



Development of a Web-Based Analytics Platform for Competitive Analysis of Direct Mail Marketing in US Financial Institutions

A project work submitted to the University Tun Abdul Razak in partial fulfilment of
the requirements for the award of

Bachelor of Information Technology
(Data Science)

by

Nur Farahanis Binti Khadir
B23761093

Guided by

Zainul Akramin Bin Mohd Drus

Bank Rakyat School of Business, Innovation, Technology &
Entrepreneurship (BRSBITE)

UNIVERSITI TUN ABDUL RAZAK
KUALA LUMPUR
MALAYSIA

Submission Month & Year
04 June 2025

UNIVERSITI TUN ABDUL RAZAK
KUALA LUMPUR
MALAYSIA

This is to certify that this project work entitled

DEVELOPMENT OF A WEB-BASED ANALYTICS PLATFORM FOR
COMPETITIVE ANALYSIS OF DIRECT MAIL MARKETING IN US
FINANCIAL INSTITUTIONS

is a bonafide record of the project work done

by

Nur Farahanis Binti Khadir
B23761093

at UNIRAZAK, Kuala Lumpur, Malaysia during the year 2025 in partial fulfilment of
the requirements for the award of the degree of

Bachelor of Information Technology (Data Science)

Zainul Akramin Bin Mohd Drus
Senior Lecturer, Programme Director
Presented and submitted to the university on **04 June 2025**.

DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

Date: **04 June 2025**

Nur Farahanis Binti Khadir
B23761093

ACKNOWLEDGMENTS

First, I would like to give all thanks to Allah Almighty, for his blessings, for granting me patience, strength and good health during this final year project. I would not have encountered or completed this journey without the guidance and mercy of Allah.

I feel especially fortunate to have Mr. Zainul Akramin Bin Mohd Drus as my supervisor. The commitment he has shown in providing me with continued consistent help, advice, feedback and reassurance has been immeasurable, during this project. During our meetings, he has challenged my thinking on several occasions, especially in relation to targeted Microsoft programming paths, giving me a reference direction

I would also like to give a heartfelt thank you to all my other lecturers, course mates and friends for offering their support through advice, comments, and encouragement. I thank you for your motivating comments and help regardless of the task or a task that was only small to them, it made this journey a lot easier and meant a lot to me.

To my dear family, I have continually appreciated your endless love, prayers, and belief in me. You have all inspired and motivated me. I dedicate this to each of you for your support and sacrifices. Finally, I would like to thank UNIRAZAK for enabling me to study at this outstanding university and gain new knowledge and experiences.

ABSTRACT

This project involves creating a web-based analytics platform for US financial institutions to perform comparative analysis of direct mail marketing effectiveness. US environmental institutions such as Bank of America, Citibank and American Express to name a few, depend heavily upon direct mail to acquire and retain customers. Currently there is no way for financial institutions to perform comparative analysis on direct mail marketing spend data across competitors at the institution level. My proposed solution is to build an accessible platform for all US financial institutions, that is centrally organized, based upon an easily defined scope, that will allow the user to specify a date range, surface monthly marketing trend data for specified time periods, to allow for report downloads. The data will be collected and stored in MySQL Workbench, organized by bank, year, month, estimated marketing spend. Data programming and processing will be done using pandas, NumPy and Matplotlib Python libraries. The web-based platform will be developed as a Streamlit application which will focus on interactive visualizations, secured access for users and downloadable CSV reports. With this platform, users will be able to find important insights that help them spot the biggest spenders, shifts in mail volume, and year-over-year performance. In addition, main goals for the system will be to be capable of providing updates in real-time, and to be designed for the most optimal user-experience for marketing analysts and decision-makers. By providing actionable insights and an overview of the competitive landscape in terms of direct mail strategies, this project will provide for greater transparency and efficiency in association with direct mail strategies and facilitate better financial marketing decision making. The product is a functional web-based dashboard fulfilling a real-world need in the US financial industry for marketing intelligence. This project illustrates how data science, database management, and web development can be combined to solve a relevant business issue, all while maintaining an academic and practical significance in the information technology or analytics space.

TABLE OF CONTENTS

	Page
DECLARATION	iii
ACKNOWLEDGMENTS	iv
ABSTRACT	v
CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
LIST OF SYMBOLS	xii
 CHAPTER I INTRODUCTION	
1.1 Research Background	121.2
	Problem Statement
	121.3
	Objectives of Research and Scope of Works
	3
 CHAPTER II LITERATURE REVIEW	
2.1 Distributed Generation	3
2.1.1 Distributed Generation	Error! Bookmark not defined.
2.1.2 Effect of Distributed Generation	Error! Bookmark not defined.
2.2 Protection Issues for Distribution Networks	
2.2.1 Short Circuit Currents	Error! Bookmark not defined.
2.2.2 Power Flow	Error! Bookmark not defined.
2.2.3 Overcurrent Protection	Error! Bookmark not defined.
	Distribution Systems
	Error! Bookmark not defined.
	2.3.1
	Review of Distribution Networks
	Error! Bookmark not defined.
	2.4
	Review of Distributed Generation
	Error! Bookmark not defined.
	2.4.1 Distribution
	System Protection
	Error! Bookmark not defined.
2.4.2 Review of Protection Methods	Error! Bookmark not defined.
2.5 Chapter Summary	
	3
 CHAPTER III DISTRIBUTION NETWORK	

3.1	Introduction	Error! Bookmark not defined.	3.2
		Radial Basis Function Neural Network	
		Error! Bookmark not defined.	3.3
		Distribution Network	
	Error! Bookmark not defined.	3.3.1	Network
	Error! Bookmark not defined.	3.3.2	
	Classification	Error! Bookmark not defined.	3.3.3
	Location	Error! Bookmark not defined.	3.3.4
	Determination	Error! Bookmark not defined.	3.3.5
	Restoration	Error! Bookmark not defined.	3.3.6
	Generation	Error! Bookmark not defined.	3.4
		Chapter Summary	
			3

CHAPTER IV PROTECTION STRATEGY

4.1	Introduction		384.2
		Protection	
	Error! Bookmark not defined.	4.2.1	
		Main and Backup	
	Error! Bookmark not defined.	4.2.2	
		Device	
	Error! Bookmark not defined.	4.3	
		Algorithm	
	Error! Bookmark not defined.	4.4	
		Proposed Strategy	
	Error! Bookmark not defined.	4.4.1	
		Main Algorithm	
	Error! Bookmark not defined.	4.4.2	
		Backup Algorithm	
	Error! Bookmark not defined.	4.5	
		Chapter Summary	
			3

CHAPTER V RESULTS AND DISCUSSION

5.1	Results of	Error! Bookmark not defined.	5.1.1
		14 Bus Test System	
	Error! Bookmark not defined.	5.1.2	
		22 Bus Test System	
	Error! Bookmark not defined.	5.1.3	
		32 Bus Test System	
	Error! Bookmark not defined.	5.1.4	
		Results for Location	
	Error! Bookmark not defined.	5.2	
		Results of Strategy	
	Error! Bookmark not defined.	5.2.1	
		Results of Strategy for the 14 bus test system	
	Error! Bookmark not defined.	5.2.2	

Results of Strategy for the 22 bus test system	
Error! Bookmark not defined.	5.2.3
Results of Strategy for the 32 bus test system	
Error! Bookmark not defined.	5.3
Chapter Summary	

3

CHAPTER VI	CONCLUSION AND FUTURE WORKS	
6.1	Conclusion	Error! Bookmark not defined.
	Significant Contributions of the Research	6.2
	Error! Bookmark not defined.	6.3
	Suggestions for Future Work	
	Error! Bookmark not defined.	

REFERENCES	4
APPENDIXES	

LIST OF TABLES

Table Number		Page
2.1	Summary of technologies	3
3.1	Fault Type Data	4
4.1	Settings of OC relays	4
4.2	The expected relay in various lines	5
5.1	14-bus test system	6

LIST OF FIGURES

Figure Number		Page
1.1	Electric power system	1
2.1	Short-circuit current	2
2.2	Network equivalent circuit of Figure 2.1	3
2.3	Thevenin equivalent circuit	4

LIST OF ABBREVIATIONS

FI: Financial Institutions

DM: Direct Mail

US: United States

W: Watt

kW: kilowatt

MW: Mega Watt

AC: Alternating current

DC: Direct Current

km: Kilometer

kV: Kilo Volt

MVA: Mega Volte Ampere

MSE: Mean Square Error

OC: Overcurrent Relay

LIST OF SYMBOLS

l : Feeder Length

d : Distance

d_{tot} : Total Feeder Length

\bar{Z} : Impedance

\bar{Z}_L : Total Line-Impedance

\bar{Z}_{DG} : The DG Impedance

\bar{Z}_S : The Source Impedance

\bar{U}_S : Voltages of the Main Source

\bar{U}_{DG} : Voltages of DG Unit

\bar{I} : Current

\bar{I}_{SC} : Short Circuit Current

$\bar{I}_{SC,S}$: The Grid Contribution of the Short Circuit Current

CHAPTER I

INTRODUCTION

1.1 BACKGROUND OF STUDY

Over the last couple of years data-driven decision-making has developed into a standard practice for strategic decisions-making across sectors, especially in financial services. Financial institutions rely on past performance data, which helps them make better decisions regarding consumer behavior, competitively understand the organization better, and help better future multimedia marketing decisions that drive ROI (Davenport & Harris, 2007). Digital channels have caused tremendous change in marketing practices, however direct mail is still a strong force in customer acquisition, especially because the banking industry is heavily regulated (Forrester, 2021). Banks are still closely monitoring their direct mail to understand ROI performance with large portions of their overall marketing budgets still being spent in direct mail. The potential for personalization with direct mail is considerable which helps mean bank institutions such as Bank of America and Citibank continue to review their envelopes, thematically and offer quality before deployment. Even when factoring in variations by placement, research has shown the average direct mail open rate is 80–90 percent, while the average email open rate is 20–30 percent (Media Logic, 2022). Furthermore, direct mail shows an average ROI of 112% which is more than any other marketing channel (DMA, 2020).

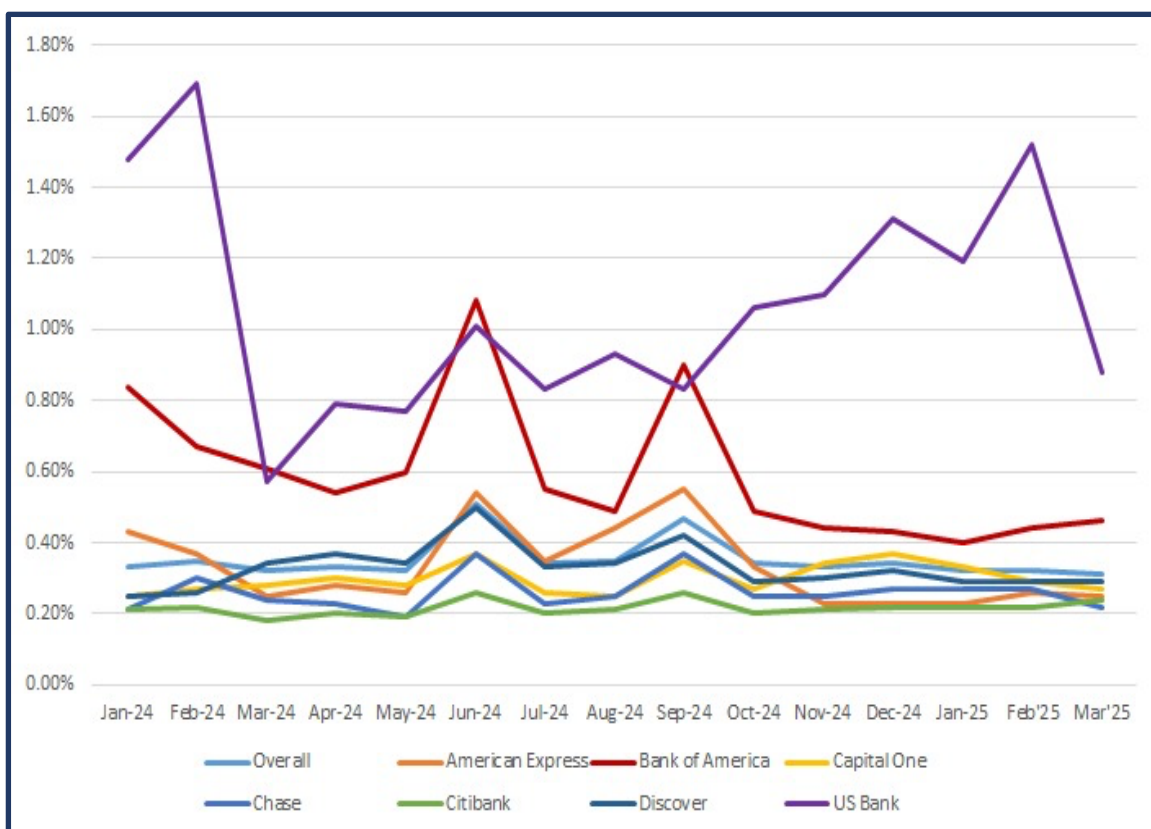
1.2 PROBLEM STATEMENT

In today's financial world, organizations like Bank of America, Citibank, and American Express continue to utilize direct mail as an important channel for customer

acquisition and retention. Regardless of the growth in digital media, direct mail has not dropped in priority since it is tactile and has a higher user-engagement rate. For example, direct mail open rates can be 80%–90%, compared to a high average financial email open rate of 31% (BAI, 2024; eMarketer, 2024). In addition, the ROI for a direct mail campaign can be 112% - again higher than other mediums (BAI, 2024). Nevertheless, a significant challenge exists like banks and credit unions typically do not have a real-time, comparative understanding of their direct mail marketing performance compared to competitors. Without benchmarking data, it will hinder decision making - especially with budget allocations, program optimizations, and marketing positioning. In fact, internal tracking initiatives may exist. However, collectively as an industry there is no standardized and centralized method to aggregate and analyze direct mail campaign data across institutions (The Financial Brand, 2020). This presents obstacles in terms of inefficiencies as well as missed opportunities to leverage trends in the market in the space of financial industries.

