Python + SQL

Dr. Villanes

A few things you are learning in Fall 1...

- What is SQL
- How to use SQL to query data in SQLite/Postgres
- Python
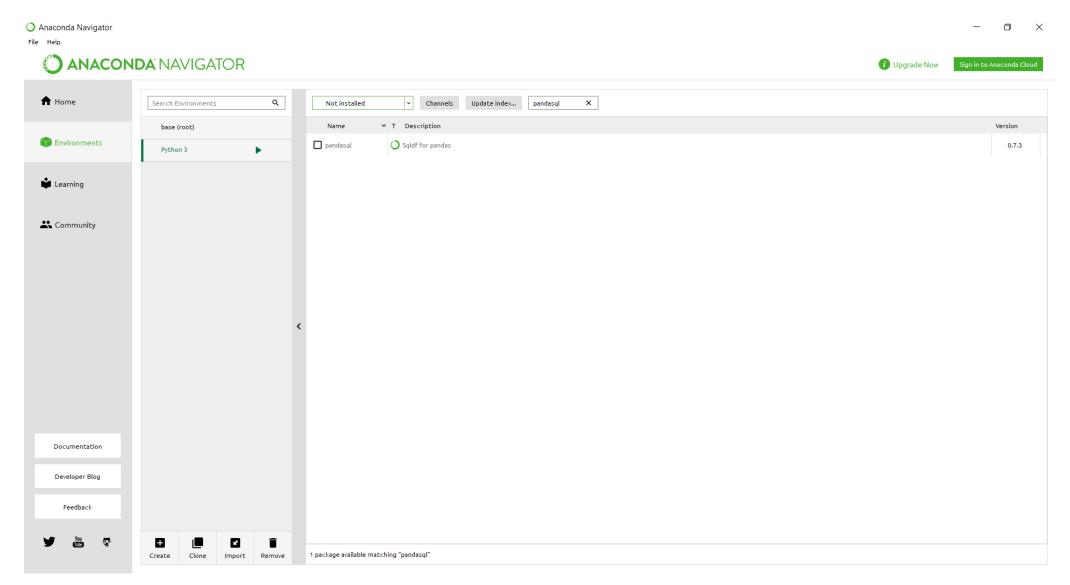- Pandas

How can we use SQL in Python?

Let's talk about Pandas…

- Pandas is one of the most preferred and widely used tools in data wrangling if not the most used one

- Built in SQL support

- Uses underlying SQLite

- You can treat Pandas dataframes as if they were tables

- Documentation: https://pypi.org/project/pysqldf/

- "Do I have to use SQL Pandas?" → https://pandas.pydata.org/docs/getting_started/comparison/comparison_with_sql.html -- Check the R, SAS links as well
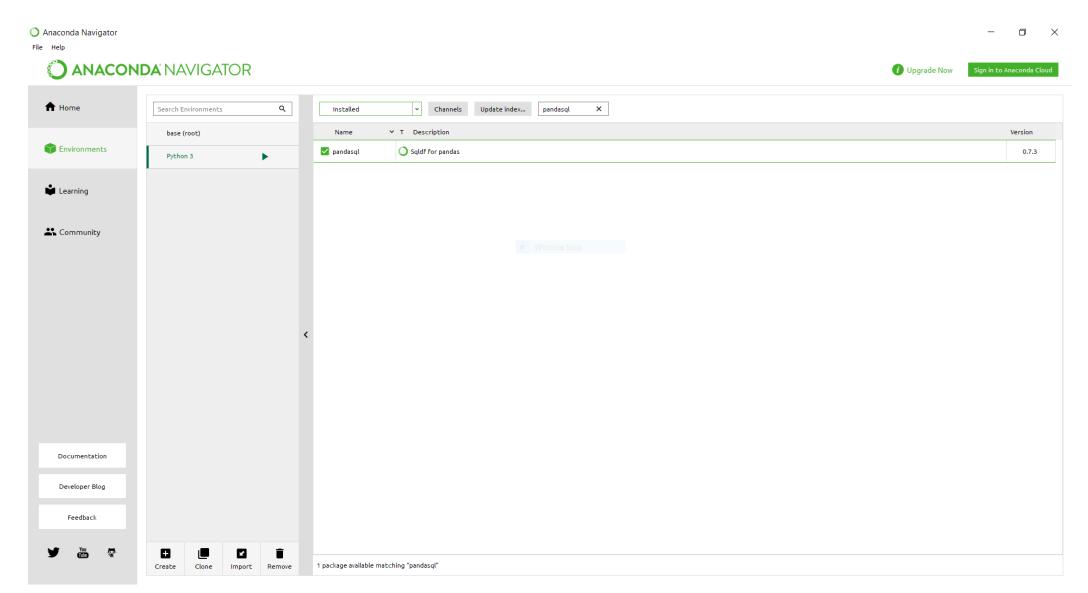
Install pandasql

```
In [1]:  import pandas as pd
         from pandasql import sqldf
         from pandasql import load_births

         births = load_births()

         print(sqldf("SELECT * FROM births where births > 250000 limit 5;", locals()))
```

```
---------------------------------------------------------------------------
ImportError                                Traceback (most recent call last)
<ipython-input-1-b72c4429f14a> in <module>()
      1 import pandas as pd
----> 2 from pandasql import sqldf
      3 from pandasql import load_births
      4
      5 births = load_births()

ImportError: No module named 'pandasql'
```
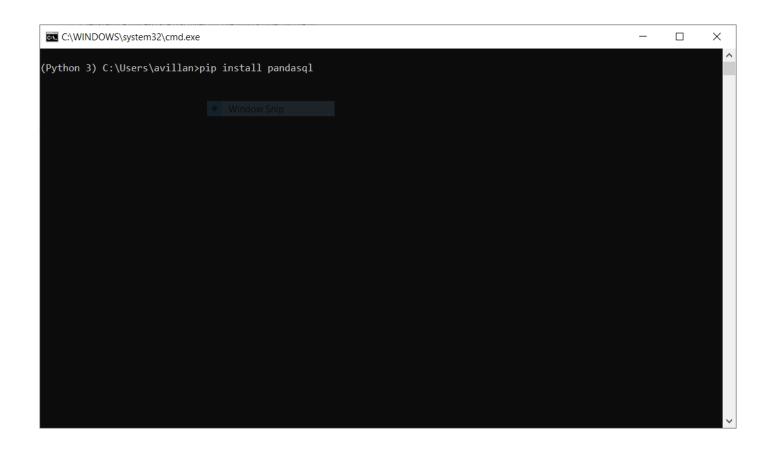
Let's install the library we will need

Once it's installed…

Another option: pip install pandasql

If you have Jupiter Notebook opened, closed everything and start again after pip install finishes

Let's go to the lab decription…

SQL: Lab 5

On Moodle, download the Lab files needed. Unzip the folder, and you will see 4 files:

- **Pandas + SQL.ipynb:** this is the example that we did during class.
- **Pandas + SQL - Exercise.ipynb:** you will complete this Jupyter Notebook, and upload it to Moodle (Lab 5 Checkout)
- **AWProduct.csv & AWProductSubcategory.csv:** the two files you need to complete the Lab

Good resource to have:

- Pandas Comparison with SQL:
  https://pandas.pydata.org/docs/getting_started/comparison/comparison_with_sql.html

Instructions:

1. Complete the **Pandas + SQL - Exercise.ipynb** questions
2. Upload your answers to Moodle Lab 5 Checkout. You should upload the Jupyter Notebook file (.ipynb)