

12_customer_seg_bankchurndata_report

December 23, 2022

1 Customer Segmentation Report

```
c:\Users\Sebas\anaconda3\envs\customer_cluster_app\lib\site-  
packages\tqdm\auto.py:22: TqdmWarning: IProgress not found. Please update  
jupyter and ipywidgets. See  
https://ipywidgets.readthedocs.io/en/stable/user\_install.html  
from .autonotebook import tqdm as notebook_tqdm
```

2 Import Data

customer_id	credit_score	country	gender	age	tenure	balance
15634602	619	France	Female	42	2	0.00
15647311	608	Spain	Female	41	1	83807.86
15619304	502	France	Female	42	8	159660.80
15701354	699	France	Female	39	1	0.00
15737888	850	Spain	Female	43	2	125510.82

customer_id	products_number	credit_card	active_member	estimated_salary
15634602	1	yes	yes	101348.88
15647311	1	no	yes	112542.58
15619304	3	yes	no	113931.57
15701354	2	no	no	93826.63
15737888	1	yes	yes	79084.10

3 Load Model Artifacts

4 Transform Dataset

customer_id	credit_score	age	tenure	balance	products_number
15634602	619.0	42.0	2.0	0.00	1.0
15647311	608.0	41.0	1.0	83807.86	1.0
15619304	502.0	42.0	8.0	159660.80	3.0
15701354	699.0	39.0	1.0	0.00	2.0

15737888	850.0	43.0	2.0	125510.82	1.0
----------	-------	------	-----	-----------	-----

customer_id	estimated_salary	country	gender	credit_card	active_member
15634602	101348.88	France	Female	yes	yes
15647311	112542.58	Spain	Female	no	yes
15619304	113931.57	France	Female	yes	no
15701354	93826.63	France	Female	no	no
15737888	79084.10	Spain	Female	yes	yes

customer_id	cluster
15634602	7
15647311	0
15619304	0
15701354	9
15737888	1

5 Data Visualization

0	1423
4	1011
6	947
7	802
10	758
3	509
5	489
13	457
12	455
11	433
16	397
1	388
2	379
14	370
8	360
15	318
9	314
18	112
17	78

Name: cluster, dtype: int64

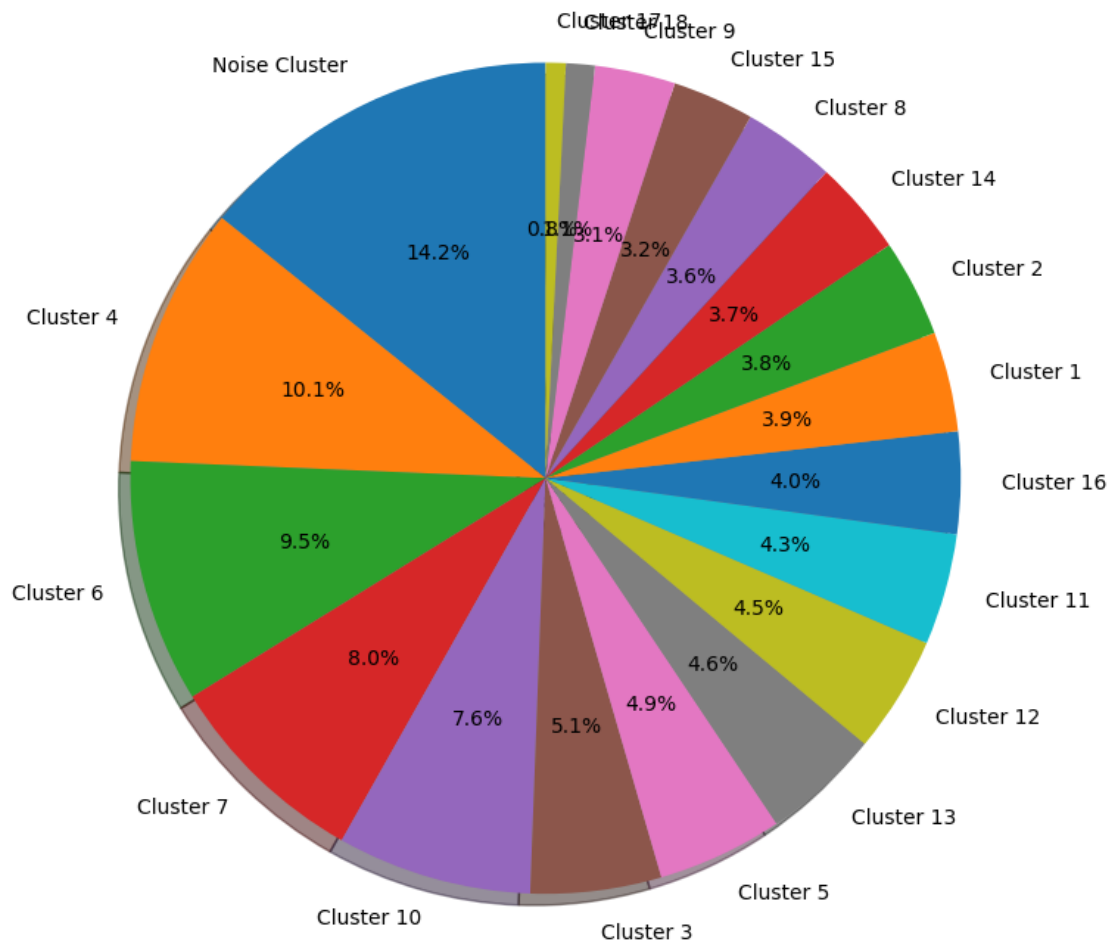
5.1 What does the Cluster Distribution Look Like?

```
['Noise Cluster',
 'Cluster 4',
 'Cluster 6',
```

```

'Cluster 7',
'Cluster 10',
'Cluster 3',
'Cluster 5',
'Cluster 13',
'Cluster 12',
'Cluster 11',
'Cluster 16',
'Cluster 1',
'Cluster 2',
'Cluster 14',
'Cluster 8',
'Cluster 15',
'Cluster 9',
'Cluster 18',
'Cluster 17']

```



6 Determine Cluster Centers

The default value of `numeric_only` in `DataFrameGroupBy.mean` is deprecated. In a future version, `numeric_only` will default to `False`. Either specify `numeric_only` or select only columns which should be valid for the function.

