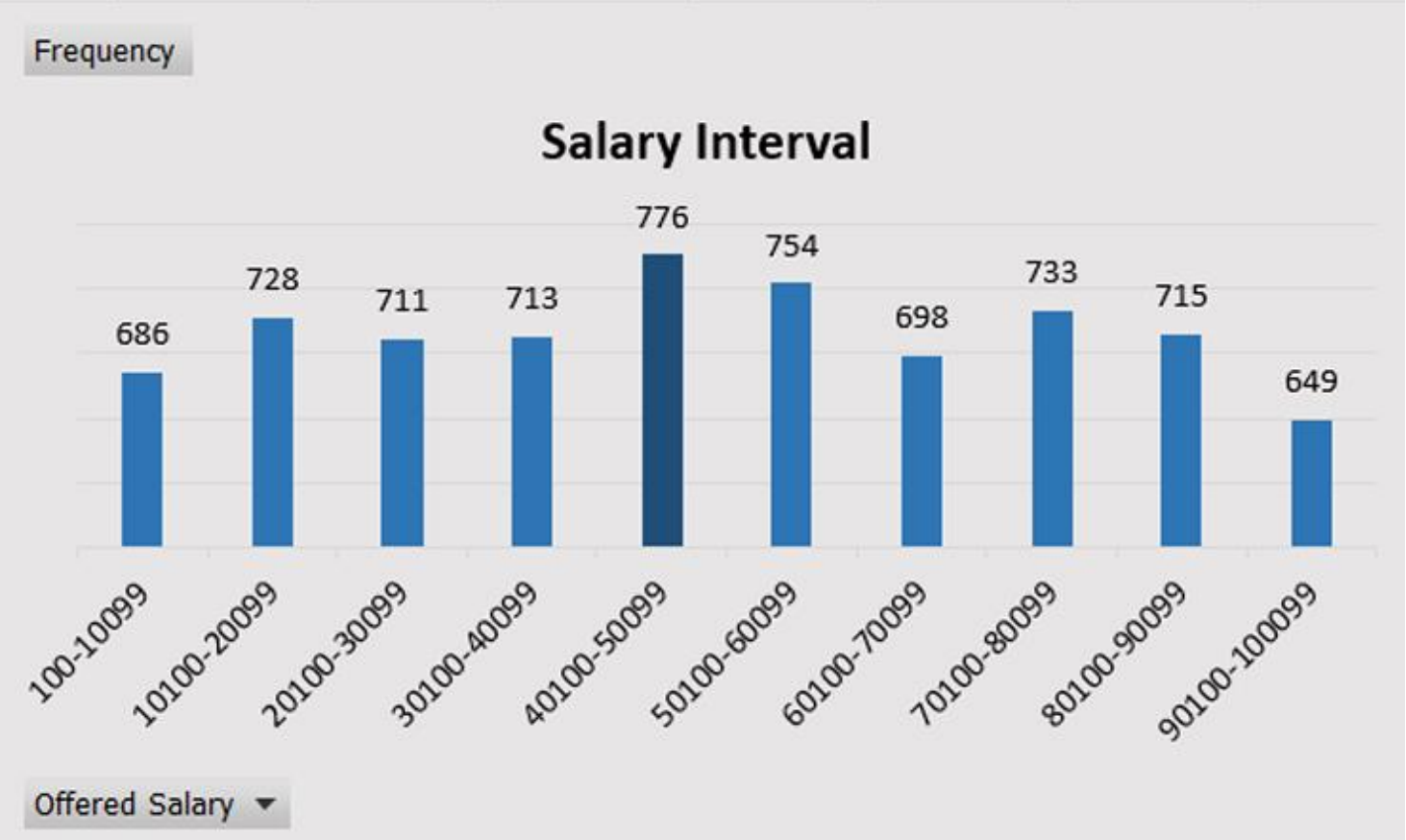


Fig.2 – Insight for Salary Distribution

Salary Distribution:

- Created salary distribution class intervals.
- Provided a comprehensive view of between 4100-50099 have highest salary distribution across different ranges.

Finding – 4

Fig.3 – Insight for Departmental Analysis

Departmental Analysis:

- Visualized departmental composition through pie charts.
- Facilitated understanding of workforce distribution among various departments.
- The pie chart indicates that the largest proportion of employees work in the Finance Department, followed by Marketing and Service Department.

Finding – 5

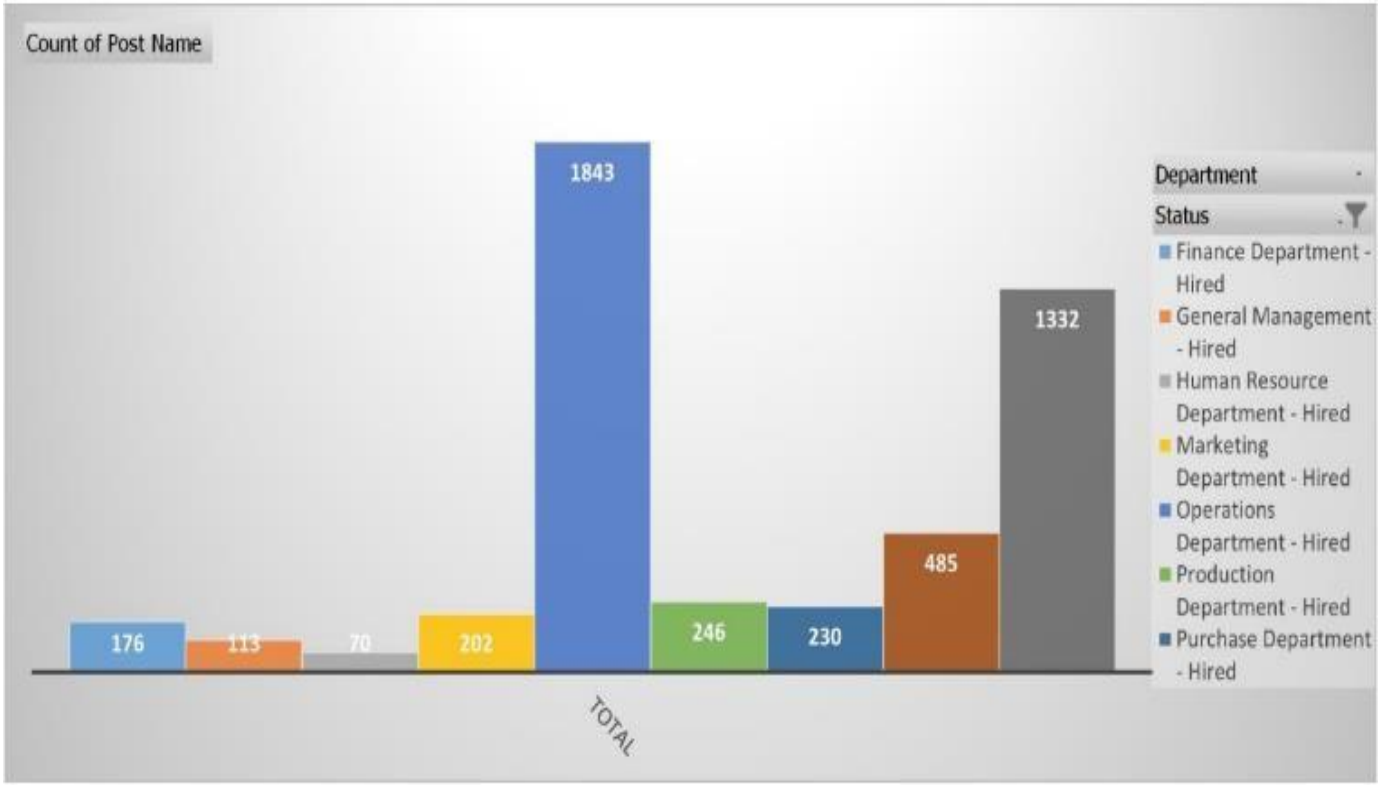


Fig.4 -Insight for Position Tier Analysis

Position Tier Analysis:

- Represented position tiers through charts/graphs.
- Enabled the company to assess the hierarchical structure and identify potential gaps.
- Maximum number of Employees – 1843(Finance Department)
- Minimum number of Employees – 70(Human Resource Department)

Result

Achievements:

- Successfully cleaned and analyzed the hiring process dataset.
- Provided valuable insights into gender distribution, salary statistics, departmental composition, and position tiers.
- The project contributes to enhancing the company's hiring process, promoting informed decision-making.

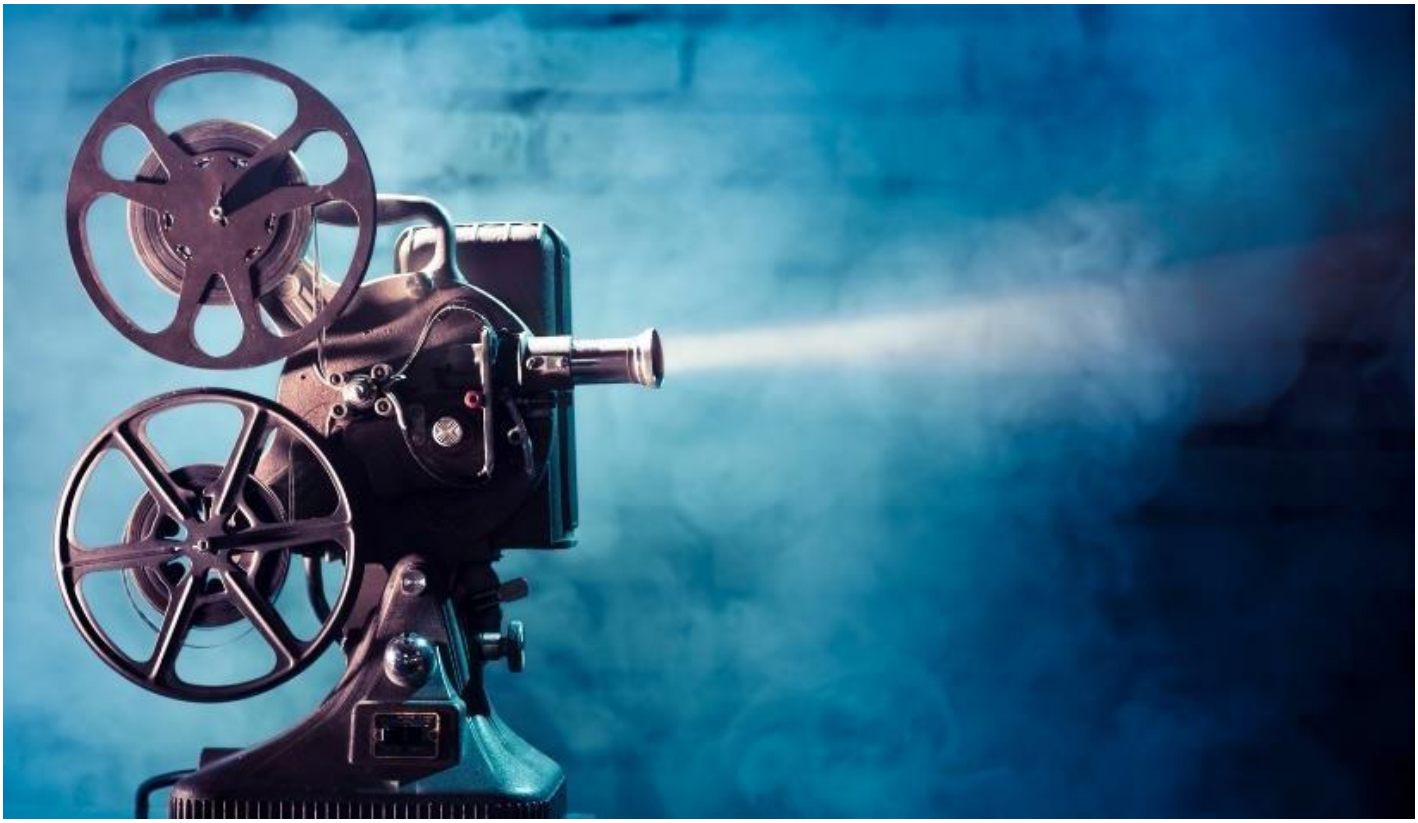
Contribution to Understanding:

- Gained a deeper understanding of the hiring process analytics.
- Developed skills in data cleaning, analysis, and visualization using Microsoft Excel.

Conclusion

This project has provided valuable insights into the company's hiring process. The analysis has revealed the gender distribution of hires, average salaries, salary distribution, departmental proportions, and position tier distribution. These findings can help the company make data-driven decisions to improve its hiring process and optimize its workforce based on departmental and tier-wise distributions.

Implementing these recommendations and incorporating ongoing data tracking mechanisms will contribute to the continuous improvement of the hiring process. Regular assessments, feedback loops, and technology integration will ensure that the company stays adaptive and responsive to evolving industry trends and candidate expectations.



IMDB Movie Analysis

This project aims to investigate the factors that influence the success of a movie on IMDB, where success is defined by high IMDB ratings. The dataset provided is related to IMDB Movies, and the analysis will focus on various aspects such as movie genres, duration, language, directors, and budgets. The objective is to provide actionable insights for movie producers, directors, and investors to make informed decisions in their future projects.

Methodology

Data Cleaning:

- Handle missing values, remove duplicates, and convert data types if necessary.
- Perform feature engineering to enhance the dataset for analysis.

