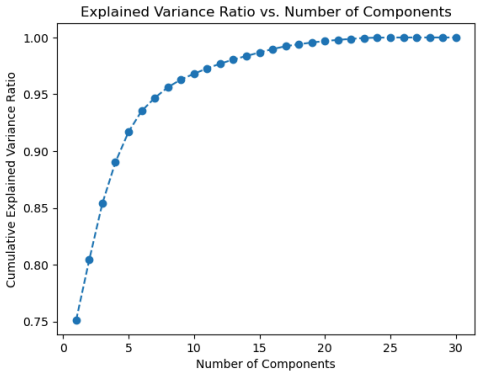
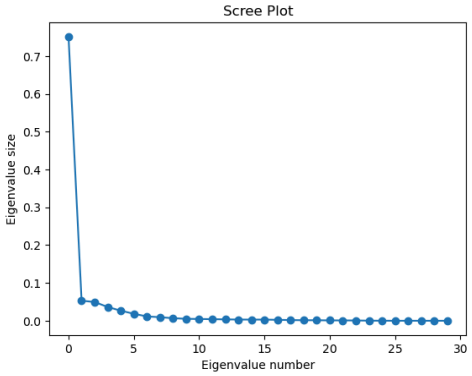
TRIAL 1

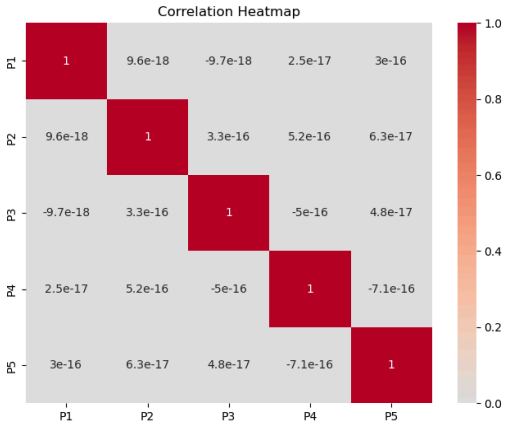
- **File Name:** clustering_trial_1.ipynb
- **Dataset:** minmax_scaled_all_except_long_lat_df.parquet
-

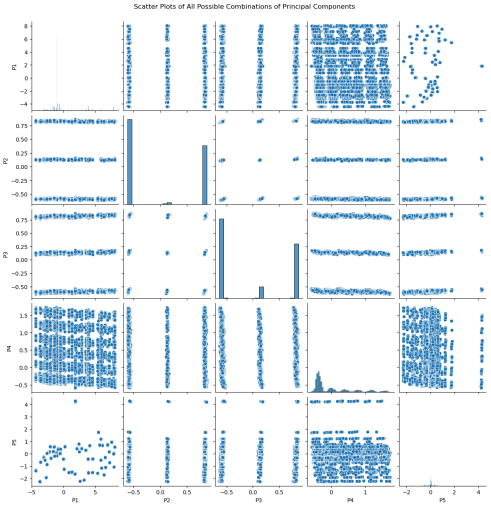
CUST_NUM	EDUCATION	AGE	TENURE	BUSINESS_OWNER	DIGITAL_FLAG	SUBSEGMENT	Frequency_cred	AMOUNT_cred	Recency_cred	LATITUDE	LONGITUDE	INCOME_SOURCE_ALLOWANCE	INCOME_SOURCE_BUSINESS	INCOME_SOURCE_COMMISSION	INCOME_SO
13401.256807	0.5	0.151163	0.031125	0.0	0.0	0.000000	0.024169	0.000418	0.000000	15.527737	120.419269	False	False	False	
4230.004965	0.0	0.139535	0.255402	0.0	0.0	0.666667	0.123867	0.011176	0.000000	14.608637	121.031947	True	False	False	
4481.937304	0.0	0.151163	0.218675	0.0	0.0	0.666667	0.061934	0.001719	0.077778	14.608637	121.031947	False	False	False	
4734.959768	0.0	0.151163	0.101645	0.0	0.0	0.000000	0.015106	0.004222	0.077778	14.608637	121.031947	False	False	False	
4828.128416	0.5	0.151163	0.146732	0.0	0.0	0.333333	0.000000	0.000176	0.466667	14.608637	121.031947	False	False	False	