Figure 1.2: DM Response Rates (%) by FI, January 2024 - March 2025

Source: [Mintel/Credit Card, Behavior](#)

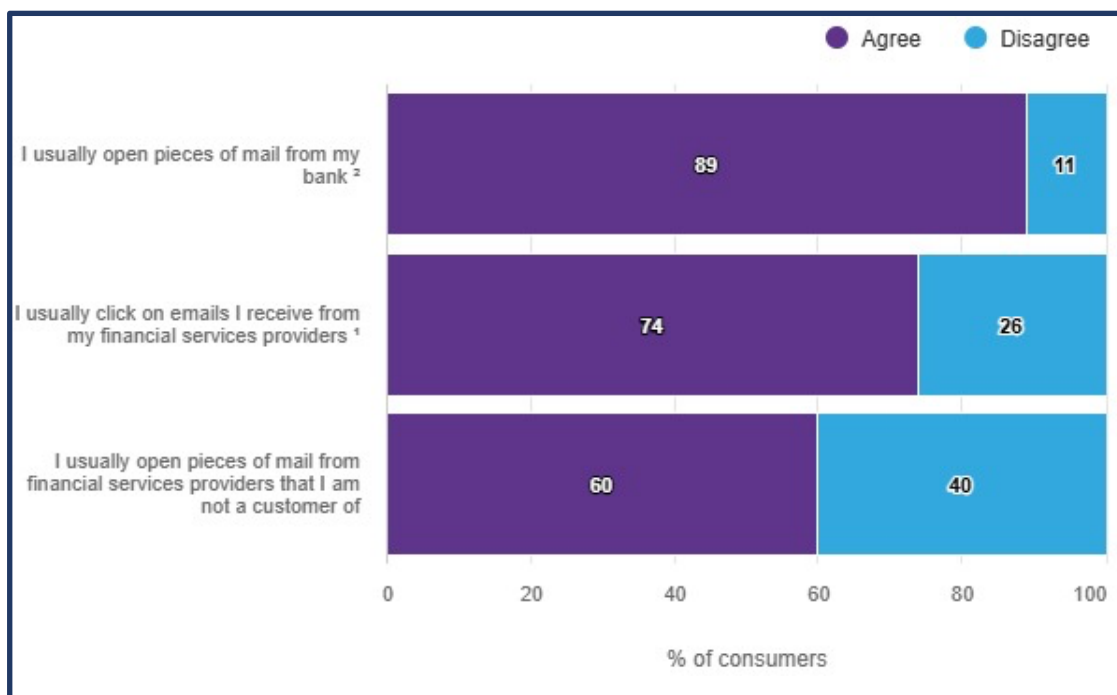


1.3 RESEARCH AIM

Financial institutions need a web-based analytics platform development for conducting direct mail marketing campaign analysis. The current data-abundant situation causes organizations to face difficulties extracting valuable insights because their data remains scattered and they have limited insight into their competitors marketing approaches. The study fulfills the requirement of building an effortless platform enabling financial organizations to add defined time periods then inspect direct mail marketing trends among significant banking institutions and discover meaningful findings. The proposed platform helps marketing analysts find essential market patterns as well as improve their campaign strategies and enhances data-based decision processes.

Figure 1.3: US Consumer Responses to DM vs. Email Communications from Financial Providers, 2024

Source: [Kantar Profiles/Mintel, January 2024](#)



1.4 RESEARCH OBJECTIVE

The research creates a web-based analytics solution focused on improving marketing decisions through strategic planning in financial institutions. This platform achieves the following:

- 1) The system organizes direct mail marketing information derived from major US banks into a structure that enables simple data queries and analysis.
- 2) The system which uses Python coupled with MySQL examines marketing expense trends and mail volume data to establish associations between marketing approaches.
- 3) Streamlit technology powers an interactive web-based dashboard within the platform that enables users to input time periods for comparing marketing trend evolutions. The system allows users to gain deeper understanding of marketing performance metrics together with dynamic monitoring abilities for better performance tracking.
- 4) The system allows users to obtain downloadable reports that enable extended analysis and strategic planning.
- 5) Real-time graphics rendering within the platform helps users visualize data trends better through updated visualizations for instant quick decision-making based on the data. Real-time data presentation through this system improves strategic planning because the data easily transforms into understandable formats.
- 6) The platform enables user data and report exportation through the download function facilitating both offline assessment and business process combination integration.

1.5 RESEARCH QUESTION

A data-driven platform which analyzes direct mail marketing trends is the subject of this research project while the listed questions serve as its investigative demands:

- 1) How can financial institutions leverage a comparative analysis of direct mail marketing trends to enhance their marketing strategies and gain a competitive edge?
- 2) What are the key patterns on monthly spending and mail volume among major US banks, and how do these trends inform marketing effectiveness?
- 3) In what possible ways enable a web-based dashboard to optimize the accessibility, visualization, and interpretation of direct mail data for marketing analysts in financial institutions?
- 4) What type of tools and technologies are most suitable for developing an intuitive, data-driven platform that facilitates seamless analysis and decision-making in direct mail marketing?

1.6 PROJECT MILESTONES

Table 1.6: Project Milestones and Process Division

Milestone No.	Date Range	Milestone Description
M1	01 May (Thurs) – 03 May (Sat)	Requirement Study Analysis <ul style="list-style-type: none"> Identify user need, define project scope and identify real problem to solve. Functional Requirements <ul style="list-style-type: none"> Outline platform features likes data input, trend visualization and CSV report. Technical and Non-Functional Spec

		<ul style="list-style-type: none"> • Define tools like MYSQL, Python, Streamlit, performance security and usability. Project Kickoff & Planning <ul style="list-style-type: none"> • Finalize proposal • Outline report structure • Create Gantt chart & milestones • Review possible data sources
M2	07 May (Wed) – 09 May (Fri)	Database Design <ul style="list-style-type: none"> • Design Entity-Relationship Diagram (ERD) • Build MYSQL schema • Build dashboard layout and data flow structure • Populate sample data for testing
M3	10 May (Sat) – 14 May (Wed)	Data Collection & Preprocessing <ul style="list-style-type: none"> • Conduct data collection and build database • Import raw datasets and build backend processing (Python) • Clean and structure data using Pandas • Explore structure
M4	15 May (Thurs) – 17 May (Sat)	Trend Analysis Development <ul style="list-style-type: none"> • Analyze direct mail spending & volume trends • Compute statistics • Display data visualizations
M5	21 May (Wed) – 23 May (Fri)	Streamlit Dashboard <ul style="list-style-type: none"> • Set up layout • Implement date filtering • Create initial trend visualizations • Add interactive features • Enable CSV report downloads

M6	24 May (Sat) – 28 May (Wed)	Testing & Debugging <ul style="list-style-type: none"> • Test date filtering, visuals, dashboard features and reports download • Validate outputs & downloads • Fix bugs
M7	29 May (Thurs) – 03 June (Tue)	Report Writing – Chapter V & VI <ul style="list-style-type: none"> • Document analysis findings • Add visuals & interpretations • Draft conclusion Appendix & Formatting <ul style="list-style-type: none"> • Add screenshots, code, ERD, structure • Organize references Supervisor Review & Amendments <ul style="list-style-type: none"> • Present to supervisor • Apply feedback
M8	4 June (Wed)	Submission & Presentation <ul style="list-style-type: none"> • Submit final report • Deliver project presentation

1.7 CONCLUSION

In conclusion, the web-based analytics platform brings solutions to financial institutions which need real-time comparative marketing performance analysis of direct mail marketing trends. The platform aggregates data from US major banks to let marketing analysts conduct effortless analyses of expenditure patterns and mail quantities and campaign performance while utilizing Python and MySQL with Streamlit for producing real-time graphical data displays and user-friendly data search functions. The platform enhances direct mail marketing outcomes by making marketing data more accessible and easier to interpret while supporting better decisions and optimized campaigns and strategic planning thus giving financial institutions better market performance.

CHAPTER II

BACKGROUND OF STUDY

2.1 INTRODUCTION TO BACKGROUND STUDY

Direct mail stands as a vital marketing pathway in the adapting financial field since it generates superior response outcomes beyond digital marketing methods (BAI, 2024). The current set of marketing analytics systems does not supply sufficient targeted functions for real-time monitoring of direct mail activities in the financial sector. This study works to build a performance evaluation system for financial institutions that combines web-based interface and lightweight design to provide real-time performance metrics about their direct mail activities compared to competitors and spending details. Through operational intelligence solutions, the proposed system creates swift decision pathways that help financial marketers transform their strategies by using existing market patterns together with competitor reactions (Oracle, 2024). Financial institutions will gain marketing improvement together with stronger customer relationships and superior market positioning in the financial services sector (ThoughtSpot, 2024).

Figure 2.1: 4 stats prove DM's marketing potential

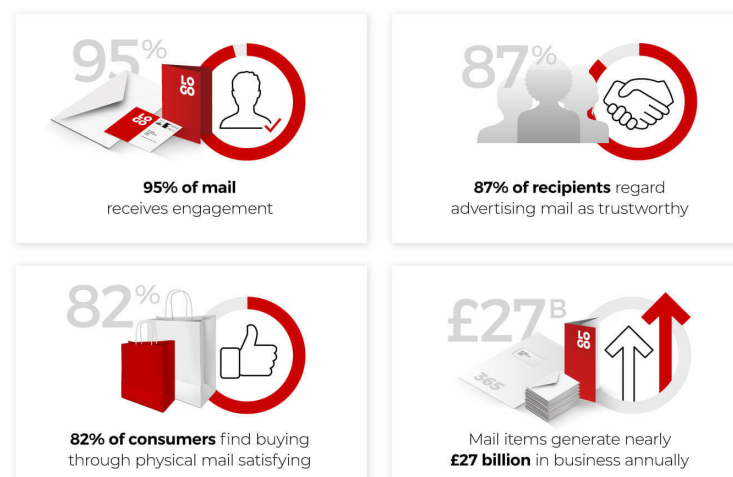
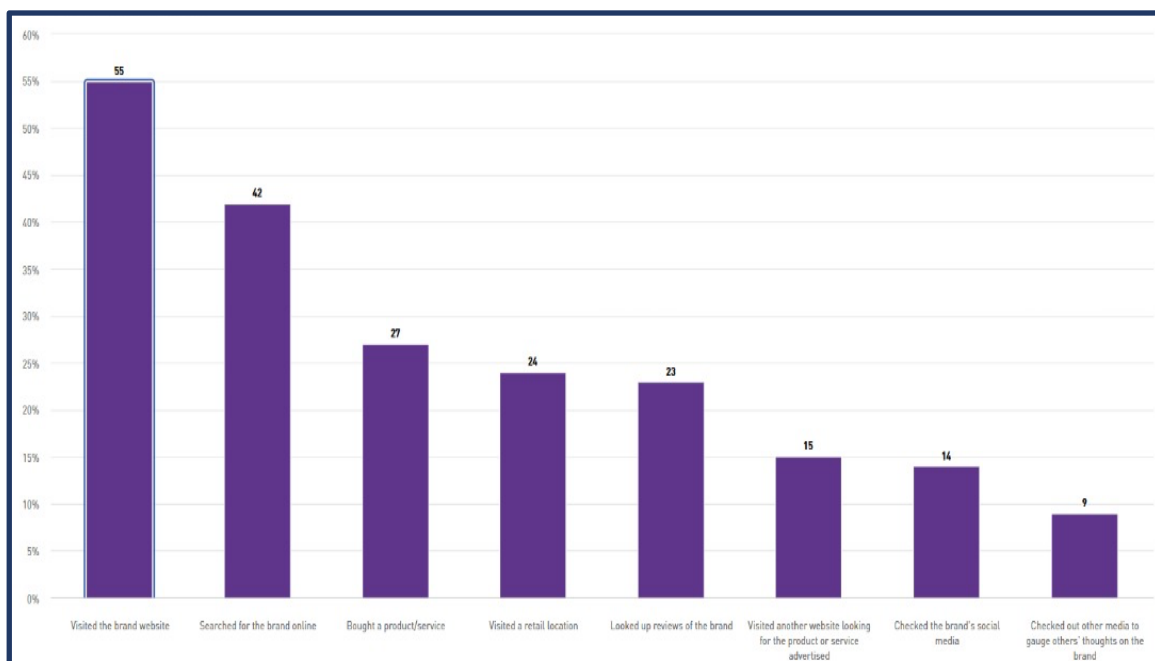


Figure 2.1: How users show how they engage in actions in DM usage that induce engagement opportunities, 2024.

