06_customer_seg_adultdata_report

December 23, 2022

1 Customer Segmentation Report

2 Import Data

	Age	Workclass	fnlwgt	Education	Education-Num	\
0	39	State-gov	77516	Bachelors	13	
1	50	Self-emp-not-inc	83311	Bachelors	13	
2	38	Private	215646	HS-grad	9	
3	53	Private	234721	11th	7	
4	28	Private	338409	Bachelors	13	

	Martial Status	Occupation	Relationship	Race	Sex	\
0	Never-married	Adm-clerical	Not-in-family	White	Male	
1	Married-civ-spouse	Exec-managerial	Husband	White	Male	
2	Divorced	Handlers-cleaners	Not-in-family	White	Male	
3	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	
4	Married-civ-spouse	Prof-specialty	Wife	Black	Female	

	Capital Gain	Capital Loss	Hours per week	Country	Income
0	2174	0	40	United-States	<=50K
1	0	0	13	United-States	<=50K
2	0	0	40	United-States	<=50K
3	0	0	40	United-States	<=50K
4	0	0	40	Cuba	<=50K

3 Load Model Artifacts

4 Transform Dataset

```
Found unknown categories in columns [0] during transform. These unknown
categories will be encoded as all zeros
```

	Age	fnlwgt	Education-Num	Capital Gain	Capital Loss	Hours per week	\