- **Characteristics:**
  - CUST_INFO and CREDIT_TRANSACTIONS dataset
  - Scaled all except Flags, Longitude, Latitude
  - MinMaxScaler()
  - Tenure (has 0.64 corr with Age) and Frequency (has 0.43 corr with Amount) included
- **PCA:**
  - Cumulative Explained Variance Ratio
  - 

  - Scree Plot (Eigenvalue size vs Eigenvalue number)
  - 

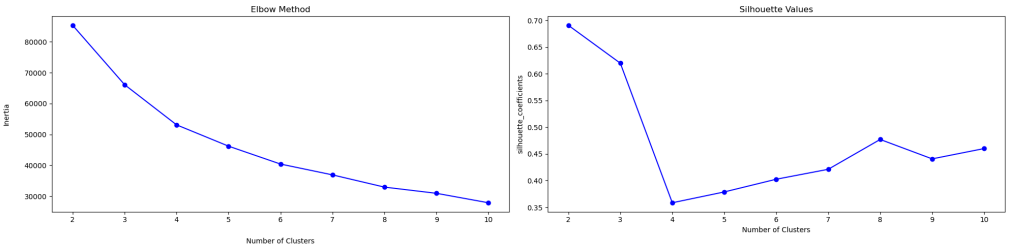
  - Chose **5** as the number of components (ok?)

- Correlation Matrix
  - 

- Pairplot
  - 

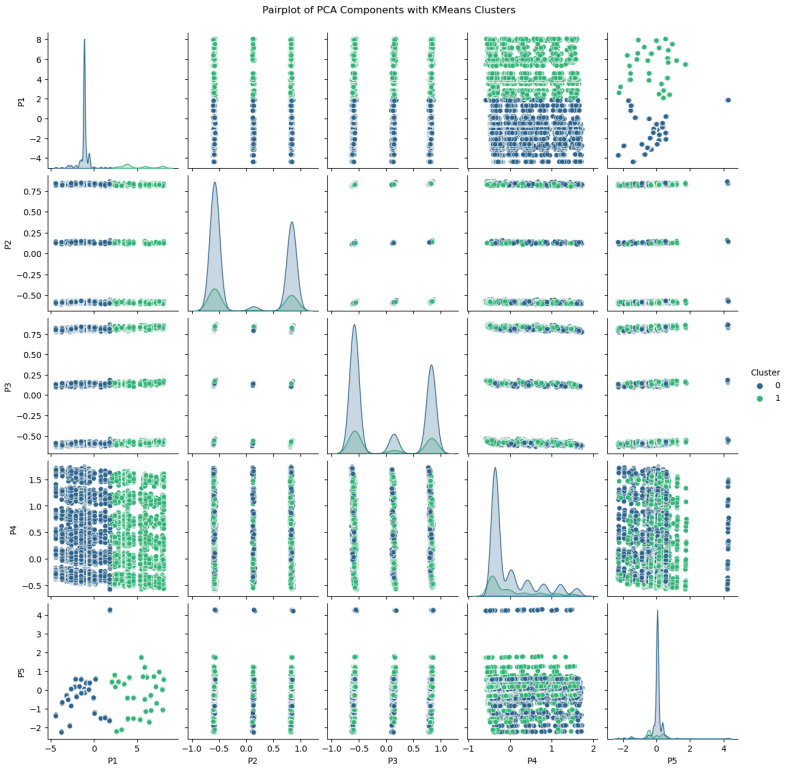
- **Hyperparameter Tuning:**
  - For this trial, we performed hyperparameter tuning
  - We set the range of numbers of clusters from 2 - 10 (ok?)
  - We use the **silhouette coefficient** as the metric
  - random_state = 42
  - Paramer Grid:
    -
```
param_grid = {
    'n_clusters': range(2, 11),
    'init': ['k-means++'],
    'n_init': [10, 20, 30],
    'algorithm': ['lloyd', 'elkan']
}
```
(ok?)
  - Best Parameters:
    - {'algorithm': 'lloyd', 'init': 'k-means++', 'n_clusters': 2, 'n_init': 10}
    - Best number of clusters: 2
    - Best silhouette score: 0.5575086885540613
- **Elbow Method and Silhouette Coefficients:**