Source: [Specific Action Taken on Direct Mail](#)



2.2 Review of Existing Systems

2.2.1 Mintel Comperemedia Direct

Figure 2.2.1: Comperemedia Direct



Mintel launched Comperemedia Direct in 1972 as its global market intelligence platform that provides enterprise data Analytics for acquisition strategies and pricing structures and targeted offers and product release projects across multiple marketing

channels including direct mail delivery. Through this platform users can display information as well as measure competitive performance while observing current market standings.

- **Strengths:** Users gain complete marketing information through this tool which helps them study competitor tactics.
- **Weaknesses:** The platform needs substantial implementation funds and employee training in order to exploit its entire functionality.
- **Operation:** Through this system users obtain access to extensive marketing material information and campaign records while creating filtering functions for specific data elements to develop strategy reports.

2.2.2 Tableau Marketing Analytics

Figure 2.2.2: Tableau Pulse Marketing Web Analytics



Tableau Software inaugurated its operation in 2003 through founders Christian Chabot and Pat Hanrahan and Chris Stolte to deliver Tableau Marketing Analytics to customers. Users of this platform integrate marketing data to obtain detailed views into their digital media expenses alongside social media reach and website traffic capabilities as well as the complete customer journey.

- **Strengths:** The platform provides users with keen data visualization features which help users make decisions based on data.
- **Weaknesses:** The software presents difficulties to customers who do not possess experience handling analytics tools.

- **Operation:** The system enables users to link multiple data sources to Tableau through its interface so they can build visualizations and distribute dashboards for marketing performance monitoring.

2.2.3 Dun & Bradstreet Marketing Analytics

Figure 2.2.3: Dun & Bradstreet Marketing Analytic



Since 1841 Dun & Bradstreet has developed marketing analytics solutions that enable businesses to search for prospects alongside prospect list quality enhancement alongside campaign performance tracking. The company develops tools that boost targeting precision and lead conversion efficacy because these elements serve as fundamental requirements for successful direct mail marketing success.

- **Strengths:** Dun & Bradstreet enables businesses to access broad business analytics and data for their marketing initiatives.
- **Weaknesses:** The company's main target market consists of B2B sectors which creates restrictions for B2C advertising campaigns.
- **Operation:** Users of Dun & Bradstreet data can use their information to conduct market segmentation as well as identify prospective clients along with analyzing their marketing campaign success.

2.2.4 Salesforce Marketing Cloud with Direct Mail Integration

Figure 2.2.4: Salesforce Marketing Cloud



The platform emerged as ExactTarget in 2000 before Salesforce purchased it in 2013 to form Salesforce Marketing Cloud. The platform helps users merge automated marketing sequences with direct mail allowing customers to receive sales promotion across multiple contact platforms.

- **Strengths:** The platform delivers one platform which manages multi-channel marketing campaigns designed for email and social media together with direct mail.
- **Weaknesses:** Technical expertise along with necessary resources may limit how readily users integrate and customize the system.
- **Operation:** Operation lets users build customer journeys which use direct mail points to deliver targeted and expedient messages across different contact methods.

2.2.5 Adobe Campaign – Direct Mail

Figure 2.2.5: Adobe Campaign



Advertising Campaign works as a part of Adobe Experience Cloud to deliver marketing automation tools that help users manage and automate their direct mail operations. Marketers can use this system to define their audience groups while creating content and creating files for mail vendors that enhance campaign management across offline platforms.

- **Strengths:** Adobe Campaign supports marketers to administer and automate direct mail marketing initiatives as part of campaign strategies.
- **Weaknesses:** The implementation of this solution requires advanced setup while its deployment needs substantial financial resources.
- **Operation:** Beside proper management of user-generated content the system enables users to produce and control direct mail efforts through audience selection and content design and mail service provider management for delivery.

2.2.6 Oracle Eloqua Marketing Automation

Figure 2.2.6: Oracle Marketing Cloud



The marketing automation platform Eloqua was established in 1999 before Oracle purchased it in 2012 to use as a cloud-based Software-as-a-Service (SaaS) solution. Eloqua delivers solutions that assist B2B marketing departments and business organizations in managing their marketing initiatives and obtaining sales leads.

- **Strengths:** Eloqua provides its users with extensive capabilities to nurture leads and to manage their marketing campaigns and conduct analytics.
- **Weaknesses:** Eloqua demonstrates limited capabilities in B2C marketing due to its main focus on B2B operations.

- **Operation:** Through the operation users deploy marketing campaigns across various channels via their available tools for email marketing and lead scoring and customer segmentation functions.

2.3 ANALYSIS ABOUT EXISTING SYSTEMS

Firstly, what I can observe is that the users cannot modify the displayed time frames in Mintel Comperemedia Direct and Dun & Bradstreet Marketing Analytics while analyzing competitive data because these platforms lack interactive dashboards. Users need to spend significant funds for getting mail comparison data through the direct marketing platform Mintel and users should possess special expertise to utilize its features effectively.

Meanwhile, Tableau and Oracle Eloqua offer powerful data visualization and analytics capabilities. Tableau allows users to benefit from its user-friendly dashboards yet its rigid setup requirements emerge since the system lacks direct mail analysis expertise and independent data source integration. The reporting system in Oracle Eloqua does not support competitive benchmarking of financial service mail as efficiently as it automates B2B lead management activities.

The direct mail automation capabilities exist in the marketing systems from Salesforce Marketing Cloud and Adobe Campaign. Their system manages campaigns internally with no external functionality for analyzing direct mail methods of competitors. Implementation complexities along with resource requirements act as barriers for financial institutions trying to gain fast market trend data rather than automated customer pathway optimization.

The present market lacks solutions that provide lightweight web-based real-time competitor direct mail benchmarking exclusively for financial institutions to track volume and expenditure data. Consumer businesses currently work with systems that either warrant broad application boundaries or monitor internal processes instead of external operations yet remain expensive to implement and require complex analytics abilities beyond basic technological skills.

2.3.1 Further Clarifications and In-Depth Analysis

The changing financial services market makes direct mail more important because it delivers connected experiences which digital platforms currently cannot provide. Direct mail produces superior response rates than email according to research findings and financial institutions utilize this method to reach 1.16% respondents in their lending promotions thus creating significant conversion prospects (CUInsight, n.d.). The rising popularity of direct mail requires financial institutions to develop strong capabilities for monitoring and analyzing their direct mail marketing strategies. An online dashboard system grants financial organizations access to real-time marketing data which lets them monitor their campaigns while comparing performance against rivals (CUInsight, n.d.).

Using direct mail with digital strategies produces an improved marketing outcome in overall terms. A proper integration of physical mail and digital follow-ups achieves response rate improvements reaching up to 62% according to KPM Group. Financial institutions achieve competitive advantage through a unified competitive analysis platform which provides them with market data about trends and customer demands and competitor movements to refine services and maintain market leadership. Platform adoption leads to better customer retention and acquisition rates as well as product and service delivery innovation through proactive measures (KPM Group, n.d.).

2.4 CONCLUSION

In conclusion, the evaluation of existing systems shows financial institutions lack proper tools which monitor their direct mail marketing trends in real-time alongside competitor data. The existing marketing analysis tools and campaign management solutions include comprehensive features yet they overlook specific comparative evaluations of direct mail expenses and distribution data between competitors operating in the financial industry. Financial institutions face a gap in the market because the proposed web-based dashboard solves this problem through its specific and easy-to-use interface which enhances decision-making while showing competitive market trends alongside efficient marketing strategy development. Through its ability to fill this missing functionality the system will generate lasting value that improves marketing

operational efficiency and strengthens market position and enables data-driven growth in the evolving financial industry.

CHAPTER III

METHODOLOGY

3.1 INTRODUCTION

This chapter explains the methodology applied to construct the web analytics platform through which financial institutions may analyze and compare direct mail marketing trends. It describes the structured steps followed in collecting the data, developing the design of the system, and in implementing the same, as well as explaining the choice of technical tools and approaches. The methodology perfectly fits the objectives of the project, which focus on data-driven insights with an interactive touch and user-oriented design. The project follows a systematic approach that incorporates steps such as: data collection, processing, visualization, and platform development. Combining prototyping skills with serious data handling through MySQL and Python, and putting the finishing touch on the web deployment through Streamlit, will ensure that the methodology makes a fine compromise among efficiency, scalability, and requirements of financial analysts.

3.2 RESEARCH METHODOLOGY

The research design for the project involves a design and development research approach, suitable for creating and assessing the functional artifact established here as the web-based analytics dashboard. This approach undertakes solving an actual-world problem, and it involves the iterative development, testing, and refining of the solution. Because the project focuses on providing a real problem faced by financial institutions in comparing and analyzing trends relating to direct-mail marketing, the design-based research framework indeed becomes relevant and useful. The research is a mixed-mode research design, employing both qualitative and quantitative strategies.

On the quantitative side, banks provide structured data on marketing expenditures and volumes, and these datasets are analyzed for trending, patterning, and comparison. On the qualitative side, stakeholder insights that can be collected via interviews and surveys inform the functional design of the platform, ensuring it implements user expectations and industry standards. This combination makes the project data-oriented and user-centered, which is something that is of great importance when building support tools for decision-making in financial analytics.

3.2.1 Agile Approach

For the real-time application of this design-oriented research, I have taken on the Agile software development technique. This method enables the iterative build, which is in harmony with the real-time nature of the dashboard and its features. If needed to perform the system through several phases such as database and data processing layer, visualization layer, and user interaction components I have used each iteration to gain feedback and thus to make a continuous polishing of the product. Such incremental way guarantees that the final solution is not only satisfying the user requirements effectively but also evolving smoothly (Clarkston Consulting, 2024).

The core of Agile, being collaboration, flexibility and speed, is just right for the technical and research purposes of the study. The Agile strategies foster regular communication and stakeholder participation at every stage, hereby guaranteeing that the end products are in line with the company's goals and the needs of the clients. This way of collaborative working raises the bar for the quality of the work output than if the work was done without such collaboration. Through this way, the product can be of high quality and based on time-adjustment as per feedback, thus more efficient and user-centric are the outcomes (Clarkston Consulting, 2024).

Relative to what is presented, Agile has a major role in risk management, allows for the early uncovering of probable issues, and continuous prioritization of tasks are two of the key benefits. As well as such structured flexibility improving response to change, particularly in data analytics projects where the needs and the insights continue to develop rapidly. In addition, Agile makes timely delivery of an implemented part the

priority, which also results in quicker creation of value, and the ones who are interested can attend the project's life cycle with visible progress and benefits (Clarkston Consulting 2024).

On the other hand, the conventional Software Development Life Cycle (SDLC) was thought of, but disqualified as it is linear. As there is always scope for adding extra features, or making enhancements to the dashboard, or even during the user testing stage, Agile is a better choice for this project. It aims at optimizing the working relationship of the developer and the users thus delivering the final output that satisfies the need perfectly.

