[Exploratory Data Analysis 1](#_Toc1721865655)

[Project Description: 1](#_Toc1880589550)

[Project Title: 1](#_Toc713711293)

[Objective: 1](#_Toc84488749)

[Dataset: 1](#_Toc1834322647)

[Methodology: 1](#_Toc1726964005)

[Tools and Technologies: 2](#_Toc875646832)

[Deliverables: 2](#_Toc1114125466)

[Descriptive Analysis 2](#_Toc698744530)

[Project Description: 3](#_Toc1988885159)

[Project Title: 3](#_Toc835655118)

[Objective: 3](#_Toc1011477313)

[Dataset: 3](#_Toc1773125350)

[Methodology: 3](#_Toc606026469)

[Tools and Technologies: 4](#_Toc865607481)

[Deliverables: 4](#_Toc179783385)

[Diagnostic Analysis 4](#_Toc1866219081)

[Project Description: Diagnostic Analysis of Sales Decline at XYZ Retail 4](#_Toc1281811206)

[Project Title: 4](#_Toc25639637)

[Objective: 4](#_Toc1856799866)

[Background: 5](#_Toc1455014181)

[Dataset: 5](#_Toc945236200)

[Methodology: 5](#_Toc983995282)

[Tools and Technologies: 6](#_Toc602443488)

[Deliverables: 6](#_Toc939073391)

[Timeline: 6](#_Toc410649154)

[Data Wrangling 6](#_Toc134895031)

[Project Description: Data Wrangling for Customer Analytics at XYZ Company 6](#_Toc583696924)

[Project Title: 6](#_Toc1532070626)

[Objective: 6](#_Toc653051912)

[Background: 7](#_Toc868241672)

[Dataset: 7](#_Toc1447253766)

[Methodology: 7](#_Toc1392815005)

[Tools and Technologies: 8](#_Toc169127571)

[Deliverables: 8](#_Toc1374045486)

[Timeline: 8](#_Toc519665074)

[Data Quality Control 9](#_Toc32734904)

[Project Description: Data Quality Control Initiative at ABC Enterprises 9](#_Toc1707066452)

[Project Title: 9](#_Toc20165743)

[Objective: 9](#_Toc1093595884)

[Background: 9](#_Toc1032443909)

[Scope: 9](#_Toc821476858)

[Methodology: 9](#_Toc831223712)

[Deliverables: 10](#_Toc1897418703)

[Timeline: 11](#_Toc1039642674)

# Exploratory Data Analysis

## Project Title:

Exploratory Data Analysis of Vancouver City Council Voting Records Using AWS Cloud Computing

## Project Description:

This project aims to perform an in-depth exploratory data analysis (EDA) of the voting records of the Vancouver City Council. The goal is to uncover trends, patterns, and insights into council voting behaviors, member participation, and decision-making processes. By leveraging the scalability and computational power of Amazon Web Services (AWS), this project will enable secure, efficient, and real-time analysis of large datasets, helping stakeholders understand key factors influencing council decisions.

## Objective:

* To analyze historical voting records of the Vancouver City Council.
* To identify voting patterns, trends, and key influencers in decision-making.
* To visualize voting behaviors and how they correlate with council members' roles, political affiliations, and agendas.
* To implement AWS cloud services for scalability, reliability, and cost-efficiency.
* To deliver a secure, reliable solution for public and private stakeholders to access these insights.

## Datasets:

* **Voting Records**: Historical voting data from the Vancouver City Council, including information on council members, votes, decisions, dates, and motion details.
* **Council Member Information**: Data on council members (e.g., name, party affiliation, tenure, and voting history).
* **Meeting Agendas**: Information related to council meetings, including topics, resolutions, and voting items.
* **Other Datasets (optional)**: Demographic data, election results, or policy impact studies to enhance analysis.

## Methodology:

1. **Data Collection & Preparation**:
   * Collect historical voting data and council member details.
   * Clean and preprocess the datasets (handling missing data, standardizing formats).
   * Use AWS S3 for storage and AWS Glue for data preparation.
2. **Exploratory Data Analysis**:
   * Perform data visualization (e.g., vote distributions, trends over time, voting by party) using AWS QuickSight or Python libraries like Matplotlib and Seaborn.
   * Analyze voting patterns and correlations (e.g., clustering council members by voting similarity).
   * Use statistical analysis techniques to explore relationships between variables.
3. **Data Visualization**:
   * Present results using AWS QuickSight dashboards and visualizations.
   * Provide public and private views with different levels of access.
4. **Cloud Infrastructure Setup**:
   * Implement AWS services for storage, computation, and visualization.
   * Use AWS Lambda and AWS Glue for data processing pipelines.
   * Ensure security using AWS Identity and Access Management (IAM) for user access control.
   * Monitor resources and performance using AWS CloudWatch.

