Accessing Databases with SQL Magic

To communicate with SQL Databases from within a JupyterLab notebook, I use the SQL "magic" provided by the ipython-sql extension. Below, I'll use the *load_ext* magic to load the ipython-sql extension. The following required modules are pre-installed in the Skills Network Labs environment. However if you run this notebook commands in a different Jupyter environment (e.g. Watson Studio or Ananconda) you may need to install these libraries by removing the # sign before !pip in the code cell below.

These libraries are pre-installed in SN Labs. If running in another environment please uncomment lines below to install them:
!pip install --force-reinstall ibm_db==3.1.0 ibm_db_sa==0.3.3
Ensure we don't load_ext with sqlalchemy>=1.4 (incompadible)
!pip uninstall sqlalchemy==1.4 -y && pip install sqlalchemy==1.3.24
!pip install ipython-sql
%load_ext sql

Now we have access to SQL magic.
Next, I connect to a Db2 database by retrieving my credentials to access my Db2 database.¶
%sql
ibm_db_sa://bcy01016:pPV11zoSkjmBw2pU@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/bludb?security=SSL

Output:
'Connected: bcy01016@bludb'

Create a table and fill it with some test data
%%sql

```
CREATE TABLE INTERNATIONAL_STUDENT_TEST_SCORES (
        country VARCHAR(50),
        first_name VARCHAR(50),
        last_name VARCHAR(50),
        test_score INT
);
INSERT INTO INTERNATIONAL_STUDENT_TEST_SCORES (country, first_name, last_name, test_score)
VALUES
('United States', 'Marshall', 'Bernadot', 54),
('Ghana', 'Celinda', 'Malkin', 51),
('Ukraine', 'Guillermo', 'Furze', 53),
('Greece', 'Aharon', 'Tunnow', 48),
('Russia', 'Bail', 'Goodwin', 46),
('Poland', 'Cole', 'Winteringham', 49),
```

('Sweden', 'Emlyn', 'Erricker', 55),
('Russia', 'Cathee', 'Sivewright', 49),
('China', 'Barny', 'Ingerson', 57),
('Uganda', 'Sharla', 'Papaccio', 55),
('China', 'Stella', 'Youens', 51),
('Poland', 'Julio', 'Buesden', 48),
('United States', 'Tiffie', 'Cosely', 58),
('Poland', 'Auroora', 'Stiffell', 45),
('China', 'Clarita', 'Huet', 52),
('Poland', 'Shannon', 'Goulden', 45),
('Philippines', 'Emylee', 'Privost', 50),
('France', 'Madelina', 'Burk', 49),
('China', 'Saunderson', 'Root', 58),
('Indonesia', 'Bo', 'Waring', 55),
('China', 'Hollis', 'Domotor', 45),
('Russia', 'Robbie', 'Collip', 46),
('Philippines', 'Davon', 'Donisi', 46),
('China', 'Cristabel', 'Radeliffe', 48),
('China', 'Wallis', 'Bartleet', 58),
('Moldova', 'Arleen', 'Stailey', 38),
('Ireland', 'Mendel', 'Grumble', 58),
('China', 'Sallyann', 'Exley', 51),
('Mexico', 'Kain', 'Swaite', 46),
('Indonesia', 'Alonso', 'Bulteel', 45),
('Armenia', 'Anatol', 'Tankus', 51),
('Indonesia', 'Coralyn', 'Dawkins', 48),
('China', 'Deanne', 'Edwinson', 45),
('China', 'Georgiana', 'Epple', 51),
('Portugal', 'Bartlet', 'Breese', 56),
('Azerbaijan', 'Idalina', 'Lukash', 50),
('France', 'Livvie', 'Flory', 54),
('Malaysia', 'Nonie', 'Borit', 48),
('Indonesia', 'Clio', 'Mugg', 47),
('Brazil', 'Westley', 'Measor', 48),
('Philippines', 'Katrinka', 'Sibbert', 51),
('Poland', 'Valentia', 'Mounch', 50),
('Norway', 'Sheilah', 'Hedditch', 53),
('Papua New Guinea', 'Itch', 'Jubb', 50),
('Latvia', 'Stesha', 'Garnson', 53),
('Canada', 'Cristionna', 'Wadmore', 46),
('China', 'Lianna', 'Gatward', 43),
('Guatemala', 'Tanney', 'Vials', 48),
('France', 'Alma', 'Zavittieri', 44),
('China', 'Alvira', 'Tamas', 50),

