Data Preprocessing:

Dependencies used:

- OpenCV cv2
- Matplotloib
- Os
- Numpy
- Pandas
- Sklearn - test_train_split

Data: I have used data given in the GitHub page in this link from where i have downloaded part 2 of the dataset and worked on it.

Dataset has following data in it.

- BrandImagesFromShelves
- ProductImagesFromShelves
- ShelfImages

I have used this data for both training and testin of my model, because the folder ProductImagesFromShelves has images where the name of the image containing the information about the particular image dimensions in a string datatype.

Creating Data Frames and importing the image files

As i have mentioned i used the part 2 data file of the dataset.. Where i have information about image dimensions in the image name itself.

First Data Frame:

Creating product data frame that contains all the information about the products inside the shelves.
I got the products data from **ProductImagesFromShelves** which has all product image dimensions in their names
As
C1_P01_N1_S2_1.JPG_1008_1552_252_376.png

Here in the above sample image name the starting point on x and y mentioned respectively. Then there is information of W(width) and H(height). The entire name is in the string form so i used string manipulation to extract the information.

Xmin = defined in the name of the image file.
Xmax = X min + width(w: there in the name)
Ymin = Defined in the name of the image file
Ymax = y min + height(h: there in the name)

With the information i got extracted i have created a Pandas Data Frame. Named as df2_products.

Second Data Frame:
My second Data frame is simply shelf images. This data frame contains both shelf id and the planogram id in its name, this was mentioned in the Readme of the problem statement.

Checking for null values or missing values:
I have checked the two dataframes for the possible occurrence of null values or missing values and datatypes of the data.

Test_train_split:
Splitting the data between training and validation in 0.7 : 0.3 splits…

Creating pickle files:
Created pkl files for saving the data of both products and shelf photos. names : photos.pkl, products.pkl.

Checking the work:

Creating rectangle boxes around the products in a shelf image using OpenCV with rectangles and dimensions of the products from the df2_products data

Xmin, Xmax, ymin, ymax..