12_customer_seg_bankchurndata_report

December 26, 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	products_number	credit_card	active_member	estimated_salary
15634602	619	France	Female	42	2	0.00	1	yes	yes	101348.88
15647311	608	Spain	Female	41	1	83807.86	1	no	yes	112542.58
15619304	502	France	Female	42	8	159660.80	3	yes	no	113931.57
15701354	699	France	Female	39	1	0.00	2	no	no	93826.63
15737888	850	Spain	Female	43	2	125510.82	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

            estimated_salary country  gender credit_card active_member  \
customer_id
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

            cluster
customer_id
15634602          8
15647311          0
15619304          0
15701354          0
15737888          1
```

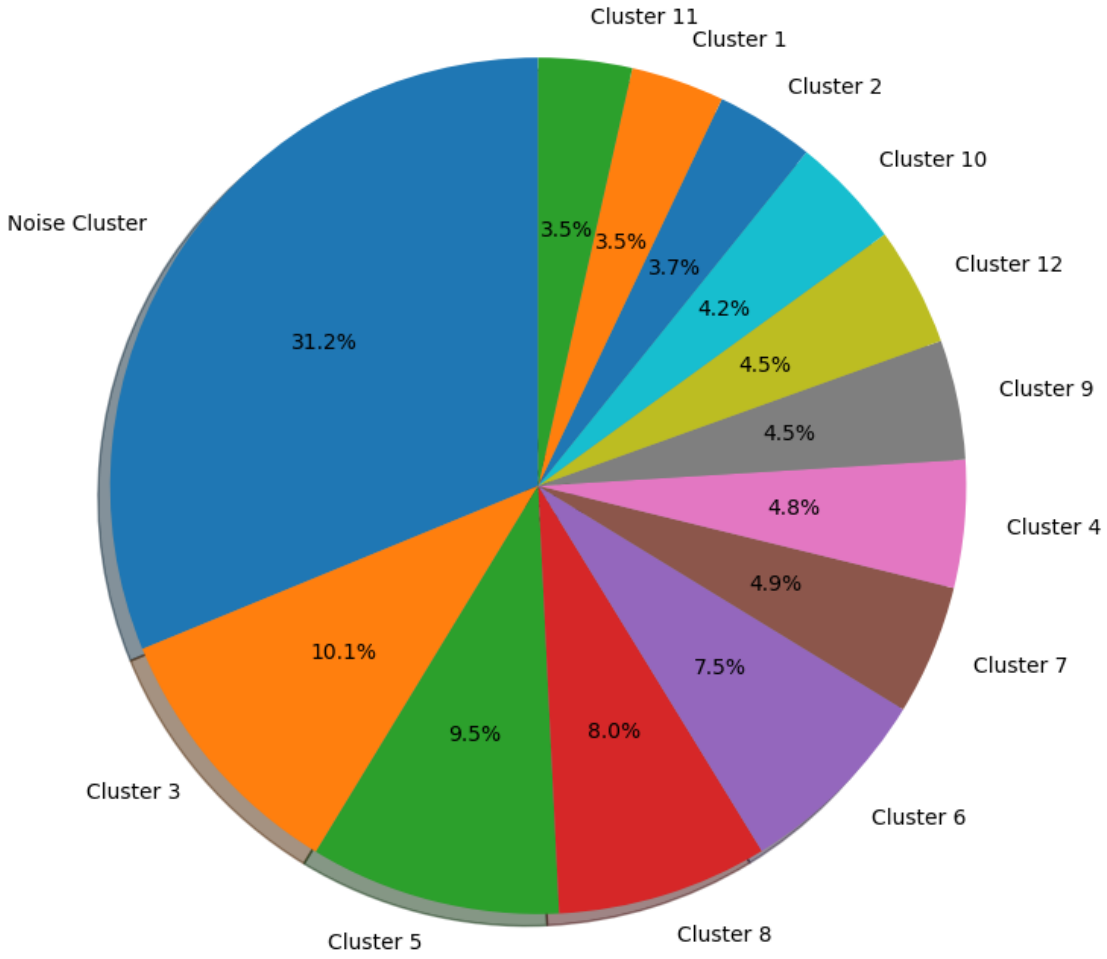
5 Data Visualization

```
0     3122
3     1009
5      947
8      797
6      751
7      490
4      482
9      452
12     449
10     423
2      372
1      353
11     353
Name: cluster, dtype: int64
```

5.1 What does the Cluster Distribution Lool Like?

```
['Noise Cluster',
 'Cluster 3',
 'Cluster 5',
 'Cluster 8',
 'Cluster 6',
 'Cluster 7',
 'Cluster 4',
 'Cluster 9',
 'Cluster 12',
```

```
'Cluster 10',
'Cluster 2',
'Cluster 1',
'Cluster 11']
```



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
```

