Case Study Report



**Tech Saksham**

Data Analytics with Power BI

**“Real-Time Analysis of Bank Customers”**

**“College Name”**

|  |  |
| --- | --- |
| **NM ID** | **NAME** |
| EBE8E703D42724EB7EC40D71B598B891 | S.Shabrintaj |

|  |  |
| --- | --- |
|  |  |
|  | Trainer Name : R UMAMAHESWARI |
|  | Master Trainer: R UMAMAHESWARI |

**ABSTRACT**

The real-time analysis of bank customers is a critical aspect of modern banking operations aimed at enhancing customer satisfaction , preventing fraud and optimizing financial services. This study explores the methodologies and technologies utilized in real-time analysis of bank customers data to extract valuable insights and improve decision-making processes. Various data sources, including transactional data, demographic information, and digital interactions, and integrated and analyzed in real-time to identify patterns, trends, and anomalies. Machine learning algorithms, such as clustering, classification, and anomaly detection, play a crucial role in analyzing vast volumes of data quickly and accurately. Furthermore, real-time analysis enables banks to personalize customer experiences, detect fraudulent activities promptly, and offer timely recommendations for financial products and services. This paper reviews the challenges, opportunities, and ethical considerations associated with real-time analysis of bank customers data and highlights the significance of continuous innovation and adaptation in the dynamic landscape of banking and finance.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Table of Contents** | **Page No.** |
| 1 | Chapter 1: Introduction | 3 |
| 2 | Chapter 2: Services and Tools Required | 7 |
| 3 | Chapter 3: Project Architecture | 9 |
| 4 | Chapter 4: Modeling and Result | 10 |
| 5 | Conclusion | 21 |
| 6 | Future Scope | 22 |
| 7 | References | 23 |
| 8 | Links | 23 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **Problem Statement**

The real-time analysis of bank customers faces several significant challenges that hinder its successful implementation and utilization:

1. Data Integration and Quality: Banks often have diverse data sources spread across multiple systems, leading to challenges in integrating data in real-time. Furthermore, ensuring the quality and accuracy of real-time data streams remains a persistent issue, impacting the reliability of insights derived from analysis.
2. Scalability and Performance: Real-time analysis requires robust infrastructure capable of processing large volumes of data rapidly. Many banks struggle with scalability and performance issues when attempting to analyze real-time data streams, leading to delays and inefficiencies in decision-making processes.
3. Security and Compliance: Real-time analysis involves the continuous processing of sensitive customer data, raising concerns about data security and privacy. Banks must ensure that real-time analytics platforms adhere to stringent security protocols and regulatory requirements to protect customer information and maintain compliance.
4. Skill Gap and Training: Real-time analysis requires specialized skills in data analytics, machine learning, and data engineering. Many banks lack the necessary talent and expertise to effectively implement and manage real-time analytics initiatives. Additionally, ongoing training and upskilling efforts are necessary to keep pace with evolving technologies and methodologies.
5. Cost and Resource Constraints: Implementing real-time analytics solutions incurs significant costs related to infrastructure, software licensing, and talent acquisition. Many banks face budgetary constraints and resource limitations, making it challenging to invest in the necessary resources and technologies to support real-time analysis effectively.
   1. **Proposed Solution**

The proposed solution to enable real-time analysis of bank customers comprises the following key components:

1. Unified Data Platform: Implement a unified data platform that integrated disparate data sources, including transactional data, customer interactions, and external data sources. This platform should support real-time data ingestion, processing, and analytics, providing a single source of truth for customer insights.
2. Advanced Analytics Capabilities: Leverage advanced analytics techniques such as machine learning, predictive modeling, and sentiment analysis to derive actionable insights from real-time data streams. Develop models to predict customer behavior, identify trends, and personalize services in real-time.
3. Real-Time Data Processing: Deploy scalable and high-performance data processing infrastructure capable of handling large volumes of data in real-time. Utilize technologies such as stream processing frameworks and in-memory databases to enable rapid data processing and analysis.
4. Dynamic Visualization and Reporting: Utilize visualization tools such as Power BI or Tableau to create dynamic dashboards and reports that provide real-time insights into customer behavior and key performance indicators. Enable stakeholders to monitor trends, detect anomalies, and make data-driven decisions in real-time.
5. Data Security and Compliance: Implement robust security measures and data governance frameworks to ensure the confidentiality, integrity, and availability of customer data. Adhere to industry regulations such as GDPR, CCPA, and PCI DSS to protect customer privacy and maintain compliance with legal requirements.
6. Talent Development and Training: Invest in talent development initiatives to build a skilled workforce capable of leveraging real-time analytics technologies effectively. Provide training programs and certification courses to empower employees with the knowledge and skills required for real-time data analysis.
   1. **Feature**

**Real-time Data Streaming:** The capability to ingest and process streaming data from various sources in real-time, including transactional data, online interactions, social media, and customer service channels.

**Predictive Analytics:** Utilization of predictive modeling techniques to forecast customer behavior, such as predicting account closures, identifying potential churn, and anticipating product preferences.

**Dynamic Customer Segmentation:** Automatic segmentation of customers based on real-time behavioral patterns, demographics, and transaction history to tailor offerings and communications effectively.

**Anomaly Detection:** Detection of unusual or suspicious activities in real-time, such as fraudulent transactions or account access, enabling immediate intervention and mitigation measures.

**Personalized Recommendations:** Generation of personalized product recommendations and marketing offers in real-time based on individual customer profiles and preferences.

**Customer Sentiment Analysis:** Analysis of customer feedback, sentiment, and social media interactions in real-time to gauge customer satisfaction levels and identify potential issues or concerns.

**Real-time Dashboards and Alerts:** Creation of interactive dashboards and alerts that provide real-time visibility into key performance metrics, allowing stakeholders to monitor trends and respond promptly to changing conditions.

**Automation and Workflow Integration:** Integration with automated workflows and backend systems to enable seamless execution of real-time decisions, such as automated fraud prevention measures or personalized customer communications.

* 1. **Advantages**

**Immediate Insights:** Real-time analysis provides banks with immediate insights into customer behavior, preferences, and interactions. This allows banks to respond promptly to changing customer needs, deliver personalized services, and capitalize on emerging opportunities.

**Enhanced Customer Engagement:** By understanding customer behavior in real-time, banks can engage with customers more effectively. Personalized offers, targeted marketing campaigns, and proactive customer support based on real-time insights help foster stronger relationships and increase customer loyalty.

