**Project report:**

Neta Ehrenberg - 213232176 & Shimon funaro - 06039805

**Overview**:

Our data focused on patients with epilepsy that underwent surgery for their condition.The dataframe had data of 443 patients. We had a column of data of what lobe the surgery took place, type of surgery, the pathology, age at first seizure and the severity of the seizures the patient had in the five years following the surgery, if at all, categorized by year. The data does have much more information but this was our main focus. We decided we wanted to find out if there is a certain parameter or parameters that could likely predict the outcome of the surgery.

**research question**:

How do certain parameters\*\*may affect the outcome of the surgery, meaning the severity of seizures at the five years following the surgery.

\*\* parameters: lobe, age of first seizure, pathology

**Cleaning:**

We needed our data to have a separate column for the “lobe” so we extracted those from a different column named “Op\_Type”, we created a new column called “lobe” and added it to the data. Our data had some patients with surgeries in more than one lobe so we went ahead and deleted them from our data. We had a column of data for each year after the surgery, some of the patients had a lot of data missing so we eliminated the patients that had over a year of data missing. What we decided to do for the patients that had just one year of missing data is to take the patient's result from the previous year and copy it instead. The ILEA categorizes the seizure from 1-5, 1 & 2 being “good” scores - meaning no seizures that year, and 3-5 being “bad” scores - meaning occurring seizures with growing severity - we decided to filter the data to “good” and “bad” by that. Deciding that each row is a good one if there are at least 4 years of 1 & 2 and a bad row if there are at least 4 years of 3-5. We were also interested in the age of the first seizure and whether it had anything to do with the outcome of the surgery so we took the column containing that data and had to make sure we could work with it, finding out the ages were given as so “15 to 19” (with some exceptions like “under 1” or “over 40”) we took care of that by taking the mean age of that group, and for the exceptions making sure it says 1 or 40 making sure to store the new data in a new column “mean\_age” and adding it to the dataframe.

**Analysis**:

Now that we had cleaned data we could start the analysis: we wanted to find out if there is any significance for the different parameters we tested between the “good” and “bad” data We decided to run a binomial test, a test that checks for significance between a “winning situation” - the good outcome to a “losing situation” - the bad outcome. For the analysis we used the filtered dataframe, after cleaning and the “good” and “bad” lists. We ran the test between the list of “good” and “bad” data to three different columns - “lobe”, “pathology” and “mean\_age”. We have the general analysis function that prints out a list to each specific group in a parameter that has a significant difference between its good and bad lists and gives us the P-value.

For example: the group “T” in the parameter “lobe” had a P-value of 0.0000000108.

**Testing:**

We tested each function in the analysis and the cleaning to make sure the functions are actually capable of doing what we want them to do and are able to catch weird values that might occur in the data.

**Ensuring data:**

We wanted to check our data to make sure all the columns had exactly what they are supposed to have with no additional things we and the testing might have missed. We have a function that print out all the unique values in each column that are important for our analysis:

lobe, pathology, mean age and all the years.

**Visualization:**

1. We decided to make some general pie graphs of the data - presenting the percent of each lobe in the different pathologies. (figure 1)
2. A bar graph that presents the “good” and “bad” lists in each different group in a parameter - good is shown in blue and bad in red (figure 2)
3. A dot graph that presents the p-values of each group in the three parameters. If its a significant value its in red else its in blue. (figure 3)

**Discussion:**

Our research question was wondering whether there are certain parameters that could perhaps predict to some level the outcome of an epilepsy surgery.

After cleaning up the data, deleting rows with too many NaN’s, making up for the missing year (if there were any) and testing the data to make sure it's all organized and makes sense for our analysis, we could move forward to the analysis - in this part we analyzed the data and ran function that tested for significance (p < 0.05) in all parameters that interested us. We found some significance in each of the categories we looked into, making them a candidate to actually have a role in the outcome of the surgery.