`['estimated_salary']` did not aggregate successfully. If any error is raised this will raise in a future version of pandas. Drop these columns/ops to avoid this warning.

	credit_score	age	tenure	balance	products_number \
cluster					
0	654.791286	40.895832	4.873507	96027.727505	1.575290
1	651.958763	40.005147	4.976804	56574.898170	1.529080
2	653.248021	39.643792	4.767810	120840.295963	1.545775
3	652.300589	39.065265	5.074656	65603.672141	1.483301
4	652.347181	39.125613	5.091988	60009.775242	1.538920
5	654.233129	39.806855	4.950920	120010.719489	1.558551
6	649.157339	36.931127	5.139388	64589.570380	1.479707
7	653.920200	39.450947	4.976309	57985.791808	1.520303
8	648.922222	38.044444	5.000000	53953.735278	1.589675
9	646.060510	36.980183	5.254777	59743.896274	1.551856
10	646.866755	37.726619	4.869393	59188.859235	1.518643
11	645.427252	37.316397	5.138568	57777.952818	1.548651
12	642.184615	38.570450	5.281319	121028.159275	1.498568
13	647.803063	40.041816	5.306346	119630.973589	1.451054
14	643.775676	37.035135	5.072973	66784.688405	1.493421
15	650.323899	38.300133	4.779874	57436.316195	1.526047
16	652.478589	39.270075	4.682620	62991.425189	1.515279
17	630.243590	35.522788	5.628205	13006.260385	1.923077
18	666.214286	38.078370	4.964286	118225.984911	1.151786

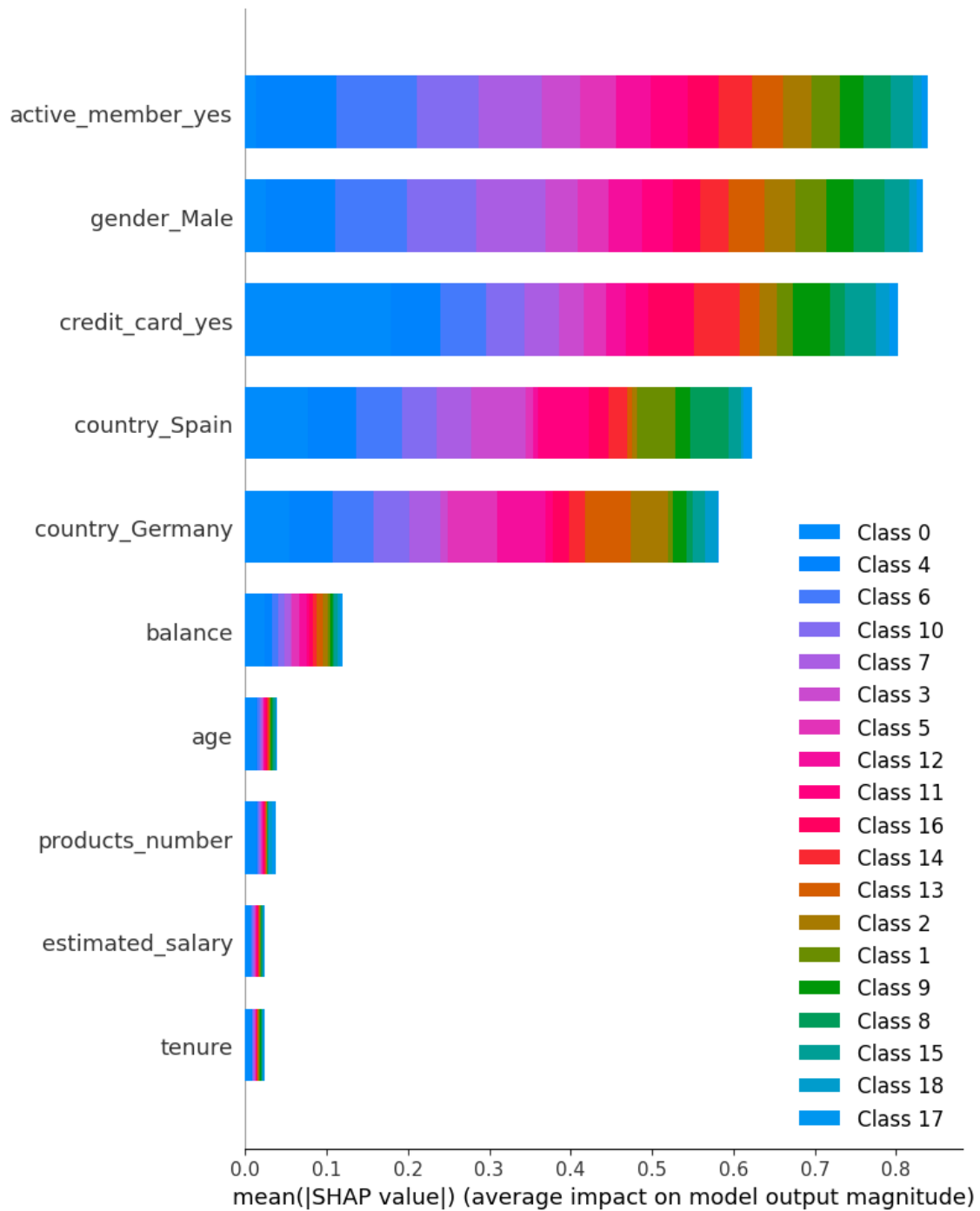
	estimated_salary	country	gender	credit_card	active_member
cluster					
0	101666.361230	Spain	Female	no	yes
1	99109.426108	Spain	Female	yes	yes
2	101213.050765	Germany	Female	yes	yes
3	95539.391591	Spain	Male	yes	yes
4	99398.792305	France	Male	yes	yes
5	100452.116708	Germany	Male	yes	yes
6	100604.350391	France	Male	yes	no
7	96281.148703	France	Female	yes	yes
8	97977.700222	Spain	Female	yes	no
9	102962.016688	France	Female	no	no
10	100490.747520	France	Female	yes	no
11	99644.155335	Spain	Male	yes	no
12	100575.535758	Germany	Male	yes	no
13	103880.558687	Germany	Female	yes	no

14	100799.680216	France	Male	no	no
15	104663.509465	France	Female	no	yes
16	99062.452040	France	Male	no	yes
17	100201.979615	Spain	Male	no	no
18	96488.275625	Germany	Male	no	yes

