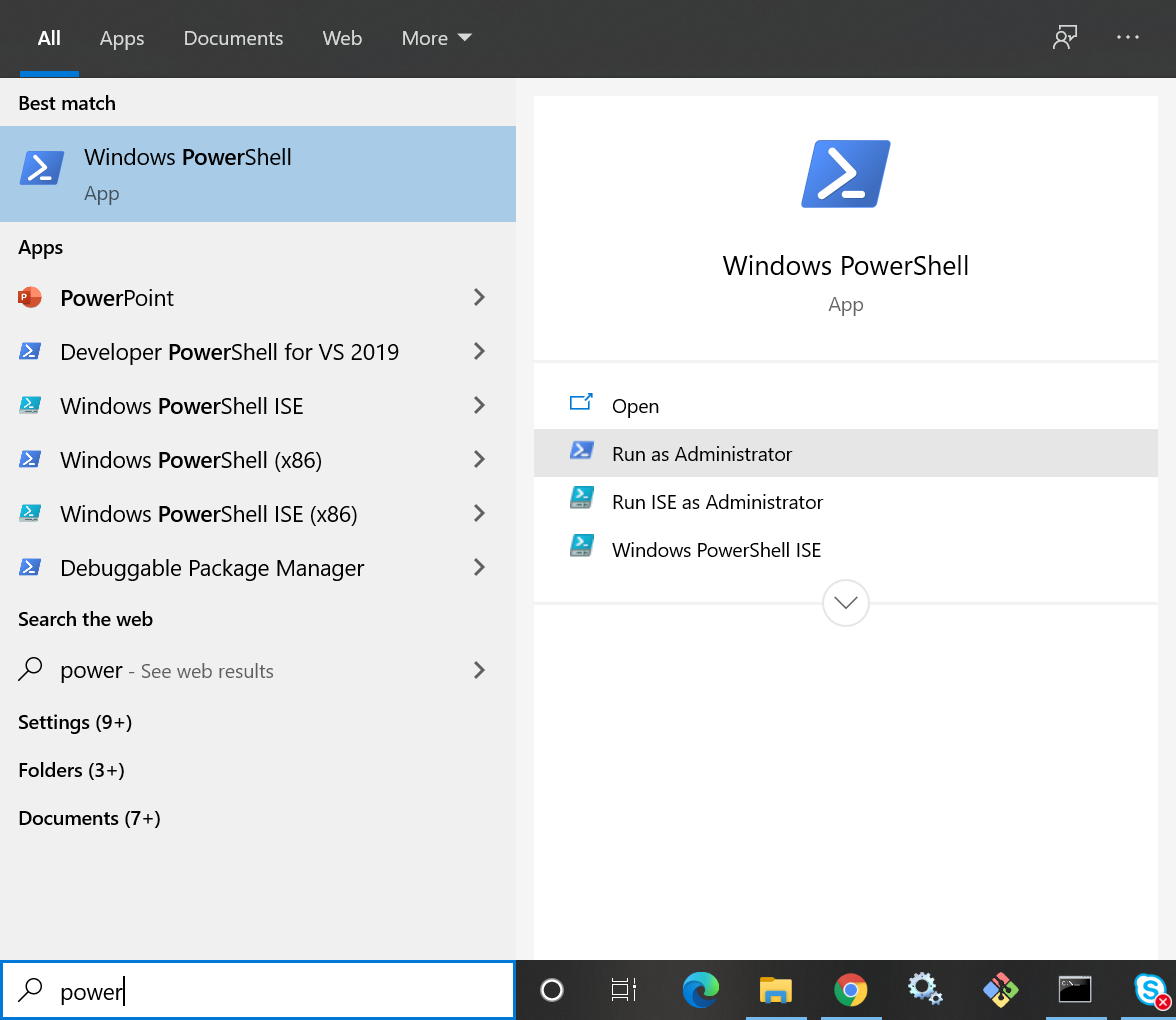
Short intro to Python: <http://do1.dr-chuck.com/pythonlearn/EN_us/pythonlearn.pdf>

