CONSOLIDATED PROJECT REPORT

Data Analytics and Business Intelligence Solutions

The domain of the Project:
Power BI and SQL

Under the guidance of
Ms. Siddhika (Software Engineer)

By:
Ms. Anjali A S (B.Tech CSE Graduate)

Period of the project

May 2024 to August 2025

Declaration

This consolidated project report encompasses three major projects undertaken under the mentorship of Ms. Siddhika, organised by SURE Trust, from May 2024 to August 2025, for the benefit of the educated unemployed rural youth for gaining hands-on experience in working on industry relevant projects that would take them closer to the prospective employer. I declare that to the best of my knowledge, I have worked on all projects successfully and enhanced my practical knowledge in the domain of data analytics, business intelligence, and database integration.

Projects Included:

1. Website Usage Analysis
2. Automated Data Pipeline: Google Forms to MySQL Integration
3. Coffee Shop Sales Analysis

Name

Ms. Anjali A S

Mentor
Software Engineer—HCL Tech

Prof. Radhakumari
Executive Director & Founder
SURE Trust

Table of contents

Executive Summary

This consolidated report presents three comprehensive data analytics projects undertaken to demonstrate proficiency in Power BI, SQL, and automated data pipeline development. The projects showcase the application of modern business intelligence tools and techniques to solve real-world data challenges across different domains: web analytics, data integration, and retail sales analysis.

The first project, Website Usage Analysis, focused on extracting actionable insights from website traffic data using Power BI dashboards. Key findings revealed that India and the USA contribute the majority of website traffic, with Chrome as the dominant browser. Engagement peaks on Mondays, and June recorded the highest monthly activity. The project successfully transformed 8.37M page views into clear visualizations showing traffic patterns, user behavior, and browser performance metrics.

The second project, Automated Data Pipeline: Google Forms to MySQL Integration, demonstrated the automation of data collection workflows. Using Python, Google Sheets API, and MySQL, the system successfully automated the transfer of form responses into a structured database. The solution eliminated manual data entry, ensured data integrity through duplicate checking, and provided a scalable framework for real-time data synchronization.

The third project, Coffee Shop Sales Analysis, analyzed retail sales performance across multiple coffee shop locations. The dashboard revealed total sales of 699K with 214K quantity sold, highlighting Hell's Kitchen, Astoria, and Lower Manhattan as top-performing locations. Coffee and Tea dominated category sales, while profit margins improved from 12% to 24% over the analysis period.

Collectively, these projects demonstrate:

- Advanced proficiency in Power BI for data visualization and dashboard creation
- Strong SQL and database management capabilities
- Python automation skills for data pipeline development
- Ability to extract actionable business insights from complex datasets
- Experience with cloud-based data integration using Google APIs

The projects underscore the critical importance of data-driven decision-making in modern organizations and showcase how affordable business intelligence tools can provide enterprise-level insights for businesses of all sizes.

Introduction

Background and Context

In today's data-driven economy, organizations across all sectors generate vast amounts of information through their digital operations, customer interactions, and business processes. The ability to transform this raw data into meaningful insights has become a critical competitive advantage, enabling evidence-based decision-making, operational optimization, and strategic planning. However, many organizations struggle to harness the full potential of their data due to inadequate analytical capabilities, lack of visualization tools, or inefficient data management processes.

This consolidated project report addresses these challenges through three distinct yet complementary projects that demonstrate comprehensive data analytics capabilities across different business scenarios. Each project tackles specific data challenges while collectively showcasing proficiency in modern business intelligence tools, database management systems, and automated data processing workflows.

Problem Statement / Goals

Organizations frequently face several critical data-related challenges:

- Large volumes of unstructured data that remain untapped for business insights
- Manual data collection and processing workflows that are time-consuming and error-prone
- Lack of real-time visibility into key performance indicators and business metrics
- Difficulty in identifying trends, patterns, and opportunities hidden within complex datasets
- Inadequate visualization tools that fail to communicate insights effectively to stakeholders

The primary goal of these projects was to demonstrate how modern data analytics tools and techniques can address these challenges by:

- Converting raw data into interactive, user-friendly dashboards and reports
- Automating data collection and processing workflows to improve efficiency and accuracy
- Providing real-time insights that support data-driven decision-making
- Creating scalable solutions that can be adapted to various business contexts