**Conclusion:**

We found some significance in the different categories we tested that might have a positive effect on the outcome of the surgery, here they are listed:

1. Lobe:

* T

1. Age of the first seizure (mean):

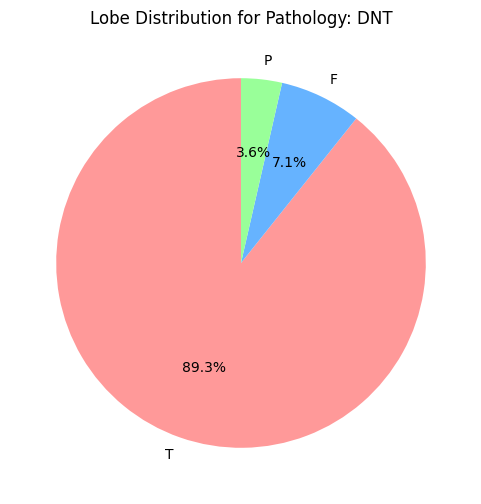
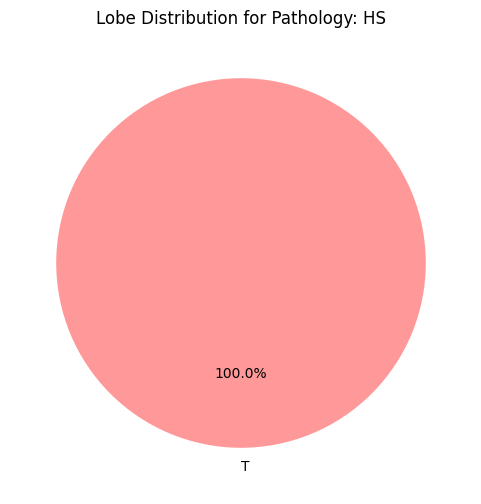
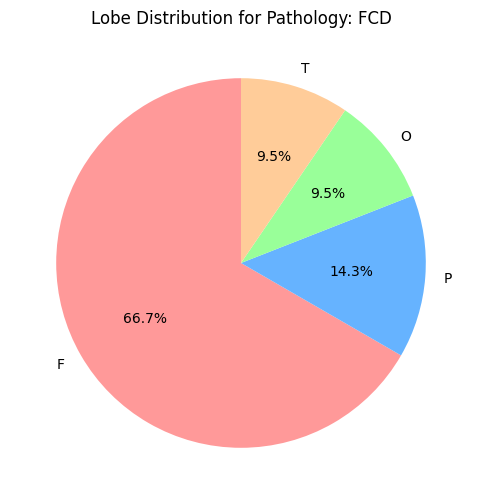
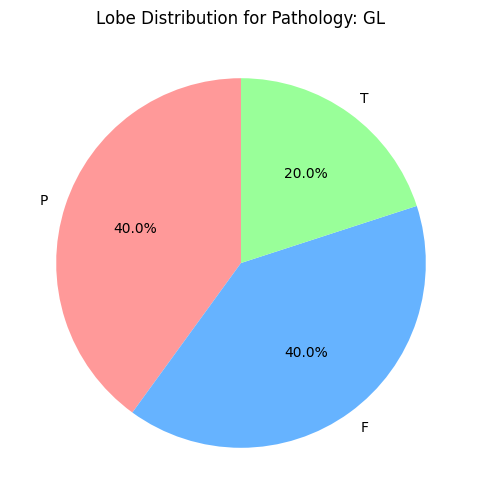
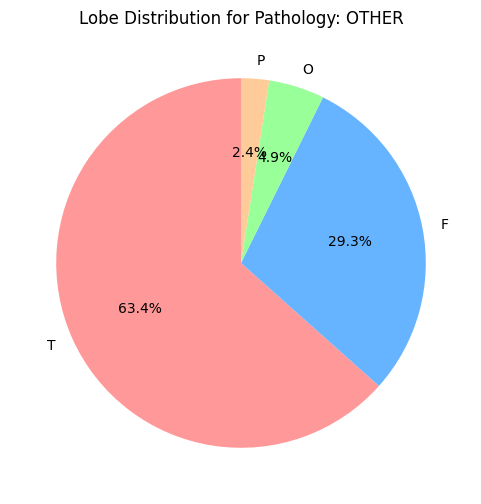
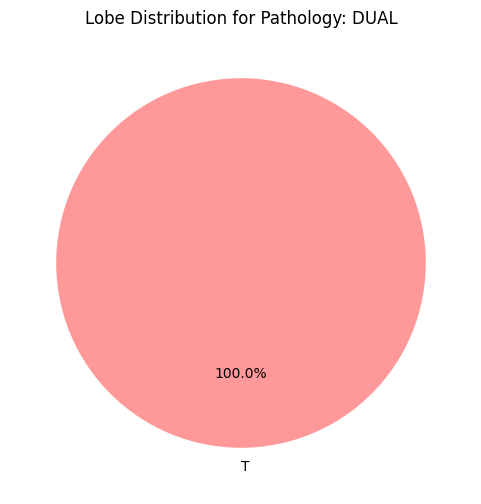
* 1.5 YO
* 3.5 YO
* 6 YO
* 22 YO

