[Video Notes](https://zoom.us/rec/play/D9SxalDPa14oqnTMqEqsGZiJSO0iCGQMNQZaTsQz8ZecVY7kZmsCBYKNrscsFE6FEYUUxQKFthflrpEi.M4NCRKivrcoCqfBI)

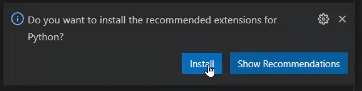
<https://docs.google.com/document/d/16T9M2-hbGyIIrUn-gVtXj62OvM1LZYivfmHRCW8pWcw/edit?usp=sharing>

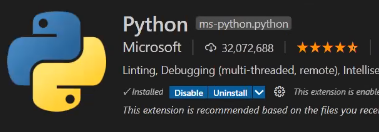
[00:00:32]

* Download Visual Studio Code from <https://code.visualstudio.com/>

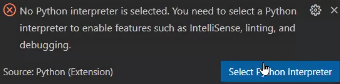
[00:06:15]

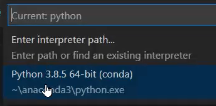
* Open a new file in VS Code and save it in .py format.
* Click on Install:





* Then select Select Python Interpreter:





* Now print hello world in python:





[00:07:54]

* Run the following code:

import pyodbc

server = 'ludsampledb.database.windows.net'

database = 'SampleDB'

username = 'sampleadmin'

password = '+U9Ly9/p'

driver= '{ODBC Driver 17 for SQL Server}'

with pyodbc.connect('DRIVER='+driver+';SERVER='+server+';PORT=1433;DATABASE='+database+';UID='+username+';PWD='+ password) as conn:

with conn.cursor() as cursor:

cursor.execute("SELECT TOP 3 name, collation\_name FROM sys.databases")

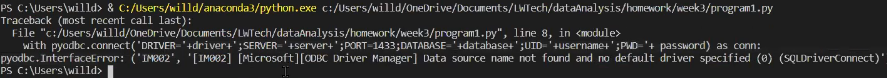
row = cursor.fetchone()

while row:

print (str(row[0]) + " " + str(row[1]))

row = cursor.fetchone()

Output:



[00:12:35]

Python:

<https://www.python.org/downloads/release/python-370/>

Get-Pip

<https://www.liquidweb.com/kb/install-pip-windows/>

ODBC Driver

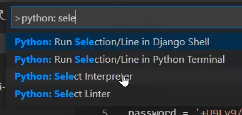
<https://www.microsoft.com/en-us/download/details.aspx?id=56567>

For Mac

<https://docs.microsoft.com/en-us/sql/connect/odbc/linux-mac/install-microsoft-odbc-driver-sql-server-macos?view=sql-server-ver15>

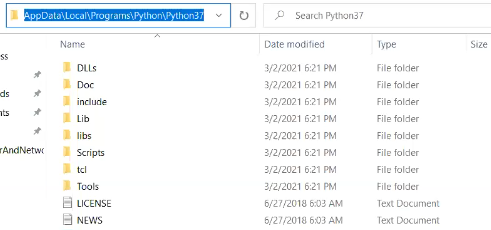
[00:15:27]

* Press CTRL+SHIFT+P and search for python:select interpreter:

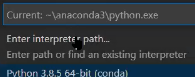


[00:20:38]

* Copy the path of the folder where python is installed:



* Then in VS Code, select Enter interpreter path and paste the copied path there:

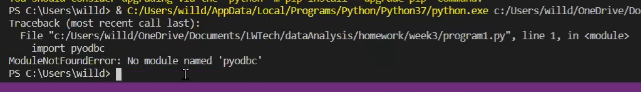


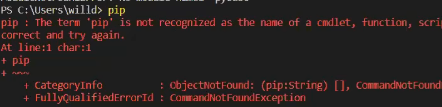




[00:21:39]

* Now try running the code:

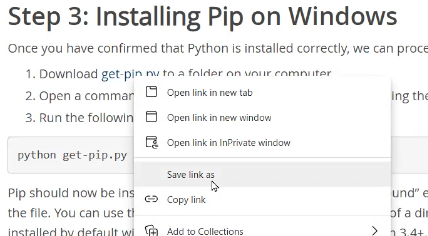




[00:25:40]

