**Project Portfolio**

Exploratory Data Analysis (EDA)

Project Title:

Exploring University Registrar and Vancouver Water Fountain Data

Project Description:

I started by examining the available datasets provided by our university registrar and Vancouver city water fountains. I wanted to understand basic qualities of the data before moving on.

Objective:

The aim of this task is to familiarize users with dataset contents quickly while detecting essential patterns alongside investigating potential data quality problems during the initial stages.

Datasets:

* A wide range of University Registrar datasets consists of Student lists in addition to Faculty enrollments and Payment transactions.
* Publicly available data on water fountains from the City of Vancouver.

Methodology:

- Uploaded the data to AWS S3.

- Conducted preliminary profiling using AWS Glue DataBrew to analyze data characteristics and quality.

Tools and Technologies:

- AWS S3

- AWS Glue DataBrew

Deliverables:

- Initial data profiles highlighting columns, data distributions, and missing values.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Descriptive Analysis

Project Title:

Summarizing University and Vancouver Data Through Descriptive Analysis

Project Description:

I performed descriptive analysis following discovery to present the essential characteristics of the datasets. The analysis used detailed attribute profiling and distribution analysis on important attributes.

Objective:

Rephrase the data to gain better insight into its distribution patterns together with frequency statistics and important features.

Methodology:

- Uploaded Registrar datasets using a Windows-based AWS EC2 instance with PowerShell scripts.

- Uploaded Vancouver datasets directly to AWS S3.

- Performed detailed data profiling with AWS Glue DataBrew.

Tools and Technologies:

- AWS EC2 with PowerShell

- AWS S3

- AWS Glue DataBrew

Deliverables:

- Comprehensive summary statistics and descriptive insights.

- Architectural diagrams clearly showing the data workflow.

A diagram of a diagram

Description automatically generated with medium confidence

A computer screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Diagnostic Analysis

Project Title:

Understanding Vancouver’s Water Fountain Distribution and Maintenance Patterns

Project Description:

The data needs to be restated for improved understanding of its distribution while keeping both frequency statistics and important characteristics.

Objective:

Identify factors leading to variations in water fountain distribution and operations.

Methodology:

- Used AWS Glue Data Catalog for structured querying.

- Ran SQL queries in AWS Athena to pinpoint key patterns and insights.

Tools and Technologies:

- AWS Athena

- AWS Glue Data Catalog

- AWS S3

Deliverables:

- Clear diagnostic insights based on query results.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Data Wrangling

Project Title:

Organizing and Preparing Data for Analytics

Project Description:

To make the data analysis ready, I completed comprehensive data wrangling. This involved structured ingestion, initial profiling, and preliminary cleaning to ensure data accuracy and readiness.

Objective:

Clean and organize the data systematically for further analysis.

Methodology:

- Used AWS EC2 and PowerShell scripts for organized data ingestion.

- Performed detailed data profiling using AWS Glue DataBrew.

- Created clear visual diagrams using Draw.io for better understanding.

Tools and Technologies:

- AWS EC2, PowerShell

- AWS Glue DataBrew

- AWS S3

- Draw.io

Deliverables:

- Structured data storage paths in AWS S3.

- Clear architecture diagrams and comprehensive profiles.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Data Quality Control

Project Title:

Ensuring High-Quality Data for Reliable Analytics

Project Description:

To guarantee data reliability, I implemented robust data quality checks using AWS Glue and DataBrew. This process ensured accuracy, consistency, and reliability of the datasets.

Objective:

Establish and maintain high data quality standards.

Methodology:

- Executed detailed AWS Glue DataBrew profiling jobs.

- Applied data quality validation rules through AWS Glue’s ETL pipeline.

Tools and Technologies:

- AWS Glue DataBrew

- AWS Glue ETL

- AWS S3

Deliverables:

- Comprehensive data quality profiles and validations.

- Clearly outlined Data Quality Control processes and pipelines.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Conclusion

The portfolio demonstrates the analytical adventure I experienced which showcases how I structured my approach to data understanding alongside quality control and deep pattern analysis. This document showcases the practical knowledge I acquired regarding cloud analytics alongside data clearance procedures and structured system handling methods.