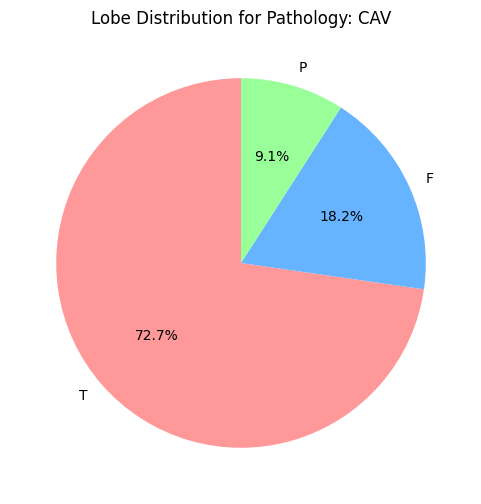
1. Pathology:

* CAV
* DUAL
* FCD
* HS

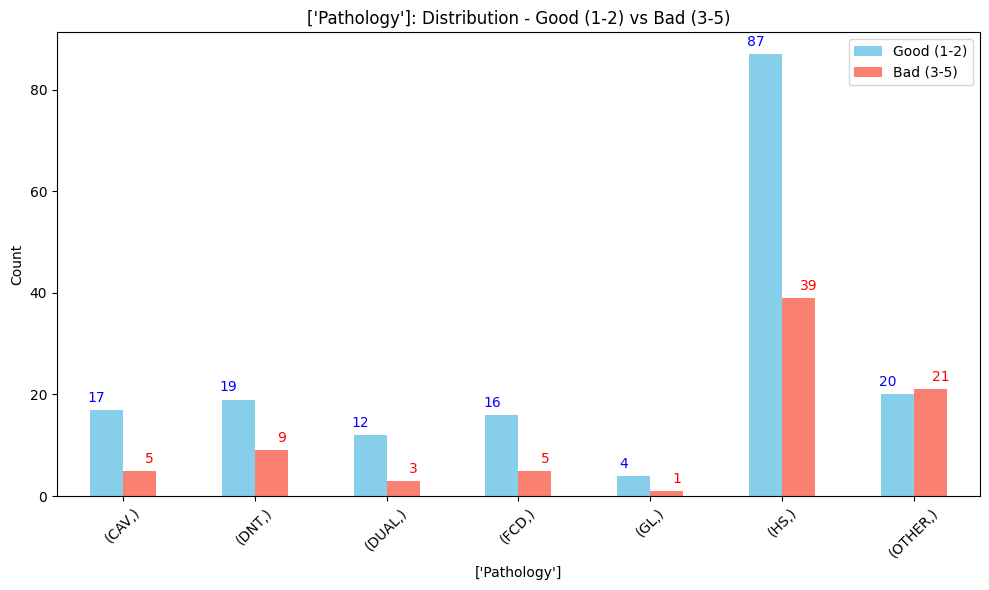
\*\* Due diligence, we did use the help of Google and AI in this project for things we did not learn or could not manage on our own but we did try to understand whatever was presented to us to the best of our ability.

