LEDA Health – Sexual Assault Survey Analysis
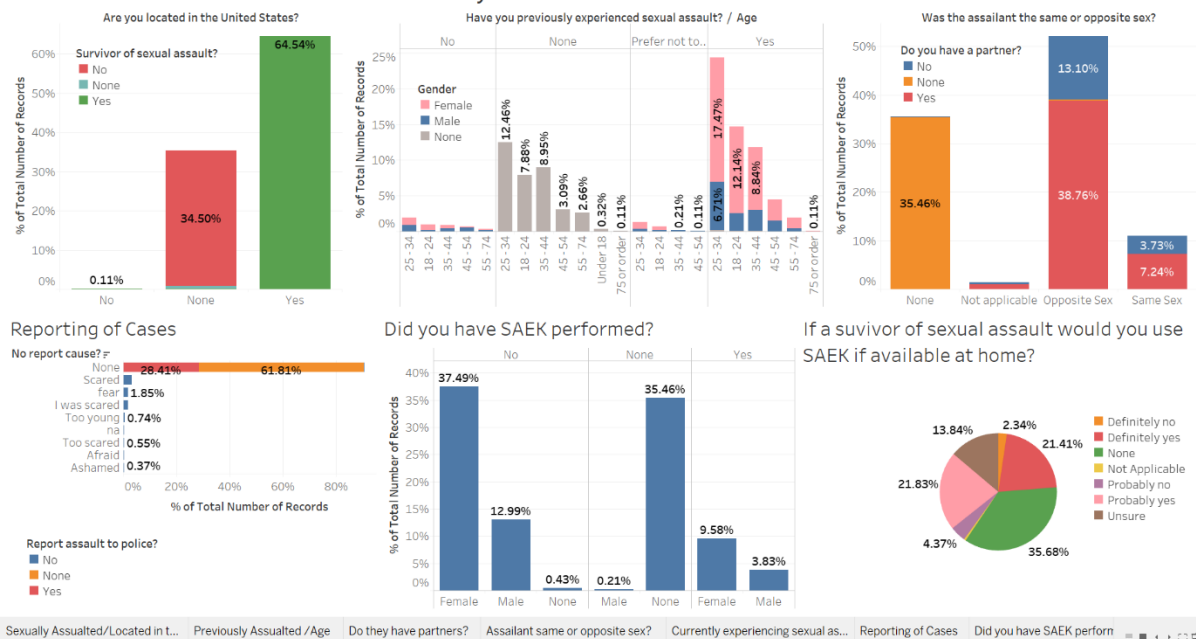
Analysis Done By: Lyzanne Erika Dsouza

Preprocessing and Plotting the Correlation:

1) The first step in my analysis was to load the dataset in excel and explore the data to get a better understanding of it.
2) I later used Python (Jupyter Notebook, Pandas and NumPy) and loaded the raw data. I performed preprocessing on the data by removing the unwanted rows and columns.
3) As the data was in string format, I first converted the data into categorical values and used the cat codes function to replace the string values with numeric values.
4) Once I got the numeric values for all the columns, I plotted the correlation matrix using Pearson's method. I further plotted a heatmap to understand the correlation between the columns.
5) I further used Tableau Desktop 10 for visualization of the data, where I plotted various univariate and bivariate charts based on the correlation matrix obtained in the earlier step.

Tableau Visualization and Analysis:

