Speaker Notes:

Hello, we are the Fly like a G6 team and our team members are Jose, Leslie, Renee, and Ben.

A little about ourselves

Jose: Currently works as a data analyst and during (INSERT PRONOUN) free time, he likes to (INSERT HERE)

Leslie (which is me): I am currently a Chemist and during my free time i enjoy crafting and being outdoors with my family.

Renee: Works at Target and she likes reading, painting, and sewing.

Last but not least
Ben: Is working in a power sports dealership and for fun enjoys fishing, researching blockchain technologies, and spending time with his fiance and dogs.

Our topic selection was the housing market

Our main purpose of choosing this topic was because we have team members that would like to buy a house in the very near future and would like to know if they should buy a house now or if they should wait.

Another reason why we choose this topic is because there was available data from trusted sources.

It is also a current trending topic.

Our main question for our housing market analysis is: Will the market crash? And if it does, will it be like the 2008 crash?

We also asked ourselves, what trends were seen in the years leading up to 2008 and how the trends are lining up to today

We wanted to know what trends are prominent, and if there was anything done to prevent another crash similar to 2008.

In order to answer our questions we needed to look at all the factors that occurred in leading up to 2008.
-These included banks being poorly regulated, high risky mortgage lending, and economic and political issues.
-Our data was sourced from housing and economic data, federal and state governments, banks and non-profit organizations. We also read a few articles such as The Crash of 2008.

Overview: During Phase 2, we ended up reject our data that was broken up into individual years. Since we felt that it wasn't a good representation of the trends for every single year.

Data Processing: We ended up finding the average of each single year and creating a dataframe with the averages. We then joining hpi and foreclosure as well as hpi and household

debt.

Feature Selection: So instead of running only years 1991, 2000, 2008, 2015, and 2022, we did all of the 30 years. Two of the datasets had data from 1991 to 2022 and our household debt data only had data from 1999 to 2021

Splitting the Data: We trained the model for every year on household debt data, foreclosure, and home price index.

The team used Visual Studio Code and Jupyter Notebook to clean the data sources, with the Pandas Library to clean datasets, remove missing rows, and extract the files. The data was then imported into PgAdmin using SQL, and joined using SQL and pandas. The data was split into years to view the Housing Price Index (HPI) and household debt. Multiple libraries were used to conduct the Machine Learning portion of the analysis.

Here we have our link to the dashboard that was created and we included graphs of our datasets. With HPI, the cost of houses has gone up from 1991-now. Household debt has seen an increase since 2000.Foreclosure rates had a big jump after 2005, right before the crash. We used k-means to see how the data would be classified. In HPI, the bumps over 2005-2008 ad 2015 are red, we see the trend… Foreclosure rates have the red on both sides of the peak, and the tails are the same color. Household debt has a small downward slope to it as it plateaus. With these three graphs we see that despite the increase in housing prices, household debt is steady and foreclosure rates are low.

On this slide we have data that was combined and used PCA to do our analysis. Using PCA Analysis, we combined HPI with the two other

In the first graph it took into consideration our Housing Price Index and household debt:
We ran PCA in 4 PCA classes

In the second graph we took the Housing Price Index and foreclosure

When looking at the machine learning module for HPI and foreclosure, we can see that the classes are similar in results as HPI and household income.

-Are their similar trends? Looking at the data, we can see that we are heading to similar trends in hpi, foreclosure, and household debt compared to 2008. Although we are not completely there yet. We are in different classes.

-Will there be a crash? We predict that there will be a crash in the near future (potentially in 2023) due to the trends being just one class behind 2008 trends. But there are too many factors and unpredicted events that needs to be considered to truly know if it will crash.

-Will the crash be as bad as 2008? Due to laws and regulations we don’t think if there is a crash, that it will be anything like 2008.

-World Events: It would have been a good idea to look at current world events and how similar past events affected the housing market. This would help us consider other factors that impact the market

-Datasets: Due to time and focus on individual years instead of years as a whole, we had limited time to work on more datasets to analyze.

-Laws: Looking into laws and regulations that were added after the 2008 crash could have also given us more insight on what was done to avoid the crash, and if there was a crash, how it would affect the housing market.

-Stock Market: Looking into all of the stock market datasets could have given us more insight and have link in current world events. For example how pharma stocks tend to go up or down depending on what health crisis is going around the world that could cause a ripple effect to the housing market.

-Data: Setting a hard deadline for ourselves on data collection and cleaning, so we could have spent more time on the analysis. With the fraud data, we added it later in the process

-Machine Learning: We had trouble being able to get results. First we broke down all the data into years and did machine learning for individual years and then we realized towards the end that we should use all of the years together to have a better analysis of the data.

-Results: At first we didn't fully understand what our results were trying to tell us when we used individual years so we spend a lot of time on individual years instead of a year's worth of data and trends.

-Topic: our first topic was spent on when the housing market would crash and we spent time trying to see how to predict when instead of answering our new question of IF it will crash.