## Tools and Technologies:

* **AWS S3**: For scalable storage of large datasets.
* **AWS Glue**: For data extraction, transformation, and loading (ETL) processes.
* **AWS Lambda**: To run serverless data processing jobs.
* **Amazon QuickSight**: For creating interactive data visualizations.
* **Amazon RDS (or Aurora)**: For structured data storage (voting records).
* **AWS CloudWatch**: For performance monitoring.
* **Python (Pandas, NumPy, Matplotlib, Seaborn)**: For data analysis and visualization.
* **AWS IAM**: For access management and security.

## Deliverables:

* **Exploratory Data Analysis Report**: Detailed analysis report of the voting records, including key trends and insights.
* **Interactive Dashboards**: AWS QuickSight-based dashboards for visualizing voting data, accessible to various stakeholders.
* **AWS Infrastructure Setup**: A well-documented, cost-optimized AWS solution for managing and analyzing voting records.
* **Security and Access Control**: IAM roles and policies to ensure secure access to the data and dashboards.
* **Scalability Plan**: A plan outlining how the solution can be scaled as new voting records are added.

# Descriptive Analysis

## Project Title:

Exploratory Data Analysis of Vancouver City Council Voting Records Using AWS Cloud Computing

## Project Description:

This project focuses on exploring the voting records of the Vancouver City Council to uncover trends, patterns, and key insights. By utilizing AWS cloud computing services, this project aims to efficiently process, store, and analyze large datasets, providing an in-depth understanding of how council members have voted on various motions over time. The analysis will help stakeholders, including researchers, public officials, and the general public, gain insights into decision-making patterns, member participation, and political alignments. The AWS infrastructure will ensure scalability, reliability, and security while offering interactive visualizations for a wide range of users.

## Objective:

* To conduct an exploratory analysis of voting records from the Vancouver City Council.
* To identify patterns, trends, and key insights regarding voting behavior and decision-making.
* To implement AWS cloud services for scalable, efficient data processing and analysis.
* To create interactive visualizations and reports for stakeholders.

## Dataset:

* **Voting Records Dataset**: Historical voting data from the Vancouver City Council, including details such as vote outcomes, motion topics, dates, and individual votes cast by council members.
* **Council Member Information**: Data about council members, including their names, roles, political affiliations, and participation in voting sessions.
* **Motion Metadata**: Information about the types of motions presented (e.g., policy changes, budget approvals) and their outcomes.

## Methodology:

1. **Data Collection**:
   * Gather historical voting data and council member information from public records or available APIs.
   * Store raw datasets in **AWS S3** for scalable storage.
2. **Data Cleaning and Preprocessing**:
   * Use **AWS Glue** to clean and preprocess the data (handling missing values, data standardization).
   * Format datasets to ensure consistency and readiness for analysis.
3. **Exploratory Data Analysis**:
   * Perform descriptive statistics to summarize the voting behavior (e.g., vote counts, most common motion types, participation rates).
   * Use clustering or correlation techniques to identify voting patterns, alliances, and common decision-making trends.
   * Investigate time-based trends (e.g., voting patterns over months or years).
4. **Visualization**:
   * Create interactive dashboards using **Amazon QuickSight** to visualize key metrics such as voting trends, council member participation, and motion success rates.
   * Allow for different user views (e.g., public access vs. internal analysis).
5. **Cloud Infrastructure**:
   * Set up **AWS Lambda** for serverless computation tasks and automate data processing workflows.
   * Use **Amazon RDS** or **Amazon Aurora** for structured data storage and querying.
   * Implement security measures with **AWS IAM** to control data access and ensure secure data processing.

## Tools and Technologies:

* **AWS S3**: Scalable storage for raw voting records and member data.
* **AWS Glue**: For data cleaning, transformation, and preparation.
* **AWS Lambda**: To enable serverless data processing workflows.
* **Amazon RDS/Aurora**: For structured storage and querying of processed data.
* **Amazon QuickSight**: For creating interactive visualizations and dashboards.
* **Python (Pandas, NumPy, Matplotlib, Seaborn)**: For data analysis and statistical exploration.
* **AWS IAM**: For managing access control and security of the cloud environment.

