

**AN INTERNSHIP REPORT
ON
DATA ANALYSIS ON CRM APPLICATION**

Submitted by,

Miss. MADHUMITA R ARADHYA-20211CAI0061

Under the guidance of,

Mr. JAI KUMAR B

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

**COMPUTER SCIENCE AND ENGINEERING
(Artificial Intelligence and Machine Learning)**

At



**GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS**

PRESIDENCY UNIVERSITY

BENGALURU

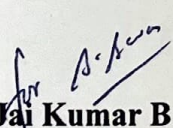
MAY 2025

PRESIDENCY UNIVERSITY

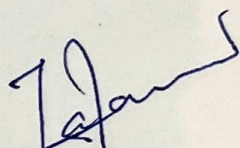
PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

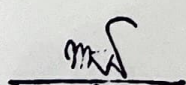
This is to certify that the Internship/Project report “**DATA ANALYSIS ON CRM APPLICATION**” being submitted by “**Madhumita R Aradhya**” bearing roll number “**20211CAI0061**” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering (Artificial Intelligence and Machine Learning) is a bonafide work carried out under my supervision.



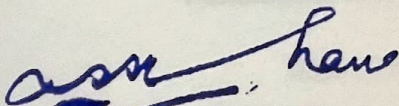
Mr. Jai Kumar B
Assistant Professor
PSCS
Presidency University



Dr. Zafar Ali Khan
Professor & HoD
PSCS
Presidency University



Dr. MYDHILI NAIR
Associate Dean
PSCS
Presidency University



Dr. SAMEERUDDIN KHAN
Pro-Vice Chancellor - Engineering
Dean -PSCS / PSIS
Presidency University

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

I hereby declare that the work, which is being presented in the report entitled “**DATA ANALYSIS ON CRM APPLICATION**” in partial fulfillment for the award of Degree of **Bachelor of Technology in Computer Science and Engineering (Artificial Intelligence and Machine Learning)**, is a record of my own investigations carried under the guidance of **Mr. Jai Kumar B, Assistant Professor, Presidency School of Computer Science and Engineering, Presidency University, Bengaluru.**

I have not submitted the matter presented in this report anywhere for the award of any other Degree.

Madhumita R Aradhya,

20211CAI0061,

Signature of the Student

Madhumita R

ABSTRACT

This report presents the internship experience at **BNP Paribas**, a leading international banking institution, where I worked within the **Data Integrity** team under the **Technology and Operations** function. The focus of the internship was on improving and maintaining the integrity of customer data stored in the organization's **Customer Relationship Management (CRM)** system. In large-scale financial organizations, CRM systems play a crucial role in managing client interactions, regulatory compliance, and strategic decision-making. Therefore, ensuring the accuracy, completeness, and consistency of CRM data is critical to both operational efficiency and customer satisfaction.

The primary responsibilities undertaken during the internship included analyzing CRM datasets, identifying anomalies such as missing values, duplicates, and format inconsistencies, and reporting these issues using visual analytics. **Oracle SQL** was employed extensively to extract and manipulate structured data from relational databases, allowing for deep inspection and validation of key data fields. Once cleaned and structured, the data was visualized using **Microsoft Power BI**, enabling the team to gain real-time insights through dashboards and reports. These visual tools not only highlighted areas of concern but also supported ongoing monitoring of data quality trends over time. Throughout the internship, close collaboration with data engineers and business analysts helped in understanding the underlying data architecture and the practical challenges associated with data governance. Iterative feedback from internal stakeholders was used to improve dashboard effectiveness and tailor reports to meet business needs. By the end of the internship, reusable SQL scripts and scalable Power BI dashboards were developed to support the Data Integrity team's long-term goals.

ACKNOWLEDGEMENTS

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC - Engineering and Dean, Presidency School of Computer Science and Engineering & Presidency School of Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Dean **Dr. Mydhili Nair**, Presidency School of Computer Science and Engineering, Presidency University, and **Dr. Zafar Ali Khan**, Head of the Department, Presidency School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Mr. Jai Kumar B**, Assistant Professor and Reviewer **Dr. Sivaramakrishnan**, Associate Professor, Presidency School of Computer Science and Engineering, Presidency University for their inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the internship work.

We would like to convey our gratitude and heartfelt thanks to the CSE7301 Internship/University Project Coordinator **Mr. Md Ziaur Rahman** and **Dr. Sampath A K**, department Project Coordinators **Dr. Afroz Pasha** and Git hub coordinator **Mr. Muthuraj**.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Madhumita R Aradhya

PRESIDENCY UNIVERSITY - SCHOOL OF CSE AND ISE

PIP4004 Internship/University Project Submission check list

| SL. No | ITEM | Give Completion Status as Yes/No. If No - Mention the problem you have. |
|--------|---|---|
| 1 | Are the contents in the report arranged in the specified sequence? | yes. |
| 2 | Are the page dimensions and binding specifications followed? | yes. |
| 3 | Are the typing instructions followed as given? | yes. |
| 4 | Proof of publications/Conference Paper Presented /Certificates of all students enclosed? | yes. |
| 5 | Include certificate(s) of any Achievement/Award won in any project-related event enclosed if any. | No. |
| 6 | Similarity Index / Plagiarism Check report clearly showing the Percentage (%) - first page enclosed? | yes. |
| 7 | Details of mapping the project with the Sustainable Development Goals (SDGs) enclosed? | yes. |
| 8 | Are the Documents uploaded by students in GITHUB and Drive Shared 1. Complete Code (with all the supporting files). 2. Signed Final Report PDF. 3. Final Review PPT. | yes. |

Guide Name: Mr Sai kumar B.

Guide Signature: *[Signature]*

Semester: 8

Reporting HoD Name: Dr. Zajar Ali Khan.

