**IST 718 Big Data Analytics Check-In**

**By Laura C. Larregui & Tiffannie Mac Donald**

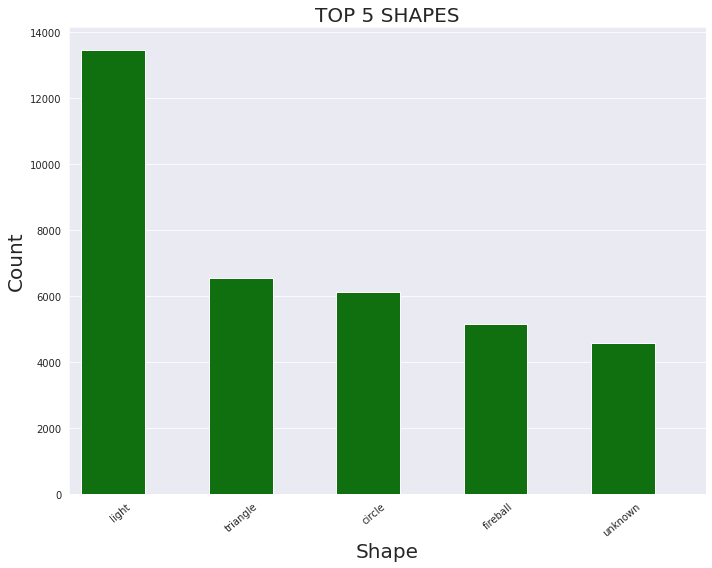
# Specification

The purpose of our project is to observe any patterns of UFO reports and be able to make predictions as to when the next may occur.

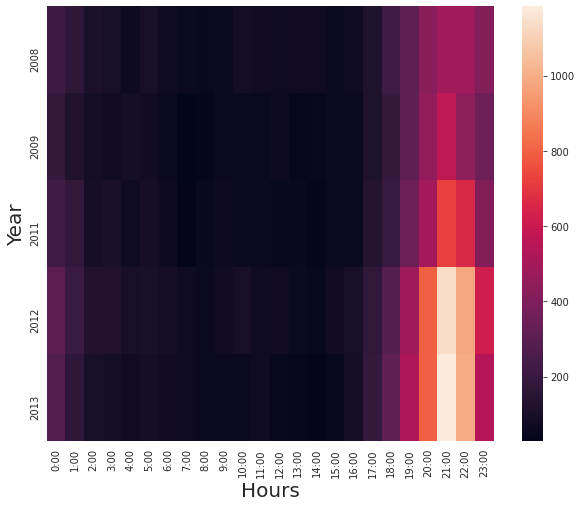
# Observation

So far, we have noticed that:

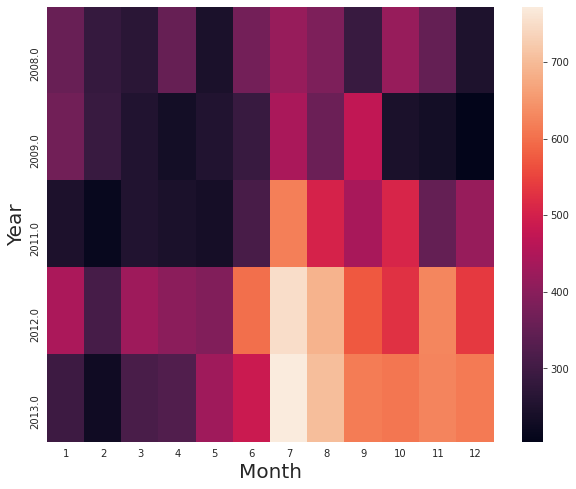
1. Light is the highest reported shaped among all UFO sightings (see **Figure 1**).
2. The years 2008, 2009, 2011, 2012, and 2013 have the highest quantity of reports.
3. Most UFO sightings occur late at night and between July and December ( see **Figure 2 and 3**).
4. California is the state that has the most UFO reports (see **Figure 4**).



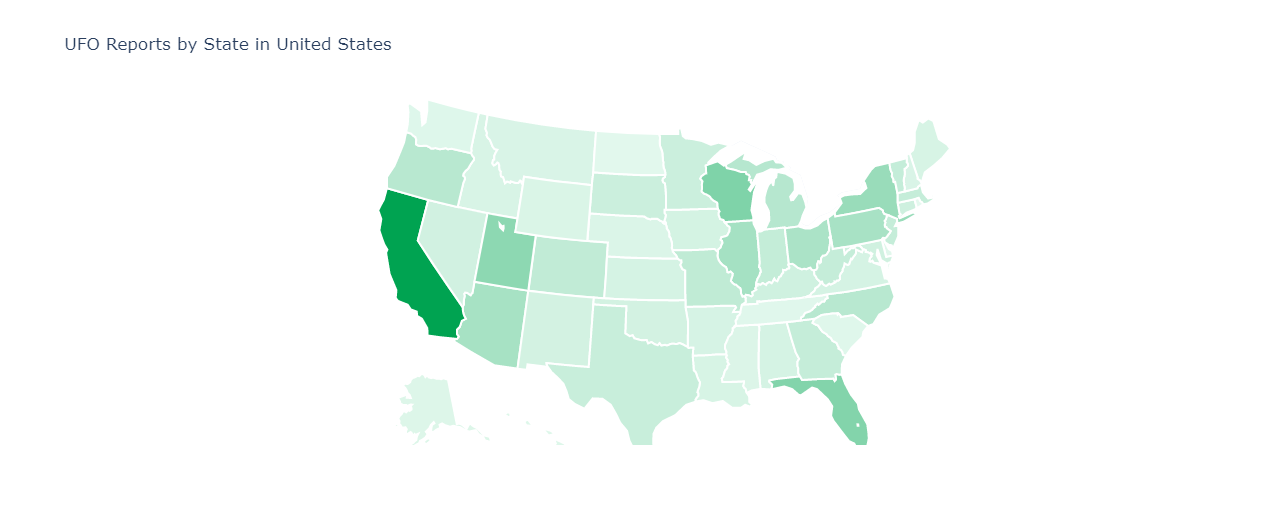
**Figure 1.** Barchart of the top 5 shapes identified at UFO sightings.



**Figure 2.** Heatmap of the hours with most UFO sightings (years with the highest encounters)



**Figure 3.** Heatmap of the months with most UFO sightings (years with the highest encounters)



**Figure 4.** Map of UFO reports by state (U.S.)

Analysis

We hope to use Association Rule Mining, Random Forest, ARIMA, and K-Means to put the pieces together and paint a better picture as to whether there is a pattern with UFO sightings.

# Recommendation

We have already completed most of the exploratory data analysis phase. We hope to move on to the modeling part of the project.