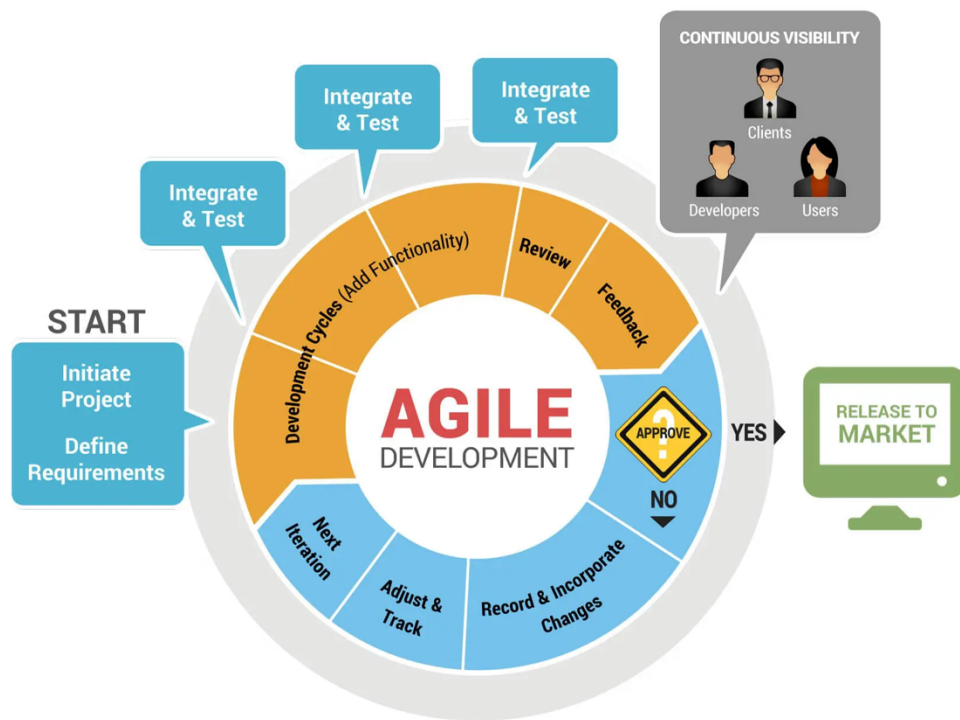
For the achievement of this project, agile methodology has been chosen due to the following reasons:

- Agile allows step-by-step developments of the web-based dashboard (Streamlit) with continuous enhancements.
- Agile facilitates the quick capture of the user's needs and the collection of data through the deployment of a paper prototype and the frequent feedback from potential users.
- Agile continuously increases the flexibility of the team to refactor the design in case of issues or the rejection of stakeholders concerning the work.
- Collaboration with the end-users, also known as financial analysts, can take place under the condition that the former is more comfortable with the Handson approach of the later.
- This also allows the application of the system where the data is being used to guide the user, the other way round, and the system is also interactive and user-friendly.
- The cycle of development being short, the period required for testing and deployment is shorter as well.

- This method is better suited for the realization of the customized features requested by the users, such as time periods that can be selected, downloading to CSV files, and the real-time tracking of trends.
- Agile further compartmentalizes the execution of the task by involving the steps of: **database (MySQL) → data processing (Python) → visualization (Streamlit).**

Figure 3.2.1: Agile Development Process

Source: [Agile Development Methodology](#)



3.3 DATA COLLECTION TECHNIQUES

3.3.1 Interviews

Guarantee to ensure that the project that can be brought to life is customer-friendly, I have conducted interviews with marketing analysts, data analysts, and people from bank that I had previously known. These discussions were fruitful and as a result

I have obtained qualitative data that helped me see the importance of real-time comparative analysis features and dashboard rankings.

According to the author of the User Interviews blog, Not only is it very flexible and adaptable, but user interviews are also very powerful tools for uncovering new opportunities and generating ideas during the discovery phase, indispensable to both qualitative and quantitative evaluative methods, and concurrent listening continuous methods to cope with users or client needs and opinions that change over time. (User Interview, 2023).

3.3.2 Questionnaires

Another approach is the method of questionnaires which in fact is a great addition to personal interviews, as it allows for the collection of a wider range of responses from a larger number of respondents without having to ask all of them directly. Through sending electronic surveys to people from the financial industry, technically competent developers can find the most important features, liked or accepted visualizations, as well as functional data and information categories that the users require on the dashboard.

Furthermore, questionnaires make it easier for one to attain the user's satisfaction and experiences during the prototype test phase. The information that is collected can then be correlated with the areas identified for improvements so that the dashboard maintains a constant evolution to meet its user's needs (SuperSurvey, 2023).

3.3.3 Observations

Observational methods are also used during prototype tests to uncover user activities on the web-based portal and likewise come up with the improvements. By watching users, who are walking through this system, where I can explore the issues of system's usability, deconstruct the users' behavior, as well as identify the areas that need improvement.

Moreover, this strategy instantly reveals the user experience limitations and the outstanding parts. In fact, the user's experience in carrying out the various tasks helps confirm that the platform is indeed enabling intuitive examination of data and identifying trends (Dovetail, 2024).

3.3.4 Third-Party Data Acquisition

To expand the exposure and consistency of the analytics platform, the project brings in outside data vendors who are specific to the direct mail marketing campaigns. This project discusses Comperemedia Direct as the vendor to be used for this reason, since they have data that is highly accurate and consistent. These outside vendors provide detailed relational databases on the volume of the campaign, format types, geographic distribution, and industry-specific benchmarking, among other important issues to be covered on how to conduct an analysis of the entire project (Deepsync, 2025).

Through this process of integrating external datasets into the database of the project, the dashboard can be provided with information from actual practice on a big scale which otherwise would require too much of an effort to be gathered in such a manner. Thus, it gets more reliable, becomes more actual and competitive, and the platform benefits from comparative benchmarking, also the dashboard can support the forecast of the trends, and the evaluation of the campaign's effectiveness (Deepsync, 2025).

3.4 PROTOTYPE OF SYSTEM DEVELOPMENT METHODOLOGY

When designing a system prototype, one of the crucial steps to take is to use an approach that is systematic and open but at the same time can easily change in line with the new requirements and at the same time, it encourages the continuous or ongoing progression. Agile was chosen as the suitable methodology because of its iterative nature, inclination for communication, and outright flexibility. The usage of Agile brings in help to the project by means of a steady stream of feedback, thus, staying close to the specification and with the user's impression of the product.

Table 3.4: Comparison of Methodology

Methodology	Description	Suitability to this project?
Waterfall	A linear and sequential approach where each phase is completed before the next phase begins.	Rigid and inflexible. It is not suitable for iterative improvements or continuous customer feedback.
Prototyping	Focus more on creating quick mockups or models that only serve to gather user feedback.	Useful for early design evaluation but lacks the structure needed for full-cycle delivery.
Agile	An iterative model that emphasizes collaboration, adaptive planning, and continuous delivery.	It is most suitable because it supports the evolving requirements process and continuous user feedback.

3.5 PROTOTYPE METHODOLOGY IMPLEMENTATION

Implementing Agile in all its effectiveness was done by the development process splitting into a series of structured sprints with each of them concentrating on particular deliverables software and their enhancements. It, in turn led to modular tracking, testing, and adaptation.

3.5.1 Sprint Planning and Execution

The project milestones which have been elaborated in Chapter 1 via Table 1.6 are the main outcomes and the time periods which are necessary for the conclusion of the project and to ensure that a structured and iterative development process was followed through, the Agile methodology was employed, which permitted the delivery of system components incrementally through several sprints. Each sprint was

meticulously synchronized with unique project milestones so that the process of the continuous tracking of the progress, the early testing, and the timely adjustments according to the feedback will become much easier.

The following table is an overview of the symmetry between Agile sprint cycles and the specified project milestones:

Table 3.5.1: Agile Sprint Mapping

Sprint	Duration	Agile Activities	Aligned Milestone
0	29 Apr - 03 May	<ul style="list-style-type: none"> • Backlog creation • Requirement analysis • Define technical stack (MySQL, Python, Streamlit) - Gantt chart + timeline setup 	<ul style="list-style-type: none"> • Requirement Study & Planning • Proposal finalization, tool selection, functional spec, project scope
1	07 May – 09 May	<ul style="list-style-type: none"> • Design ERD & DB schema • Build initial Streamlit layout • Create MySQL database and populate test data 	<ul style="list-style-type: none"> • Database Design • ERD, schema creation & frontend structure planning
2	10 May - 14 May	<ul style="list-style-type: none"> • Backend data collection & import • Panda's preprocessing • Handle missing/invalid data 	<ul style="list-style-type: none"> • Data Collection & Preprocessing • Use Python to clean and structure direct mail datasets
3	15 May - 17 May	<ul style="list-style-type: none"> • Implement trend analysis logic • Statistical summaries • Visualizations (line charts, bar charts) 	<ul style="list-style-type: none"> • Trend Analysis Development Spending trends, mail volume insights, visualization logic
4	21 May - 23 May	<ul style="list-style-type: none"> • Finalize Streamlit dashboard UI • Add filters, dropdowns, user controls 	<ul style="list-style-type: none"> • Streamlit Dashboard • Full interface with interactivity and reporting

		<ul style="list-style-type: none"> • Implement CSV download button 	
5	24 May - 28 May	<ul style="list-style-type: none"> • Test full app (filters, visuals, downloads) • Debug any UI or backend issues • Validate user scenarios 	<ul style="list-style-type: none"> • Testing & Debugging • Fix usability/logic bugs, QA validation
6	29 May - 03 June	<ul style="list-style-type: none"> • Finalize Chapter V & VI • Include charts, analysis results • Add appendix, code screenshots • Format citations and references 	<ul style="list-style-type: none"> • Report Writing & Supervisor Review • Results write-up, conclusion & formatting
7	04 June	<ul style="list-style-type: none"> • Submit final version • Conduct presentation with dashboard walkthrough 	<ul style="list-style-type: none"> • Submission & Presentation • Final delivery of thesis and demo

3.6 PROGRAMMING LANGUAGE AND JUSTIFICATION

Below is a combination of technologies that have been selected based on their effectiveness for data processing, visualization and web usage to make this project a success:

Table 3.6: Programming Language Used

No	Technology	Role in Project	Justification
1	Python	Back-end logic, data analysis	A rich ecosystem of libraries like Pandas (for data manipulation), NumPy (for numerical computing) and Matplotlib (for data

			visualization) efficient for handling large data sets and performing complex analyses.
2	Streamlit	Web application framework	Python-based open source framework specifically designed to create interactive and responsive web dashboards.
3	MYSQL Workbench	Relational database management	Powerful and user-friendly tools for designing, managing, and querying relational databases.
4	CSV export	Reporting and data downloads	Allows users to download analyzed reports and upload raw data for processing.

3.7 CONCLUSION

In conclusion, this project was designed in a way that a data-focused web analytics platform for the financial sector could be developed and used to compare and analyze the direct mail marketing effect. Through the methodology of the Agile framework, the project could capitalize on the manifestation of incremental enhancements, the collaboration with stakeholders, and the adaptability to the feedback of the product's users. The rich information which the platform was imparted due to the interviews, questionnaires, observations, and the use of external data sources made it even more useful by creating value for the users. The step-by-step managing of the project tasks, which began with the creation of the database and ended with the deployment of the Streamlit dashboard, was so well-coordinated as to result in the timely delivery of the components with scalability, usability, and real-time insights.

CHAPTER IV

DATA COLLECTION AND TREND ANALYSIS

4.1 INTRODUCTION

This chapter explains the complete process from start to finish of merchandising the dataset that supported creating the internet-based competitive analysis tool for direct mail marketing in U.S. financial institutions. It all begins with gathering of raw data, that includes monthly estimated marketing expenditure and mail volume from top brands such as Bank of America, Citibank, and American Express. This basic data mainly of what helps to know expenditure of financial institutions and the way they are using direct mail campaigns over time.

Then the analyzed data part surfaces presented in a relational database that was created and administered by MySQL Workbench followed by a step-by-step guide on the same data preprocessing and cleaning stage carried out through the usage of Python libraries such as Pandas. The data, after it has been cleaned and formatted, is taken for analysis to reveal the various spending and mail volume trends, identify the outlying data points, and compute comparative statistics. Observations that constitute the functional part of the visualized Streamlit dashboard. The interactiveness of the precise data characteristics allows the banking institutions to use it effectively for tracking, comparing, and ultimately optimizing the direct mail marketing performance.

4.2 DATA COLLECTION

To start preparing an analytics dashboard, need to load the data in MySQL database and combine it into a table at the same time as Data collection in Chapter 3 in section [3.3.4: Third-Party Acquisition](#) has been described. Major details among other

things like Company Name, Product, Month, estimated Mail Volume, and Spending must be gathered to be able to do the additional analysis. The information is also gathered from the biggest banks in the US which are Bank of America, Citibank, American Express, and so on. These are employed marketing analysts for a competitive analysis in the future.