- Used the best parameters in hyperparameter tuning and used elbow method to check

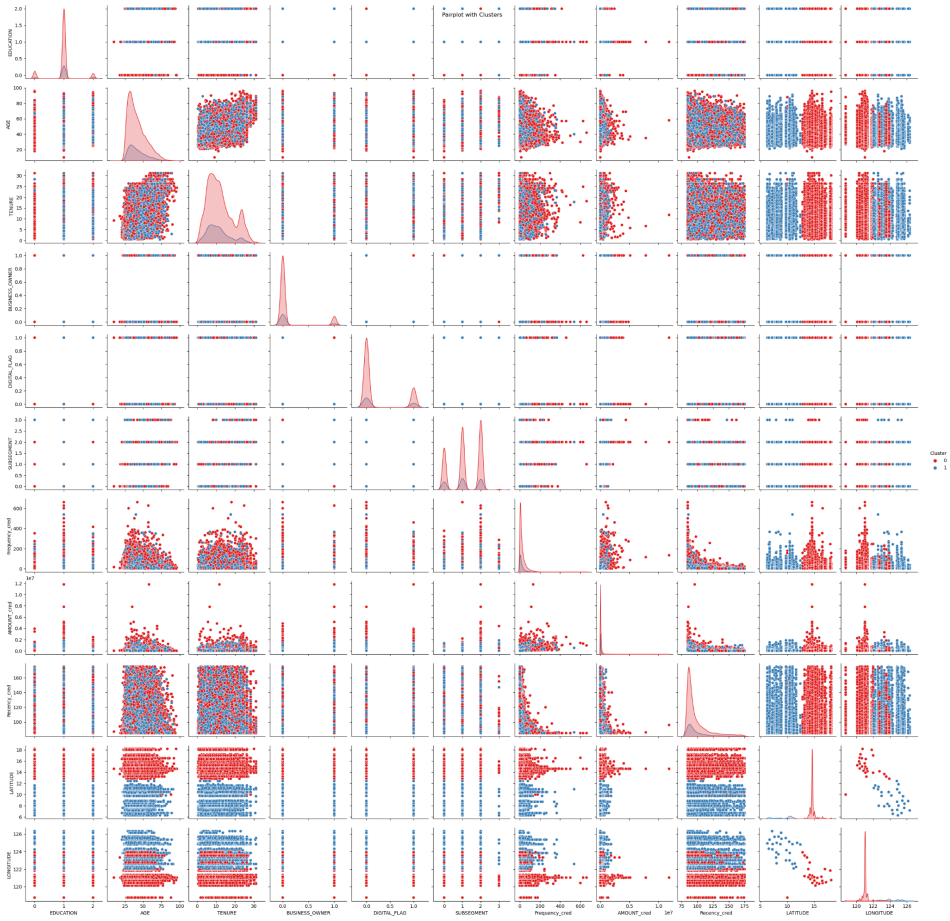


- Silhouette Value for n_clusters = 2 is high but it does not appear to be the elbow. **We look at n_clusters = 2 in the Cluster EDA.**