Using Tableau Desktop 10, I created a Dashboard for Leda Health – Sexual Assault Survivor Survey.
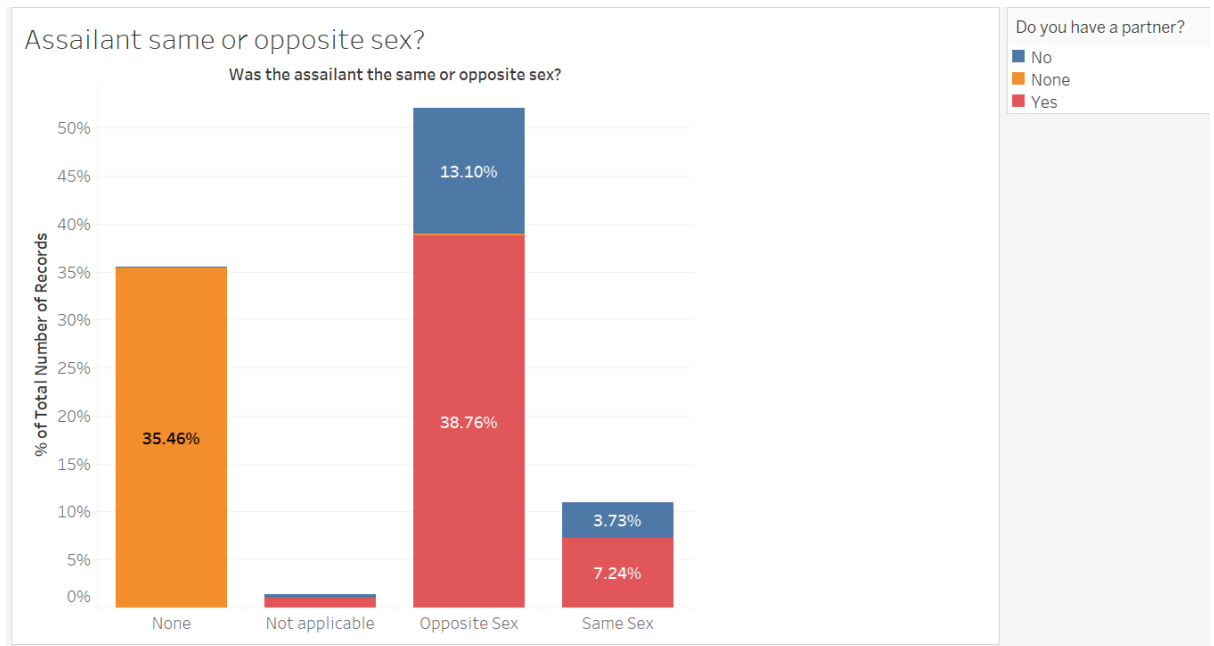
The dashboard mainly consists of 6 graphs:

1) The first graph displays if a person has been sexually assaulted or not and if the person is located in the United States. From the graph it can be seen that nearly 64.54% of the people that are located in the United states have been a survivor of sexual assault. The reason for plotting this graph was that these two columns had high correlation in the correlation matrix plotted earlier.
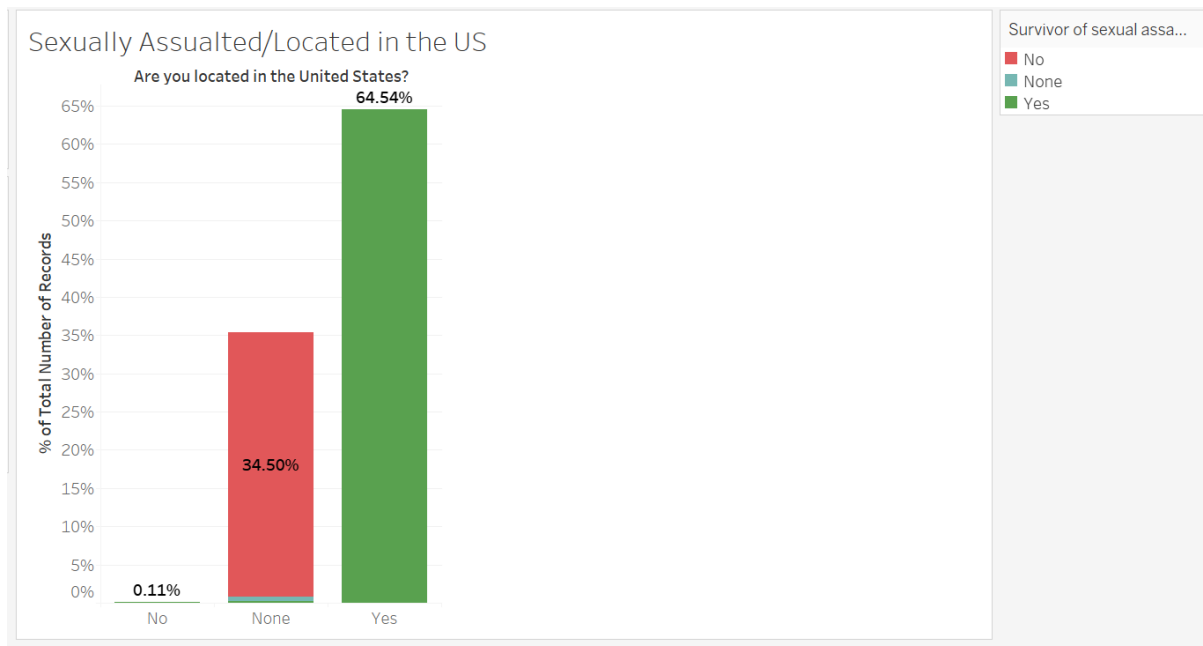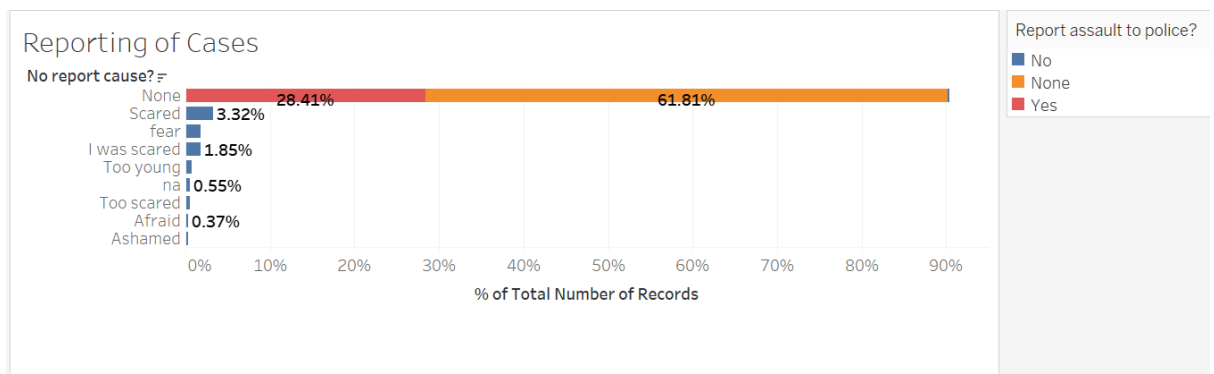
2) The second graph displays data if a person has been previously assaulted or not and has been classified by their age. To add more value to the graph, I have added gender as a color filter. From the below graph it can be seen that a greater number of Females have been assaulted previously and a majority of them are of the age group 25 to 34 years.
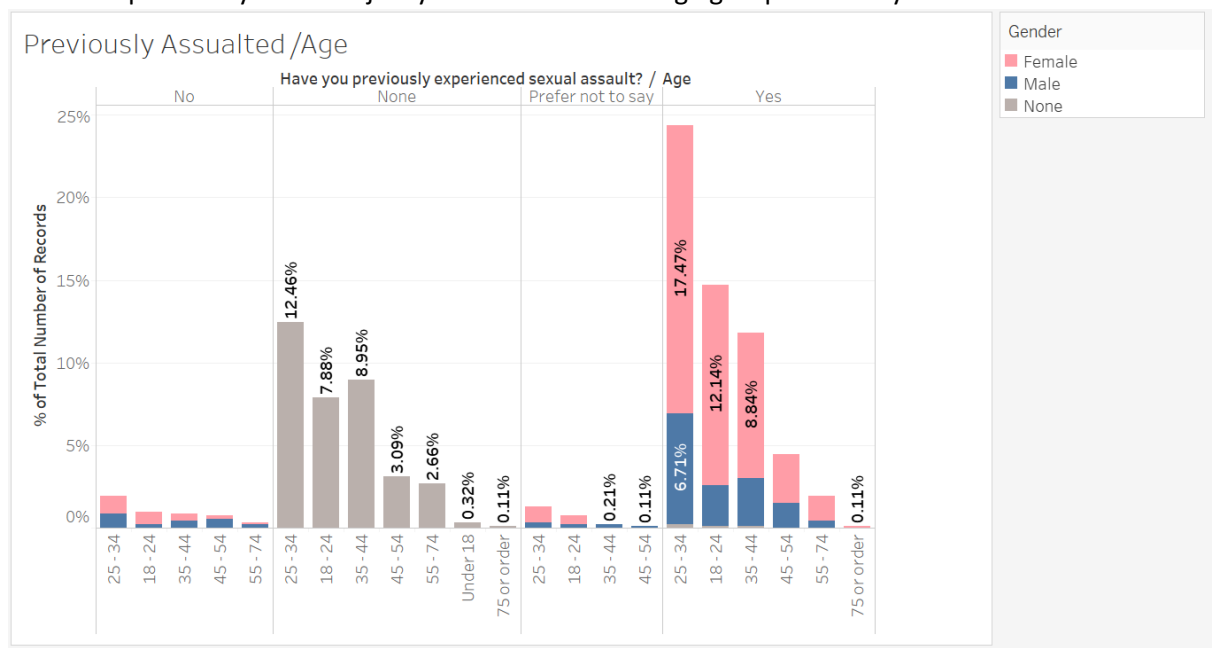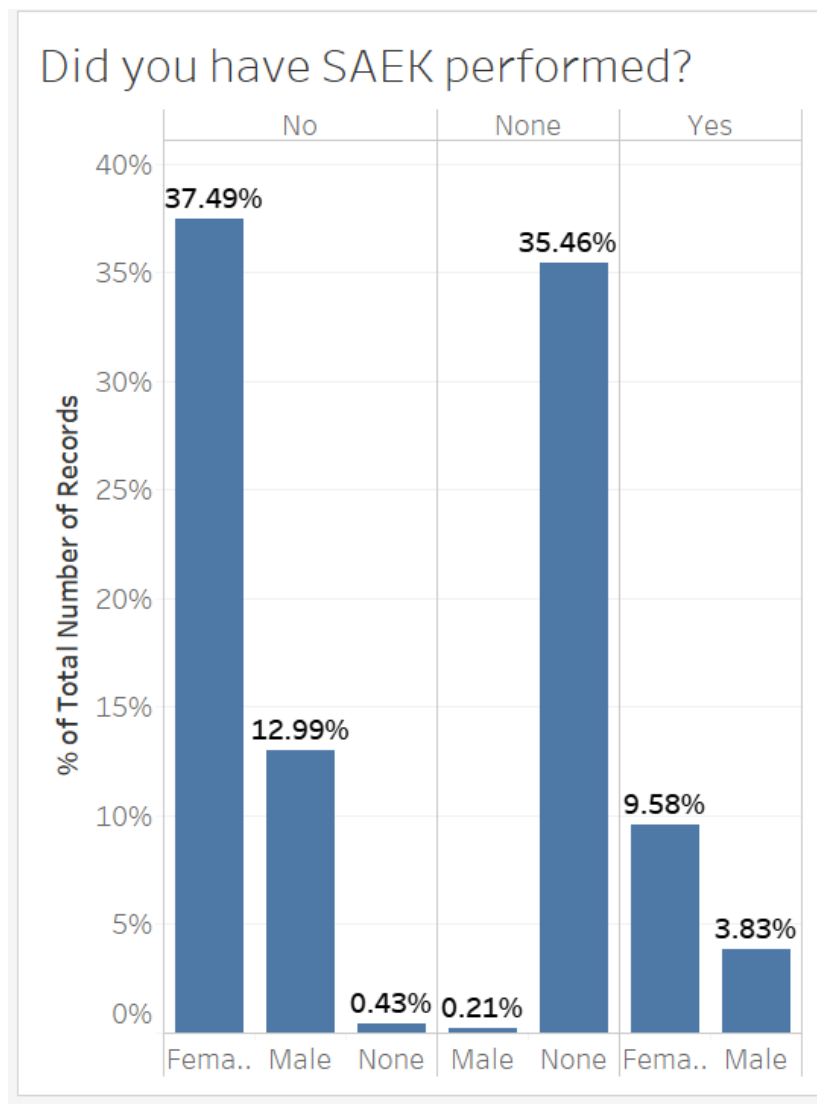

