**Data Analysis:**

In this Data Analysis Step, I use AWS Glue DataBrew for analysis purposes leading to structured high-quality datasets for process of cleaning and transforming data sets. AWS Glue DataBrew enabled streamlining of data preparation through automatic data profiling and cleaning functions as well as transformation and enrichment operations before analysts could begin their analysis Also, I used Amazon S3 which allowed me to manage system and user-uploaded data storage efficiently while enabling seamless integration with Amazon Athena querying. By utilizing Amazon S3 for data storage confirmed scalability, durability, and affordable data management for user-uploaded along with system-generated datasets. The serverless query service Athena provided me with direct access to analyze data kept in S3 through SQL operations which proves beneficial when managing extensive analytical requests.

By using AWS DataBrew, S3 and Athena, I streamlined data preparation, storage, enabling efficient analysis and querying of structured and unstructured datasets.

The screenshots of the process that I have done is inserted below:

A screenshot of a computer

AI-generated content may be incorrect.

A computer screen with a white screen

AI-generated content may be incorrect.

A computer screen with a white screen

AI-generated content may be incorrect.

A computer screen with a chat box

AI-generated content may be incorrect.

**Data Security:**

After analysing data, for data security I did implementation of AWS KMS (Key Management Service) encryption keys you developed data protection measures which use strong cryptographic methods. By using default encryption in S3 my system automatically encrypted every stored piece of information while providing added security protection. All objects that Amazon S3 receives after default encryption activation become encrypted without the opportunity for human error in security configurations. Bucket versioning is enabled through my initiative because it tracks data changes and prevents accidental bucket deletions. Data safety is ensured by this bucket versioning feature because users can retrieve previous document states when they delete or modify information by mistake. The implementation of replication rules on my S3 buckets guaranteed data redundancy while enhancing high availability across locations thus safeguarding your organization in case of failure or data loss requirements and it also ensured data redundancy and high availability across geographically distributed locations. And I repeat enabling data encryption, enabling bucket versioning and replication rules for 2 more buckets (raw and curated bucket) that I created in S3.

By integrating KMS, S3 encryption, bucket versioning, and replication rules, I reinforced data protection, and making sure that sensitive data remains recoverable and encrypted.

The needed screenshots are present below:

A computer screen with a white background

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A computer screen shot of a computer

AI-generated content may be incorrect.

A computer screen with a white screen

AI-generated content may be incorrect.

A computer screen shot of a computer screen

AI-generated content may be incorrect.

A computer screen with text on it

AI-generated content may be incorrect.

A computer screen with a white background

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Data Governance:**

In this Data Governance, I used AWS Glue which allowed me to build an ETL workflow for data governance which extracted data and performed transformations before loading it into structured storage. The combination of AWS Glue with Amazon S3 shows use a structured data management system and efficient method to maintain high-quality dataset quality. AWS Glue enabled you to establish an ETL (Extract Transform Load) data workflow for metadata-enabled data management. The ETL pipeline served as a fundamental tool to collect data from different sources while applying important data adjustments followed by structured storage system insertion. The adoption of this approach both improved system data processing speed and provided better metadata organization through AWS Glue Data Catalog for dataset discovery across AWS services.. And in other step, S3 bucket categorization was established as "passed" and "failed" to develop data quality protocols which allowed me to discard defective records while keeping data integrity intact. The process ensures data verification because it eliminates inaccurate and inaccurate datasets from moving forward toward analysis.

The screenshots for the step of data governance is inserted below:

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Data Monitoring:**

In Data Monitoring, I use Cloudwatch and create dashboards to track metrics such as BucketSizeBytes and ResourceUsage. BucketSizeBytes visualizes the size of storage of an S3 bucket over the time and ResourceUsage shows the resource utilization metrics. The dashboards are set to 15 minutes intervals for monitoring which means data are refreshed in each 15 minutes. Also, by enabling time zones of 3 months helps to show real time visibility into resource utilization and storage consumption. After that, I create alarms in Cloudwatch to ensure proactive monitoring which allow me to detect optimize system performance and anomalies. Moreover, In AWS CloudTrail users could achieve security compliance by maintaining an audit trail which logged API calls and tracked and monitored user activities. Enabled trail logging status shows actively recording API activity and the log are stored in S3 buckets. This approach enabled both accountability tracking and unauthorized access detection along with the improvement of overall security profile.

The process for data monitoring and controlling screenshots are inserted below:

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.