

Industry Project Report

On

Data Analysis and Power BI Dashboard

Developed By:-

Dhyan Patel (20162101009)

Guided By:-

Dr. Rohit Patel

Mr. Vishal Yadav

Submitted to
Department of Computer Science &
Engineering Institute of Computer
Technology



Ganpat
University

॥ विद्यया समाजोत्कर्षः ॥

Institute of
Computer
Technology



Year: 2024

CERTIFICATE

This is to certify that the **Industry Project** work entitled “**Data Analyst Tanning**” by Dhyan Patel (20162101009) of Ganpat University, towards the fulfilment of requirements of the degree of Bachelor of Technology – Computer Science and Engineering, carried out in the CSE(CBA) Department at Edulyt India . The results/findings contained in this Project have not been submitted in part or full to any other University / Institute for award of any other Degree/Diploma.

Name & Signature of Internal Guide

Name & Signature of Head

Place: ICT - GUNI

Date:12/05/2024

ACKNOWLEDGEMENT

An Industry Internship project is a golden opportunity for learning and self-development. I consider myself very lucky and honoured to have so many wonderful people lead me through in completion of this project. First and foremost, I would like to thank Dr. Rohit Patel, Principal, ICT, and Prof. Dharmesh Darji, Head, ICT who gave us an opportunity to undertake this project. My grateful thanks to Dr. Rohit Patel & Mr. Vishal Yadav (Internal & External Guides) for their guidance in project work, who despite being extraordinarily busy with academics, took time out to hear, guide, and keep us on the correct path. We do not know where would have been without his/her help. The CSE department monitored our progress and arranged all facilities to make life easier. We choose this moment to acknowledge their contribution gratefully.

Dhyan Patel (20162101009)

INDEX

Chapter	Content	Page No
1.	Introduction	6
2.	Introduction to MySQL and Basic Operations	7
3.	MySQL Advanced Concepts and Practical Applications	8
4.	Data Cleaning and Processing with MySQL	9
5.	Projects Overview	10
	5.1 Project1: credit card data analysis	10
	5.2 Project2: Indian credit Card Spending Analysis	15
6.	Power-BI Learning	19
7.	Course: Microsoft Power BI Data Analyst	20
8.	E-commerce platform sales Dashboard	21
	8.1 Data Gathering and Preparation	21
9.	Dashboard Development and Analysis	22
10.	Dashboard to Improve Employee Performance and Retention	23
	10.1 Data Gathering and Preparation	23
11.	Dashboard Planning and Development	24
12.	Dashboard Development and Analysis	26
13.	Dashboard Development and Analysis	28
14.	Sales Analysis Dashboard	29
	14.1 Data Gathering and Preparation	29
15.	Sales overview report and Customer Details Development	30
16.	Product Details report Development	32
	Conclusion	33
	References	34

Chapter - 1

Introduction

In the dynamic realm of data analytics, the role of a Data Analyst Intern at Edulyt India has been both illuminating and transformative. This report chronicles the journey undertaken during my internship, where the focus transcended traditional learning boundaries. Edulyt India, a pioneering start-up established in 2015, has been at the forefront of disrupting conventional education management practices, particularly in the domain of analytics. My internship aligns seamlessly with the organization's mission to bridge the gap between education and employment, emphasizing the cultivation of fresh graduates into industry-ready professionals.

As a Data Analyst Intern, the exploration delves deep into the intricacies of analytics within the Banking, Financial Services, and Insurance (BFSI) sector. This introduction sets the stage for a comprehensive narrative, detailing the objectives, methodologies, tools, and technologies employed, as well as the practical implementation of analytical processes. The journey undertaken speaks to the synthesis of theoretical knowledge and hands-on application, navigating real-world challenges in the pursuit of data-driven insights. Through this report, I aim to convey not just the technical intricacies of data analytics but also the profound impact and learning derived from contributing to Edulyt India's mission and vision.

Chapter - 2

Introduction to MySQL and Basic Operations

- Focused on learning the basics of MySQL, including its role within relational databases and the significance of managing structured data. Following that, I downloaded and installed MySQL, making sure to configure it for optimal performance.
- Focus shifted towards creating databases and tables. I delved into understanding various data types and the enforcement of constraints. Additionally, I familiarized myself with fundamental commands such as SELECT, INSERT, UPDATE, and DELETE, which formed the basis for data manipulation.
- I dedicated my efforts to mastering the SELECT statement. I delved into the intricacies of filtering data using WHERE clauses, sorting results with ORDER BY, and utilizing aggregate functions for data analysis. This newfound knowledge significantly bolstered my capability to extract meaningful insights from databases.

Chapter - 3

Advanced Concepts and Practical Applications

- In this Chapter I deep dive into advanced concepts. I explored various types of joins such as INNER, LEFT, and RIGHT, which enabled me to merge data from multiple tables effectively. Additionally, I learned about the power of subqueries as a tool for crafting more intricate and efficient queries. These concepts opened up new avenues for manipulating and extracting insights from relational databases.
- My focus was on gaining practical experience with data modification operations. I learned to execute insertions, updates, and deletions of records while ensuring data consistency. Understanding transactions became pivotal as I delved into their role in preserving data integrity. This hands-on experience provided valuable insights into managing and safeguarding database information.
- Optimization and Review The final day of the two-week period focused on optimization techniques. I learned about indexing strategies and performance considerations to enhance database efficiency. The week concluded with a comprehensive review of the learned concepts, solidifying my understanding of MySQL fundamentals

Chapter - 4

Data Cleaning and Processing with MySQL

- This Chapter kicked off with an in-depth exploration of data cleaning concepts within MySQL. I focused on grasping the significance of data cleanliness, identifying prevalent data quality issues, and implementing strategies to rectify anomalies and inconsistencies within the database. This understanding laid a strong foundation for ensuring the accuracy and reliability of the data stored in MySQL.
- The focus transitioned to data processing techniques utilizing MySQL. I gained proficiency in aggregating and transforming data to extract meaningful insights. Techniques like GROUP BY, HAVING, and CASE statements were utilized to manipulate and process data effectively. This knowledge empowered me to derive valuable insights and make informed decisions based on the data stored in MySQL.
- I engaged in hands-on exercises focused on data cleaning and processing. I applied the concepts learned to real-world datasets, tackling messy data and implementing processing techniques to prepare it for analysis or reporting. These practical exercises provided valuable experience in managing data effectively within MySQL and ensured that I could confidently handle various data challenges in a real-world context.

Chapter - 5

Project Overview

5.1) Project-1: Credit Card Data Analysis

Purpose: The purpose of the Credit Card Data Analysis project is to perform a comprehensive analysis of credit card transaction data, with the goal of extracting meaningful insights regarding spending patterns and financial trends. The primary objectives revolve around the extraction, transformation, and loading (ETL) of the credit card data, followed by Exploratory Data Analysis (EDA) to unveil valuable information. This process involves cleansing and pre-processing the data to ensure its accuracy and reliability, then employing various analytical techniques to identify patterns, trends, and anomalies within the dataset. Ultimately, the project aims to provide actionable insights that can inform decision-making processes for financial institutions or individuals.