**Fraud Detection and Prevention:** Real-time analysis enables banks to detect and prevent fraudulent activities as they occur. By monitoring transactional data and detecting anomalies in real-time, banks can implement immediate fraud prevention measures, protecting both customers and the institution from financial losses.

**Optimized Cross-Selling and Upselling:** Real-time analysis allows banks to identify cross-selling and upselling opportunities in real-time. By analyzing customer transactions and preferences in real-time, banks can offer relevant products and services to customers at the right moment, increasing revenue and customer satisfaction.

**Operational Efficiency:** Real-time analysis streamlines banking operations by automating decision-making processes and reducing manual interventions. This leads to improved efficiency, reduced processing times, and lower operational costs for the bank.

**Risk Management:** Real-time analysis helps banks better manage risks by identifying and mitigating potential risks as they emerge. By monitoring customer behavior, market trends, and external factors in real-time, banks can proactively manage risks and ensure regulatory compliance.

**Competitive Advantage:** Real-time analysis provides banks with a competitive advantage in the market. Banks that can effectively leverage real-time insights to deliver superior customer experiences, innovate new products and services, and respond quickly to market changes are better positioned to outperform competitors and capture market share.

* 1. **Scope**

**Data Collection and Integration:** The scope begins with the collection and integration of data from various sources, including transactional data, customer interactions, social media, and external sources. This involves establishing data pipelines and integrating disparate data sources to create a unified view of customer information.

**Real-Time Data Processing:** Real-time analysis requires the capability to process streaming data rapidly and efficiently. The scope includes implementing technologies such as stream processing frameworks and in-memory databases to enable real-time data processing and analysis.

**Predictive Analytics:** Predictive analytics plays a crucial role in real-time analysis by forecasting customer behavior and trends. The scope includes developing predictive models to identify patterns, predict outcomes, and personalize offerings in real-time based on customer data.

**Dynamic Customer Segmentation:** Real-time analysis enables dynamic customer segmentation based on real-time behavioral patterns, demographics, and transaction history. The scope encompasses the automatic segmentation of customers to tailor offerings and communications effectively.

**Anomaly Detection and Fraud Prevention:** An essential aspect of real-time analysis is the detection of anomalies and fraudulent activities as they occur. The scope includes implementing algorithms and techniques for real-time anomaly detection and fraud prevention to safeguard customer assets and maintain trust.

**Personalized Recommendations:** Real-time analysis enables the generation of personalized product recommendations and marketing offers in real-time. The scope involves leveraging customer data to offer relevant and timely recommendations, improving customer satisfaction and engagement.

**Integration with Business Processes:** Real-time analysis should be seamlessly integrated with existing business processes and workflows. The scope includes integrating real-time insights into decision-making processes, customer interactions, and operational workflows to drive business outcomes effectively.

**CHAPTER 2**

**SERVICES AND TOOLS REQUIRED**

**2.1 Services Used**

**Streaming Data Platforms:** Services like Apache Kafka, Amazon Kinesis, or Azure Stream Analytics are used for ingesting and processing large volumes of streaming data from various sources, ensuring real-time data availability for analysis.

**Cloud Data Warehouses:** Cloud-based data warehousing solutions such as Amazon Redshift, Google BigQuery, or Azure Synapse Analytics provide scalable storage and processing capabilities for analyzing structured and semi-structured data in real-time.

**Real-time Analytics Engines:** Technologies like Apache Flink, Apache Spark Streaming, or Azure Stream Analytics are employed for performing real-time analytics on streaming data, enabling the derivation of insights and actionable intelligence in near real-time.

**Machine Learning Services:** Platforms such as Amazon SageMaker, Google Cloud AI Platform, or Azure Machine Learning facilitate the development and deployment of machine learning models for predictive analytics, anomaly detection, and personalized recommendations based on real-time customer data.

**Visualization and Dashboarding Tools:** Visualization tools like Tableau, Power BI, or Google Data Studio are utilized for creating interactive dashboards and visualizations that enable stakeholders to monitor key metrics and trends in real-time.

**Data Integration Platforms:** Integration platforms such as Apache Nifi, Talend, or Informatica enable the seamless integration of data from disparate sources, ensuring data consistency and accuracy for real-time analysis.

**2.2 Tools and Software used**

**Tools**:

* **PowerBI**: The main tool for this project is PowerBI, which will be used to create interactive dashboards for real-time data visualization.
* **Power Query**: This is a data connection technology that enables you to discover, connect, combine, and refine data across a wide variety of sources.

**Software Requirements**:

* **PowerBI Desktop**: This is a Windows application that you can use to create reports and publish them to PowerBI.
* **PowerBI Service**: This is an online SaaS (Software as a Service) service that you use to publish reports, create new dashboards, and share insights.
* **PowerBI Mobile**: This is a mobile application that you can use to access your reports and dashboards on the go.

**CHAPTER 3**

**PROJECT ARCHITECTURE**

**3.1 Architecture**

**USER FRONTEND BACKEND**

|  |  |  |
| --- | --- | --- |
|  | **HTML 5** | **NODEJS 14.0**  **Database** |

Here’s a high-level architecture for the project:

1. **Data Collection**: Real-time customer data is collected from various sources like bank transactions, customer interactions, etc. This could be achieved using services like Azure Event Hubs or AWS Kinesis.
2. **Data Storage**: The collected data is stored in a database for processing. Azure SQL Database or AWS RDS can be used for this purpose.
3. **Data Processing**: The stored data is processed in real-time using services like Azure Stream Analytics or AWS Kinesis Data Analytics.
4. **Machine Learning**: Predictive models are built based on processed data using Azure Machine Learning or AWS SageMaker. These models can help in predicting customer behavior, detecting fraud, etc.
5. **Data Visualization**: The processed data and the results from the predictive models are visualized in real-time using PowerBI. PowerBI allows you to create interactive dashboards that can provide valuable insights into the data.
6. **Data Access**: The dashboards created in PowerBI can be accessed through PowerBI Desktop, PowerBI Service (online), and PowerBI Mobile.

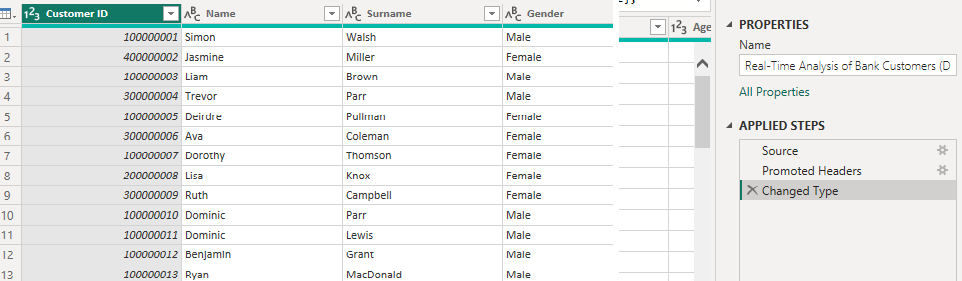
This architecture provides a comprehensive solution for real-time analysis of bank customers. However, it’s important to note that the specific architecture may vary depending on the bank’s existing infrastructure, specific requirements, and budget. It’s also important to ensure that all tools and services comply with relevant data privacy and security regulations.