7 Explain Clusters

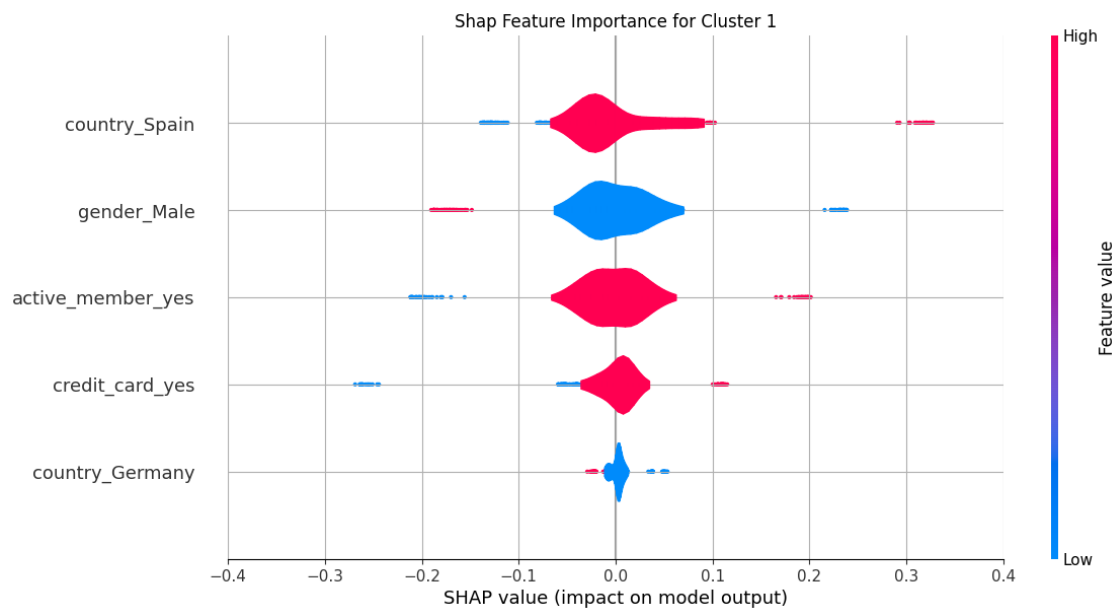
8 Custom Functions

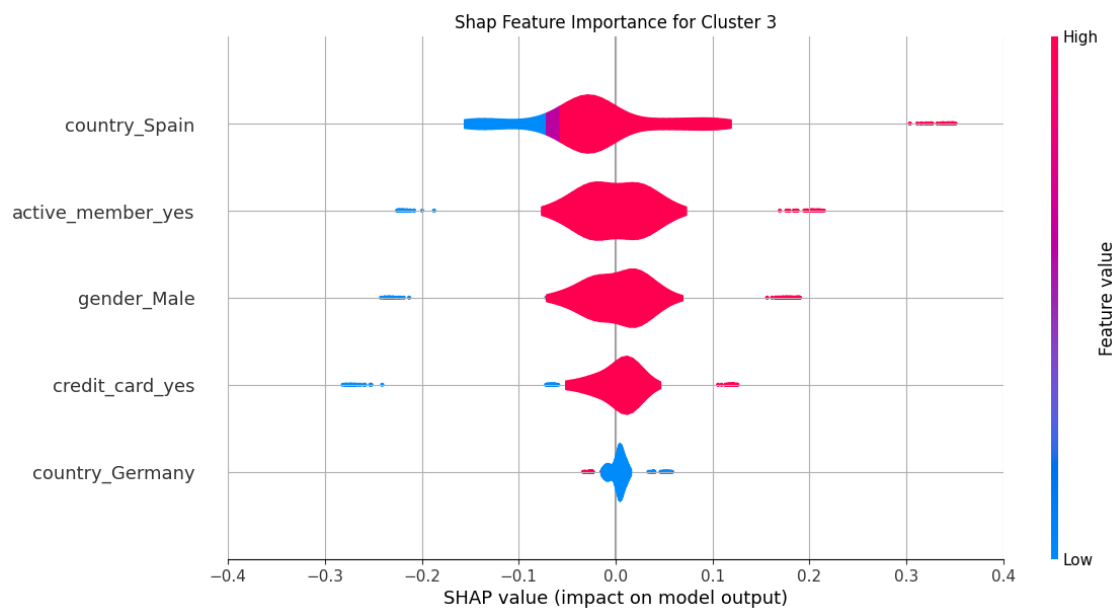
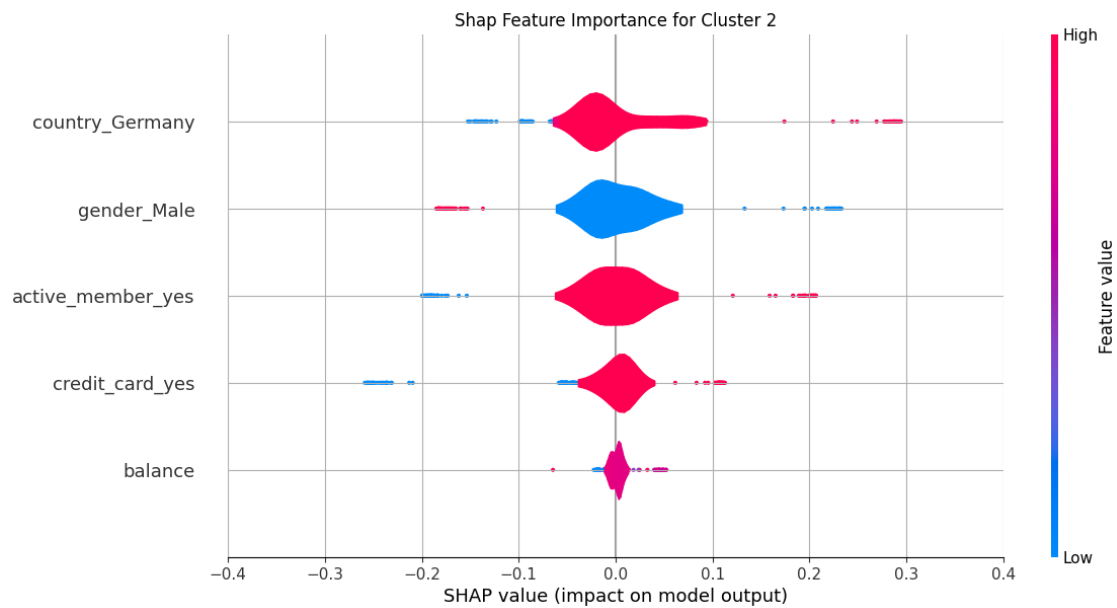
8.1 Average Feature Importance on Cluster Building