4.2.1 Data Source

The dataset includes:

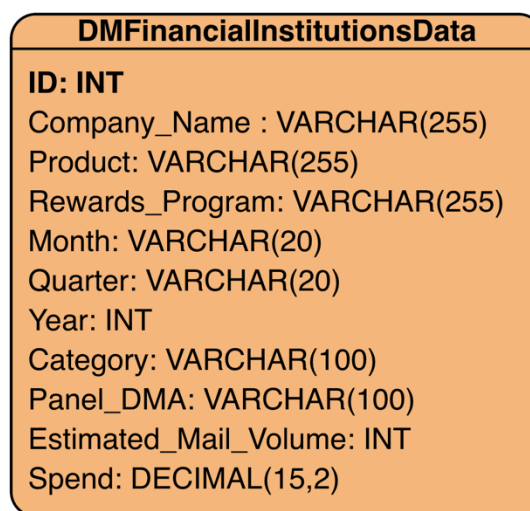
- 1) Company Name
- 2) Product Type
- 3) Rewards Program
- 4) Month
- 5) Date
- 6) Panel DMA
- 7) Estimated Mail Volume
- 8) Spend (USD)

4.2.2 Database Design with MySQL

Since the original data were unstructured at the beginning and present in a variety of formats, it was needed to restructure the data to fit the results of accurate and efficient analysis. In order to do this, a relational database was designed and created using MySQL Workbench. This method not only reaches through to data normalization but also it is able to reduce redundancy as well as easing the data maintenance and retrieval operations. For instance, by carving out clear, comprehensible, and simple links between the round-up tables and the bank names that they represent, the data itself for the analyst becomes moderately, adaptably, and instantly accessible for research.

The utilization of MySQL will offer the benefit of advanced data management tools such as indexing, filtering and aggregating for large datasets, which are necessary when handling time-series data like monthly marketing expenditure and mail volume. Furthermore, a well-structured schema will make it easier for marketing analysts and decision-makers to uncover insights, perform different comparisons or integrate data with visualization tools in the latter part of the project.

Figure 4.2.2: Entity Relationship Diagram (ERD)



Below is the SQL script used to create the database structure for DM Financial Institutions Data:

Figure 4.2.2: SQL "CREATE TABLE" Script for DM Financial Institutions Data

```

CREATE TABLE DMFinancialInstitutionsData (
  ID INT PRIMARY KEY,
  Company_Name VARCHAR(255),
  Product VARCHAR(255),
  Rewards_Program VARCHAR(255),
  Month VARCHAR(20),
  Quarter VARCHAR(20),
  Year INT,
  Category VARCHAR(100),
  Panel_DMA VARCHAR(100),
  Estimated_Mail_Volume INT,
  Spend DECIMAL(15,2)
);
  
```

Figure 4.2.2: Inserting Data into DM Financial Institutions Table Using SQL

```

INSERT INTO DMFinancialInstitutionsData (
  ID, Company_Name, Product, Rewards_Program, Month, Quarter, Year, Category, Panel_DMA, Estimated_Mail_Volume, Spend
) VALUES
(1, 'Citibank', 'Citi Checking Account', 'Not Specified', 'Nov-22', '2022 Q4', 2022, 'Checking', 'New York, NY', 2989937, 1566231.00),
(2, 'American Express', 'American Express High Yield Savings Account', 'Not Specified', 'May-24', '2024 Q2', 2024, 'Savings', 'Los Angeles, CA', 2667156, 782620.00),
(3, 'American Express', 'American Express High Yield Savings Account', 'Not Specified', 'Jan-24', '2024 Q1', 2024, 'Savings', 'New York, NY', 2534503, 741431.00),
(4, 'American Express', 'American Express High Yield Savings Account', 'Not Specified', 'Mar-25', '2025 Q1', 2025, 'Savings', 'Los Angeles, CA', 2532016, 735639.00),
(5, 'American Express', 'American Express High Yield Savings Account', 'Not Specified', 'May-24', '2024 Q2', 2024, 'Savings', 'New York, NY', 2469819, 722183.00),
(6, 'American Express', 'High Yield Savings Account', 'Not Specified', 'Nov-24', '2024 Q4', 2024, 'Savings', 'New York, NY', 2353946, 681925.00),
(7, 'American Express', 'High Yield Savings Account', 'Not Specified', 'Aug-22', '2022 Q3', 2022, 'Savings', 'New York, NY', 2298227, 675257.00),
(8, 'American Express', 'High Yield Savings Account', 'Not Specified', 'Jan-23', '2023 Q1', 2023, 'Savings', 'Los Angeles, CA', 2235329, 649165.00),
(9, 'American Express', 'American Express High Yield Savings Account', 'Not Specified', 'Mar-25', '2025 Q1', 2025, 'Savings', 'New York, NY', 2232363, 653657.00),

```

4.2.3 How This Schema Helps in The Reordering of Data for Analytics

1) Clear Column Structure

Each column like **Company_Name**, **Product**, **Month**, **Year**, **Spend** stores a particular type of data so that querying, filtering, and analysis will not be a very big issue.

2) Unique Identification

The ID column determines the primary key, ensuring that the row will be made distinct and as well as fast indexing and subsequent retrieval be made easy.

3) Time-Based Analysis

Features like **Month**, **Quarter**, and **Year** are the key factors to the wave of time that assist in trend analysis, seasonal comparisons, and time-based aggregations.

4) Categorical Grouping

Attributes like **Company_Name**, **Product**, and **Panel_DMA** are act as a pivot for grouping and filtering by company, product type, or region.

5) Quantitative Metrics

Estimated_Mail_Volume and **Spend** is numerical data that provides data accuracy if analysts want to use it when they make calculations, such as totals, averages, or return on investment.

6) Consistent and Scalable

Technical matters about the data type and a straightforward structure are the factors that allow the schema to be used consistently and to provide the opportunity for the schema to continue growing without having to change it frequently.

7) Supports Advanced Queries

Schema supports a variety of queries that are in advance, specifically the one for instance that can be put in an SQL language query to get insights such as a sum of the total **Spend** by region or **Mail Volume**.

4.3 DATA PREPROCESSING

Data preprocessing was performed using Python and the Pandas library, which resulted in a clean, organized data set that was ready for analysis. Several important steps were part of the process, such as replacing missing data, sorting out incorrect data formats, giving columns more meaningful names, and changing data types for tool compatibility. The data set was simplified by removing unnecessary columns and the variable categories were standardized to ensure consistent records. It is through this preprocessing action that data quality is improved, noise is reduced, and a structured format that enables fast queries, visualizations, and machine learning is created.

4.3.2 Cleaning and Structuring

From the given data, the process of data cleaning and structuring was carried out very meticulously and had several vital phases to ensure the authenticity and consistency of the analysis:

- Removed duplicate rows to get rid of the same records and to have the validity of the records remained intact.

- Changed the bank names like "Bank of America", "BOA", "BofA" into one unified format, so that proper filtering, grouping, and comparison of institutions are possible.
- Transformed the date variables into a **Month**, **Quarter**, and **Year** format which are more suitable and convenient for Time-based research (or analysis) like trend detection and seasonal impact. It also made the values consistent across columns.

Figure 4.3.2: Python code snippet for Data Cleaning & Structuring

```
# Data Cleaning -----

# Standardize column names
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")

# Strip whitespace from object columns
for col in df.select_dtypes(include='object').columns:
    df[col] = df[col].str.strip()

# Remove Duplicates -----

df = df.drop_duplicates()

# Standardize Bank Names -----

bank_name_map = {
    "BOA": "Bank of America",
    "BofA": "Bank of America",
    "Bank Of America": "Bank of America",
    "Amex": "American Express",
    "American Express Bank": "American Express",
    "Citi": "Citibank",
    "Chase Bank": "Chase",
    "JP Morgan Chase": "Chase"
}
df["company_name"] = df["company_name"].replace(bank_name_map)

# Format Date Columns -----

# Convert 'month' to datetime
df["month"] = pd.to_datetime(df["month"], errors="coerce")

# Ensure 'year' is numeric
df["year"] = pd.to_numeric(df["year"], errors="coerce").astype("Int64")

# Standardize 'quarter' format
df["quarter"] = df["quarter"].astype(str).str.strip()
df["quarter"] = df["quarter"].str.replace(r"(\d{4})\s*[Qq](\d)", r"\1 Q\2", regex=True)
```

Figure 4.3.2: Result for Data Cleaning & Structuring

```
1) Final Cleaned Data Shape: (6847, 11)
2) Unique Banks: 9
3) Month Range: 2022-05-01 00:00:00 to 2025-04-01 00:00:00
```

4.3.3 Handling Missing Values and Outliers

- After cleaning the data, the column names were made uniform by removing spaces, changing to lowercase, and replacing spaces with underscores.
- I counted the missing values and then filled in the nulls before and after cleaning by using the `isnull().sum()` method and explained the percentages for ease of understanding.
- I also eliminated those rows that had no vital fields like **Company_Name**, product, month, **Estimated_Mail_Volume**, and **Spend** to avoid tampering with the data quality.
- To achieve the perfectness of data quality while working on the most necessary fields, such as **Company_Name**, **Product**, **Month**, **Estimated_Mail_Volume**, and **Spend**, I've completed the missing values with "Unknown" using `.fillna()` that allowed the result to have the non-critical attributes full.
- Using `.fillna()` in the optional fields (e.g. **Rewards_Program**) will ensure that the non-critical attributes are also complete, for this reason, these fields were filled with "Unknown".
- I've applied `pd.to_numeric()` to convert the data types of string fields such as the **Product**, **Month**, and **Company_Name** details, while others like **Estimated_Mail_Volume** and **Spend** were converted to numeric by simply using the appropriate attributes.

- A somewhat similar operation was achieved with the `pd.to_numeric()` method but this time, after converting text data fields such as **Product**, **Month** and **Company_Name** entries to numbers, float and int fields such as **Estimated_Mail_Volume** and **Spend** were converted directly without using any additional information.

Figure 4.3.3: Python code snippet for Handling Outliers (IQR Method)

```
# Handle Outliers (IQR Method) -----

numeric_cols = ["estimated_mail_volume", "spend"]

for col in numeric_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    df[col] = np.where(df[col] < lower, lower, np.where(df[col] > upper, upper, df[col]))

# Clean column names
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")

# Strip whitespace from object columns
for col in df.select_dtypes(include='object').columns:
    df[col] = df[col].str.strip()

# Standardize and clean numeric columns (just for plotting)
df = df.dropna(subset=["estimated_mail_volume", "spend"])
df["estimated_mail_volume"] = pd.to_numeric(df["estimated_mail_volume"], errors="coerce")
df["spend"] = pd.to_numeric(df["spend"], errors="coerce")

# Plot histograms and boxplots to visualize outliers
sns.set(style="whitegrid")
fig, axes = plt.subplots(2, 2, figsize=(14, 10))

# Histogram and Boxplot for Estimated Mail Volume
sns.histplot(df["estimated_mail_volume"], bins=30, ax=axes[0, 0], kde=True, color='skyblue')
axes[0, 0].set_title("Histogram of Estimated Mail Volume")

sns.boxplot(x=df["estimated_mail_volume"], ax=axes[0, 1], color='lightgreen')
axes[0, 1].set_title("Boxplot of Estimated Mail Volume")

# Histogram and Boxplot for Spend
sns.histplot(df["spend"], bins=30, ax=axes[1, 0], kde=True, color='salmon')
axes[1, 0].set_title("Histogram of Spend")

sns.boxplot(x=df["spend"], ax=axes[1, 1], color='orange')
axes[1, 1].set_title("Boxplot of Spend")

plt.tight_layout()
plt.show()
```

