



Hala Albahloul

Clustering

E-Commerce Data



Data Analysis & Preprocessing

**Elbow Method &
K-means Clustering**

**Silhouette Method &
K-means Clustering**

**Hierarchical Agglomerative
Clustering**

Agenda

Dataset

[Back to Agenda Page](#)

All the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/9/2011 12:50	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/9/2011 12:50	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/9/2011 12:50	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/9/2011 12:50	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/9/2011 12:50	4.95	12680.0	France

541909 rows × 8 columns

Dataset

[Back to Agenda Page](#)

InvoiceNo	0
StockCode	0
Description	1454
Quantity	0
InvoiceDate	0
UnitPrice	0
CustomerID	135080
Country	0

	Quantity	UnitPrice	CustomerID
count	541909.000000	541909.000000	406829.000000
mean	9.552250	4.611114	15287.690570
std	218.081158	96.759853	1713.600303
min	-80995.000000	-11062.060000	12346.000000
25%	1.000000	1.250000	13953.000000
50%	3.000000	2.080000	15152.000000
75%	10.000000	4.130000	16791.000000
max	80995.000000	38970.000000	18287.000000

Dataset

[Back to Agenda Page](#)



Analysis the Customers based on 3 factors:

[Back to Agenda Page](#)

**Total amount a
customer spent**

Quantity * UnitPrice

**Number of
transaction**

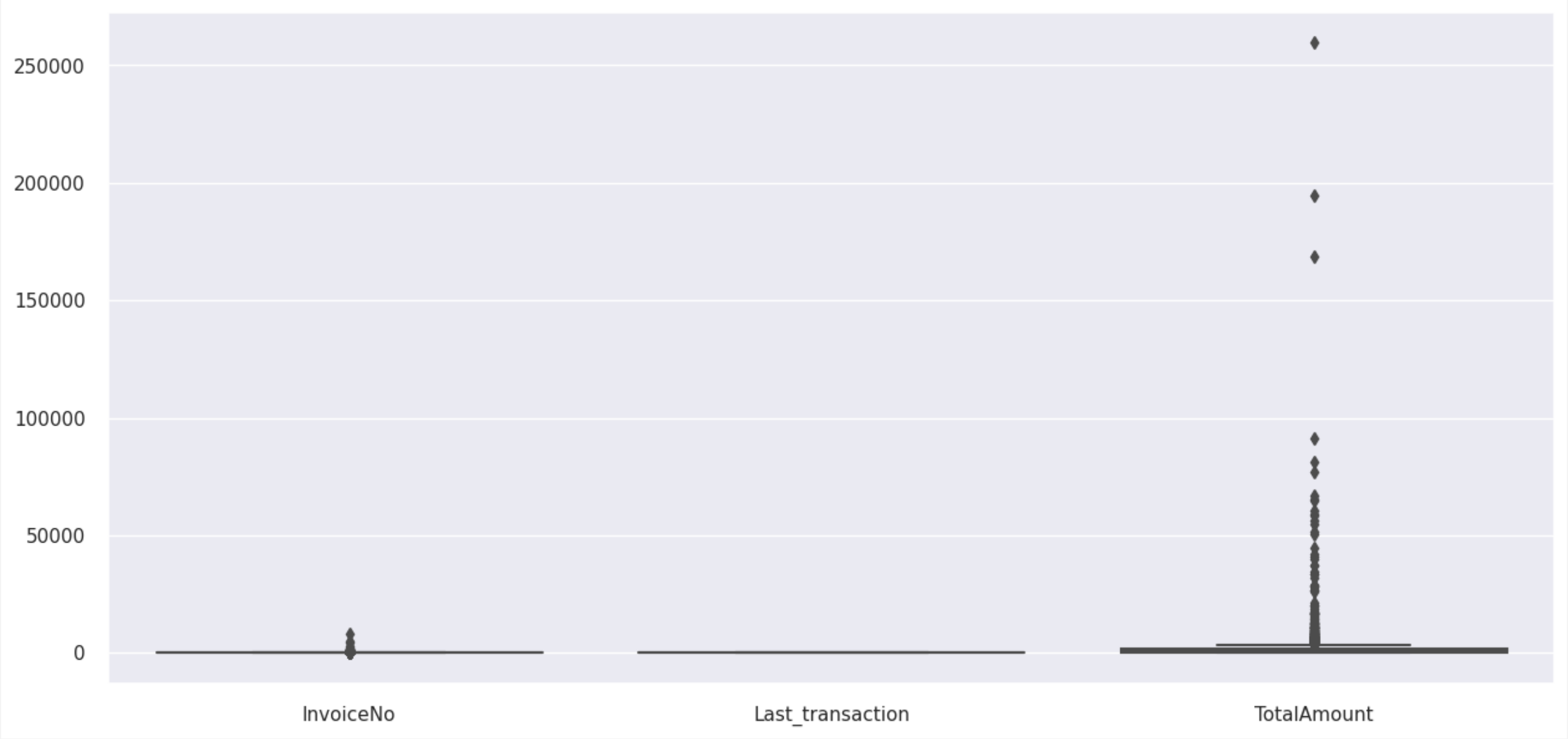
Number of InvoiceNo by Customer

**Last
transaction:**

days since last transaction

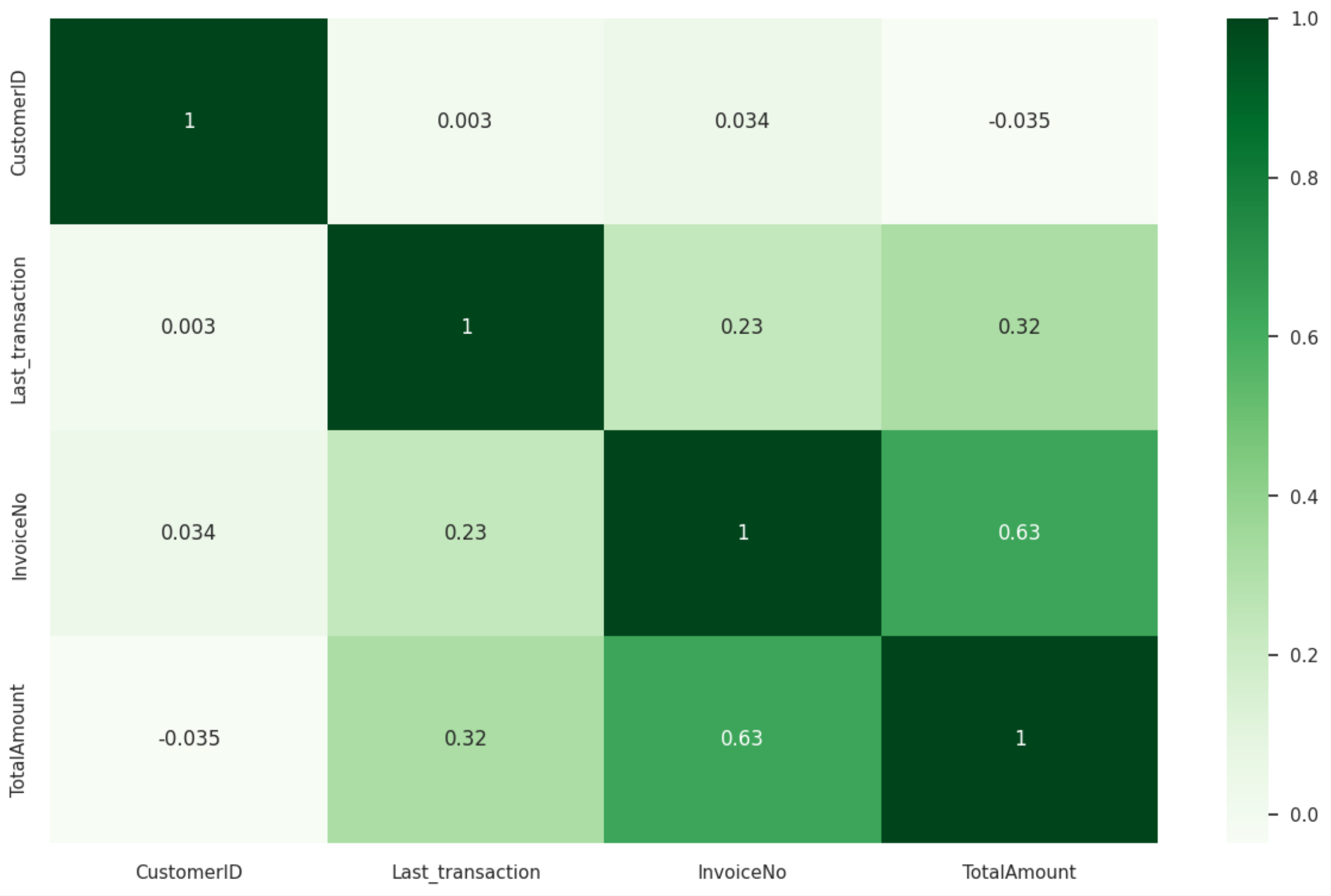
Outliers

[Back to Agenda Page](#)