Scope and Limitations

The scope of this consolidated project encompasses three key areas of data analytics:

1. Web Analytics: Analysis of website usage patterns, user behavior, and traffic optimization
2. Data Integration: Automation of data pipelines connecting cloud-based forms with relational databases
3. Retail Analytics: Sales performance analysis across multiple dimensions including location, products, and time

Limitations across the projects include:

- Use of static datasets rather than real-time data streams for some analyses
- Focus on descriptive analytics with limited predictive modeling
- Exclusion of advanced machine learning techniques or AI-powered insights
- Limited demographic and behavioral segmentation in user analysis

Innovation Component

The innovation across these projects lies in the integration of multiple modern data tools and techniques to create comprehensive analytical solutions. Key innovative aspects include:

- Development of automated data pipelines that bridge cloud-based collection tools with traditional databases
- Creation of interactive dashboards that combine multiple data dimensions into single, cohesive views
- Implementation of real-time data synchronization mechanisms using API integrations
- Design of scalable analytical frameworks that can be adapted across different industries and use cases

These projects demonstrate how affordable, accessible tools can be combined to create enterprise-level business intelligence capabilities, making advanced analytics available to organizations of all sizes.

Objectives and Goals

- Objectives and Goals
- The consolidated project objectives span three critical areas of modern data analytics and business intelligence:
- **Website Usage Analysis Objectives**:
  - Systematically analyse raw website logs and aggregated usage tables to reveal user demographics, content consumption patterns, and behavioral insights
  - Calculate and profile core metrics including page views, sessions, unique users, bounce rate, average session duration, and conversion events
  - Identify top-performing pages and underperforming content areas to guide optimization strategies
  - Build comprehensive Power BI dashboards with executive overview, trend analysis, and segmentation controls
  - Implement documented data-preparation workflows that can be replicated for future data loads
  - Provide drill-through capabilities for detailed investigation of page-level metrics and user cohorts
- **Data Pipeline Integration Objectives**:
  - Establish secure connections with Google Sheets using cloud service account credentials
  - Automate the transfer of Google Form responses into structured MySQL databases
  - Implement data cleaning and validation processes to ensure data quality and consistency
  - Convert timestamps and format data for database compatibility
  - Create duplicate detection mechanisms to maintain data integrity
  - Develop scalable automation scripts that can be scheduled for periodic execution
- **Sales Analysis Objectives:**
  - Analyze sales performance across multiple dimensions including location, products, and time periods
  - Evaluate sales quantity, revenue amounts, and average performance metrics
  - Identify high-performing store locations and areas requiring improvement
  - Understand product-level and category-level performance to guide inventory decisions
  - Monitor profitability through margin analysis over time
  - Create interactive dashboards that support dynamic exploration of sales data

Expected Outcomes and Deliverables

- Website Usage Analysis Deliverables:
  - Comprehensive Power BI dashboard (.pbix file) with multiple analytical pages:
    - Executive Overview with KPIs and top-line trends
    - Traffic Trends showing daily, weekly, and monthly patterns
    - Geographic and Source Analysis with country-level insights
    - Browser and Device Performance comparisons
    - Content Analysis highlighting top pages and exit rates
    - Drill-through diagnostic capabilities
  - Identification of peak activity periods and optimization opportunities
  - Documented methodology and recommendations for performance improvements
- Data Pipeline Integration Deliverables:
  - Fully functional Python automation script for Google Forms to MySQL integration
  - Secure API authentication setup using Google Cloud service accounts
  - MySQL database schema optimized for form response storage
  - Data validation and error handling mechanisms
  - Documentation of the automated workflow and deployment instructions
  - Demonstration of real-time data synchronization capabilities
- Sales Analysis Deliverables:
  - Interactive Power BI sales dashboard showcasing:
    - Key performance indicators and sales metrics
    - Location-based sales performance analysis
    - Product and category breakdown with profitability insights
    - Time-series analysis of sales trends and seasonality
    - Margin analysis and operational efficiency metrics
  - Strategic recommendations for business optimization
  - Identification of key revenue drivers and growth opportunities

1. Methods / Technology Used

Project 1: Website Usage Analysis

Methods / Technology Used :