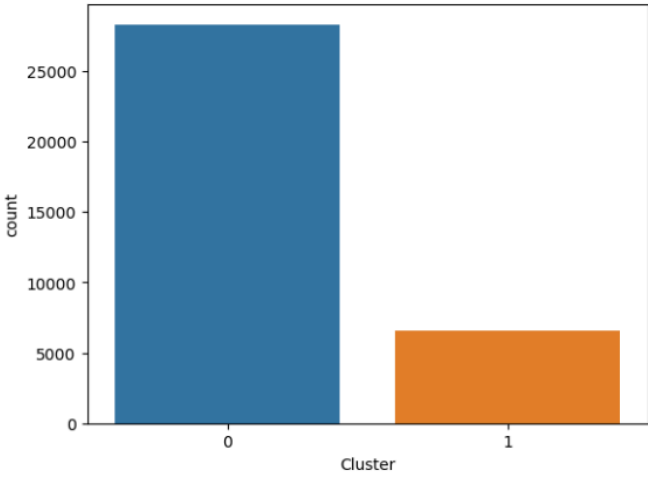
- **Cluster EDA:**
  - **PCA Pairplot:**

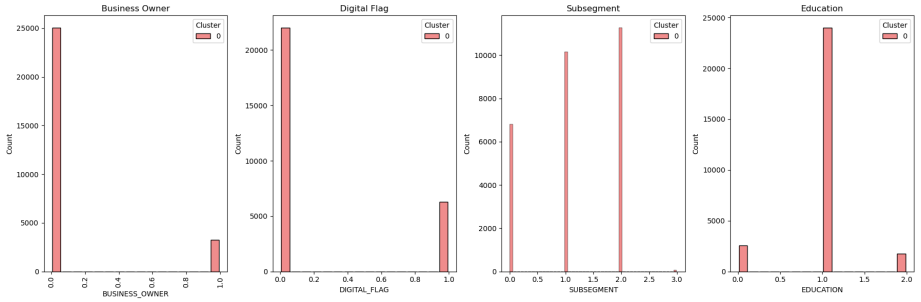


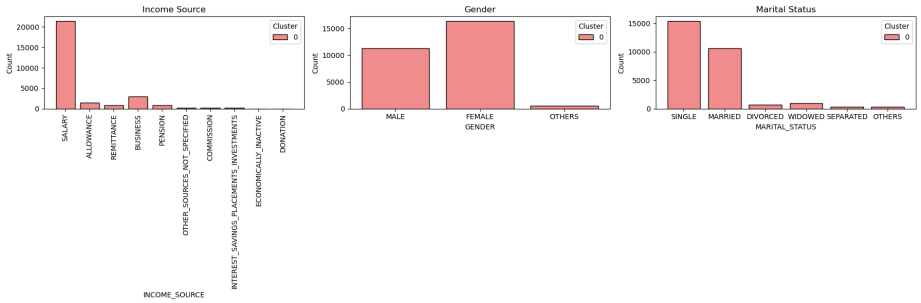
- **Pairplot (Original Merged Dataset):**