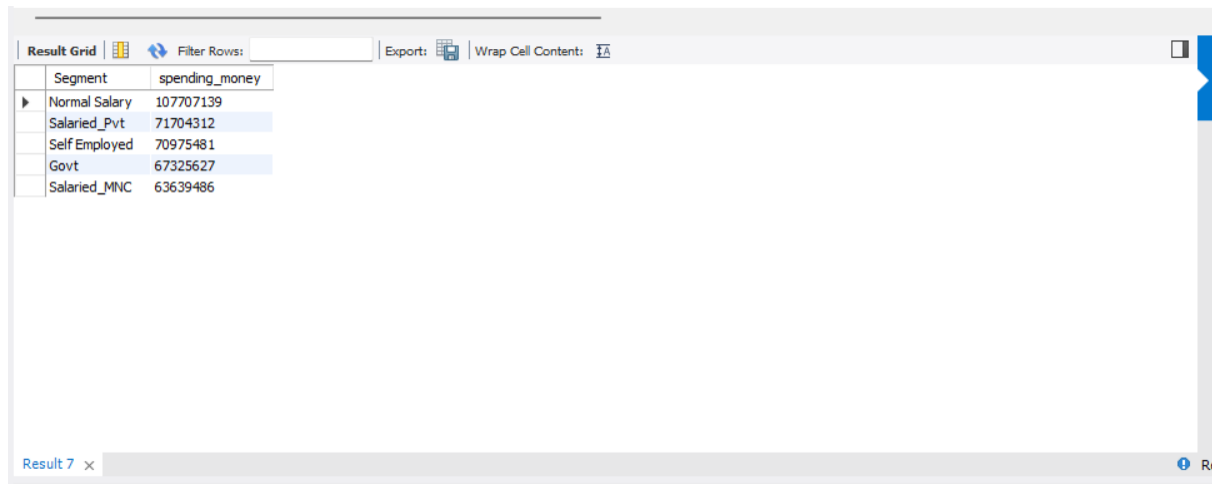
Scope: This project is primarily focused on comprehending the patterns and behaviours inherent in credit card transactions, with a special emphasis on factors like transaction volumes, spending categories, and anomalies. The objective extends to identifying trends and insights that can guide strategic decision-making concerning credit card usage. By analysing transaction data in depth, the project aims to uncover actionable insights that can help financial institutions and individuals make informed decisions regarding credit card management, risk assessment, and financial planning.

- started with a thorough review of the project requirements for Credit Card Data Analysis. I began by immersing myself in the dataset, comprehending its structure, and pinpointing pertinent variables for analysis. Through this initial exploration, I gleaned valuable insights to lay the groundwork for subsequent steps in the project. This process ensured a solid understanding of the data and set the stage for more in-depth analysis and exploration.
- A considerable portion of the week was allocated to data cleaning and preparation for the Credit Card Data Analysis project. I meticulously tackled missing values, outliers, and inconsistencies present within the dataset. Various transformations were applied to guarantee data accuracy and readiness for analysis. This rigorous process was essential for ensuring the reliability and integrity of the data, laying a solid foundation for subsequent analysis and interpretation.
- With a pristine dataset at hand, I delved into exploratory data analysis (EDA) during this phase. Leveraging statistical measures and visualizations, I meticulously uncovered patterns, trends, and potential correlations within the credit card data. This crucial step laid the groundwork for informed decision-making in subsequent stages of the analysis. By gaining insights into the underlying dynamics of the data, I set the stage for more in-depth analysis and interpretation.

A sneak peek at the project solution highlights:

Q1. Which age group is spending more money?

```
SELECT Segment ,sum(Amount) as spending_money  
FROM cb join spend on spend.customer = cb.customer  
group by Segment  
order by spending money desc;
```



The screenshot shows a SQL query result grid with two columns: 'Segment' and 'spending_money'. The results are ordered by spending money in descending order. The segments and their corresponding spending money values are:

Segment	spending_money
Normal Salary	107707139
Salaried_Pvt	71704312
Self Employed	70975481
Govt	67325627
Salaried_MINC	63639486

Fig 5.1 solution of Q1

Q2. Which age group is spending more money?

```
SELECT SUM(Amount) AS total_spending ,  
CASE  
WHEN Age < 18 THEN 'Under 18'  
WHEN Age >= 18 AND Age < 30 THEN '18-29'  
WHEN Age >= 30 AND Age < 40 THEN '30-39'  
ELSE '40 and above'  
END AS age_group  
from cb  
join spend on cb.customer = spend.customer  
GROUP BY age_group  
ORDER BY total_spending DESC;
```

total_spending	age_group
223263092	40 and above
93433334	Under 18
46088736	18-29
18566883	30-39

Fig 5.2 solution of Q2

Q3.Which is the most profitable segment?

SELECT Segment, SUM(spend.Amount) ,SUM(repayment.Amount) ,

CASE

WHEN SUM(repayment.Amount) > SUM(spend.Amount) then 'segment_profit'

ELSE '0'

END AS segment_profit

FROM cb

join spend on cb.customer = spend.costomer

join repayment on cb.customer =repayment.costomer

group by Segment

ORDER BY segment_profit DESC;

Segment	SUM(spend.Amount)	SUM(repayment.Amount)	segment_profit
Self Employed	1772234101	1803292711.5068533	segment_profit
Salaried_Pvt	1888313184	177123681.553834	0
Govt	1615819450	1549329701.8628302	0
Normal Salary	2532062414	2476206234.4449975	0
Salaried_MNC	1643150424	1599110209.7495804	0

Fig 5.3 solution of Q3

Q4. In which category the customers are spending more money?

```
SELECT type, sum(Amount) as Spending FROM spend
```

```
GROUP BY type
```

```
ORDER BY Spending DESC
```

```
;
```