The website usage analysis employed a descriptive analytics approach to transform raw web traffic data into actionable business intelligence. The methodology followed these key stages:

- Data Cleaning & Preprocessing – Power Query was used to handle missing values, remove duplicate entries, and ensure consistent data formatting across all metrics
- Data Modeling – The dataset was structured in Power BI to support efficient relational models and enable dynamic filtering across multiple dimensions
- Visualization & Analysis – Interactive dashboards were designed to present key performance indicators and usage patterns in an executive-friendly format
- Insight Extraction – Visual exploration techniques were used to identify traffic drivers, engagement periods, and browser performance issues

Tools / Software Used

- Power BI – Primary visualization platform for dashboard creation and data analysis
- Kaggle – Open-source data platform providing the raw website usage dataset
- Power Query – Data transformation and cleaning tool integrated within Power BI

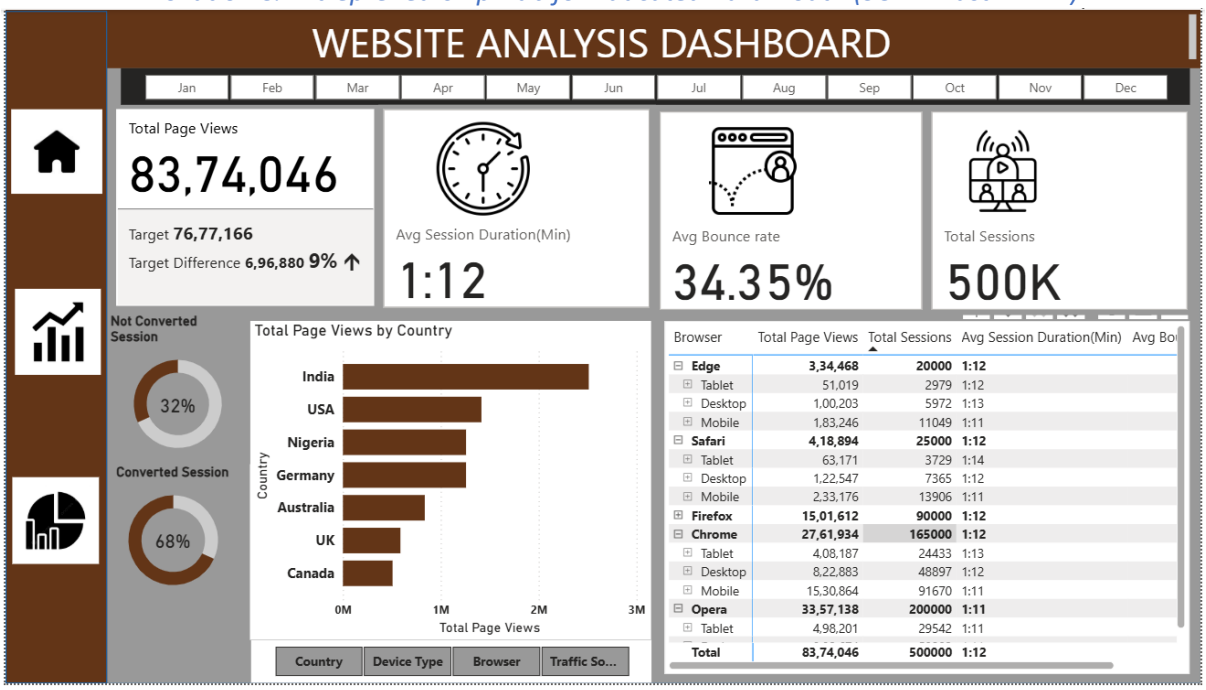
Data Collection Approach The dataset was sourced from Kaggle, containing comprehensive website metrics including page views, sessions, bounce rates, session durations, and conversion data. Raw data underwent rigorous quality assessment to identify and resolve inconsistencies, missing values, and duplicate records before analysis.