0	39.0	77516.0	13.0	2174.0	0.0	40.0	
1	50.0	83311.0	13.0	0.0	0.0	13.0	
2	38.0	215646.0	9.0	0.0	0.0	40.0	
3	53.0	234721.0	7.0	0.0	0.0	40.0	

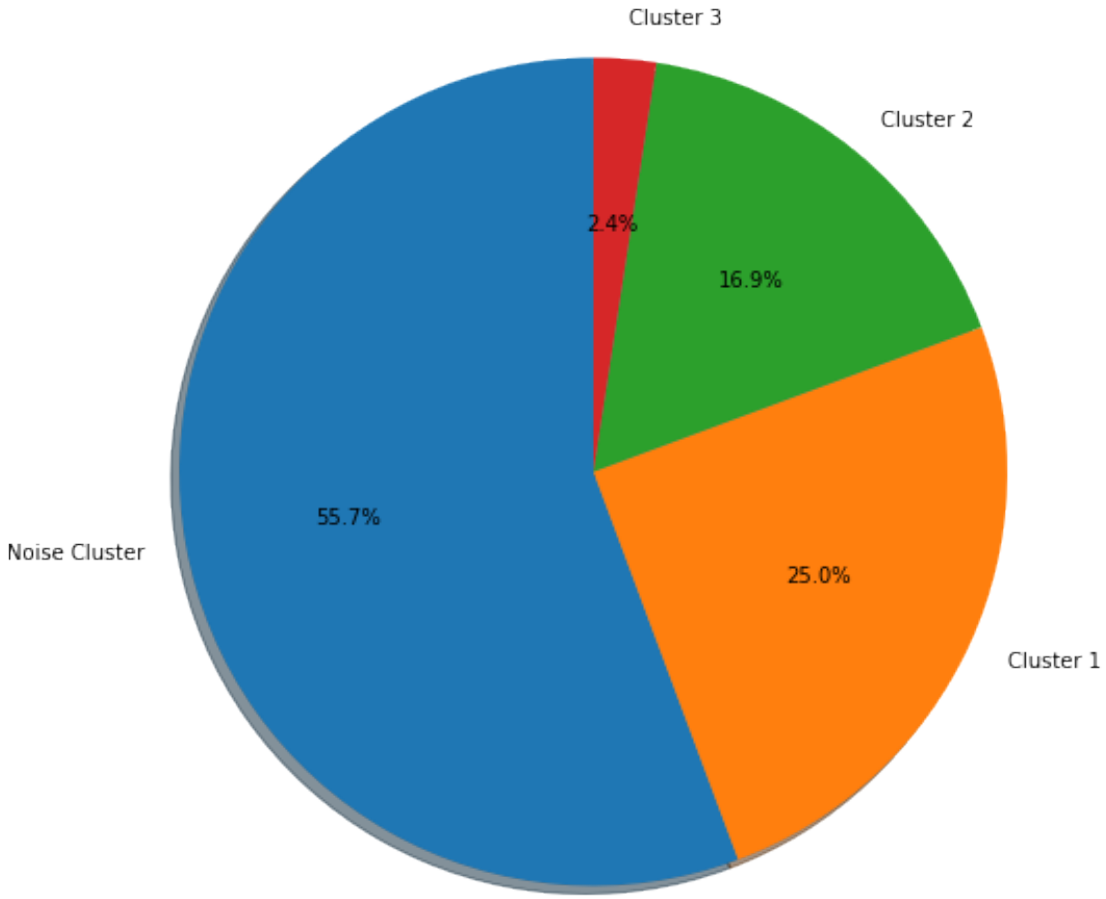
```
4  28.0  338409.0           13.0           0.0           0.0            40.0

          Workclass  Education      Martial Status       Occupation  \
0         State-gov  Bachelors       Never-married     Adm-clerical
1  Self-emp-not-inc  Bachelors  Married-civ-spouse  Exec-managerial
2           Private    HS-grad            Divorced             Rare
3           Private       Rare  Married-civ-spouse             Rare
4           Private  Bachelors  Married-civ-spouse   Prof-specialty

    Relationship   Race     Sex        Country Income  cluster
0  Not-in-family  White    Male  United-States  <=50K        0
1        Husband  White    Male  United-States  <=50K        0
2  Not-in-family  White    Male  United-States  <=50K        2
3        Husband  Black    Male  United-States  <=50K        0
4           Wife  Black  Female           Rare  <=50K        0
```

