

UFO_Analysis_Report

To begin, within the global dataset, recording over 80,000 UFO sightings, the average encounter duration time was found to be approximately 9,017 seconds. Converting seconds to minutes, and minutes to hours, the average encounter duration was over two and a half hours. We found this to be interesting because from movies and shows, we would think a UFO encounter would be quick, ten minutes or less. In media, UFOs are depicted as stealthy and fast moving, stop to put on your shoes and you might miss the sighting; however, this data implies UFOs move in less stealthy, snail-like speeds.

Moreover, the most common word the witnesses used to describe the shape of the UFOs was “light”. Specifically, 15,656 witnesses recall the UFOs as being in the shape of light; of which, a whopping 1,983 were from California. To put that number in perspective, in 2012, California had an amassing official population of 37,944,551, almost 38 million. For a state that big, less than 2000 sightings of “light” is just a drop in the bucket; the witnesses of the sightings composed less than 1% of the state’s population.

Our last shocking observations are in regard to the number of sightings. The largest quantity of UFO sightings for a given year was recorded in 2012, for a global total of 7,357 sightings in a single calendar year. For the collective data, July is the month with more sightings than any other month. The total amount of sightings

throughout July (from all years) was 9,520 sightings. The second highest record was in the following month, August (from all years), with 8,636 sightings. We found this data interesting because 2012 was dubbed “The End of the World” in ancient Maya Prophecy; if ever there would be an Unidentified Flying Objects spike, it would be right before “The End of the World”. Also, with July being mid-summer in the US (where there were most of the sightings), it only makes sense for the most UFOs to be sighted during this month. Most people will be out with family or friends whether it is day or night which allows for more opportunity to sight the UFOs.

Our whole programmatic approach was to use pandas to amend our data frame or to use our current data frame to create new ones. Modifying the given data frame was a necessity to get the information that we needed. This included changing the format in which that date was entered into a ‘datetime’ format, modifying the ‘duration (seconds)’ column, creating a data frame focusing specifically on the ‘light’ shaped sightings and amending the given data frame adding a year column. Formatting the date was the simplest. We only had to write a few lines of code that allowed us to do so. Changing the ‘duration (seconds)’ column was the most challenging. It took us a while to figure out why `‘ufo_df[‘duration (seconds)’] = pd.to_numeric(ufo_df[‘duration (seconds)’])’` wasn’t initially working. After reading the error code we came to the conclusion that in a few of the entries of duration there were more than just numbers, so we then cleaned up that data changing those entries to their intended numbers and then after we were able to convert the column from an object to an integer. Making a dataframe that only consisted of data from sightings that included ‘light’ under the shape

column was possible using the function `'light_shape_df = ufo_df.loc[ufo_df['shape']=='light']` once that was done we could then use the `'value_counts()'` function after specifying that we wanted the count for the `'state'` column to figure out which state had the most `'light'` sightings. Finally we amending the original data frame to add a year category so that we could get a count for which year had the most sightings in the data frame this was possible by using the function `'ufo_df['year'] = pd.DatetimeIndex(ufo_df['datetime']).year'` and then using the `'value_counts()'` function specifying the column `'year'`. To collect our fun fact for the month in which there are most of the sightings, we decided to do a countplot through seaborn. We started out by plotting the date time month through `'ax = sns.countplot(df.datetime.dt.month)'` and continued by creating a title through `plt.title()` and named our plot "Ufo Sighting by months of year". We then gave our axis a label through `'ax.set(xlabel='Month of Year', ylabel='Sightings')'`. We then created a for loop in order to create the patches and have mock objects for our values. Lastly we wanted our graph to be larger so we used `'plt.figure(figsize=(18,12))'` to make it larger.