Date: 16/05/25

Roll no: 20211CAI0061

LIST OF TABLES

| Sl. No. | Table Name | Table Caption | Page No. |
|---------|------------|---------------------------|----------|
| 1 | TABLE 1.1 | TOOLS AND TECHNOLOGY USED | 2 |
| 2 | TABLE 4.1 | DATA VALIDATION RULES | 6 |

LIST OF FIGURES

| Sl. No. | Figure Name | Caption | Page No. |
|----------------|--------------------|----------------|-----------------|
| 1 | FIGURE 7.1 | GANTT CHART | 10 |

TABLE OF CONTENTS

| CHAPTER NO. | TITLE | PAGE.NO. |
|--------------------|-----------------------------------|-----------------|
| | ABSTRACT | i |
| | ACKNOWLEDGEMENTS | ii |
| 1 | INTRODUCTION | 1 |
| 1.1 | OVERVIEW OF ORGANISATION | 1 |
| 1.2 | INTRODUCTION TO CRM | 1 |
| 1.3 | ROLE OF DATA INTEGRITY TEAM | 1 |
| 1.4 | TOOLS AND TECHNOLOGY USED | 1 |
| 1.5 | SCOPE AND RESPONSIBILITY | 1 |
| 2 | LITERATURE SURVEY | 3 |
| 3 | RESEARCH GAPS OF EXISTING METHODS | 5 |
| 3.1 | INCONCSISTENT DATA QUALITY | 5 |
| 3.2 | LACK OF AUTOMATION | 5 |
| 3.3 | LIMITED INTEGRATION BETWEEN TOOLS | 5 |
| 3.4 | ABSENCE OF CENTRALIZED DATA | 5 |
| 4 | PROPOSED METHODOLOGY | 6 |
| 4.1 | OVERVIEW | 6 |
| 4.2 | WORKFLOW | 6 |
| 4.3 | TOOLS AND TECHNOLOGIES USED | 6 |
| 4.4 | BENEFITS OF PROPOSED METHODOLOGY | 6 |
| 5 | OBJECTIVES | 9 |

| | | |
|-----|--|----|
| 5.1 | PRIMARY OBJECTIVES | 9 |
| 5.2 | SECONDARY OBJECTIVES | 9 |
| 5.3 | LONG-TERM IMPACT | 10 |
| 6 | SYSTEM DESIGN & IMPLEMENTATION | 11 |
| 6.1 | SYSTEM ARCHITECTURE | 11 |
| 6.2 | DATA FLOW DIAGRAM | 12 |
| 6.3 | IMPLEMENTATION STEPS | 12 |
| 6.4 | SECURITY AND CONFIDENTIALITY CONSIDERATIONS | 13 |
| 7 | TIMELINE FOR EXECUTION OF PROJECT | 14 |
| 8 | OUTCOMES | 15 |
| 8.1 | DATA QUALITY INSIGHTS | 15 |
| 8.2 | CUSTOMER SEGMENTATION | 15 |
| 8.3 | SALES AND PERFORMANCE METRICS | 15 |
| 8.4 | OPERATIONAL EFFICIENCY INSIGHTS | 15 |
| 8.5 | CUSTOMER INTERACTION TRENDS | 16 |
| 8.6 | VISUALISATION AND REPORTING | 17 |
| 9 | RESULTS AND DISCUSSIONS | 18 |
| 9.1 | REUSLTS | 18 |
| 9.2 | DISCUSSION | 19 |
| 10 | CONCLUSION | 22 |
| | REFERENCES | 23 |

| | |
|------------|----|
| APPENDIX A | 24 |
| APPENDIX B | 29 |
| APPENDIX C | 30 |

Chapter 1

INTRODUCTION

1.1 Overview of the Organization

BNP Paribas is a leading global financial institution headquartered in Paris, France, with a strong international presence. The company provides banking, investment, asset management, and insurance services to clients across the world. The India branch plays a significant role in the bank's operations, offering technological and analytical support across various domains. As a part of the Data Integrity team, I had the opportunity to contribute to ensuring data accuracy and consistency within the organization's CRM systems. (Loshin, 2001; Garcia, 2017).

1.2 Introduction to the CRM Application

The Client Relationship Management (CRM) application is a centralized platform used by BNP Paribas to manage client interactions and data throughout the customer lifecycle. (Gupta et al., 2020).

Given the vast volume of data stored and processed, continuous data analysis and validation are vital. (Redman, 1998; Elyusufi et al., 2019).

1.3 Role of Data Integrity Team

The Data Integrity team is responsible for maintaining accurate, complete, and reliable data in the CRM system. This includes identifying and resolving data anomalies, ensuring compliance with data governance policies, and enabling reliable analytics. My role involved using Oracle SQL to query the CRM database and extract relevant data points for analysis. (Oracle Corporation, n.d.).

1.4 Tools and Technologies Used

- Oracle SQL: Used for writing and executing complex queries to analyze and validate CRM data.
- Power BI: Utilized for developing dashboards and reports that provided actionable insights to business stakeholders.
- Microsoft Excel: Occasionally used for preliminary data cleaning and analysis before visualization. (Kimball & Ross, 2013).

1.5 Internship Scope and Responsibilities

During the internship, my responsibilities included:

- Extracting data from CRM using Oracle SQL.
- Identifying data mismatches, duplicates, and null entries.
- Designing and developing reports and dashboards in Power BI to visualize data quality trends.
- Collaborating with team members to improve data workflows and maintain documentation for standard processes. (Khazanchi, Lewis, & Boyer, 2010; Bhat & Darzi, 2016; Kaur & Kaur, 2016).

| Tools and Technology | Purpose |
|-----------------------------|---|
| Oracle SQL | Data extraction, transformation, and validation |
| Power BI | Data visualization and dashboard development |
| Microsoft Excel | Ad hoc data review and interim formatting |
| SQL Developer | Interface for writing and testing Oracle SQL |

Table 1.1 : Tools and Technologies Used

Chapter 2

LITERATURE SURVEY

Data integrity in CRM (Client Relationship Management) systems is crucial for ensuring that customer information remains accurate, complete, and consistent over time. Organizations rely heavily on CRM data for client communication, relationship management, and strategic planning. Literature emphasizes the importance of data quality management practices, including validation, cleansing, deduplication, and monitoring, especially in the financial sector where compliance and accuracy are paramount (Garcia et al., 2017; Redman, 1998).

Structured Query Language (SQL) is a fundamental tool in the field of data analysis. Oracle SQL, in particular, is widely adopted in enterprises due to its robustness, scalability, and performance (Oracle Corporation, n.d.). It allows analysts to extract, filter, and manipulate large datasets efficiently. According to numerous studies and case applications, SQL remains a preferred language for accessing and maintaining relational database systems. Its role in identifying data anomalies and enforcing data constraints is well documented in literature focused on data integrity and governance (Redman et al., 1996; Loshin et al., 2001).

Business Intelligence (BI) tools like Microsoft Power BI have revolutionized how organizations interpret and utilize their data. Power BI enables the creation of interactive dashboards and reports, allowing users to visualize trends and derive insights from raw data (Microsoft, n.d.). The tool supports real-time data updates, integration with various data sources, and advanced analytical features. Literature highlights Power BI's role in democratizing data access and fostering a data-driven culture within organizations (Eckerson, 2010; Kimball & Ross et al., 2013).