## Deliverables:

* **Exploratory Data Analysis Report**: A detailed report summarizing voting patterns, trends, and key findings.
* **Interactive Dashboards**: AWS QuickSight-based visualizations of voting data, trends, and council member participation.
* **Cloud Infrastructure Setup**: A scalable and secure AWS-based architecture for managing voting data.
* **Processed Datasets**: Cleaned and preprocessed datasets stored in AWS for further analysis or public access.
* **Access Control Policies**: Secure user access policies using AWS IAM to ensure privacy and data integrity.

# Diagnostic Analysis

## Project Title:

Diagnostic Data Analysis of Vancouver City Council Voting Records Using AWS Cloud Computing

## Project Description:

This project focuses on performing diagnostic data analysis on the voting records of the Vancouver City Council to identify factors influencing voting behaviors and outcomes. The goal is to analyze voting patterns, council member alignments, and the drivers behind specific decision-making processes. By leveraging AWS cloud computing, the project will enable secure and scalable processing of large datasets, allowing for the identification of underlying reasons for particular voting outcomes. This diagnostic approach will help uncover correlations and causal relationships, providing valuable insights to city officials, researchers, and the general public.

## Objective:

* To identify the causes and influencing factors behind the voting outcomes in Vancouver City Council.
* To diagnose patterns of alignment between council members and their voting behavior.
* To implement AWS cloud infrastructure for processing, analyzing, and visualizing diagnostic data efficiently.
* To deliver insights on how council members' affiliations, motion types, and session contexts affect voting behavior.

## Background:

City councils play a crucial role in making decisions that shape the policies and development of a city. Understanding not just what decisions were made, but why and how council members arrived at those decisions, can reveal important dynamics within the council. Vancouver City Council’s voting records provide a rich dataset to explore these questions. By performing a diagnostic analysis, we aim to uncover the reasons behind voting outcomes, patterns of political or personal alignment, and contextual factors that influence council decisions.

## Dataset:

* **Voting Records Dataset**: Detailed data on council votes, including:
  + Vote outcomes (pass/fail).
  + Types of motions (e.g., policy changes, financial decisions).
  + Dates, times, and durations of council sessions.
  + Individual votes of each council member.
* **Council Member Data**: Information about each council member’s:
  + Name, party affiliation, role, and tenure.
  + Historical voting patterns and participation in council sessions.
* **Motion Metadata**: Data related to the context and content of each motion (e.g., public response, debate duration, political sensitivity).
* **Contextual Data (optional)**: External factors such as media coverage, public opinion, or external political pressures at the time of voting.

## Methodology:

1. **Data Collection and Storage**:
   * Gather voting records, council member profiles, and motion metadata from public or official sources.
   * Store raw data in **Amazon S3** for scalable and secure access.
2. **Data Cleaning and Preparation**:
   * Use **AWS Glue** to clean, standardize, and organize the datasets.
   * Handle missing values, incorrect data entries, and outliers.
3. **Diagnostic Analysis**:
   * Perform **correlation analysis** to identify relationships between council members' votes and specific factors such as political affiliation, motion type, or council session context.
   * Conduct **regression analysis** to model how different variables (e.g., motion type, member affiliation) affect voting outcomes.
   * Use clustering techniques to identify groups of council members who vote similarly and explore the factors driving these alignments.
4. **Pattern and Cause Identification**:
   * Explore patterns of voting consistency (e.g., council members consistently supporting or opposing specific types of motions).
   * Investigate external factors (e.g., media, public opinion) that may correlate with voting trends.
   * Use time-series analysis to understand how patterns change over time or in specific periods (e.g., election cycles, budget seasons).
5. **Data Visualization**:
   * Create interactive dashboards using **Amazon QuickSight** to visualize the factors influencing voting outcomes.
   * Provide drill-downs for individual motions, council members, and sessions, showing the diagnostics of specific voting outcomes.
6. **Cloud Infrastructure Setup**:
   * Implement **AWS Lambda** for automated data processing tasks.
   * Use **Amazon RDS** or **Amazon Aurora** for structured data storage and querying.
   * Ensure data privacy and security with **AWS IAM** for role-based access control.

