For this project, we utilized data from the National Longitudinal Survey of Youth 1997, sponsored by the U.S. 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 because it is closer (and therefore more relevant) to the ages of the majority of our classmates. The 1997 study also included variables that were of interest, like subjects’ cohabitation.

We utilized Pandas, MatPlotLib, and Plotly to explore the data and determine the needs and focus of our model. This allowed us to visualize the potential relationships between certain variables, as you will see <a href=’#exporation’>below</a>. 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 coined the term and engineered the Longitudinal Chaos Index (LCI).

**LCI**

The LCI is a measure of the objective chaos of a subject’s love life throughout the study. The more changes in your relationship/marital status, the higher your LCI. To create the LCI feature, we looked at reported marital status over the span of the study. The specific interview question used in the study was “Respondent's marital status in this month in [1994-2016]” (which will now be referred to as the “research question”) and was recorded every month after the subject’s fourteenth birthday. There were six possible responses: “Never Married, Not Cohabitating,” “Never Married, Cohabitating,” “Married,” “Legally Separated,” “Divorced,” and “Widowed.”

The LCI was designed to start at zero for each subject and then increase or decrease according to any change in response to the research question. All but two changes result in an increase by one point. The first exception is the change from “Never Married, Cohabitating” to “Married,” which we determined would not increase a subject’s chaos and therefore does not change the LCI. The second exception is the change from “Married” to “Divorced,” which increases the LCI by two points. After the survey subjects’ LCIs were calculated, 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.