School of Information Studies
SYRACUSE UNIVERSITY

**IST 718 Big Data Analytics Project Proposal**
**By Laura C. Larregui & Tiffannie Mac Donald**

# Background

While there is an open debate about the existence of extraterrestrial life, the study of UFOs is important to non-believers and believers alike. The study of UFO sightings help us understand the unknown and unusual occurrences of the world.[1]

# Questions

- What areas of the country (USA) are more likely to have UFO sightings?
- Are there any noticeable trends over time?
- Do UFO sightings tend to increase when alien movies come out?
- Are UFO sightings more prevalent near airports?

# Data

The main dataset used in this project is from Kaggle, but it originated from the National UFO Reporting Center (NUFORC). The dataset was scraped, geolocated, and time standardized by Sigmond Axel. The dataset includes 11 attributes and 80,332. Since the project will only focus in UFO sightings in the USA, the dataset was decreased 65,114 records.

Data Source: NUFORC (2019). UFO Sightings (Version 2) [CSV file]. Retrieved October 21, 2020 from https://www.kaggle.com/NUFORC/ufo-sightings/metadata.

Data regarding movie date releases will also be collected from the TMDb API. The Movie Database (TMDb) is a community built movie and TV database. Every piece of data has been added by the movie lover community dating back to 2008.

Data Source: https://www.themoviedb.org/documentation/api

[1] Disclaimer: The members of this project do not seek to prove the existence of extraterrestrial life or supernatural creatures.

The locations of airports will be collected from the OpenFlights Airports Database. This database contains over 10,000 airports, train stations, and ferry terminals spanning the globe.

Data Source:  https://openflights.org/data.html

# Initial Strategies

- *Exploratory Data Analysis*: (histograms, barplots, treemaps)
- *Correlation Matrix*: Which predictor variables go together?
- *Text Mining*
- Any model discussed in class that seems relevant to the project.