**Tools and Technologies:**

* **AWS S3**: For scalable and secure storage of voting records, council member data, and motion metadata.
* **AWS Glue**: For cleaning, transforming, and preparing data for analysis.
* **AWS Lambda**: For automating diagnostic analysis tasks in a serverless environment.
* **Amazon RDS/Aurora**: For relational database management and querying.
* **Amazon QuickSight**: For building interactive visualizations and dashboards to present diagnostic insights.
* **Python (Pandas, NumPy, Scikit-learn)**: For statistical analysis, regression modeling, and clustering algorithms.
* **AWS IAM**: For managing secure access to data and services.

## Deliverables:

* **Diagnostic Analysis Report**: A comprehensive report detailing the factors influencing voting outcomes and member behavior.
* **Correlation and Regression Models**: Results from statistical analyses, highlighting key drivers of voting behavior.
* **Interactive Dashboards**: AWS QuickSight dashboards that allow stakeholders to explore the factors affecting specific votes, council members, and motions.
* **Cleaned and Processed Data**: Organized datasets available for further analysis or public sharing.
* **Scalable Cloud Infrastructure**: A fully implemented AWS-based solution for performing ongoing diagnostic analysis as new voting records are added.
* **Security and Access Management**: Secure access policies via AWS IAM to ensure data privacy and controlled access to sensitive information.

## Timeline:

* **Week 1-2: Data Collection and Storage**:
  + Gather datasets and store them in AWS S3.
* **Week 3-4: Data Cleaning and Preparation**:
  + Use AWS Glue to clean, preprocess, and organize the data for analysis.
* **Week 5-6: Diagnostic Analysis**:
  + Perform correlation, regression, and clustering analyses to identify patterns and causes of voting behavior.
* **Week 7-8: Data Visualization and Reporting**:
  + Build interactive dashboards with Amazon QuickSight.
  + Generate a detailed report on the findings of the diagnostic analysis.
* **Week 9: Cloud Infrastructure Setup and Security**:
  + Implement AWS infrastructure, automate processes with Lambda, and secure the environment with IAM.
* **Week 10: Final Testing and Deliverables**:
  + Test the system, finalize the report, and deliver dashboards and processed data.

Top of Form

Bottom of Form

# Data Wrangling

## Project Title:

Data Wrangling for Vancouver City Council Voting Records Using AWS Cloud Computing

## Project Description:

This project focuses on performing data wrangling for the Vancouver City Council’s voting records using AWS cloud services. Data wrangling involves transforming and preparing the raw voting data into a structured format suitable for analysis. This process includes cleaning, reshaping, and standardizing the data, allowing for more efficient querying and visualization. The project aims to establish a robust AWS-based infrastructure that supports the preparation of large voting datasets for further diagnostic, descriptive, and predictive analyses. The processed data will be used to uncover insights into council voting behavior, motion outcomes, and trends over time.

## Objective:

* To clean, organize, and standardize the Vancouver City Council voting records for further analysis.
* To implement an AWS-based infrastructure for scalable data wrangling.
* To ensure that the dataset is ready for descriptive, diagnostic, and predictive analysis by transforming it into a usable format.
* To automate the data wrangling process to support future voting data imports.

## Background:

Vancouver City Council voting records contain rich information about council members' decisions and the motions presented over time. However, raw data from such sources often contain inconsistencies, missing values, and variations in structure that make analysis challenging. Data wrangling is essential to convert this data into a clean, structured format that can support various types of analysis. By leveraging AWS cloud technologies, this project ensures the scalability and security needed to manage large datasets, while automating the data wrangling process to handle future records efficiently.

## Dataset:

* **Voting Records Dataset**: The primary dataset, containing records of each vote cast in Vancouver City Council, including:
  + Voting outcomes (pass/fail).
  + Vote timestamps and council session identifiers.
  + Individual council members’ voting decisions.
  + Motion metadata (e.g., motion type, description, category).
* **Council Member Information**: Data containing information about council members such as:
  + Names, party affiliations, roles, and their participation in voting sessions.
* **Motion Metadata**: Descriptions of the types of motions discussed, including their categories (e.g., financial, environmental) and outcomes.

