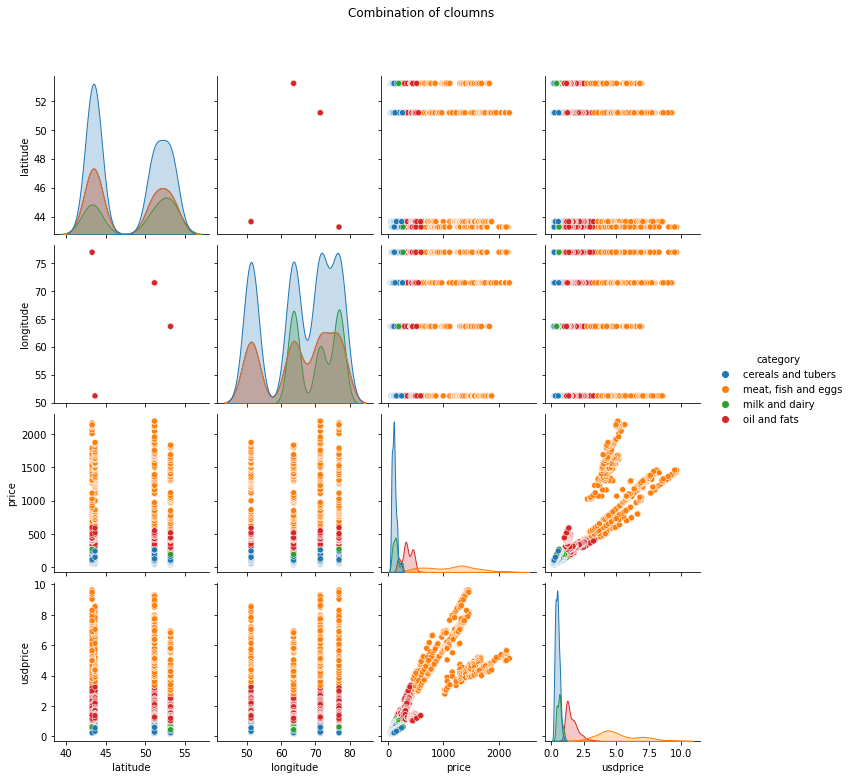
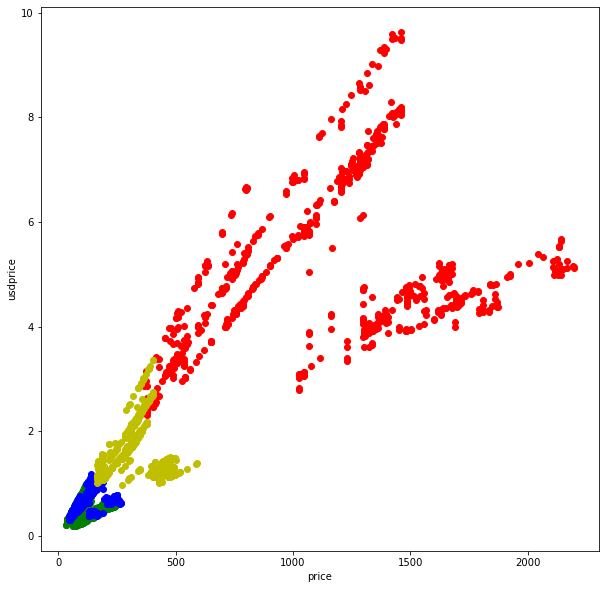
In this stage lets classify products. **.** For classification we need target column it would be “category”. However for solving this problem we need illustrate data. To better show how data is distributed between each other in python, seaborn library has **pairplot.** It shown below.



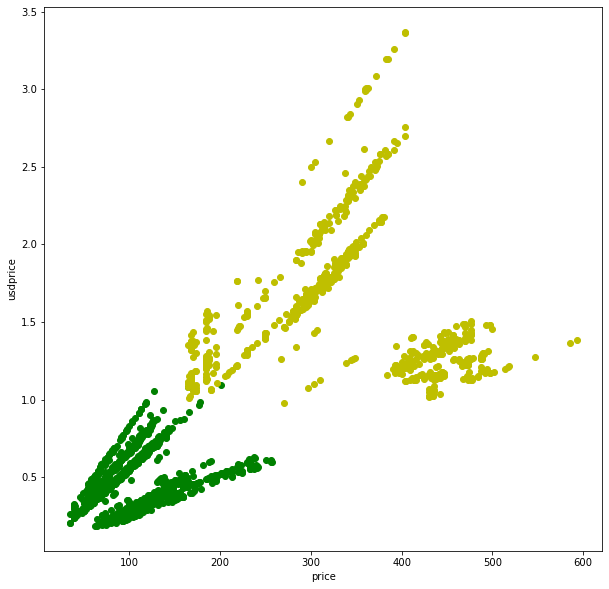
If we saw these graphs we can identify that columns “price” and “usdprice” are somehow distributed. In addition to that much closer for classification.

Let’s show the graph in a bigger form. Here each color represents categories.

Red: meat, fish and egg ; Green: cereals and tubers ; Blue : milk and dairy ; Yellow : oil and fats



Meat and oil intersects on the most of data. When it comes milk and cereals it is impossible to take them at least in this situation. For classification, it would be better if we take cerals and oil.



So we prepare proper data. To classify them we can you two algorithms KNN and SVM. After that compare which of them fit better for this data.

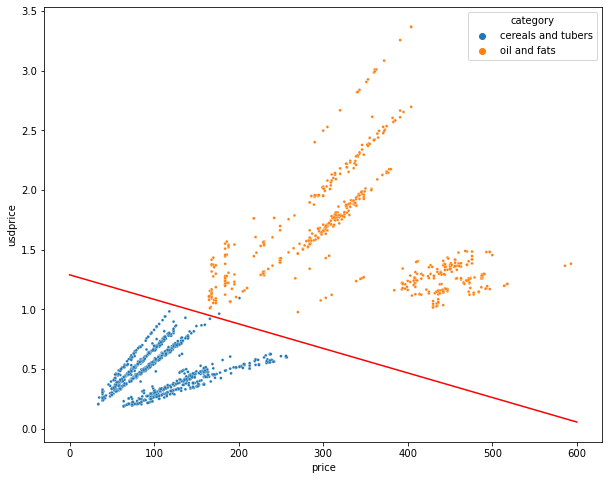


**KNN - K-Nearest Neighbours**

The first step starts with dividing data into train and test data with percentages of 80 and 20. Further, we need to normalize data to make the graph more suitable. After implementing, all required functions we compare it with inbuilt functions. It shows 100 percent of similarity. In addition to that with an accuracy of 97 percent.

**SVM - Support vector machine**

For here we need to take already separated data in the KNN part. When we find accuracy inbuilt python libraries. It gave a 67 percent of accuracy.



For this sort of data we can conclude that KNN better suit.