5 Data Visualization

```
0    27198
1    12208
2     8246
3     1190
Name: cluster, dtype: int64
```

5.1 What does the Cluster Distribution Lool Like?

```
['Noise Cluster', 'Cluster 1', 'Cluster 2', 'Cluster 3']
```



6 Determine Cluster Centers

cluster	Age	fnlwgt	Education-Num	Capital Gain	Capital Loss
0	40.911580	192300.261221	10.287227	806.467609	77.508181
1	42.688991	180531.280159	9.951016	851.457887	42.345697
2	26.558951	186843.276619	9.152559	78.423230	27.467518
3	28.379832	188394.921701	13.015126	110.082353	48.792694

cluster	Hours per week	Workclass	Education	Martial Status
0	40.389680	Private	Rare	Married-civ-spouse
1	43.643237	Private	HS-grad	Married-civ-spouse
2	35.059433	Private	HS-grad	Never-married
3	40.818258	Private	Bachelors	Never-married

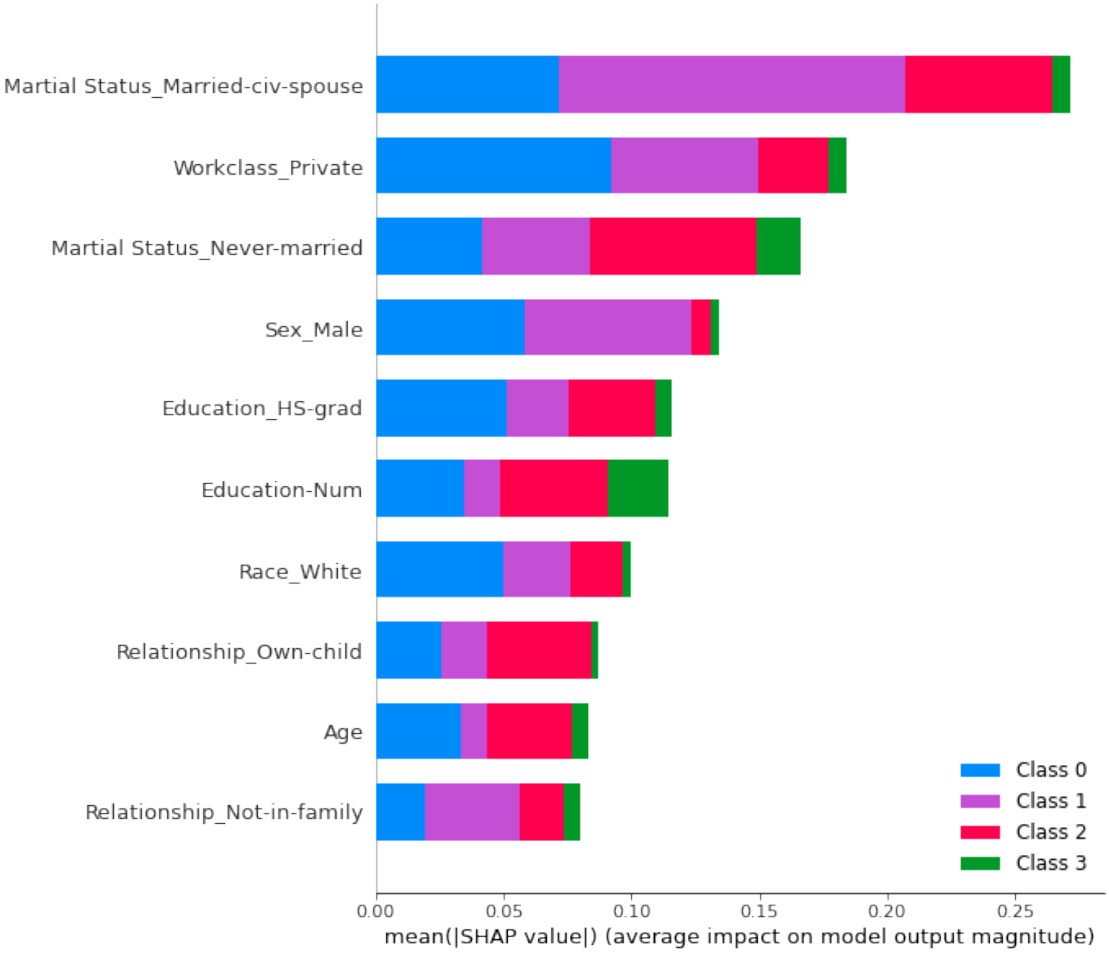
Occupation	Relationship	Race	Sex	Country	Income

cluster						
0	Exec-managerial	Not-in-family	White	Male	United-States	<=50K
1	Craft-repair	Husband	White	Male	United-States	<=50K
2	Rare	Own-child	White	Male	United-States	<=50K
3	Exec-managerial	Not-in-family	White	Female	United-States	<=50K