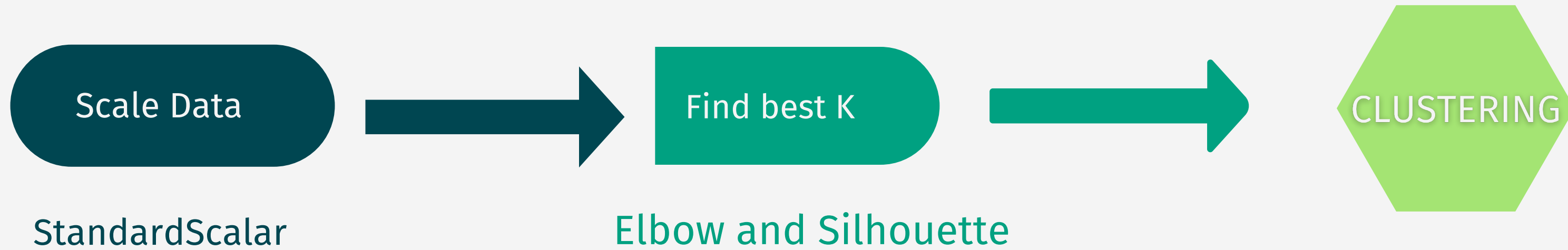


correlation

[Back to Agenda Page](#)

