University of Belgrade
Faculty of Mathematics
Departament of Astronomy
MASS - Introduction to AGN

SDSS Galaxies: Learning SQL

Tutorial #1

Nicolás Guerra-Varas
Professor Dragana Ilić
Tutor Isidora Jankov
25 de marzo de 2023

The goal of this tutorial was to get familiarised with the Sloan Digital Sky Survey (SDSS) data and learn how to access it online.

a) Initial SQL Query

First, I selected the galaxies and quasars I needed with the following SQL query in the search tool of the Data Release 18 (DR18) of the SDSS [1]:

Code 1: Initial SQL Query

```
1  SELECT s.z AS redshift, g.h_beta_eqw, g.h_beta_flux,
2       (g.oiii_5007_flux / h_beta_flux),
3       (g.h_beta_flux / h_gamma_flux),
4       (g.oiii_5007_flux / h_gamma_flux),
5       g.spectofiber
6       --count(*)
7
8  FROM GalSpecLine AS g
9  JOIN SpecObj AS s ON
10       g.specobjid = s.specobjid
11
12 WHERE
13       (s.class = 'QSO' or s.class = 'GALAXY')
14       AND s.z between 0.05 and 0.30
15       AND s.snmedian_g > 35
16       AND (g.oiii_5007_flux / g.oiii_5007_flux_err) > 1
17       AND (g.h_beta_flux / g.h_beta_flux_err) > 1
18       AND (g.h_gamma_flux / g.h_gamma_flux_err) > 1
19       AND g.oiii_5007_flux <> 0
20       AND g.h_beta_flux <> 0
21       AND g.h_gamma_flux <> 0
22       AND 2.355 * g.sigma_balmer > 1000
```

I selected the g filter for the condition on the signal-to-noise because the Hα line lies in its wavelength range (line 14 in 1). To ensure that the spectral lines [OIII] (5007 Å), Hβ (4863 Å) and Hγ (4341 Å) were present, I created two conditions for the flux of the line: it must be different than 0 (lines 18, 19, 20); and it must be larger than it's error (lines 15, 16, 17). Finally, to make sure that the Full Width Half Maximum (FWHM) of the Hβ line was larger than 1000 km s^{-1}, I calculated it with [2]:

$$\mathrm{FWHM} = 2\sqrt{2\ln 2} \times \sigma \approx 2.355 \times \sigma, \tag{1}$$

where σ is the standard deviation. In the case of spectral lines, this is measured by the velocity dispersion. I obtained this from the `sigma_balmer` column of the `GalSpecLine` table.

b) Number of Objects

This query resulted in 217 objects in total (see the `sdss_sql_HW1a.csv` file).

To find out which condition discards the most objects, I ran the query 1 with only `count(*)` in the `SELECT` line eight different times: once with no imposed conditions, and then with only one of the conditions activated at a time (class, redshift, signal to noise, [OIII], Hβ, Hγ, FWHM). The search that resulted in the least objects and thus discards the most objects was when the condition on the signal to noise was activated (90822 objects).

To double check this, I ran search an additional seven times, each time deactivating one condition but leaving the rest active. I got the most objects when deactivating the condition on the signal to noise (12286 objects), indicating once again that this is the condition that narrows the results down the most.

c) AGN Sub-class

According to the details of the tables indicated in the schema browser [3], the column `class` in the `SpecObj` table, which indicates the spectroscopic class of the objects, only has three possible classes: quasars, galaxies or stars. However, there is another column called `subClass`. Thus, I ran a new query with the same conditions as the initial query 1, but now only requesting `s.class` and `s.subClass` the columns (`SELECT` line). I downloaded the result of this query (see the `sdss_sql_HW1_subclasses.csv` file) and used `Python` to quickly check if there are any sub-classes (see File `subclasses.py`). There are three AGN subclasses. Out of the 217 objects, 141 are quasars, which have the following subclasses: starburst broadline (70 objects), broadline (61 objects) and AGN broadline (10).