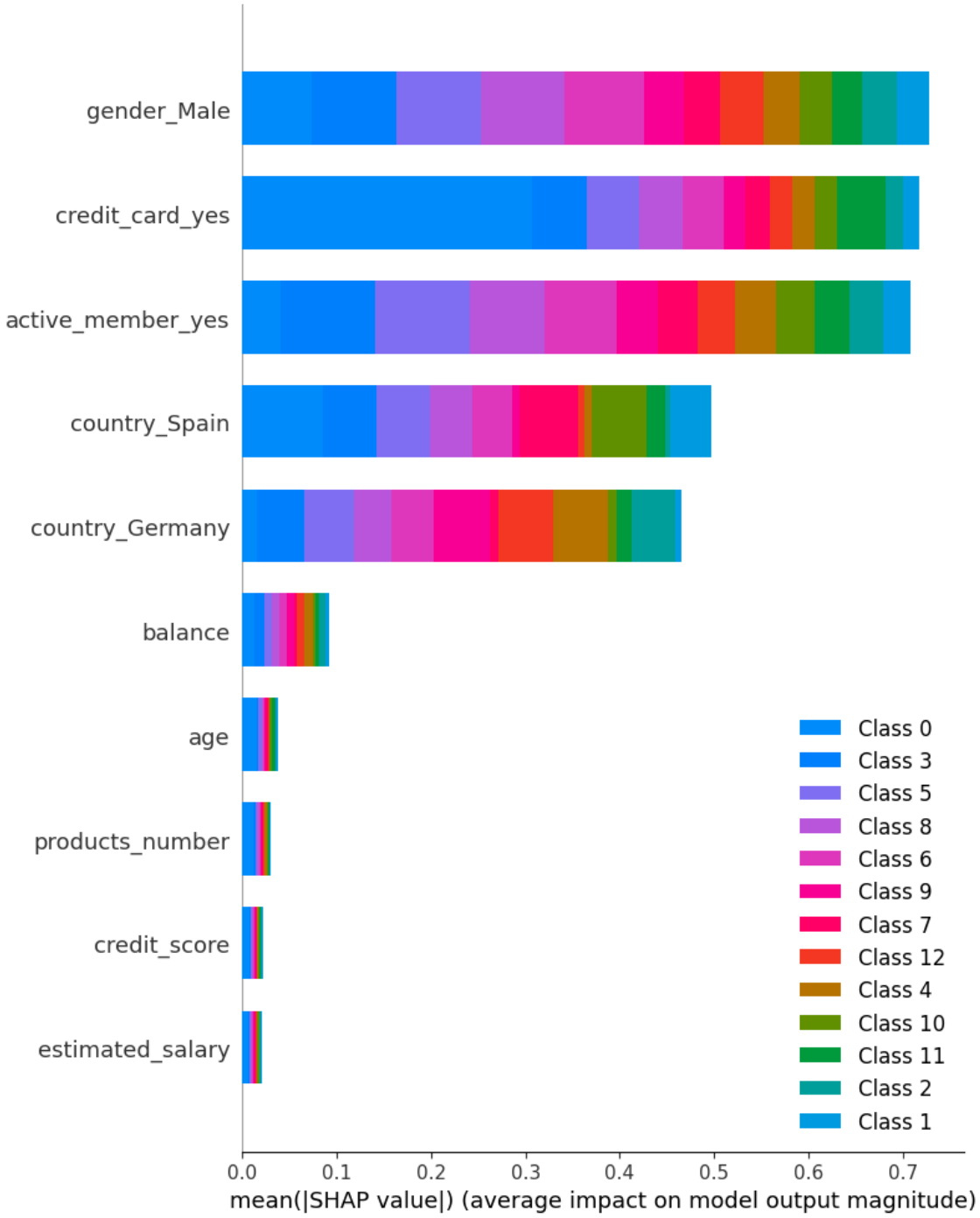
0	651.380205	39.824430	4.918642	78420.609439	1.561050
1	653.838527	38.651869	4.920680	51977.250057	1.553209
2	654.142473	39.474307	4.817204	120496.708226	1.537227
3	652.417245	39.066617	5.093162	59901.590654	1.538007
4	655.946058	39.628625	4.919087	120029.813714	1.546817
5	649.157339	36.931127	5.139388	64589.570380	1.479707
6	645.973369	37.673472	4.874834	58768.362170	1.507700
7	654.632653	38.605438	5.108163	64541.406020	1.479592
8	654.148055	39.362461	4.973651	58013.876487	1.515684
9	641.526549	38.602998	5.278761	121102.218739	1.490190
10	644.808511	37.080378	5.148936	56860.158771	1.549801
11	644.640227	36.422096	5.070822	64619.360057	1.508685
12	647.643653	39.989109	5.309577	119816.335234	1.424423

cluster	estimated_salary	country	gender	credit_card	active_member
0	101445.890753	Spain	Female	no	yes
1	97997.731133	Spain	Female	yes	yes
2	100514.897849	Germany	Female	yes	yes
3	99409.058642	France	Male	yes	yes
4	99893.466079	Germany	Male	yes	yes
5	100604.350391	France	Male	yes	no
6	100651.166897	France	Female	yes	no
7	95255.682653	Spain	Male	yes	yes
8	96226.088670	France	Female	yes	yes
9	100434.335155	Germany	Male	yes	no
10	99041.084539	Spain	Male	yes	no
11	100823.305637	France	Male	no	no
12	103877.611359	Germany	Female	yes	no