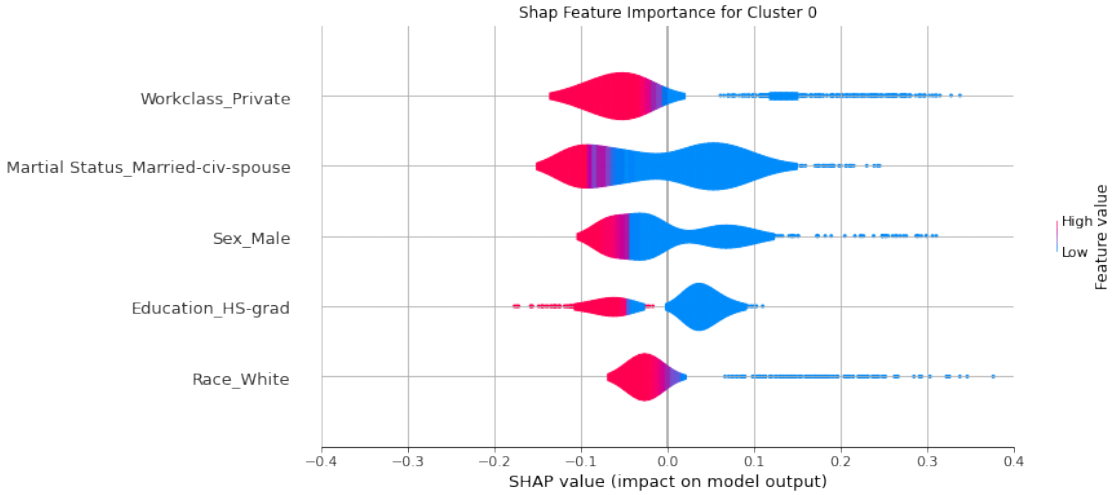
7 Explain Clusters

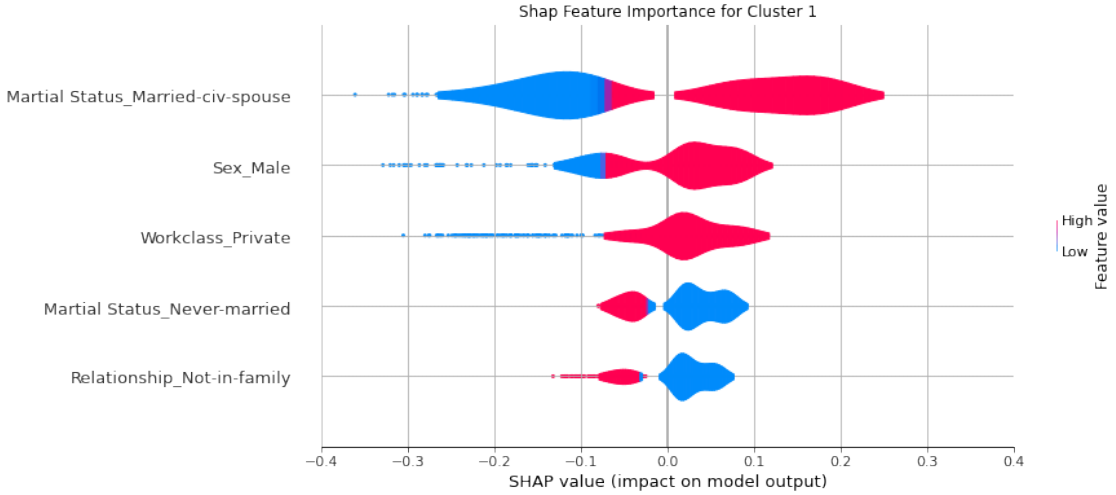
8 Custom Functions

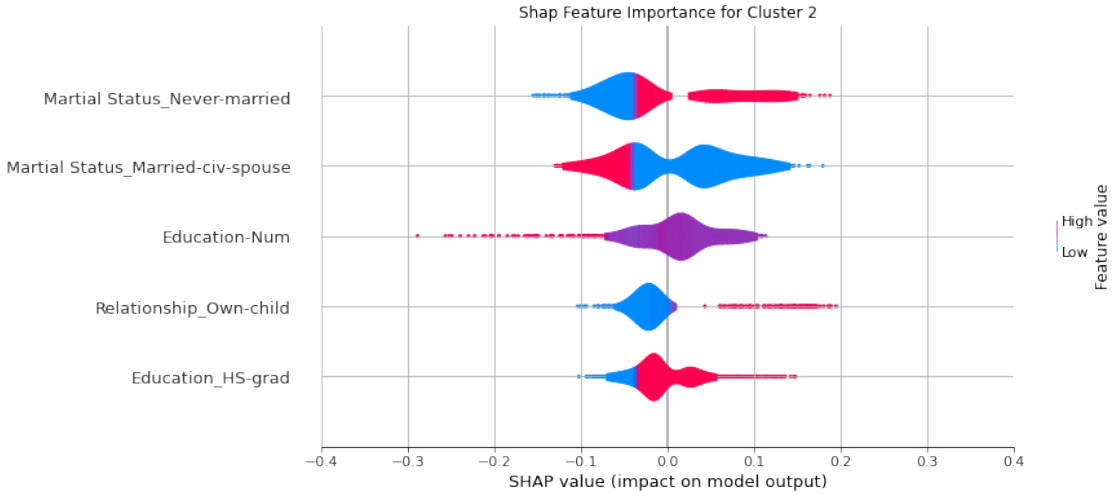
8.1 Average Feature Importance on Cluster Building