Figure 4.3.3: Python code snippet for Handling Missing Values

```
# Handle Missing Values -----

# Clean column names
df_raw.columns = df_raw.columns.str.strip().str.lower().str.replace(" ", "_")

# Step 1: Missing Summary BEFORE Cleaning ===
missing_before = df_raw.isnull().sum().reset_index()
missing_before.columns = ["column_name", "missing_before"]
missing_before["missing_before_pct"] = (missing_before["missing_before"] / len(df_raw)) * 100

# Step 2: Clean the data ===
df_clean = df_raw.copy()

# Drop rows with missing essential values
df_clean = df_clean.dropna(subset=["company_name", "product", "month", "estimated_mail_volume", "spend"])

# Fill optional fields
if "rewards_program" in df_clean.columns:
    df_clean["rewards_program"] = df_clean["rewards_program"].fillna("Unknown")

# Step 3: Missing Summary AFTER Cleaning ===
missing_after = df_clean.isnull().sum().reset_index()
missing_after.columns = ["column_name", "missing_after"]
missing_after["missing_after_pct"] = (missing_after["missing_after"] / len(df_clean)) * 100

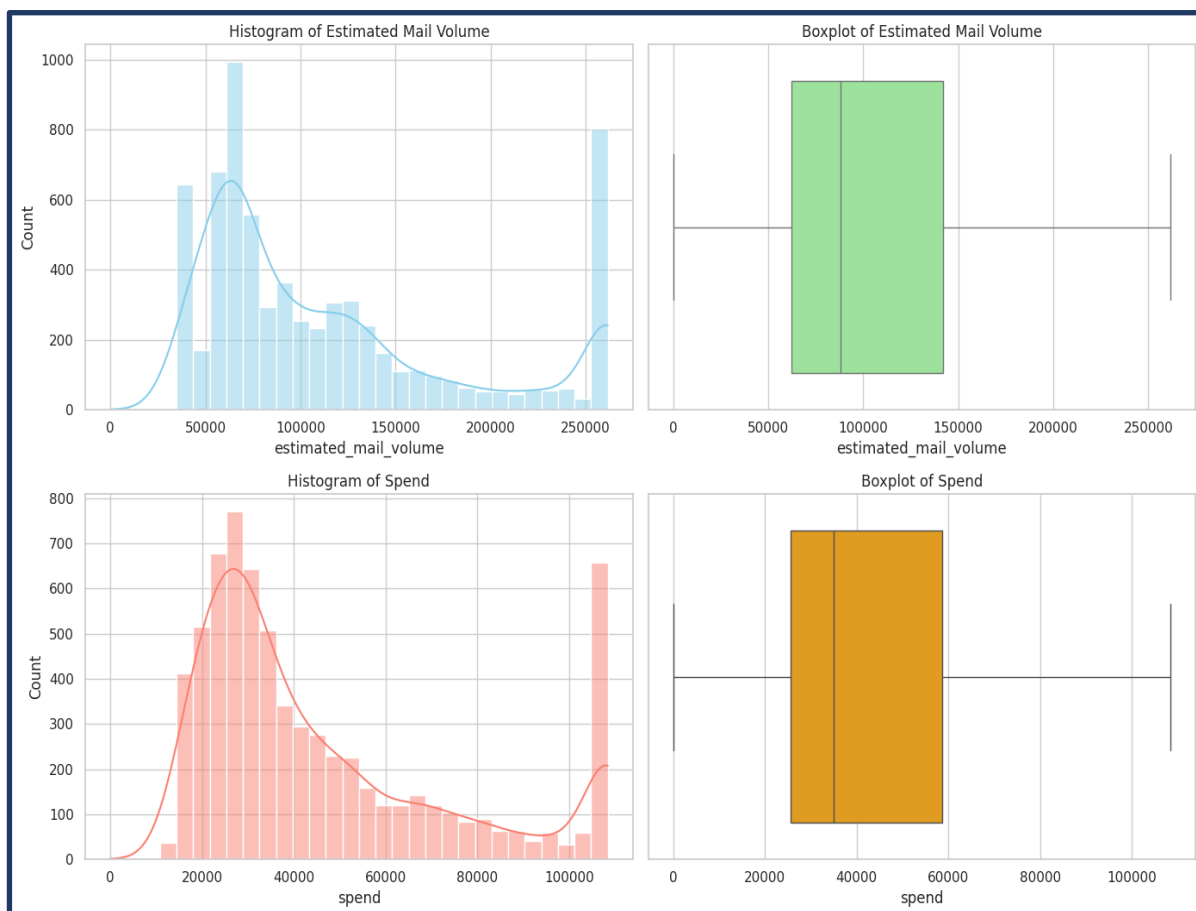
# Step 4: Merge & Format Table ===
missing_summary = pd.merge(missing_before, missing_after, on="column_name")
missing_summary = missing_summary.sort_values(by="missing_before", ascending=False)

# Format percentages for display
missing_summary["missing_before_pct"] = missing_summary["missing_before_pct"].map("{:.2f}%".format)
missing_summary["missing_after_pct"] = missing_summary["missing_after_pct"].map("{:.2f}%".format)
```

Figure 4.3.3: Result for Handling Missing Values

```
4) Missing Value Table:
  column_name  missing_before  missing_before_pct  missing_after  missing_after_pct
    id          0              0.00%                0              0.00%
  company_name  0              0.00%                0              0.00%
    product     0              0.00%                0              0.00%
  rewards_program  0              0.00%                0              0.00%
    month       0              0.00%                0              0.00%
    quarter     0              0.00%                0              0.00%
    year        0              0.00%                0              0.00%
    category    0              0.00%                0              0.00%
    panel_dma   0              0.00%                0              0.00%
  estimated_mail_volume  0              0.00%                0              0.00%
    spend       0              0.00%                0              0.00%
```

Figure 4.3.3: Distribution and Outlier Detection of Estimated Mail Volume & Spend Using Histograms and Boxplots



- **Showing Data Patterns** - The histogram shows that most banks send less Mail and spend less, while only a few send or spend a lot.
- **Detecting Unusual Values** - Box plots help identify values that are significantly higher than normal, these are called outliers.
- **Outliers Can Mislead Results** - If outliers are not handled, the data can make the average appear higher than it really is and lead to incorrect conclusions.
- **Determines the Process of Cleanliness** - In doing so the image assists in the selection of whether the cleaning of the data will be done by fixing the extreme values or by limiting the use of the IQR method.

- **Interpreting the Outcomes of Unsymmetric Data** - Given that there is a disparity in the data, using only the average would not be enough and so we have to use the median or change the scale.
- **Data in Better Quality** - By this, the gathering of the data is made sure that it is without dirt and hence, is more reliable for analysis so that the conclusions can be safely reached.

4.4 TREND ANALYSIS

Once the dataset had been through the data cleaning and data restructuring process, the U.S. financial institution's direct mail marketing history was to be studied by analyzing the data on temporal and categorical patterns. The database was enriched by creating several new date-related columns (**Year**, **Year_Month**) and through naming columns in a consistent way. To spot both high-level trends and detailed patterns in spend and mail volume, several charts were plotted using the Matplotlib and Seaborn libraries. Thus, a very complete picture was created of the timepoints, the companies that were at the top of the heap, and the changes in year-over-year behaviors in the industry.

4.2.2 Spending Trends

In this part, a chart displayed the monthly marketing expenses space in top financial institutions from the years 2022 until 2025. A line graph was employed to indicate changes in the expenses following a period which was selected to show the top participants with the biggest daily expenses and thus assist in focusing on the details. Function of **.nlargest()** was applied to get the top 5 spenders per month and later on, these data were filtered for visualization, showing the seasonality and the most successful months. Also, a bar chart was used for comparing total various company expenses over the whole data set and it gave a birds-eye view of the major players.

Figure 4.4.1: Monthly Spending Across Bank (2022 - 2023)

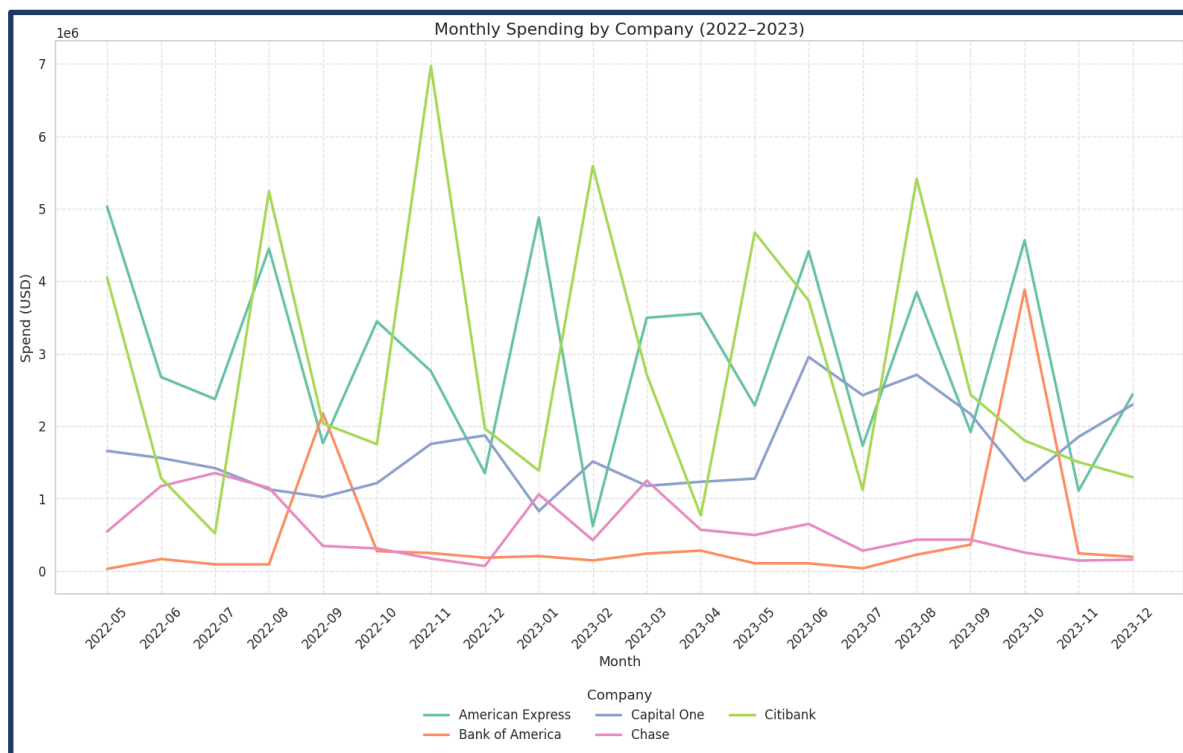


Figure 4.4.1: Monthly Spending Across Bank (2024 - 2025)

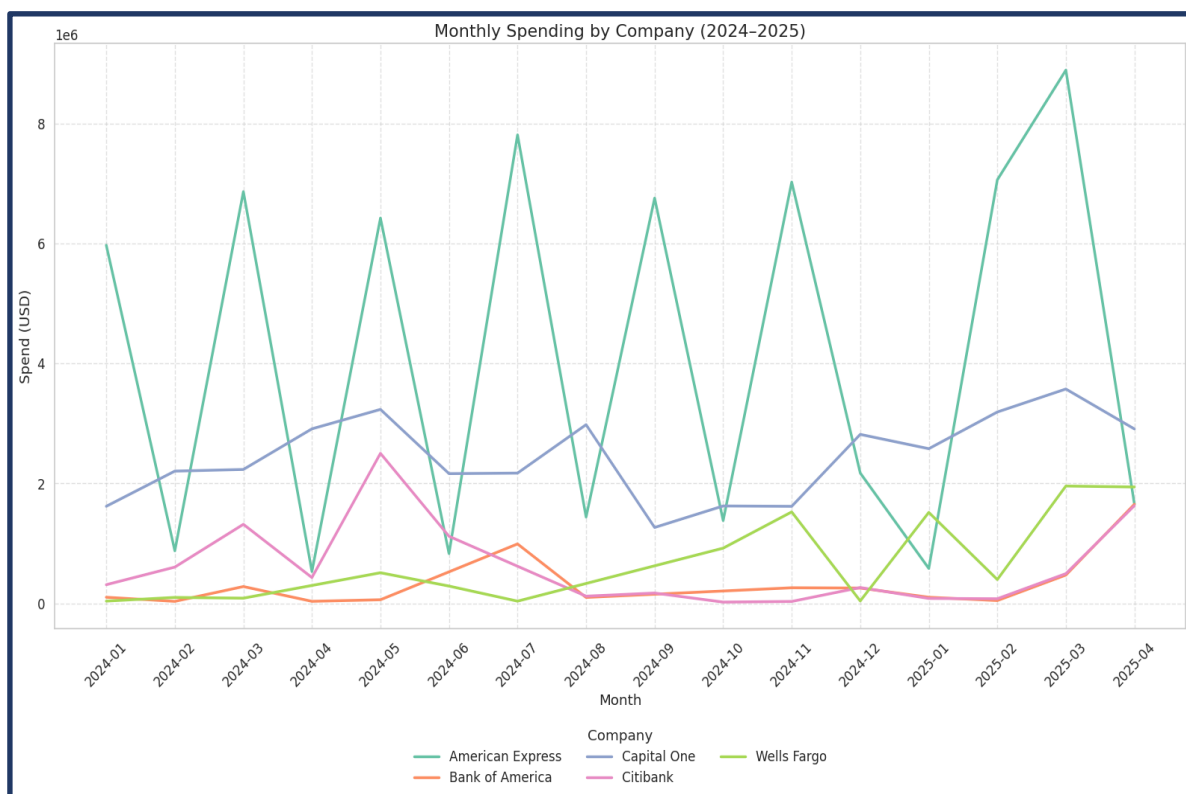


Figure 4.4.1: Total Spending Across Bank (2022 - 2025)