7 Explain Clusters

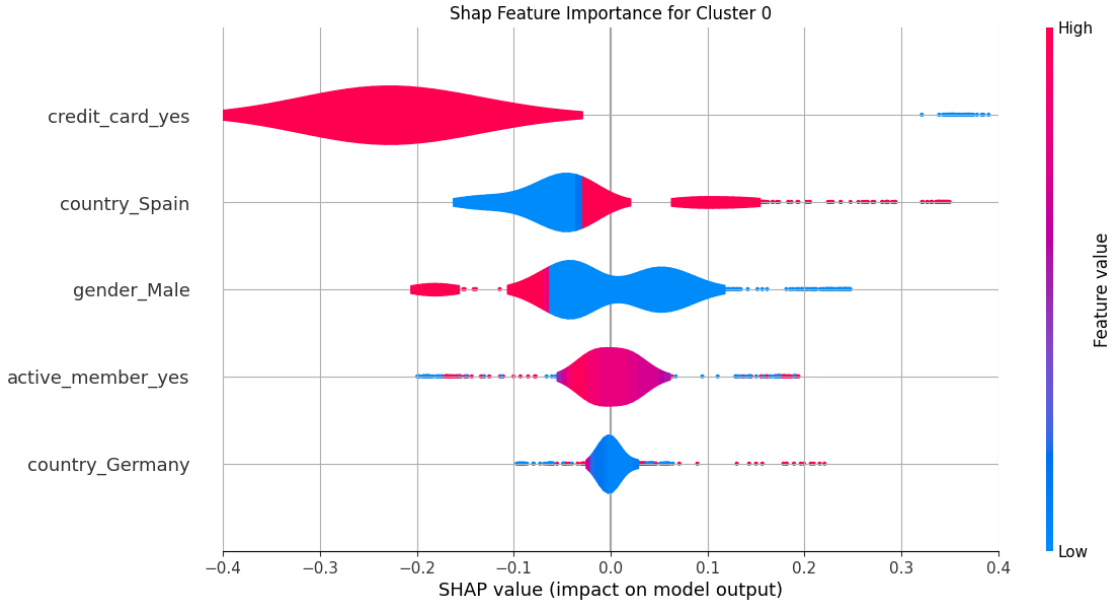
8 Custom Functions

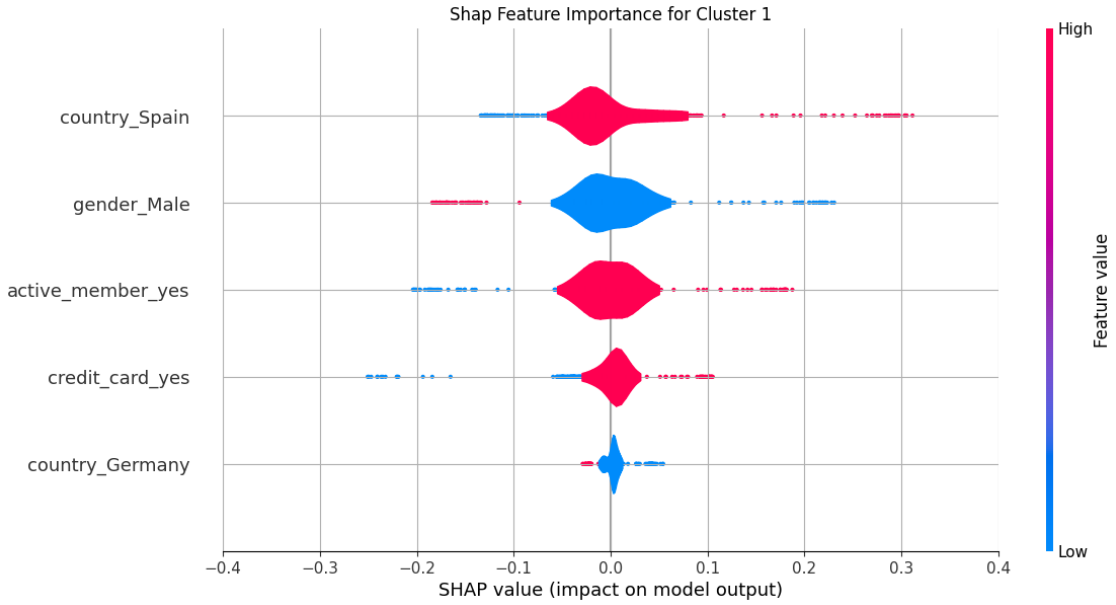
8.1 Average Feature Importance on Cluster Building

