ZOO 4926

Matt Gitzendanner

Ashley T. Zeledon Bodden

December 10, 2018

**Problem Set 4**

Find a dataset that is currently in a table that you think could be turned into a multi-table database.

1. Describe the data. Write a metadata document that describes the source of the data, what the data are, what each column is, etc.

# Active Businesses Metadata

# Information about the Active Businesses Data Set

This file contains City of Gainesville licensed businesses public domain information from the City of Gainesville Data Organization's website. The dataset is from December 7, 2018.

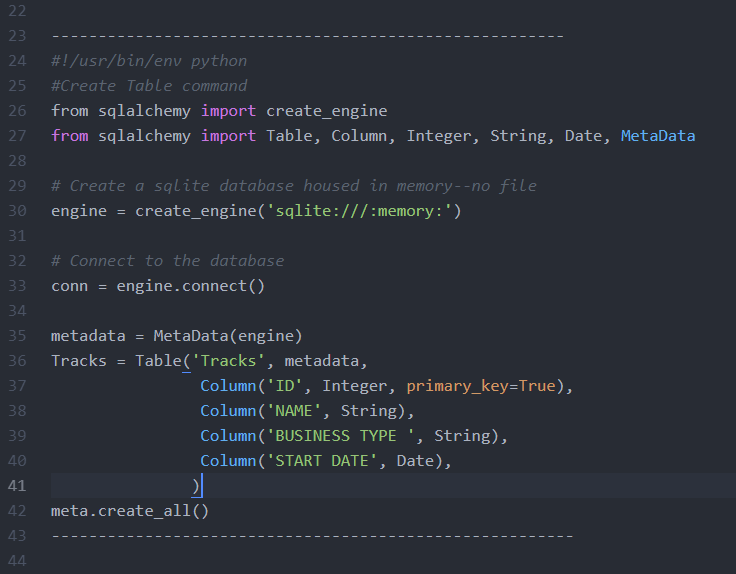
The columns in the dataset are in the table below.

|  |  |
| --- | --- |
| Column Header | Description |
| BUSINESS TYPE | Type of business (STRING) |
| ID | Business Identification number (INTEGER) |
| NAME | Business Name (STRING) |
| START DATE | Effective date of business license (DATETIME) |
| PHYSICAL ADDRESS | Physical address where the business is located (STRING) |
| MAILING ADDRESS | (STRING) |
| MAILING CITY | (STRING) |
| BUSINESS PHONE | (INTEGER) |
| EMAIL | Business contact email (STRING) |
| CONTACT | Business contact (STRING) |
| LOCATION | (STRING) |

1. Normalize the data. Write descriptions (words, drawings, or tables are fine--not code at this point) of the tables you could use to normalize these data. Describe why you chose the normalize the data in this manner. Include the column names, data types for each column, primary and foreign keys you would use. (8pts.)

This data set can be normalized by creating a table with **ID** and **NAME** relating **BUSINESS TYPE** with **START DATE** to identify similar businesses and their active business license effective dates. This can help identify the time at which similar businesses began their effective dates. One can use this to identify businesses with the oldest **START DATE** which may prove helpful. Column names may be **ID** (INTEGER) with **Primary Key** so the database can keep track of the data with number; **NAME** (STRING) providing identification; **BUSINESS TYPE** (STRING UP TO 10 CHARACTERS) to save space in the database; **START DATE** (DATETIME). No **FOREIGN KEY** seems necessary for this table at the moment.

1. Either directly in sqlite or in Python with SQLAlchemy, write the code needed to define the tables above. (6pts.)



#!/usr/bin/env python

from sqlalchemy import create\_engine

from sqlalchemy import Table, Column, Integer, String, Date, MetaData

# Create a sqlite database housed in memory--no file

engine = create\_engine('sqlite:///:memory:')

# Connect to the database

conn = engine.connect()

#Create Table command

metadata = MetaData(engine)

Tracks = Table('Tracks', metadata,

Column('ID', Integer, primary\_key=True),

Column('NAME', String),

Column('BUSINESS TYPE ', String),

Column('START DATE', Date),

)

meta.create\_all()