WEEKLY EXTRA ACTIVITY – III

01) The data is collected about anemia disease prediction dataset from Kaggle website. Data present in the URL below is being generated and the data in it is not Genuine and copyrights and trademarks are infringed in between. There are 1000 Samples are in the data. The components included in the data are Sex of the individual, Red color pixel intensity, Blue color pixel intensity, Green color pixel intensity and Hemoglobin. All the variables except sex column are numeric.

	Number	Sex	%Red Pixel	%Green pixel	%Blue pixel	Hb	Anaemic
0	1	M	43.26418	30.83892	25.89959	6.297293	Yes
1	2	F	43.14483	30.17140	26.69300	8.608315	Yes
2	3	F	46.50649	27.43091	26.05113	9.713010	Yes
3	4	F	44.96398	30.51920	24.49916	4.809385	Yes
4	5	M	45.06947	31.08938	23.85352	8.995228	Yes
...	...	...	...	...	...	...	...
995	996	F	45.70484	27.55714	26.71248	13.615370	No
996	997	F	44.90627	30.37529	24.71365	11.594260	No
997	998	F	43.43062	29.77783	26.78202	12.407690	No
998	999	F	43.30468	29.80503	26.87594	12.102480	No
999	1000	F	47.64893	26.80250	25.55574	15.083690	No

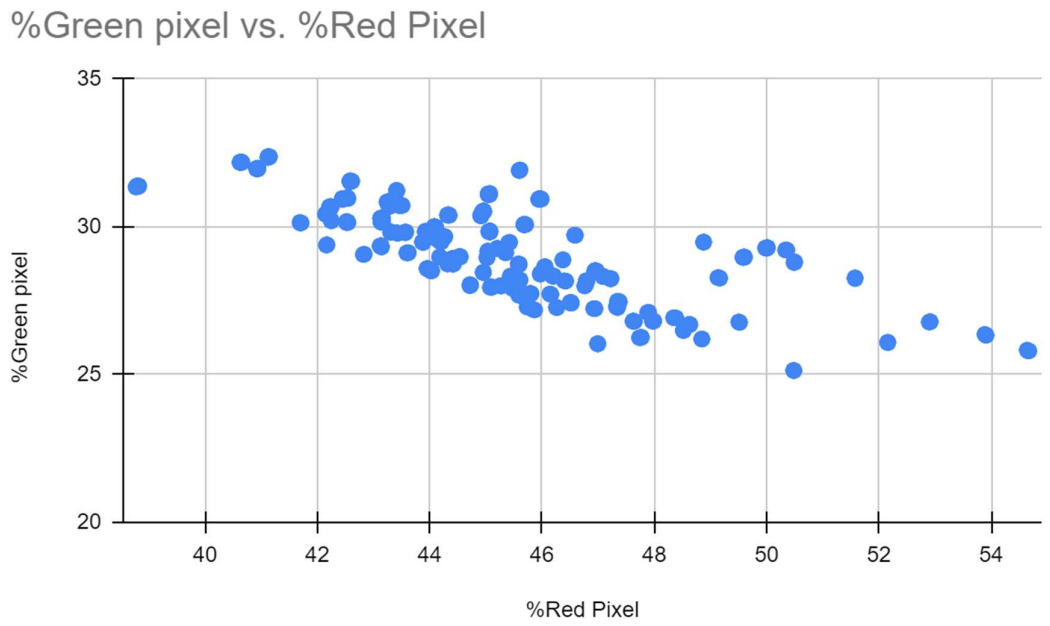
1000 rows × 7 columns

02) The data is transferred to Google Sheet and can be accessed from the following link

URL:
https://docs.google.com/spreadsheets/d/1fq9t84NTsvmLAi_uv4B7WXrkWGWRWc83de6ghx41cP8/edit?usp=sharing