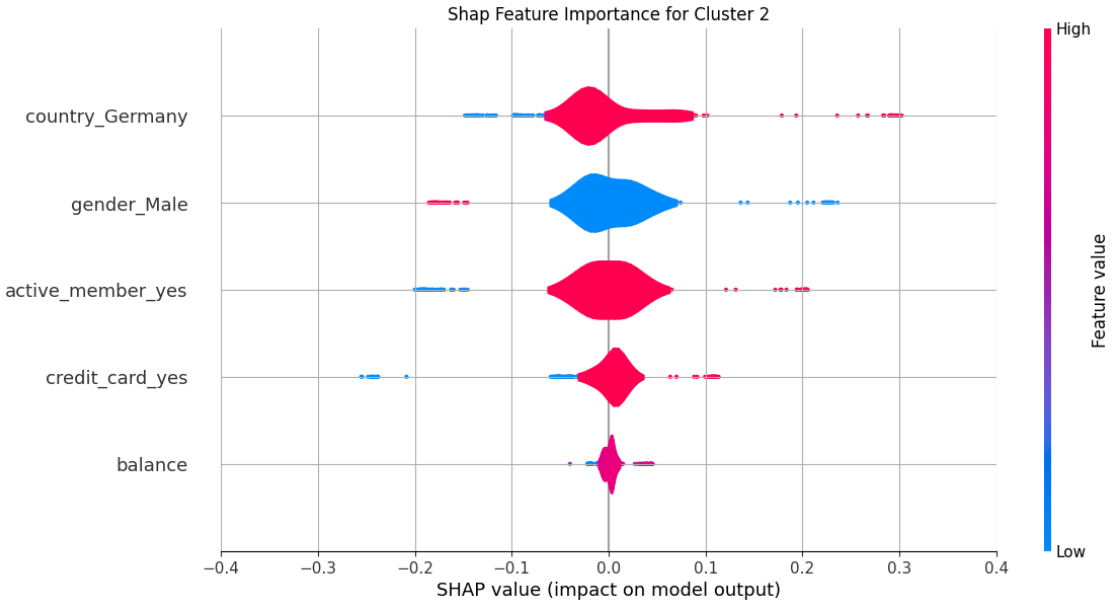


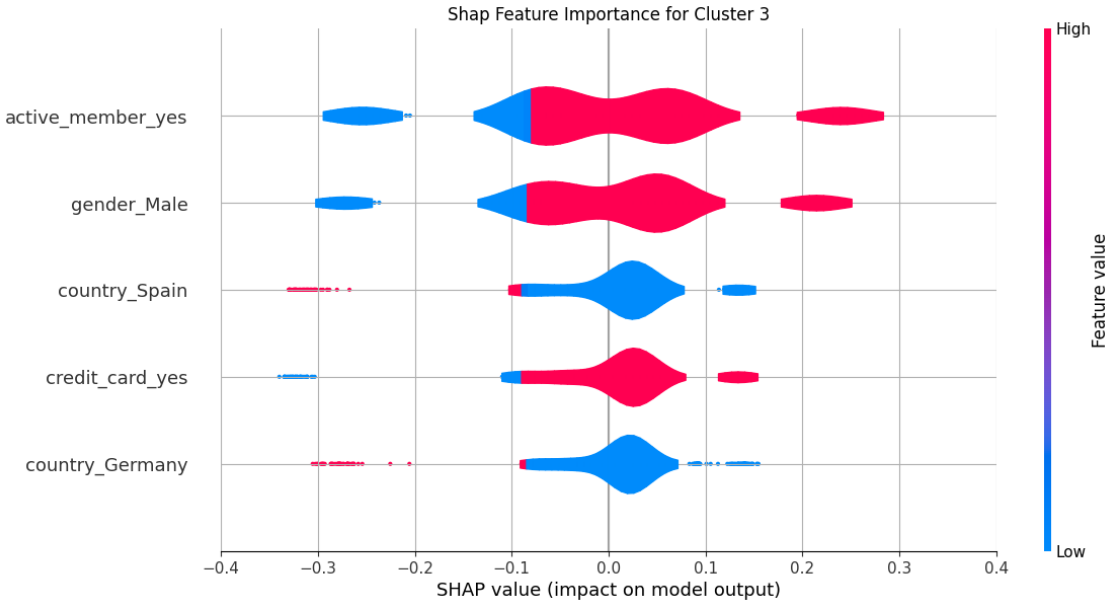
8.2 Explain Single Cluster vs Rest

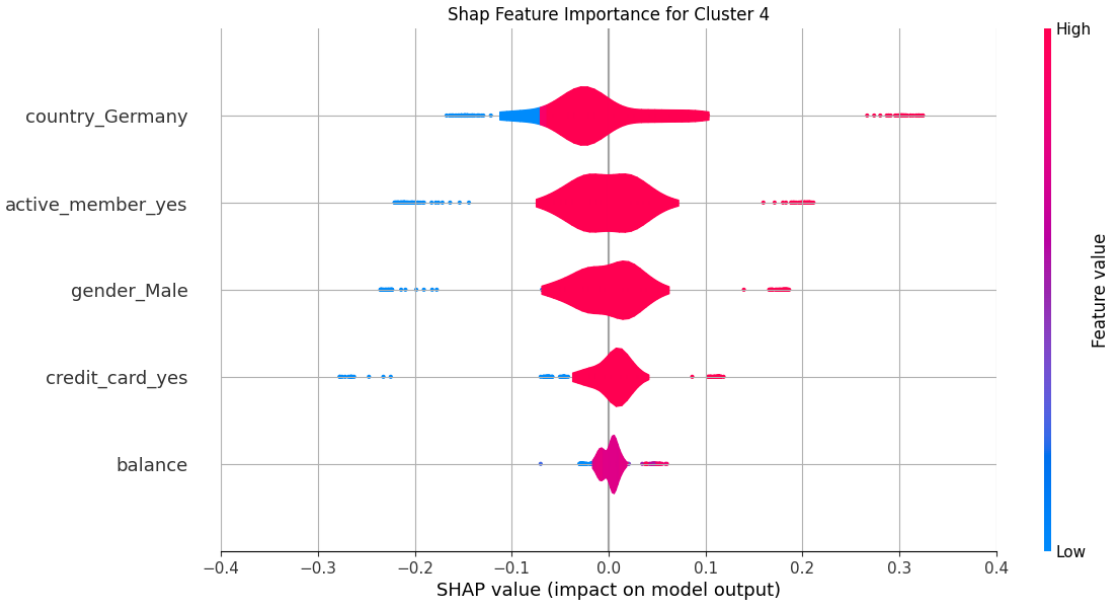
```
No data for colormapping provided via 'c'. Parameters 'vmin', 'vmax' will be
ignored
```