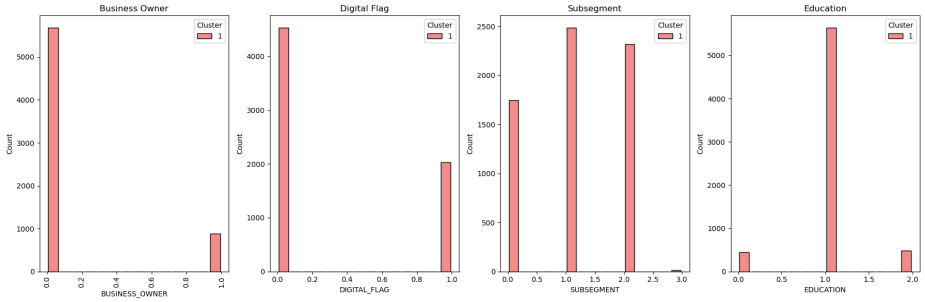


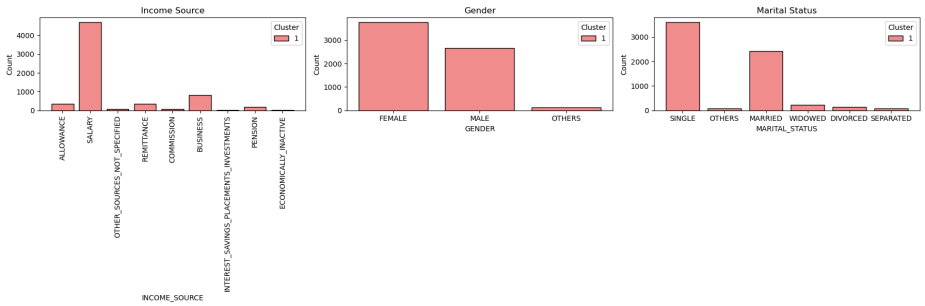
- Clustering became based on Longitude and Latitude (maybe due to them not being scaled)
- Cluster 1 = Luzon, Cluster 2 = Visayas and Mindanao
- Same distribution of Ages, Tenure, Recency for both clusters
- Cluster 1 tends to spend more
- Cluster 1 tends to be more frequent



- **Cluster 0 Histograms:**
  - 

  - 

  - Mostly No Business, Mostly Digital, Mostly lower to upper middle class, mostly mid educated, mostly salaried, mostly male and female (slightly more female), mostly single and married (slightly more single)
- **Cluster 1 Histograms:**
  - 

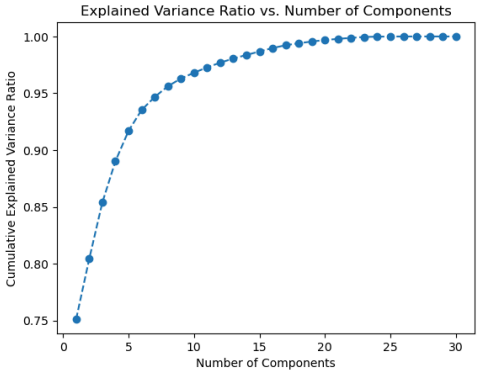
  - 

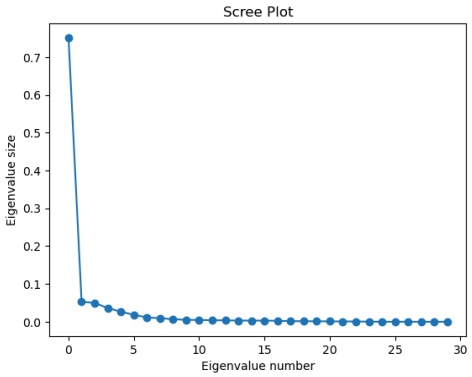
  - Almost same as Cluster 0

TRIAL 2

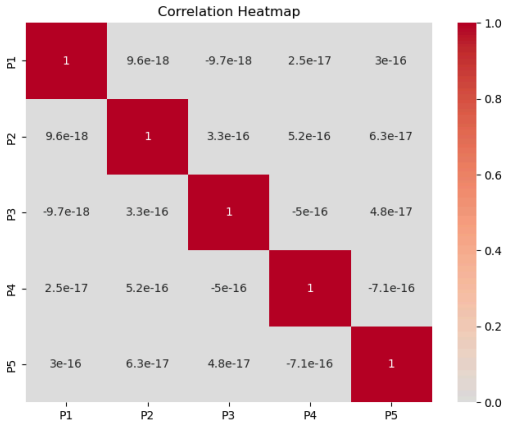
- **File Name:** clustering_trial_2.ipynb
- **Dataset:** minmax_scaled_all_except_long_lat_df.parquet
-

