**Cosmos (Or How I Learned to Love the Night Sky)**

Space. The final frontier. These are the stars that top scientific minds have concluded that exoplanets are orbiting around them. Using this database, students, teachers, amateur star gazers, stoners who watch too much Science Channel, and parents who want to get their children excited about the great unknown can explore data and prepare themselves to take that one small step for a man, one giant leap for mankind.

Over the last decade or two, scientists have gone from theorizing that exoplanets - that is, planets that orbit suns not our own - exist, to theorizing that not only do they exist, but it’s likely that every star in the sky has planets orbiting them. The methods they use to locate and identify these strange new worlds vary, from observing dips in the star’s light as the body crosses in front of its parent star to measuring the gravitational impact that an orbiting body has on it’s parent star, but the end result is the same: a long list of new, unexplored worlds to dazzle and inspire the imaginations of generations to come.

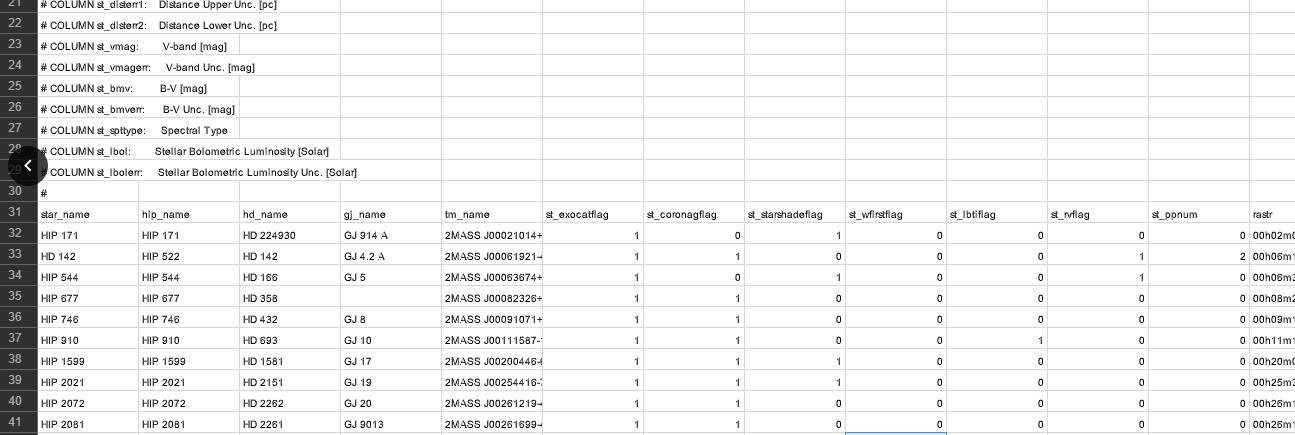
There is just one issue, and that is the availability of the data specifically for an audience that is looking for an introduction to extrasolar solar systems and space exploration. For our project, Group Cosmic Degenerates decided to step in and fulfill this need.

Before we begin, we did want to acknowledge one major limitation that we encountered at the outset of our project and it’s one that may surprise you. From the priests of Ancient Egypt to the Oracles at Delphi to the Druids of Celtic Britain, people have been cataloguing and observing the locations and movements of the celestial bodies. There is literally thousands of years of data from civilizations all across the globe to parse through associated with the stars in the sky. There is, undoubtedly, a plethora of information we could have included given more time to thoroughly explore the data and study some of the more technical aspects that scientists have developed to measure stellar properties, however we feel that this is a great starter database for cadets of StarFleet.

Without further ado, our ETL Project: “Cosmos: Or How I Learned to Love The Sky”.