d) Cross ID

Next, I searched for spectra in the object cross-id tool [4]. For this, I used the `287-plate-MJD-fiberID.txt` list that was provided and ran the following SQL query:

Code 2: SQL Query for Cross-ID

```sql
SELECT s.z AS redshift, g.h_beta_eqw, g.h_beta_flux,
     (g.oiii_5007_flux / h_beta_flux),
     (g.h_beta_flux / h_gamma_flux),
     (g.oiii_5007_flux / h_gamma_flux),
     g.spectofiber,
     s.plate, s.mjd, s.fiberID
     --count(*)

FROM #upload u
      JOIN
      SpecObjAll AS s
```

```
12      ON (s.plate=u.up_plate AND s.mjd=u.up_mjd AND s.fiberID=u.up_fiber)
13      JOIN
14      GalSpecLine AS g
15      ON g.specobjid = s.specobjid
16
17  WHERE
18      (s.class = "QSO" or s.class = "GALAXY")
19      AND s.z between 0.05 and 0.60
20      AND s.snmedian_g > 35
21      AND (g.oiii_5007_flux / g.oiii_5007_flux_err) > 1
22      AND (g.h_beta_flux / g.h_beta_flux_err) > 1
23      AND (g.h_gamma_flux / g.h_gamma_flux_err) > 1
24      AND g.oiii_5007_flux <> 0
25      AND g.h_beta_flux <> 0
26      AND g.h_gamma_flux <> 0
27      AND 2.355 * g.sigma_balmer > 1000
```

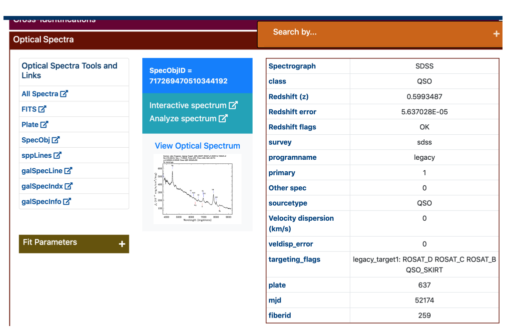
This version of the query has three main differences from the one in part a 1:

1. A syntax difference is that to indicate the use of strings, only double are accepted. So I changed ' to" where needed.
2. More importantly, this query required an additional `JOIN` in the `FROM` prompt. The main table where objects were selected from is the uploaded `.txt` file, which in the query appears as `#upload u`. Only then it can be joined with the other two tables.
3. Finally, the sixth line was added. This enables a button to appear in the query search result page, which directs to the Science Archive Service (SAS) [5].

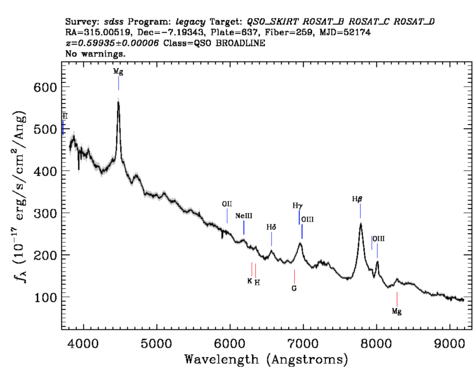
Note that the twelfth line was generated automatically.
This query resulted in 47 objects (see the `sdss_sql_HW1_crossid_pmf.csv` file).

e) Checking Spectra

I checked the spectra of one of the objects in the Explore tool [6] (see Figure 1).



(a) Screenshot of the Explore tool website.