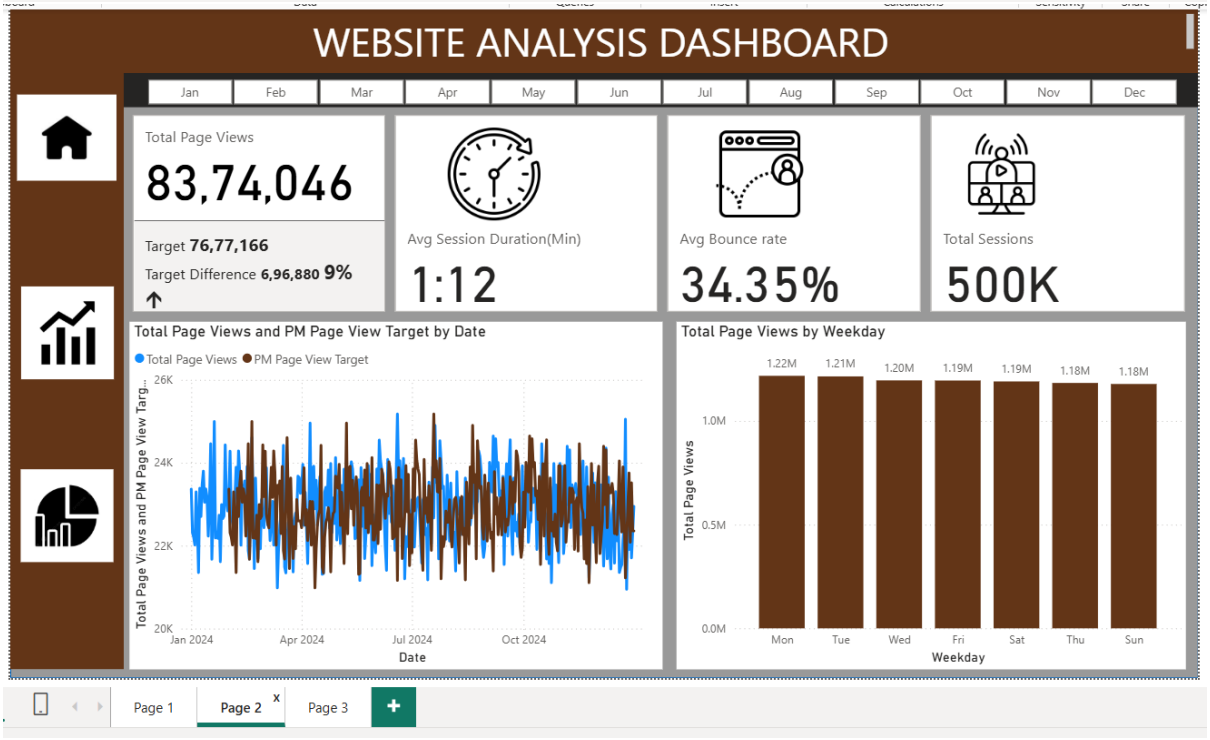
Project Architecture

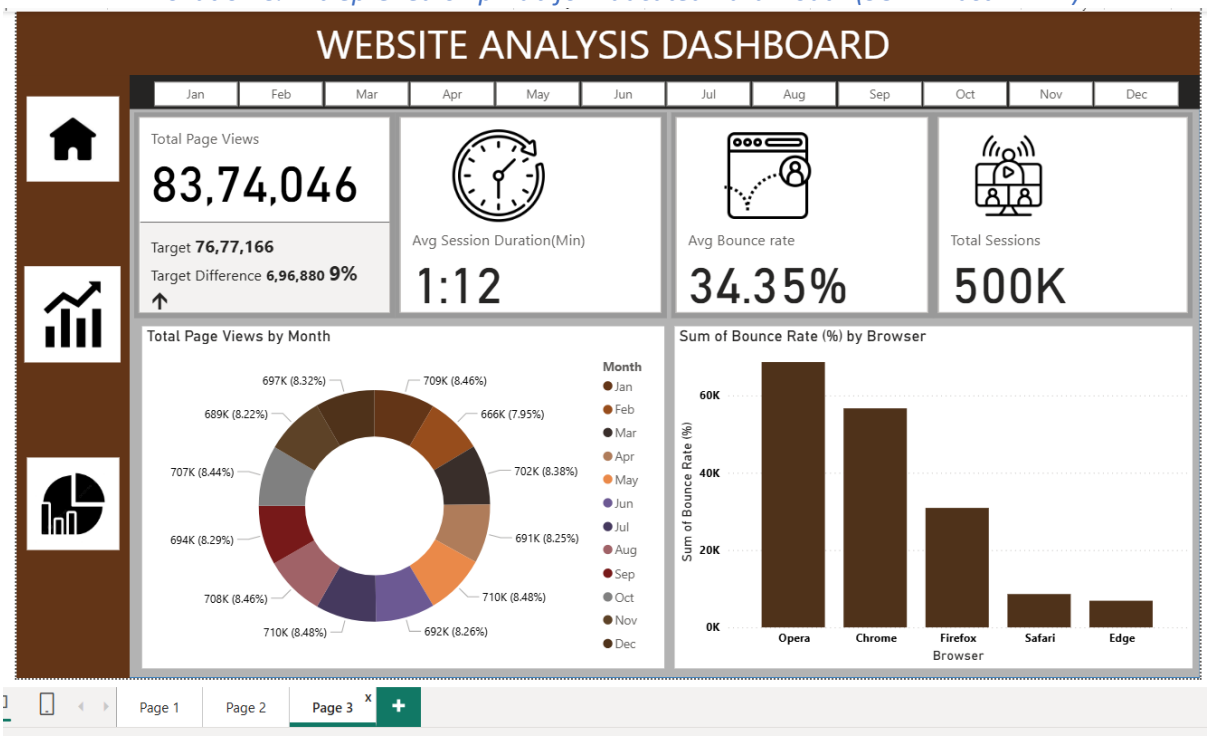
1. Data Collection – Website usage dataset acquired from Kaggle platform
2. Data Preparation – Cleaning and preprocessing using Power Query (duplicate removal, missing data treatment, normalization)
3. Data Modeling – Structuring data in Power BI with DAX formulas for KPIs such as bounce rate and session duration
4. Visualization & Insights – Multi-view dashboards designed for overview, trends, and detailed analysis

Final Results

- Total page views analyzed: 8.37M with 34.35% bounce rate and 1m 12s average session duration
- Geographic analysis revealed India and USA as primary traffic sources
- Chrome dominated browser usage, while Opera showed highest bounce rate
- Monday recorded peak engagement (1.22M page views), June showed highest monthly activity (710K views)
- Interactive dashboards successfully converted complex metrics into actionable insights


WEBSITE ANALYSIS DASHBOARD
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Total Page Views
83,74,046
Target 76,77,166
Target Difference 6,96,880 9% ↑
Avg Session Duration(Min)
1:12
Avg Bounce rate
34.35%
Total Sessions
500K
Not Converted Session
32%
Converted Session
68%