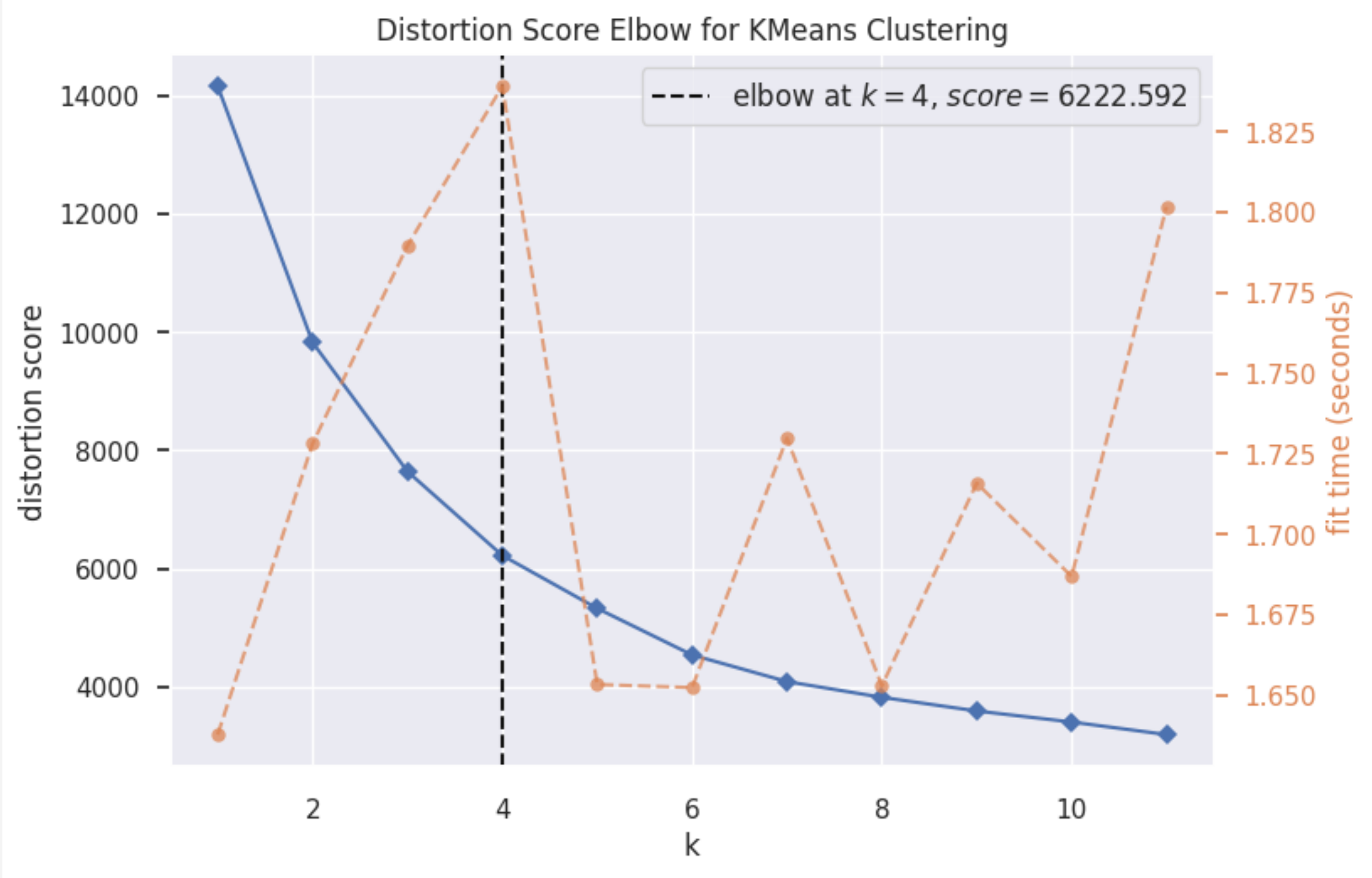


K-means and Steps



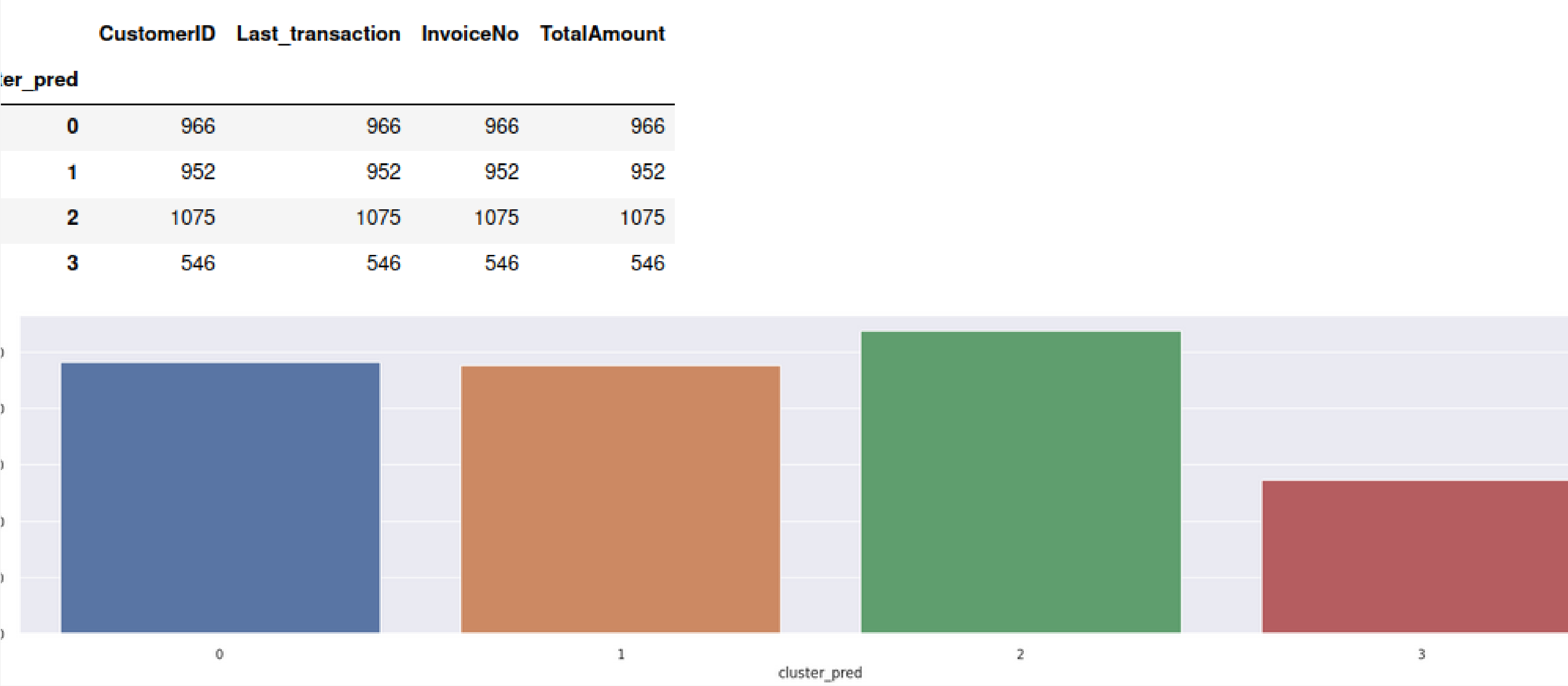
Elbow method

[Back to Agenda Page](#)



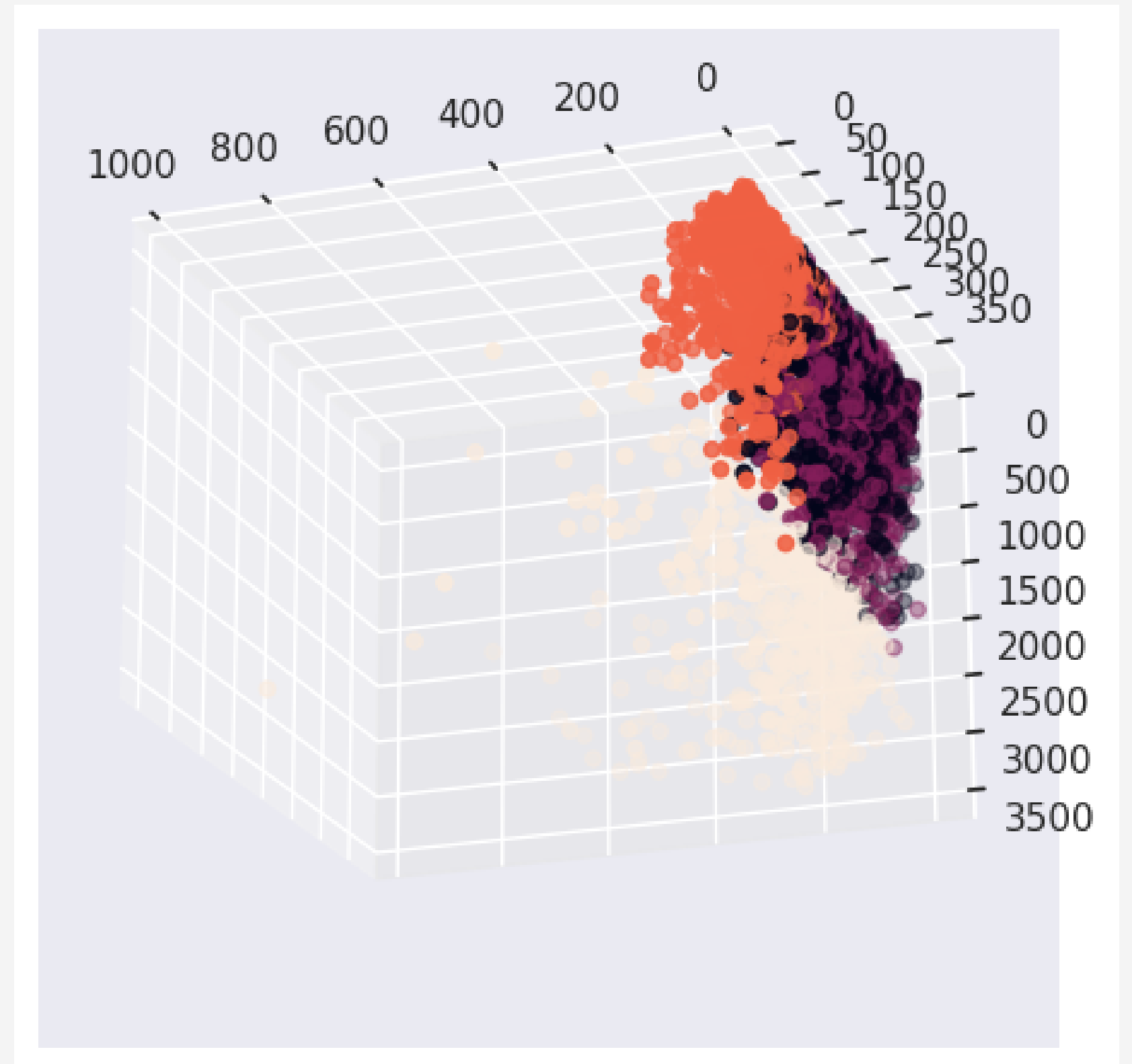
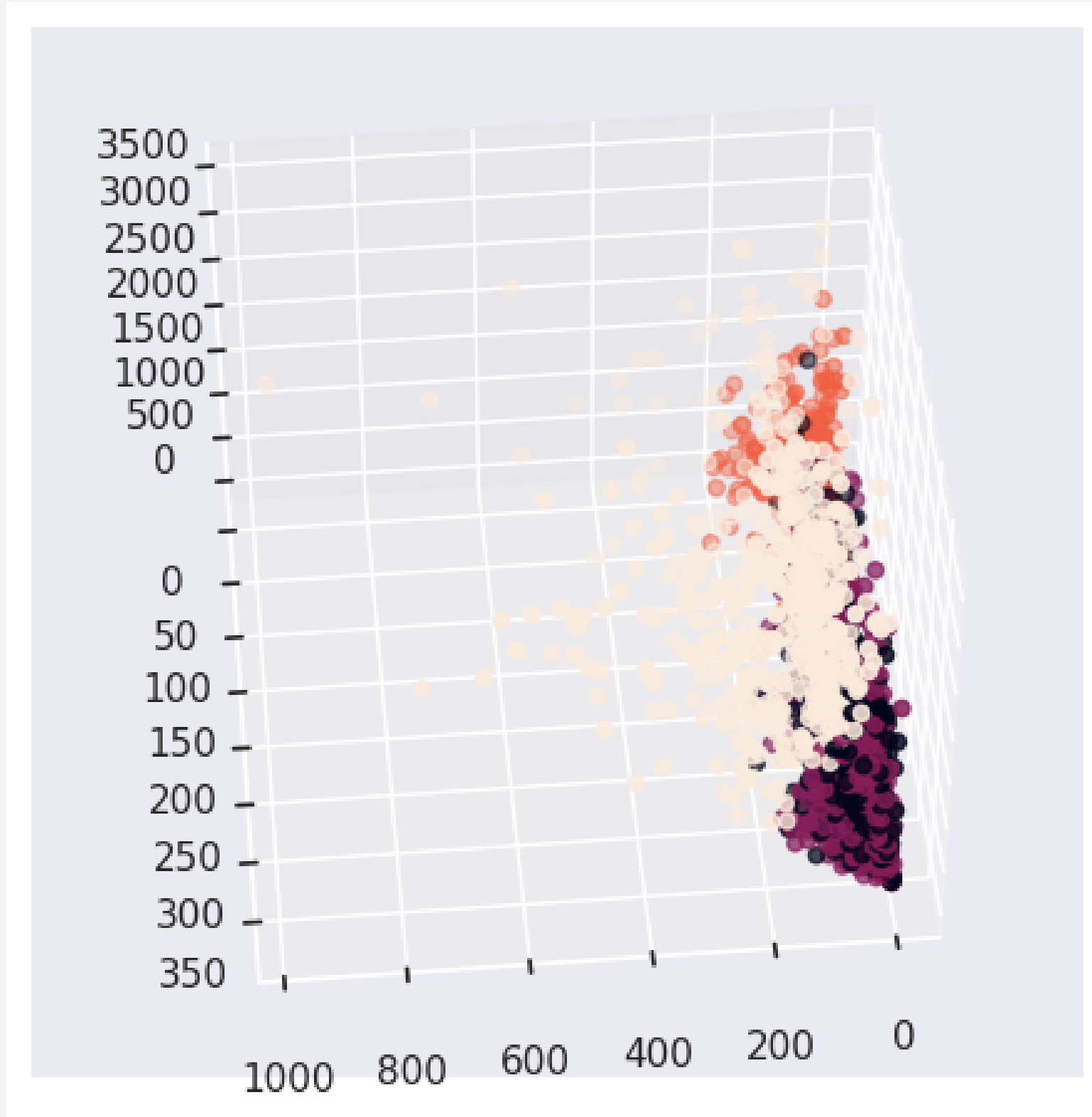
K-means

