**Seattle Pet License dimensional modeling and Loading via Talend - Workshop implementation by Individuals**

**Neha Bhutkar**

**Table of Contents**

1. **Introduction**
   * Project Overview
   * Data Source Description
2. **Data Profiling and Understanding**
   * Schema Examination
   * Content Assessment
   * Data Quality Assessment
3. **Business Requirements**
   * Creating Violation Category
   * Calculating Violation Score
   * Determining Inspection Results
   * Handling Location Field
4. **Data Preparation (Part 1)**
   * Data Retrieval
   * Data Format Selection
   * Data Loading into Stage Table
   * Staging Guidelines and Audit Columns
   * Data Profiling Outcomes
   * Data Cleansing Methods
5. **Dimensional Data Modeling (Part 2)**
   * Identification of Facts and Dimensions
   * Grain Definition
   * Handling Inspection Codes
   * Surrogate Keys, Relationships, and Associations
   * Appropriate Data Types
   * Date Dimension and Format
   * ER Studio/Navicat Modeling
   * DDL Script Generation
6. **Data Loading (Part 3)**
   * Data Preparation Workflow
   * Talend Jobs Execution
   * Business Logic Documentation
   * Repository Usage

**Project Overview:**

The Food Inspections Assignment involves the analysis of food inspection data from the Sonoma County Department of Health Services. This documentation provides a comprehensive overview of the project's key components, data sources, and deliverables.

**Step 1: Discussed the Schema and Content of the data provided.**

* Understood the layout of tables, fields, and the relationships between different data elements. It helped us in understanding how the data is organized.

**Step 2: Data Profiling in Alteryx:**

**Title**: **Seattle Pet License dimensional modeling and Loading via Talend**

**Description of Dataset:** A list of active/current Seattle pet licenses, including animal type (species), pet's name, breed, and the owner's ZIP code.

**Aim of the Assignment:**

The objective of this assignment is to examine the dataset and document any identified issues, including but not limited to missing data, data discrepancies, and inconsistencies.

**Alteryx Profiling workflow screenshot:**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

**Observation:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Datatype | % missing | Unique values | Observation | Min Value | Max Value |
| License Issue Date | V\_String | 0.0% | 931 | Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together. | 1 | 190 |
| License Number | V\_String | 0.0% | 42192 | Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together. | 1 | 2 |
| Animal’s Name | V\_String | 0.07% | 12696 | Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together. | 1 | 395 |
| Species | V\_String | 0.0% | 4 | Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together. | 2 | 29173 |
| Primary Breed | V\_String | 0.0% | 237 | Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together. | 1 | 7242 |
| Secondary Breed | V\_String | 36.0% | 259 | This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together. | 1 | 17830 |
| Zip Code | V\_String | 0.3% | 208 | Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together. | 1 | 3873 |

Talend Implementation:

Talend Workflow screenshots:

1. Loading stage table

A screenshot of a computer

Description automatically generated

1. Dimension Breed

A screenshot of a computer

Description automatically generated

Dimension Location:

A screenshot of a computer

Description automatically generated

Dimension Date:

A screenshot of a computer

Description automatically generated

Loading Fact Table:

MySQL:

A computer screen shot of a computer program

Description automatically generated

Azure:

A screenshot of a computer

Description automatically generated

Database screenshots:

MySQL loading tables:

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

Azure:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