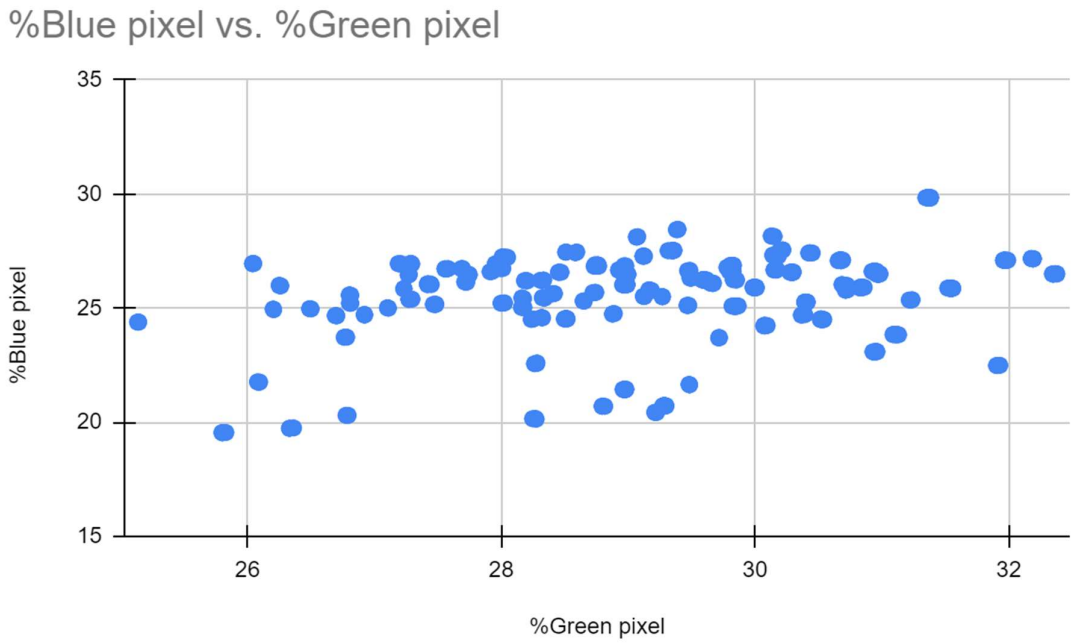
03) (i). To perform the activity, %Red Pixel, %Green pixel, and %Blue pixel are chosen as the numerical variables X1, X2 and X3, respectively from the dataset.

WEEKLY EXTRA ACTIVITY – III



Interpretation :

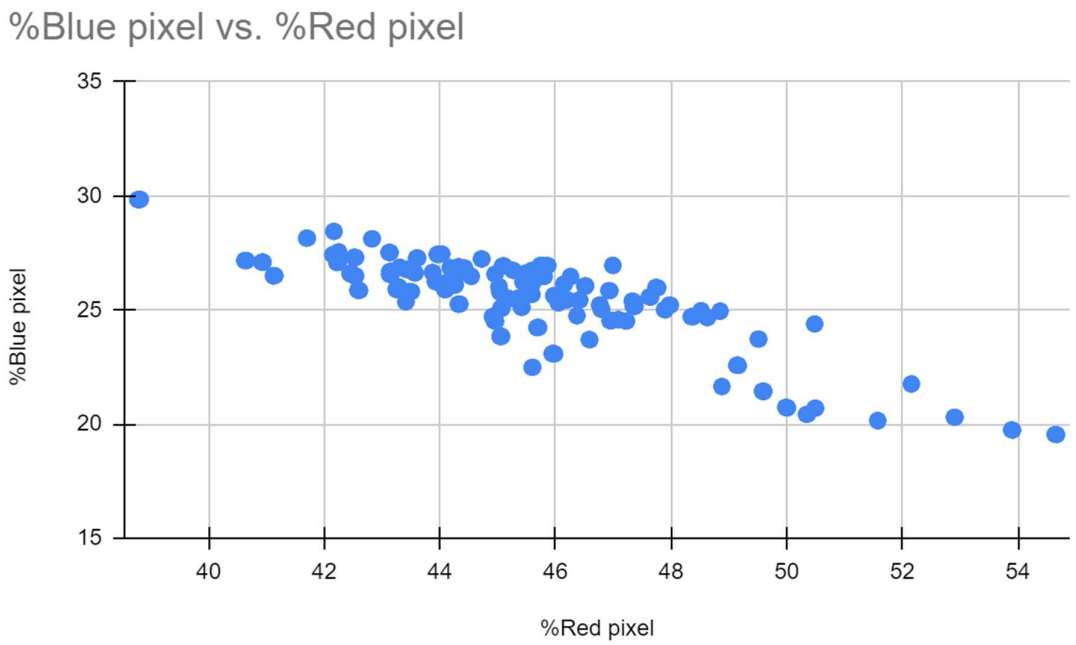
From the above scatter plot, we can conclude that there is a weak relation between Green Pixel percentage and Red Pixel Percentage. This indicates a close relationship between those to numerical variables. Thus if the value of one variable increases then another decreases…