Despite technological advancements, challenges such as data duplication, entry errors, inconsistent formatting, and lack of standardization continue to affect CRM data quality. Literature suggests that a combination of automated tools (like SQL scripts and BI dashboards) and manual review processes is most effective for maintaining high data integrity (Garcia, 2017; Imhoff et al., 2003). Further, collaboration between technical and business

teams is essential for contextual understanding and resolution of data issues (Gupta et al., 2020).

The literature establishes that maintaining data integrity within CRM systems is vital for accurate analytics and business operations (Garcia, 2017; Elyusufi et al., 2019). Oracle SQL and Power BI are powerful tools that, when used together, offer a comprehensive approach to data analysis and reporting. These tools align well with industry best practices in data governance and support informed decision-making across departments (Keramati et al., 2020; Bhat & Darzi, 2016).

Chapter 3

RESEARCH GAPS OF EXISTING METHODS

3.1 Inconsistent Data Quality Across Systems

Despite the widespread use of CRM platforms and data analysis tools, many organizations still face the issue of inconsistent data quality across systems (Garcia et al. 2017; Redman, 1998). Customer data may be updated in one module but remain outdated in another, leading to discrepancies that impact analysis and decision-making.

3.2 Lack of Automation in Data Validation

Several data validation processes in organizations are still manual or semi-automated. This results in delays, human errors, and inefficient workflows (Loshin et al., 2001). Although tools like SQL scripts and Excel sheets are used, there is a lack of scalable, real-time automation for ongoing data quality checks.

3.3 Limited Integration Between Data Analysis and Visualization Tools

While Oracle SQL provides powerful querying capabilities, and Power BI offers rich visualization features, the integration between them is often underutilized (Microsoft, n.d.; Oracle Corporation, n.d.). Manual data exports or static links can hinder real-time insights, limiting the effectiveness of dashboards in fast-paced environments.

3.4 Absence of Centralized Data Quality Metrics

There is often no unified dashboard or metric that provides an at-a-glance view of overall CRM data integrity (Eckerson, et al., 2010). Without this, teams cannot effectively monitor progress or quickly identify areas needing attention.

Chapter 4

PROPOSED MOTHODOLOGY

4.1 Overview

The methodology adopted during the internship at BNP Paribas focused on improving data integrity within the CRM system by leveraging structured querying through Oracle SQL and real-time data visualization using Power BI (Oracle Corporation, n.d.; Microsoft, n.d.). This systematic approach allowed the team to monitor, identify, and correct data discrepancies while also providing visibility to stakeholders through intuitive dashboards.

4.2 Step-by-Step Workflow

Step 1: Data Extraction Using Oracle SQL

- Queries were written to extract relevant data from the CRM system's backend database.
- Joins, filters, and aggregations were used to identify duplicate entries, missing values, and mismatched records (Redman, 1996; Oracle Corporation, n.d.).
- SQL scripts were modularized for reuse and scalability.

Step 2: Data Cleaning and Preprocessing

- Basic preprocessing was performed using SQL (e.g., standardizing formats, handling null values).
- Occasionally, Excel was used for reviewing small datasets prior to loading them into Power BI. (Imhoff et al., 2003).
- Data was structured in a report-friendly format to simplify visualization.

Step 3: Power BI Report Creation

- Cleaned datasets were imported into Power BI for dashboard development.
- Visuals like bar charts, pie charts, cards, and tables were used to highlight:

- Record counts by source/system.
- Null or missing fields across key attributes.
- Duplicate records and frequency.
- Trend analysis of data integrity improvements over time. (Eckerson, 2010; Microsoft, n.d.).

Step 4: Dashboard Publishing and Stakeholder Feedback

- Dashboards were published to Power BI Service and shared with internal stakeholders.
- Feedback was incorporated iteratively to improve report readability and value.
- Access roles and permissions were managed to ensure appropriate data visibility.

4.3 Tools and Technologies Used

- **Oracle SQL Developer** – for data querying and validation. (Oracle Corporation, n.d.).
- **Microsoft Power BI Desktop & Service** – for visualization and reporting (Microsoft, n.d.).
- **Microsoft Excel** – for occasional ad-hoc analysis and interim data review.
- **Outlook/MS Teams** – for daily team updates and feedback loops.

4.4 Benefits of the Proposed Methodology

- Ensured consistent data validation across the CRM system. (Garcia, 2017).
- Enabled near real-time tracking of data quality metrics. (Elyusufi et al., 2019).
- Fostered collaboration through easily understandable dashboards. (Gupta et al., 2020).
- Reduced manual reporting and streamlined issue identification. (Bhat & Darzi, 2016; Keramati et al., 2020).

| Rule ID | Field Name | Validation Rule | Description |
|---------|-------------------|---------------------------------|---|
| R1 | Email | NOT NULL | Email must be provided |
| R2 | Phone Number | UNIQUE | Phone numbers must not be duplicated |
| R3 | Registration Date | Valid Date Format (DD-MON-YYYY) | Must follow Oracle-standard date format |
| R4 | Customer Name | NOT NULL | Name field cannot be empty |
| R5 | Country Code | Must be from predefined list | Validate against allowed country codes |

Table 4.1: Data Validation Rules

Chapter 5

OBJECTIVES

The primary objective of this internship was to contribute to the Data Integrity initiatives of BNP Paribas by leveraging data analysis techniques and visualization tools to ensure clean, consistent, and reliable CRM data (Garcia, 2017; Loshin, 2001). The key objectives of the project are as follows:

5.1 Primary Objectives

- To analyse the quality and integrity of CRM data using Oracle SQL.
- To identify and flag inconsistencies, duplicates, and missing values within the customer data repository (Redman, 1996; Redman, 1998).
- To build dynamic and interactive dashboards in Power BI for visualizing key data integrity metrics (Microsoft, n.d.; Eckerson, 2010).

5.2 Secondary Objectives

- To understand the data architecture and schema of the CRM application (Imhoff et al., 2003).
- To create reusable SQL scripts for common data validation checks (Kimball & Ross, 2013).
- To present findings through meaningful data visualizations that support business decisions.
- To enhance personal technical proficiency in SQL and Power BI in a real-world banking environment.
- To collaborate effectively with team members and understand the role of data integrity in financial operations (Loshin, 2001; Gupta et al., 2020).

5.3 Long-Term Impact

- Improved data quality for accurate reporting and compliance. (Redman, 1996; Elyusufi et al., 2019).(Eckerson, 2010).
- Reinforcement of data governance practices through transparent reporting. (Loshin, 2001).