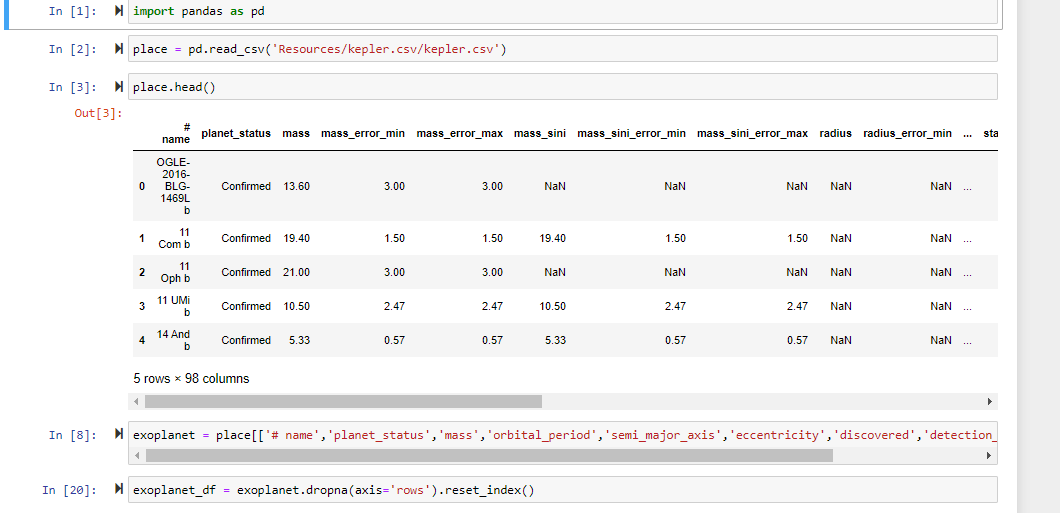
**Extract**

There are numerous sources out there for information regarding both exoplanets and the stars in the sky. We decided to use NASA’s Exoplanet API (<https://exoplanetarchive.ipac.caltech.edu/docs/data.html>), which describes what stars astronomers have focused on in their search for extrasolar worlds and gives a lot of detailed information about the star including position in the night sky, luminosity, and star type. From Kaggle, we were able to find Kepler data regarding exoplanets and their properties. The Kepler data is as surefire source as the NASA data, as the Kepler mission’s primary goal is to discover and ascertain information about exoplanets. The data was available in many types including csv, json, and others. We went with csv files.



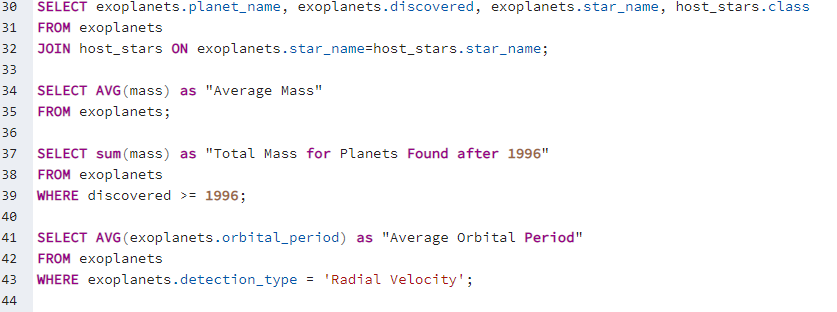
**Transform**

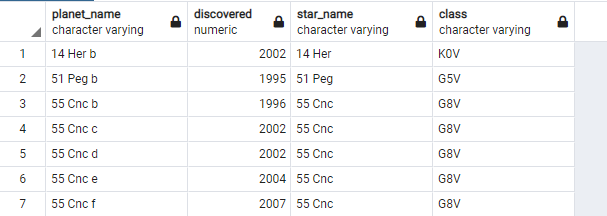
Once we had our datasets, we began the process of getting the data ready for queries. This involved removing rows with null or incomplete values, filtering out columns that we felt didn’t add anything meaningful to the conversation, and determining keys that would be used to relate our cleaned exoplanet table to our cleaned star table. We also changed the datatypes of the values in columns so our target users will be able perform aggregate queries when necessary.



**Load**

We had our sources from which we collected our raw data, we transformed that data into usable things, now it was time to create our database. We opted to use SQL for a few reasons. First and foremost, it is - as far as we know, at least - unparalleled when it comes to ease of use and performing relational data queries. Using our database “Cosmos”, students, teachers, and anyone with an inkling to explore the vast poisonous radioactive wasteland that is outer space and look for diamonds in that rough. For more information, the included ReadMe file has a more detailed breakdown of the database. Below are a few example queries that students can run while exploring the database. One provides the name of the exoplanet along with when it was discovered and it’s sun’s stellar type. The other is the average mass of all discovered exoplanets after 1996.



****

**Finitione**

There is little to no doubt that mankind’s trajectory into the future lies above us. Using the Cosmos database we created, anyone from young students and teachers to old people near death will be able to query aspects of both exoplanets and the suns that rise in their sky. The only question left is “Where to start?”, to which our answer is “Second star on the right, and straight on ‘til morning.”