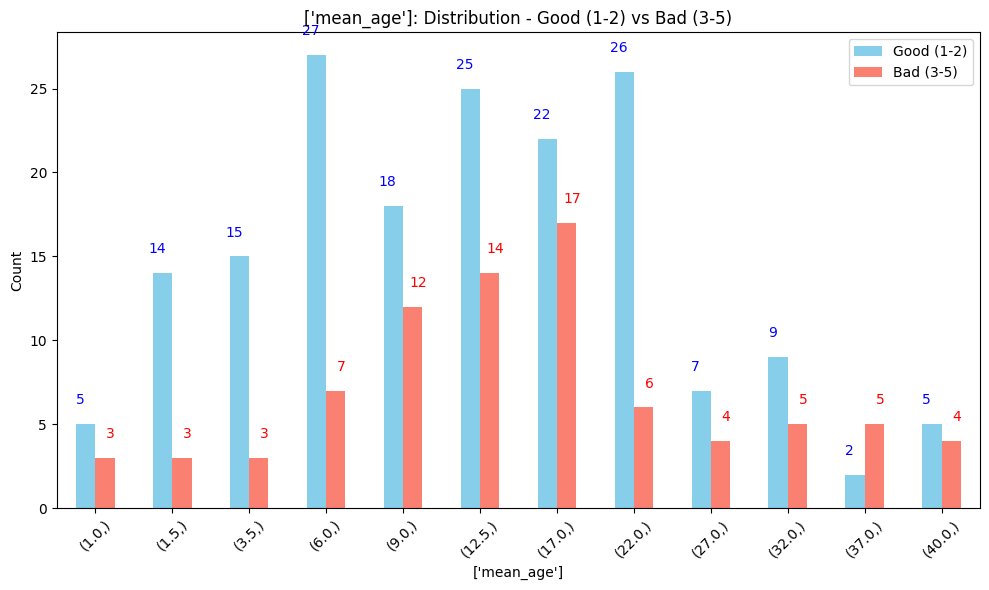
**Figure 1:**

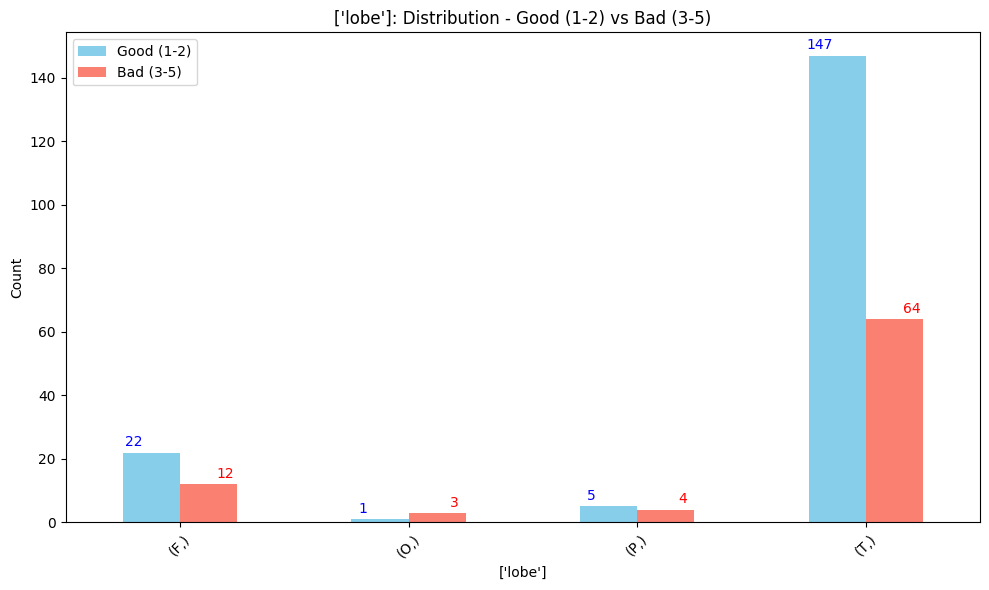




**Figure 2:**







**Figure 3:**