[Back to Agenda Page](#)



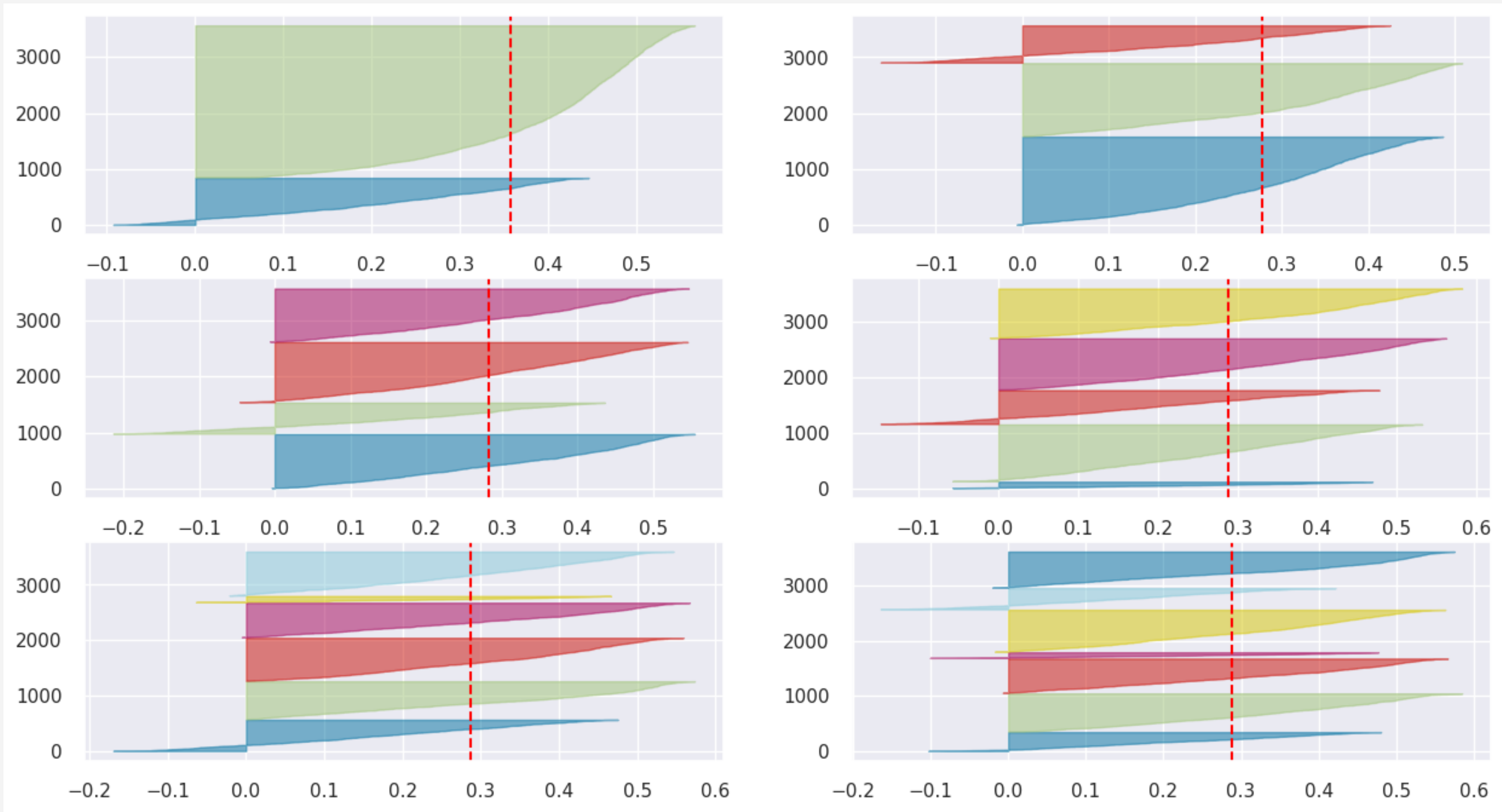
3D - Clusters

[Back to Agenda Page](#)