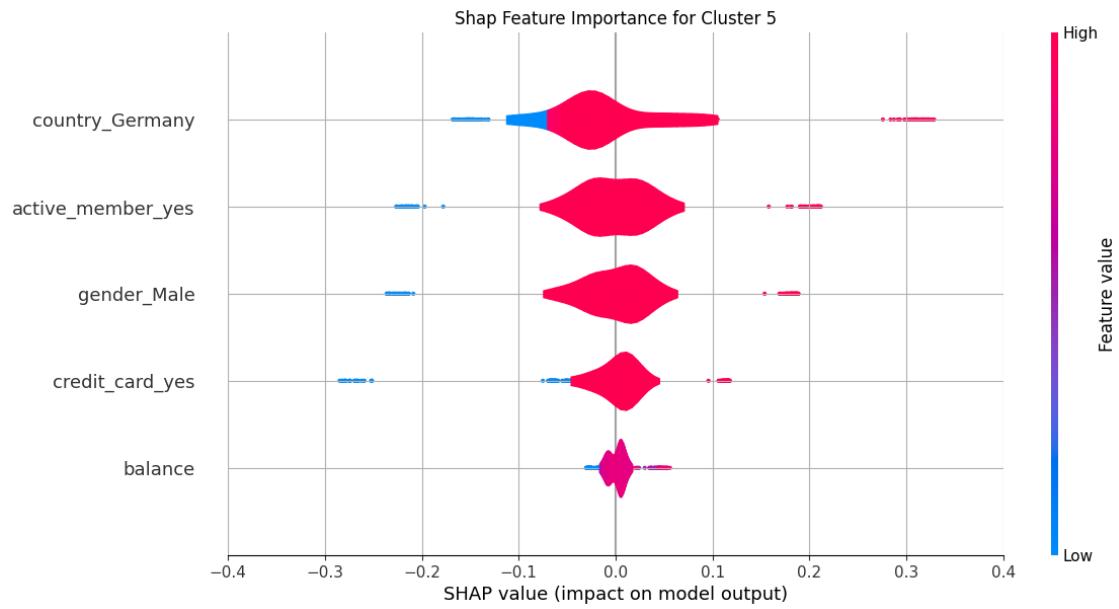
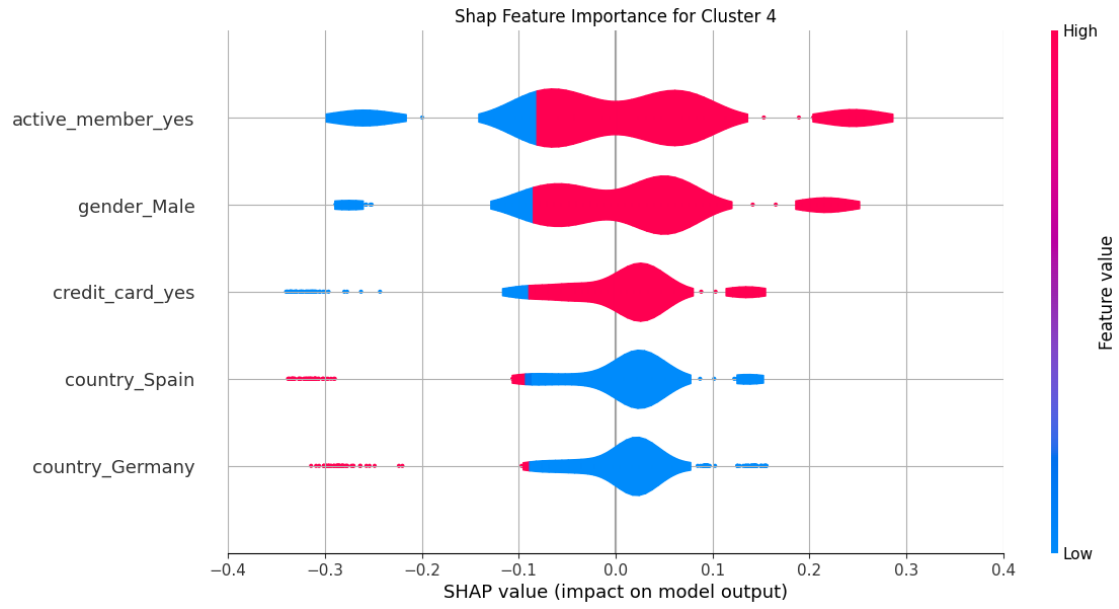


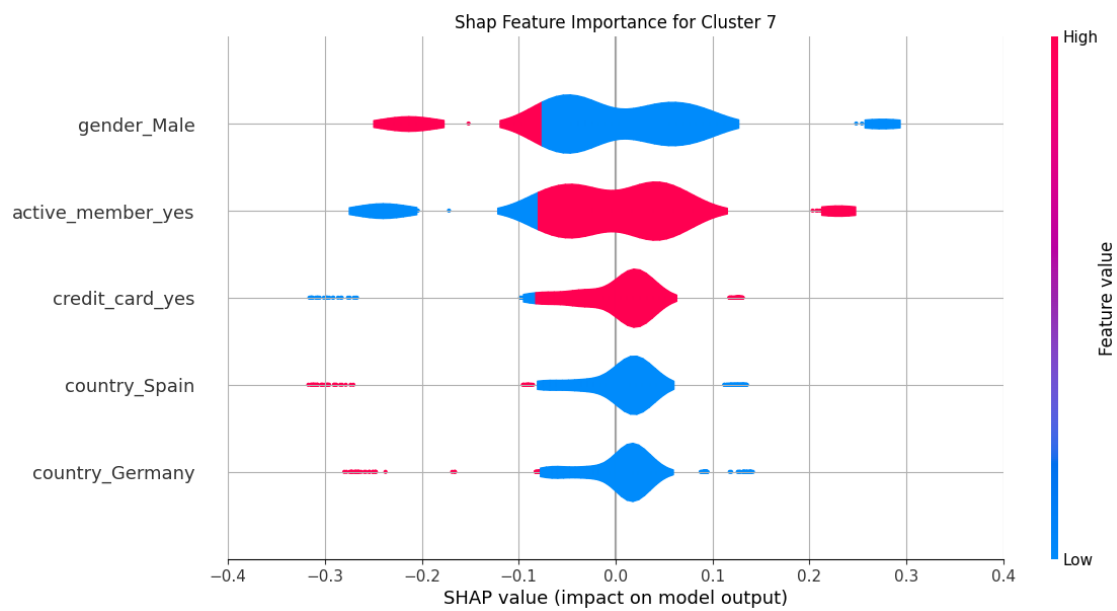
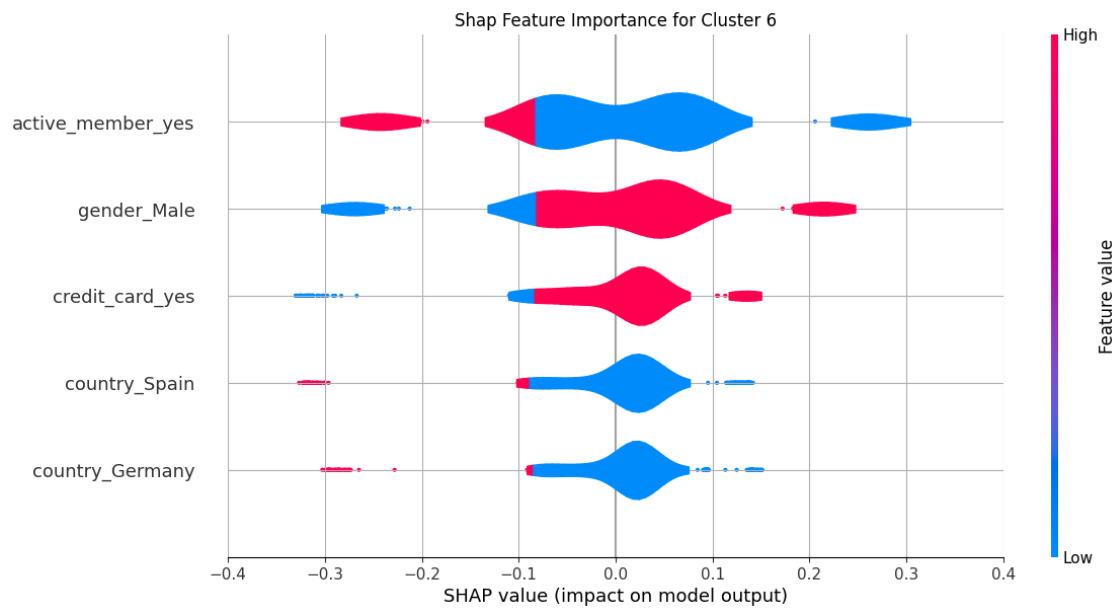
8.2 Explain Single Cluster vs Rest

