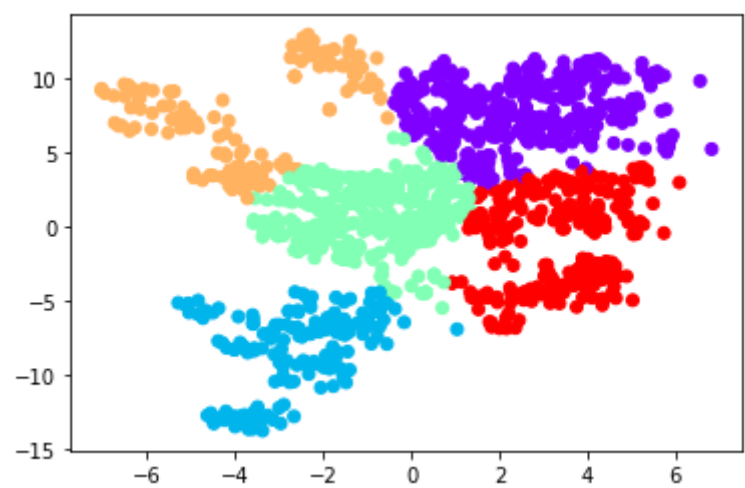


Final report

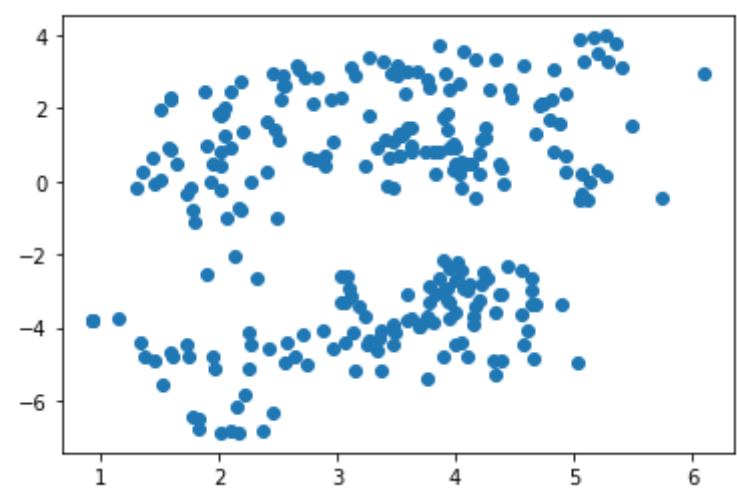
The objective of this data science project is to segment the data so that the customer can take the appropriate measures to optimize their company.

After the exploratory analysis steps and the assembly of the machine learning model, we arrived at an optimal number of 5 well-defined clusters, whose characteristics will be detailed below.



Cluster 0

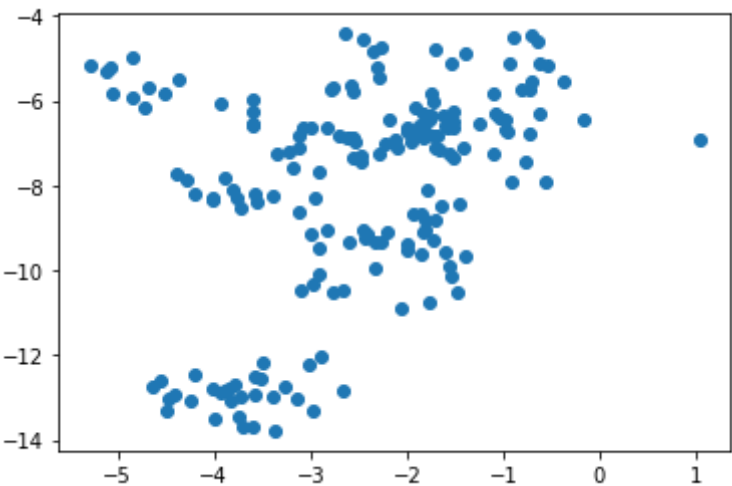
	Cluster_Num	V1	V2
count	272.0	272.000000	272.000000
mean	0.0	3.372109	-0.894426
std	0.0	1.114514	2.969554
min	0.0	0.929700	-6.880400
25%	0.0	2.442100	-3.716475
50%	0.0	3.536300	-0.051741
75%	0.0	4.156025	1.560725
max	0.0	6.091900	3.983400



Summary of cluster distribution and statistics, represented by the red color in the graph.