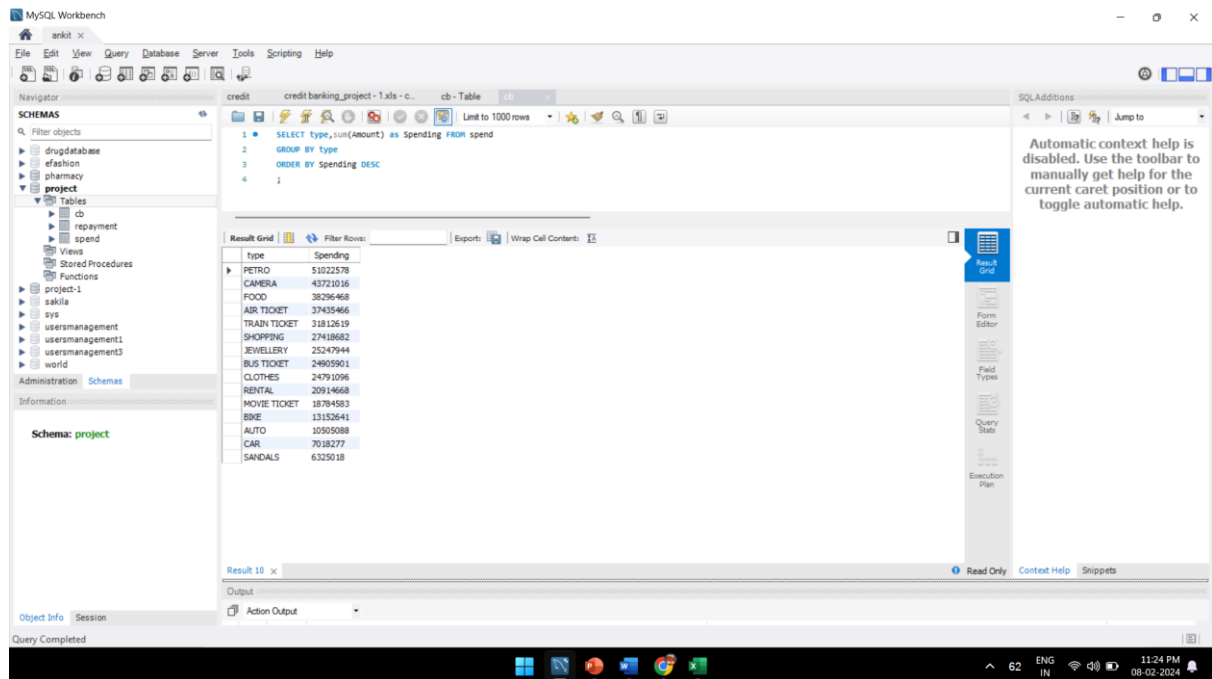


Fig 5.4 solution of Q4

Q5. Monthly profit for the bank.

```
SELECT
```

```
MONTH(month) AS monthly,
```

```
SUM(Amount) AS monthly_spend,
```

```
`limit`, -- Enclose the column name in backticks
```

```
CASE
```

```
WHEN SUM(Amount) > `limit` THEN (`limit` * 0.02)
```

```
ELSE 0
```

```
END as bank_profit
```

```
FROM
```

```
spend as a
```

```
JOIN
```

cb as b ON a.customer = b.customer

GROUP BY

MONTH(month);

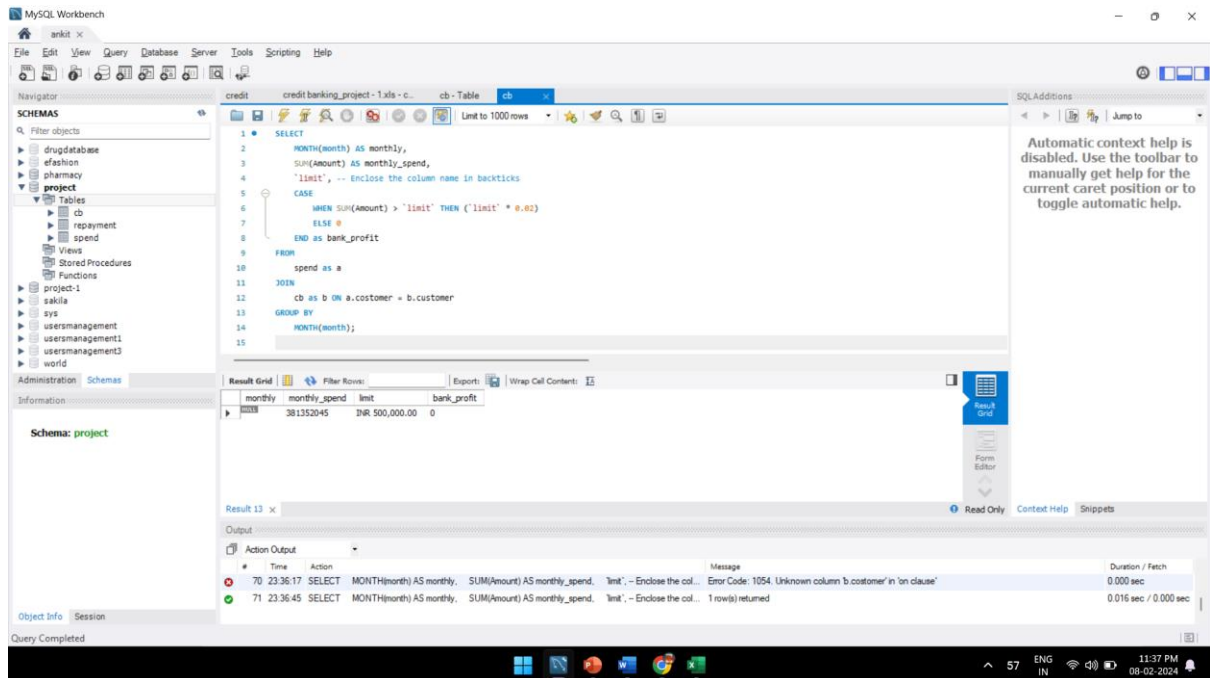


Fig 5.5 solution of Q5

5.2)Project - 2 Overview: Indian Credit Card Spending Analysis

Purpose:

The Indian Credit Card Spending Analysis project aims to explore a dataset providing insights into credit card transactions conducted across various cities in India. The primary objective is to develop a thorough understanding of spending habits by utilizing data on transaction dates, card types, expense categories, gender, and transaction amounts. The project entails leveraging SQL queries to extract meaningful patterns and trends from the dataset. By employing SQL techniques, such as aggregations, joins, and subqueries, the project aims to uncover valuable insights that can inform decision-making processes related to financial planning, marketing strategies, and risk assessment.

Scope:

The scope of this project involves delving into spending behaviors within Indian credit card transactions, with a keen focus on cities, card types, expense categories, and gender-related insights. The overarching objective is to pinpoint trends, anomalies, and key contributors to credit card spending, thereby facilitating strategic decision-making for various stakeholders. By conducting in-depth analyses and employing statistical and data visualization techniques, the project aims to unearth actionable insights that can guide financial institutions, businesses, and policymakers in optimizing their strategies, mitigating risks, and enhancing customer experiences in the realm of credit card usage.

A sneak peek at the project solution highlights:

1. Write a query to print highest spend month and amount spent in that month for each card type.

```
SELECT [Card type], month_name, year_name, total_spend
FROM (
    SELECT [Card type],
           DATEPART(year, [date]) AS year_name,
           DATENAME(month, [date]) AS month_name,
           SUM(amount) AS total_spend,
           DENSE_RANK() OVER (PARTITION BY [Card type] ORDER BY SUM(amount)
                               DESC) AS rank
    FROM credit_card_transactions
    GROUP BY [Card type], DATEPART(year, [date]), DATENAME(month, [date])
) AS card_type_data
WHERE rank = 1;
```