Total Page Views by Country
India
USA
Nigeria
Germany
Australia
UK
Canada
Country
0M 1M 2M 3M
Total Page Views
Country Device Type Browser Traffic So...
Browser Total Page Views Total Sessions Avg Session Duration(Min)
Edge 3,34,468 20000 1:12
Tablet 51,019 2979 1:12
Desktop 1,00,203 5972 1:13
Mobile 1,83,246 11049 1:11
Safari 4,18,894 25000 1:12
Tablet 63,171 3729 1:14
Desktop 1,22,547 7365 1:12
Mobile 2,33,176 13906 1:11
Firefox 15,01,612 90000 1:12
Chrome 27,61,934 165000 1:12
Tablet 4,08,187 24433 1:13
Desktop 8,22,883 48897 1:12
Mobile 15,30,864 91670 1:11
Opera 33,57,138 200000 1:11
Tablet 4,98,201 29542 1:11
Total 83,74,046 500000 1:12
Page 1 Page 2 Page 3


WEBSITE ANALYSIS DASHBOARD
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Total Page Views
83,74,046
Target 76,77,166
Target Difference 6,96,880 9%
↑
Avg Session Duration(Min)
1:12
Avg Bounce rate
34.35%
Total Sessions
500K
Total Page Views and PM Page View Target by Date
Total Page Views
PM Page View Target
Total Page Views and PM Page View Targ...
26K
24K
22K
20K
Jan 2024
Apr 2024
Jul 2024
Oct 2024
Date
Total Page Views by Weekday
1.22M 1.21M 1.20M 1.19M 1.19M 1.18M 1.18M
1.0M
0.5M
0.0M
Total Page Views
Mon Tue Wed Fri Sat Thu Sun
Weekday
Page 1 Page 2 Page 3