('United States', 'Shanon', 'Peres', 45),
('Sweden', 'Maisey', 'Lynas', 53),
('Indonesia', 'Kip', 'Hothersall', 46),
('China', 'Cash', 'Landis', 48),
('Panama', 'Kennith', 'Digance', 45),
('China', 'Ulberto', 'Riggeard', 48),
('Switzerland', 'Judy', 'Gilligan', 49),
('Philippines', 'Tod', 'Trevaskus', 52),
('Brazil', 'Herold', 'Heggs', 44),
('Latvia', 'Verney', 'Note', 50),
('Poland', 'Temp', 'Ribey', 50),
('China', 'Conroy', 'Egdal', 48),
('Japan', 'Gabie', 'Alessandone', 47),
('Ukraine', 'Devlen', 'Chaperlin', 54),
('France', 'Babbette', 'Turner', 51),
('Czech Republic', 'Virgil', 'Scotney', 52),
('Tajikistan', 'Zorina', 'Bedow', 49),
('China', 'Aidan', 'Rudeyeard', 50),
('Ireland', 'Saunder', 'MacLice', 48),
('France', 'Waly', 'Brunstan', 53),
('China', 'Gisele', 'Enns', 52),
('Peru', 'Mina', 'Winchester', 48),
('Japan', 'Torie', 'MacShirrie', 50),
('Russia', 'Benjamen', 'Kenford', 51),
('China', 'Etan', 'Burn', 53),
('Russia', 'Merralee', 'Chaperlin', 38),
('Indonesia', 'Lanny', 'Malam', 49),
('Canada', 'Wilhelm', 'Deeprose', 54),
('Czech Republic', 'Lari', 'Hillhouse', 48),
('China', 'Ossie', 'Woodley', 52),
('Macedonia', 'April', 'Tyer', 50),
('Vietnam', 'Madelon', 'Dansey', 53),
('Ukraine', 'Korella', 'McNamee', 52),
('Jamaica', 'Linnea', 'Cannam', 43),
('China', 'Mart', 'Coling', 52),
('Indonesia', 'Marna', 'Causbey', 47),
('China', 'Berni', 'Daintier', 55),
('Poland', 'Cynthia', 'Hassell', 49),
('Canada', 'Carma', 'Schule', 49),
('Indonesia', 'Malia', 'Blight', 48),
('China', 'Paulo', 'Seivertsen', 47),
('Niger', 'Kaylee', 'Hearley', 54),
('Japan', 'Maure', 'Jandak', 46),
('Argentina', 'Foss', 'Feavers', 45),

('Venezuela', 'Ron', 'Leggitt', 60),
('Russia', 'Flint', 'Gokes', 40),
('China', 'Linet', 'Conelly', 52),
('Philippines', 'Nikolas', 'Birtwell', 57),
('Australia', 'Eduard', 'Leipelt', 53)

Explore the data table
SELECT country, first_name, last_name, test_score FROM
INTERNATIONAL_STUDENT_TEST_SCORES;

Output:

country	first_name	last_name	test_score
United States	Marshall	Bernadot	54
Ghana	Celinda	Malkin	51
Ukraine	Guillermo	Furze	53
Greece	Aharon	Tunnow	48
Russia	Bail	Goodwin	46
Poland	Cole	Winteringham	49
Sweden	Emlyn	Erricker	55
Russia	Cathee	Sivewright	49
China	Barny	Ingerson	57
Uganda	Sharla	Papaccio	55
China	Stella	Youens	51
Poland	Julio	Buesden	48
United States	Tiffie	Cosely	58
Poland	Auroora	Stiffell	45
China	Clarita	Huet	52
Poland	Shannon	Goulden	45
Philippines	Emylee	Privost	50
France	Madelina	Burk	49
China	Saunderson	Root	58
Indonesia	Bo	Waring	55
China	Hollis	Domotor	45
Russia	Robbie	Collip	46
Philippines	Davon	Donisi	46
China	Cristabel	Radeliffe	48
China	Wallis	Bartleet	58
Moldova	Arleen	Stailey	38
Ireland	Mendel	Grumble	58
China	Sallyann	Exley	51
Mexico	Kain	Swaite	46
Indonesia	Alonso	Bulteel	45
Armenia	Anatol	Tankus	51
Indonesia	Coralyn	Dawkins	48
China	Deanne	Edwinson	45
China	Georgiana	Epple	51
Portugal	Bartlet	Breese	56
Azerbaijan	Idalina	Lukash	50
France	Livvie	Flory	54
Malaysia	Nonie	Borit	48
Indonesia	Clio	Mugg	47
Brazil	Westley	Measor	48
Philippines	Katrinka	Sibbert	51
Poland	Valentia	Mounch	50
Norway	Sheilah	Hedditch	53
Papua New Guinea	Itch	Jubb	50
Latvia	Stesha	Garnson	53
Canada	Cristionna	Wadmore	46
China	Lianna	Gatward	43
Guatemala	Tanney	Vials	48
France	Alma	Zavittieri	44
China	Alvira	Tamas	50
United States	Shanon	Peres	45
Sweden	Maisey	Lynas	53
Indonesia	Kip	Hothersall	46
China	Cash	Landis	48
Panama	Kennith	Digance	45
China	Ulberto	Riggeard	48
Switzerland	Judy	Gilligan	49
Philippines	Tod	Trevaskus	52
Brazil	Herold	Heggs	44
Latvia	Verney	Note	50
Poland	Temp	Ribey	50
China	Conroy	Egdal	48
Japan	Gabie	Alessandone	47
Ukraine	Devlen	Chaperlin	54
France	Babbette	Turner	51
Czech Republic	Virgil	Scotney	52
Tajikistan	Zorina	Bedow	49
China	Aidan	Rudeyeard	50
Ireland	Saunder	MacLice	48
France	Waly	Brunstan	53
China	Gisele	Enns	52
Peru	Mina	Winchester	48
Japan	Torie	MacShirrie	50
Russia	Benjamen	Kenford	51
China	Etan	Burn	53
Russia	Merralee	Chaperlin	38
Indonesia	Lanny	Malam	49
Canada	Wilhelm	Deeprose	54
Czech Republic	Lari	Hillhouse	48
China	Ossie	Woodley	52
Macedonia	April	Tyer	50
Vietnam	Madelon	Dansey	53
Ukraine	Korella	McNamee	52
Jamaica	Linnea	Cannam	43
China	Mart	Coling	52
Indonesia	Marna	Causbey	47
China	Berni	Daintier	55
Poland	Cynthia	Hassell	49
Canada	Carma	Schule	49
Indonesia	Malia	Blight	48
China	Paulo	Seivertsen	47
Niger	Kaylee	Hearley	54
Japan	Maure	Jandak	46
Argentina	Foss	Feavers	45
Venezuela	Ron	Leggitt	60
Russia	Flint	Gokes	40
China	Linet	Conelly	52
Philippines	Nikolas	Birtwell	57
Australia	Eduard	Leipelt	53

# I have a python variable country with a value of "Canada". I use this variable in a SQL query to find all the rows of students from Canada.
country = "Canada"
%sql select * from INTERNATIONAL_STUDENT_TEST_SCORES where country = :country

Output:

country	first_name	last_name	test_score
Canada	Cristionna	Wadmore	46
Canada	Wilhelm	Deeprose	54
Canada	Carma	Schule	49

# I use an SQL query to retrieve the distribution of test scores (i.e. how many students got each score). I assign the result of this query to the python variable test_score_distribution.
test_score_distribution = %sql SELECT test_score as "Test Score", count(*) as "Frequency" from INTERNATIONAL_STUDENT_TEST_SCORES GROUP BY test_score;
test_score_distribution

Output:

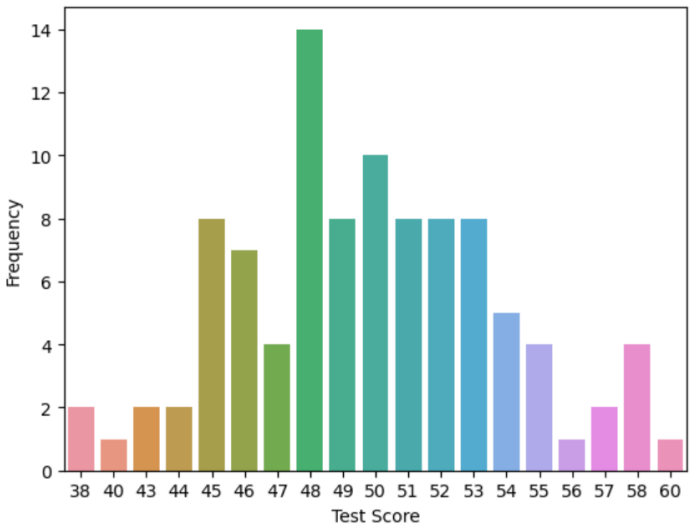
Test Score	Frequency
38	2
40	1
43	2
44	2
45	8
46	7
47	4
48	14
49	8
50	10
51	8
52	8
53	8
54	5
55	4
56	1
57	2
58	4
60	1

```
# Convert the SQL query result to a pandas dataframe. Then, plot the test score distribution.
dataframe = test_score_distribution.DataFrame()

%matplotlib inline
# uncomment the following line if you get an module error saying seaborn not found
# !pip install seaborn==0.9.0
import seaborn

plot = seaborn.barplot(x='Test Score',y='Frequency', data=dataframe)

# Output:
```




```
# Analysis: The plot shows that the most common test score was a score of 48 points, and most of the test scores lie in the range between 45 and 53 points.
```