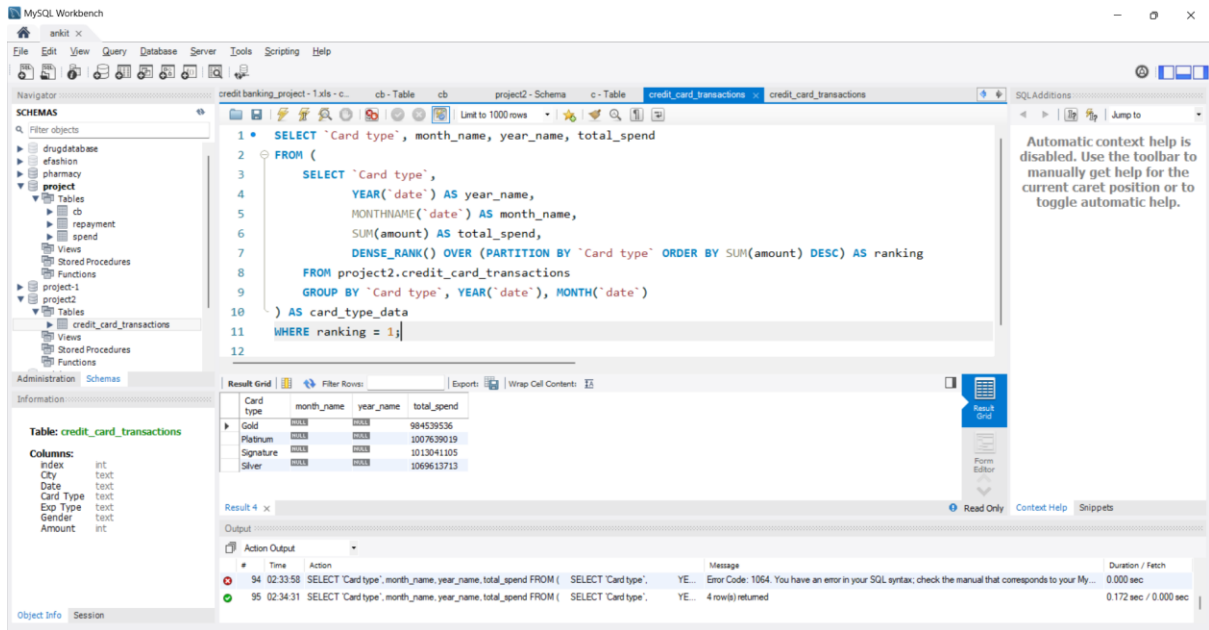


Fig 5.6 solution of Q1

2. Write a query to find city which had lowest percentage spend for gold card type.

WITH silver_amt_citywise AS (

SELECT city, SUM(amount) AS silver_amount

FROM credit_card_transactions

WHERE `Card type` = 'Silver'

GROUP BY city

),

total_amt_citywise AS (

SELECT city, SUM(amount) AS total_amount

FROM credit_card_transactions

GROUP BY city

),

contribution AS (

SELECT s.city, s.silver_amount, t.total_amount, ROUND(100 * s.silver_amount / t.total_amount, 2) AS percentage_contribution

FROM silver_amt_citywise AS s

INNER JOIN total_amt_citywise AS t ON s.city = t.city

)

SELECT *

FROM contribution

ORDER BY percentage_contribution ASC

LIMIT 1;

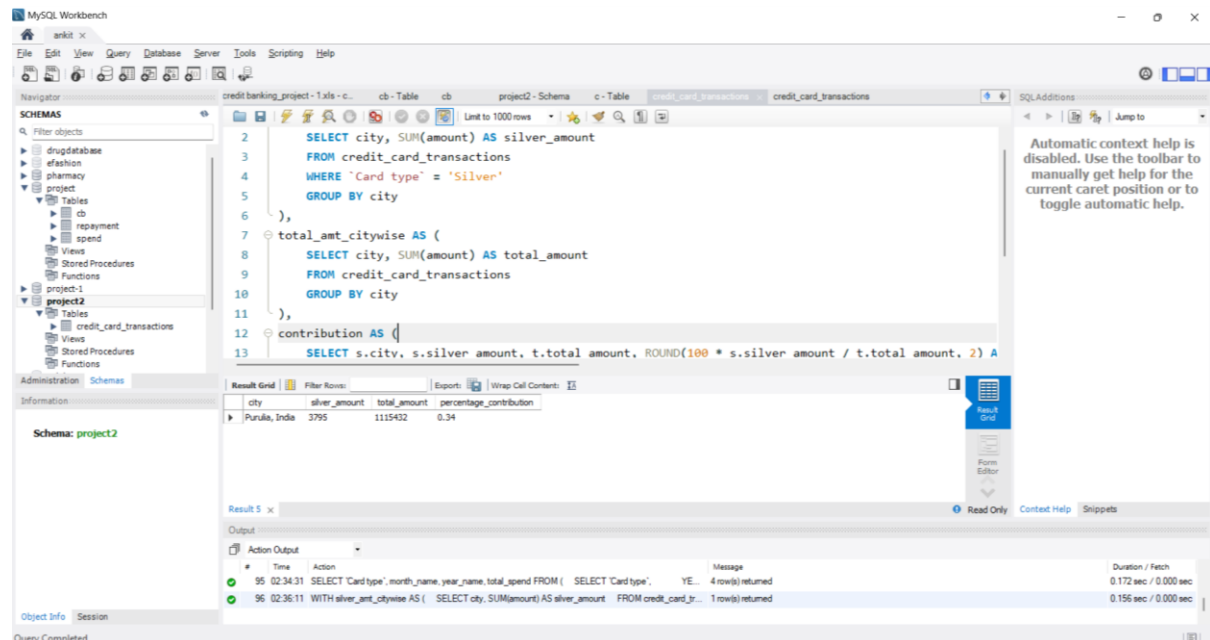


Fig 5.7 solution of Q2

3. During weekends which city has highest total spend to total no of transaction's ratio?

SELECT

City,

SUM(Amount) / COUNT(*) AS SpendToTransactionRatio

FROM (

SELECT

City,

STR_TO_DATE(Date, '%d-%b-%y') AS TransactionDate,

Amount

FROM

credit_card_transaction -- Replace with your actual table name

) AS converted_data

WHERE

DAYOFWEEK(TransactionDate) IN (1, 7) -- Assuming 1 is Sunday and 7 is Saturday

GROUP BY

City

ORDER BY

SpendToTransactionRatio DESC

LIMIT 5;

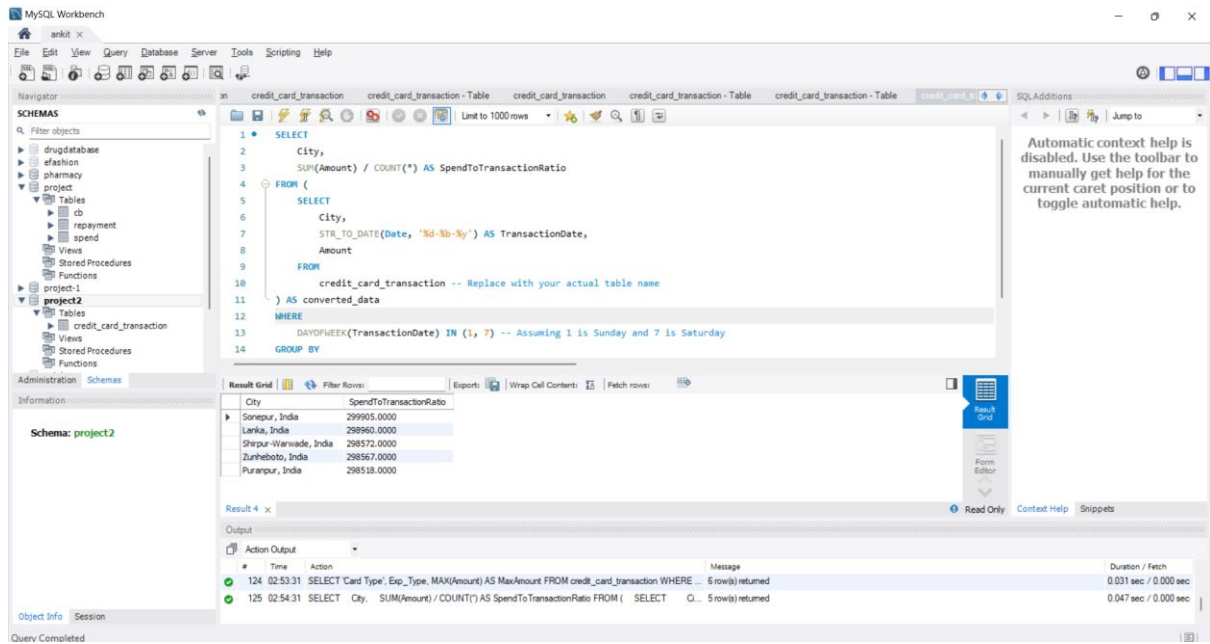


Fig 5.8 solution of Q3

Chapter - 6

Power-BI Learning

- Introduction to Power BI and its components.
- Connecting to various data sources and importing data.
- Basics of data modeling and creating relationships.
- Visualization techniques and dashboard creation..
- Advanced data modeling with DAX language.
- Complex visualization techniques and interactivity.
- Data governance, security, and compliance.