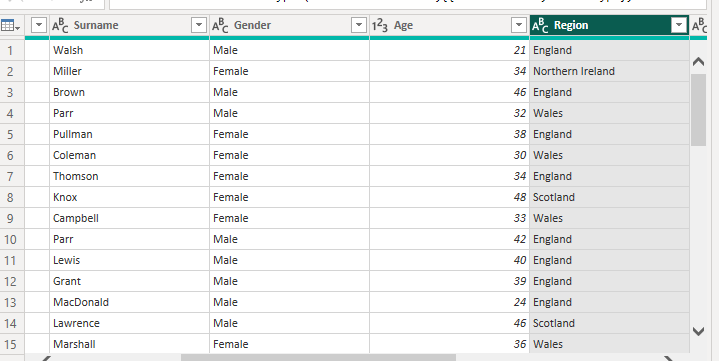
**CHAPTER 4**

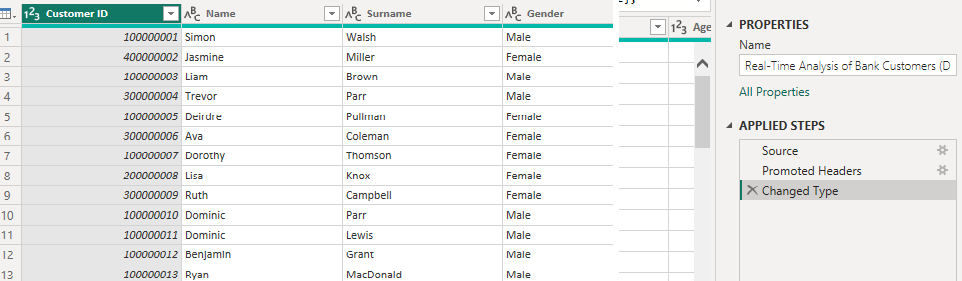
**MODELING AND RESULT**

**Manage relationship**

The “disp” file will be used as the main connector as it contains most key identifier (account id, client id and disp id) which can be use to relates the 8 data files together. The “district” file is use to link the client profile geographically with “district id”





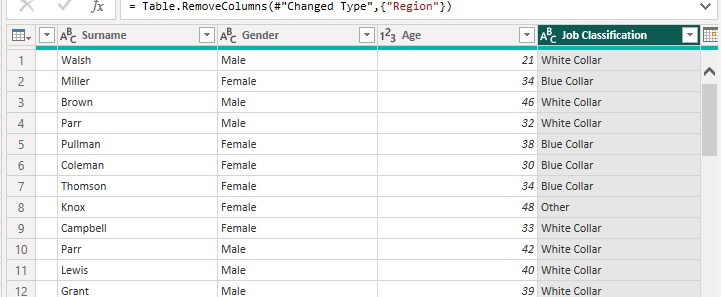


**Modelling for Gender and Age data**

Notice that the Gender and age of the client are missing from the data. These can be formulated from the birth number YYMMDD where at months (the 3rd and 4th digits) greater than 50 means that client is a Female. We can create a column for Gender.



For birthday, we need to reduce the birth month of the female by 50 and then change the date format to DD/MM/YYYY adding 1900 to the year.

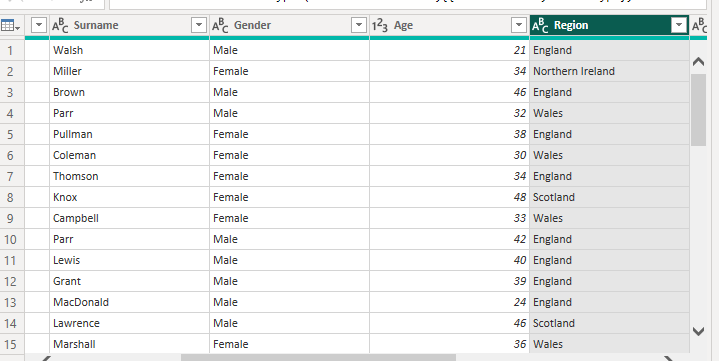


For Age, we shall assume it is year 1999 as explain previously and use it to minus from the birth year.



**Replacing values**

Set some fields to English for easy understanding, we replace values to English with the Power Query Editor.



Changing the order of Region name at Power Query

Duplicate the “district /region” then split column using space as delimiter.



Then merge column by Region and direction. Refer to applied steps for details.



**Grouping of age by ranges**

As the customers’ age ranges from 12 to 88, we shall group them into different generation age range for easier profiling, we will group the ages into 5 groups.

The Gen Y are youths,

Gen X are young working adults, some starting their families

Baby Boomer are working adults with families.

The silent Generations some are working and retired, living on pensions.

The greatest Generation, retired elderly living on pensions.



**Credit Rating and Loan Status**

As the Loan status uses A, B, C, D which are not reader friendly. We can add a column to represent what it stands for, we also simplify the classification of those with late or default on payment as bad credit, refer to the table below for details on the new columns added.