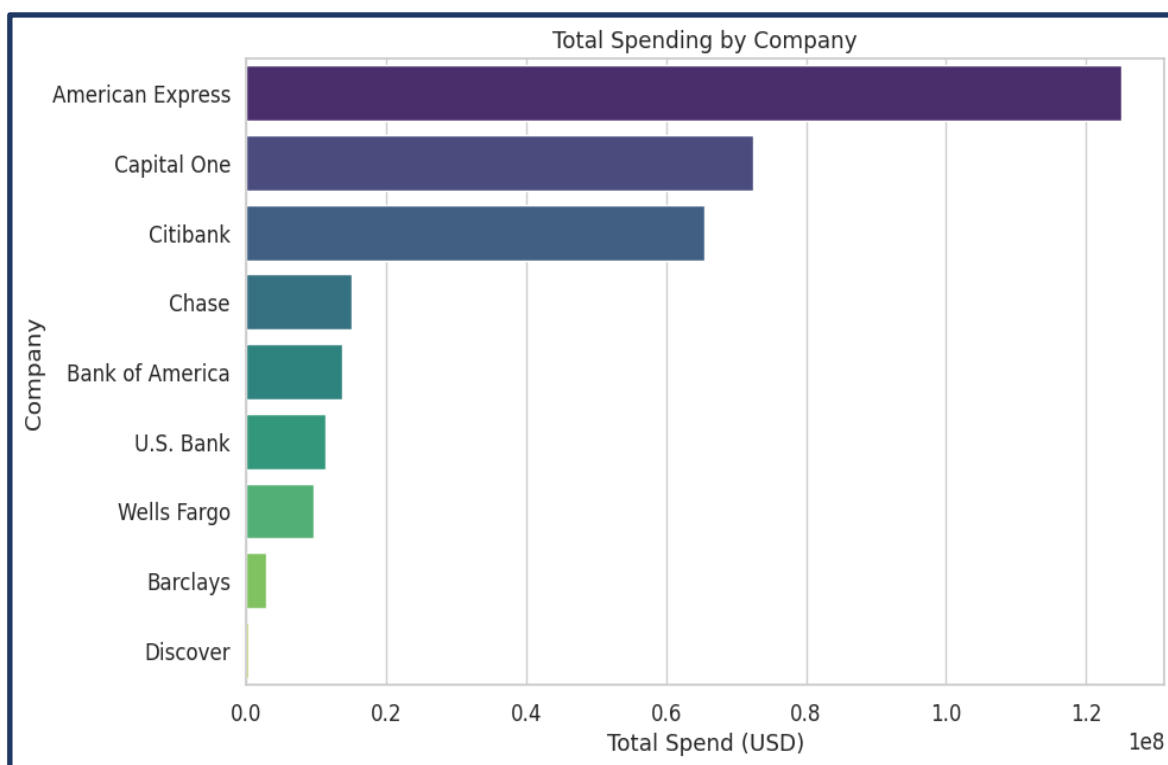
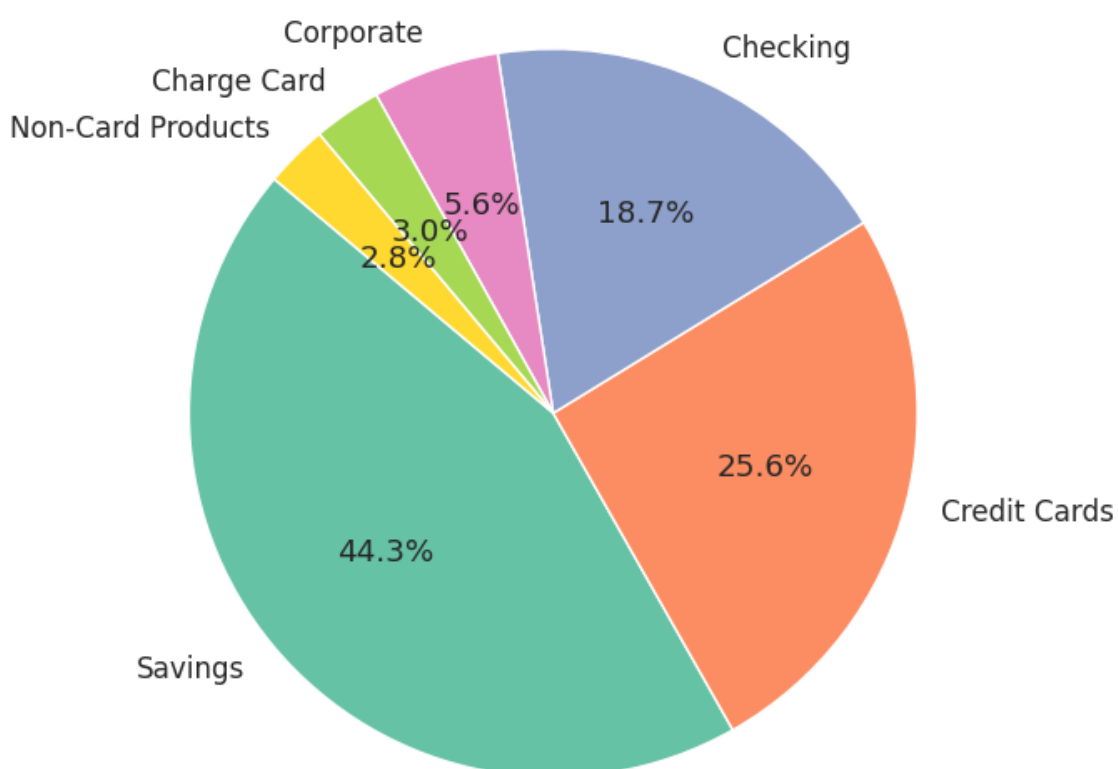


Figure 4.4.1: Top Card Categories by Spend Across Bank (2022 - 2025)



4.2.3 Mail Volume Trends

Trends in estimated mail volumes for a specific period of time and by institution were taken into consideration. A bar chart on a monthly basis was utilized to show the ebbs and flows of the mail volume throughout the different **Year_Month** periods. This of course had to do with the people's changing marketing behavior during the various seasons of the year. An additional bar chart was compiled reflecting the companies at the top of the email volume list. This bar chart was set next to the other one for a more detailed comparison. In addition, a pie chart was prepared to specify the Panel DMA regions which were the top 5 by mail volume.

Figure 4.4.2: Monthly Estimated Mail Volume for all Bank (2022 - 2025)

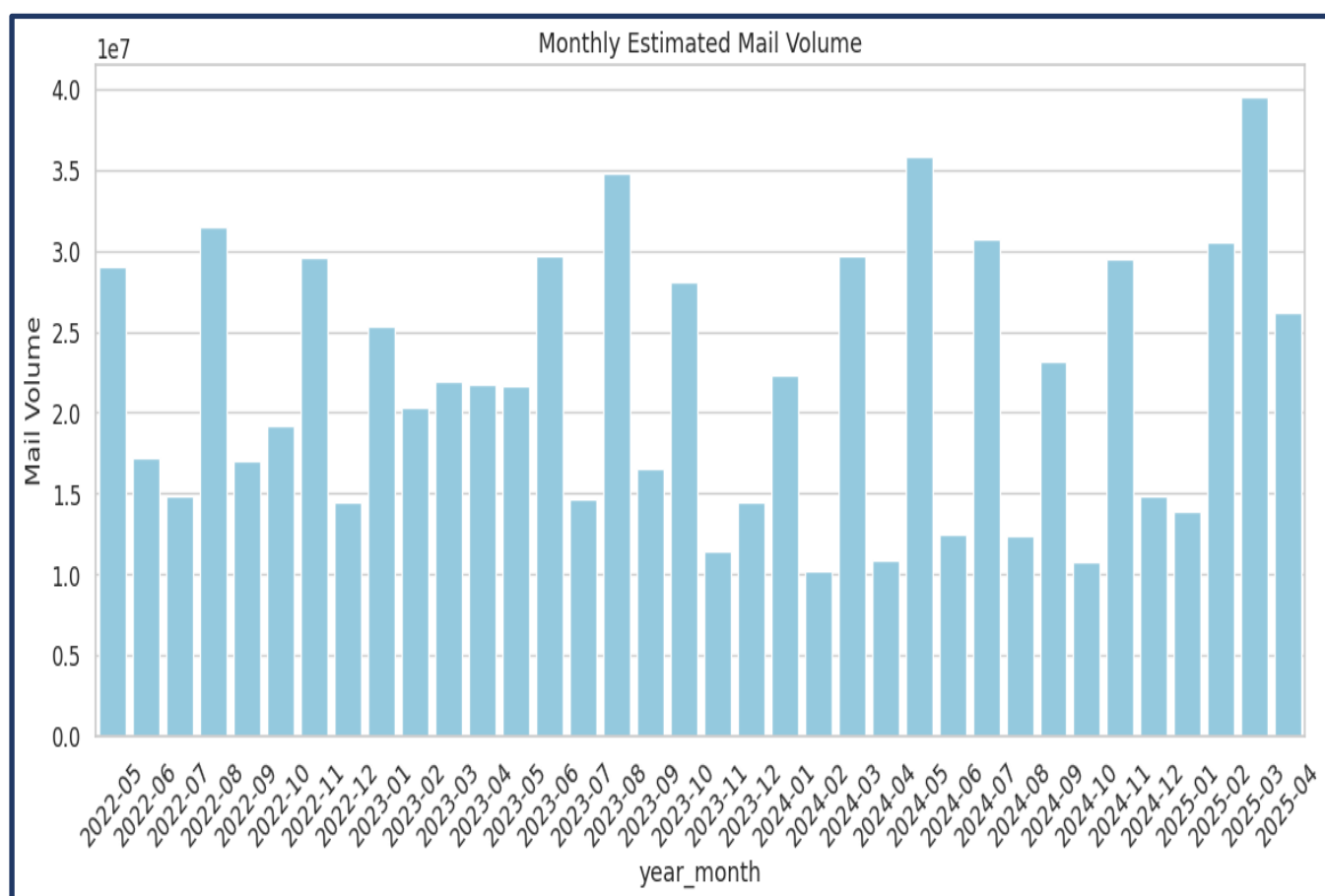


Figure 4.4.2: Total Estimated Mail Volume Across Bank (2022 - 2025)