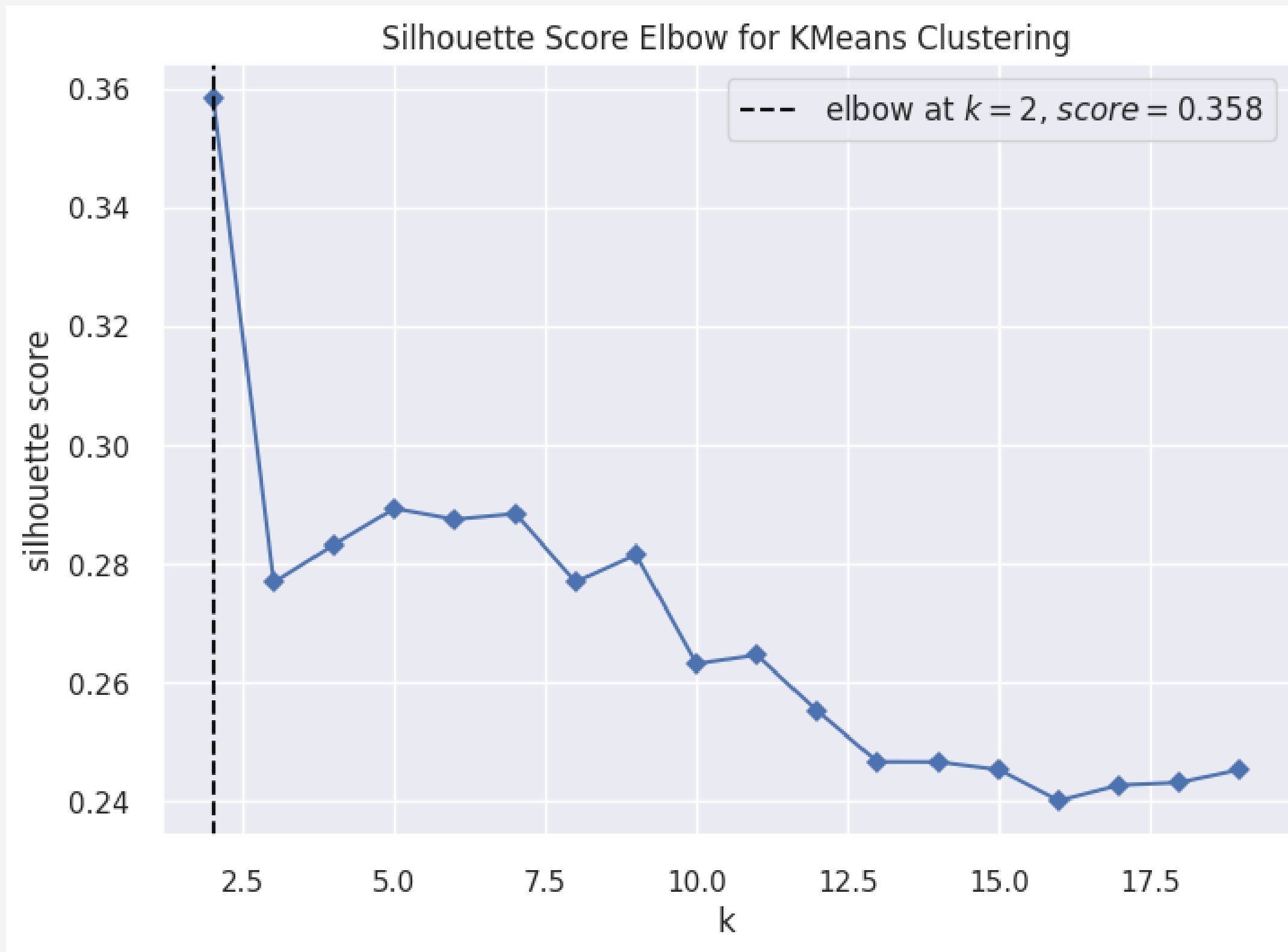
Silhouette method

[Back to Agenda Page](#)



Silhouette method

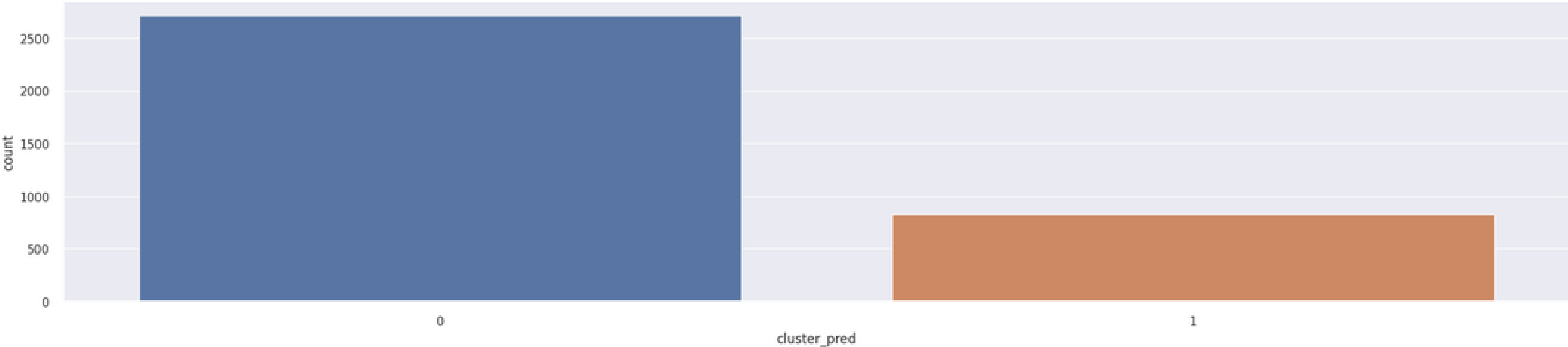
[Back to Agenda Page](#)



Silhouette method

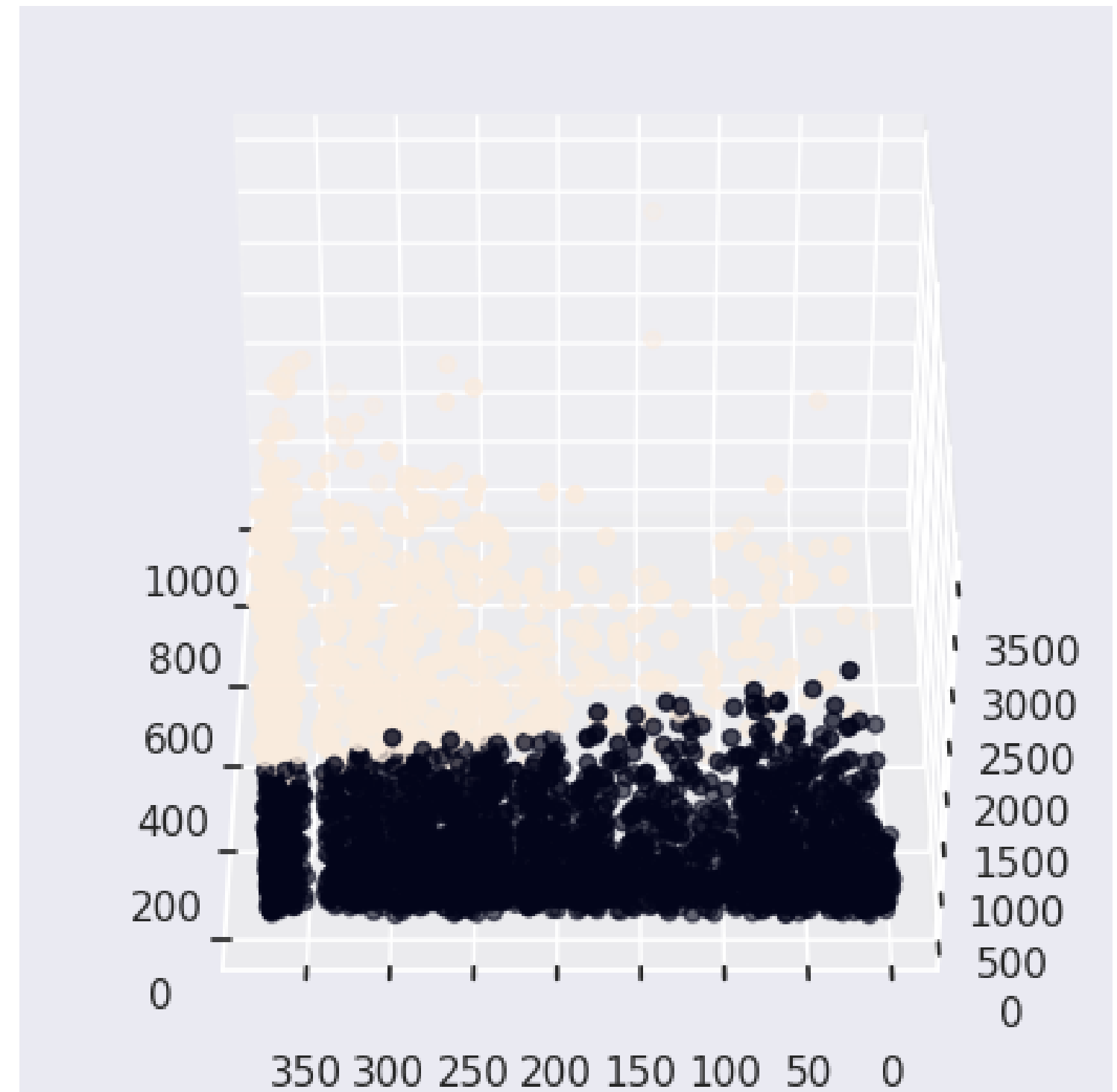
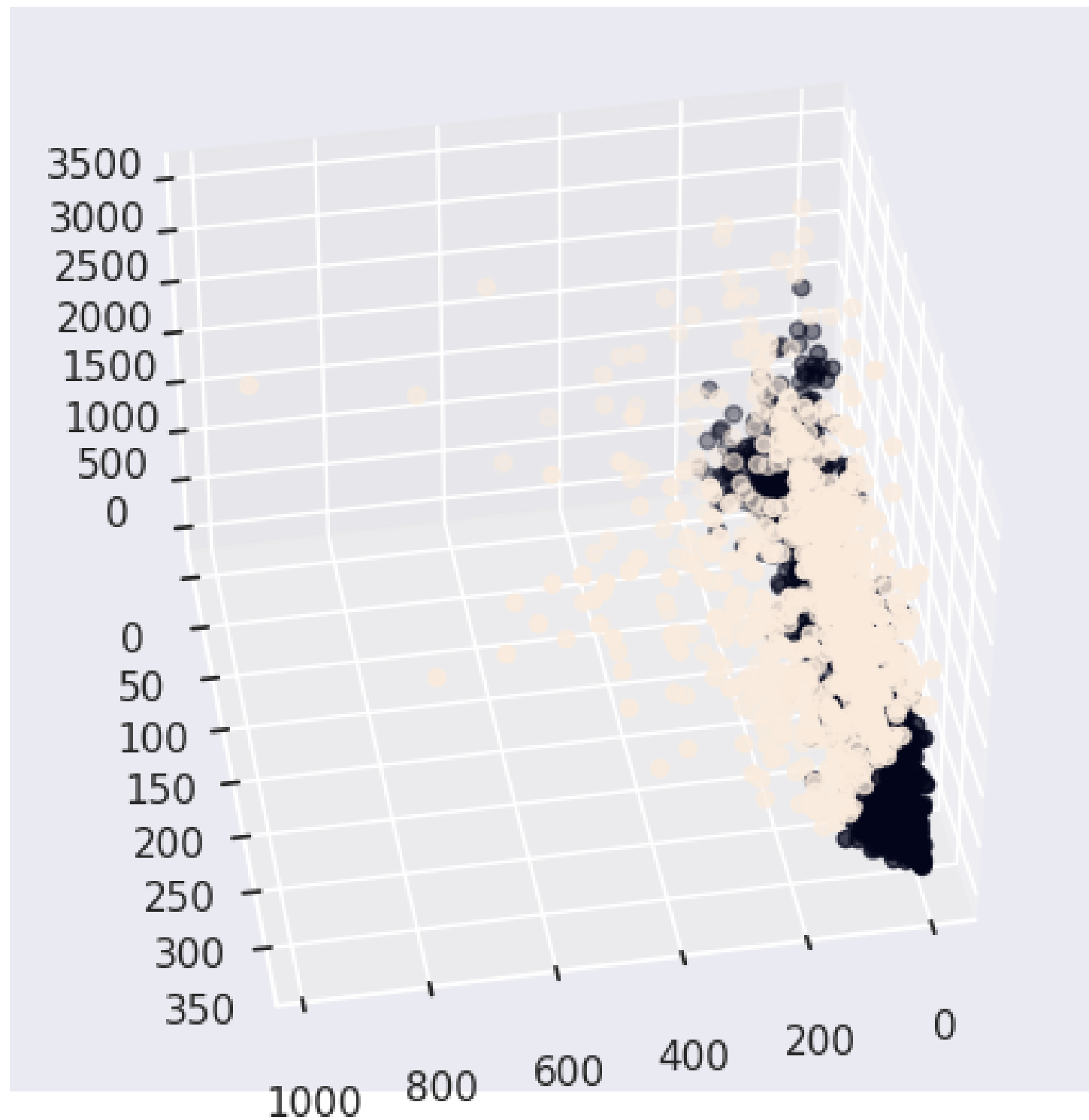
[Back to Agenda Page](#)

