|  |  |
| --- | --- |
| **Slide** | **Script** |
| **Slide 6: Project Flow overview** | If you look at the slide on the screen you can see the steps we took to perform our analysis on the screen. We first loaded the dataset from Kaggle into Postrgres and connected the database to jupyter notebook where we cleaned the data. We then applied machine learning algorithms to analyze the relationships between house features and price. Finally, we created visualizations in tableau. |
| **Slide 7: Questions for us to answer** | We would like the machine learning model to answer the following questions:   * Which cities/suburbs in and around Austin are the fastest growing in regards to real estate? * What factors influence the increase in housing prices (such as schools, population, square footage of the house, built year, etc.)? * How has the Austin housing market changed over the past three years? |
| **Slide 8: Exploring the Data** | After connecting our database to postgres and loading it into jupyter notebook we began to explore the data. There were 2 major components: cleaning the data and creating a correlation matrix to narrow down the features we were observing.  1st we cleaned the data set. We pre-processed the data by dropping new values and columns that didn’t fit. Wethen created a correlation matrix to illustrate how features contribute to the house price.  There are a lot of factors and considerations that go into the price of an individual house; we wanted to focus on the factors that generally had the most impact on all house prices in the Austin area. We ultimately dropped the columns for Number of Price Changes, Street Address, Latitude, Longitude, Latest Price Source, and Home Image |