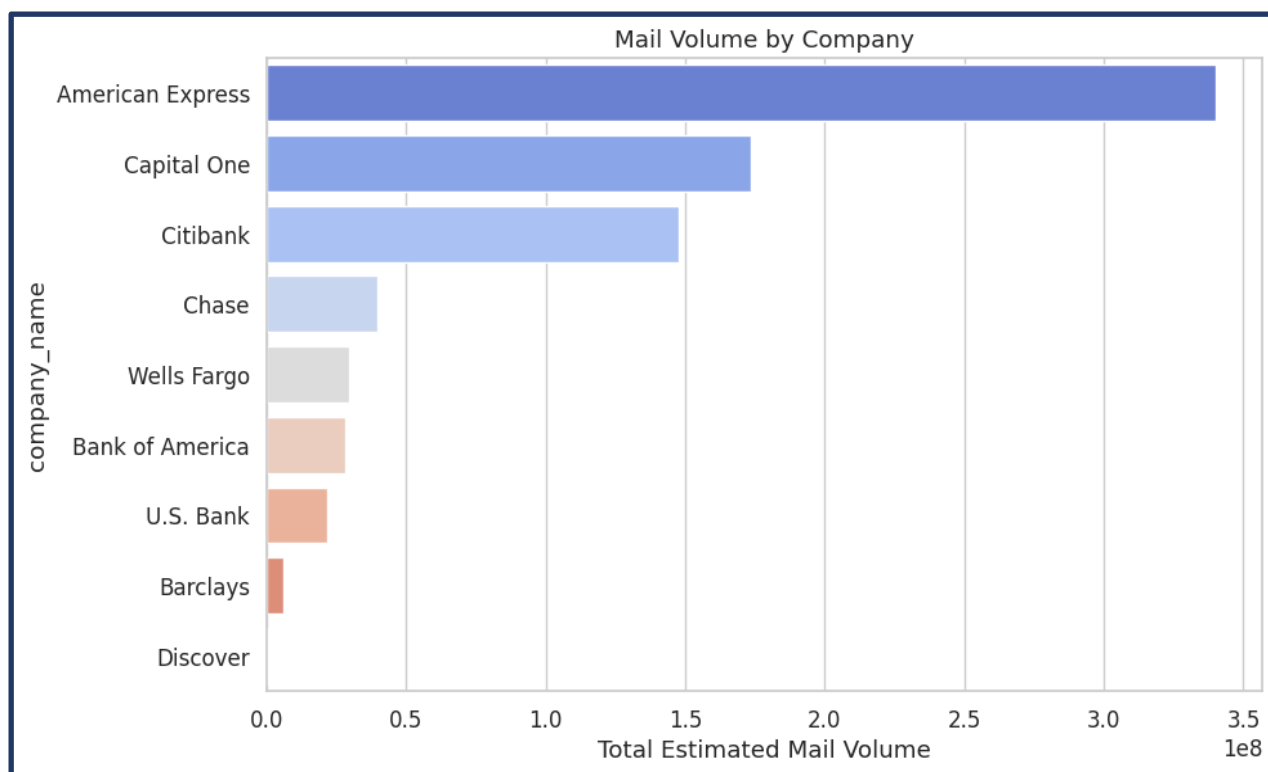
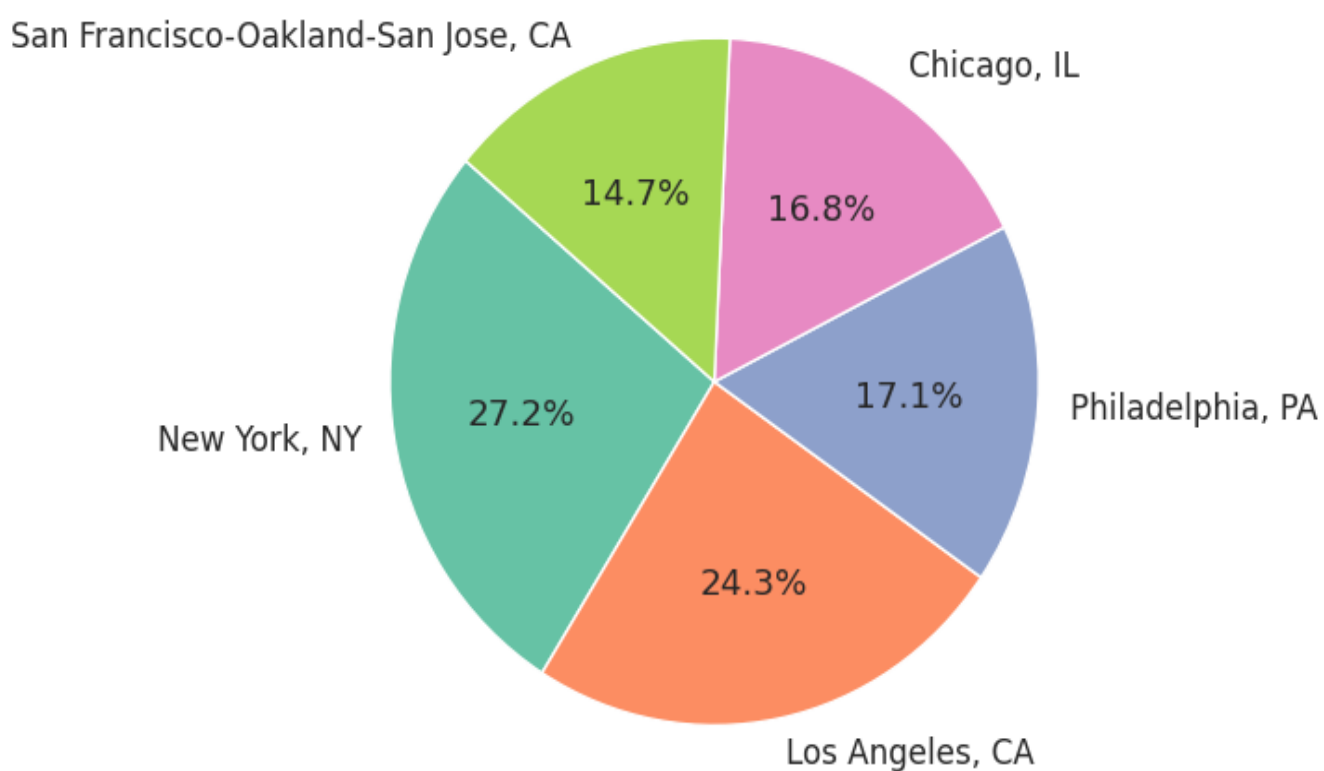


Figure 4.4.2: Top 5 Panel_DMA by Mail Volume (2022 - 2025)



4.2.4 Comparative Statistics

The dataset often used to compare businesses and years across is comprehensive. The total spending according to the loyalty program has been calculated and shown through a table. So the comparative statistics activity is very important in an organization, especially for Financial Institutions. For each year, the data is used to create a pivot table with the annual spending for each bank and then to evaluate the first three major companies that can be used as a research analysis. Not only that, the company also explores the list of top spending companies to make it easier to plan. In addition, the quarter-to-quarter (QoQ) growth rate is also calculated and aggregated month by month 2023 to illustrate the development along the timeline of a brand in terms of marketing. Through this presentation, it is an intertextual table below, it is one of the methods that is often evaluated and studied to achieve wisdom in decision-making.

Figure 4.4.3: Top Reward Programs (2022 - 2025)

Reward Program	Total Spend
Not Specified	\$225,411,378
Cashback/cash rebate	\$61,440,294
Points	\$21,237,152
Miles	\$4,176,800
Miles, Not Specified	\$2,632,050
Cashback/cash rebate, Not Specified	\$816,604
Cashback/cash rebate, Miles	\$108,424

Figure 4.4.3: Total Spend into DM Marketing by Bank

Bank Name	2022	2023	2024	2025
American Express	\$23,846,384	\$34,852,567	\$48,083,425	\$18,189,228
Bank of America	\$3,260,711	\$6,034,294	\$2,125,457	\$2,278,148
Barclays	\$452,010	\$736,702	\$975,380	\$733,604
Capital One	\$11,621,450	\$21,663,133	\$26,847,313	\$12,253,015
Chase	\$5,115,423	\$6,148,621	\$3,725,692	\$113,498
Citibank	\$23,817,428	\$32,419,623	\$6,905,240	\$2,288,395
Discover	\$52,177	\$89,438	\$108,424	\$0
U.S. Bank	\$1,930,264	\$6,972,163	\$2,075,795	\$431,067
Wells Fargo	\$186,675	\$120,658	\$3,555,245	\$5,814,049

Figure 4.4.3: Top 3 Spenders on DM Marketing by Year's

Year	Bank Name	Total Spend
2022	American Express	\$23,846,384
	Citibank	\$23,817,428
	Capital One	\$11,621,450
2023	American Express	\$34,852,567
	Citibank	\$32,419,623
	Capital One	\$21,663,134
2024	American Express	\$48,083,425
	Capital One	\$26,847,313
	Citibank	\$6,905,240
2025	American Express	\$18,189,228
	Capital One	\$12,253,016
	Wells Fargo	\$5,814,049

Figure 4.4.3: 2023 MoM Growth Summary on DM Marketing

Bank Name	Total Spend (2023)	MoM Growth (%)
American Express	\$34,852,567	+74.3%
Citibank	\$32,419,623	+70.3%
Capital One	\$21,663,134	+12.4%
U.S. Bank	\$6,972,163	+inf%
Chase	\$6,148,621	+121.3%
Bank of America	\$6,034,294	+114.2%
Barclays	\$736,702	+inf%
Wells Fargo	\$120,658	+inf%
Discover	\$89,438	+inf%

4.3 CONCLUSION

In conclusion, obtaining the data on spending behavior and mail volume of U.S. financial institution's direct mail marketing through trend analysis has shown us a number of new facts and their changes during the period. One of the findings of a monthly visualization of the money spent and the number of pieces mailed through the analysis is the seasonal pattern changes and the players who are the main drivers of most of the activity. Comparative statistics are used showing yearly discrepancies in spending, the fact that the top spender dominated the market, and their continuous performance improvement. The results of this research form a solid basis for decision-making, as it gives banking institutions the power to analyze data and performance and thereafter, allocate resources to the most profitable marketing trends.

REFERENCES

- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
- DMA (Data & Marketing Association). (2020). *Response rate report: Performance and cost metrics across direct marketing channels*. Retrieved from <https://thedma.org/>
- Forrester. (2021). *The future of direct mail in digital transformation*. Retrieved from <https://go.forrester.com/blogs>
- Media Logic. (2022). *Modern direct mail marketing for financial institutions*. Retrieved from <https://www.medialogic.com/blog/financial-services-marketing/modern-direct-mail-marketing/>
- Statista. (2023). *Advertising spending in the banking sector worldwide from 2015 to 2023*. Retrieved from <https://www.statista.com/statistics/269078/advertising-expenditure-in-the-banking-industry-worldwide/>
- BAI. (2024). *Direct mail's triumphant return to bank marketing*. Retrieved from <https://www.bai.org/banking-strategies/direct-mails-triumphant-return-to-bank-marketing/>
- eMarketer. (2024). *Banking customers enjoy direct mail marketing when it's done right*. Retrieved from <https://www.emarketer.com/content/banking-customers-enjoy-direct-mail-marketing>
- The Financial Brand. (2020). *5 Ways Financial Marketers Can Be Sure They're Using the Right Metrics*. Retrieved from <https://thefinancialbrand.com/news/bank-marketing/financial-marketers-metrics-channels-digital-media-data-attribution-91477/>
- Adobe. (n.d.). *Adobe Campaign – Direct Mail*. Retrieved from <https://business.adobe.com/products/campaign/direct-mail.html>
- Demodia. (2016, December 14). *Marketing automation review: Oracle Eloqua*. Retrieved from <https://demodia.com/articles/revenue-operations/marketing-automation-review-oracle-eloqua>
- Dun & Bradstreet. (n.d.). *Marketing Analytics*. Retrieved from <https://www.dnb.com/en-us/resources/marketing/marketing-analytics.html>
- Mintel. (n.d.). *Comperemedia Direct*. Retrieved from <https://www.mintel.com/products/comperemedia-direct/>
- Postalytics. (2018, March 5). *Announcing Salesforce Direct Mail from Postalytics*. Retrieved from <https://www.postalytics.com/blog/announcing-salesforce-direct-mail-from-postalytics/>
- Tableau Software. (n.d.). *Marketing Analytics*. Retrieved from <https://www.tableau.com/solutions/marketing-analytics>
- TechTarget. (n.d.). *What is Eloqua?* Retrieved from <https://www.techtarget.com/whatis/definition/Eloqua>
- Wikipedia contributors. (2023, April 10). *Salesforce Marketing Cloud*. In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Salesforce_Marketing_Cloud
- Wikipedia contributors. (2023, August 1). *Tableau Software*. In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Tableau_Software
- CUIInsight. (n.d.). *The pros and cons of direct mail campaigns for financial institutions*. Retrieved from <https://www.cuinsight.com/the-pros-and-cons-of-direct-mail-campaigns-for-financial-institutions/>
- KPM Group. (n.d.). *Why direct mail delivers for financial services marketing*.

- Retrieved from <https://kpmgroup.co.uk/why-direct-mail-delivers-for-financial-services-marketing/>
- BAI. (2024, October 15). *Direct mail's triumphant return to bank marketing*. BAI. <https://www.bai.org/banking-strategies/direct-mails-triumphant-return-to-bank-marketing/>
- Oracle. (2024, September 20). *12 benefits of real-time analytics for businesses*. Oracle. <https://www.oracle.com/lk/mysql/real-time-analytics-benefits/>
- ThoughtSpot. (2024, November 5). *Business analytics dashboard: Benefits, examples, and more*. ThoughtSpot. <https://www.thoughtspot.com/data-trends/dashboard/business-analytics-dashboard>
- Patterson, M., & Regnerus, B. (2024, June 26). *Considerations for applying Agile principles to analytics*. Clarkston Consulting. <https://clarkstonconsulting.com/insights/applying-agile-principles-to-analytics/>
- User Interviews. (2023, March 18). *User interviews for UX research: What, why & how*. <https://www.userinterviews.com/ux-research-field-guide-chapter/user-interviews>
- SuperSurvey. (2023). *50+ survey questions for dashboard feedback*. <https://www.supersurvey.com/LPD-dashboard-user-experience>
- Dovetail. (2024). *Ultimate prototype usability testing guide*. <https://dovetail.com/ux/prototype-usability-testing-guide/>
- Oberoi, K. (2024, May 22). *Triangulation in qualitative research: A comprehensive guide*. LoopPanel. <https://www.looppanel.com/blog/triangulation-in-qualitative-research>
- Deepsync. (2025, January 20). *What is third-party data? The ultimate guide for businesses*. <https://deepsync.com/what-is-third-party-data/>
- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Thomas, D. (2001). *Manifesto for Agile Software Development*. Retrieved from <https://agilemanifesto.org>