Interpretation :

WEEKLY EXTRA ACTIVITY – III

From the above scatter plot, we can conclude that there is a strong relation between Blue Pixel percentage and Green Pixel percentage. This indicates a close relationship between those to numerical variables.



Interpretation :

From the above scatter plot, we can conclude that there is a weak relation between Green Pixel percentage and Red Pixel Percentage. This indicates a close relationship between those to numerical variables. Thus if the value of one variable increases then another decreases…

03) (ii) Calculation of Covariance between pair of variables:

We can determine the covariance between the pair of variables using the covariance matrix, as illustrated below:

	%Red Pixel	%Green pixel	%Blue pixel
%Red Pixel	7.727981301	-3.281897191	-4.44625908
%Green pixel	-3.281897191	2.485432592	0.7959396341
%Blue pixel	-4.44625908	0.7959396341	3.651324115

Interpretation:

WEEKLY EXTRA ACTIVITY – III

The covariance of Red pixel percentage vs Green pixel percentage is Negative. The covariance of Red pixel percentage vs Blue pixel percentage is Negative. The covariance of Blue pixel percentage vs Green pixel percentage is Positive. By observing the values we can say that,

- If the Red pixel percentage increases Green pixel percentage decreases,
- If the Red pixel percentage increases Blue pixel percentage decreases,
- If the Blue pixel percentage increases Green pixel percentage also increases.

04) (I) Calculation of mean and standard deviations:

Mean and standard deviations for the variables X1, X2, and X3 have been calculated as below :

	Mean	Std Deviation
%Red Pixel	45.03124758	2.781315699
%Green pixel	29.39409918	1.577314335
%Blue pixel	25.5743227	2.824116829

Interpretation :

All the three variables have different Mean values and different Standard deviation with Blue pixel percentage having the least pixel percentage and Red pixel percentage having the highest pixel percentage. This may be because of different individual samples may contain different values according to their lifestyle and diet.

04) (ii). Calculation of bounds by using Chebyshev's inequality:

One of the practical questions that comes to the investors mind is what is the upper and lower bound in the probability that these stocks will make a move of certain standard deviation from the mean.

1). Calculation of bounds for %Red Pixel

a). What is the Upper bound on the probability that Red pixel Percentage will make a move of more than 40 from its mean in the coming time given the data?

$P(|X - \mu| \geq k\sigma) \leq 1/k^2$

For Blue pixel percentage,

$\mu = 45.03$, $\sigma = 2.78$,

$P(|X - 45.03| \geq 40) \leq 1/k^2$

$P(|X - 45.03| \geq 2.78*14.39) \leq 1/(14.39)^2$

$P(|X - 45.03| \geq 2.78*14.39) \leq$ **0.005**

Interpretation:

The chance that Red Pixel percentage will make a move above 40 from mean is at most 0.005. Hence, there is a very low probability that there will be a move of that magnitude in Red pixel percentage.

b). What is the Lower bound on the probability that Red pixel Percentage will make a move of less than 40 from its mean in the coming time given the data?