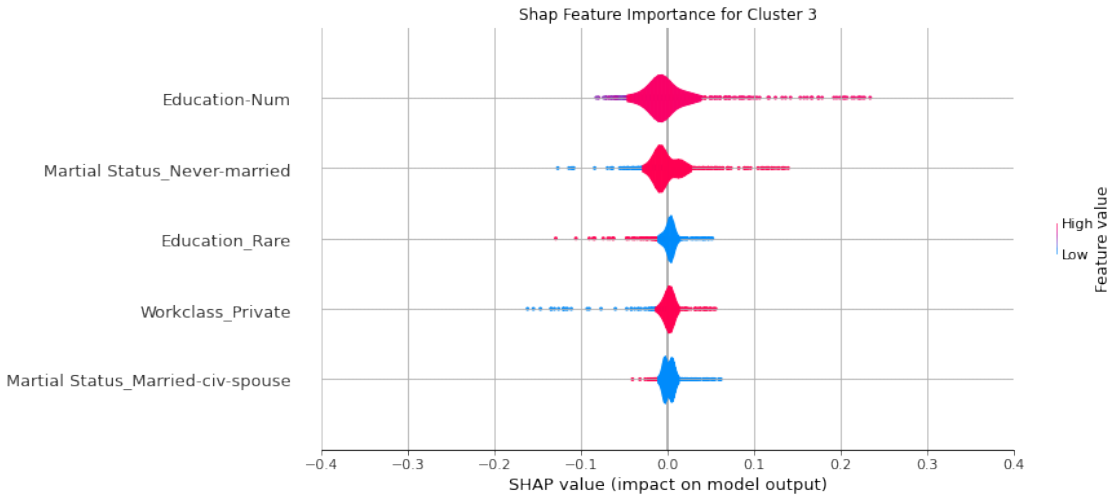


8.2 Explain Single Cluster vs Rest







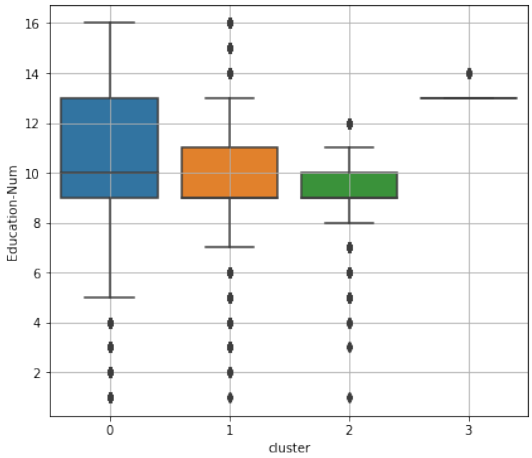


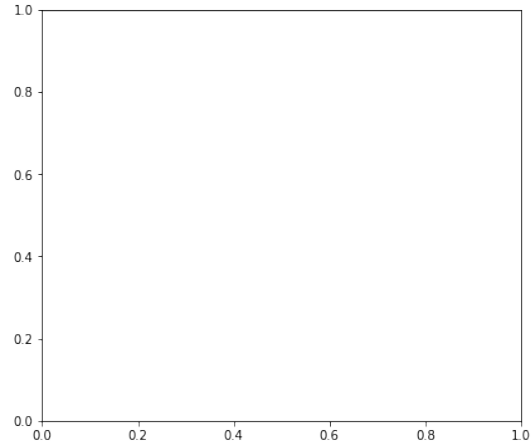

```
['Martial Status_Never-married',
 'Martial Status_Married-civ-spouse',
 'Education-Num',
 'Relationship_Own-child',
 'Education_HS-grad']

['Relationship_Not-in-family',
 'Education_HS-grad',
 'Martial Status_Never-married',
 'Education_Rare',
 'Workclass_Private',
 'Relationship_Own-child',
 'Sex_Male',
```

```
 'Martial Status_Married-civ-spouse',
 'Education-Num',
 'Race_White']
```