3) The third graph shows if the assailant is of the same sex or not. It also displays if the sexual assault survivor has a partner. As the correlation of these columns was high, I chose to plot this data to get useful information. It can be seen from the below graph that the highest percentage of assailants are of the opposite sex and most of the survivors did have partners.
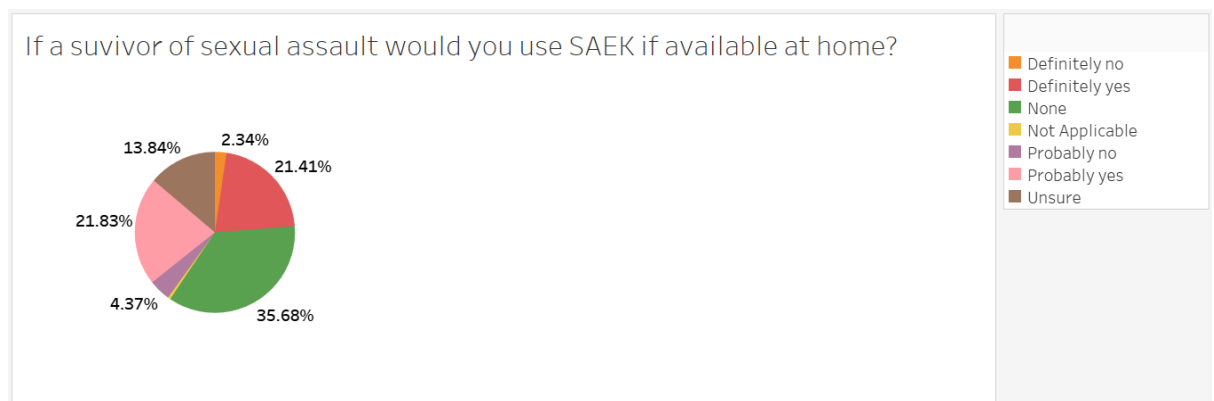
Assailant same or opposite sex?