Project 2: Coffee Shop Sales Analysis

Methods / Technology Used :

The sales analysis project employed comprehensive business intelligence techniques:

- Data Preparation – Dataset import with structural verification and necessary transformations
- Data Modeling – Relationships built between fact tables (sales) and dimension tables (products, stores, dates)
- Visualization Design – Charts, KPIs, and tree maps created to present trends and patterns clearly
- Performance Analysis – Multi-dimensional exploration of sales behavior across stores, products, and time periods

Tools / Software Used

- Power BI – Primary analytical and visualization platform
- Kaggle – Source platform for coffee sales dataset

Data Collection Approach Coffee sales data was sourced from Kaggle, containing comprehensive metrics across multiple store locations, product categories, and time periods. Data quality checks addressed inconsistencies and missing values before analysis.

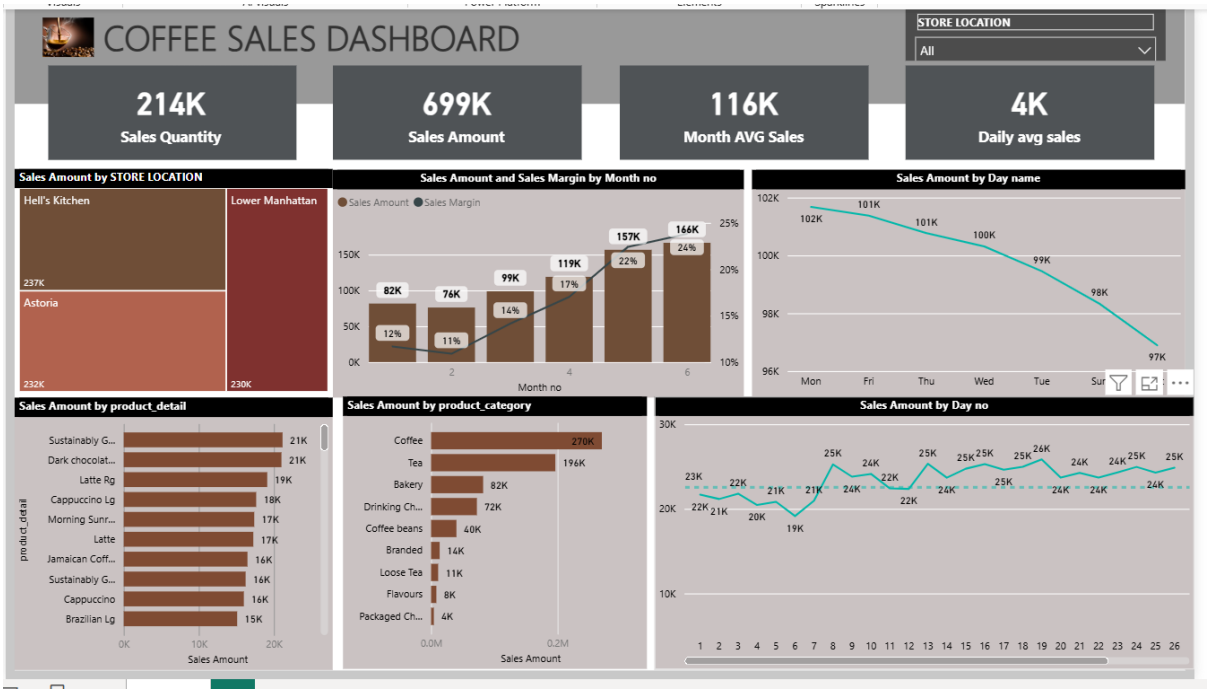
Project Architecture

1. Data Source – Sales dataset from Kaggle

2. Data Import & Preparation – Power BI transformations including field renaming and date formatting
3. Data Modeling – Fact/dimension relationships with DAX calculations for averages and margins
4. Dashboard Creation – Interactive visualizations for KPIs, store performance, and trend analysis

Results

- Business Performance: 214K sales quantity, 699K total sales amount, 116K monthly average, 4K daily average
- Location Analysis: Hell's Kitchen (237K), Astoria (232K), Lower Manhattan (230K) as top performers
- Product Insights: Coffee (270K) and Tea (196K) dominated sales, specialty items showed strong performance
- Trend Analysis: Month 6 peak at 166K, profit margins improved from 12% to 24%
- Weekly Patterns: Monday peak engagement with gradual decline toward Sunday
- Daily Volatility: Sales fluctuated 19K-26K with peaks on 7th, 11th, and 15th of months



