**Databases for Data Science**

**Term Project**

The first project consists of two parts. In the first part, you will create a database called “Portal Project Ecological Database.” This is a database from a scientific project for long-term monitoring and experimental manipulation of a Chihuahuan desert ecosystem near Portal, Arizona. Hence, they are real scientific data gathered over a long period of time. You will be installing this database in your SQLite database system instance that you have already created and will be using it for performing many data manipulation operations.

In the second part, you will create a database called “Northwind,” which consists of a schema for small businesses. This is a popular database used for learning about relational databases. You will use the SQLite system to install this database and perform data querying using SQL.

You can find all of my project files here: [**https://github.com/nrdeering/Data710/tree/master/Term\_Project**](https://github.com/nrdeering/Data710/tree/master/Term_Project)

1. Portal Project: The data for the Portal Project can be found at the figshare site:

<https://figshare.com/articles/Portal_Project_Teaching_Database/1314459>

Create a folder called PortalProject, and download the Download All zip file (called 1314459.zip). Unzip it and you will see a set of .csv files in there as well as a .sqlite file.

I don’t want you to use the .sqlite file. Instead, create three tables (species, plots, surveys) using CREATE statements, and then upload the files in the .csv file into the database. Save the database in a file called YourName\_PortalProject.db, and upload that also as part of the project submission. You will also do some SQL queries and data manipulation as mentioned later.

1. Northwind: The data for Northwind is found in GitHub at the following location:

<https://github.com/jpwhite3/northwind-SQLite3>

Create a folder called Northwind, and then clone or download the zip file in there and unzip it. The website contains a READMe.md file, which shows the schema for the database. The file is also found in your directory as Northwind\_ERD. Unzip the Northwind\_large.sqlite.zip file. This contains the database that you want to upload into SQLite. In this case, you can open the .sqlite file as your database (no need for creating tables or uploading CSV files). You will do some SQL queries as mentioned next as part of the project submission.

**Queries to Be Answered**

Portal Project:

1. Find all genus and species of the rodent taxa.

|  |  |
| --- | --- |
| **species** | **genus** |
| harrisi | Ammospermophilus |
| taylori | Baiomys |
| merriami | Dipodomys |
| ordii | Dipodomys |
| spectabilis | Dipodomys |
| sp. | Dipodomys |
| albigula | Neotoma |
| sp. | Neotoma |
| leucogaster | Onychomys |
| torridus | Onychomys |
| sp. | Onychomys |
| baileyi | Chaetodipus |
| eremicus | Peromyscus |
| flavus | Perognathus |
| hispidus | Perognathus |
| intermedius | Chaetodipus |
| leucopus | Peromyscus |
| maniculatus | Peromyscus |
| penicillatus | Chaetodipus |
| sp. | Chaetodipus |
| fulvescens | Reithrodontomys |
| megalotis | Reithrodontomys |
| montanus | Reithrodontomys |
| sp. | Reithrodontomys |
| fulviventer | Sigmodon |
| hispidus | Sigmodon |
| ochrognathus | Sigmodon |
| spilosoma | Spermophilus |
| tereticaudus | Spermophilus |
| sp. | Sigmodon |
| sp. | Rodent |

1. Find all species (genus, species, and taxa) that were found by survey in 1987.

|  |  |  |
| --- | --- | --- |
| **species** | **genus** | **taxa** |
| megalotis | Reithrodontomys | Rodent |
| merriami | Dipodomys | Rodent |
| maniculatus | Peromyscus | Rodent |
| leucogaster | Onychomys | Rodent |
| eremicus | Peromyscus | Rodent |
| torridus | Onychomys | Rodent |
| albigula | Neotoma | Rodent |
| ordii | Dipodomys | Rodent |
| bilineata | Amphispiza | Bird |
| spectabilis | Dipodomys | Rodent |
| chlorurus | Pipilo | Bird |
| harrisi | Ammospermophilus | Rodent |
| brunneicapillus | Campylorhynchus | Bird |
| spilosoma | Spermophilus | Rodent |
| penicillatus | Chaetodipus | Rodent |
| audubonii | Sylvilagus | Rabbit |
| hispidus | Sigmodon | Rodent |
| viridis | Crotalus | Reptile |
| melanocorys | Calamospiza | Bird |
| sp. | Rodent | Rodent |
| flavus | Perognathus | Rodent |
| sp. | Pipilo | Bird |