**SET UP**

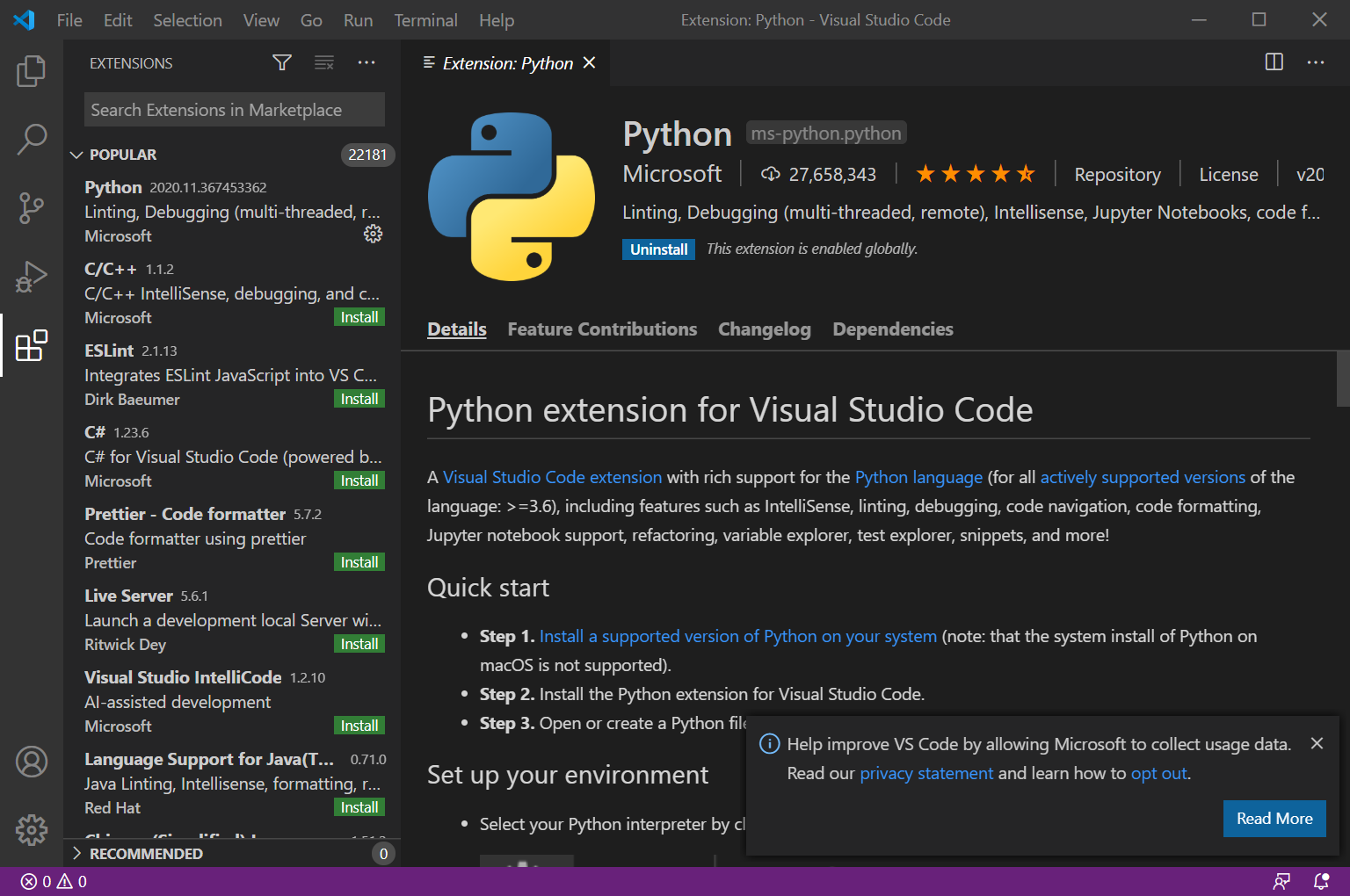
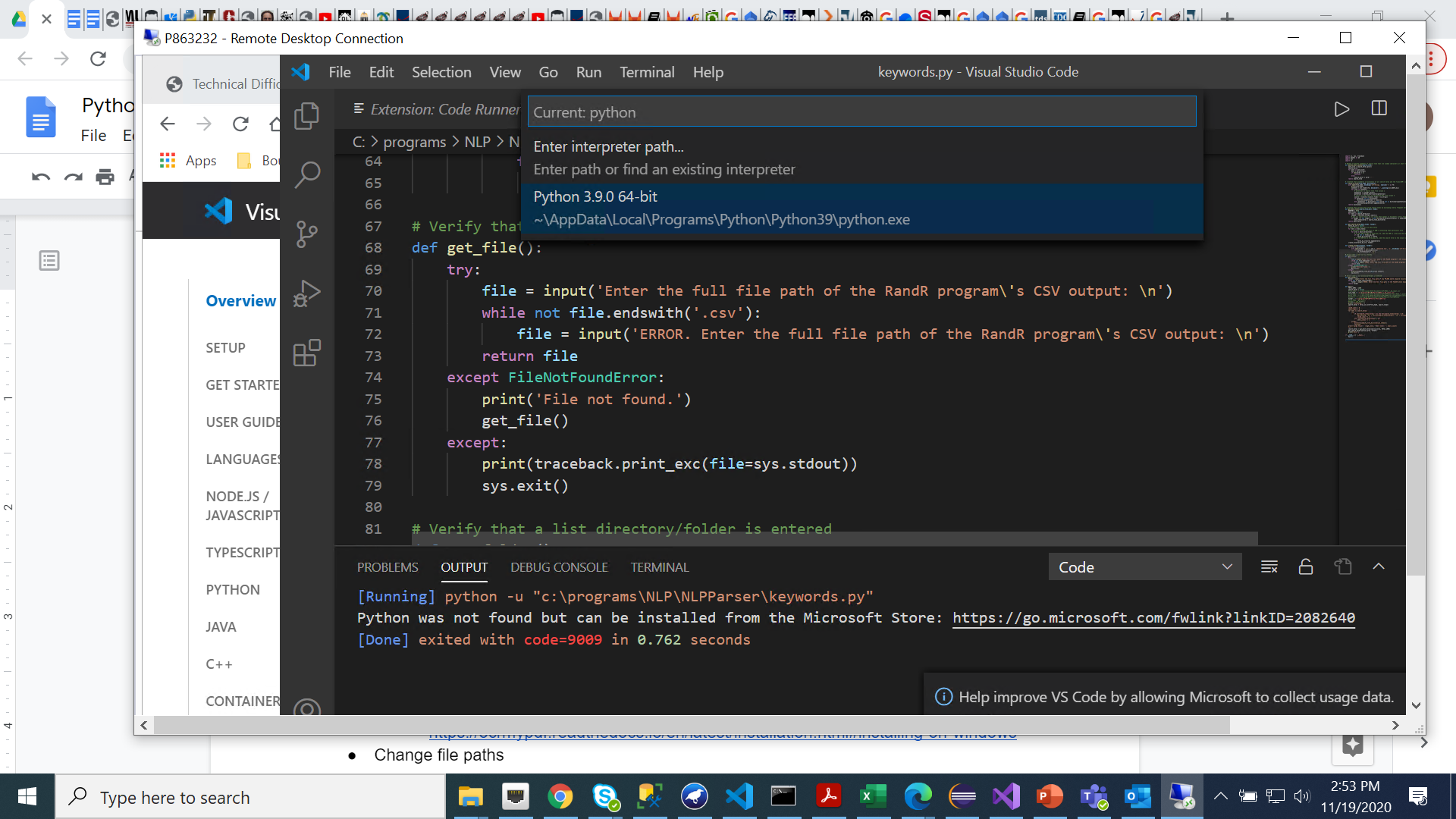
**Install Python/Chocolatey package manager:**