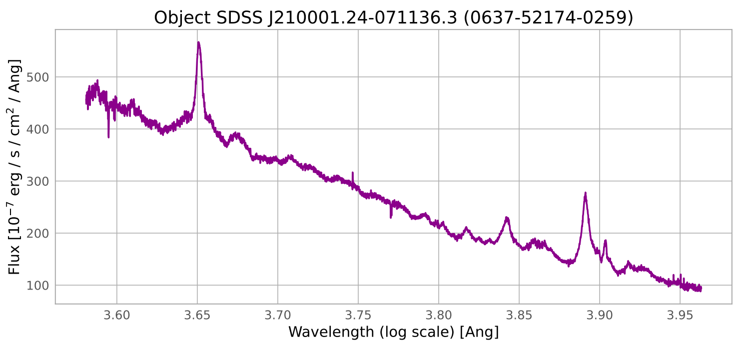
(b) Screenshot of the image of the spectrum shown in the website.

Figure 1: Spectra of the object plate-mjd-fiberID 637-52174-259 shown in the Explore tool.

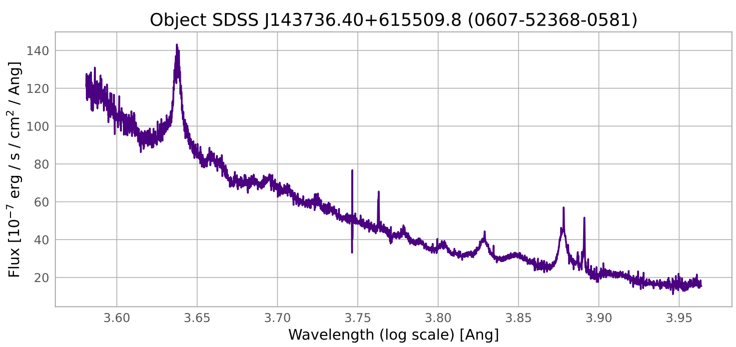
Then, I downloaded them from the SAS using `wget` (see the `spectra.zip` file). This required many tries since the server was overloaded.

f) Plotting Spectra

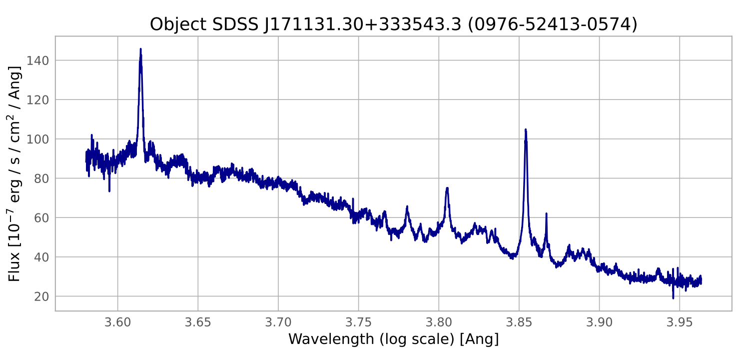
Finally, I used the `astropy` package to read the `.fits` files and plot the spectra of three objects. The code I used is in the `plotting_spectra.ipynb` file. The spectra I plotted are shown in Figure **??**. I randomly selected these three objects, and it coincided that they are all broadline quasars. So their spectra are noticeably similar. In addition, the first object is the same one as in Figure 1.



(a) Spectra of the object SDSS J210001.24-071136.3



(b) Spectra of the object SDSS J143736.40+615509.8



(c) Spectra of the object SDSS J171131.30+333543.3

Figure 2: Plotted spectra of three objects.

References

[1] “Sdss dr18 skyserver.” https://skyserver.sdss.org/dr18, Date of access: March 2023.

[2] “Gaussian function, wolfram alpha.” https://mathworld.wolfram.com/GaussianFunction.html, Date of access: March 2023.

[3] “Schema browser.” https://skyserver.sdss.org/dr18/MoreTools/browser#, Date of access: March 2023.

[4] “Object crossid.” https://skyserver.sdss.org/dr18/CrossMatchTools/ObjectCrossID, Date of access: March 2023.

[5] “Science archive service.” https://dr18.sdss.org/home, Date of access: March 2023.

[6] “Explore tool.” https://skyserver.sdss.org/dr18/VisualTools/explore/summary, Date of access: March 2023.