For this project, we utilized data from the National Longitudinal Survey of Youth 1997, sponsored by the Bureau of Labor Statistics. This study was dedicated to tracking the labor market and other life experiences of American men and women.

The 1997 study was chosen over the 1979 study, as we felt it was more relevant to the majority of our classmates and included a number of variables we were specifically interested in.

In order to get a better understanding of the data and begin to decipher what we needed to focus our model on, we utilized pandas, matplotlib, and plotly. This allowed us to visualize the potential relationships between certain variables, as you will see below. These variables included age of first incarceration, marital status, personality, highest level of education, and even favorite ice cream flavor. Out of this exploration, we chose to engineer a feature termed the Longitudinal Chaos Index (LCI).

**LCI**

The LCI is a measure of how objectively chaotic someone's love life is over the course of the study. The more changes in your relationship/marital status, the higher your LCI. To create the LCI feature, we looked at the marriage variable over the span of the study. The specific question was listed as “Respondent's marital status in this month in [1994-2016]” was calculated for each month beginning with the month the respondent turned 14 years old. There were six possible responses: “Never Married, Not Cohabitating,” “Never Married, Cohabitating,” “Married,” “Legally Separated,” “Divorced,” and “Widowed.”

Depending on whether or not the respondent changed the answer to this question from the previous month, indicated whether or not the LCI would change. Except for two specific life events, all changes result in a positive 1 being added to the LCI. The first exception has to do with moving from “Never Married, Cohabitating” to “Married.” This does not cause the LCI to change. The second exception occurs when moving from “Married” to “Divorced,” which causes increases the LCI by 2 points.

After we had calculated the LCI of all the respondent's in the survey, we created scoring divisions. These scores detail a final, qualitative measure of a person's LCI. The categories are: Sad and Alone (SA), Happily Married (HM), It’s About the Journey (AJ), Train Wreck (TW).

**MODEL CREATION**

After the formation of the LCI, we moved on to creating a model that could predict a person's LCI, based on life events. To find the variables that would be the most predictive, we created a simple model that evaluated a single variable. After completing the model for the single variable, we added an iterator to loop through all of the variables included in the longitudinal study in order to find which ones were most influential to a person's LCI.

Finally, based on our findings, we were then able to create a predictive model to forecast LCI based on certain life events. With this, we were able to take user input and calculate an LCI based on the life events the current user had lived through at that time.