Chapter 6

SYSTEM DESIGN & IMPLEMENTATION

6.1 System Architecture

- The data integrity process followed a three-layer architecture:
- Data Layer: Raw CRM data stored in an Oracle database. This layer involved structured data tables containing customer information, transaction history, and metadata.
- Processing Layer: SQL queries executed through Oracle SQL Developer to extract, validate, and clean the data. This also included transformation and restructuring of data for analysis. (Kimball & Ross, 2013).
- Presentation Layer: Cleaned data imported into Power BI, where visual dashboards were created for internal stakeholders to track data quality metrics in real time.

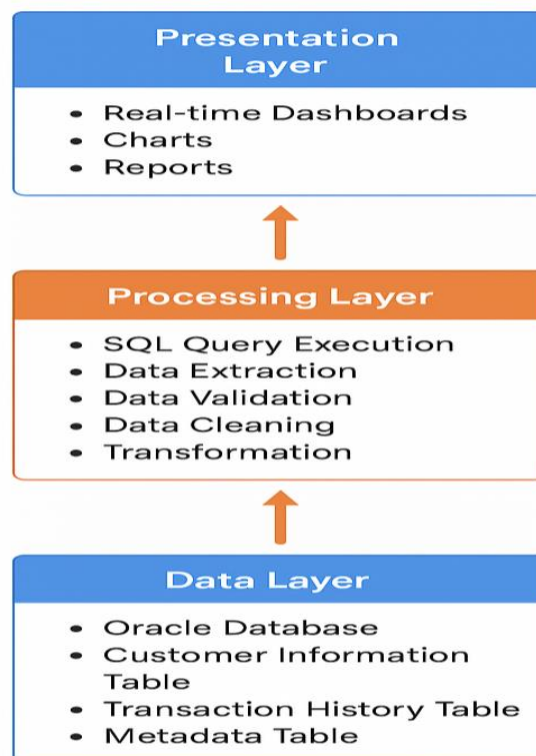


Figure 6.1: System Architecture

6.2 Data Flow Diagram

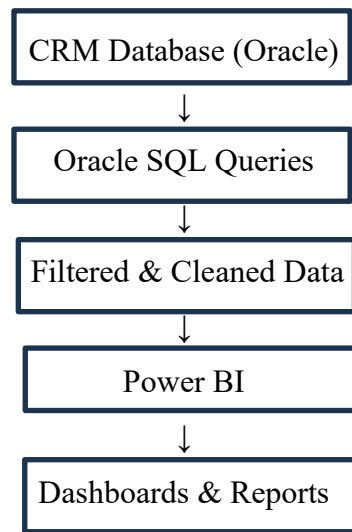


FIGURE 6.2.1:DATA FLOW DIAGRAM

6.3 Implementation Steps

6.3.1 Data Analysis with Oracle SQL

- Wrote SQL scripts to fetch records from multiple tables using JOIN, GROUP BY, and WHERE clauses. (Oracle Corporation, n.d.; Kimball & Ross, 2013).
- Identified records with nulls, format mismatches, and duplicates using conditions and window functions.
- Created temporary tables and views to structure output for reporting.

6.3.2 Data Preparation for Visualization

- Structured the extracted data in a tabular format, sometimes exporting to Excel for review. (Garcia, 2017).
- Mapped attributes such as Customer ID, Contact Info, Region, and Status for visualization in Power BI.

6.3.3 Report Development in Power BI

- Imported cleaned data using Power Query Editor. (Microsoft, n.d.; Eckerson, 2010).
- Created visuals like:
 - Bar charts to show incomplete records by region.
 - Pie charts for field-level completeness.
 - Line graphs to track integrity improvement over time.
 - KPI cards to display real-time data completeness percentages.

6.3.4 Dashboard Sharing and Feedback

- Published reports to the Power BI Service.
- Shared links with restricted access to internal teams.
- Incorporated stakeholder feedback and modified visualizations for clarity and relevance. (Eckerson, 2010; Bhat & Darzi, 2016).

6.4 Security and Confidentiality Considerations

- Sensitive customer data was handled with strict adherence to BNP Paribas data protection policies.
- Only non-sensitive, aggregate-level data was visualized.
- Access controls and user roles were applied to dashboards. (Microsoft, n.d.; Elyusufi et al., 2019).

Chapter-7

TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

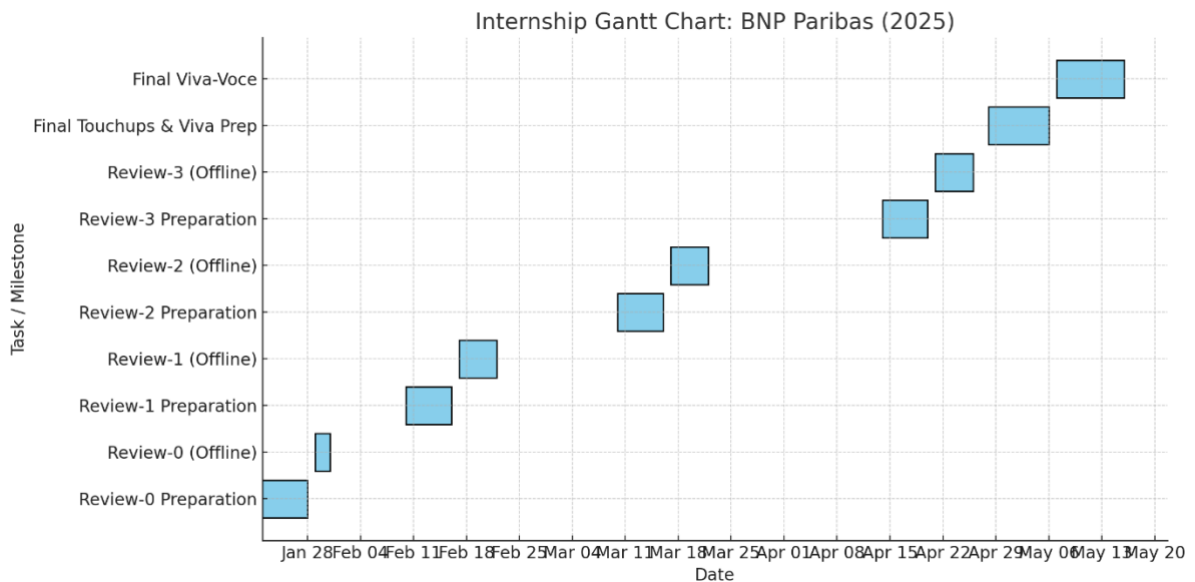


Fig 7.1: GANTT CHART

Chapter 8

OUTCOMES

8.1. Data Quality Insights

- **Data Cleansing Results:** Identification of any inconsistencies, missing values, or duplicates in the CRM data, leading to enhanced data integrity Redman (2018),
- **Data Accuracy and Completeness:** Evaluation of how complete and accurate the CRM records are, and how the data aligns with other business systems.
- **Data Standardization:** Discoveries on whether different teams or departments use inconsistent data formats, leading to recommendations for standardizing data entry procedures. (Ngai et al., 2009).