Chapter - 7

Course PL-300T00--A: Microsoft Power BI Data Analyst

This course covers the various methods and best practices that are in line with business and technical requirements for modelling, visualizing, and analyzing data with Power BI. The course will show how to access and process data from a range of data sources including both relational and non-relational sources. Finally, this course will also discuss how to manage and deploy reports and dashboards for sharing and content distribution.

Outcome of the course:

- **Proficiency in Microsoft Power BI:** Participants gain a comprehensive understanding of the Power BI platform, including its various features, functionalities, and tools for data analysis and visualization.
- **Data Analysis Skills:** Learners develop strong data analysis skills, enabling them to extract insights from complex datasets using Power BI's advanced analytics capabilities.
- **Data Modelling:** Participants learn how to create effective data models within Power BI, allowing them to structure data for analysis and visualization purposes.
- **Report Creation and Visualization:** The course equips learners with the skills to design and build compelling reports and dashboards using Power BI's intuitive interface and visualization tools.
- **Collaboration and Sharing:** Students learn how to collaborate with team members and share insights by publishing reports to the Power BI service and leveraging collaboration features.
- **Data Connectivity:** Participants understand how to connect to various data sources, including databases, files, and online services, to import and transform data for analysis in Power BI.
- **Power BI Desktop and Power BI Service:** The course covers both Power BI Desktop and Power BI Service, ensuring participants are proficient in using both tools for end-to-end data analysis and visualization workflows.

Chapter - 8

E-commerce platform sales Dashboard

Project 1 - Analyse the E-commerce platform sales

Focus: The project aims to analyze their online sales across India.

Goal: Develop an interactive Power BI dashboard for actionable insights.

Objective: owner of this store wants to track and analyze their online sales across India.

8.1) Data Gathering and Preparation

- **Data collection** imported the data from online resources.
- **Data cleaning** Conducted data pre-processing tasks, including cleaning, formatting, and transforming the data to make it suitable for analysis.
- **Data manipulation** through this one extra column average order value added in data report by dax in power bi.
- Implemented the data model in Power BI for analysis.

Chapter - 9

Dashboard Development and Analysis

- Developed an interactive Power BI dashboard.
- Developed an interactive dashboard in Power BI, created bar column chart ,donut chart, pie chart, line chart, slicer ,card etc. for visualization.
- Implemented navigation features for user exploration.
- Add interactivity to the dashboard elements, such as filters, and drill-down functionalities, to allow users to explore the data dynamically.

Screenshot:



Fig 9.1 E-commerce sales Dashboard

Chapter - 10

Dashboard to Improve Employee Performance and Retention

Project 2 : Analytics Dashboard to Improve Employee Performance and Retention

Objective: The purpose of this project is to help an organization to improve employee performance and reduce attrition by creating an HR analytics dashboard. The dashboard provides valuable insights into employee data, which can be used to make data-driven decisions and improve employee satisfaction and retention.

10.1) Data Gathering and Pre-processing

- **Data Source :** The raw data for this project was obtained in the form of a CSV file with 38 columns and approximately 1.5k rows. The data contained information about employee demographics, job roles, salaries, and tenure, among others.
- **Data Cleaning and Processing:** Several steps were taken to prepare the data for analysis. Null values were removed and duplicate values were eliminated.
- **Key Performance Indicators (KPIs):** To start the analysis, I have identified the key performance indicators (KPIs) to track and monitor employee performance and attrition.

Chapter - 11

Dashboard Planning and Development

Planned and designed interactive dashboards in Power BI to visualize used several charts and visualizations.

Developed Demographics dashboards with appropriate visualizations, including charts, graphs, and matrix table, to showcase key insights effectively.

Screenshots:

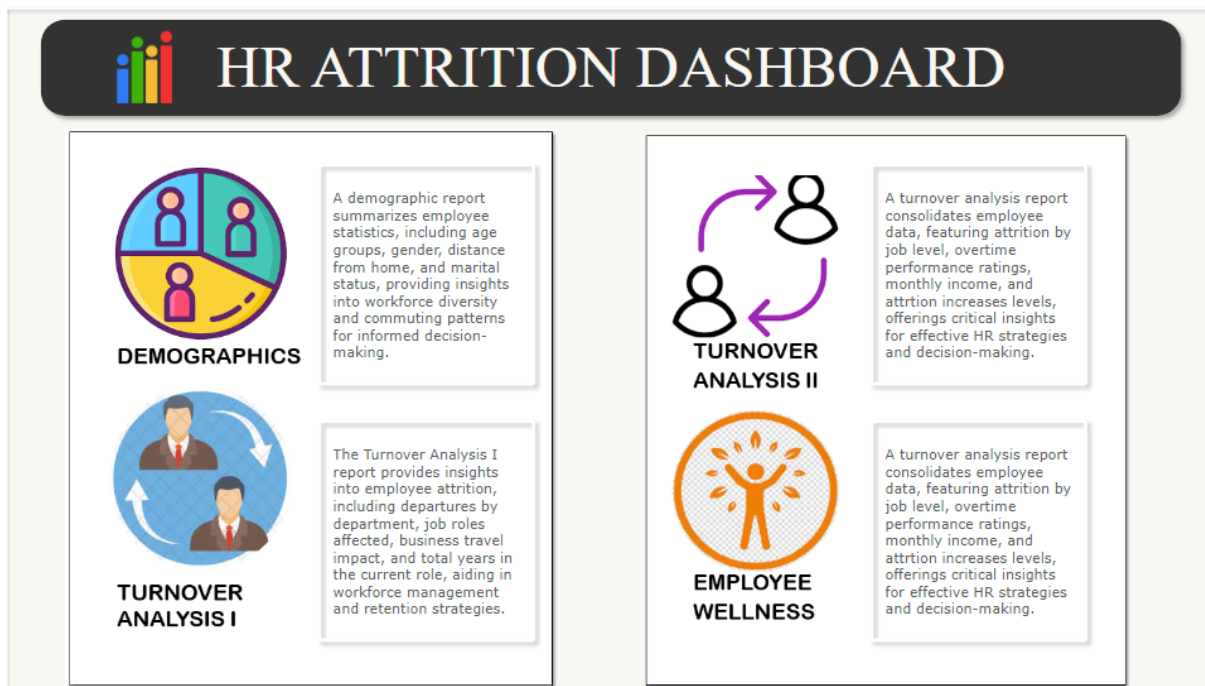


Fig 11.1 HR Attrition Dashboard

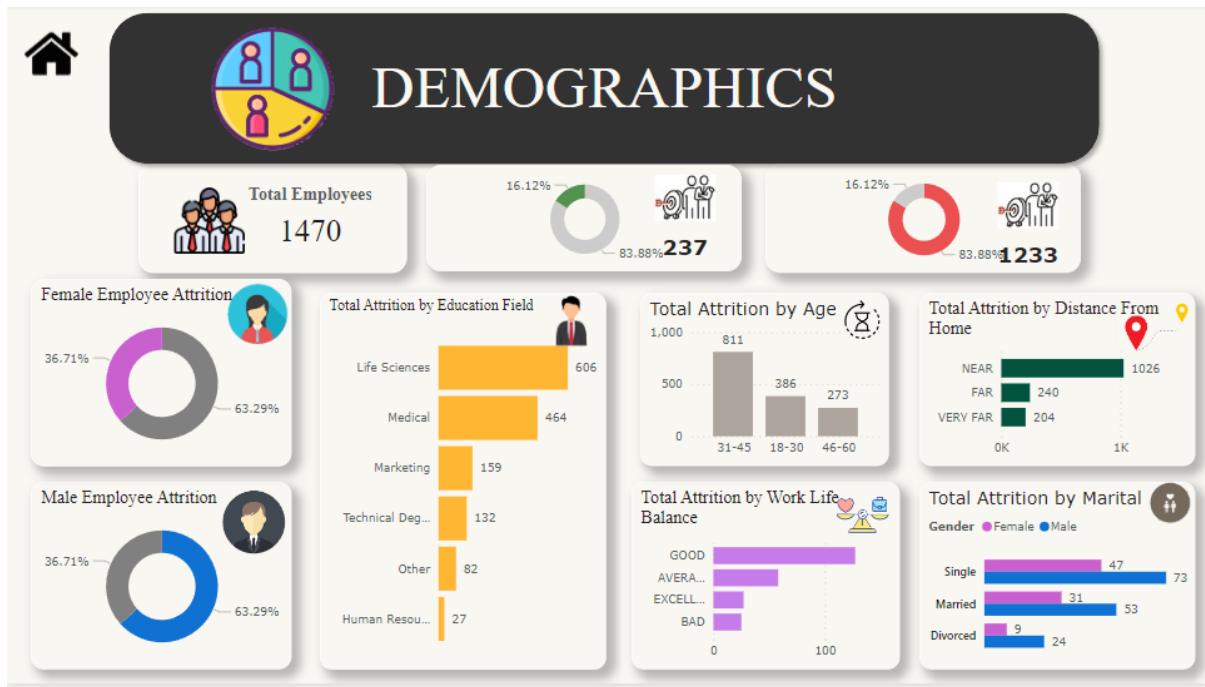


Fig 11.2 Demographics

Chapter - 12

Dashboard to Improve Employee Performance and Retention

Created Turnover Analysis 1 report provides insights into employee attrition including departures by department, job roles affected, business travel impact and retention strategies.

Screenshot:

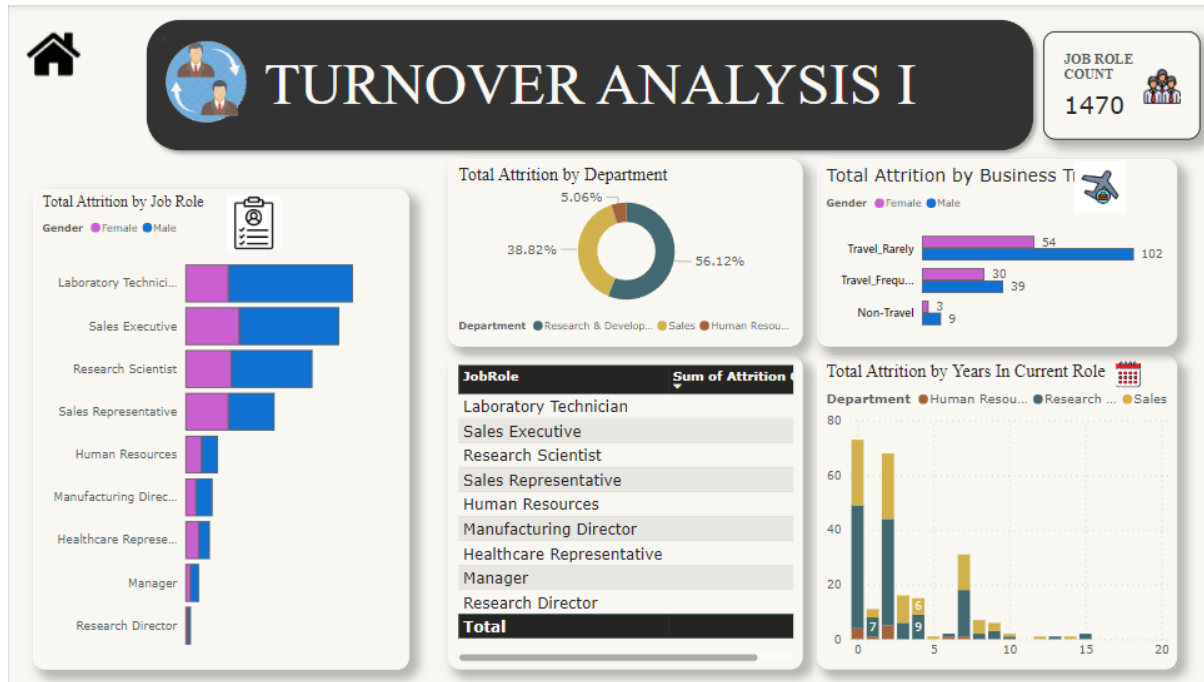


Fig 12.1 Turnover Analysis 1 Dashboard

Created a turnover analysis 2 report consolidates employee data, featuring attrition by job level, overtime performance ratings, monthly income, and attrition increases levels, offerings critical insights for effective HR strategies and decision-making.

Screenshot:

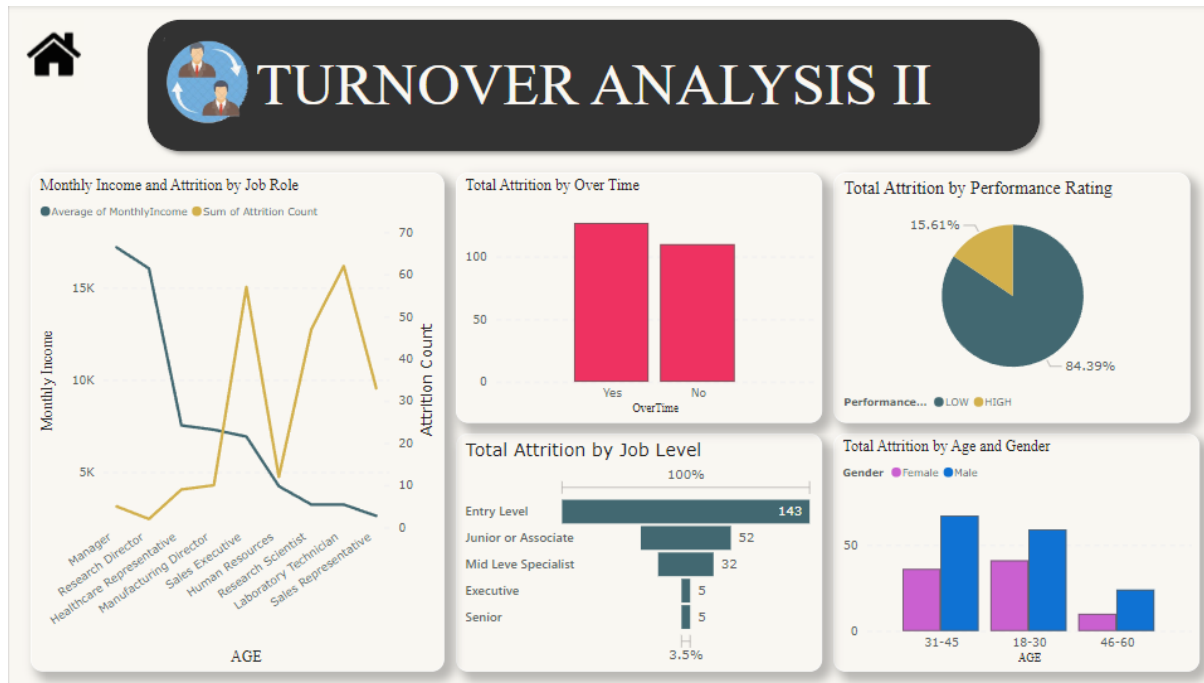


Fig 12.2 Turnover Analysis 2 Dashboard

Chapter - 13

Dashboard to Improve Employee Performance and Retention

Created a Employee wellness report

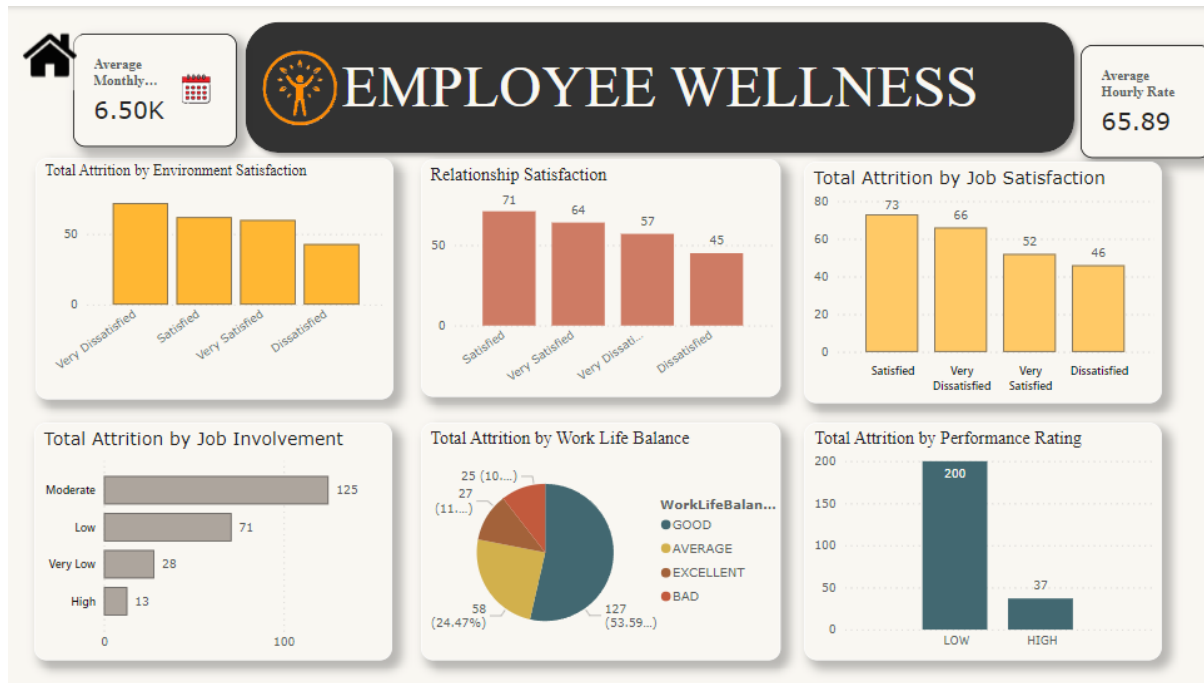


Fig 13.1 Employee Wellness Dashboard

Conclusion:

This HR analytics dashboard showed important information about employees that can help make better decisions and keep employees happy and working for the organization. By taking actions to address the issues that employees face and creating a positive work environment, the HR department can help employees perform better and stay in their jobs longer. This can help the organization make more money and be more successful.

Chapter – 14

Sales Analysis Dashboard

Overview: The Sales Analysis Dashboard project aims to provide comprehensive insights into sales performance, customer behavior, and product trends through interactive visualization using Power BI. By integrating sales overview, customer details, and product details, this dashboard empowers stakeholders to make informed decisions, optimize strategies, and enhance business outcomes.

14.1) Data Collection and Preparation.

- Get data from Kaggle
- Load CSV to PowerBI
- Clean data eliminating null/blank
- Identify metrics to extract from the report
- Present the data visually with key metrics

Chapter - 15

sales overview report Customer Details Development

worked on sales overview report

- Visualizations depicting sales metrics such as revenue, units sold, profit margins, and sales growth over time.
- Comparative analysis of sales performance across regions, products, and time periods.
- Key performance indicators (KPIs) highlighting targets versus actual performance.

Screenshot:

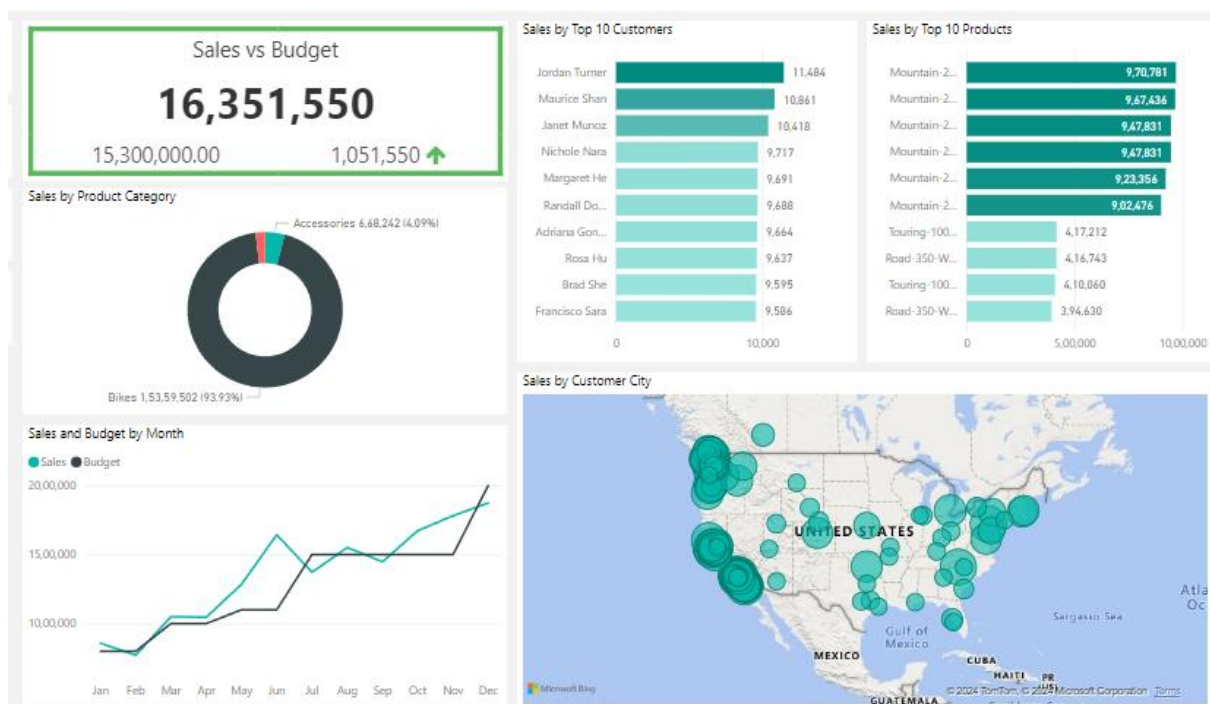


Fig 15.1 Sales Report

Worked on Customer Details

- Customer segmentation based on demographics, purchasing frequency, and buying preferences.
- Customer lifetime value (CLV) analysis to identify high-value customers.
- Visualization of customer churn rates and retention strategies.

