Loading Data in Python
A Jupyter Notebook Example

Data Exploration and Visualization

Introduction

Welcome to Data Exploration and Visualization.
This presentation serves as a walk-through and detailed explanation of the provided Example 1 Jupyter Notebook.
This Notebook loads a database of cereals and various properties about the cereals and then computes and displays the average Sugar content (in grams) of each manufacturer's cereals.
Here, you will step through the example code and learn each block's function.

Part 1

```
import sqlite3
```

This single line of code includes the libraries needed for the rest of the example. In this simple example, you will need only include the sqlite3 library. Future examples and assignments will require additional libraries.

Part 2

```
db_filename = 'cereals.db'
```

This single line of code establishes the name of the database that contains the relevant data for the example. While you will use a database containing cereal information here, your assignment will use a different database.

Part 3

```
conn = sqlite3.connect(
   db_filename)
c = conn.cursor()
```

These two lines of code establish the connection with the database that was named in Part 2 (line 1) and the create a cursor object. The cursor object is necessary to execute queries on the database.

Part 4

```
c.execute("SELECT DISTINCT
    Manufacturer
    FROM cereals")
```

This line of code perform a simple query on the database to return a list of all the manufacturers represented in the database. Note that while this line executes the query, it does not retrieve the results.

Part 5

```
for manufacturer in c.fetchall():
  c.execute("SELECT avg(Sugars)
    FROM cereals WHERE
    Manufacturer='{0}'".format(
   manufacturer[0]))
  print ("{0}:{1:.2f}".
    format( manufacturer[0],
    c.fetchall()[0][0]))
```

In this block of code, the results of the previous query are retrieved by using the ‘fetchall()’ function, and then iterated over in sending the second query, which returns the average Sugar value of the specified manufacturer. The print statement then displays the results.

Conclusion

This example demonstrates the process of loading a dataset and performing simple statistics on that dataset.