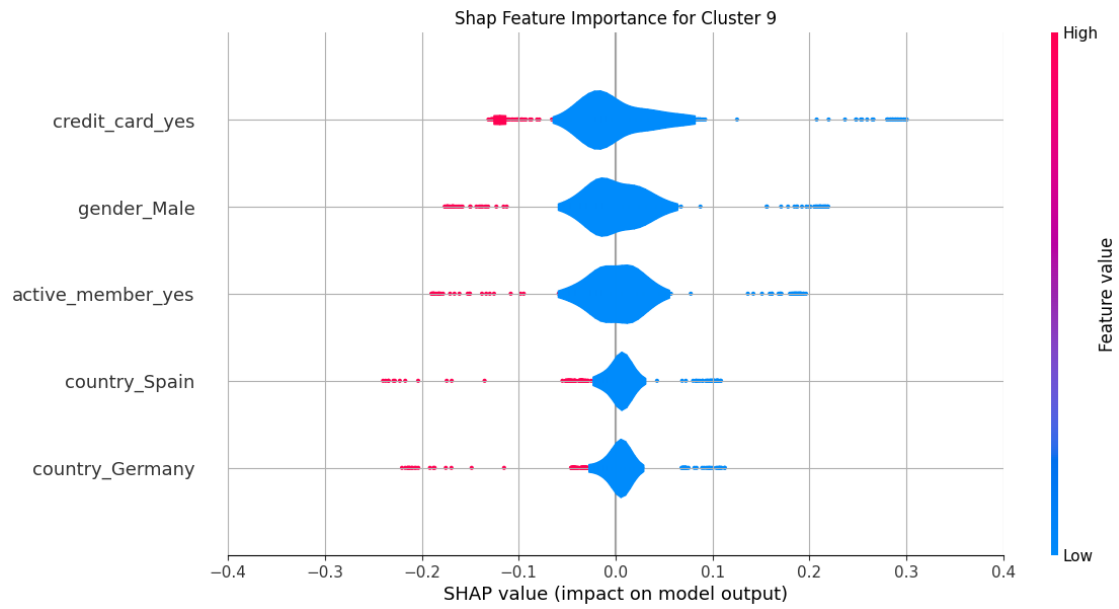
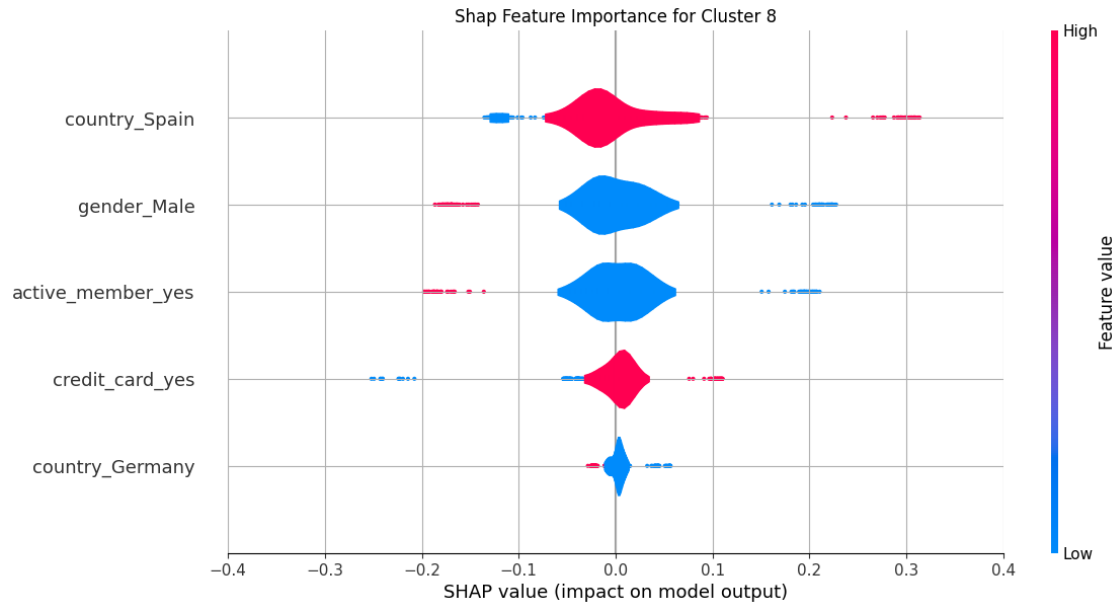
No data for colormapping provided via 'c'. Parameters 'vmin', 'vmax' will be ignored