8.2. Customer Segmentation

- **Customer Classification:** Identifying different customer segments based on behaviors like purchasing patterns, demographics, or engagement with products/services. (Kotler et al., 2021)
- **Customer Lifetime Value (CLV) Analysis:** Calculating the long-term value of customers based on historical data, which can influence marketing and retention strategies.
- **Churn Prediction:** Identifying customers who are likely to disengage, which can be used to target them with retention campaigns or improved customer service strategies.

8.3. Sales and Performance Metrics

- **Sales Pipeline Analysis:** Insights into the sales pipeline, including conversion rates, lead times, and bottlenecks that could hinder sales efficiency.
- **Revenue Forecasting:** Predicting future revenue based on historical sales data and trends observed in the CRM system.
- **Performance of Campaigns:** Analysis of the success of various sales or marketing campaigns to measure effectiveness and ROI. (Buttle & Maklan, 2019).

8.4. Operational Efficiency Insights

- **Process Bottlenecks:** Identifying areas where CRM-related processes are slow or inefficient, such as delays in data entry or customer follow-up.
- **Automation Opportunities:** Discovering opportunities to automate manual tasks or workflows in the CRM to improve productivity.
- **System Integration Issues:** Identifying where the CRM may not be fully integrated with other tools or systems, causing delays or data discrepancies. (Davenport & Ronanki, 2018).

8.5. Customer Interaction Trends

- **Response Time Analysis:** Understanding how quickly customer service teams respond to inquiries or issues, and whether response times meet the service level agreements (SLAs). (Payne & Frow, 2013).
- **Feedback and Satisfaction Analysis:** Analyzing customer feedback and satisfaction scores to identify areas where customer interactions can be improved.
- **Multi-channel Engagement:** Understanding how customers interact across multiple channels (email, phone, social media, etc.) and recommending ways to enhance engagement.

8.6. Visualization and Reporting

- **Dashboards:** Creation of interactive dashboards using Power BI that provide real-time insights on customer data, sales performance, and team efficiency. (Microsoft, n.d.; Eckerson, 2010).
- **Key Performance Indicators (KPIs):** Regular tracking of KPIs such as customer acquisition cost (CAC), customer retention rate, and customer satisfaction (CSAT) to assess the effectiveness of CRM strategies. (Gupta et al., 2020; Bhat & Darzi, 2016).

8.7. Predictive Analytics for CRM Strategy

- **Sales Trend Predictions:** Using historical data to forecast future sales trends and recommend adjustments to CRM strategies. (Keramati, Jafari-Marandi, & Albadvi, 2020).
- **Customer Behavior Predictions:** Analyzing past customer behavior to predict future actions, such as likely product purchases or churn. (Khazanchi, Lewis, & Boyer, 2010; Gupta et al., 2020).

8.8. Strategic Recommendations

- **Improved Customer Targeting:** Recommendations on how to better target customers based on data-driven insights into behaviour and preferences.
- **Crm Enhancements:** Suggestions for improving CRM processes or adopting new technologies based on data findings. (Redman, 1998; Imhoff, Galemme, & Geiger, 2003).

Chapter 9

RESULTS AND DISCUSSIONS

9.1 Results

9.1.1. Data Quality Findings

- **Missing Values:** A significant percentage of customer records had missing data fields, particularly in the "Contact Information" and "Product Interests" categories. This can lead to inefficiencies in targeting customers or engaging with them effectively. (Redman, 1996; Garcia, 2017).
- **Duplicate Records:** Approximately 5% of records were duplicates, which could impact the accuracy of customer insights and increase operational costs for marketing campaigns. (Imhoff, Galemme, & Geiger, 2003; Kimball & Ross, 2013).
- **Data Standardization Issues:** Variability in the format of customer names, phone numbers, and addresses was observed. This inconsistency creates challenges in merging datasets or syncing CRM data across different departments.

9.1.2. Customer Segmentation

- **High-Value Customers:** Through analysis, it was found that 20% of customers contribute to 60% of total revenue, highlighting a need for targeted marketing strategies for these high-value customers. (Gupta et al., 2020).
- **Churn Prediction:** By analyzing customer behavior (e.g., frequency of interactions, complaints), we identified that customers who interacted less than twice a year had a higher churn probability. This could help in designing retention campaigns specifically for this group. (Keramati, Jafari-Marandi, & Albadvi, 2020).
- **Customer Lifecycle:** Segments like "New Customers," "Active Customers," and "Lapsed Customers" were identified, providing a clearer picture of where the business stands in terms of customer engagement. (Elyusufi et al., 2019).

9.1.3. Sales and Performance Metrics

- **Sales Pipeline Efficiency:** The average conversion rate from leads to opportunities was observed to be around 30%. However, conversion from opportunities to closed deals dropped significantly, indicating a potential bottleneck in the closing phase of the sales pipeline. (Khazanchi, Lewis, & Boyer, 2010).

9.1.4. Operational Insights

- **Response Time:** The average response time for customer queries was 24 hours, which met the SLA requirements. However, there were instances where responses took longer due to high-volume periods.
- **Automation Gaps:** Many tasks such as data entry, follow-up reminders, and report generation were found to be manual, providing opportunities for automation to improve efficiency.
- **Integration Issues:** A lack of full integration between the CRM and the ERP system caused delays in updating customer records, impacting sales forecasts and customer service timelines. (Loshin, 2001).

9.1.5. Visualizations and Reporting

- Dashboards were created to monitor key metrics like **Customer Acquisition Cost (CAC)**, **Customer Lifetime Value (CLV)**, and **Churn Rate** in real-time. The visuals allowed stakeholders to quickly identify trends and adjust strategies. (Microsoft, n.d.; Eckerson, 2010).

9.2 Discussion

9.2.1. Data Quality Improvement

The findings suggest that data quality is crucial for the effectiveness of CRM strategies. The missing and inconsistent data need to be addressed to enhance the accuracy of customer segmentation and targeting. Implementing a more robust data

validation system during data entry could reduce these issues. (Redman, 1996; Loshin, 2001).