	CustomerID	Last_transaction	InvoiceNo	TotalAmount
cluster_pred				
0	2710	2710	2710	2710
1	829	829	829	829



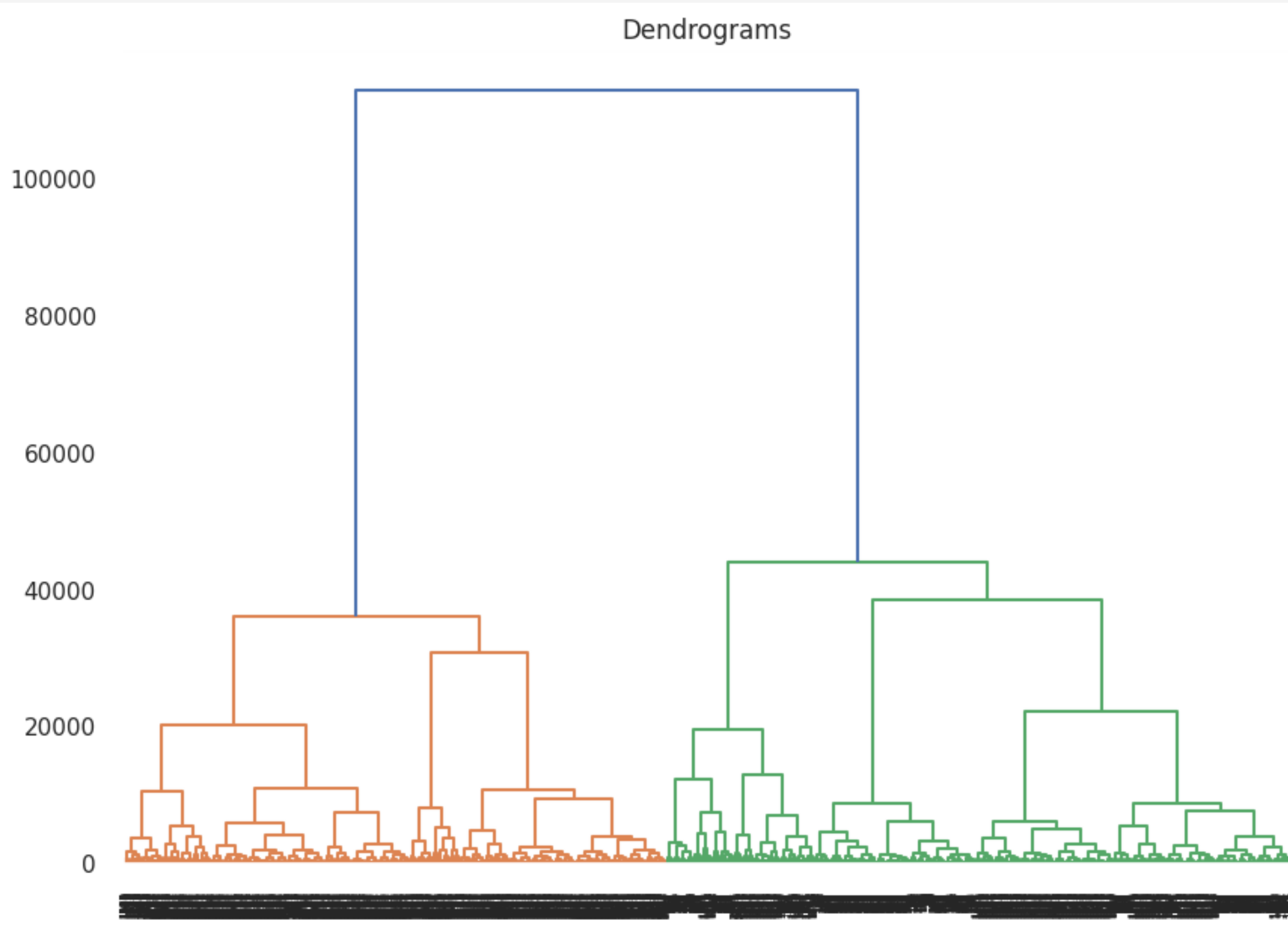
3D - Clusters

[Back to Agenda Page](#)



