**Project Group:** Nupur Harmalkar

**Literature review (including citation), identifying the gap in the literature**

Sloan Digital Sky Survey current [DR16](https://www.sdss.org/dr16) Server Data release with Galaxies, Stars and Quasars. The Sloan Digital Sky Survey has created the most detailed three-dimensional maps of the Universe ever made, with deep multi-color images of one third of the sky, and spectra for more than three million astronomical objects. The Sloan Digital Sky Survey has been working for more than 20 years to make a map of the Universe and will continue for many years to come.

Measurements of large-scale structure in SDSS maps of galaxies, quasars, and intergalactic gas have become a central pillar of the standard cosmological model that describes the understanding of the history and future of the Universe. SDSS data have helped to demonstrate that the Universe is dominated by unseen dark matter and pervasive dark energy and seeded with structure by quantum fluctuations in the infant cosmos. Those fluctuations have grown into the large-scale structure we see today.

The table results from a query which joins two tables:

* "*PhotoObj*" which contains photometric data
* "*SpecObj*" which contains spectral data.

There are Total 17 variables and 1 additional variable (char) 'class'.  
A class object can be predicted from the other 17 variables.

**Variables description:**  
**objid** = Object Identifier  
**ra** = J2000 Right Ascension (r-band)  
**dec** = J2000 Declination (r-band)  
**u** = better of deV/Exp magnitude fit (u-band)  
**g** = better of deV/Exp magnitude fit (g-band)  
**r** = better of deV/Exp magnitude fit (r-band)  
**i** = better of deV/Exp magnitude fit (i-band)  
**z** = better of deV/Exp magnitude fit (z-band)  
**run** = Run Number  
**rerun** = Rerun Number  
**camcol** = Camera column  
**field** = Field number  
**specobjid** = Object Identifier  
**class** = object class (galaxy, star or quasar object)  
**redshift** = Final Redshift  
**plate** = plate number  
**mjd** = MJD of observation  
**fiberid** = fiberID

A four-color UVGR intermediate-band photometric system (Thuan-Gunn astronomic magnitude system) is discussed in [1].

The Sloan Digital Sky Survey (SDSS) photometric system, a new five-color (u′ g′ r′ i′ z′) wide-band CCD system is described in [2]

The variables 'run', 'rerun', 'camcol' and 'field' features which describe a field within an image taken by the SDSS. A field is basically a part of the entire image corresponding to 2048 by 1489 pixels. A field can be identified by: - run number, which identifies the specific scan, - the camera column, or "camcol," a number from 1 to 6, identifying the scanline within the run, and the field number. The field number typically starts at 11 (after an initial rampup time) and can be as large as 800 for particularly long runs. - An additional number, rerun, specifies how the image was processed.

➢ **Research questions, and the significance of the study**

We will be classifying the class object based on 17 variables to predict whether the celestial object is a star or a galaxy or a quasar.

With the help of our study we will try to obtain results and insights about this classification, what photometric system results to a star, a galaxy or a quasar.

The study will help to rationalize the difference between the three classification objects.

***➢* Descriptive statistics, and visualizations**

* Exploratory Data Analysis is done in Python with Pandas, Matplotlib and NumPy.
* We have also used Tableau for some visualization.

***➢* References**

* Searching for Data: A Tutorial. (n.d.). Retrieved from http://skyserver.sdss.org/dr16/en/help/howto/search/searchhowtohome.aspx [1]
* Thuan, T. X., & Gunn, J. E. (2016, May 11). IOPscience. Retrieved from https://iopscience.iop.org/article/10.1086/129982 [2]
* Talking Photometry: Understanding Photometric Data Formats. (n.d.). Retrieved from http://www.photometrictesting.co.uk/File/understanding\_photometric\_data\_files.php