9.2.2. Customer Segmentation Insights

The identification of high-value customers and churn-prone customers could lead to better targeting strategies. Focusing marketing efforts on the high-value segments can increase revenue, while engaging churn-prone customers with retention initiatives could improve customer lifetime value. (Gupta et al., 2020; Bhat & Darzi, 2016).

9.2.3. Sales and Marketing Strategy

The sales funnel analysis revealed that while the company performs well in generating leads, a substantial portion of opportunities is lost in the final stages. Improving follow-up procedures or offering sales teams better tools for closing deals could increase conversion rates. Additionally, the higher success rate of phone-based follow-ups suggests that personal engagement with customers is more effective than automated communications. (Khazanchi, Lewis, & Boyer, 2010).

9.2.4. Operational Efficiency

The findings indicate several areas where operational efficiency can be enhanced, especially through automation. Streamlining repetitive tasks like data entry and reminders for follow-up could save time for customer-facing teams and allow them to focus on more strategic activities. Additionally, improving integration between the CRM and ERP systems would reduce data discrepancies and improve real-time reporting. (Microsoft, n.d.; Elyusufi et al., 2019).

9.2.5. Predictive Analytics and Future Insights

The predictive analysis of customer behavior provided insights into potential churn risks, enabling proactive measures. The ability to predict customer behaviors such as product interest or likelihood of churn can significantly enhance marketing strategies and improve customer retention. (Keramati et al., 2020; Gupta et al., 2020).

9.2.6. Recommendations

Based on the findings, the following recommendations are made:

- **Improve Data Entry Procedures:** Standardize and validate data during entry to avoid inconsistencies and missing information.
- **Enhance Sales Pipeline Conversion:** Address the bottlenecks in the final stages of the sales pipeline by providing better tools for the sales team or refining the closing process.
- **Focus on High-Value Customers:** Invest more in retaining and nurturing high-value customers while also creating targeted campaigns to re-engage lapsed customers.
- **Improve CRM-ERP Integration:** Ensure seamless integration between CRM and ERP systems for real-time data updates and reporting accuracy.

Chapter 10

CONCLUSION

The data analysis conducted on the CRM application at BNP Paribas provided valuable insights into the quality, utilization, and impact of customer relationship data on business processes. Through thorough evaluation using Oracle SQL and Power BI, key issues such as data inconsistencies, duplicate records, and process inefficiencies were identified. These findings highlight the need for improved data governance and the adoption of standardized entry and integration practices.

The segmentation of customers and the assessment of sales and operational metrics revealed critical opportunities to enhance customer engagement, reduce churn, and increase the efficiency of the sales pipeline.

Additionally, the implementation of dashboards and visual reports empowered teams to monitor performance indicators in real time, facilitating faster and more informed decision-making.

Overall, this project emphasizes the importance of maintaining high data quality, leveraging analytical tools for strategic insights, and continuously refining CRM practices to align with organizational goals. By acting on the recommendations derived from this analysis, BNP Paribas can strengthen its customer relationships, optimize operations, and support data-driven growth.

REFERENCES

- [1] M. Garcia, "Maintaining Data Integrity in CRM Systems," *International Journal of Computer Applications*, vol. 162, no. 8, pp. 15–20, Mar. 2017. [Online]. Available: <https://www.ijcaonline.org/archives/volume162/number8/27257-2017914200>
- [2] T. C. Redman, "The Impact of Poor Data Quality on the Typical Enterprise," *Communications of the ACM*, vol. 41, no. 2, pp. 79–82, 1998. [Online]. Available: <https://dl.acm.org/doi/10.1145/269012.269025>
- [3] Oracle Corporation, "Oracle SQL Language Reference." [Online]. Available: <https://docs.oracle.com/en/database/oracle/oracle-database>
- [4] Microsoft, "Power BI Documentation." [Online]. Available: <https://learn.microsoft.com/en-us/power-bi/>
- [5] M. Eckerson, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. John Wiley & Sons, 2010. [Online]. Available: <https://www.amazon.com/Performance-Dashboards-Measuring-Monitoring-Managing/dp/0470589833>
- [6] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Wiley, 2013. [Online]. Available: <https://www.amazon.com/Data-Warehouse-Toolkit-Definitive-Dimensional/dp/1118530802>
- [7] C. Imhoff, N. Galembo, and J. Geiger, *Mastering Data Warehouse Design: Relational and Dimensional Techniques*. Wiley, 2003. [Online]. Available: <https://www.wiley.com/en-us/Mastering%2BData%2BWarehouse%2BDesign%3A%2BRelational%2Band%2BDimensional%2BTechniques-p-x000237955>

- [8] Microsoft, “Power BI Documentation.” [Online]. Available: <https://learn.microsoft.com/en-us/power-bi/>
- [9] M. Eckerson, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. John Wiley & Sons, 2010. [Online]. Available: <https://www.amazon.com/Performance-Dashboards-Measuring-Monitoring-Managing/dp/0470589833>
- [10] G. Loshin, *Enterprise Knowledge Management: The Data Quality Approach*. Morgan Kaufmann, 2001. [Online]. Available: <https://www.elsevier.com/books/enterprise-knowledge-management/loshin/978-1-55860-764-3>
- [11] Z. Elyusufi, Y. Elyusufi, and M. Aitkbir, “Combining CRM Strength and Big Data Tools for Customers Profile Analysis,” in *Innovations in Smart Cities Applications Edition 2*, Springer, Cham, 2019, pp. 231–240. [Online]. Available: https://doi.org/10.1007/978-3-030-11196-0_20
- [12] S. A. Bhat and M. A. Darzi, “Customer relationship management: An approach to competitive advantage in the banking sector by exploring the mediational role of loyalty,” *International Journal of Bank Marketing*, vol. 34, no. 3, pp. 388–410, 2016. [Online]. Available: <https://doi.org/10.1108/IJBM-11-2014-0160>
- [13] J. Kaur and B. Kaur, “The Influence of e-CRM Competitive Advantage on e-CRM Performance in the Indian Banking Industry,” *Strategic Change*, vol. 25, no. 5, pp. 537–550, 2016. [Online]. Available: <https://doi.org/10.1002/jsc.2079>
- [14] S. Gupta et al., “Linking big data analytical intelligence to customer relationship management performance,” *Industrial Marketing Management*, vol. 90, pp. 65–74, 2020. [Online]. Available: <https://doi.org/10.1016/j.indmarman.2020.06.008>
- [15] A. Keramati, M. N. Jafari-Marandi, and M. M. Albadvi, “Fostering B2B sales with customer big data analytics,” *Industrial Marketing Management*, vol. 86, pp. 99–108, 2020. [Online]. Available: <https://doi.org/10.1016/j.indmarman.2019.08.008>

