* Each group must produce three interesting facts they discovered in the data through their analysis.
* Each group must write a short report of their investigation (the fun facts and the approach they followed to achieve these results).
  + The report is either a Word file or a PDF file.

The cult of Sarria

Our group decided to analyze the ufo.csv file. The first thing we had to do was clean up our data set and make it more understandable. Another thing we decided to do was convert the data posted from a string to a date time variable. This was done by using the pd.to\_datetime function on the Date posted column. Before we could analyze our data, we decided to modulate the data to be more concise. This was done by converting the strings to floats regarding the seconds variable. We did this so the duration is not a solid integer.

A cluster of questions we decided to answer using the data was: What city had the most sightings, Average Duration of sighting, what times, months, years were sightings popular? Most common Shape? Any change over the years?

Application

Description automatically generated with low confidence

The most common sighting of UFO’s happened in Seattle Washington with 525 occurrences, 454 UFO sightings in Phoenix Arizona, and third most sightings happened in Portland Oregon. To find this information, we used the value\_counts() function and then used the .head(3) function on the most\_common\_cities. The value\_counts() function takes the most\_common\_cities variable and counts the number of inputs for each variable inside of the data frame. The .head(3) function visually shows the three most common outputs.

A picture containing chart

Description automatically generatedA picture containing table

Description automatically generatedThe most common shapes of sightings used the same functions. The only difference we used was the variable that pulled the data. The variable used was the most\_common\_shapes. The result of this analysis was that the most common was an undefined shape giving off light.

We then wanted to find out what was the average amount of time individuals viewed a sighting. So, we simply took the data set and grabbed the duration in seconds column and did .mean() to give us the average seconds. After finding the average seconds we wanted it into mins so we did a conversion by dividing it by 60.

Lastly, to figure out if there was a change over the years, we decided to make a time series graph that would show the years, and number of sightings seen. In order to make the graph we sorted the data into a new data frame. Next, we created a new variable named onlyyears. The variable groups the data frame by year, counts the sightings in years. Once the new data frame has been created, we created a time series using plt functions. On the graph we noticed there is weird data in 2014. This is caused by the data being incomplete in 2014 with months 1-5 only included.