Was the assailant the same or opposite sex?

Do you have a partner?
- No
- None
- Yes

13.10%
38.76%
35.46%
3.73%
7.24%

None | Not applicable | Opposite Sex | Same Sex

4) The fourth graph shows if the sexual assault survivor reported the case to the police. It also displays the top reasons for survivors not reporting the cases. It can be seen that most survivors didn't report the cases as they were scared.

Reporting of Cases

No report cause?

Report assault to police?
- No
- None
- Yes

None — 28.41% — 61.81%
Scared — 3.32%
fear
I was scared — 1.85%
Too young
na — 0.55%
Too scared
Afraid — 0.37%
Ashamed

% of Total Number of Records

5) The fifth graph shows if the survivors had SAEK performed. It is also classified by gender as they had a high correlation. The below graph shows that nearly 37.49% Females didn't have SAEK performed.

Did you have SAEK performed?

	No	None	Yes

% of Total Number of Records

- 37.49% — Fema.. (No)
- 12.99% — Male (No)
- 0.43% — None (No)
- 0.21% — Male (None)
- 35.46% — None (None)
- 9.58% — Fema.. (Yes)
- 3.83% — Male (Yes)

6) The sixth graph shows if a survivor of sexual assault would use SAEK if made available at home. It can be seen from the below graph that a relatively a greater number of survivors would use the SAEK if made available at home.

If a suvivor of sexual assault would you use SAEK if available at home?

- 13.84%
- 2.34%
- 21.41%
- 21.83%
- 4.37%
- 35.68%

Legend:
- Definitely no
- Definitely yes
- None
- Not Applicable
- Probably no
- Probably yes
- Unsure

Finally, to add more value and gain better insights on the data and their correlations, I added the above graphs onto a dashboard. I even added an interactive action filter which displays the association between all the graphs above and helps one understand the data more efficiently.

Here, is an example of the action filter applied on the dashboard:
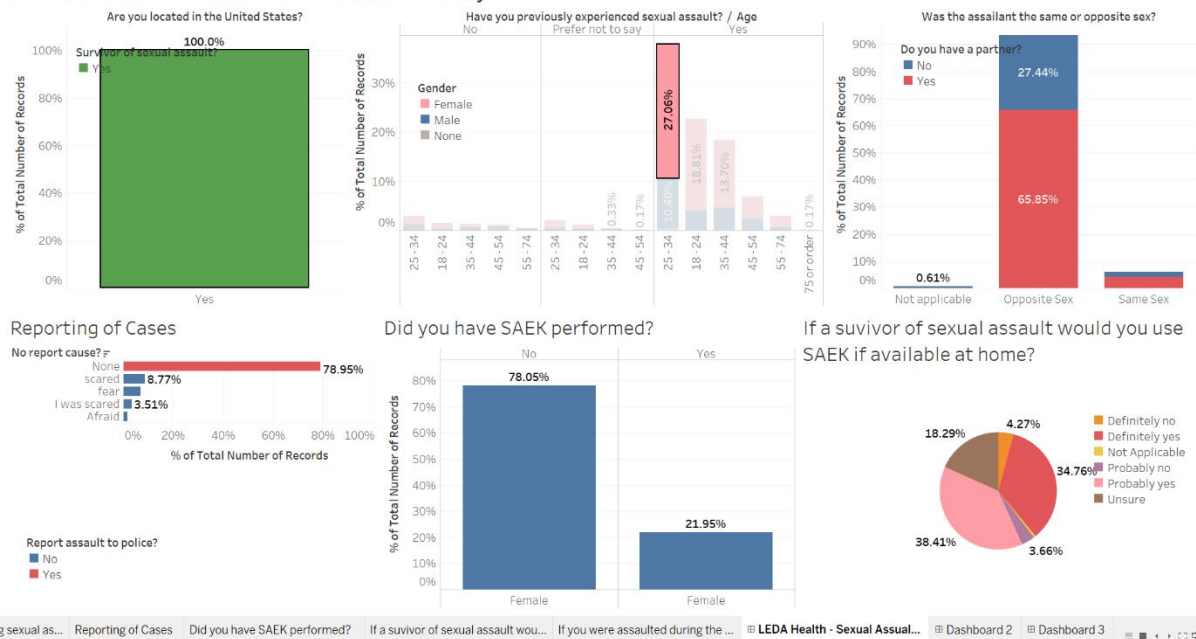
Here, I have selected all the sexually assaulted people located in the United states and belonging to the age group 25-34 and are Females. We can see that the charts have also been filtered accordingly.