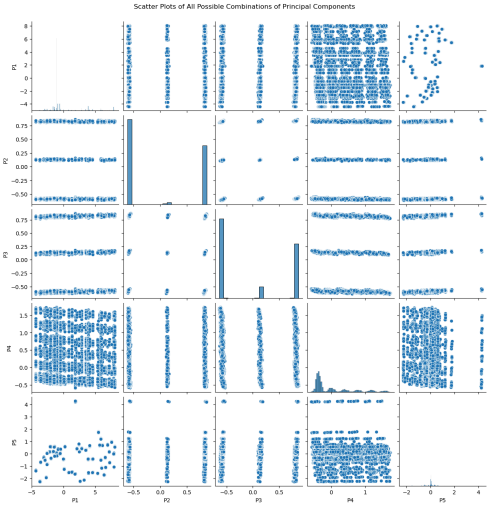
CUST_NUM	EDUCATION	AGE	TENURE	BUSINESS_OWNER	DIGITAL_FLAG	SUBSEGMENT	Frequency_cred	AMOUNT_cred	Recency_cred	LATITUDE	LONGITUDE	INCOME_SOURCE_ALLOWANCE	INCOME_SOURCE_BUSINESS	INCOME_SOURCE_COMMISSION	INCOME_SO
13401.256807	0.5	0.151163	0.031125	0.0	0.0	0.000000	0.024169	0.000418	0.000000	15.527737	120.419269	False	False	False	
4230.004965	0.0	0.139535	0.255402	0.0	0.0	0.666667	0.123867	0.011176	0.000000	14.608637	121.031947	True	False	False	
4481.937304	0.0	0.151163	0.218675	0.0	0.0	0.666667	0.061934	0.001719	0.077778	14.608637	121.031947	False	False	False	
4734.959768	0.0	0.151163	0.101645	0.0	0.0	0.000000	0.015106	0.004222	0.077778	14.608637	121.031947	False	False	False	
4828.128416	0.5	0.151163	0.146732	0.0	0.0	0.333333	0.000000	0.000176	0.466667	14.608637	121.031947	False	False	False	

- **Characteristics:**
  - CUST_INFO and CREDIT_TRANSACTIONS dataset
  - Scaled all except Flags, Longitude, Latitude
  - MinMaxScaler()
  - Tenure (has 0.64 corr with Age) and Frequency (has 0.43 corr with Amount) included
- **PCA:**
  - Cumulative Explained Variance Ratio
  - 

  - Scree Plot (Eigenvalue size vs Eigenvalue number)
  - 

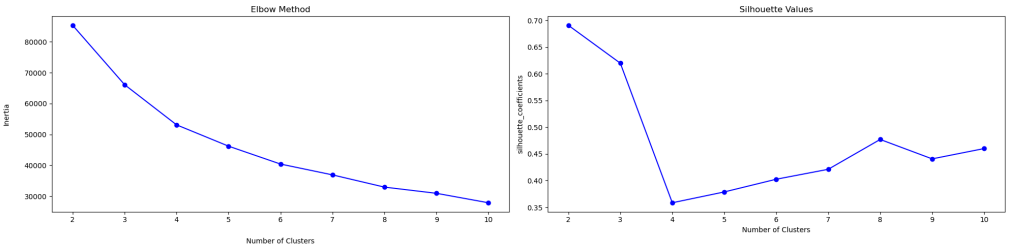
  - Chose **5** as the number of components (ok?)
  - Correlation Matrix