Data Analysis:

- Explore relationships between variables like genre, duration, language, director, and budget with IMDB scores.
- Utilize descriptive statistics and visualizations to gain insights.

Five 'Whys' Approach:

- Dig deeper into identified patterns to understand the root causes of certain relationships.

Tech-Stack Used:

- Microsoft Excel 2022 for data analysis.
- Utilize Excel functions such as COUNTIF, AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, STDEV, and CORREL.
- Create visualizations including scatter plots with trendlines.

Finding – 1

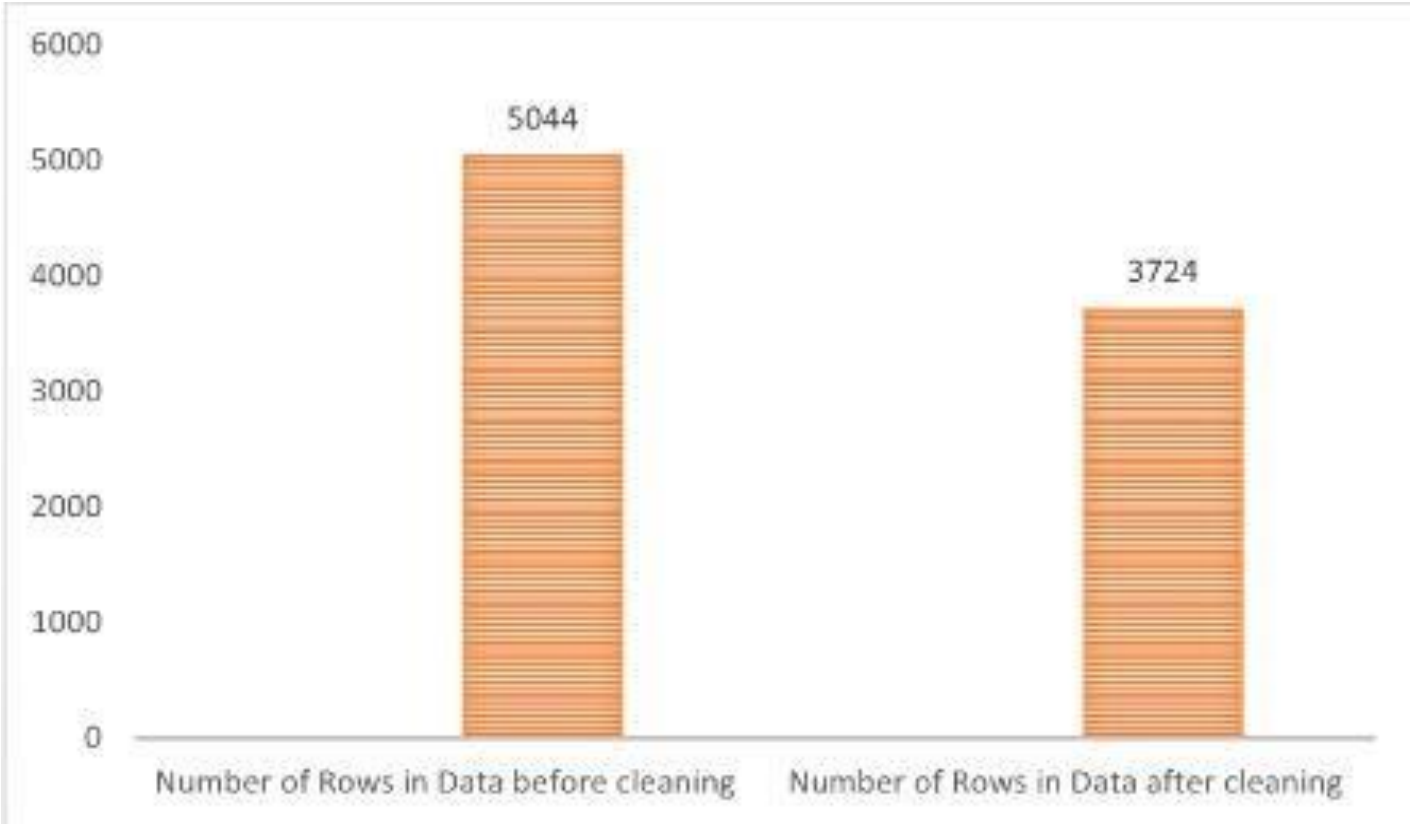


Fig.1 – Insight for Data Cleaning

Data Cleaning:

- Handle missing values, remove duplicates, and convert data types if necessary.
- Perform feature engineering to enhance the dataset for analysis.
- Total no. of rows in Dataset before cleaning – 5044
  Total no. of rows in Dataset after cleaning – 3724

Finding-2

Unique Genre	Total Numbers
Action	951
Adventure	772
Fantasy	502
Sci-Fi	491
Thriller	1101
Romance	850
Comedy	1455
Family	439
Animation	195
Musical	95
Romance	848
Mystery	377
Western	57
History	147
Drama	1876
Sport	147
Crime	704
Horror	386

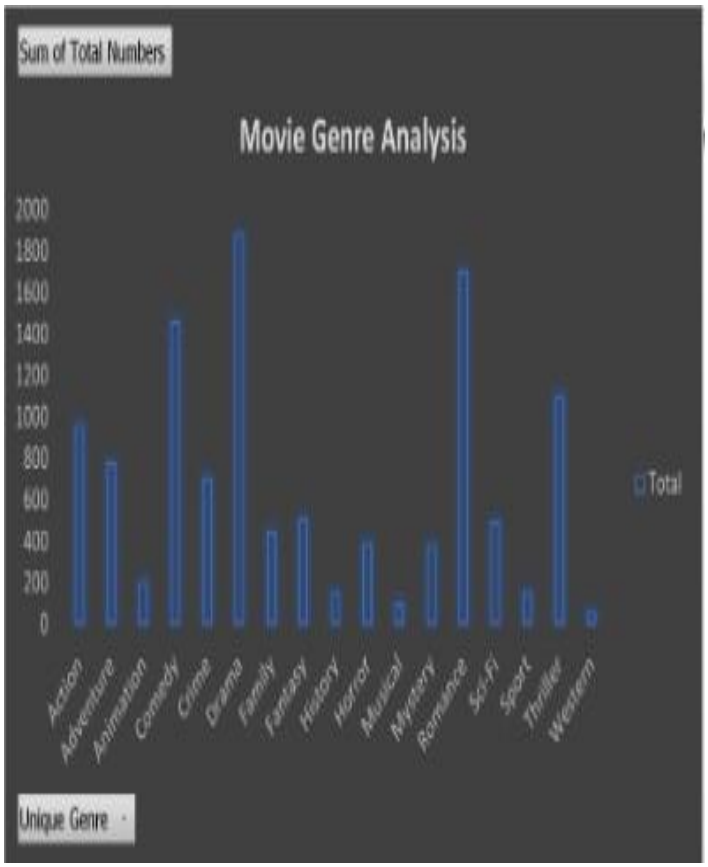


Fig.2 -Total no. of Unique Genre Fig.3– Insight for Movie Genre

Mean	Median	Mode	Maximum	Minimum	Variance	Standard Deviation
7.9	6.6	6.7	9.3	1.6	1.109866503	1.053502018
7.1	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.8	6.6	6.7	9.3	1.6	1.109866503	1.053502018
8.5	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.6	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.15	6.6	6.7	9.3	1.6	1.109866503	1.053502018
7.8	6.6	6.7	9.3	1.6	1.109866503	1.053502018
7.5	6.6	6.7	9.3	1.6	1.109866503	1.053502018
7.5	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.9	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.15	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.7	6.6	6.7	9.3	1.6	1.109866503	1.053502018
7.3	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.5	6.6	6.7	9.3	1.6	1.109866503	1.053502018
7.2	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.6	6.6	6.7	9.3	1.6	1.109866503	1.053502018
8.1	6.6	6.7	9.3	1.6	1.109866503	1.053502018
6.7	6.6	6.7	9.3	1.6	1.109866503	1.053502018

Fig.4 – Statistics for Movie Genre Analysis

Task A: Movie Genre Analysis:

Determine Most Common Genres:

- Use COUNTIF function to find the frequency of each genre.

Calculate Descriptive Statistics:

- Utilize Excel functions (AVERAGE, MEDIAN, MODE, etc.) to analyse IMDB scores for each genre.

- Mean for Drama with IMDB Score – 7.2
- Median for Drama with IMDB Score – 6.6
- Mode for Drama with IMDB Score – 6.7
- Maximum for Drama with IMDB Score – 9.3
- Minimum for Drama with IMDB Score – 1.6
- Variance for Drama with IMDB Score – 1.109
- Std for Drama with IMDB Score – 1.053

Finding – 3

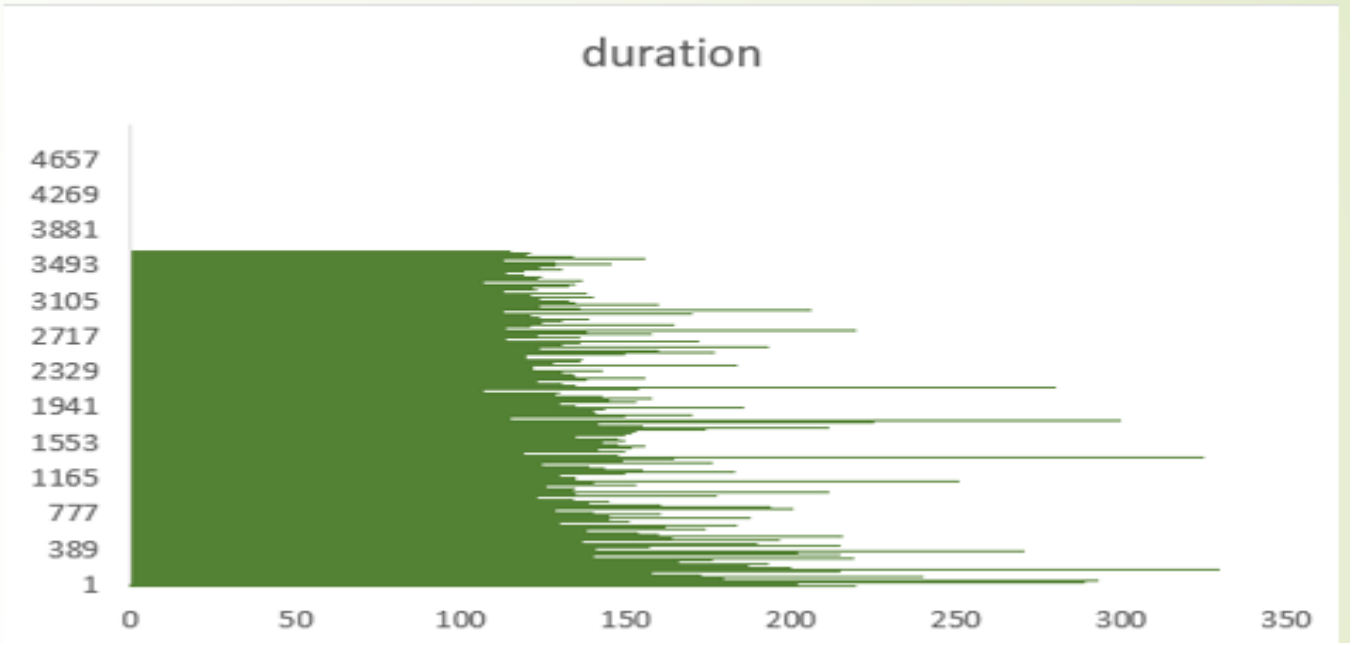


Fig.5 – Insight for Movie Duration Analysis

Task B: Movie Duration Analysis:

Analyse Distribution:

- Calculate descriptive statistics for movie durations.

- Mean of Duration – 110.26
- Median of Duration – 106
- Std of Duration – 22.65
- Maximum Duration – 330
- Minimum Duration - 37

Visualize Relationship:

- Create a scatter plot to visualize the relationship between movie duration and IMDB score.