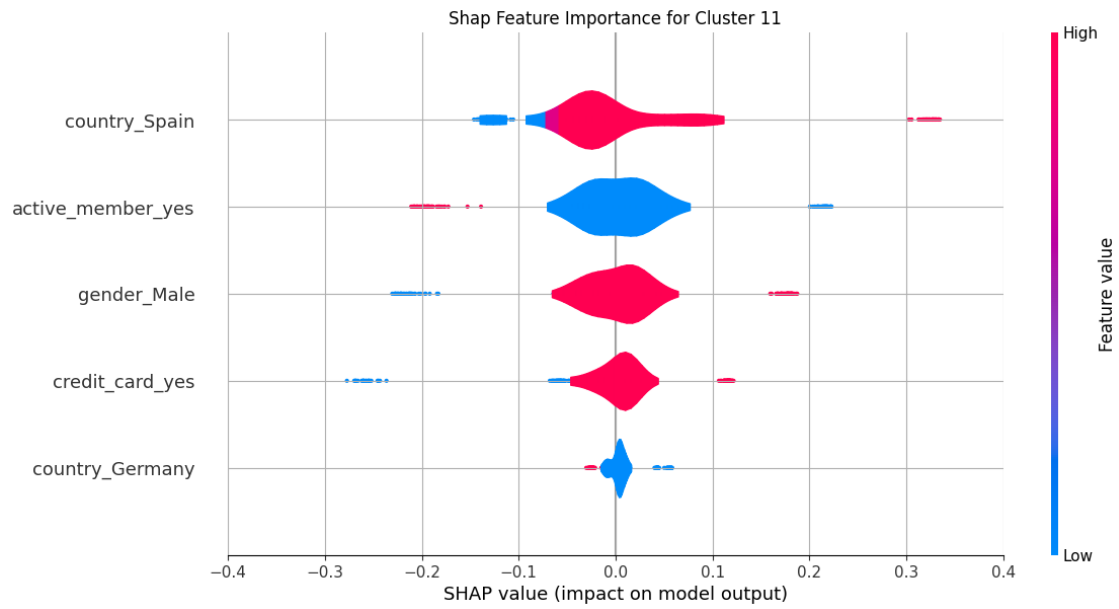
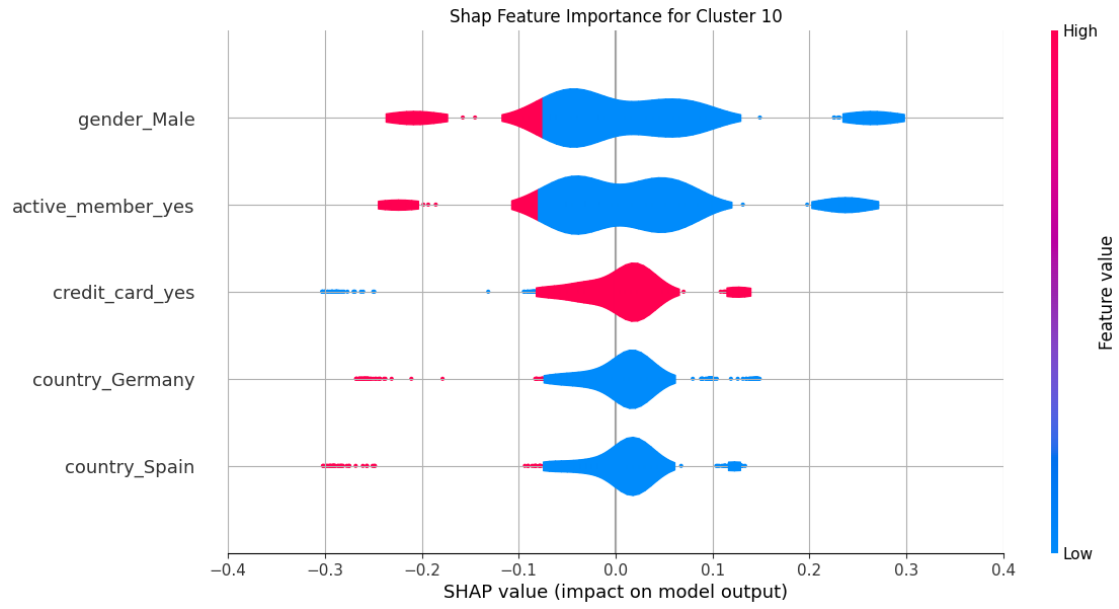


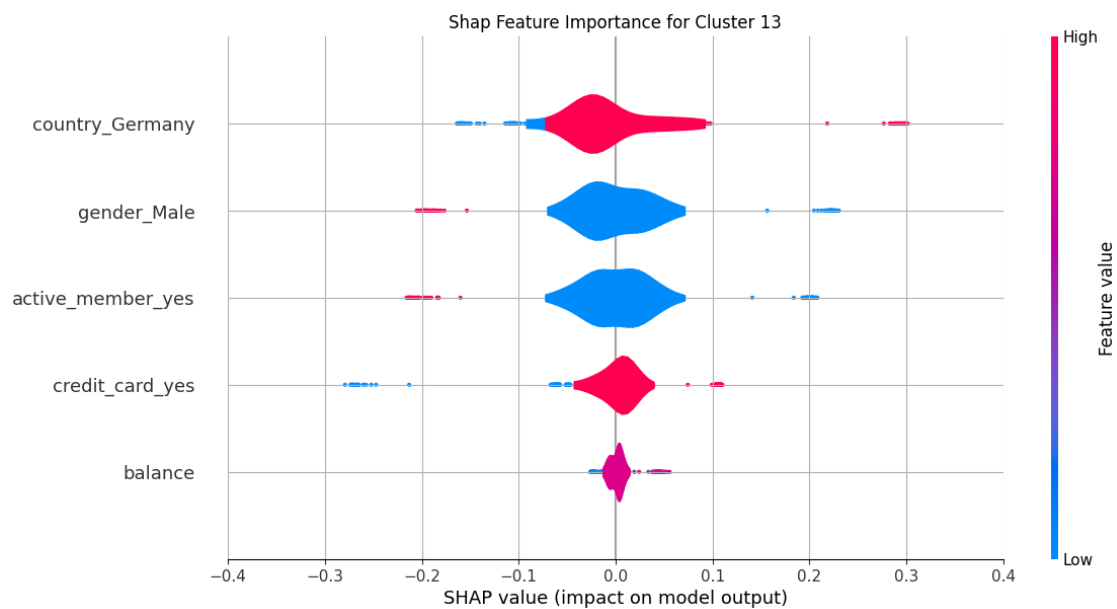
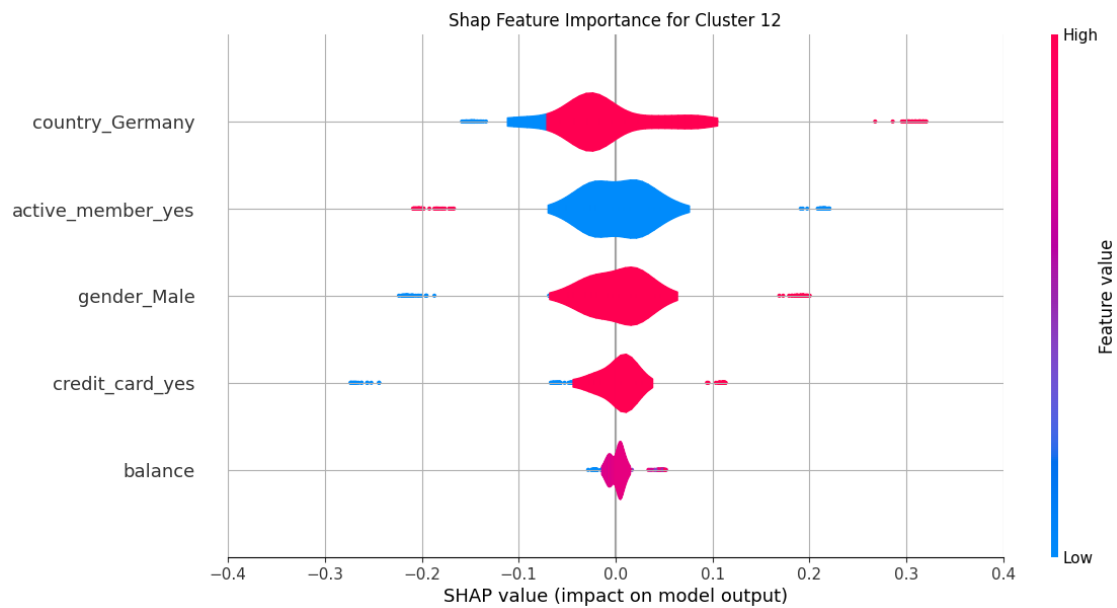