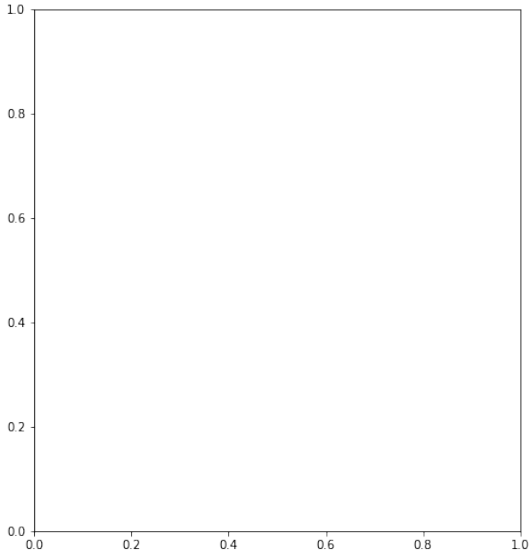
```
['Education-Num']
```

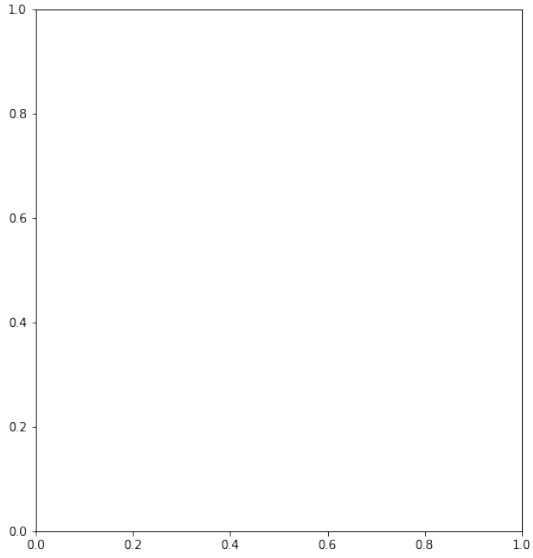
```
['Relationship', 'Martial Status', 'Education', 'Workclass', 'Race', 'Sex']
```