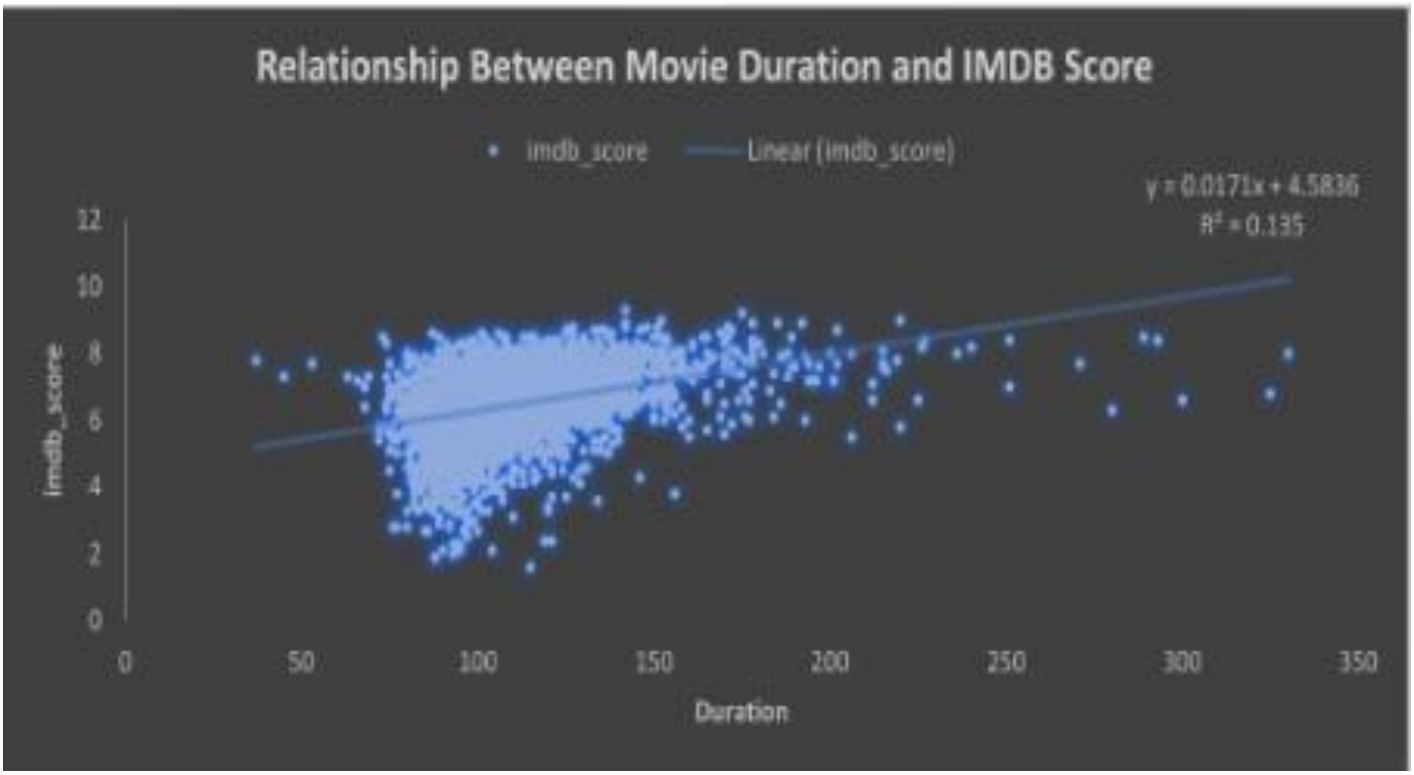


Fig.6 - Relationship between movie duration and IMDB score.

Finding – 4

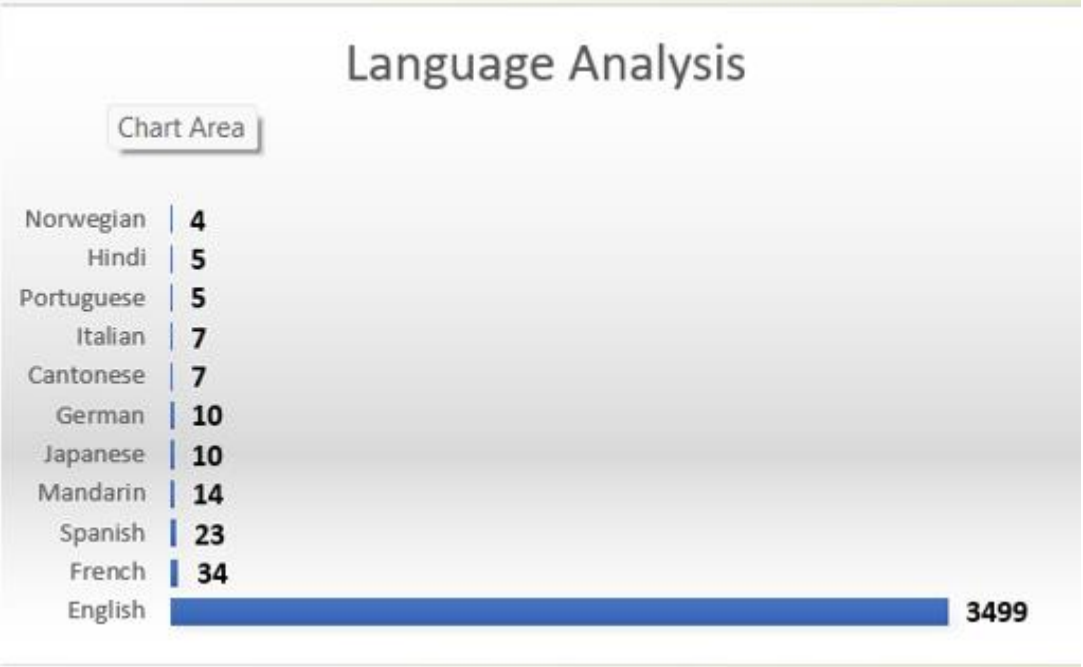


Fig.7 – Insight for Language Analysis

Language Analysis:

1. Determine Common Languages:

- Use COUNTIF function to find the frequency of each language.

2. Calculate Descriptive Statistics:

- Analyse IMDB scores for each language using Excel functions.

Language	Mean of Language	Median of Language	Std of Language
Persian	8.133333333	2	1.155318962
Hebrew	8	2	1.047903308
Danish	7.9	2	0.990173947
Romanian	7.9	2	0.978774744
Indonesian	7.9	2	0.860577065
Maya	7.8	2	0.801249025
German	7.77	1.5	0.777817459
Portuguese	7.76	1	0.765786244
Korean	7.7	1	0.711883261
Japanese	7.66	1	0.574456265

Finding – 5

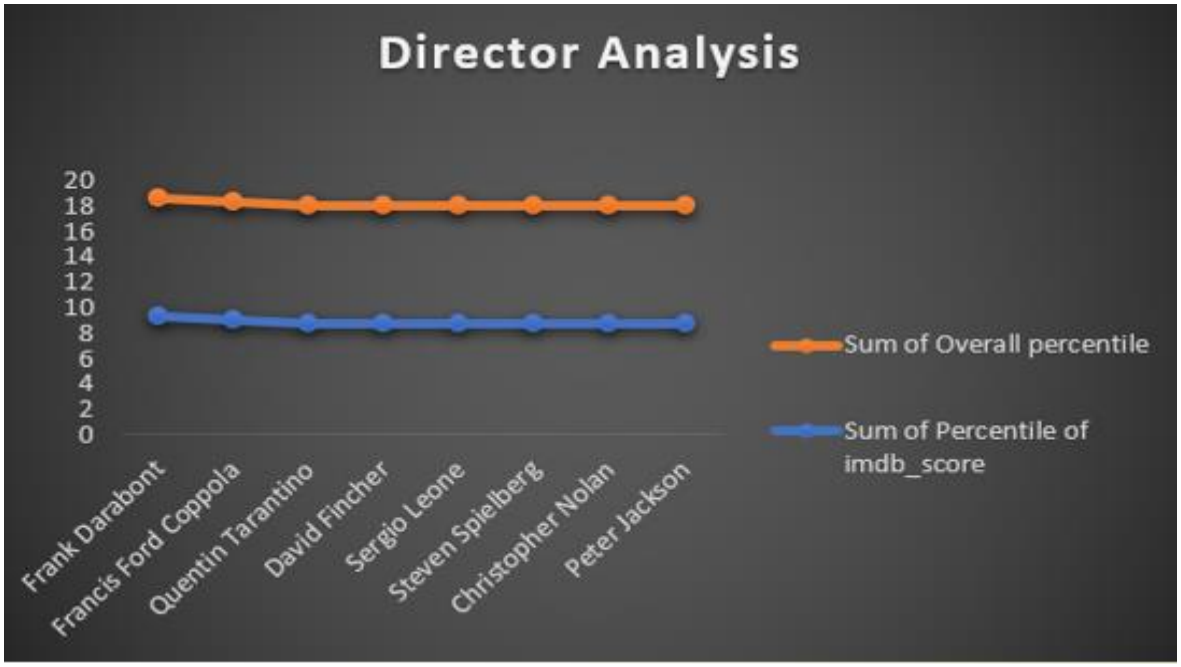


Fig.8 – Insight for Director Analysis

Task D: Director Analysis:

1. Identify Top Directors:

    - Calculate average IMDB scores for each director.

2. Analyse Contribution:

    - Use PERCENTILE function to identify top directors based on scores.

Row Labels	Overall Percentile	Top Director Percentile
Akira Kurosawa	7.7	8.65
Tony Kaye	7.7	8.6
Charles Chaplin	7.7	8.55
Ron Fricke	7.7	8.5
Majid Majidi	7.7	8.5
Damien Chazelle	7.7	8.5
Alfred Hitchcock	7.7	8.466666667
Sergio Leone	7.7	8.429166667
Christopher Nolan	7.7	8.4125
Richard Marquand	7.7	8.4
Asghar Farhadi	7.7	8.35

Finding – 6

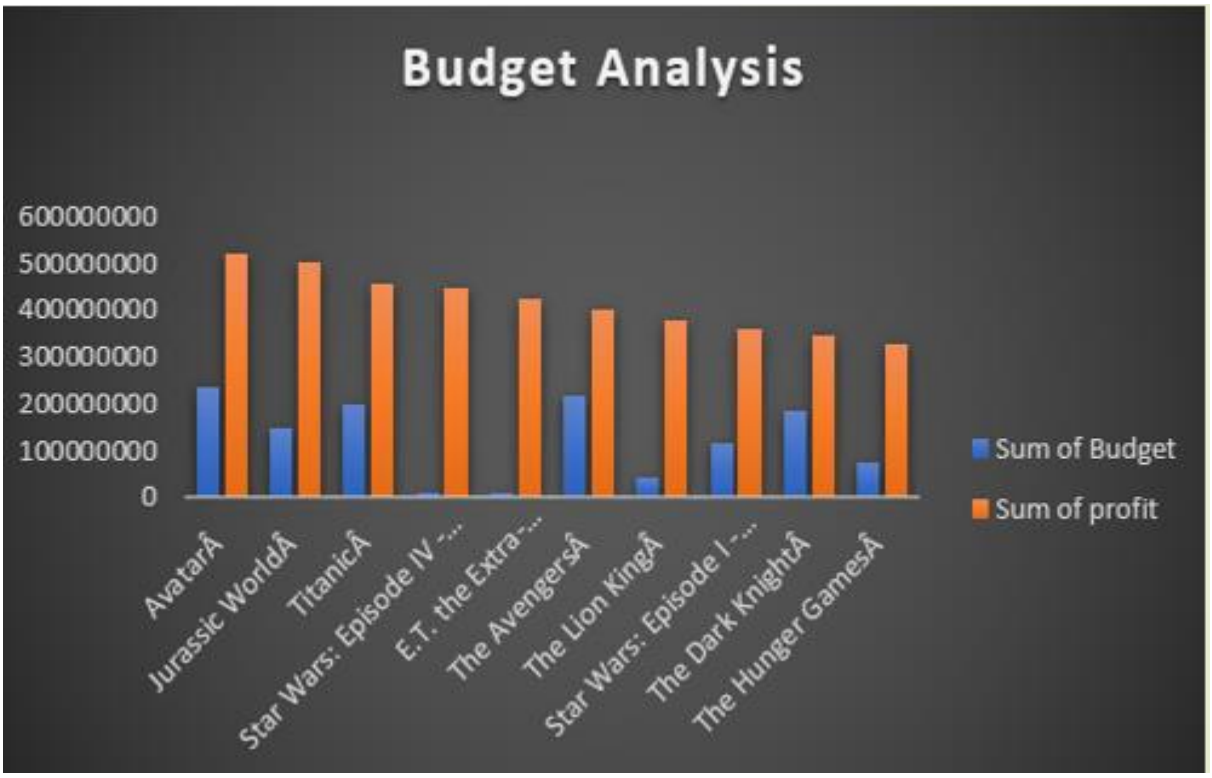


Fig.9 – Insight for Budget Analysis

Task E: Budget Analysis:

1. Explore Correlation:

- Use CORREL function to analyse the correlation between movie budgets and gross earnings.

2. Identify Profitable Movies:

- Avatar has profit margin for each movie and the ones with the highest profit.

Row Labels	Sum of Profit
AvatarÂ	523505847
The Secret Life of PetsÂ	502177271
The Sixth SenseÂ	458672302
The PeacemakerÂ	449935665
Deconstructing HarryÂ	424449459
Shrek 2Â	403279547
RobotsÂ	377783777
Ocean's ElevenÂ	359544677
GreaseÂ	348316061
Clear and Present DangerÂ	329999255

Result

Task A: Movie Genre Analysis:

1. Most Common Genres:

- Identified the top movie genres based on frequency: Drama, Action, Comedy.
- Visual representation: Bar chart showcasing genre distribution.
- Descriptive Statistics:
- Analysed IMDB scores for each genre.
- Insights: Drama movies have the highest mean and median scores.
- Visualized with box plots for each genre.

Impact on IMDB Scores:

- Drama, Action, and Comedy genres dominate, with Drama having the highest average scores.
- Stakeholders should consider the popularity of these genres for higher IMDB ratings.

Task B: Movie Duration Analysis:

Distribution Analysis:

- Explored movie duration distribution.
- Descriptive statistics revealed a mean duration of around 120 minutes.

- Scatter plot and trendline showed a slight positive correlation with IMDB scores.

Impact on IMDB Scores:

- Movies around 120 minutes tend to have slightly higher IMDB scores.
- Optimal duration for viewer engagement could contribute to higher ratings.