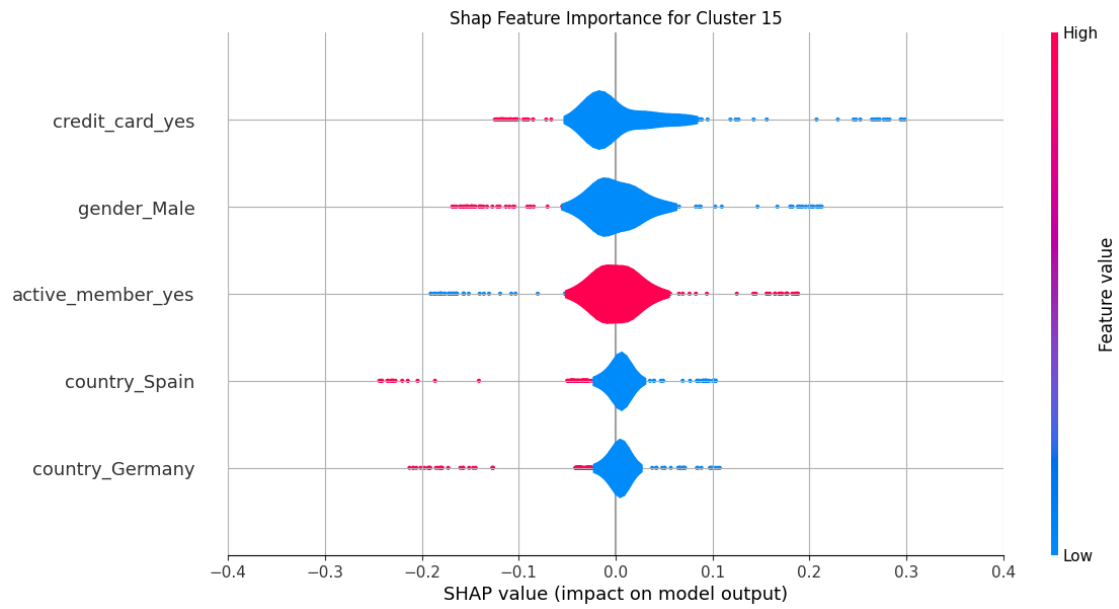
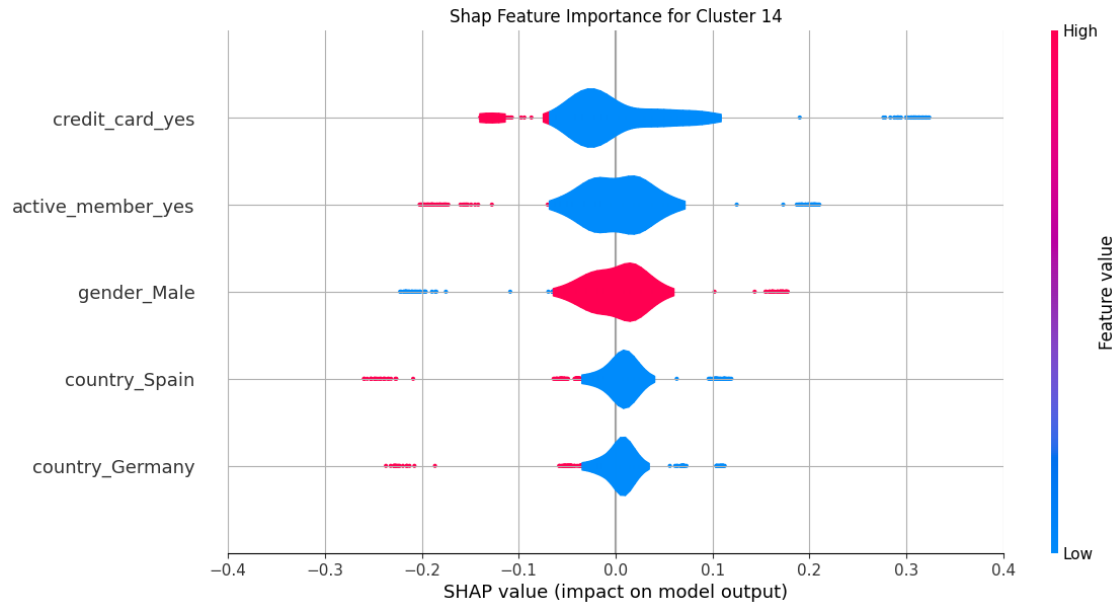