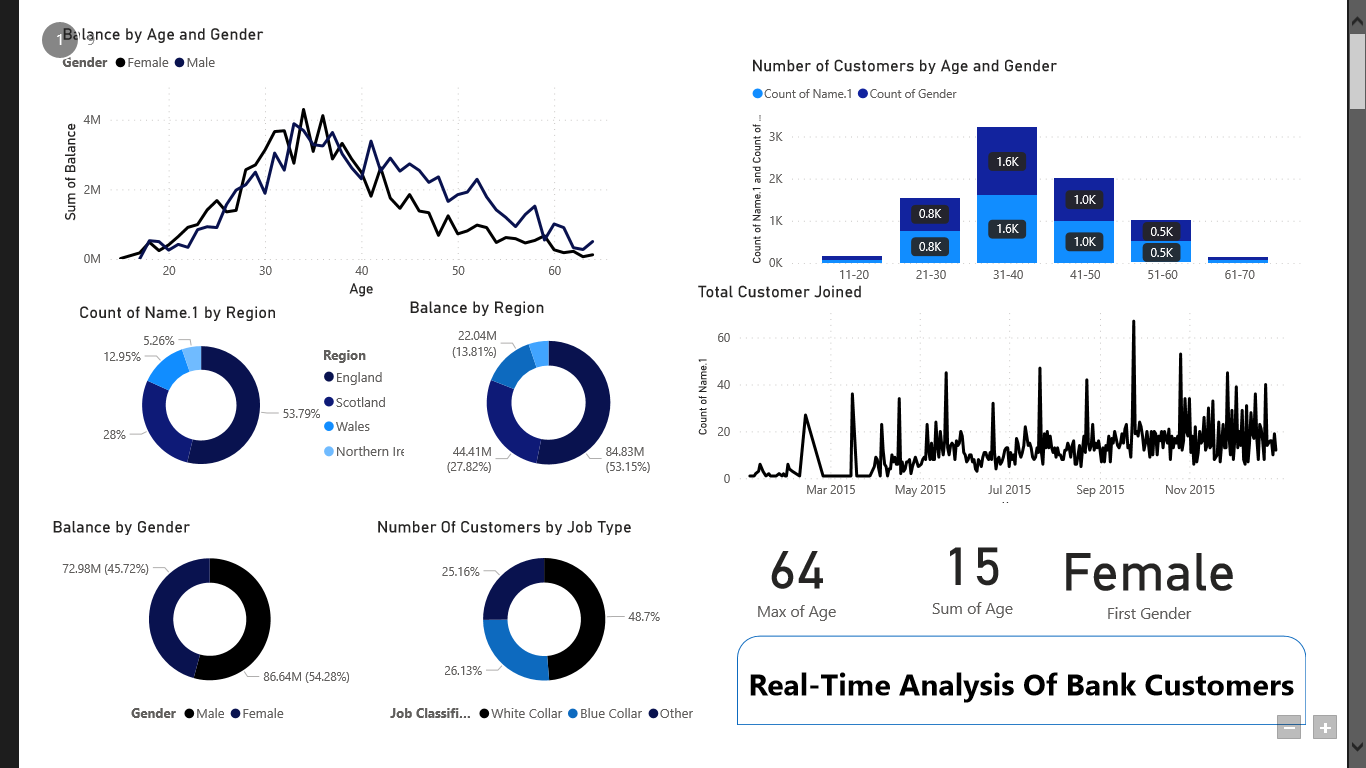
Values of such as “account Id” have also been set as Text.

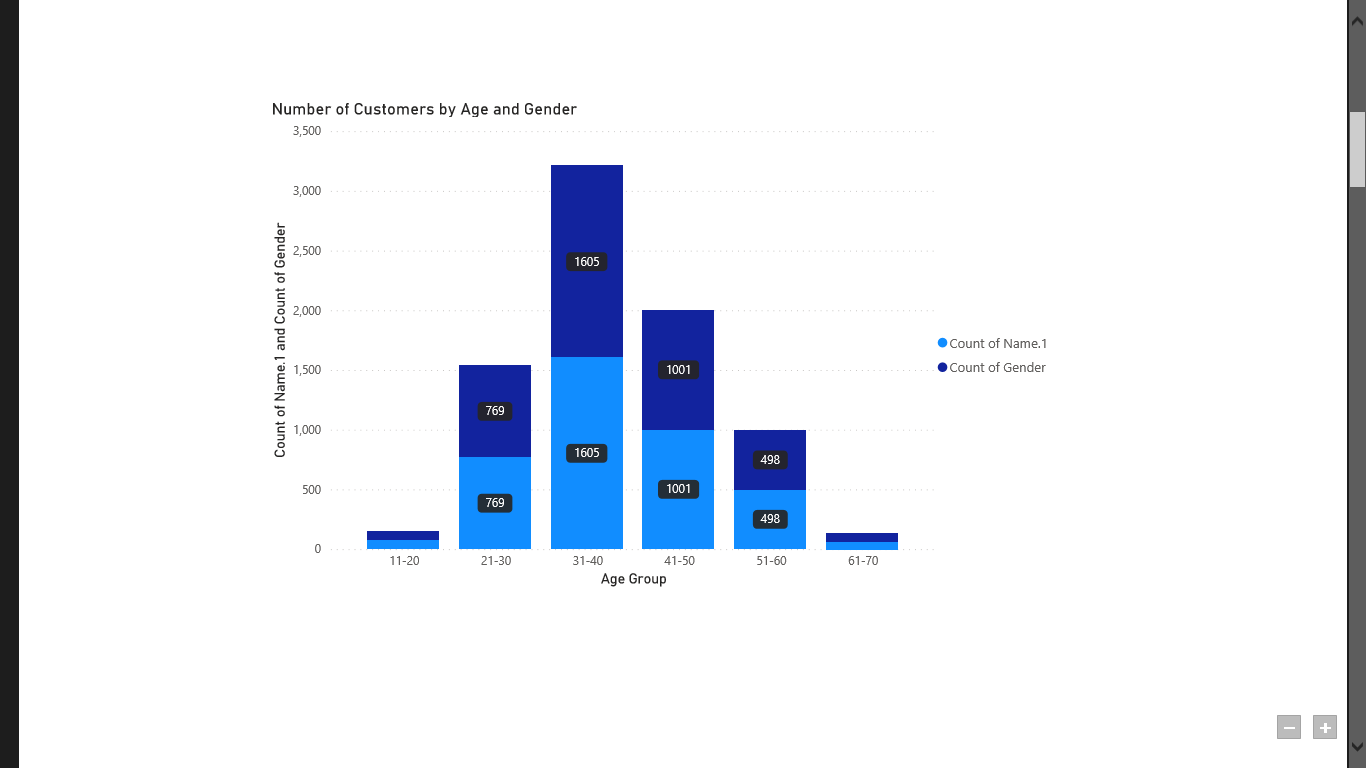
And District name have been categorized as place to be use for the map to show the sum of the inhabitants in each region.