Cluster 1

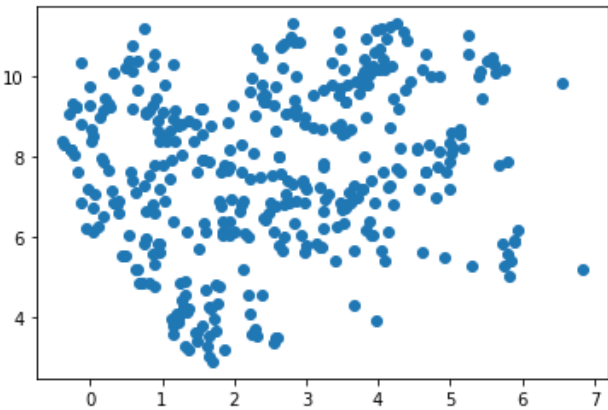
	Cluster_Num	V1	V2
count	185.0	185.000000	185.000000
mean	1.0	-2.527072	-8.134872
std	0.0	1.196272	2.556141
min	1.0	-5.294300	-13.773100
25%	1.0	-3.506000	-9.367600
50%	1.0	-2.434900	-7.240400
75%	1.0	-1.710100	-6.348600
max	1.0	1.040000	-4.415900



Summary of cluster distribution and statistics, represented by the blue color in the graph.

Cluster 2

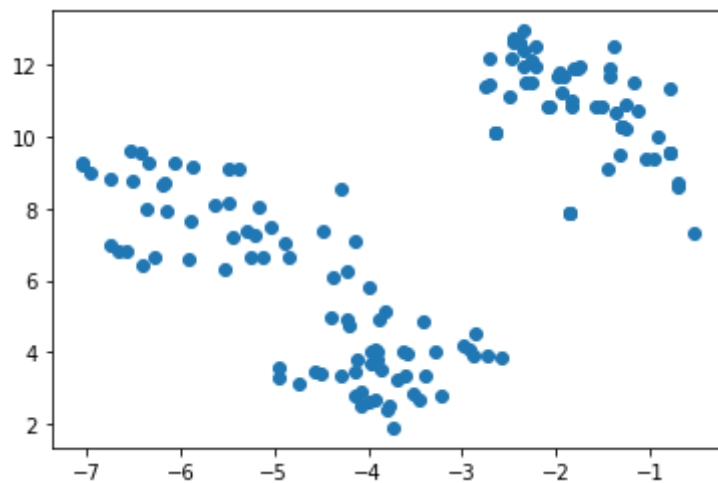
	Cluster_Num	V1	V2
count	388.0	388.000000	388.000000
mean	2.0	2.482843	7.633778
std	0.0	1.628885	2.062630
min	2.0	-0.382140	2.909400
25%	2.0	1.158050	6.138500
50%	2.0	2.410200	7.717350
75%	2.0	3.776025	9.274500
max	2.0	6.824800	11.304500



Summary of cluster distribution and statistics, represented by the purple color in the graph.

Cluster 3

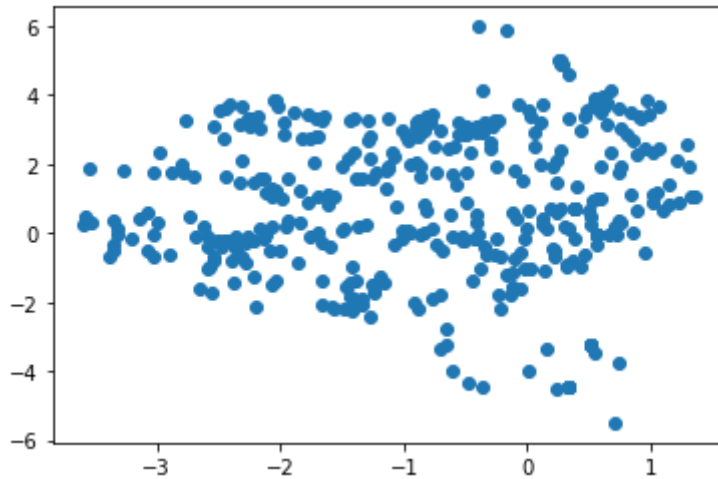
	Cluster_Num	V1	V2
count	142.0	142.000000	142.000000
mean	3.0	-3.514487	7.764252
std	0.0	1.780747	3.225657
min	3.0	-7.042100	1.903700
25%	3.0	-4.874375	4.278775
50%	3.0	-3.591300	8.084700
75%	3.0	-1.973075	10.718825
max	3.0	-0.539660	12.951600



Summary of cluster distribution and statistics, represented by the yellow color in the graph.

Cluster 4

	Cluster_Num	V1	V2
count	385.0	385.000000	385.000000
mean	4.0	-0.828329	0.834479
std	0.0	1.228363	2.098236
min	4.0	-3.593300	-5.498100
25%	4.0	-1.821500	-0.436630
50%	4.0	-0.735100	0.709800
75%	4.0	0.265170	2.696300
max	4.0	1.351800	5.978100



Summary of cluster distribution and statistics, represented by the green color in the graph.

With all clusters well defined and detailed graphically and statistically, we are able to identify the pattern that the data is showing us and with the knowledge of the customer's business, he will be able to take the necessary measures to improve his company.

With this segmentation, the customer will be able to better identify the counterfeit bills and thus improve his company.