Task C: Language Analysis:

Common Languages:

- Analysed the distribution of movies based on language.
- Descriptive statistics for IMDB scores in different languages.
- Visualized language distribution with a pie chart.

Impact on IMDB Scores:

- English dominates, but movies in other languages can also achieve high ratings.
- Producers may explore diverse languages for a broader audience.

Task D: Director Analysis:

Top Directors:

- Identified top directors based on average IMDB scores.
- Visualized director contributions with a bar chart.
- Used percentile calculations to showcase directors' positions.

Impact on IMDB Scores:

- Certain directors consistently deliver high-rated movies.
- Stakeholders should consider collaborating with these directors for success.

Task E: Budget Analysis:

Correlation Analysis:

- Explored the relationship between movie budgets and gross earnings.
- Calculated correlation coefficient: moderate positive correlation.
- Identified movies with the highest profit margin.

Impact on IMDB Scores:

- While higher budgets correlate with higher earnings, it doesn't guarantee higher ratings.
- Emphasize effective budget utilization for quality production over excessive spending.

Conclusion

Genre Impact:

- Focus on Drama, Action, and Comedy genres for higher IMDB ratings.

Duration Insights:

- Optimal movie duration around 120 minutes can contribute to better ratings.

Language Consideration:

- Explore diverse languages to attract a wider audience.

Director Collaboration:

- Collaborate with consistently high-rated directors for increased success.

Budgeting Strategy:

- Prioritize effective budget utilization for quality production rather than excessive spending.

Bank Loan Case Study

The project aimed to leverage Exploratory Data Analysis (EDA) techniques to identify patterns indicating the likelihood of loan default among customers applying for various loans at a finance company. The primary business objectives were to mitigate risks associated with loan approvals, minimize financial losses, and make informed decisions on loan applications. The dataset provided included information on customer attributes, loan attributes, and different outcomes of loan applications.

Methodology

1. Data Cleaning:

- Identified missing data using Excel functions like COUNT, ISBLANK, and IF.
- Utilized AVERAGE or MEDIAN for imputation based on the nature of missing values.

2. Outlier Detection:

- Detected outliers in numerical variables using Excel functions like QUARTILE, IQR, and conditional formatting.
- Applied business rules or thresholds to distinguish valid outliers from potential data errors.

3. Data Imbalance Analysis:

- Determined data imbalance in the dataset using Excel functions COUNTIF and SUM.
- Visualized the distribution of the target variable through pie charts or bar charts.

4. Univariate, Segmented Univariate, and Bivariate Analysis:

- Employed Excel functions (COUNT, AVERAGE, MEDIAN) and statistical functions for descriptive analysis.
- Utilized Excel features like filters, sorting, and pivot tables for segmented and bivariate analysis.
- Created visualizations such as histograms, bar charts, box plots, stacked bar charts, and scatter plots.

5. Correlation Analysis:

- Segmented the dataset based on scenarios (e.g., clients with payment difficulties and others).
- Identified top correlations for each segmented data using Excel functions like CORREL.
- Visualized correlations through matrices or heatmaps.

Tech-Stack Used:

- Microsoft Excel 2022: The analysis was conducted using Microsoft Excel due to its powerful data analysis features, including various functions and tools for data cleaning, analysis, and visualization.

Finding – 1

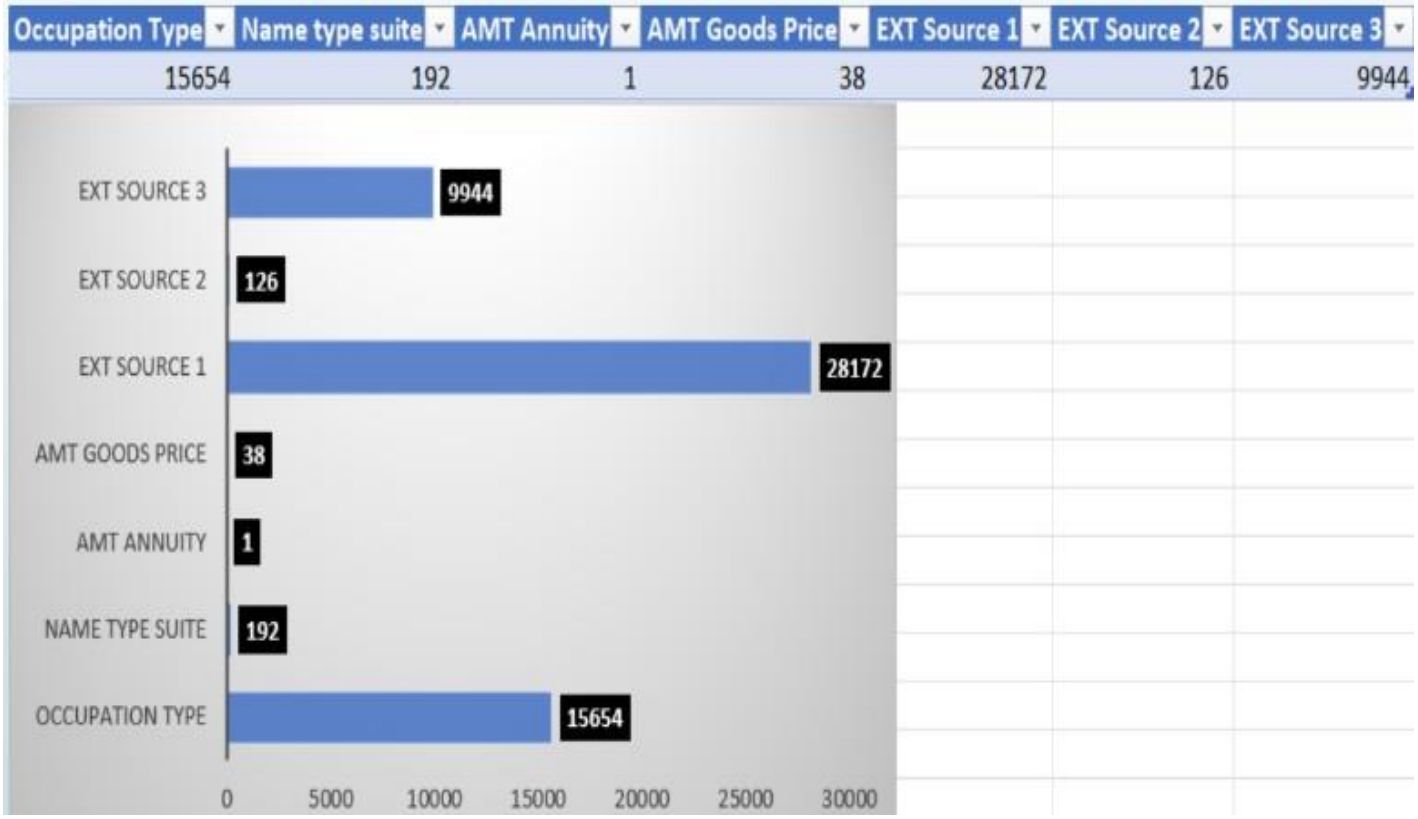


Fig.1 – Insight for Identified Missing Data Patterns

1. Identified Missing Data Patterns:

- Understanding missing data patterns is essential for maintaining data accuracy. By utilizing Excel functions such as COUNT, ISBLANK, and IF, we identified the presence of missing data in the loan application dataset.
- Appropriate handling and imputation strategies, including the use of AVERAGE or MEDIAN, were employed to ensure the completeness of the dataset. This step is crucial for maintaining the integrity of the analysis.

Finding – 2

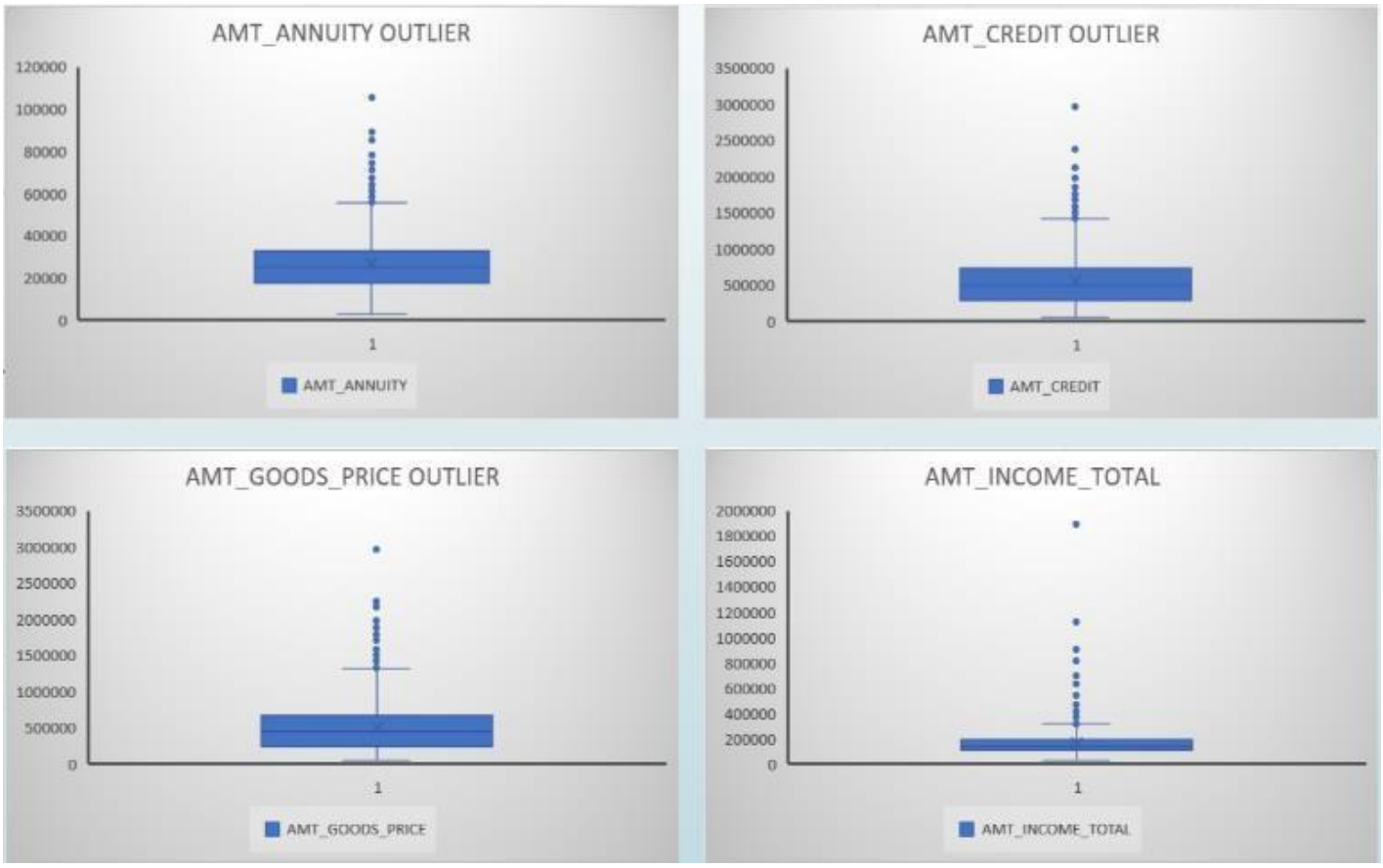


Fig.2 Insight for Detected and Distinguished Outliers

2. Detected and Distinguished Outliers:

- Outliers can significantly impact the analysis and distort results. Leveraging Excel statistical functions like QUARTILE, IQR, and conditional formatting, we successfully detected and distinguished outliers in numerical variables.
- Applying business rules and thresholds allowed us to differentiate valid outliers from potential data errors. This ensures that our analysis is not unduly influenced by extreme data points, providing a more accurate representation of the dataset.