Project 2: Automated Data Pipeline - Google Forms to MySQL Integration

Methods / Technology Used :

The data pipeline project utilized Python programming combined with cloud API integration to automate form response processing:

- Python Automation – Scripts developed for data fetching, cleaning, formatting, and database insertion
- Google Sheets API Integration – gspread library used for secure access to form responses
- Data Validation – Header standardization, missing field handling, and timestamp conversion for MySQL compatibility
- Database Operations – MySQL integration with duplicate prevention using unique identifiers

Tools / Software Used

- Python 3.13 – Core programming language for automation
- MySQL Workbench – Database creation, management, and querying
- Google Cloud Console – Service account creation and API credential management
- Key Libraries: gspread, oauth2client, mysql-connector-python

Data Collection Approach Form responses were automatically collected through Google Forms and stored in linked Google Sheets. Python scripts accessed these sheets via API credentials, validated data quality, and transferred only new responses to prevent duplication.

Project Architecture

1. User Submission – Responses submitted via Google Forms
2. Cloud Storage – Automatic storage in Google Sheets
3. API Integration – Python script connects securely to Google Sheets
4. Data Processing – Cleaning, validation, and timestamp conversion
5. Database Insert – New responses inserted into MySQL with duplicate checking

Results

- Successful automation of form response transfer with 100% accuracy
- Real-time data synchronization between Google Sheets and MySQL
- Effective duplicate prevention through unique identifier checking
- Scalable solution handling multiple responses without errors
- Timestamp conversion successfully implemented for all entries

Attendance form (Responses)

File Edit View Insert Format Data Tools Extensions Help

A1 | Timestamp

Responses

	Timestamp	email	meeting_id	rating	feedback
2	31/08/2025 19:13:05	gopikang2020@gmail.cc	6859	4	Good
3	31/08/2025 19:42:37	anjalias2004@gmail.cor	5346346	5	good
4	31/08/2025 21:13:26	Aswini as	Aswini	5	Nice
5	01/09/2025 00:03:52	nipunonepiece@gmail.c	Monkeydluffy	5	It was great class
6	01/09/2025 10:35:00	anjalisuresh5237@gmm	56578	5	good
7	01/09/2025 17:07:08	Rs45@gmail.com	Rohit sharma	5	It was a clearly understa
8	01/09/2025 17:13:06	dominic@gmail.com	ghw567g	5	nice
9	01/09/2025 20:32:35	maya123@gmail.com	ugh356i	3	Less content
10	01/09/2025 20:33:09	anu2009@gmail.com	hhsegyuw76	3	Satisfied
11	01/09/2025 20:34:38	karthik308@gamil.com	bgdsgy889	2	Not bad
12	01/09/2025 20:35:35	tim2020@gmail.com	abc123	3	Not bad
13	01/09/2025 20:36:03	manoj123@gmail.com	abc123	4	Nice
14	01/09/2025 20:36:44	aswinias2005@gmail.cc	abc123	5	Nice
15	01/09/2025 20:38:46	syam2001@gmail.com	abc123	5	Loved it
16					

```sql
CREATE DATABASE IF NOT EXISTS meeting_feedback;
USE meeting_feedback;

CREATE TABLE IF NOT EXISTS form_responses (
  id INT AUTO_INCREMENT PRIMARY KEY,
  email VARCHAR(255) NOT NULL,
  meeting_id VARCHAR(100) NOT NULL,
  rating TINYINT NULL,
  feedback TEXT NULL,
  submitted_at DATETIME NOT NULL,
  is_processed TINYINT DEFAULT 0,
  UNIQUE KEY uq_response (email, meeting_id, submitted_at)
);

select * from form_responses ;
```