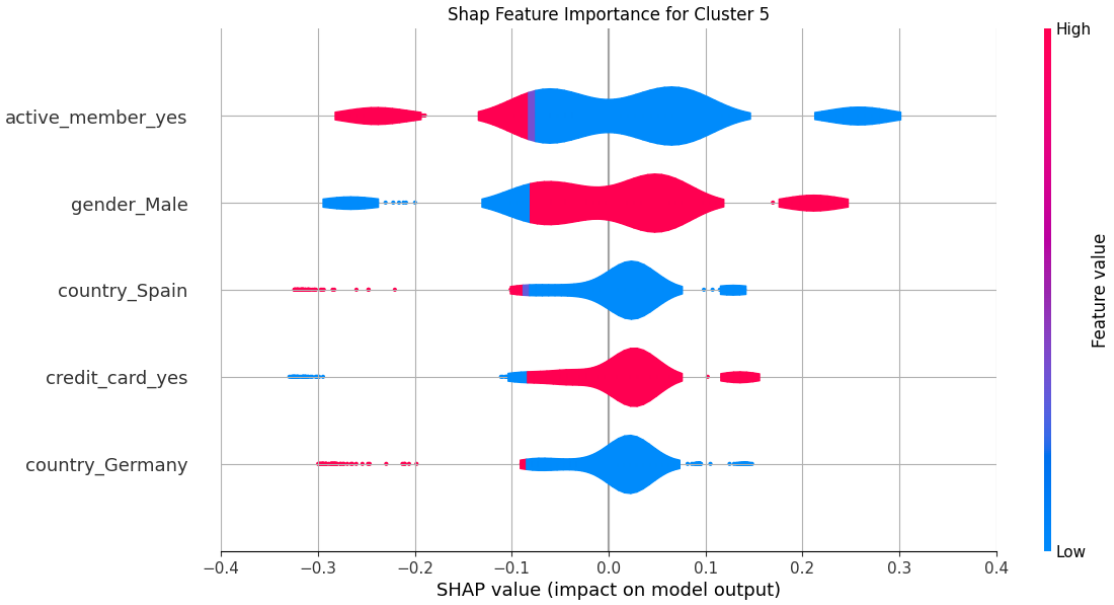


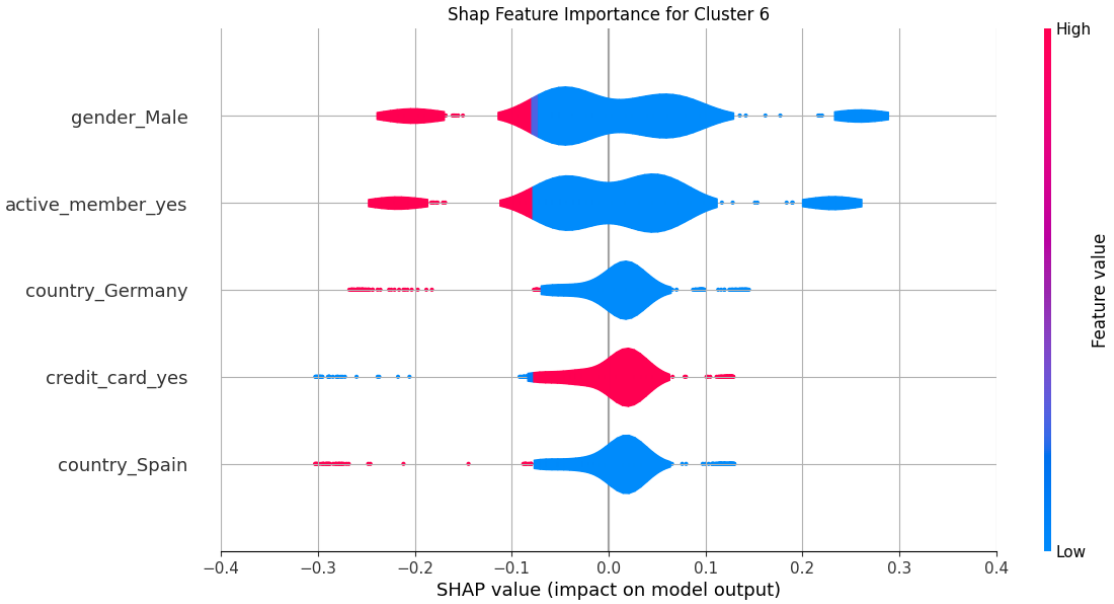


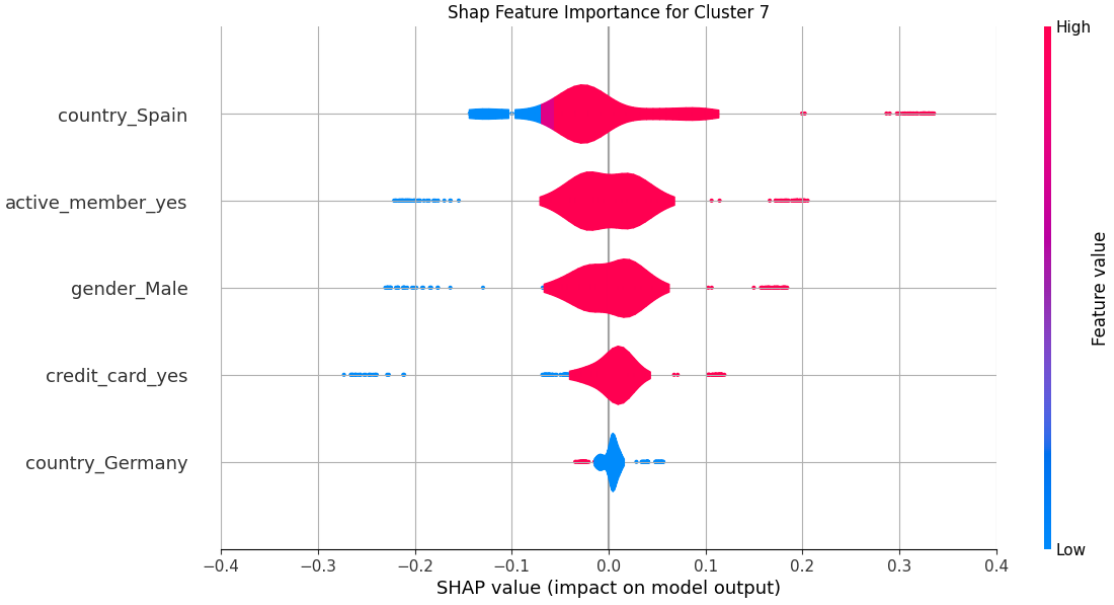

Shap Feature Importance for Cluster 2
country_Germany
gender_Male
active_member_yes
credit_card_yes
balance
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

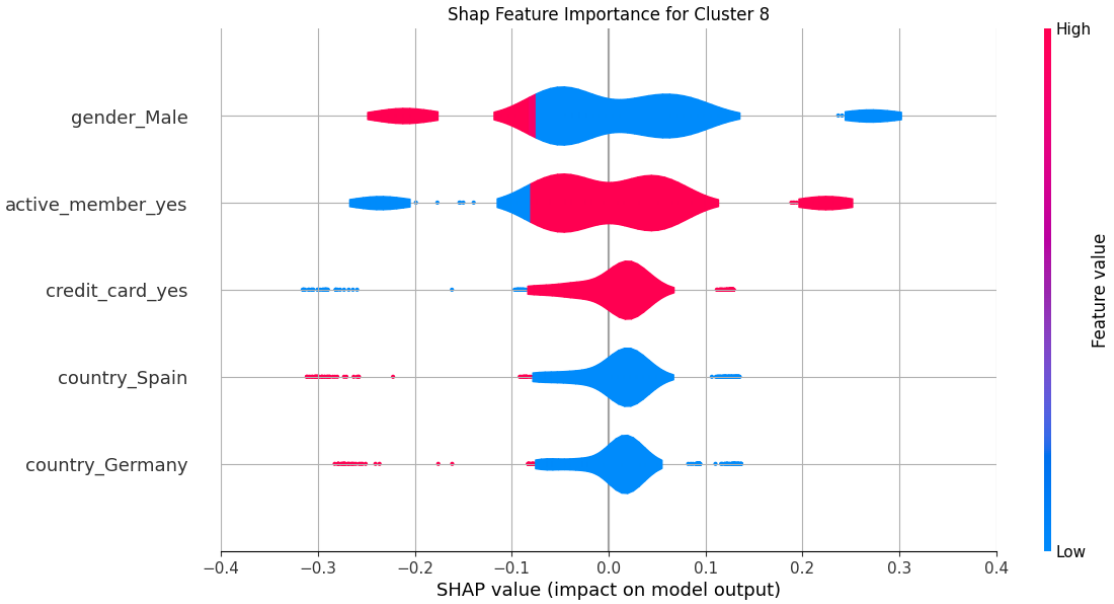

Shap Feature Importance for Cluster 3
active_member_yes
gender_Male
country_Spain
credit_card_yes
country_Germany
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