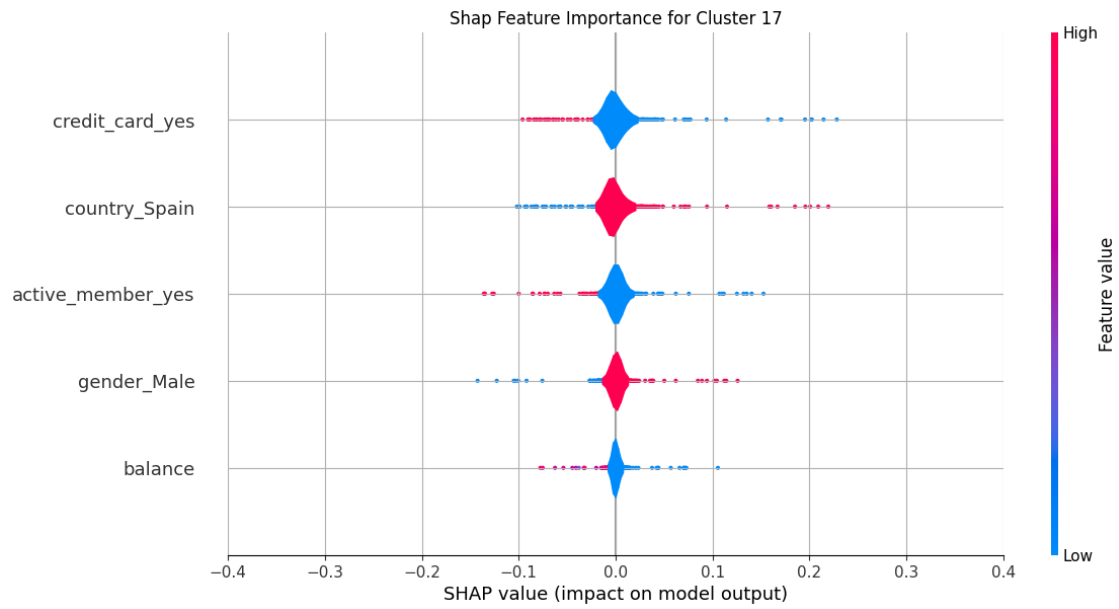
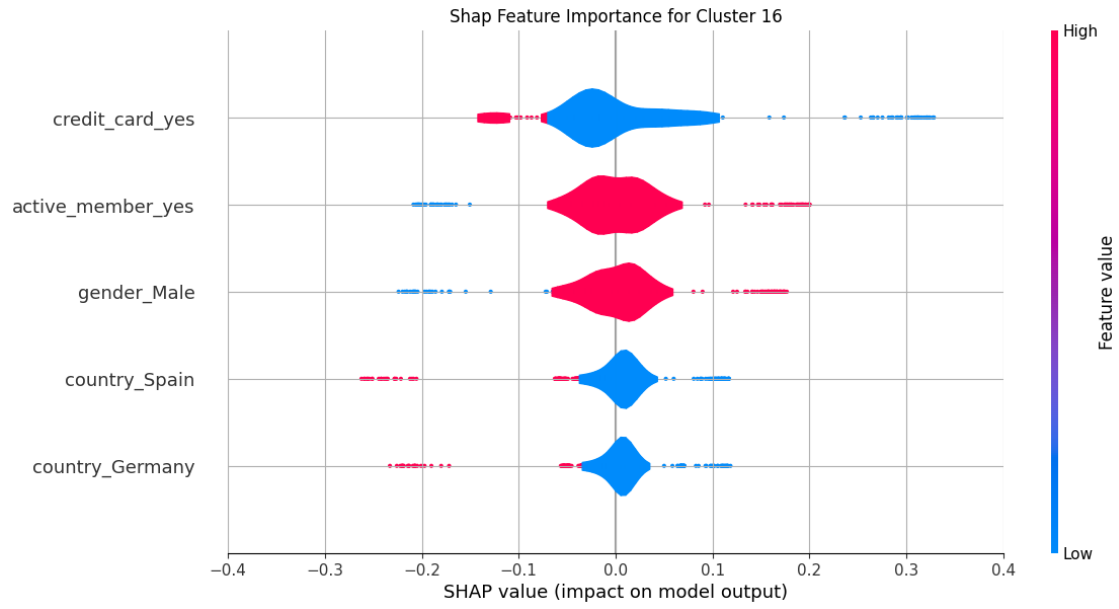


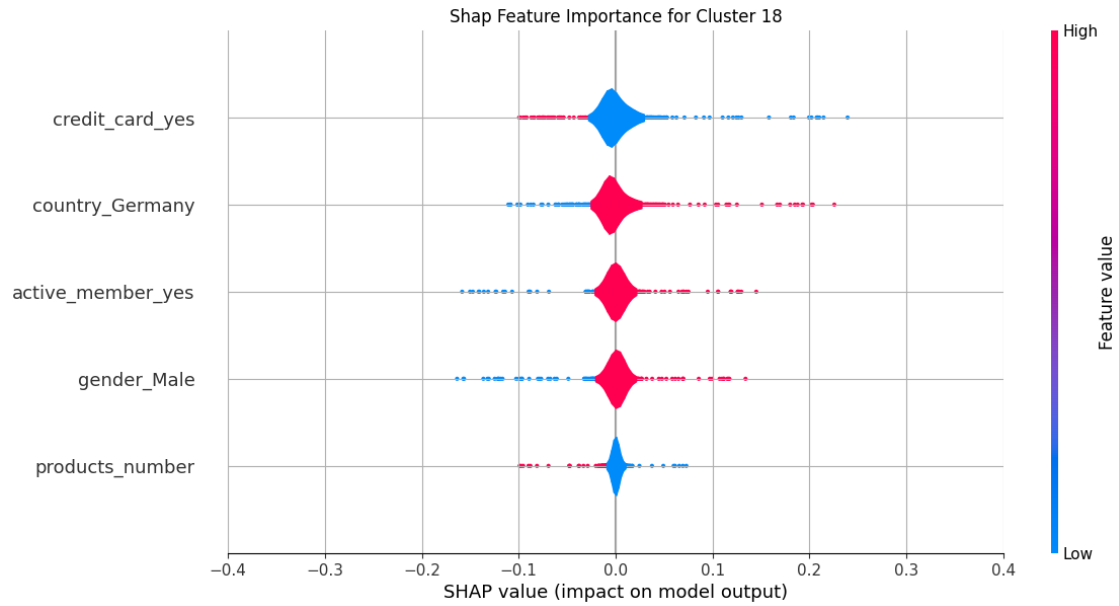










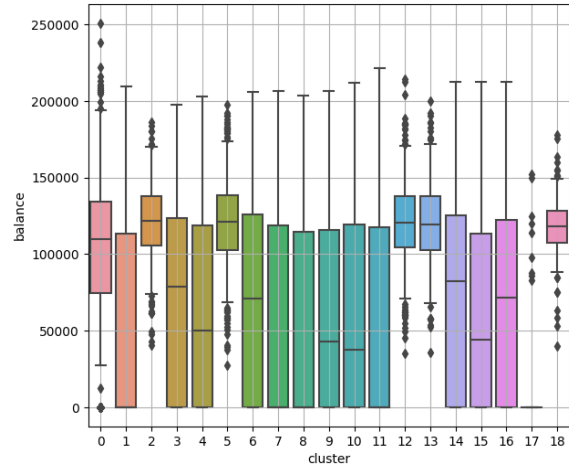
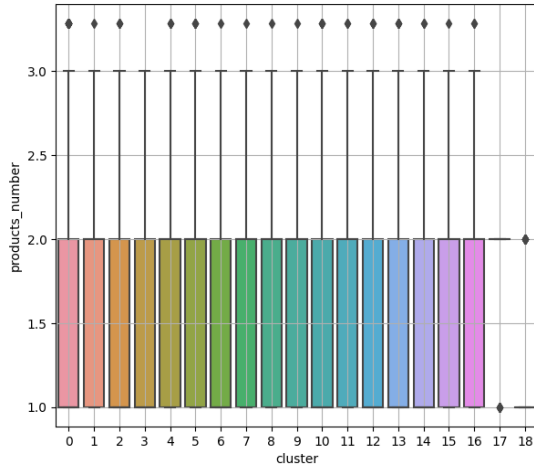


```
[ 'country_Germany',
  'gender_Male',
  'active_member_yes',
  'credit_card_yes',
  'balance']
```

```
[ 'active_member_yes',
  'country_Spain',
  'products_number',
  'balance',
  'credit_card_yes',
  'country_Germany',
  'gender_Male']
```

```
[ 'products_number', 'balance']
```

```
[ 'country', 'gender', 'credit_card', 'active_member']
```

1