Result Grid

id	email	meeting_id	rating	feedback	submitted_at	is_processed
1	gopikang2020@gmail.com	6859	4	Good	2025-08-31 19:13:05	0
2	anjalias2004@gmail.com	5346346	5	good	2025-08-31 19:42:37	0
3	Aswini as	Aswini	5	Nice	2025-08-31 21:13:26	0
7	nipunonepiece@gmail.com	Monkeydluffy	5	It was great class	2025-09-01 00:03:52	0
16	anjalisuresh5237@gmmail.com	56578	5	good	2025-09-01 10:35:00	0
22	Rs45@gmail.com	Rohit sharma	5	It was a clearly understanding session	2025-09-01 17:07:08	0
29	dominic@gmail.com	ghw567g	5	nice	2025-09-01 17:13:06	0
37	maya123@gmail.com	ugh356i	3	Less content	2025-09-01 20:32:35	0
38	anu2009@gmail.com	hhsegyuw76	3	Satisfied	2025-09-01 20:33:09	0
39	karthik308@gamil.com	bgdsgy889	2	Not bad	2025-09-01 20:34:38	0

form_responses6

Output

Action Output

#	Time	Action	Message
5	10:35:24	select * from form_responses LIMIT 0, 1000	4 row(s) returned

```sql
CREATE DATABASE IF NOT EXISTS meeting_feedback;
USE meeting_feedback;

CREATE TABLE IF NOT EXISTS form_responses (
  id INT AUTO_INCREMENT PRIMARY KEY,
  email VARCHAR(255) NOT NULL,
  meeting_id VARCHAR(100) NOT NULL,
  rating TINYINT NULL,
  feedback TEXT NULL,
  submitted_at DATETIME NOT NULL,
  is_processed TINYINT DEFAULT 0,
  UNIQUE KEY uq_response (email, meeting_id, submitted_at)
);
```

Learning and Reflection

The three projects offered extensive learning across technical, analytical, and professional areas, significantly enhancing both technical proficiency and practical problem-solving skills.

Technical Skills

- **Power BI**: Advanced data modeling, DAX for complex metrics, Power Query for cleaning, and interactive dashboards.
- **Database Management**: MySQL design, querying, cloud authentication, and duplicate-prevention mechanisms.
- **Python & Automation**: API integration, error handling, secure authentication, scheduling, and database connectivity.

Analytical Skills

- Pattern recognition across web traffic, sales, and form responses.
- Data quality management through handling missing values, duplicates, and inconsistencies.
- Translating technical findings into actionable business insights.

Professional Skills

- **Project Management**: Time allocation, task breakdown, documentation, and quality assurance.
- **Communication**: Visual storytelling with dashboards, writing executive summaries, and presenting actionable insights.
- **Industry Awareness**: Exposure to web analytics, survey automation, and retail analytics.

Overall, the projects boosted confidence in managing real-world data challenges, simulating professional work environments, and applying technical solutions to business problems.

Conclusion and Future Scope

The successful completion of these three projects illustrates strong capability in applying modern data analytics and automation tools to solve real-world problems. Each project achieved its stated goals while also opening opportunities for future enhancements.

Key Achievements

- *Website Usage Analysis*: Transformed 8.37M raw web traffic records into actionable insights through dashboards, highlighting geographic trends, browser performance, and peak engagement periods.
- *Automated Data Pipeline*: Built a secure, automated Google Forms–to–MySQL integration that eliminated manual entry, ensured data integrity, and enabled real-time synchronization.
- *Coffee Shop Sales Analysis*: Analyzed 699K in retail sales data, identified top-performing outlets, highlighted coffee and tea as major revenue drivers, and tracked profit margin growth from 12% to 24%.

Future Scope

- **Real-Time Dashboards**: Linking live data streams for continuous monitoring of website traffic, survey responses, and sales performance.
- **Predictive Analytics**: Applying forecasting models to predict future web traffic, sales demand, and customer trends.
- **Customer Segmentation & Behavior Analysis**: Incorporating demographic and loyalty data to generate more personalized business insights.
- **Scalability**: Extending pipeline automation to handle larger datasets and integrating with enterprise cloud platforms.
- **Cross-Domain Insights**: Combining analytics from multiple domains (web, retail, and survey data) to build a holistic view of business performance.

Conclusion

Together, these projects provided practical exposure to **Power BI, SQL, Python, and API integration**, while strengthening project management and communication skills. The integrated experience demonstrates how data-driven solutions can directly support **digital growth, operational efficiency, and strategic decision-making** across diverse business contexts.