WEEKLY EXTRA ACTIVITY – III

$P(|X - \mu| \geq k\sigma) \leq 1/k^2$
For Blue pixel percentage,
$\mu = 45.03$, $\sigma = 2.78$,
$P(|X – 45.03| < 40) > 1 - 1/k^2$
$P(|X – 45.03| < 2.78*14.39) > 1 - 1/(14.39)^2$
$P(|X – 45.03| < 2.78*14.39) >$ **0.995**

Interpretation:

The chance that Red Pixel percentage will make a move above 40 from mean is at least 0.995. Hence, there is a somewhat high probability that there will be a move of that magnitude in Red pixel percentage.

2). Calculation of bounds for %Blue Pixel

a). What is the upper bound on the probability that Blue pixel Percentage will make a move of more than 40 from its mean in the coming time given the data?

Interpretation :

$P(|X - \mu| \geq k\sigma) \leq 1/k^2$

For Blue pixel percentage,

$\mu = 25.57$, $\sigma = 2.82$,

$P(|X – 25.57| \geq 40) \leq 1/k^2$

$P(|X – 45.03| \geq 14.18*2.82) \leq 1/(14.18)^2$

$P(|X – 45.03| \geq 14.18*2.82) \leq 0.005$

Interpretation:

The chance that Blue Pixel percentage will make a move above 40 from mean is at most 0.26. Hence, there is a somewhat low probability that there will be a move of that magnitude in Blue pixel percentage.

b). What is the lower bound on the probability that Blue pixel Percentage will make a move of less than 40 from its mean in the coming time given the data?

Interpretation :

$P(|X - \mu| \geq k\sigma) \leq 1/k^2$

For Blue pixel percentage,

$\mu = 25.57$, $\sigma = 2.82$,

$P(|X – 25.57| < 40) > 1 - 1/k^2$
$P(|X – 25.57| < 14.18*2.82) > 1 - 1/(14.39)^2$
$P(|X – 25.57| < 14.18*2.82) >$ **0.995**

Interpretation:

The chance that Blue Pixel percentage will make a move above 40 from mean is at least 0.99. Hence, there is a very high probability that there will be a move of that magnitude in Blue pixel percentage.

WEEKLY EXTRA ACTIVITY – III

3). Calculation of bounds for %Green Pixel

a). What is the Lower bound on the probability that Green pixel Percentage will make a move of more than 40 from its mean in the coming time given the data?

Interpretation :

$P(|X - \mu| \geq k\sigma) \leq 1/k^2$

For Green pixel percentage,

$\mu = 29.39$, $\sigma = 1.58$,

$P(|X - 29.39| \geq 40) \leq 1/k^2$

$P(|X - 45.03| \geq 1.58*25.32) \leq 1/(25.32)^2$

$P(|X - 45.03| \geq 1.58*25.32) \leq$ **0.002**

Interpretation:

The chance that Green Pixel percentage will make a move above 40 from mean is at most 0.005. Hence, there is a somewhat low probability that there will be a move of that magnitude in Green pixel percentage.

b). What is the Lower bound on the probability that Green pixel Percentage will make a move of less than 40 from its mean in the coming time given the data?

Interpretation :

$P(|X - \mu| \geq k\sigma) \leq 1/k^2$

For Green pixel percentage,

$\mu = 29.39$, $\sigma = 1.58$,

$P(|X - 29.39| < 40) > 1 - 1/k^2$

$P(|X - 29.39| < 1.58*25.32) > 1 - 1/(25.32)^2$

$P(|X - 29.39| < 1.58*25.32) >$ **0.998**

Interpretation:

The chance that Green Pixel percentage will make a move above 40 from mean is at least 0.005. Hence, there is a high probability that there will be a move of that magnitude in Green pixel percentage.