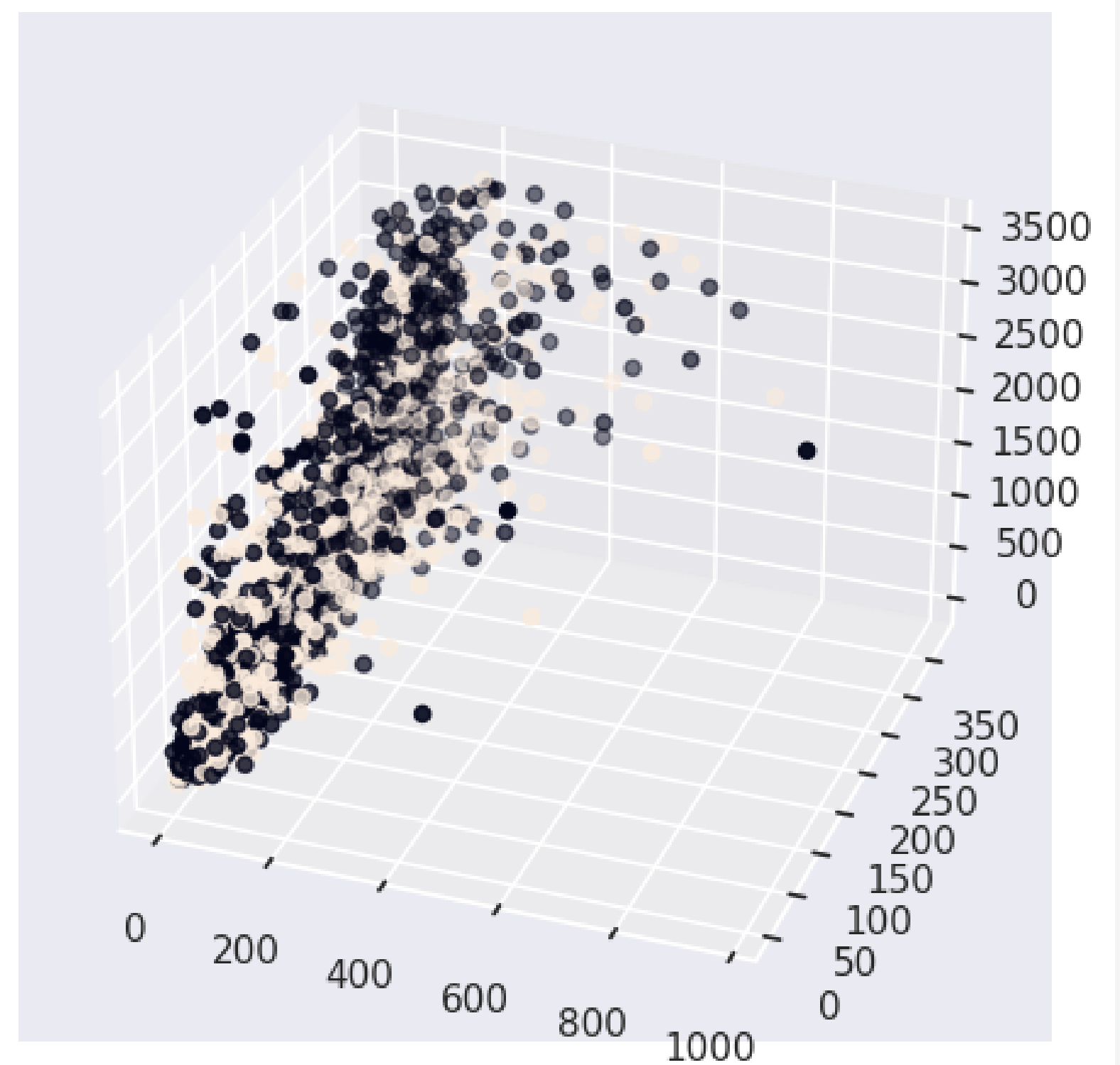
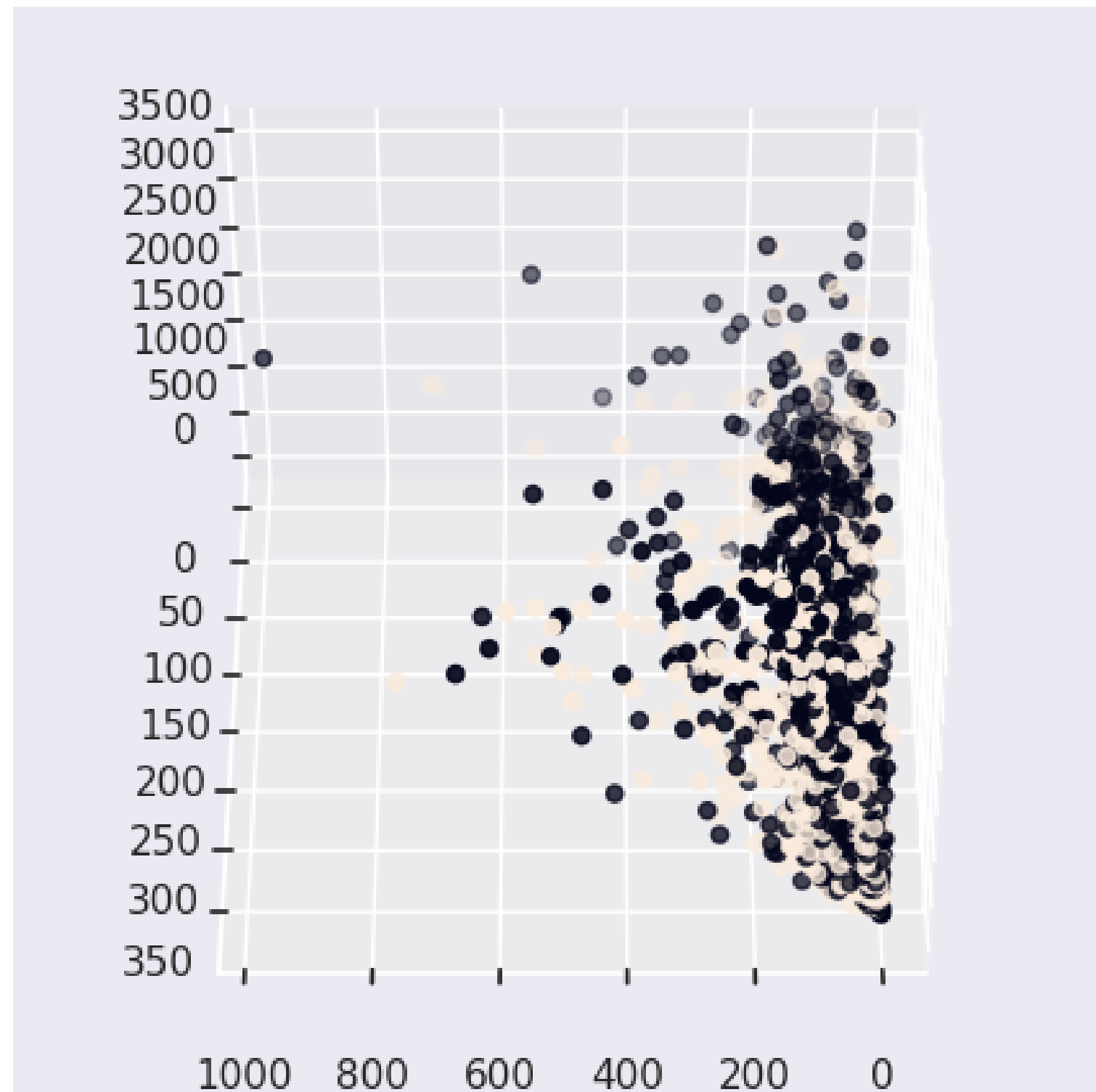
Hierarchical Agglomerative Clustering

[Back to Agenda Page](#)