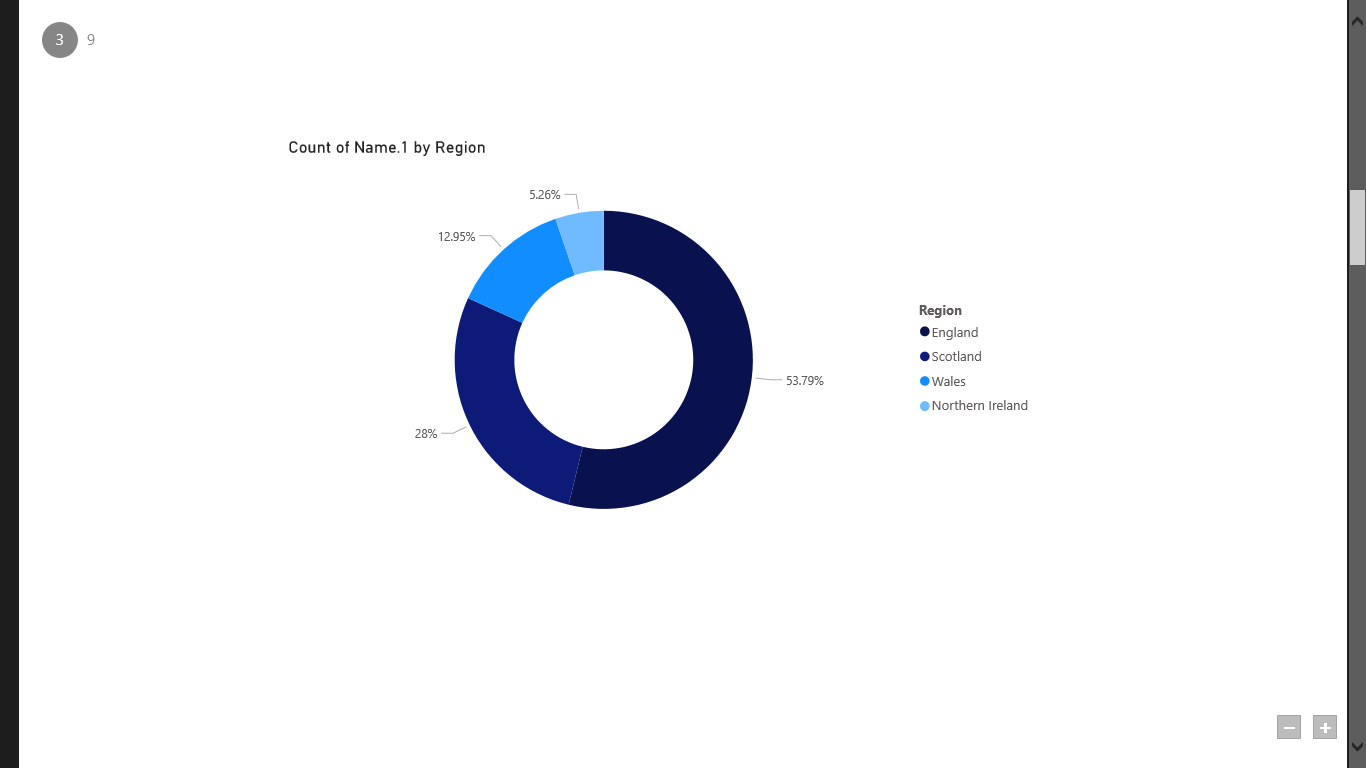
**Dashboard**

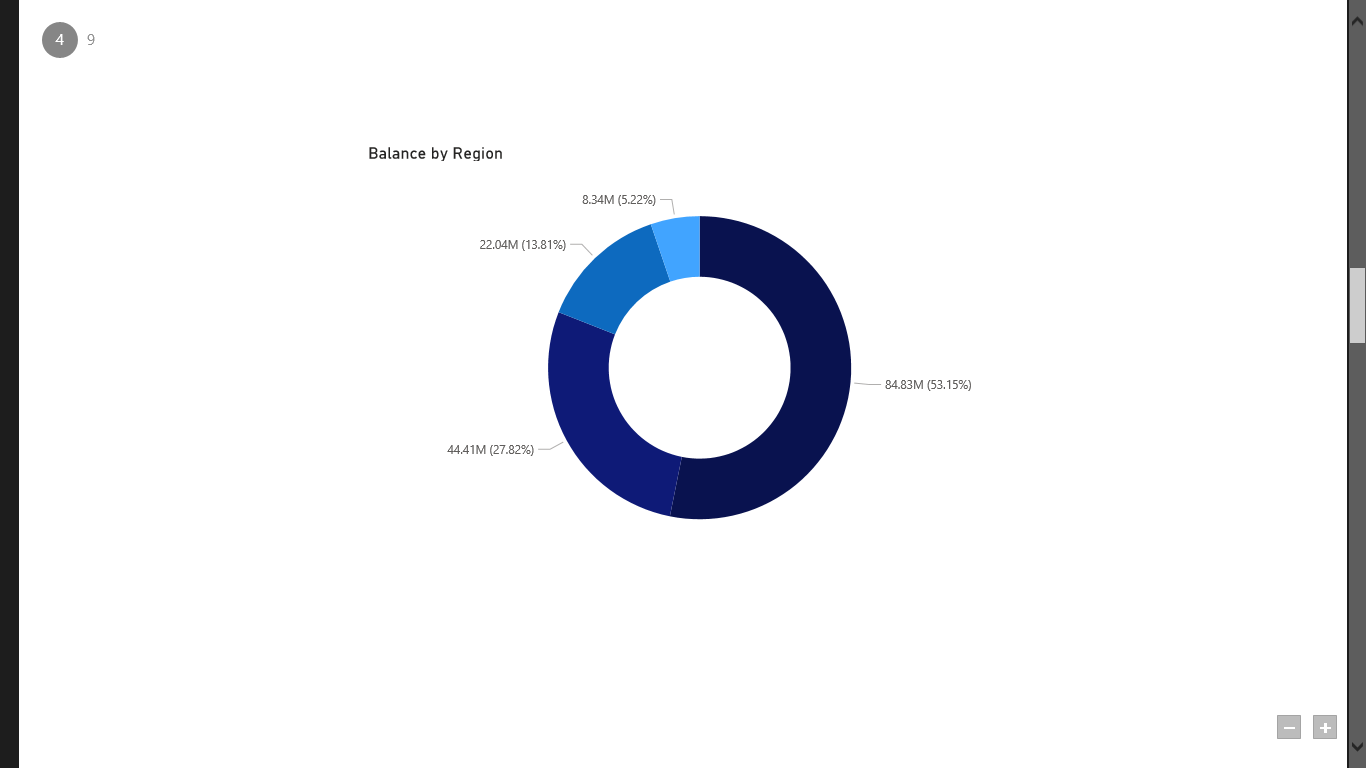
A screenshot of a credit card

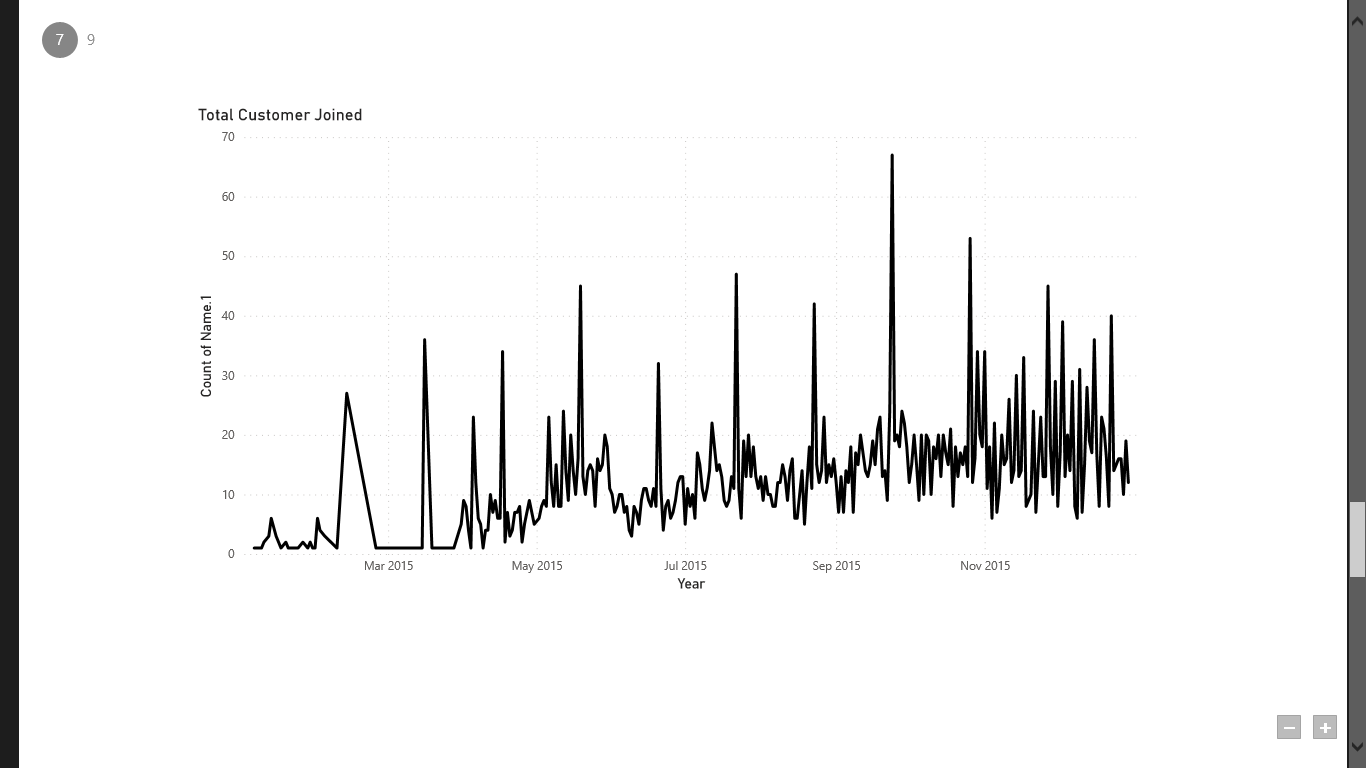
Description automatically generated

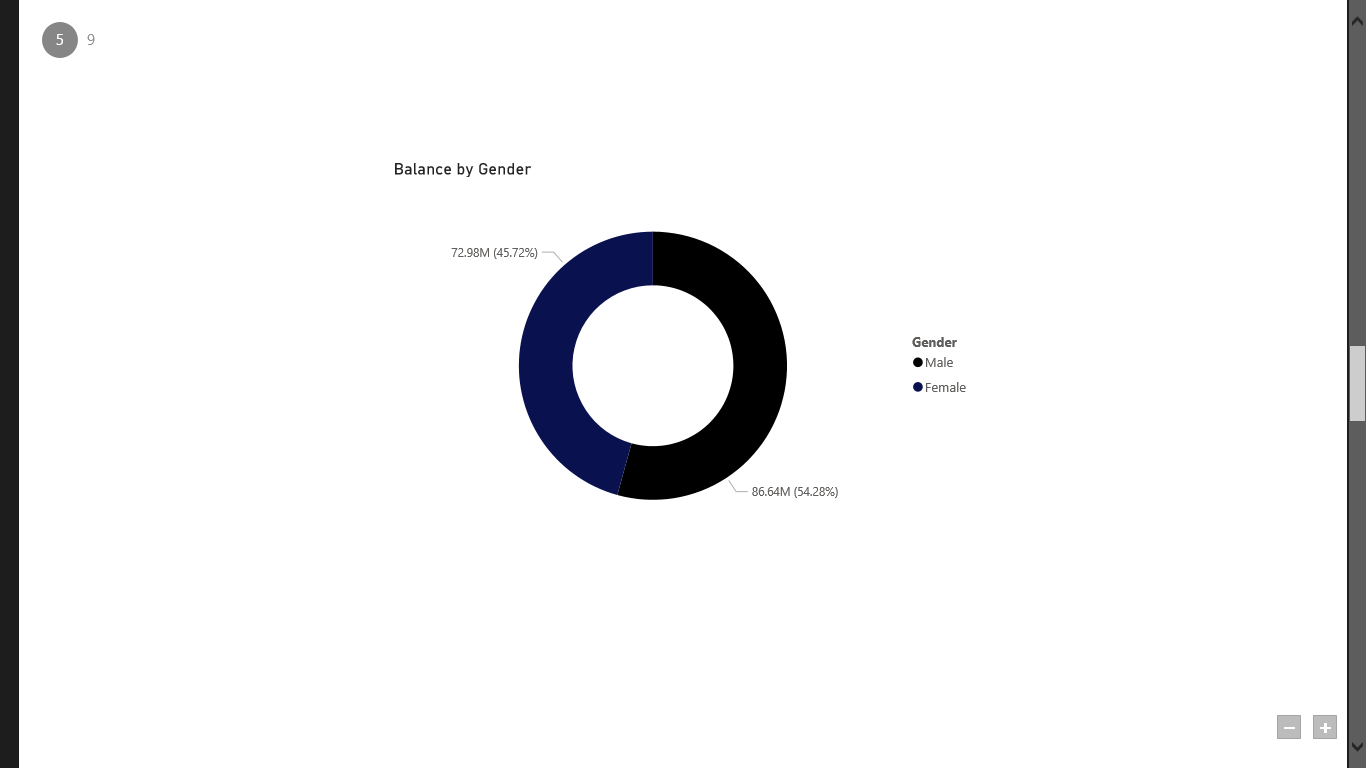


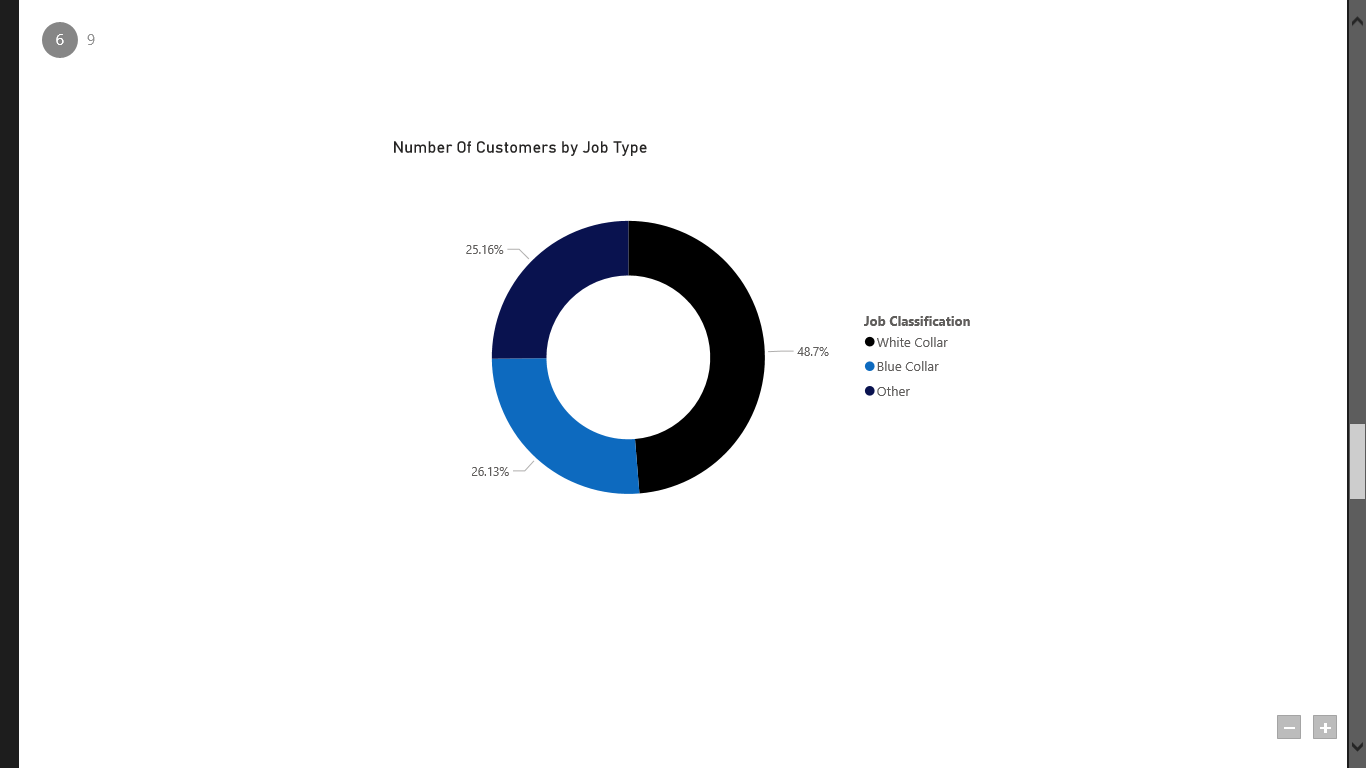


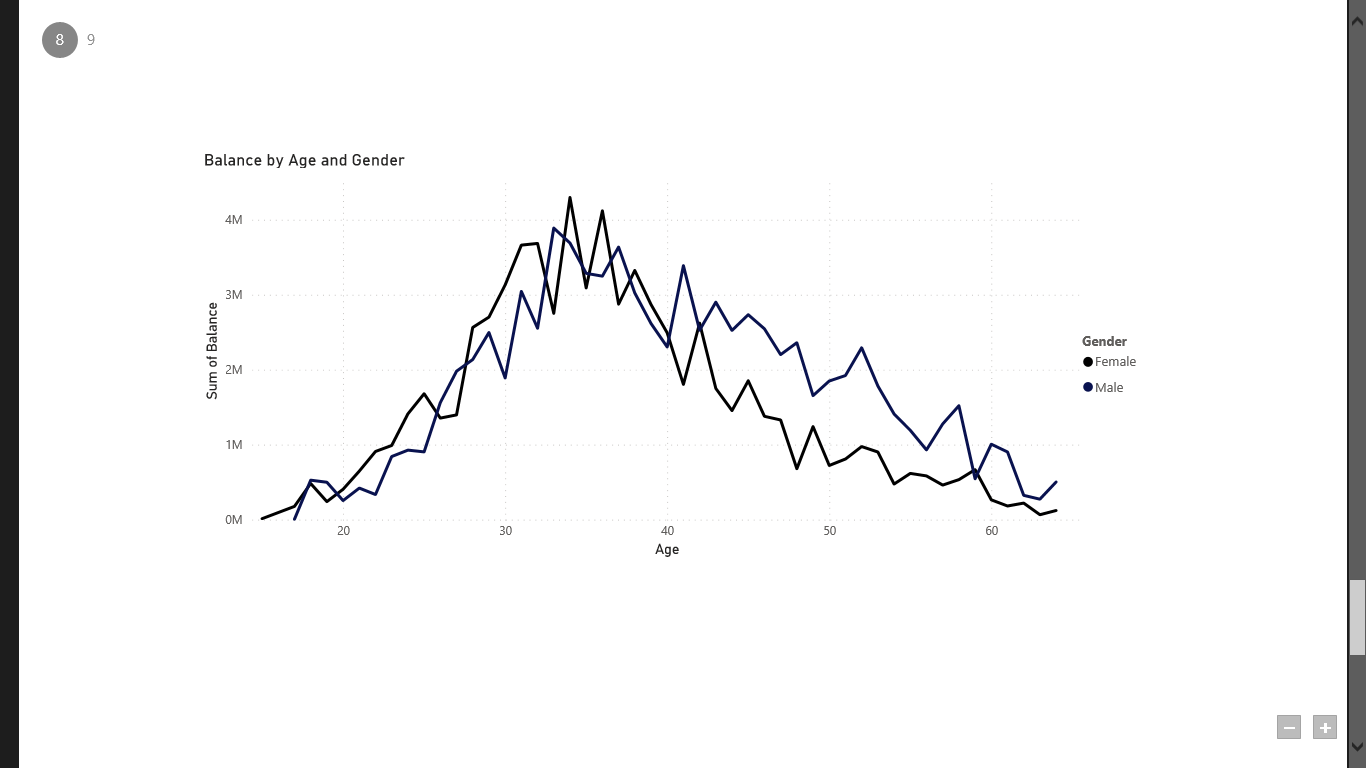


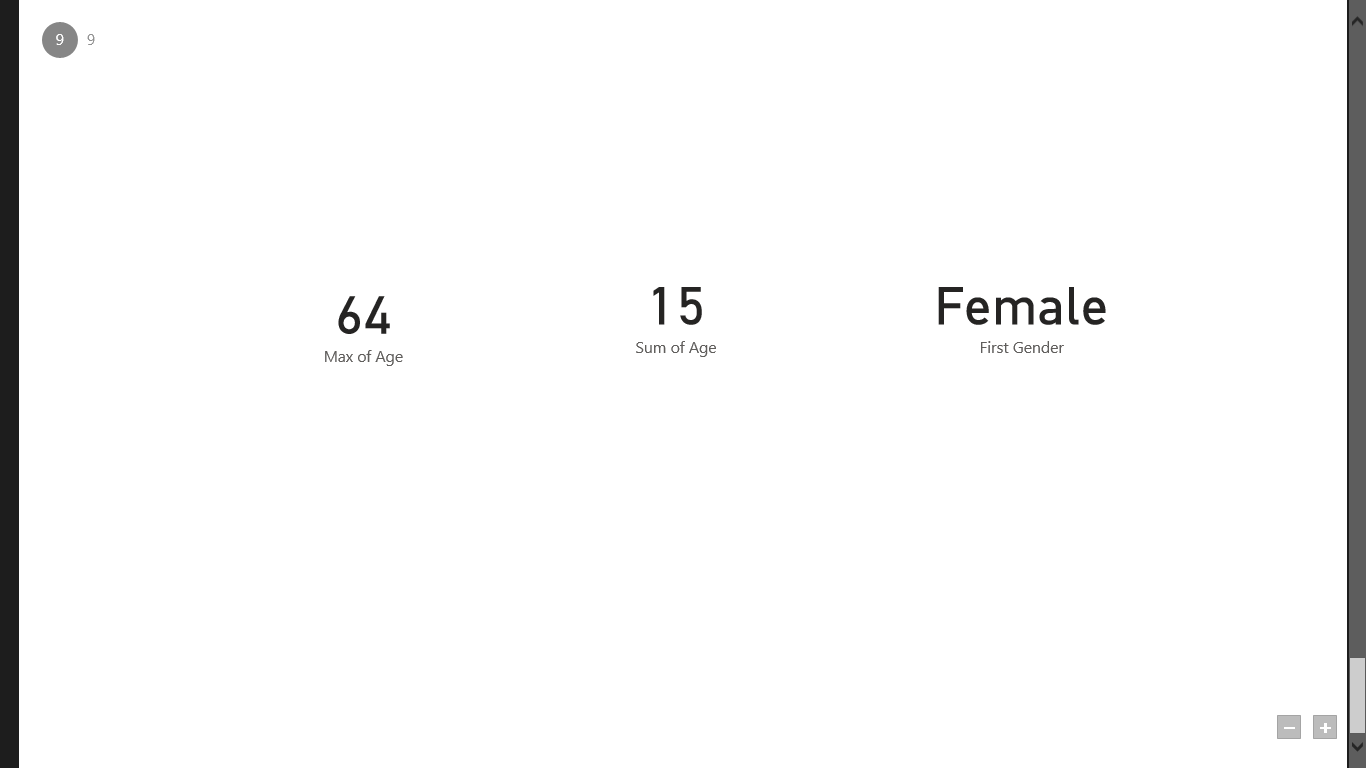
****

****

****

****

****

****

**CONCLUSION**

The project “Real-Time Analysis of Bank Customers” using PowerBI has successfully demonstrated the potential of data analytics in the banking sector. The real-time analysis of customer data has provided valuable insights into customer behavior, preferences, and trends, thereby facilitating informed decision-making. The interactive dashboards and reports have offered a comprehensive view of customer data, enabling the identification of patterns and correlations. This has not only improved the efficiency of data analysis but also enhanced the bank’s