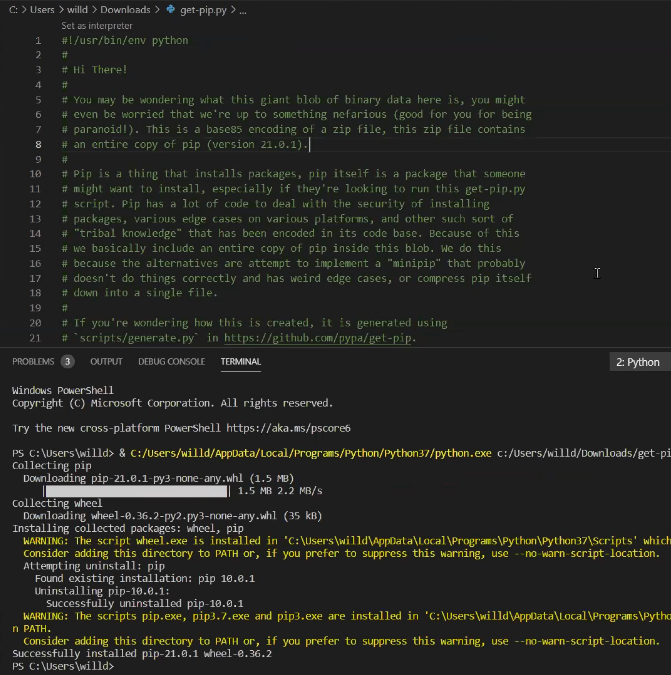
Get-Pip

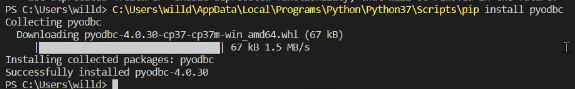
<https://www.liquidweb.com/kb/install-pip-windows/>



[00:30:57]

* Open new file in VS Code and paste content of get pip in that and then run it:





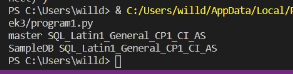
[00:36:56]

ODBC Driver

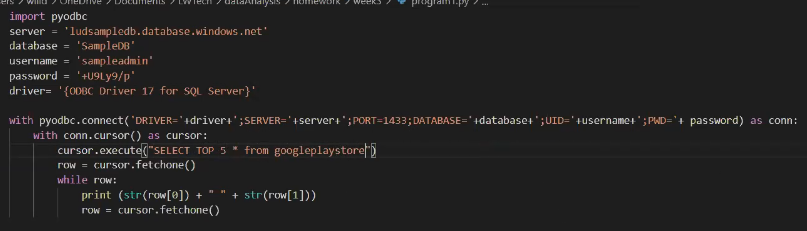
<https://www.microsoft.com/en-us/download/details.aspx?id=56567>



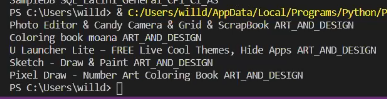
* After installing, run the code again:



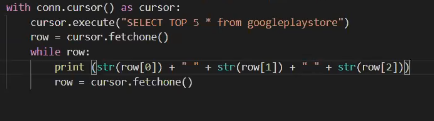
[00:38:43]



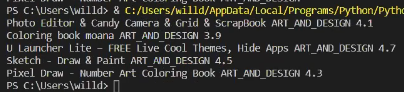
Output:

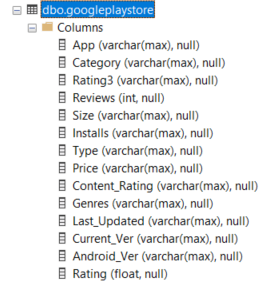


* Add str(row[2]) to get another column in output:



Output:





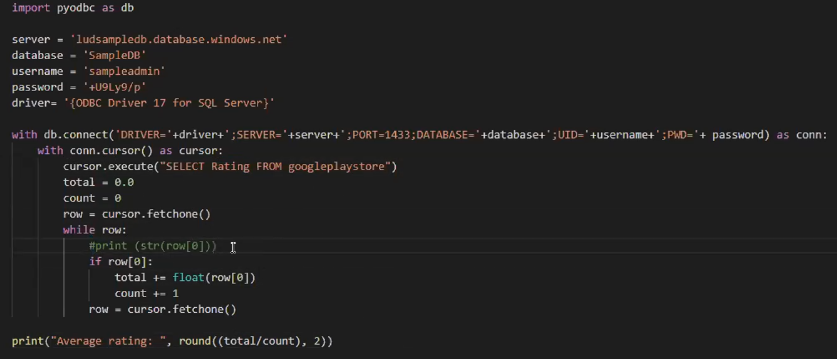
[00:41:18]

**Problem 1**:

Average Rating (float, null) NOT Rating3, round to 2 digits

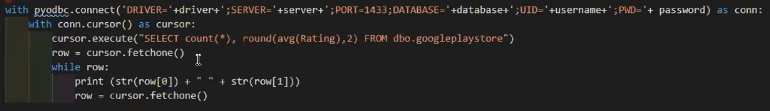
Across all the apps (all the rows).

