Group 3 Final Project

      For our final project we are going to focus on wine ratings and pricing. We have found a data set that contains four different csv files. These include red wine, white wines, sparkling wines and rose wines. The important columns in our data sets include regions, countries, ratings, price, year, and number of ratings. For the number of ratings we are going to focus on wines that received above 100 ratings. For the price column we are going to convert the original price from the data set (in EUR) to American dollars. That way it is easier to understand what the price really would be. We are also going to combine all the tables into one so that all our data is in one place. The best wine years were 1975, 1976, 1983, 1995, and 2019. These were the best wine years for vintage wines. The newer years for good wines are 2007, 2010, 2012, 2014, and 2017. Red wines are the only wines that get better with age. This means that red wine is better to be left alone for years and then drunk. Meaning that the production year for these wines should be long before the selling year. For our models we are going to focus on using supervised machine learning. We are going to incorporate pandas into our supervised machine learning as well. We will create a tableau story to show our different visualizations we create and tell the story of wine. For our supervised machine learning we are planning on creating a linear regression model to show the price of each wine versus the year that it was produced. We are also considering looking at the region versus the price and the country versus the price. For our tableau visualizations we will create a map to show the different regions in each country and see how many bottles of wine they have produced. This will give us an idea of which regions produce the most wines and which countries produce the most wine as well. We will also analyze which countries produce the best wine by looking at the rating given. This is the starting point for our final project.