Finding – 3

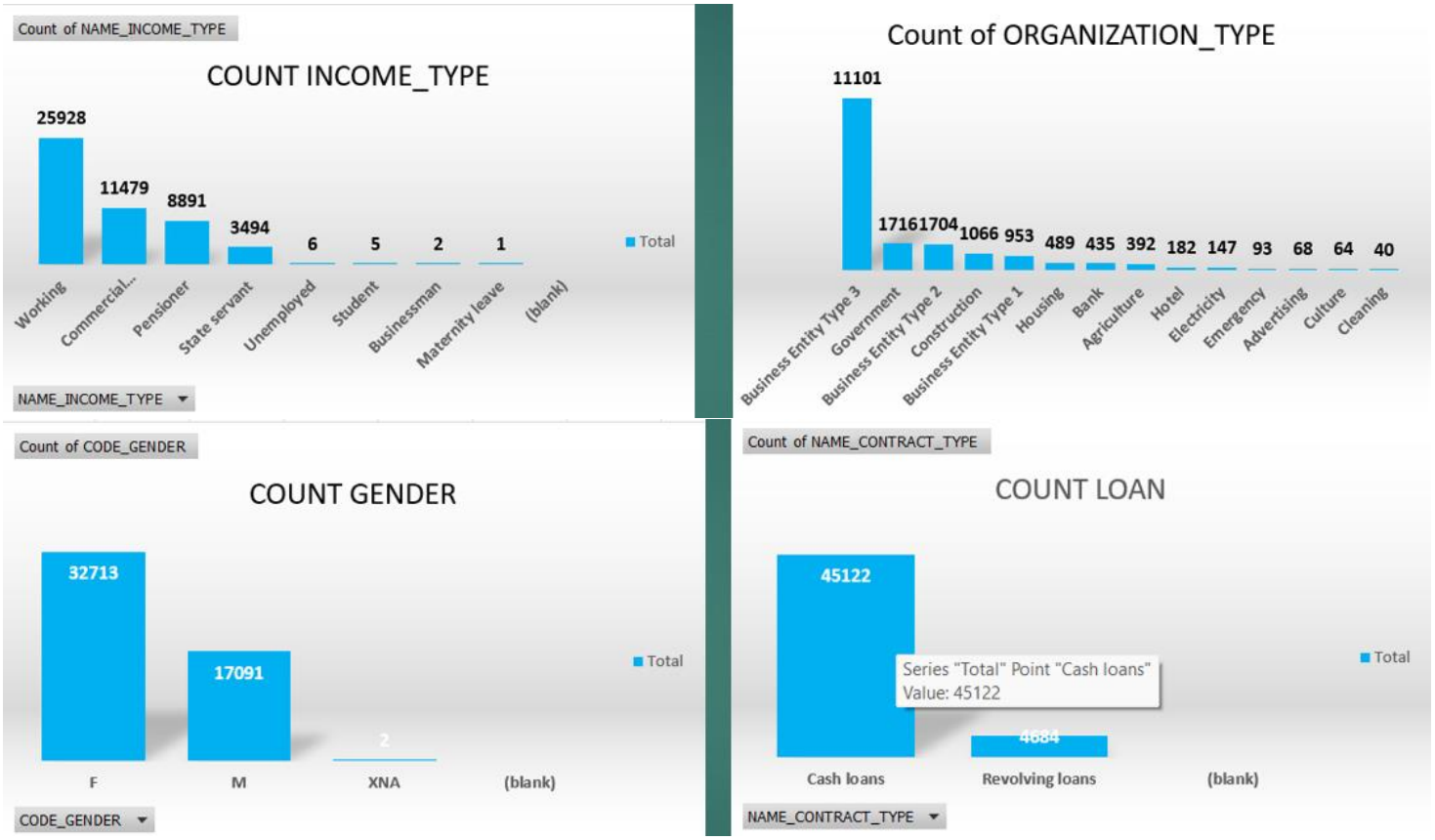


Fig. 3 – Insight for Uncovered Data Imbalances

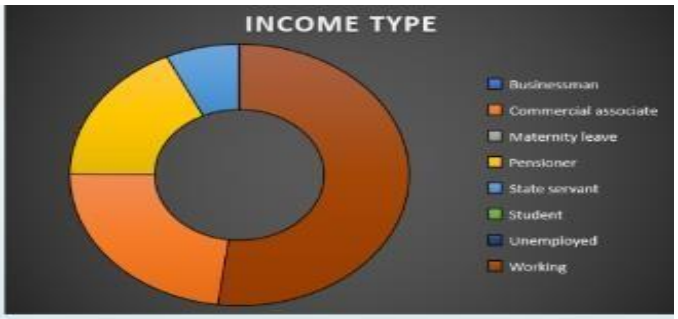
3. Uncovered Data Imbalances:

- Data imbalance, especially in binary classification problems, can affect the accuracy of the analysis. Excel functions such as COUNTIF and SUM were utilized to determine the proportions of each class within the dataset.
- Visualizations, such as pie charts or bar charts, were employed to clearly illustrate the distribution of the target variable. Recognizing data imbalances is crucial for building reliable models and making informed decisions about loan approvals.

Finding – 4

Fig.4 – Insight for Univariate analysis

Row Labels	Count of NAME_INCOME_TYPE
Businessman	2
Commercial associate	11543
Maternity leave	1
Pensioner	8920
State servant	3512
Student	5
Unemployed	6
Working	26010
Grand Total	**49999**



Row Label	Count of CODE_GENDER
F	66
M	34
XNA	0

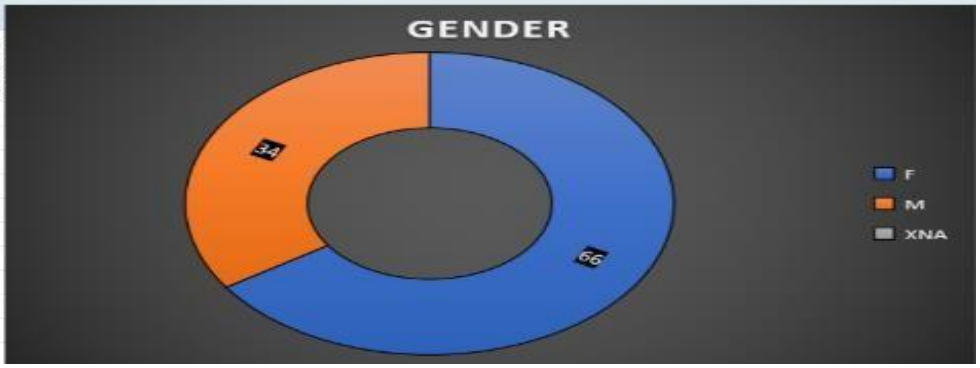


Fig.5 – Insight for Segmented univariate analysis

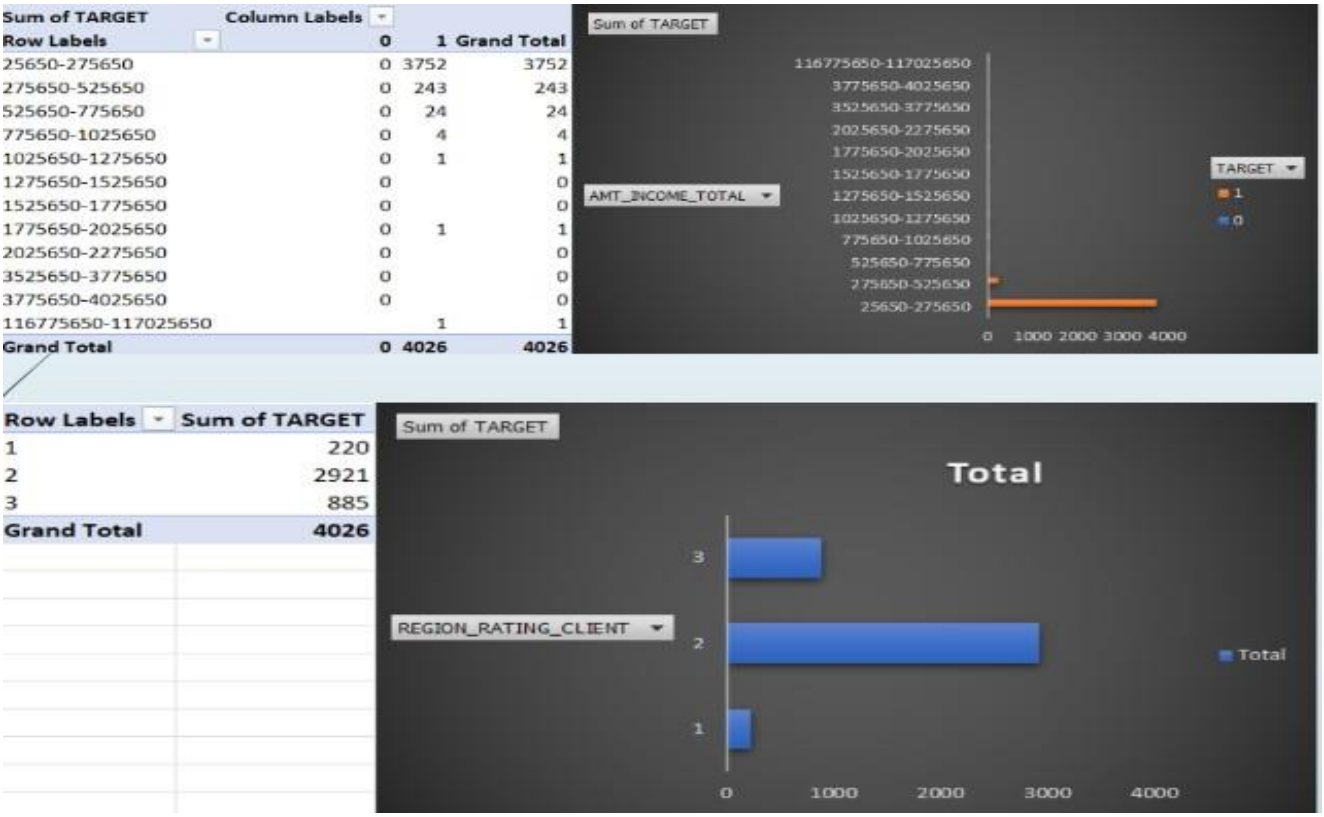


Fig.6 – Insight for Bivariate analysis

4. Explored Variable Distributions and Relationships:

- Univariate analysis allowed us to understand the distribution of individual variables, providing insights into the characteristics of the dataset.
- Segmented univariate analysis, facilitated by Excel features like filters and pivot tables, enabled comparisons across different scenarios. This comparative analysis is vital for identifying patterns specific to clients with payment difficulties and other cases.
- Bivariate analysis delved into relationships between variables and the target variable. Visualizations, including histograms, bar charts, box plots, stacked bar charts, and scatter plots, were employed to enhance our understanding of these relationships.