## Methodology:

1. **Data Collection and Ingestion**:
   * Gather voting records, council member information, and motion metadata from public sources, APIs, or internal databases.
   * Store the raw data in **AWS S3** to provide scalable storage and easy access to the datasets.
2. **Data Cleaning**:
   * Use **AWS Glue** to identify and handle missing values, duplicates, and inconsistencies in the voting records.
   * Standardize formats (e.g., date formats, member names) to ensure uniformity across the dataset.
   * Remove irrelevant or erroneous data entries that do not contribute to the analysis.
3. **Data Transformation**:
   * Normalize the data using **Python (Pandas)** to ensure consistency across all records.
   * Reshape the data into a structured format (e.g., converting long-format data into wide-format where necessary).
   * Use feature engineering techniques to extract additional insights, such as session duration or vote alignment scores.
4. **Schema Definition and Data Standardization**:
   * Create a unified schema for all datasets, ensuring consistent column names, data types, and relationships between datasets (e.g., council members to voting sessions).
   * Store the processed, structured data in **Amazon RDS** or **Amazon Aurora** to enable querying and further analysis.
5. **Automation and Workflow Management**:
   * Implement **AWS Lambda** to automate data ingestion and transformation workflows, enabling the system to handle new voting data as it becomes available.
   * Schedule regular data wrangling processes to keep the dataset updated and accurate.

## Tools and Technologies:

* **AWS S3**: For scalable, reliable storage of raw and processed datasets.
* **AWS Glue**: For cleaning, transforming, and preparing the data for analysis.
* **AWS Lambda**: To automate data wrangling tasks and create event-driven workflows.
* **Amazon RDS/Aurora**: For structured data storage, querying, and relational database management.
* **Python (Pandas, NumPy)**: For performing data cleaning, reshaping, and feature engineering tasks.
* **AWS IAM**: For managing access control and securing the cloud environment during data processing.

## Deliverables:

1. **Cleaned and Standardized Dataset**: A fully cleaned, organized, and transformed voting record dataset, ready for analysis.
2. **Unified Data Schema**: A well-defined schema for the voting data, ensuring consistency across all data points.
3. **Automated Data Wrangling Pipeline**: A fully implemented AWS Lambda-based pipeline that automates the wrangling process for future datasets.
4. **AWS Infrastructure Setup**: A cloud infrastructure that supports scalable data processing, automation, and storage.
5. **Data Wrangling Report**: A detailed report outlining the steps taken for data cleaning, standardization, and transformation.
6. **Access-Controlled Data Storage**: Secure access to the processed datasets, managed via AWS IAM for different user roles (e.g., analysts, researchers, public).
7. **Documentation**: Comprehensive documentation outlining the data wrangling process, the AWS architecture, and how to manage and maintain the system.
8. **Testing Framework**: Validation tools to ensure data quality and consistency for ongoing datasets.

## Timeline:

* **Week 1-2: Data Collection and Ingestion**:
  + Gather the raw datasets, upload them to AWS S3, and begin initial exploration.
* **Week 3-4: Data Cleaning**:
  + Use AWS Glue and Python scripts to clean the data, handle missing values, and standardize formats.
* **Week 5-6: Data Transformation and Schema Definition**:
  + Normalize and reshape the data into the required format, creating a well-defined schema for relational storage.
* **Week 7-8: Automation and Workflow Setup**:
  + Implement AWS Lambda to automate data wrangling tasks and set up scheduled workflows for future data imports.
* **Week 9-10: Testing and Final Deliverables**:
  + Test the pipeline, finalize the cleaned dataset, and deliver the report, infrastructure, and documentation.

Top of Form

Bottom of Form

# Data Quality Control

## Project Title:

Quality Control for Vancouver City Council Voting Records Using AWS Cloud Computing

## Project Description:

This project focuses on ensuring the accuracy, reliability, and consistency of Vancouver City Council’s voting records through a robust quality control process. By implementing cloud computing solutions via AWS, the project aims to identify and rectify errors, missing data, and inconsistencies within the voting records dataset. The quality control framework will ensure that the data is accurate, validated, and ready for downstream analysis and reporting. The AWS platform will provide scalable, secure, and automated tools to conduct data validation and ensure continuous monitoring of data quality as new voting records are added.

## Objective:

* To establish a comprehensive quality control process for Vancouver City Council’s voting records.
* To ensure the accuracy, consistency, and completeness of the dataset for further analysis.
* To implement AWS cloud solutions that automate the identification of data quality issues and allow for ongoing quality monitoring.
* To create a framework for validation rules and testing to ensure data integrity.