1. Print hindfoot length and weight of all species found in “Long-term Krat Exclosure.”

There ended up being 820 rows for this question. I grabbed the top 5:

|  |  |  |
| --- | --- | --- |
| **hindfoot\_length** | **weight** | **survey\_cnt** |
|  |  | 477 |
| 16.0 | 8.0 | 111 |
| 16.0 | 9.0 | 105 |
| 17.0 | 10.0 | 80 |
| 15.0 | 7.0 | 66 |

1. Find the dominant species in “Spectab exclosure.” What if you restrict to only 1998?

*Limited results to top 5*

Merriami is the dominant species:

|  |  |
| --- | --- |
| **species** | **species\_cnt** |
| merriami | 1584 |
| ordii | 693 |
| spectabilis | 520 |
| penicillatus | 322 |
| torridus | 188 |

In 1998, it is also the Merriami:

|  |  |
| --- | --- |
| **species** | **species\_cnt** |
| merriami | 57 |
| ordii | 38 |
| penicillatus | 18 |
| torridus | 7 |
| baileyi | 6 |

Northwind Project:

1. Find all suppliers who supply a beverage (category).

|  |
| --- |
| **CompanyName** |
| Exotic Liquids |
| Refrescos Americanas LTDA |
| Bigfoot Breweries |
| Aux joyeux ecclésiastiques |
| Leka Trading |
| Pavlova, Ltd. |
| Plutzer Lebensmittelgroßmärkte AG |
| Karkki Oy |

1. Find suppliers (print supplier companyname and region) who supplied both condiments and confections (categories).

|  |  |
| --- | --- |
| **CompanyName** | **Region** |
| Exotic Liquids | British Isles |
| New Orleans Cajun Delights | North America |
| Grandma Kelly's Homestead | North America |
| Mayumi's | Eastern Asia |
| Pavlova, Ltd. | Victoria |
| Specialty Biscuits, Ltd. | British Isles |
| Heli Süßwaren GmbH & Co. KG | Western Europe |
| Leka Trading | South-East Asia |
| Zaanse Snoepfabriek | Northern Europe |
| Karkki Oy | Scandinavia |
| Forêts d'érables | North America |
| Plutzer Lebensmittelgroßmärkte AG | Western Europe |

1. Find all orders (print only orderId, orderdate, Product Name) by customer “Piccolo und mehr” through the employee named “Michael Suyama”.

751 lines total, only posting first 5:

|  |  |  |
| --- | --- | --- |
| **OrderID** | **OrderDate** | **ProductName** |
| 10489 | 2017-03-28 | Queso Cabrales |
| 10489 | 2017-03-28 | Pavlova |
| 10747 | 2017-11-19 | Gorgonzola Telino |
| 10747 | 2017-11-19 | Jack's New England Clam Chowder |
| 10747 | 2017-11-19 | Vegie-spread |

1. Count the number of shipments handled by each shipping company. What if you group by year?

Part 1:

|  |  |
| --- | --- |
| **CompanyName** | **order\_count** |
| Federal Shipping | 5461 |
| Speedy Express | 5335 |
| United Package | 5486 |

Part 2:

|  |  |  |
| --- | --- | --- |
| **order\_yr** | **CompanyName** | **order\_count** |
| 2012 | Speedy Express | 216 |
| 2012 | Federal Shipping | 219 |
| 2012 | United Package | 219 |
| 2013 | United Package | 443 |
| 2013 | Speedy Express | 450 |
| 2013 | Federal Shipping | 458 |
| 2014 | Speedy Express | 429 |
| 2014 | United Package | 447 |
| 2014 | Federal Shipping | 475 |
| 2015 | Federal Shipping | 477 |
| 2015 | United Package | 478 |
| 2015 | Speedy Express | 494 |
| 2016 | Speedy Express | 482 |
| 2016 | United Package | 500 |
| 2016 | Federal Shipping | 524 |
| 2017 | Speedy Express | 567 |
| 2017 | United Package | 605 |
| 2017 | Federal Shipping | 608 |
| 2018 | Federal Shipping | 501 |
| 2018 | Speedy Express | 506 |
| 2018 | United Package | 542 |
| 2019 | Speedy Express | 439 |
| 2019 | United Package | 453 |
| 2019 | Federal Shipping | 470 |
| 2020 | United Package | 447 |
| 2020 | Speedy Express | 463 |
| 2020 | Federal Shipping | 466 |
| 2021 | Speedy Express | 458 |
| 2021 | Federal Shipping | 475 |
| 2021 | United Package | 487 |
| 2022 | Federal Shipping | 430 |
| 2022 | Speedy Express | 449 |
| 2022 | United Package | 473 |
| 2023 | Federal Shipping | 358 |
| 2023 | Speedy Express | 382 |
| 2023 | United Package | 392 |