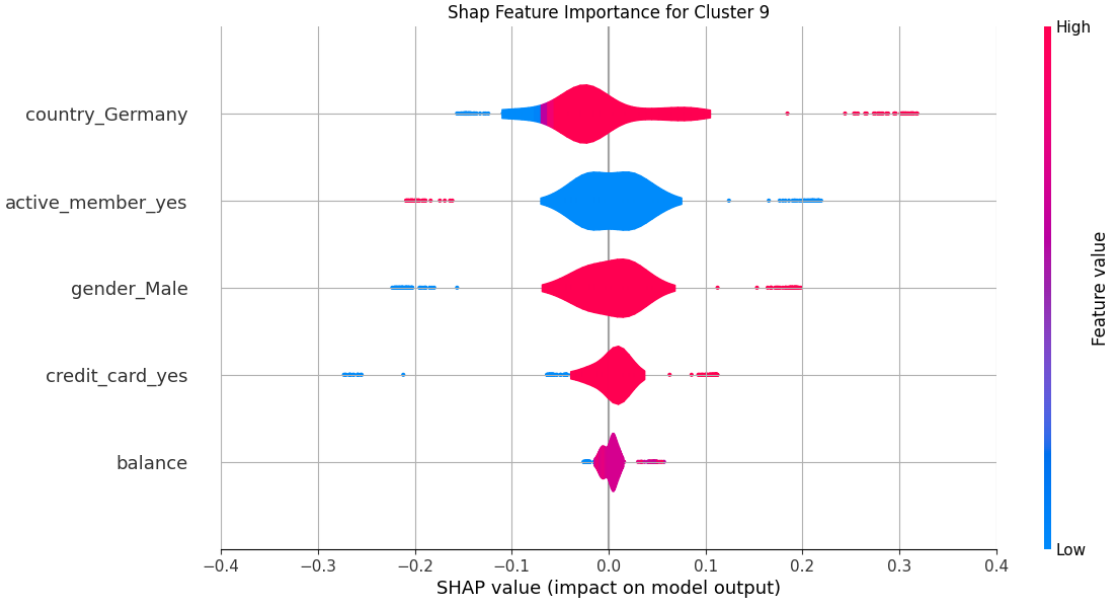

Shap Feature Importance for Cluster 4
country_Germany
active_member_yes
gender_Male
credit_card_yes
balance
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

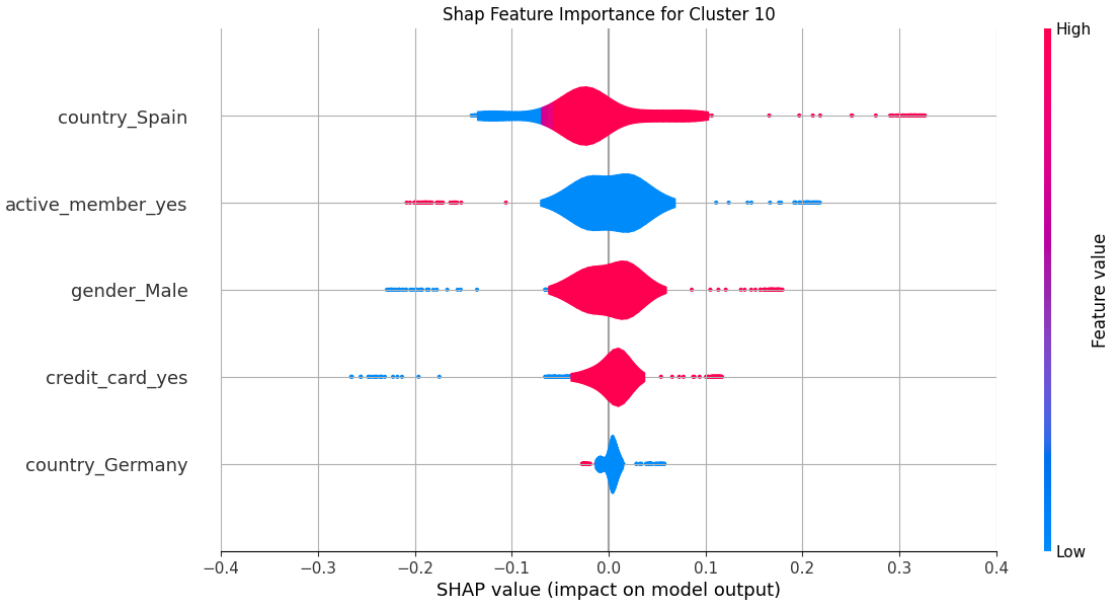

Shap Feature Importance for Cluster 5
active_member_yes
gender_Male
country_Spain
credit_card_yes
country_Germany
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

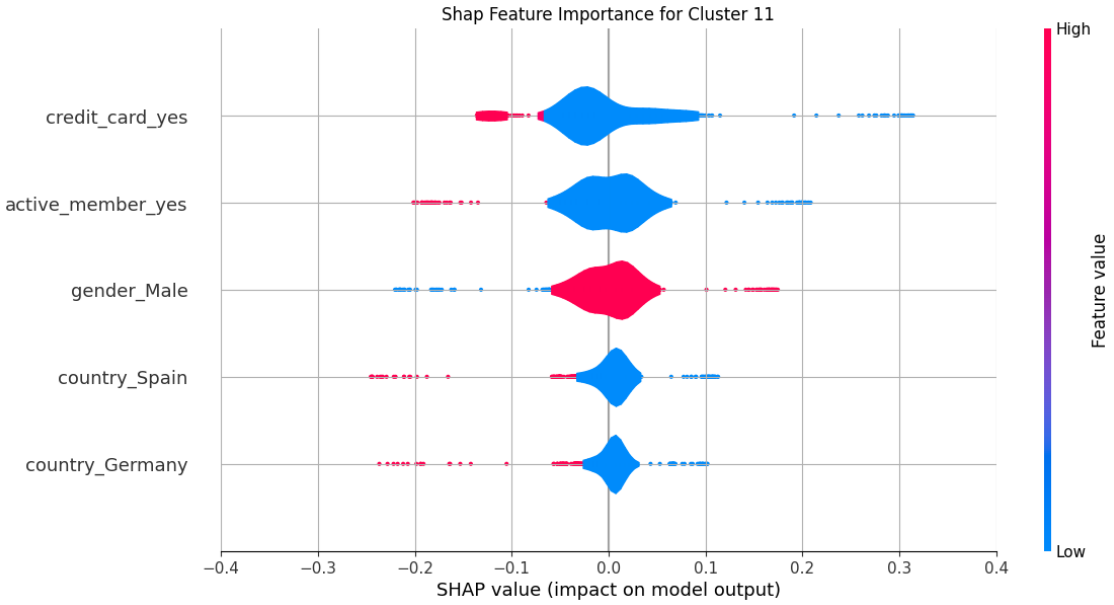

Shap Feature Importance for Cluster 6
gender_Male
active_member_yes
country_Germany
credit_card_yes
country_Spain
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