* Go to Windows Powershell → Run as Administrator



* Install Chocolatey package manager following the instructions here: <https://chocolatey.org/install>
* Install necessary packages using these instructions: <https://ocrmypdf.readthedocs.io/en/latest/installation.html#installing-on-windows>
* In the same location/from the Powershell (make sure to exit Powershell in order to use newly installed packages), pip install additional packages that will be used in the program:
  + pip3 install pandas
  + pip3 install tika
  + pip3 install docx2txt
  + If you get a ModuleNotFoundError for anything when you run the program, simply type pip3 install [moduleName] into the Powershell (or command line) to install whatever is missing

**Download** [**VS Code**](https://code.visualstudio.com/download) **or another code editor/IDE that will run Python**

* Click on .exe file.
* Click Run.
* Accept the license agreement & click Next.
* Select Installation location and click Next.
* Choose name of Start Menu folder and click next (the default is fine).
* Select desired options and click Install.
* Add Python and Code Runner Extensions to VS Code
  + 
* Click on the VS Code option to select Python interpreter. Select version 3.9.
  + 
* You should then be prompted to install Linter pylint. Click install.

**To run NLPParser Program, see Google Doc called “Run R&R Parser”. It explains how to:**

* Convert PDF’s & .DOCX files to .TXT
* Parse TXT files
* Edit RandR output to prepare it to be entered into the database
* Run Keywords program that generates lists of keywords for each SRM