

### Group Assignment:

Muhammad Anser Sohaib (367628)

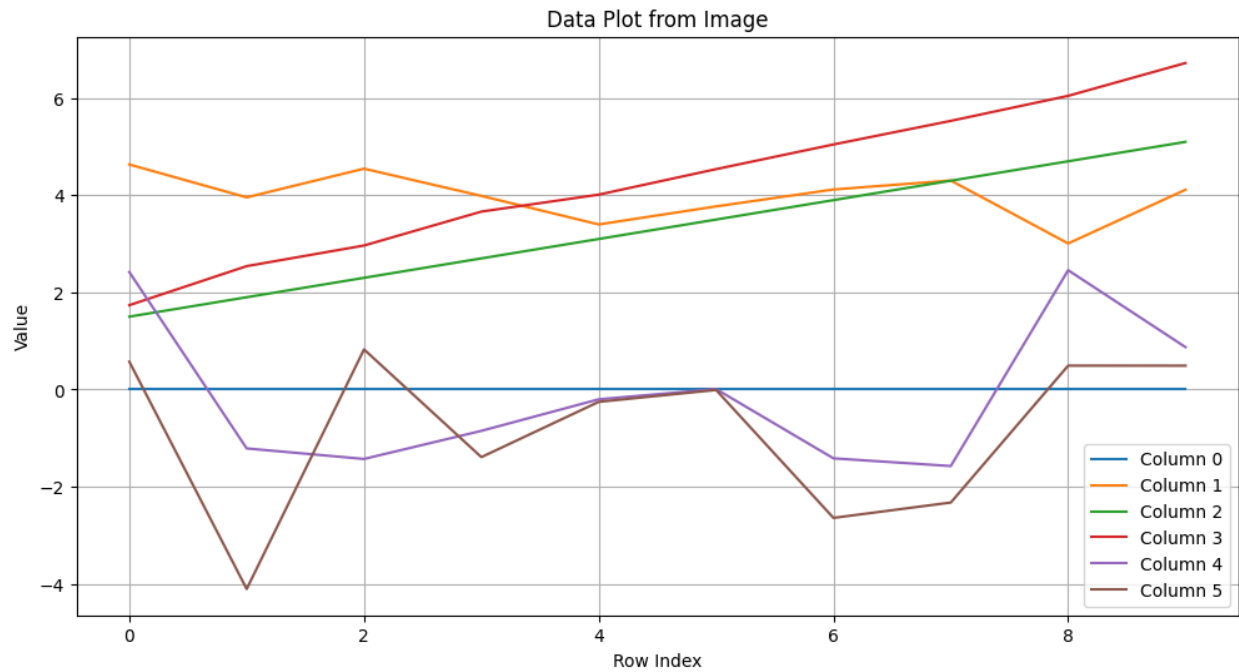
Zuha Fatima (385647)

Using the given data, we calculated the following statistical values

Which columns are useless and which are useless? why?

0	4.6345	1.5000	1.7364	2.4188	0.5726
0	3.9570	1.9000	2.5405	-1.2137	-4.1087
0	4.5477	2.3000	2.9652	-1.4305	0.8252
0	3.9868	2.7000	3.6637	-0.8517	-1.3928
0	3.3993	3.1000	4.0134	-0.2021	-0.2547
0	3.7709	3.5000	4.5384	0.0084	-0.0098
0	4.1190	3.9000	5.0463	-1.4180	-2.6444
0	4.3046	4.3000	5.5324	1.5765	-2.3284
0	3.0078	4.7000	6.0477	2.3509	0.1352
0	4.1139	5.1000	6.7239	0.8723	0.4903

Column	1	2	3	4	5	6	
Mean	0	3.98415	3.3	4.38079	0.21109	-0.87155	
Variance	0	0.2219780065	1.32	2.4410477769	2.0474423709	2.4972078485	
Standard Deviation	0	0.4711454197	0.1489125293	1.5623852844	1.4308886647	1.5802556276	



### Our Reasoning:

- Column 1 has a 0 variance, so it's not useful at all.
- Column 2 has a little variance as compared to other columns; it is not that useful to keep it. Removing it will be better.
- Columns 3 is the scaled version of column 4, we'll keep column 4 since it has more variance than 3.
- Columns 4,5,6 having the highest variance should be kept and they are useful.

### Conclusion:

- Column 1,2, 3 are not useful and columns 4, 5, 6 are useful.