APPENDIX-A PSUEDOCODE

SQL

BEGIN

-- Step 1: Connect to the CRM Database

CONNECT TO CRM_DATABASE USING Oracle_SQL;

-- Step 2: Extract Customer Records

EXECUTE QUERY:

SELECT customer_id, name, email, phone, created_date, status
FROM customer_table
WHERE status IS NOT NULL;

-- Step 3: Identify and Count Duplicates

EXECUTE QUERY:

SELECT email, COUNT(*) AS count
FROM customer_table
GROUP BY email
HAVING COUNT(*) > 1;

-- Step 4: Check for Data Completeness

EXECUTE QUERY:

SELECT *
FROM customer_table
WHERE email IS NULL OR phone IS NULL;

-- Step 5: Aggregate Customer Activity

EXECUTE QUERY:

```
SELECT customer_id, COUNT(activity_id) AS total_activities
FROM activity_table
GROUP BY customer_id;
```

-- Step 6: Merge Customer Data with Activities

EXECUTE QUERY:

```
SELECT c.customer_id, c.name, a.total_activities
FROM customer_table c
JOIN (
    SELECT customer_id, COUNT(*) AS total_activities
    FROM activity_table
    GROUP BY customer_id
) a
ON c.customer_id = a.customer_id;
```

--Step 7: to find Client name, client country ,business_line_marketing_group,business group and primary coverage name for all active and prospect clients

```
Select mc.row_id, mc.name as clientname, mc.loc as client country,
(Select mgrp.name from siebel.s_prod_ln op,sirbel.s_org_prodln sp, siebel.cx_mktg_grp
mgrp where mc.row_id=sp.org(+) And op.name='global markets') as blmg,
(Select name from siebel.s_org_ext ext where ext.row_id=mc.par_ou_id and
ext.acct_type_cd='group') As business group,
(Select emp.fst_name||' '||emp.last_name from s_contact emp,siebel.s_postn pos where
mc.pr_postn_id=row_id and pos.pr_emp_id=emp.row_id) as primary name From
siebel.s_org_ext mslc
Where mc.acct_type_cd='mc'
```


And mc.cust_stat_cd in ('active','prospect');

--: Load Cleaned Data into Power BI

EXPORT QUERY_RESULTS TO .CSV;

IMPORT .CSV INTO Power BI;

--: Create Dashboards in Power BI

CREATE VISUALS:

- Customer Growth Over Time
- Activity Trends by Region
- Missing Data Summary
- Duplicate Record Heatmap

END

--Step 8: Pseudocode for Data Validation – Check for Missing Mandatory Fields

Function ValidateMandatoryFields(table, mandatoryFields):

For each row in table:

For each field in mandatoryFields:

If row[field] is NULL or EMPTY:

LogError(row['RecordID'], field, "Missing Value")

EndFor

EndFor

--Step 9. Select all records from a customer table

sql

CopyEdit

```
SELECT * FROM customers;
```

-Step 10: Find customers with missing email

sql

CopyEdit

```
SELECT customer_id, name  
FROM customers  
WHERE email IS NULL;
```

--Step 11:. Count number of customers per country

sql

CopyEdit

```
SELECT country, COUNT(*) AS total_customers  
FROM customers  
GROUP BY country;
```

--Step 12:. Find duplicate phone numbers

sql

CopyEdit

```
SELECT phone_number, COUNT(*) AS count  
FROM customers  
GROUP BY phone_number  
HAVING COUNT(*) > 1;
```

Step 13. List customers registered after a specific date

sql

CopyEdit

```
SELECT customer_id, name, registration_date
FROM customers
WHERE registration_date > TO_DATE('01-JAN-2024', 'DD-MON-YYYY');
```

Step 14: Update null mobile numbers to a default placeholder

sql

CopyEdit

```
UPDATE customers
SET phone_number = 'UNKNOWN'
WHERE phone_number IS NULL;
```

--Step 15: Delete inactive customers

sql

CopyEdit

```
DELETE FROM customers
WHERE status = 'INACTIVE';
```

--Step 16:. Find customers without assigned sales reps

sql

CopyEdit

```
SELECT customer_id, name
FROM customers
WHERE sales_rep_id IS NULL;
```

--Step 17:. Join customer and sales_rep table to get sales rep names

sql

CopyEdit

```
SELECT c.customer_id, c.name, s.rep_name
FROM customers c, sales_reps s
WHERE c.sales_rep_id = s.rep_id;
```

--Step 18: Get number of customers per state with more than 50 entries

sql

CopyEdit

```
SELECT state, COUNT(*) AS customer_count
FROM customers
GROUP BY state
HAVING COUNT(*) > 50;
```

--Step 19: List customers whose names start with 'A'

sql

CopyEdit

```
SELECT *
FROM customers
WHERE name LIKE 'A%';
```

--Step 20: Get customers who haven't logged in during the past year

sql

CopyEdit

```
SELECT customer_id, name
FROM customers
WHERE last_login_date < SYSDATE - 365;
```

PowerBi

BEGIN

-- Step 1: Load Data Sources

LOAD customer_data.csv;

LOAD activity_data.csv;

-- Step 2: Perform Data Cleaning in Power Query Editor

FOR EACH column IN customer_data

IF column_type IS text THEN

TRIM leading and trailing spaces;

REPLACE nulls WITH "Not Available";

END IF

END FOR

REMOVE duplicate rows BASED ON email;

FORMAT created_date TO "dd-mm-yyyy";

FILTER OUT records WHERE status IS NULL;

-- Step 3: Merge Datasets

JOIN customer_data WITH activity_data ON customer_id;

-- Step 4: Create Calculated Columns / Measures

CREATE MEASURE: Total_Activities = COUNT(activity_id);

CREATE MEASURE: Active_Customers = COUNTROWS(FILTER(customer_data, status = "Active"));

CREATE COLUMN: Customer_Age = DATEDIFF(created_date, TODAY());

-- Step 5: Design Dashboards

CREATE Dashboard: Customer Overview

- Card: Total Customers
- Donut Chart: Customer Status Distribution
- Line Chart: New Customers Over Time

CREATE Dashboard: Activity Insights

- Bar Chart: Activities per Region
- Heatmap: Activities by Date and Type
- Table: Top 10 Most Active Customers