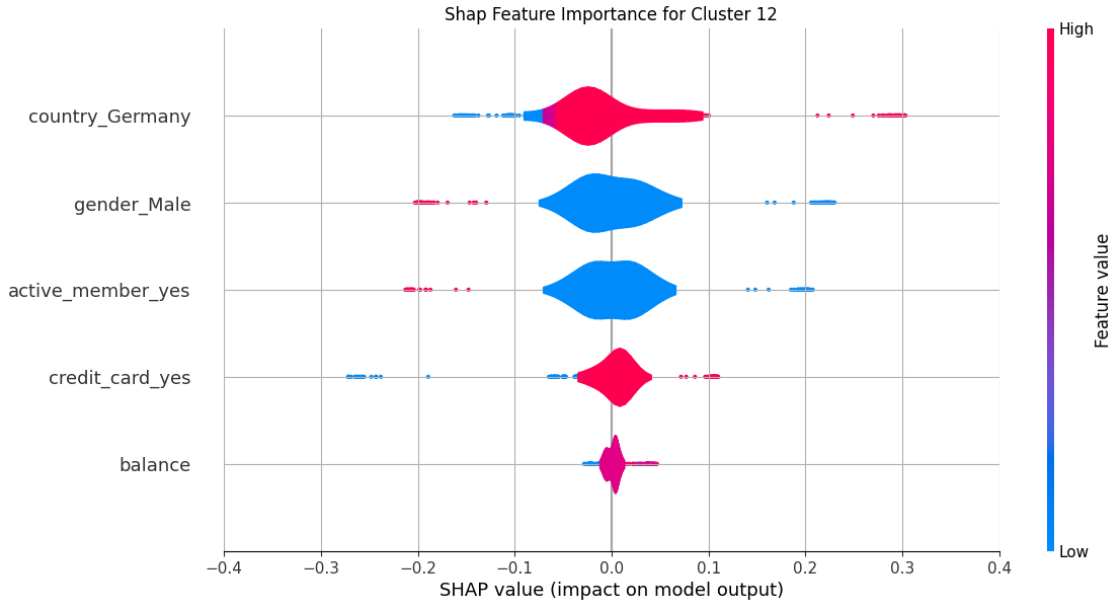

Shap Feature Importance for Cluster 7
country_Spain
active_member_yes
gender_Male
credit_card_yes
country_Germany
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low


Shap Feature Importance for Cluster 8
gender_Male
active_member_yes
credit_card_yes
country_Spain
country_Germany
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

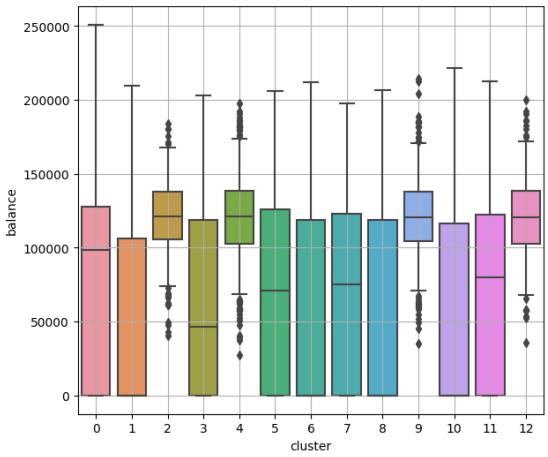

Shap Feature Importance for Cluster 9
country_Germany
active_member_yes
gender_Male
credit_card_yes
balance
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

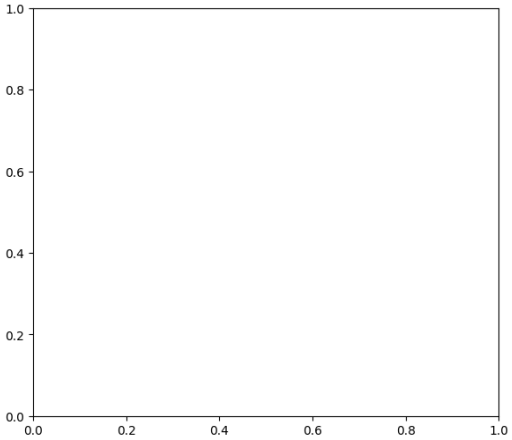

Shap Feature Importance for Cluster 10
country_Spain
active_member_yes
gender_Male
credit_card_yes
country_Germany
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low


Shap Feature Importance for Cluster 11
credit_card_yes
active_member_yes
gender_Male
country_Spain
country_Germany
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
SHAP value (impact on model output)
High
Feature value
Low

