**Lesson 8**

How to pull data from a microsoft sql database

In [1]:

*# Import libraries*

**import** **pandas** **as** **pd**

**import** **sys**

**from** **sqlalchemy** **import** create\_engine, MetaData, Table, select

In [2]:

print 'Python version ' + sys.version

print 'Pandas version: ' + pd.\_\_version\_\_

Python version 2.7.5 |Anaconda 2.1.0 (64-bit)| (default, Jul 1 2013, 12:37:52) [MSC v.1500 64 bit (AMD64)]

Pandas version: 0.15.2

**Version 1¶**

In this section we use the ***sqlalchemy*** library to grab data from a sql database. Make sure to use your own ***ServerName***, ***Database***, ***TableName***.

In [14]:

*# Parameters*

ServerName = "RepSer2"

Database = "BizIntel"

TableName = "DimDate"

*# Create the connection*

engine = create\_engine('mssql+pyodbc://' + ServerName + '/' + Database)

conn = engine.connect()

*# Required for querying tables*

metadata = MetaData(conn)

*# Table to query*

tbl = Table(TableName, metadata, autoload=**True**, schema="dbo")

*#tbl.create(checkfirst=True)*

*# Select all*

sql = tbl.select()

*# run sql code*

result = conn.execute(sql)

*# Insert to a dataframe*

df = pd.DataFrame(data=list(result), columns=result.keys())

*# Close connection*

conn.close()

print 'Done'

Done

Select the contents in the dataframe.

In [15]:

df.head()

Out[15]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **DateSK** | **Date** | **Day** | **DaySuffix** | **DayOfWeek** | **DOWInMonth** | **DayOfYear** | **WeekOfYear** | **WeekOfMonth** | **Month** | **MonthName** | **Quarter** | **QuarterName** | **Year** | **StandardDate** | **HolidayText** |
| **0** | 20000101 | 2000-01-01 | 1 | 1st | Saturday | 1 | 1 | 1 | 1 | 1 | January | 1 | First | 2000 | 01/01/2000 | New Year's Day |
| **1** | 20000102 | 2000-01-02 | 2 | 2nd | Sunday | 1 | 2 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/02/2000 | None |
| **2** | 20000103 | 2000-01-03 | 3 | 3rd | Monday | 1 | 3 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/03/2000 | None |
| **3** | 20000104 | 2000-01-04 | 4 | 4th | Tuesday | 1 | 4 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/04/2000 | None |
| **4** | 20000105 | 2000-01-05 | 5 | 5th | Wednesday | 1 | 5 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/05/2000 | None |

In [16]:

df.dtypes

Out[16]:

DateSK int64

Date datetime64[ns]

Day int64

DaySuffix object

DayOfWeek object

DOWInMonth int64

DayOfYear int64

WeekOfYear int64

WeekOfMonth int64

Month int64

MonthName object

Quarter int64

QuarterName object

Year object

StandardDate object

HolidayText object

dtype: object

Convert to specific data types. The code below will have to be modified to match your table.

In [17]:

*# Convert data types*

df.StandardDate = pd.to\_datetime(df.StandardDate)

df.Year = df.Year.astype('int')

print 'Data Types'

print df.dtypes

Data Types

DateSK int64

Date datetime64[ns]

Day int64

DaySuffix object

DayOfWeek object

DOWInMonth int64

DayOfYear int64

WeekOfYear int64

WeekOfMonth int64

Month int64

MonthName object

Quarter int64

QuarterName object

Year int32

StandardDate datetime64[ns]

HolidayText object

dtype: object

**Version 2¶**

In [18]:

**import** **pandas.io.sql**

**import** **pyodbc**

In [19]:

*# Parameters*

server = 'repser2'

db = 'BizIntel'

*# Create the connection*

conn = pyodbc.connect('DRIVER={SQL Server};SERVER=' + server + ';DATABASE=' + db + ';Trusted\_Connection=yes')

*# query db*

sql = """

SELECT top 5 \*

FROM DimDate

"""

df = pandas.io.sql.read\_sql(sql, conn)

df.head()

Out[19]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **DateSK** | **Date** | **Day** | **DaySuffix** | **DayOfWeek** | **DOWInMonth** | **DayOfYear** | **WeekOfYear** | **WeekOfMonth** | **Month** | **MonthName** | **Quarter** | **QuarterName** | **Year** | **StandardDate** | **HolidayText** |
| **0** | 20000101 | 2000-01-01 | 1 | 1st | Saturday | 1 | 1 | 1 | 1 | 1 | January | 1 | First | 2000 | 01/01/2000 | New Year's Day |
| **1** | 20000102 | 2000-01-02 | 2 | 2nd | Sunday | 1 | 2 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/02/2000 | None |
| **2** | 20000103 | 2000-01-03 | 3 | 3rd | Monday | 1 | 3 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/03/2000 | None |
| **3** | 20000104 | 2000-01-04 | 4 | 4th | Tuesday | 1 | 4 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/04/2000 | None |
| **4** | 20000105 | 2000-01-05 | 5 | 5th | Wednesday | 1 | 5 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/05/2000 | None |

**Version 3¶**

In [20]:

**from** **sqlalchemy** **import** create\_engine

In [21]:

*# Parameters*

ServerName = "RepSer2"

Database = "BizIntel"

*# Create the connection*

engine = create\_engine('mssql+pyodbc://' + ServerName + '/' + Database)

df = pd.read\_sql\_query("SELECT top 5 \* FROM DimDate", engine)

df

Out[21]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **DateSK** | **Date** | **Day** | **DaySuffix** | **DayOfWeek** | **DOWInMonth** | **DayOfYear** | **WeekOfYear** | **WeekOfMonth** | **Month** | **MonthName** | **Quarter** | **QuarterName** | **Year** | **StandardDate** | **HolidayText** |
| **0** | 20000101 | 2000-01-01 | 1 | 1st | Saturday | 1 | 1 | 1 | 1 | 1 | January | 1 | First | 2000 | 01/01/2000 | New Year's Day |
| **1** | 20000102 | 2000-01-02 | 2 | 2nd | Sunday | 1 | 2 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/02/2000 | None |
| **2** | 20000103 | 2000-01-03 | 3 | 3rd | Monday | 1 | 3 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/03/2000 | None |
| **3** | 20000104 | 2000-01-04 | 4 | 4th | Tuesday | 1 | 4 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/04/2000 | None |
| **4** | 20000105 | 2000-01-05 | 5 | 5th | Wednesday | 1 | 5 | 2 | 2 | 1 | January | 1 | First | 2000 | 01/05/2000 | None |