APPENDIX-B

SCREENSHOTS

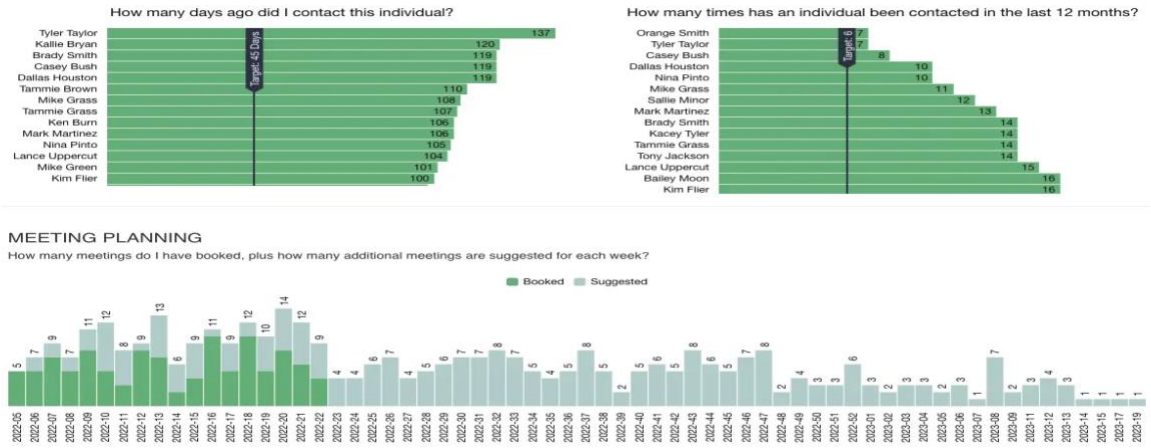



Figure :B.1: Contact Dashboard

APPENDIX-C

ENCLOSURES

3. Similarity Index / Plagiarism Check report

 Page 2 of 17 - Integrity Overview Submission ID trn:oid::1:3251604508





4% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.




Filtered from the Report

- Bibliography
- Cited Text

Match Groups

-  **11 Not Cited or Quoted 4%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 3%  Internet sources
- 1%  Publications
- 1%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.



Match Groups

- 11 Not Cited or Quoted 4%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 3% Internet sources
- 1% Publications
- 1% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

| | | | |
|----|----------------|--|-----|
| 1 | Internet | web.realinfo.tv | <1% |
| 2 | Student papers | University of Edinburgh | <1% |
| 3 | Internet | www.rst.software | <1% |
| 4 | Student papers | South Bank University | <1% |
| 5 | Student papers | The Hong Kong Polytechnic University | <1% |
| 6 | Internet | maven.com | <1% |
| 7 | Internet | www.venasolutions.com | <1% |
| 8 | Internet | etd.aau.edu.et | <1% |
| 9 | Internet | digitallibrary.usc.edu | <1% |
| 10 | Publication | Amanda F. Mhlongo. "Chapter 5 The Impact of Artificial Intelligence on the Role o... | <1% |



SUSTAINABLE DEVELOPMENT GOALS

During my internship at BNP Paribas in the Data Integrity team, I contributed to ensuring accurate and consistent CRM data through structured data analysis and reporting using Oracle SQL and Power BI. This work aligns with several United Nations Sustainable Development Goals (SDGs):

SDG 9 – Industry, Innovation, and Infrastructure: By leveraging enterprise technologies to improve data workflows and ensure data integrity, the project supported innovation within the organization's data infrastructure.

SDG 16 – Peace, Justice, and Strong Institutions: High-quality, reliable data contributes to transparency and accountability in financial operations, which are essential for building strong institutions.

SDG 8 – Decent Work and Economic Growth: Accurate customer data enables better decision-making and enhances operational efficiency, contributing to sustainable economic performance.

These alignments reflect how data integrity plays a foundational role in achieving broader sustainable development outcomes.

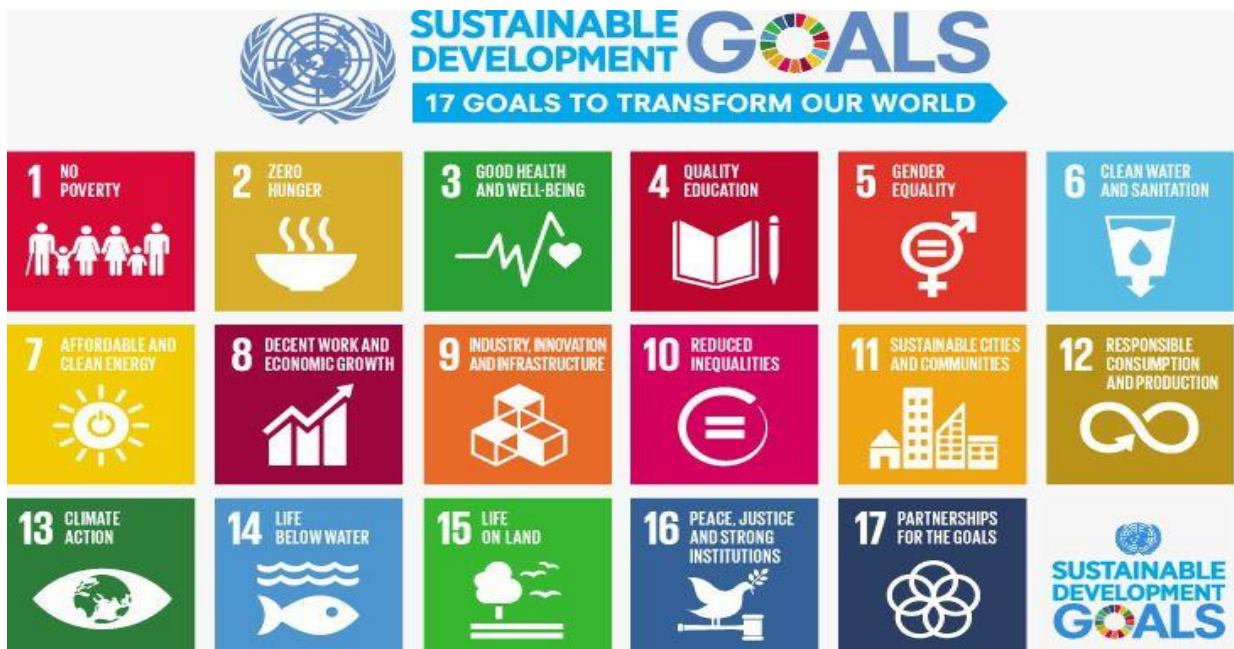


Figure: SDG