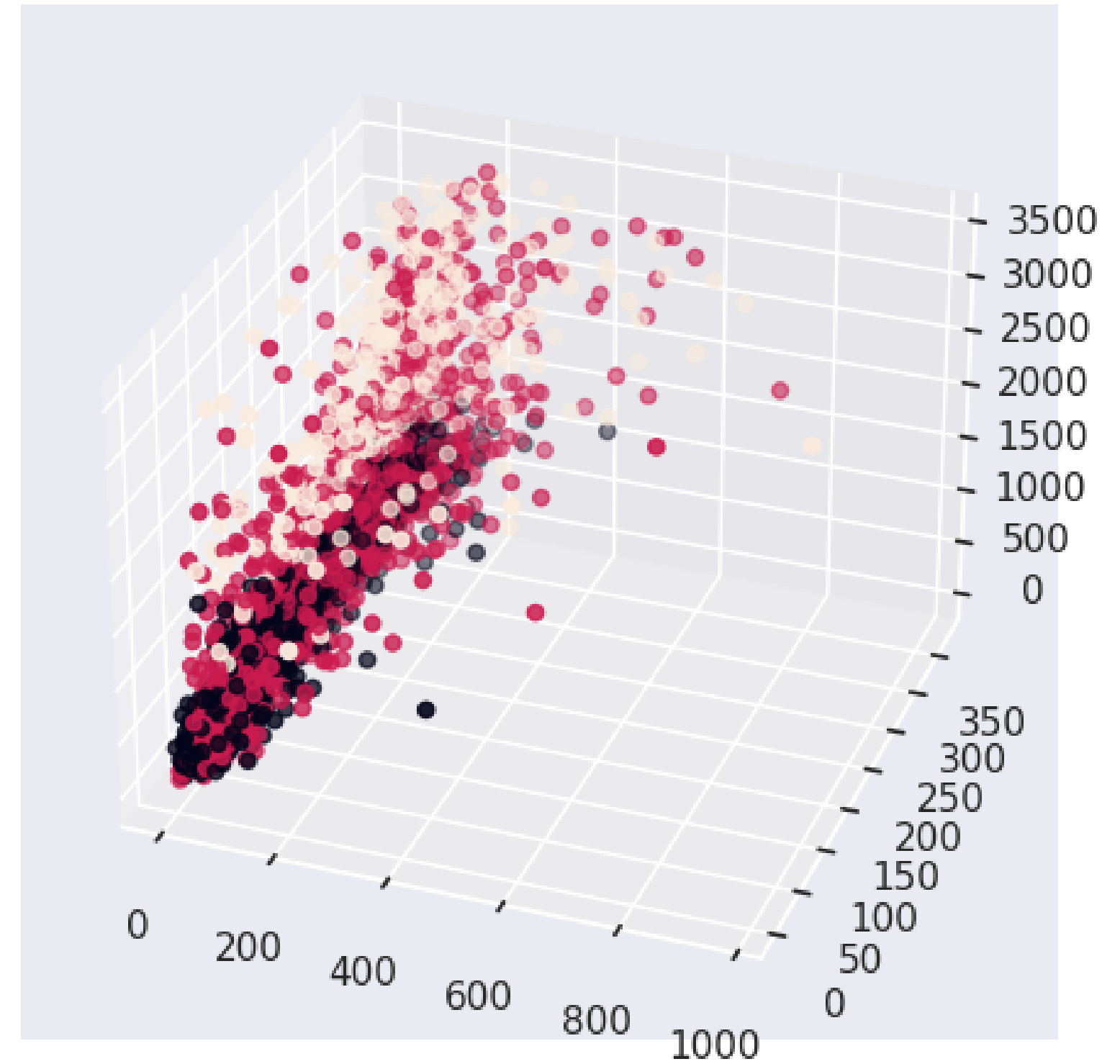
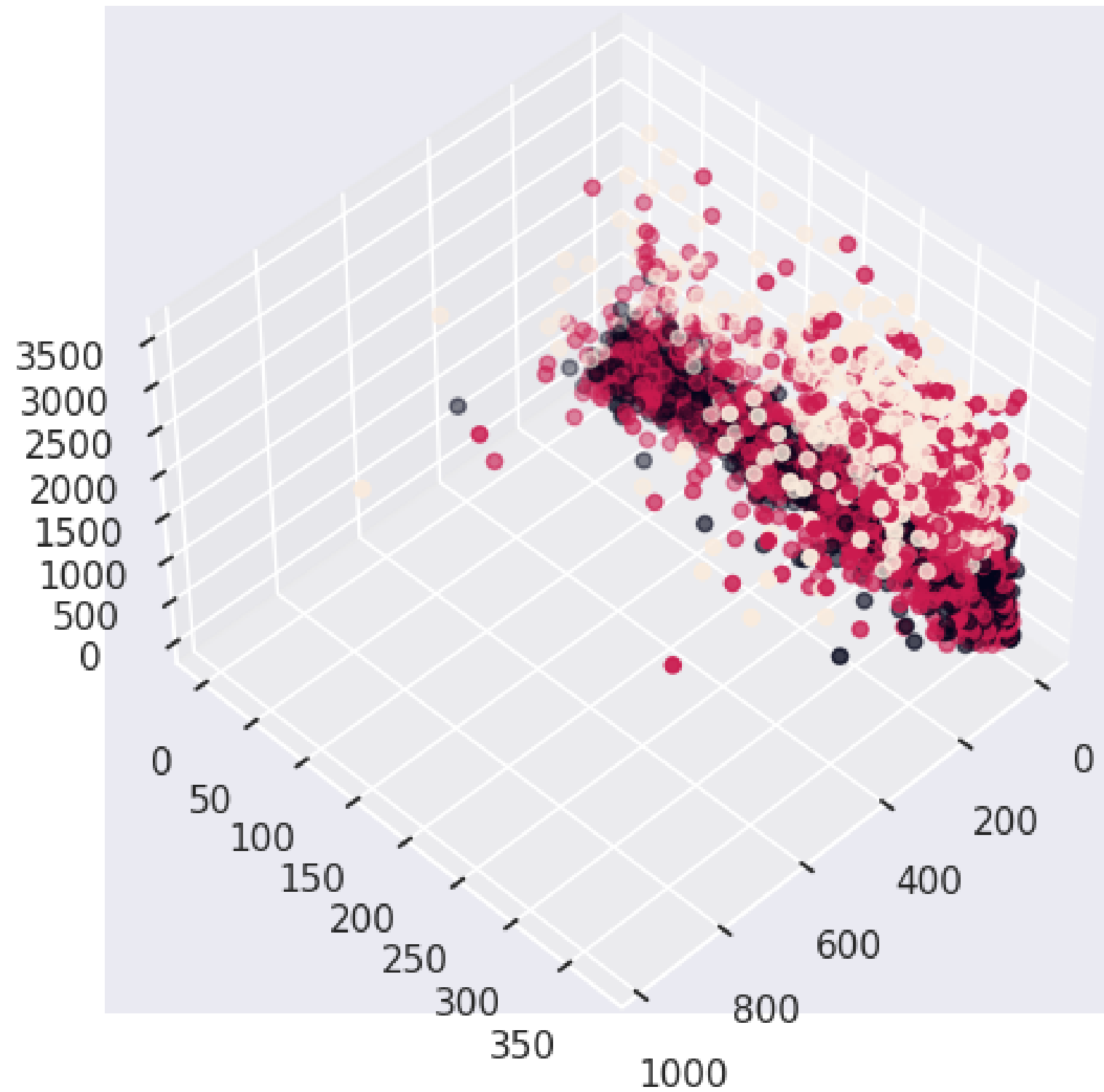
3D - Clusters


[Back to Agenda Page](#)



3D - Clusters

[Back to Agenda Page](#)



A decorative graphic on the right side of the slide consists of several overlapping hexagons in various shades of teal, dark blue, light green, and yellow, arranged in a diagonal pattern from the top left towards the bottom right.

Do you have any questions?

Thank you!

[Back to Agenda Page](#)