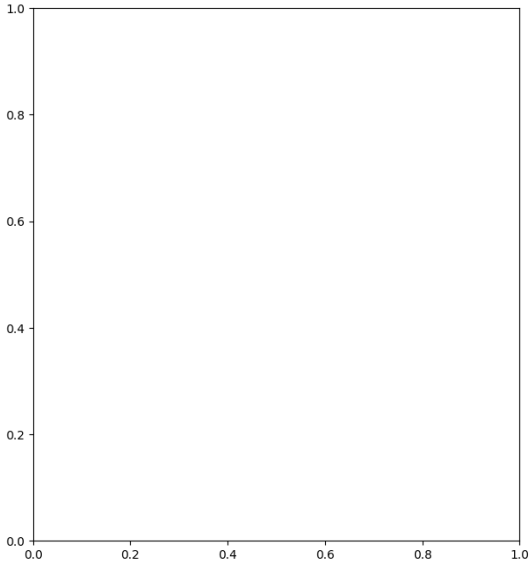


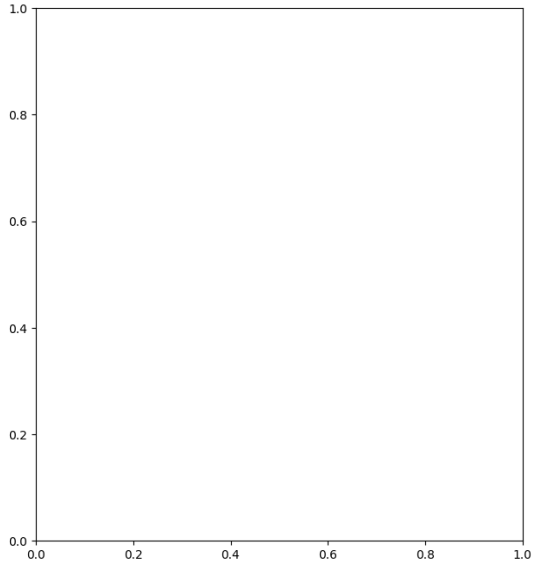

```
['country_Germany',
 'gender_Male',
 'active_member_yes',
 'credit_card_yes',
 'balance']

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

['balance']

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

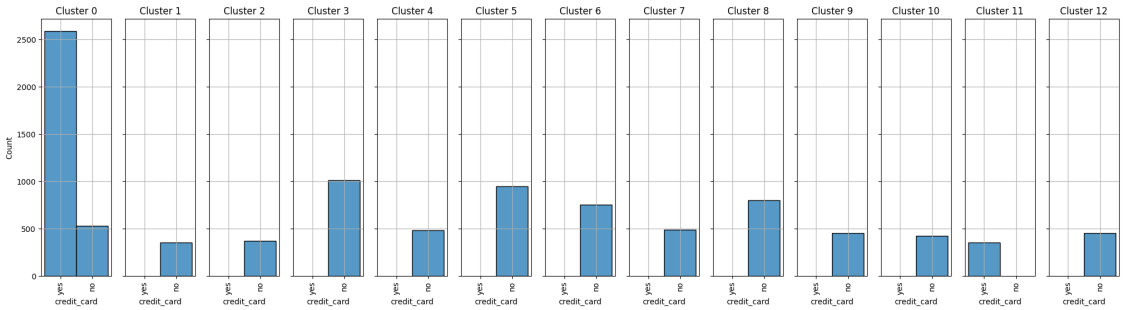

balance
cluster

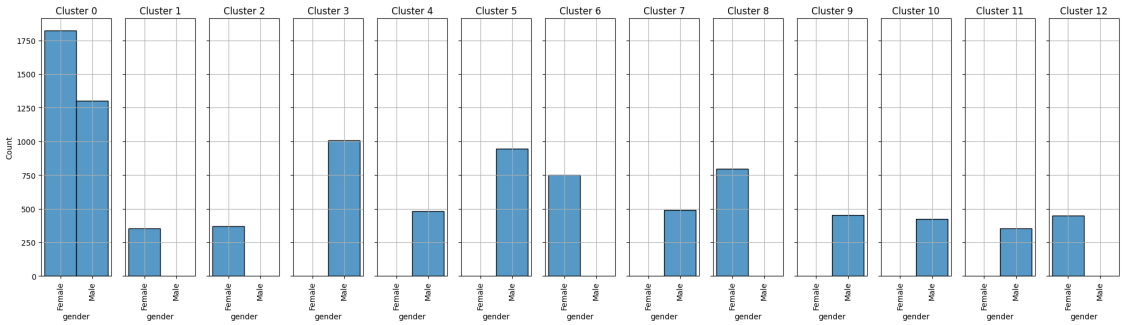


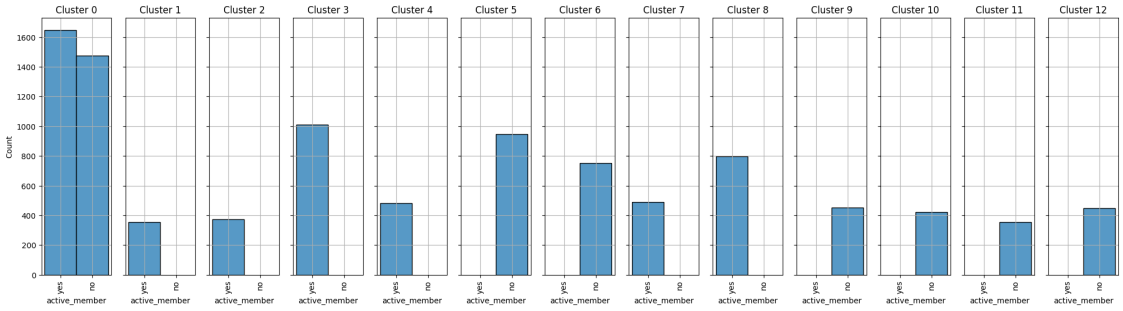
0

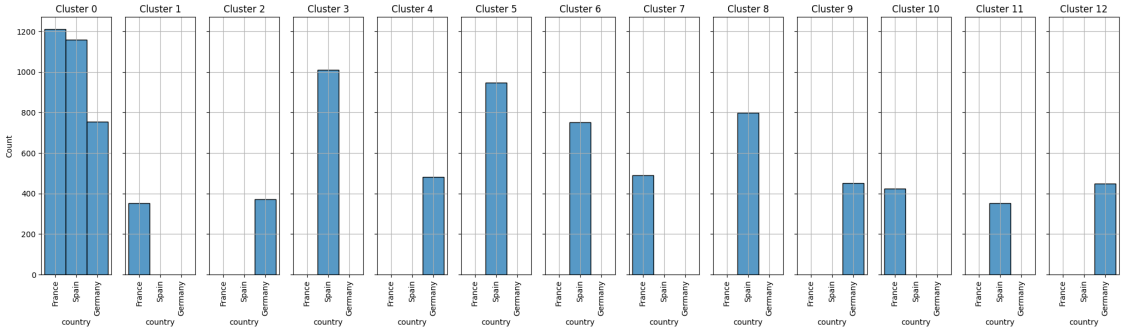





Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Count
credit_card
yes
no


Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Count
0
250
500
750
1000
1250
1500
1750
Female
Male
gender


Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Count
0
200
400
600
800
1000
1200
1400
1600
yes
no
active_member


Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Count
0
200
400
600
800
1000
1200
France
Spain
Germany
country