## Background:

Vancouver City Council’s voting records contain critical information used for analyzing council decision-making, voting patterns, and governance trends. However, raw data often contains errors, missing entries, or formatting inconsistencies that can lead to incorrect conclusions or hinder meaningful analysis. A quality control process is necessary to address these issues before data can be reliably used. AWS offers powerful cloud-based solutions for managing large datasets and automating quality checks, making it ideal for handling ongoing quality control tasks.

## Scope:

This project will focus on the following aspects:

* Identifying and correcting missing, duplicated, or inconsistent voting records.
* Ensuring that council member profiles, vote outcomes, and motion details are complete and accurate.
* Implementing automated validation rules and monitoring systems for future data entries to maintain data quality.

## Methodology:

1. **Data Collection and Initial Assessment**:
   * Store raw voting data in **AWS S3** and perform an initial assessment to identify common issues such as missing fields, inconsistent formats, and duplicated records.
   * Conduct exploratory analysis to highlight specific data quality challenges (e.g., incomplete council member profiles, vote duplication).
2. **Data Validation and Error Detection**:
   * Implement **AWS Glue** to perform data cleansing, addressing missing values, inconsistent formats, and outliers.
   * Create validation rules to ensure data accuracy and completeness, such as checks for:
     + Consistent vote outcomes (pass/fail).
     + Accurate date formats and timestamps for voting sessions.
     + Cross-checking vote records against council member profiles.
3. **Automated Quality Control**:
   * Use **AWS Lambda** to automate recurring quality control tasks, such as running data validation scripts on newly ingested data.
   * Schedule automated processes for checking data consistency and completeness at regular intervals (e.g., weekly or monthly).
4. **Monitoring and Reporting**:
   * Use **Amazon CloudWatch** to set up alerts for data quality issues, ensuring timely detection and correction.
   * Implement **Amazon QuickSight** to generate dashboards that monitor data quality metrics (e.g., missing values, data inconsistencies) over time.
   * Provide automated reports on data quality status, summarizing detected issues and corrective actions.
5. **Data Repair and Standardization**:
   * For any identified issues, repair and standardize the voting records to ensure consistency across the entire dataset.
   * Use feature engineering and data transformation techniques to restructure problematic sections of the data, ensuring alignment with council session details and member profiles.
6. **Continuous Improvement**:
   * Create a framework for ongoing data quality monitoring and improvements, ensuring that as new data is added, the quality control process automatically validates and corrects any issues.

## Tools and Technologies:

* **AWS S3**: For storing raw and processed voting records data.
* **AWS Glue**: For data cleaning, transformation, and applying validation rules.
* **AWS Lambda**: For automating recurring quality control tasks and ensuring continuous data validation.
* **Amazon CloudWatch**: For monitoring and setting up alerts for data quality issues.
* **Amazon QuickSight**: For building dashboards and generating reports on data quality metrics.
* **Python (Pandas, NumPy)**: For writing scripts to detect and address data quality issues.
* **AWS IAM**: For securing access and ensuring controlled access to data during quality control processes.

## Deliverables:

1. **Quality-Controlled Dataset**: A cleaned, standardized, and validated dataset of Vancouver City Council’s voting records, ready for analysis.
2. **Automated Validation Pipeline**: AWS Lambda-based automation for continuous quality checks and error detection.
3. **Data Quality Dashboards**: Real-time monitoring dashboards via Amazon QuickSight, providing insights into the current status of data quality.
4. **Data Quality Reports**: Regularly generated reports summarizing the issues found and actions taken to address data quality.
5. **AWS Cloud Infrastructure Setup**: A fully implemented AWS-based infrastructure that supports scalable data quality management.
6. **Data Quality Framework Documentation**: Comprehensive documentation outlining validation rules, quality control procedures, and AWS architecture for future maintenance.

## Timeline:

* **Week 1-2: Data Collection and Initial Assessment**:
  + Upload raw data to AWS S3, perform an initial data audit, and identify data quality issues.
* **Week 3-4: Data Validation Setup**:
  + Develop validation rules and scripts using AWS Glue and Python to detect missing values, duplicates, and formatting errors.
* **Week 5-6: Automation and Monitoring Setup**:
  + Implement AWS Lambda for automated quality control tasks, set up CloudWatch alerts, and build QuickSight dashboards for real-time monitoring.
* **Week 7-8: Data Repair and Final Testing**:
  + Correct identified data issues, finalize the quality-controlled dataset, and test the automated validation processes.
* **Week 9-10: Reporting and Documentation**:
  + Generate the first quality control reports and provide complete documentation for the entire process, along with the final deliverables.