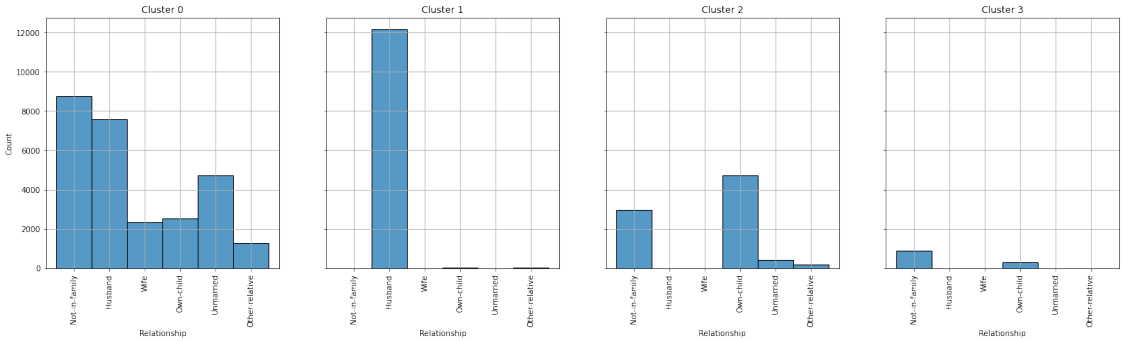


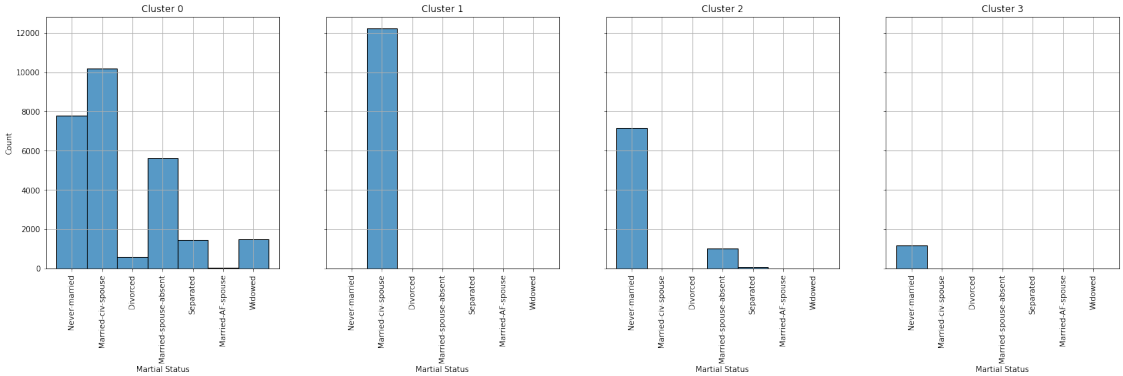


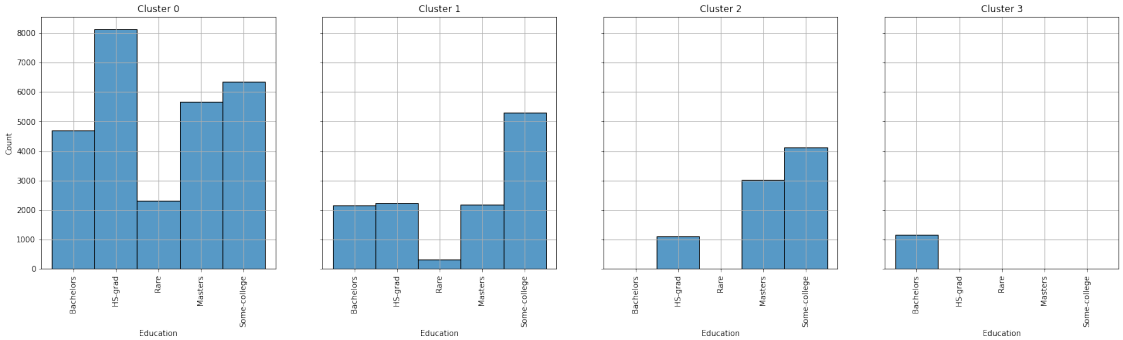
0

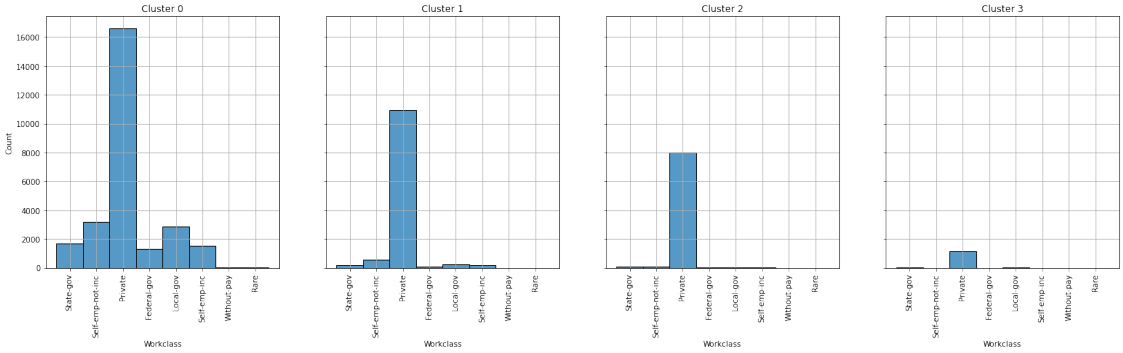


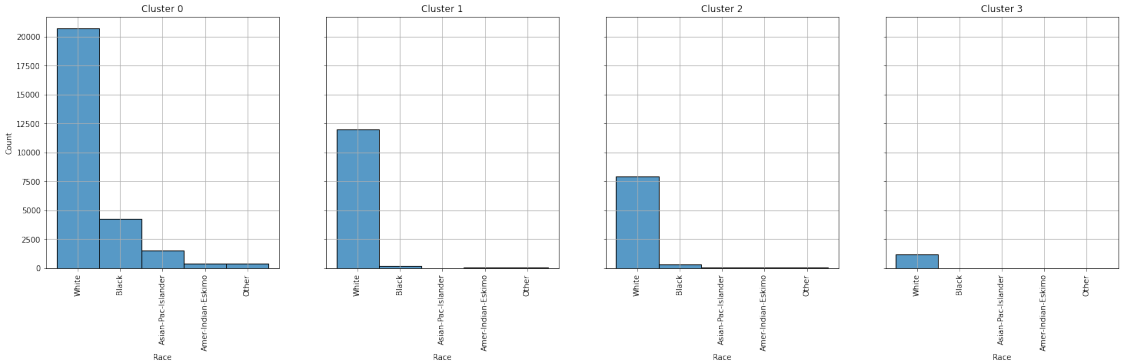


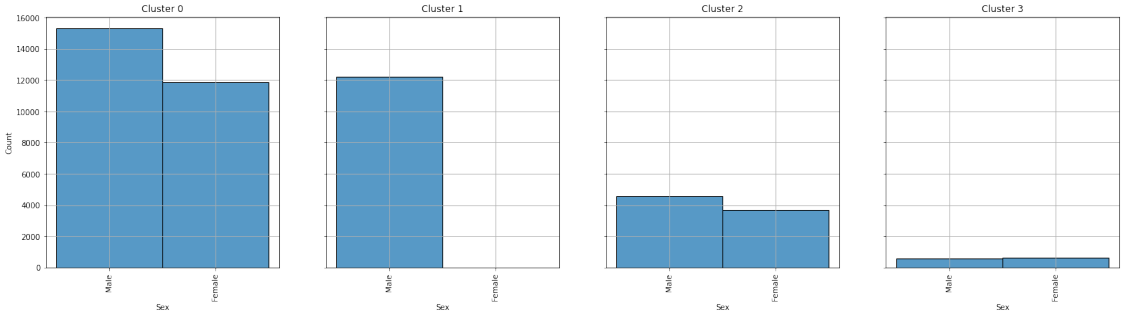

Cluster 0
Cluster 1
Cluster 2
Cluster 3
Count
Relationship
Not-in-family
Husband
Wife
Own-child
Unmarried
Other-relative


Cluster 0
Cluster 1
Cluster 2
Cluster 3
Count
Martial Status
Never-married
Married-civ-spouse
Divorced
Married-spouse-absent
Separated
Married-AF-spouse
Widowed


Cluster 0
Cluster 1
Cluster 2
Cluster 3
Count
Education
Bachelors
HS-grad
Rare
Masters
Some-college








```
C:\Users\Nutzer\.conda\envs\customer_analytics\lib\site-
packages\traitlets\traitlets.py:2389: FutureWarning: Supporting extra quotes
around strings is deprecated in traitlets 5.0. You can use '../reports' instead
of "'../reports'" if you require traitlets >=5.
  warn(
[NbConvertApp] Converting notebook 06_customer_seg_adultdata_report.ipynb to pdf
[NbConvertApp] Support files will be in 06_customer_seg_adultdata_report_files\
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
```

```
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Making directory .\06_customer_seg_adultdata_report_files
[NbConvertApp] Writing 36851 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', 'notebook.tex', '-quiet']
[NbConvertApp] Running bibtex 1 time: ['bibtex', 'notebook']
[NbConvertApp] WARNING | b had problems, most likely because there were no
citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 525291 bytes to
..\reports\06_customer_seg_adultdata_report.pdf
```