[01:12:01]





[01:14:57]





[01:15:41]

* When doing Data Analysis, if you can do SQL do SQL only for the parts you can’t do SQL do Python.
* For a linear regression SQL can technically do it with a lot SQL (pick python)

[01:17:53]

<https://stackoverflow.com/questions/39835770/read-data-from-pyodbc-to-pandas>

How to turn your SQL result into pandas df:

data = pandas.read\_sql(sql,cnxn)`

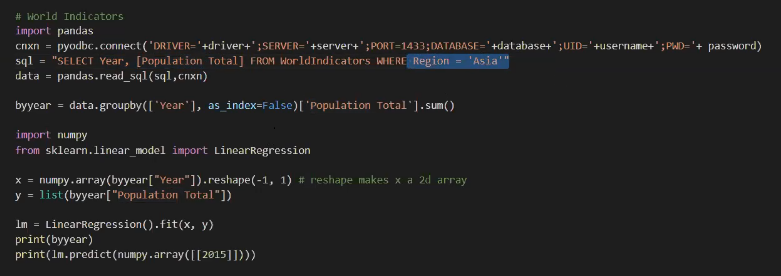
[01:35:53]

****

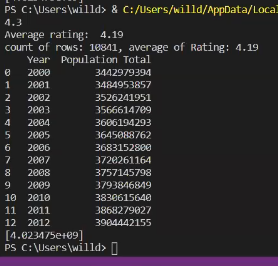
**Problem 2**:

Use a combination of SQL and Python to predict the population of Asia in 2015 based on the data on the table.

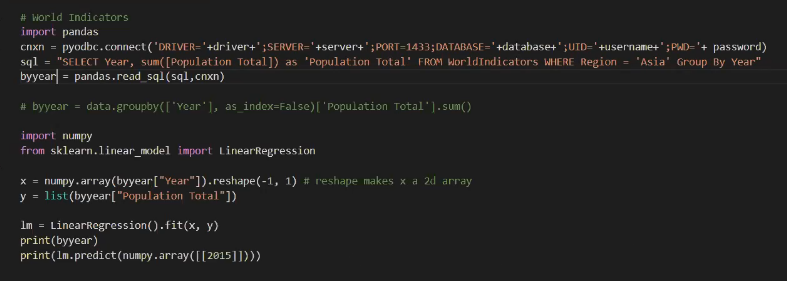
[02:11:05]



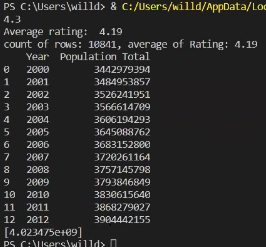
Output:



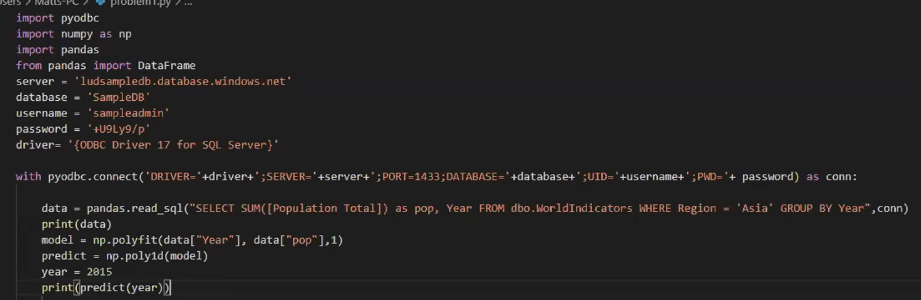
Using group by:



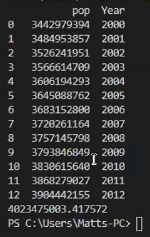
Output:



[02:16:22]



Output:



[02:17:12]

Data from SQL, use as much SQL as possible to prepare the data. Then use python to do complex statistics calculations