Screenshot:

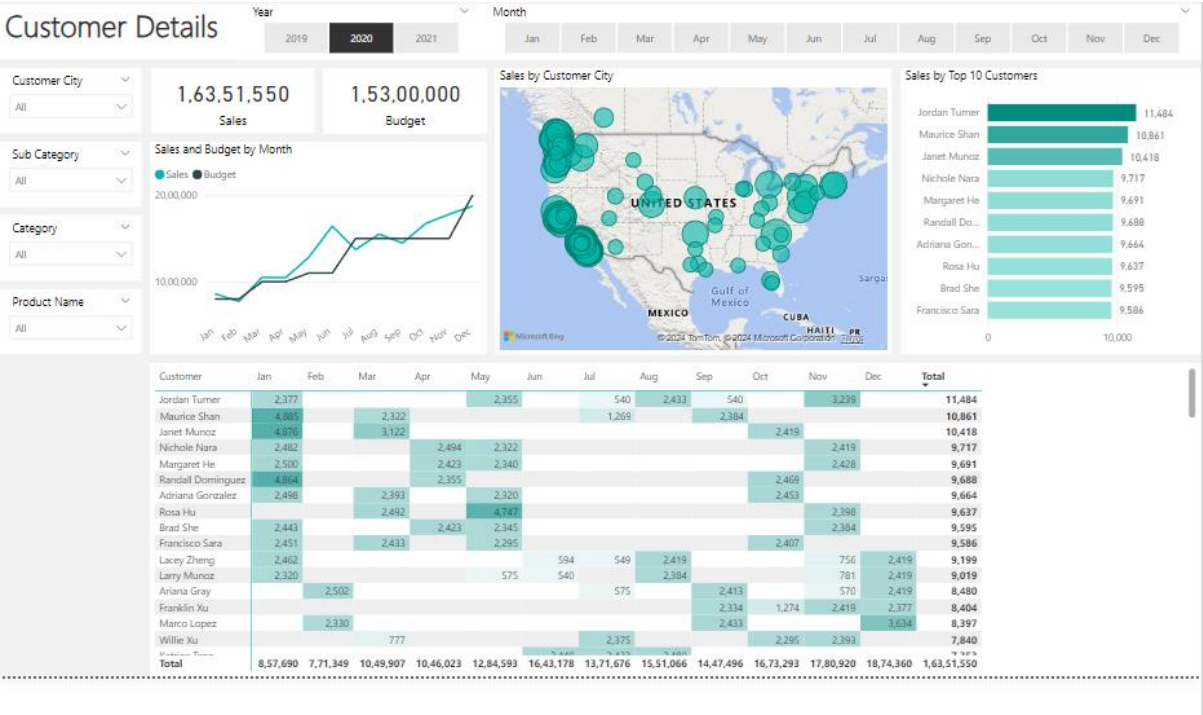


Fig 15.2 Customer Details

Chapter - 16

product Details Report Development

Worked on product Details report

- Product sales distribution by category, brand, or SKU.
- Analysis of product profitability, including gross margin and contribution to overall revenue.
- Product performance trends, highlighting top-selling items, slow movers, and seasonality effects.

Screenshot:

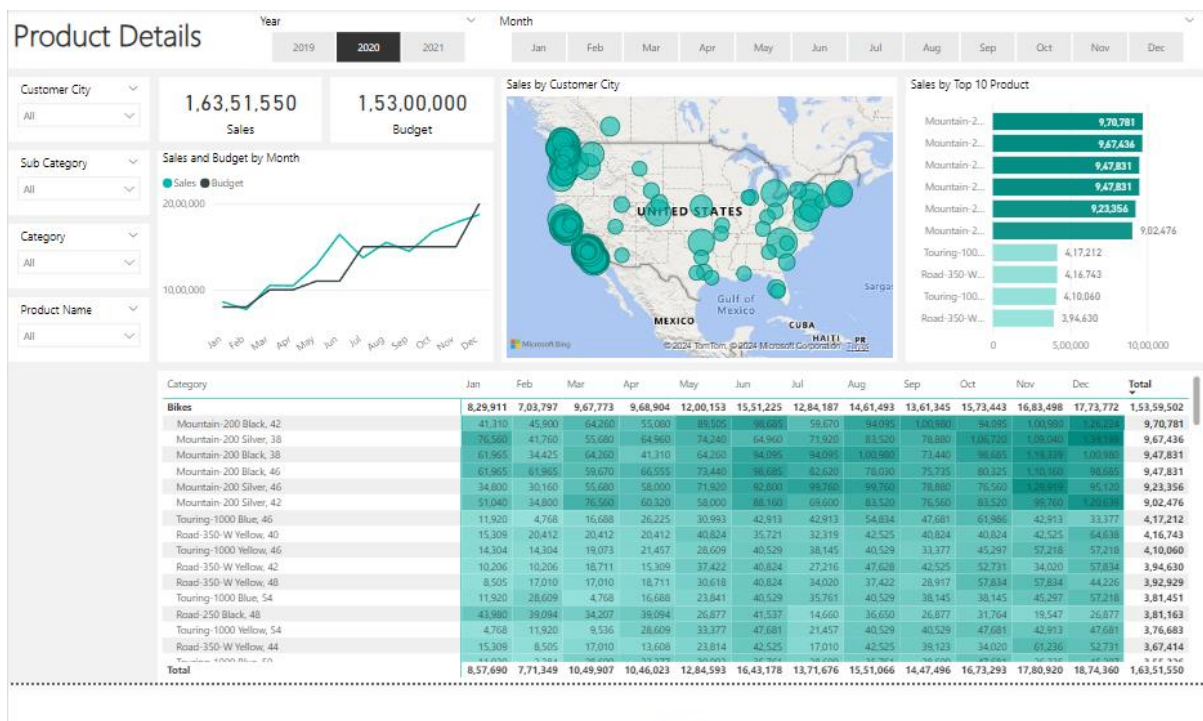


Fig 16.1 Product Details Report

Conclusion:

The Sales Analysis Dashboard serves as a powerful tool for driving sales performance, understanding customer dynamics, and optimizing product strategies. By leveraging the capabilities of Power BI, this project empowers organizations to harness the full potential of their sales data and gain a competitive edge in today's dynamic market landscape.

Conclusion

During my internship at Edulyt India, a finance firm, I gained invaluable understanding about the significance of information management in the workplace. Through extensive work with diverse datasets and crafting numerous Power BI dashboards, I developed a distinct viewpoint on data analysis within the financial industry. My time at Edulyt revealed the broad spectrum of tasks a data analyst can handle, utilizing technical skills such as web development and SQL, in conjunction with statistical expertise, to tackle assigned duties proficiently. Skilled in developing proficient Power BI dashboards.

References

Power BI YouTube Channel: <https://www.youtube.com/user/mspowerbi>

Power BI Course : <https://learn.microsoft.com/en-us/training/courses/pl-300t00>

Power BI Documentation: <https://learn.microsoft.com/en-us/power-bi/>

Plagiarism Report