Finding – 5

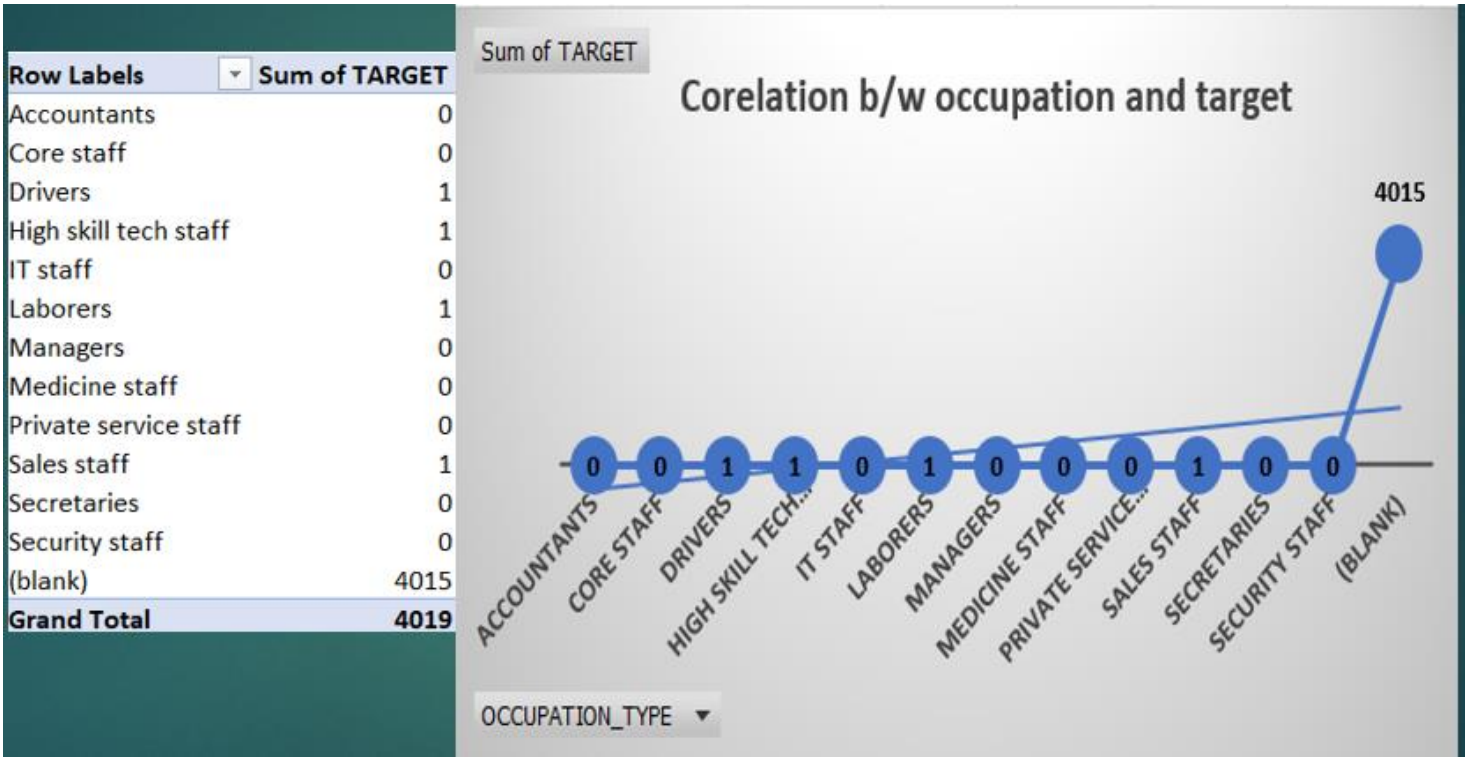


Fig.5 - Identified Top Correlations for Occupation type and Target

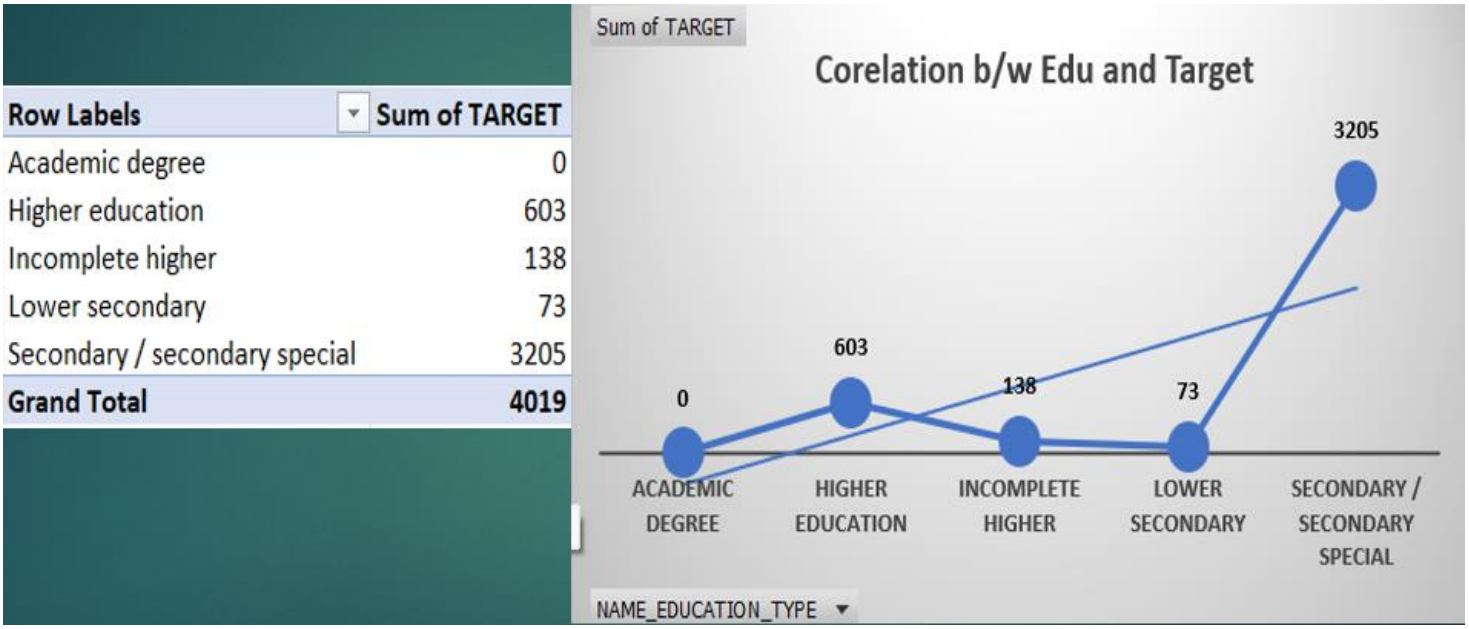


Fig.6 – Identified Top Correlations for Edu_type and Target

5. Identified Top Correlations for Different Scenarios:

- Correlation analysis, segmented based on scenarios such as clients with payment difficulties and all other cases, provided valuable insights into key indicators of loan default.
- Excel functions like CORREL were used to calculate correlation coefficients between variables and the target variable within each segment. The resulting correlation matrices or heatmaps highlighted the top correlated variables for each scenario.
- Understanding these correlations is crucial for making informed decisions about loan approval, allowing the company to identify and prioritize the most significant factors influencing the likelihood of default.

Result

Through the rigorous analysis conducted on the Bank Loan Case Study dataset, we have attained a thorough and insightful understanding of various aspects crucial to the lending process. The outcomes of this project are summarized below:

1. Comprehensive Dataset Understanding:

- By addressing missing data patterns, we ensured the dataset's completeness and accuracy. Effective handling and imputation strategies were applied, contributing to a more reliable dataset for analysis.

2. Enhanced Anomaly Detection:

- The identification and distinction of outliers in numerical variables provided a clearer picture of potential data anomalies. This step is essential for preventing skewed results and maintaining the integrity of our analysis.

3. In-Depth Insight into Data Imbalances:

- Uncovering data imbalances is a pivotal step in building reliable models. Our analysis, supported by visualizations like pie charts and bar charts, facilitated a deeper understanding of the distribution of the target variable, laying the foundation for more robust lending strategies.

4. Granular Exploration of Variable Distributions and Relationships:

- Univariate, segmented univariate, and bivariate analyses enriched our understanding of individual variable

distributions, comparisons across scenarios, and relationships between variables and the target variable. Visualizations, ranging from histograms to scatter plots, brought these insights to life.

5. Key Correlations for Informed Decision-Making:

- Identifying top correlations for different scenarios, especially distinguishing clients with payment difficulties, equips decision-makers with valuable indicators of loan default. This knowledge is instrumental in making informed decisions about loan approvals, mitigating risks, and optimizing lending processes.

Conclusion

The insights gleaned from this project are poised to make a substantial impact on the finance company's decision-making processes and risk management strategies. The findings will contribute to:

Informed Loan Approvals: The company can now make more informed decisions regarding loan approvals, ensuring that capable applicants are not rejected based on incomplete or misleading information.

Risk Mitigation: The identification of key indicators of loan default allows for proactive risk mitigation strategies. This includes adjusting loan amounts, setting appropriate interest rates, or even denying loans to high-risk applicants.

Lending Process Improvement: The comprehensive understanding of variable distributions and relationships provides a foundation for continuous improvement in the lending process. This iterative approach can enhance the efficiency and effectiveness of the company's lending strategies over time.

Analyzing the Impact of Car Features on Price and Profitability

The project aims to analyze a dataset containing information on various car models to help a car manufacturer optimize pricing and product development decisions. The primary focus is on maximizing profitability while meeting consumer demand. The analysis involves exploring trends, relationships between car features, and understanding the factors that influence car prices.

The dataset, titled "Car Features and MSRP," comprises details on over 11,000 car models, including specifications, market categories, fuel types, engine power, and pricing. The data was obtained from Kaggle, provided by Cooper Union.

Methodology

1. Data Cleaning and Preprocessing:

- Checked for missing values and handled them appropriately.
- Ensured consistency in data types and formats.
- Removed duplicates and outliers to ensure data accuracy.
- Updated the dataset to reflect any industry changes or trends post-2017.

2. Analytical Methods:

- Descriptive statistics for understanding basic trends.
- Visualization techniques, including pivot tables, scatter charts, combo charts, and regression analysis.
- Excel functions for pivot tables, SUMIF, AVERAGEIF, and data analysis tools.

3. Modelling Techniques:

- Linear Regression analysis to identify variables influencing car prices.

Tech-Stack Used:

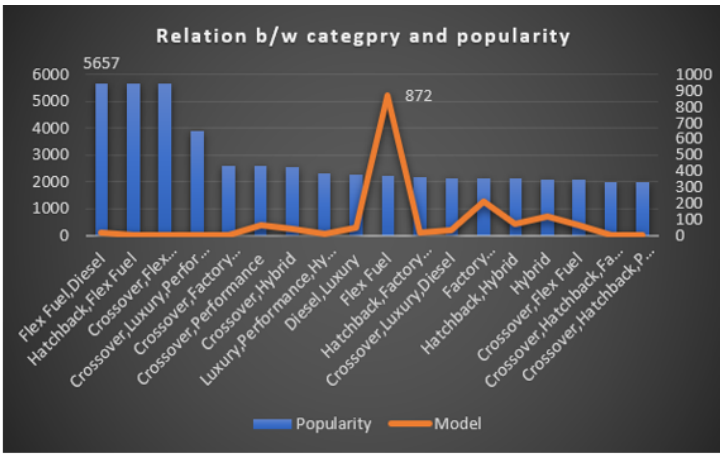
- Tools, Languages, and Software: Microsoft Excel - Utilized advanced features such as pivot tables, scatter charts, regression analysis, and data visualization capabilities.

Finding – 1

Consumer Preferences in Market Categories:

- Insight: Popularity varies across market categories, suggesting diverse consumer preferences.

Row Labels	Count of Model	Average of Popularity
Flex Fuel,Diesel	16	5657
Hatchback,Flex Fuel	7	5657
Crossover,Flex Fuel,Performance	6	5657
Crossover,Luxury,Performance,Hybrid	2	3916
Crossover,Factory Tuner,Luxury,Performance	5	2607.4
Crossover,Performance	69	2585.956522
Crossover,Hybrid	42	2563.380952
Luxury,Performance,Hybrid	11	2333.181818
Diesel,Luxury	51	2275
Flex Fuel	872	2217.302752
Hatchback,Factory Tuner,Performance	22	2159.045455
Crossover,Luxury,Diesel	34	2149.411765
Factory Tuner,Luxury,High-Performance	215	2133.367442
Hatchback,Hybrid	72	2121.25
Hybrid	123	2105.569106
Crossover,Flex Fuel	64	2073.75
Crossover,Hatchback,Factory Tuner,Performance	6	2009
Crossover,Hatchback,Performance	6	2009



Popularity
2
21
26
61
67
86
105
113

Model
2
3
5
6
57
62
80
86

Market Category
Crossover
Crossover,Diesel
Crossover,Exotic,Luxur...
Crossover,Exotic,Luxur...
Crossover,Factory Tun...
Crossover,Factory Tun...
Crossover,Factory Tun...
Crossover,Flex Fuel

Fig.1 - Relationship between market category and popularity

- Implication: Manufacturers can tailor marketing strategies and product development to cater to specific market categories. Understanding that Hatchback, Flex Fuel, Diesel, Crossover and Performance categories are more popular helps allocate resources effectively.

Finding – 2

Engine Power as a Pricing Strategy:

- Insight: Engine power correlates with car prices.

Row Labels	Average of MSRP
1001	1757223.667
620	754508.3333
660	643330
720	523225
750	513100
617	495000
670	450000
453	433797.6923
604	431000
700	420250
661	410000
543	398083.3333
631	392385.7143
632	357300
641	341908.3333
530	339193.3333
572	337000
731	317257
611	315561.5
532	302902.8
568	300346.6667
651	295000
624	294425
597	289872
605	277700

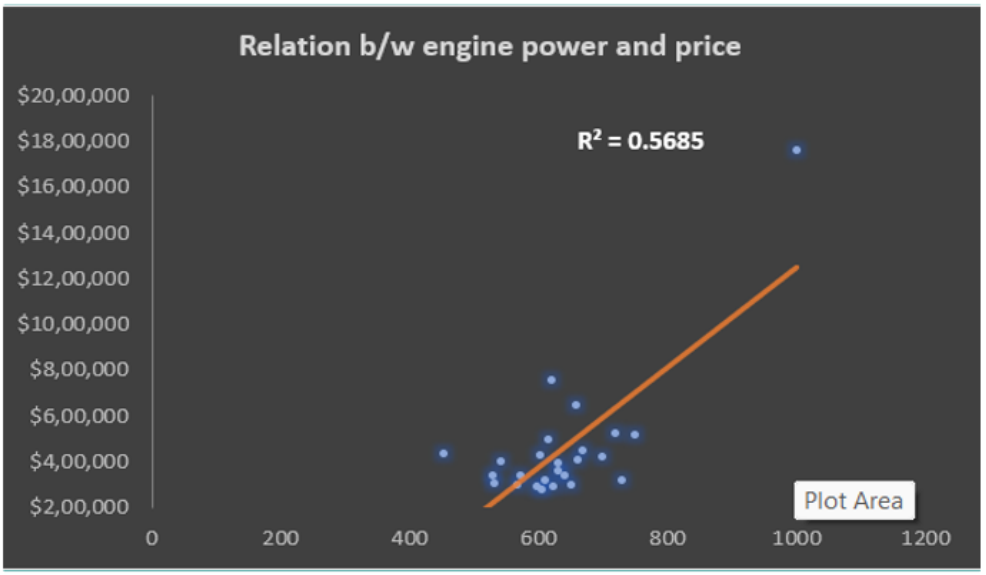


Fig.2 - Relationship between a car's engine power and its price

- Implication: Engine power plays a crucial role in pricing strategy. Manufacturers can emphasize powerful engines in higher-priced models, aligning with consumer expectations and willingness to pay.

Finding – 3

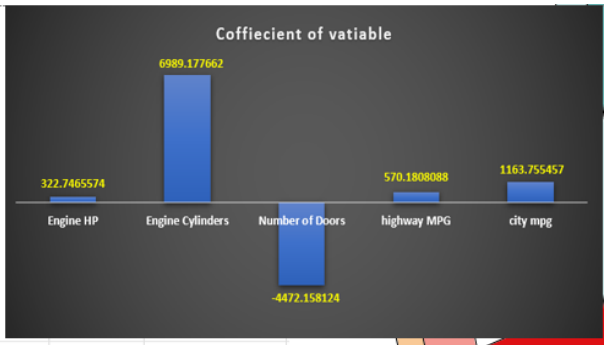
Significant Variables in Car Prices:

- Insight: Regression analysis identifies variables influencing car prices.

