**Project Description**

Descriptive Analysis of Animal Services Business Licenses in Vancouver

**Project Title**

Analyzing Business License Status, Revenue & Workforce in Vancouver's Animal Services Sector

**Objective**

This project aims to build a scalable, cloud-based Data Analytics Platform using AWS services to analyze business license data from Vancouver's Animal Services sector. In order to help city authorities and business planners make data-driven decisions, the main objective is to extract insights regarding licensing statuses, revenue patterns, and workforce distribution.

**Dataset**

The Open Data Catalog of the City of Vancouver provided the dataset for this project. It includes comprehensive data about companies in the Animal Services industry, such as their staff size, operational status, and licensing details.  
**Key attributes in the dataset include**  
• LicenceNumber – Unique identifier for each business license  
• BusinessName – Name of the registered business  
• BusinessType – Type of animal services (e.g., grooming, pet daycare)  
• Status – License status (Issued, Pending, Inactive)  
• IssuedDate – Date when the license was granted  
• ExpiredDate – The expiry date of the license  
• NumberofEmployees – Total staff working at the business  
• FeePaid – Total licensing fee paid by the business  
The dataset was downloaded in CSV format and uploaded to Amazon S3 under the path:  
s3://business-raw-aks/licenses/animal-services/year=2025/server=BLGVS-AKS/

**Methodology**

**Data Collection and Preparation**

The first step involved ingesting the Animal Services business license dataset from the City of Vancouver’s Open Data Catalog and preparing it for cloud-based processing.  
**Steps Undertaken**  
• I downloaded the raw dataset (CSV format) from the Open Data Portal.  
• Created an S3 bucket named business-raw-aks and uploaded the dataset under a structured path:  
 s3://business-raw-aks/licenses/animal-services/year=2025/server=BLGVS-AKS/  
• I then created a project in AWS Glue DataBrew and connected it to the raw S3 data to begin the profiling and cleaning workflow. A screen shot of a computer

AI-generated content may be incorrect.

**Data Profiling**

I used AWS Glue DataBrew to do data profiling before I started examining the Animal Services dataset. Prior to cleaning and transformation, this process aimed to comprehend the data's structure, quality, and completeness.  
**Steps Undertaken**  
• Created a transformed bucket in S3 named business-trf-aks to store cleaned datasets separately from the raw data.  
• In DataBrew, I set up a new project and connected it to the dataset stored in the raw bucket.  
• Encountered a CSV delimiter issue and resolved it by changing the delimiter from commas to semicolons.  
• Ran a profiling job to generate a summary of the dataset.  
**Key Findings**  
• Missing values in FeePaid and NumberofEmployees.  
• Redundant columns like Postal Code, GeoPoint, and Unit.  
• Inconsistent data types in date and financial fields.  
• Placeholder values like 0 or NULL in critical columns.

**Data Cleaning**

A diagram of a computer

AI-generated content may be incorrect.

After profiling the Animal Services business license dataset, I proceeded with data cleaning using AWS Glue DataBrew.  
**Steps Undertaken**  
• Removed unnecessary columns: Geom, GeoPoint, Street, Postal Code, Unit, Unit Type, BusinessTradeName.  
• Dropped records where FeePaid or NumberofEmployees were null or zero.  
• Removed rows with missing BusinessType.  
• Standardized data formats (DECIMAL, INTEGER, YYYY-MM-DD).  
• Renamed columns for clarity.  
• Filtered to include only Animal Services businesses with Status = 'Issued'.  
• Saved cleaned data in transformed S3 bucket (business-trf-aks).

**Data Cataloging**

A diagram of a computer

AI-generated content may be incorrect.

After cleaning, I cataloged the dataset using AWS Glue Data Catalog to make it accessible and query-ready with Amazon Athena.  
**Steps Undertaken**  
• Created Glue database: business-data-catalog-aks.  
• Used Glue Crawler to crawl transformed S3 bucket and detect schema.  
• Created tables with correct headers, data types, and formats.  
• Partitioned table by Status to improve query performance.

**Data Summarization**

A diagram of a computer

AI-generated content may be incorrect.

To extract insights, I summarized the data using SQL queries in Amazon Athena.  
**Steps Undertaken**  
• Wrote queries to calculate:  
 1. Total number of licenses per status.  
 2. Total licensing revenue by BusinessType.  
 3. Average employees per BusinessType.  
• Built a visual ETL pipeline in DataBrew for final summarization.  
• Stored summarized data in curated S3 bucket: business-cur-aks/licenses/animal-services/

**Tools and Technologies**

This project utilized a range of AWS cloud-based services to ensure secure, scalable, and efficient data processing:  
• Amazon S3 – Cloud storage for raw, cleaned, and curated datasets.  
• AWS Glue DataBrew – For visual data profiling, cleaning, and transformation workflows.  
• AWS Glue Data Catalog – Managed metadata and enabled structured, schema-aware queries.  
• Amazon Athena – Used for writing and executing SQL queries to analyze data directly in S3.  
• Amazon EC2 – Provided scalable compute resources to support back-end data operations.  
• AWS Key Management Service (KMS) – Ensured encryption and secure access to sensitive datasets.  
• AWS CloudTrail – Monitored user activities and maintained audit logs for compliance.  
• AWS CloudWatch – Enabled real-time monitoring of AWS service usage and configured cost and performance alerts.

**Deliverables**

• A structured and cleaned dataset stored in Amazon S3, ready for analysis and reporting.  
• SQL queries and summarized insights executed through Amazon Athena.  
• A metadata catalog in AWS Glue for easy access to partitioned, query-optimized data.  
• A visual ETL pipeline in AWS Glue DataBrew for cleaning and aggregating business metrics.  
• A comprehensive monitoring dashboard in CloudWatch and activity tracking through CloudTrail.  
• A secure architecture with encryption, replication, and access control using AWS KMS.  
• Estimated cost summary and budget tracking for cloud resource usage.

**Insights and Recommendations**

**Insights**  
- Most Animal Services businesses operate with fewer than 5 employees.  
- Issued licenses dominate the dataset.  
- A few business types contribute the majority of licensing revenue.  
- Over 95% of expiration dates were outdated.  
**Recommendations**  
- Improve update frequency of expiration records.  
- Focus policies on high-revenue business types.  
- Integrate spatial data for more geographic insights.  
- Explore ML for forecasting license renewals or spotting anomalies.  
This project helped me apply AWS tools in a real-world scenario, transforming raw data into valuable business insights.

**Conclusion**

This descriptive analysis project aims to provide a comprehensive understanding of licensing patterns and business activity within Vancouver’s Animal Services sector. By leveraging AWS tools, the project delivered insights into license status distribution, workforce size, and revenue generation. For both public and commercial stakeholders, these insights can help with improved planning, compliance monitoring, and operational enhancements. Additionally, the scalable and secure architecture lays the groundwork for future growth, encompassing geographic mapping and predictive analytics.

**Certification**  
 AWS Academy Cloud Foundations – Successfully completed the certification, demonstrating foundational knowledge of cloud computing concepts, AWS core services, security, architecture, and pricing models.

A certificate of completion

AI-generated content may be incorrect.

— Akshita Katoch