ability to provide personalized services to its customers. The project has also highlighted the importance of data visualization in making complex data more understandable and accessible. The use of PowerBI has made it possible to present data in a visually appealing and easy-to-understand format, thereby aiding in better decision-making.

In conclusion, leveraging Power BI for real-time analysis of bank customers offers a transformative approach to understanding and serving customers in the banking industry. Power BI's robust capabilities empower banks to gather, process, analyze, and visualize customer data streams in real-time, enabling them to derive actionable insights and make informed decisions promptly.

**FUTURE SCOPE**

The future scope of real-time analysis of bank customers is vast and promising. Here are several potential areas of growth and development:

**Enhanced Personalization:** Real-time analysis allows banks to understand their customers' behavior, preferences, and needs instantly. This data can be leveraged to offer highly personalized services and product recommendations, leading to increased customer satisfaction and loyalty.

**Fraud Detection and Prevention:** Real-time analysis enables banks to detect and prevent fraudulent activities as they occur. By continuously monitoring transactions and customer behavior patterns, banks can quickly identify suspicious activities and take immediate action to mitigate risks.

**Risk Management:** Real-time analysis helps banks assess and manage various types of risks, including credit risk, market risk, and operational risk. By analyzing real-time data on market trends, economic indicators, and customer behavior, banks can make better-informed decisions and optimize their risk management strategies.

**Customer Experience Optimization:** Real-time analysis allows banks to monitor and improve the overall customer experience. By analyzing customer feedback, website interactions, and service usage patterns in real-time, banks can identify areas for improvement and implement changes to enhance the customer experience.

**Predictive Analytics:** Real-time analysis combined with predictive analytics can help banks anticipate customer needs and preferences. By analyzing historical data and real-time interactions, banks can identify trends and patterns that enable them to offer proactive services and personalized recommendations.

**Operational Efficiency:** Real-time analysis can help banks streamline their operations and improve efficiency. By analyzing real-time data on transaction processing, customer service interactions, and resource utilization, banks can identify bottlenecks and inefficiencies and take corrective actions in real-time.

**REFERENCES**

<https://medium.com/analytics-vidhya/analysis-of-bank-customers-using-dashboard-in-power-bi-a366f2b3e563>

**LINK**

<https://github.com/githubtraining/hellogitworld.git>