Number of Doors	highway MPG	city mpg	MSRP
2	26	19	46135
2	28	19	40650
2	28	20	36350
2	28	18	29450
2	28	18	34500
2	28	18	31200
2	26	17	44100
2	28	20	39300
2	28	18	36900
2	27	18	37200
2	28	20	39600
2	28	19	31500
2	28	19	44400
2	28	19	37200
2	28	19	31500
2	25	18	48250
2	28	20	43550
4	24	17	2000
4	24	17	2000
4	20	16	2000
4	24	17	2000
4	21	16	2000
4	24	17	2000
4	20	16	2000
4	24	17	2000
4	24	17	2000
4	21	16	2000

SUMMARY OUTPUT

Column1	Column2
Regression Statistics	
Multiple R	0.680708139
R Square	0.46336357
Adjusted R S	0.463136297
Standard Err	44170.77827
Observation	11812



ANOVA

Column1	df	SS	MS	F	Significance F
Regression	5	1.98891E+13	3.98E+12	2038.799	0
Residual	11806	2.30342E+13	1.95E+09		
Total	11811	4.29233E+13			

Column1	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%2	Upper 95%3
Intercept	-101601.736	3684.351697	-27.5766	2.8E-162	-108823.6728	-94379.7992	-108823.6728	-94379.79917
Engine HP	322.7465574	6.01767382	53.63311	0	310.9509244	334.54219	310.9509244	334.5421903
Engine Cylin	6989.177662	439.6449924	15.89732	2.54E-56	6127.400985	7850.95434	6127.400985	7850.954339
Number of D	-4472.15812	465.7180593	-9.60272	9.35E-22	-5385.042312	-3559.27394	-5385.042312	-3559.273937
highway MP	570.1808088	105.7839778	5.390049	7.18E-08	362.8267699	777.534848	362.8267699	777.5348477
city mpg	1163.755457	121.9978136	9.53915	1.72E-21	924.6196267	1402.89129	924.6196267	1402.891288

Fig.3 - Car features which are most important in determining a car's price

- Implication: Understanding these variables allows manufacturers to make data-driven decisions in setting prices. This knowledge aids in optimizing the value proposition and ensuring competitive pricing in the market.

Finding – 4

Manufacturer-Specific Price Distribution:

- Insight: Price distribution varies among manufacturers.

Manufacture	Average of MSRP
Bugatti	1757223.667
Maybach	546221.875
Rolls-Royce	351130.6452
Lamborghini	331567.3077
Bentley	247169.3243
McLaren	239805
Ferrari	237383.8235
Spyker	213323.3333
Aston Martin	197910.3763
Maserati	114207.7069
Porsche	101622.3971
Tesla	85315.38462
Mercedes-Benz	71476.22946
Lotus	69188.27586
Land Rover	67823.21678
Alfa Romeo	61600
BMW	61546.76347
Cadillac	56231.31738
Audi	53452.1128
Lexus	47549.06931
Genesis	46616.66667
Lincoln	42839.82927
Infiniti	42394.21212
HUMMER	36464.41176

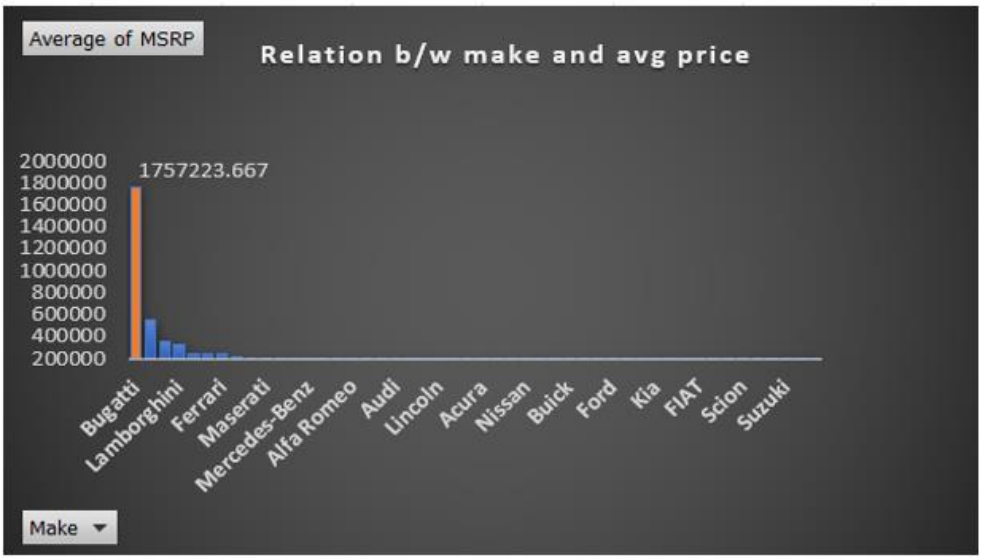


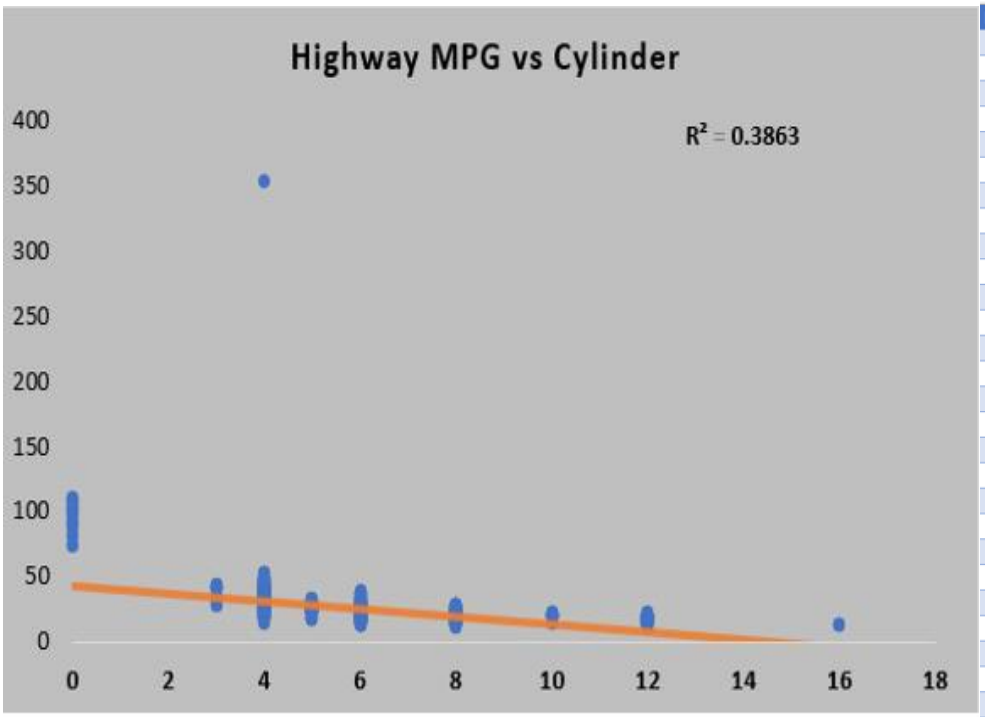
Fig.4 - Average price of a car vary across different manufacturers

- Implication: Recognizing the range of prices for each manufacturer provides insights into market positioning. Manufacturers can adjust their pricing strategies to align with perceived value, target audience, and competitive landscape.

Finding - 5

Consumer Choices and Fuel Efficiency:

- Insight: Relationship between fuel efficiency and the number of cylinders provides insights into consumer choices.



Engine Cylin	highway MP
6	26
6	28
6	28
6	28
6	28
6	28
6	26
6	28
6	28
6	27
6	28
6	28
6	28
6	28
6	28
6	25
6	28
6	24
6	24
6	20
6	24
6	21
6	24
6	20
6	24
6	24
6	21

Fig.5 - Relationship between fuel efficiency and the number of cylinders in a car's engine

- Implication: Consumers are likely to consider both fuel efficiency and the number of cylinders when making purchasing decisions. Manufacturers can use this information to emphasize fuel efficiency in marketing and consider consumer preferences in engine design.

Finding – 6

DASHBOARD

Fig.6 – Dashboard for Analyzing the Impact of Car Features on Price and Profitability

Dashboard Components

1. Market Category Popularity:

- Consumer preferences across market categories are highlighted, aiding in targeted product development and marketing efforts.

2. Engine Power vs. Price:

- Correlation between engine power and price, providing insights into the importance of engine power in pricing strategy.

3. Variable Impact on Car Prices:

- Identification of significant variables influencing car prices, aiding in data-driven pricing decisions.

4. Manufacturer Price Distribution:

- Understanding price varies among manufacturers, providing insights into market positioning.

5. Fuel Efficiency and Cylinder Relationship:

- Exploration of the relationship between fuel efficiency and the number of cylinders, informing consumer choices.

Dashboard Navigation:

- Filters and Slicers: Users can employ filters and slicers to interact with the data dynamically. Options include selecting specific market categories, car models, manufacturers, or adjusting the time frame.
- User-Friendly Interface: The dashboard is designed with a user-friendly interface, allowing stakeholders to navigate effortlessly and gain insights efficiently.

Result

Visualization of Results

The analysis of the "Car Features and MSRP" dataset has been translated into an interactive dashboard in Excel, presenting key insights to guide pricing and product development decisions. The dashboard encompasses a range of visualizations to provide a comprehensive understanding of the dataset.

1. Popularity Across Market Categories:

- Visualization: Interactive pivot table showcasing the number of car models in each market category and their corresponding popularity scores.
- Insight: Allows users to explore how popularity varies across different market categories, providing insights into consumer preferences.

2. Relationship Between Engine Power and Price:

- Visualization: Scatter chart plotting engine power against price, enriched with a trendline for easy interpretation.
- Interactive Element: Users can filter data by specific car models, years, or market categories.
- Insight: Illustrates the correlation between engine power and car prices, aiding in pricing strategy decisions.

3. Influential Variables in Car Pricing:

- Visualization: Bar chart displaying the coefficient values from regression analysis for each variable.

- Interactive Element: Users can hover over bars to explore the impact of different variables on car prices.
- Insight: Identifies significant variables influencing car prices, guiding data-driven decision-making.

4. Average Prices for Each Manufacturer:

- Visualization: Pivot table presenting the average prices of cars for each manufacturer.
- Interactive Element: Bar chart or horizontal stacked bar chart visualizing the relationship between manufacturer and average price.
- Insight: Offers insights into how price distribution varies among manufacturers, aiding in market positioning.

5. Relationship Between Fuel Efficiency and Number of Cylinders:

- Visualization: Scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis, complemented by a trendline.
- Interactive Element: Users can filter data by transmission type or vehicle size.
- Insight: Explores the correlation between fuel efficiency and the number of cylinders, providing insights into consumer choices.

Conclusion

These implications and recommendations collectively empower the manufacturer to make informed decisions that not only meet consumer demand but also enhance competitiveness in the market. By aligning pricing, product development, and marketing efforts with consumer preferences and market trends, the manufacturer can establish a stronger foothold in the automotive industry, driving sustained growth and profitability.

The interactive dashboard offers a user-friendly interface to explore and interpret key insights from the analysis. Stakeholders can seamlessly navigate through visualizations, gaining a deeper understanding of consumer preferences, market dynamics, and influential factors in car pricing. This visualization tool is a powerful asset for the car manufacturer to make informed decisions and stay competitive in the dynamic automotive industry.

ABC Call Volume Trend Analysis

The project focuses on analyzing the inbound call data of ABC Insurance Company over a 23-day period. The dataset includes information on agents, queue time, call time, call duration, and call status. The primary objectives are to determine the average call duration for each time bucket, visualize call volume trends, propose a manpower plan to reduce abandon rates during regular hours, and address the issue of unanswered calls during night hours.

Methodology

The analysis was conducted using Microsoft Excel 2022. The dataset was imported into Excel, and various functions and features were utilized to derive insights. The project followed a step-by-step approach, starting with calculating average call durations, creating visualizations for call volume analysis, proposing a manpower plan for regular hours, and addressing night shift concerns.

Tech-Stack Used:

Software: Microsoft Excel 2022

Purpose: Excel was chosen for its powerful data analysis capabilities, ease of use, and familiarity. It provides essential functions for statistical analysis, chart creation, and data manipulation.

Finding – 1

Row Labels	Average of Call_Seconds (s)
10_11	203
11_12	198
12_13	191
13_14	194
14_15	193
15_16	199
16_17	199
17_18	199
18_19	203
19_20	203
20_21	201
9_10	196
Grand Total	**198**

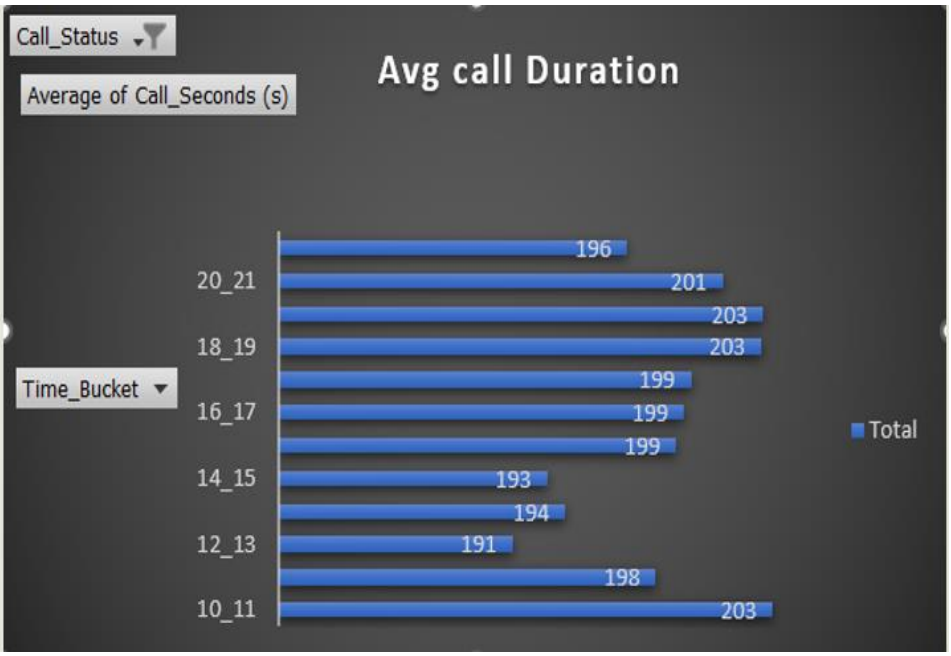


Fig.1 – Insight for Average Call Duration

Average Call Duration:

- Average call durations were calculated for each time bucket.
- Insights gained: Variability in call durations across different time buckets, potential patterns in customer behavior.

Finding – 2

Row Labels	Count of Customer_Phone_No	Count of Call_Seconds (s)
20_21	2420	3.45%
9_10	3520	5.02%
19_20	3872	5.52%
10_11	4894	6.98%
18_19	5523	7.88%
11_12	6714	9.58%
14_15	6777	9.67%
17_18	6821	9.73%
15_16	6853	9.77%
16_17	7037	10.04%
13_14	7626	10.88%
12_13	8054	11.49%
Grand Total	70111	100.00%

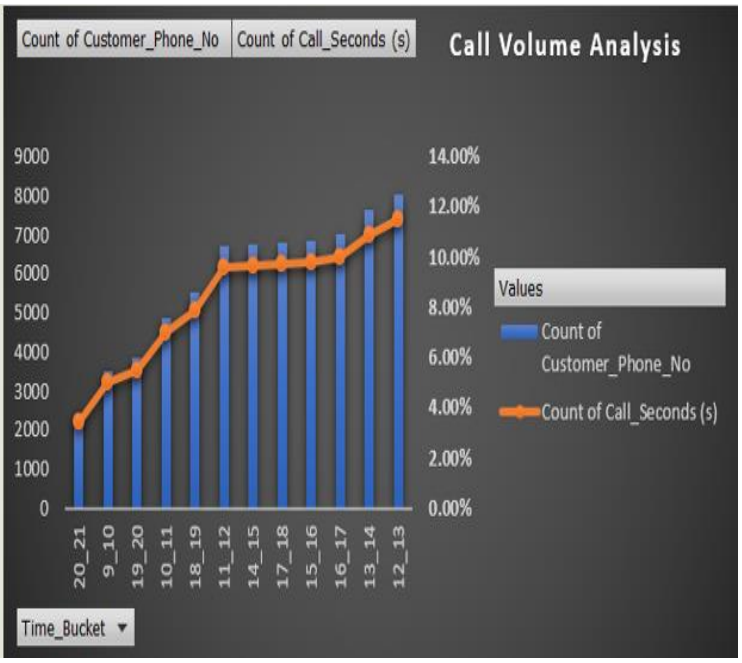


Fig.2 – Insight for Call Volume Analysis

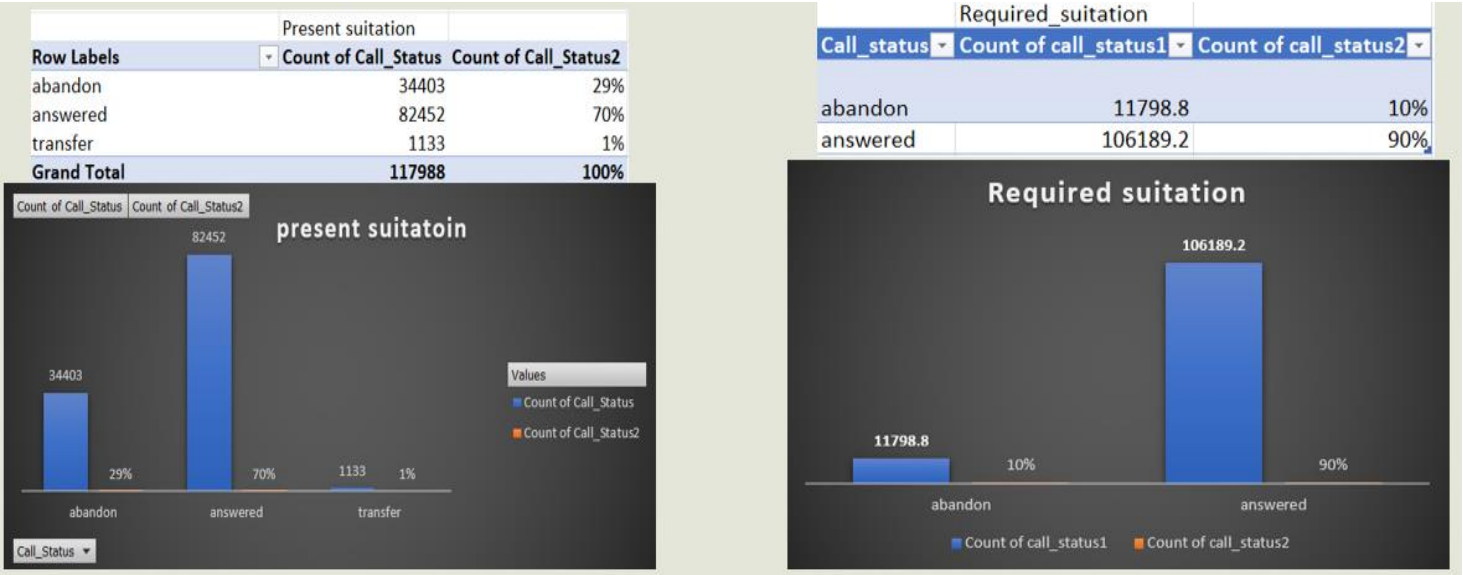
Call Volume Analysis:

- A graph was created to visualize the total number of calls received in each time bucket.
- Insights gained: Identification of peak call hours, understanding overall call distribution throughout the day.

Finding – 3

Manpower Planning:

- Proposed a plan to allocate agents during each time bucket to reduce the abandon rate to 10%.



Present suitation		
Row Labels	Count of Call_Status	Count of Call_Status2
abandon	34403	29%
answered	82452	70%
transfer	1133	1%
Grand Total	117988	100%

Required_suitation		
Call_status	Count of call_status1	Count of call_status2
abandon	11798.8	10%
answered	106189.2	90%

Fig.3 – Statistics for Manpower Planning

- Insights gained: Understanding the relationship between agent availability and call abandonment, optimizing manpower for better customer service.

Row Labels	Sum of Agent_Work	Sum of Agent_needed
18_19	3	3
20_21	2	3
19_20	2	3
15_16	3	4
17_18	3	4
16_17	3	4
14_15	4	5
13_14	4	5
9_10	3	5
10_11	5	6
12_13	5	6
11_12	5	7
Grand Total	42	55

Fig. 4 – Insight for Relationship b/w Agents Working vs Agent Needed

Finding – 4

Night Shift Manpower Planning:

- Proposed a manpower plan for night hours to address unanswered calls.

Row Labels	Call_distribution	Time_distribution	Agent_req
10_11	3	10	1.5
11_12	2	15	1
12_13	2	15	1
13_14	1	30	0.5
14_15	1	30	0.5
15_16	1	30	0.5
16_17	1	30	0.5
17_18	3	10	1.5
18_19	4	7.5	2
19_20	4	7.5	2
20_21	5	6	2.5
9_10	3	10	1.5
Grand Total	**30**	**1**	**15**

Task4	
avg no. of call in night=	1539
To increase hour call rate in night to 90%=	76
no. of agent required in night fr	17
Avg of call duration in sec=	198
Working hour of agent=	4.5

Fig.5 – Statistics for Night Shift Manpower Planning

Insights gained:

- Recognizing the importance of 24/7 customer service, ensuring a seamless experience for customers.

Row Labels	Sum of Agent_req
10_11	1.5
11_12	1
12_13	1
13_14	0.5
14_15	0.5
15_16	0.5
16_17	0.5
17_18	1.5
18_19	2
19_20	2
20_21	2.5
9_10	1.5
Grand Total	**15**

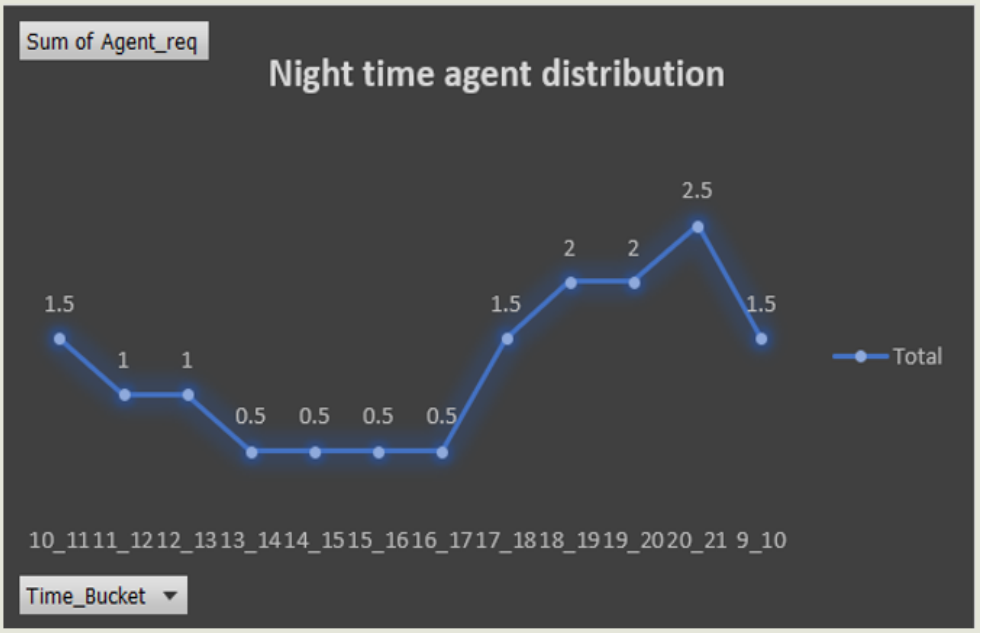


Fig. 6 – Insight for Night Time Agents Distribution

Result

1. Insightful Analytics:

- Call Duration Patterns: In-depth analysis revealed significant variations in call durations across different time buckets, providing insights into customer engagement and agent performance.
- Call Volume Trends: Identification of peak call hours and understanding the overall distribution helped in predicting high-demand periods, enabling proactive resource allocation.
- Agent Availability Impact: Analysis showcased the correlation between agent availability and abandonment rates, emphasizing the importance of optimizing workforce management.

2. Optimized Manpower Allocation:

- Abandon Rate Reduction: A pragmatic plan was formulated to decrease abandon rates during regular hours by optimizing the number of agents per time bucket. This ensures a balanced workload and improved customer service.
- Efficiency Enhancement: The proposed strategy minimizes customer wait times, increasing the likelihood of call resolution and overall satisfaction.

3. Enhanced Customer Experience:

- Night Shift Manpower Planning: Addressing concerns about unanswered calls during non-business hours, a

comprehensive plan was proposed to allocate resources efficiently during the night shift.

- Continuous Service Improvement: Ensuring 24/7 coverage contributes to an enhanced customer experience, demonstrating a commitment to customer satisfaction and loyalty.

Conclusion

The ABC Call Volume Trend Analysis project has successfully uncovered key insights and proposed actionable strategies to optimize agent allocation, reduce abandon rates, and improve the overall customer experience. By implementing the suggested recommendations and maintaining a focus on continuous improvement, ABC Insurance Company is poised to enhance its customer support operations and solidify its position as a customer-centric organization.

Appendix

- Bank Loan Excel sheet And Presentation Link : -

https://1drv.ms/x/c/e1641d3ba3ec422b/Efy1pc8Ae8FDunC4JB83J70B9UOfG1GKAqhlH0-gt6bNUg?e=qw0V2W

- Bank Loan Presentation Link :-

https://docs.google.com/present

ation/d/1d7e2AQrIqljhqFug_wNN

beaxGTc6Bj9M/edit?usp=drive_li

nk&ouid=118127998133245231361&rtpof=true&sd=true

ABC Call Volume Trend Analysis Excel sheet Link :-

https://1drv.ms/x/c/e1641d3ba3ec422b/EbhNXZOQQM1EsmzwNR7xWNEBoG72J6PtIArVV2mDW3rwTQ?e=HQnlQP

https://1drv.ms/x/c/e1641d3ba3ec422b/EeF223dbV99ChAn9nEhvDe8BQ474ZiTYjOuXC8DmibSpQA?e=VcNRiN

ABC Call Volume Trend Analysis Presentation Link :-

https://docs.google.com/presentation/d/1df05gaVR8zaz9JTQNVtAlpFyyK-L3hbf/edit?usp=drive_link&ouid=118127998133245231361&rtpof=true&sd=true

Analyzing the Impact of Car Features on Price and Profitability

Excel sheet Link :-

Car_features_excel_report.xlsx

Analyzing the Impact of Car Features on Price and Profitability
Presentation Link :-

https://docs.google.com/presentation/d/1s26O9W8dRzoVNVPzc3VqWhvfGbVz_fMo/edit?usp=drive_link&ouid=118127998133245231361&rtpof=true&sd=true