- 

  - Pairplot
  - 

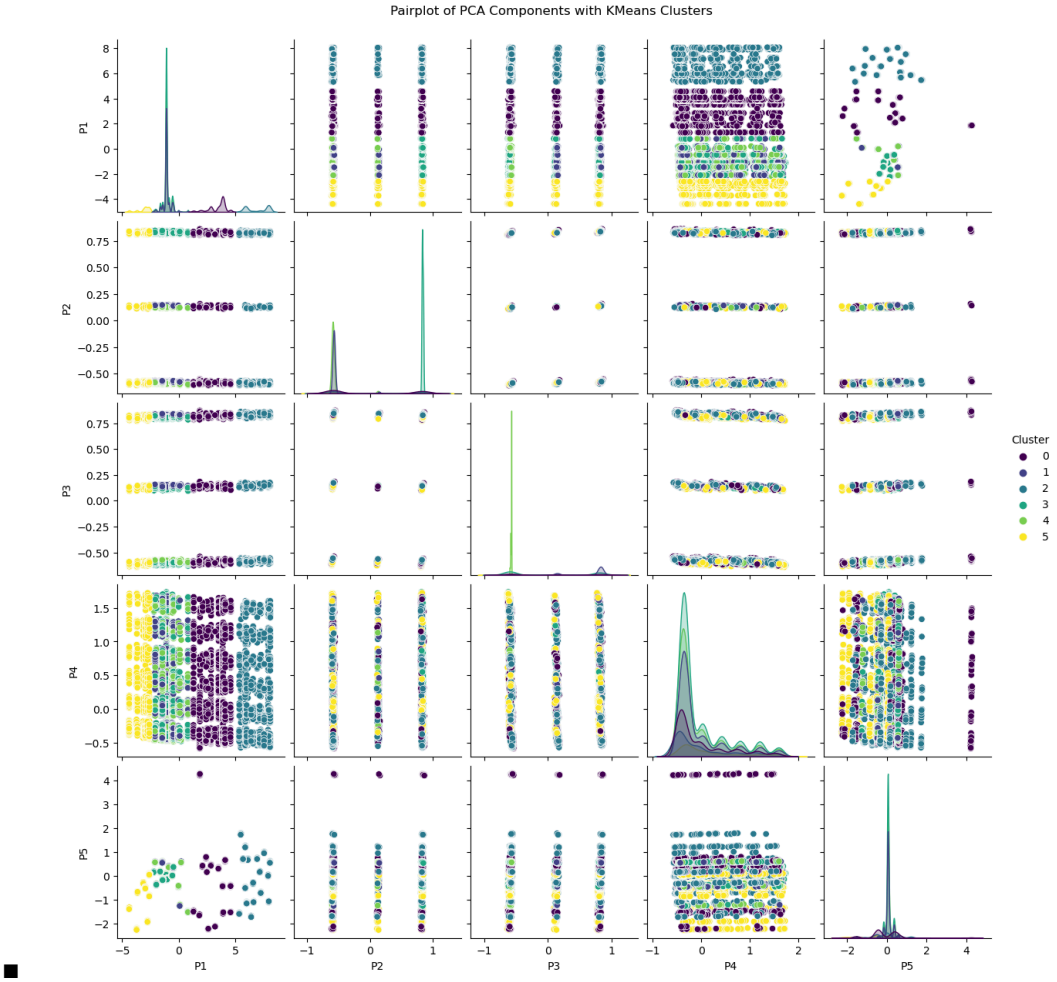
- **Hyperparameter Tuning:**
  - For this trial, we performed hyperparameter tuning
  - We set the range of numbers of clusters from 2 - 10 (ok?)
  - We use the **silhouette coefficient** as the metric
  - random_state = 42
  - Paramer Grid:
    -
```
param_grid = {
    'n_clusters': range(2, 11),
    'init': ['k-means++'],
    'n_init': [10, 20, 30],
    'algorithm': ['lloyd', 'elkan']
}
```
(ok?)
  - Best Parameters:
    - {'algorithm': 'lloyd', 'init': 'k-means++', 'n_clusters': 2, 'n_init': 10}
    - Best number of clusters: 2
    - Best silhouette score: 0.5575086885540613
- **Elbow Method and Silhouette Coefficients:**

- Used the best parameters in hyperparameter tuning and used elbow method to check

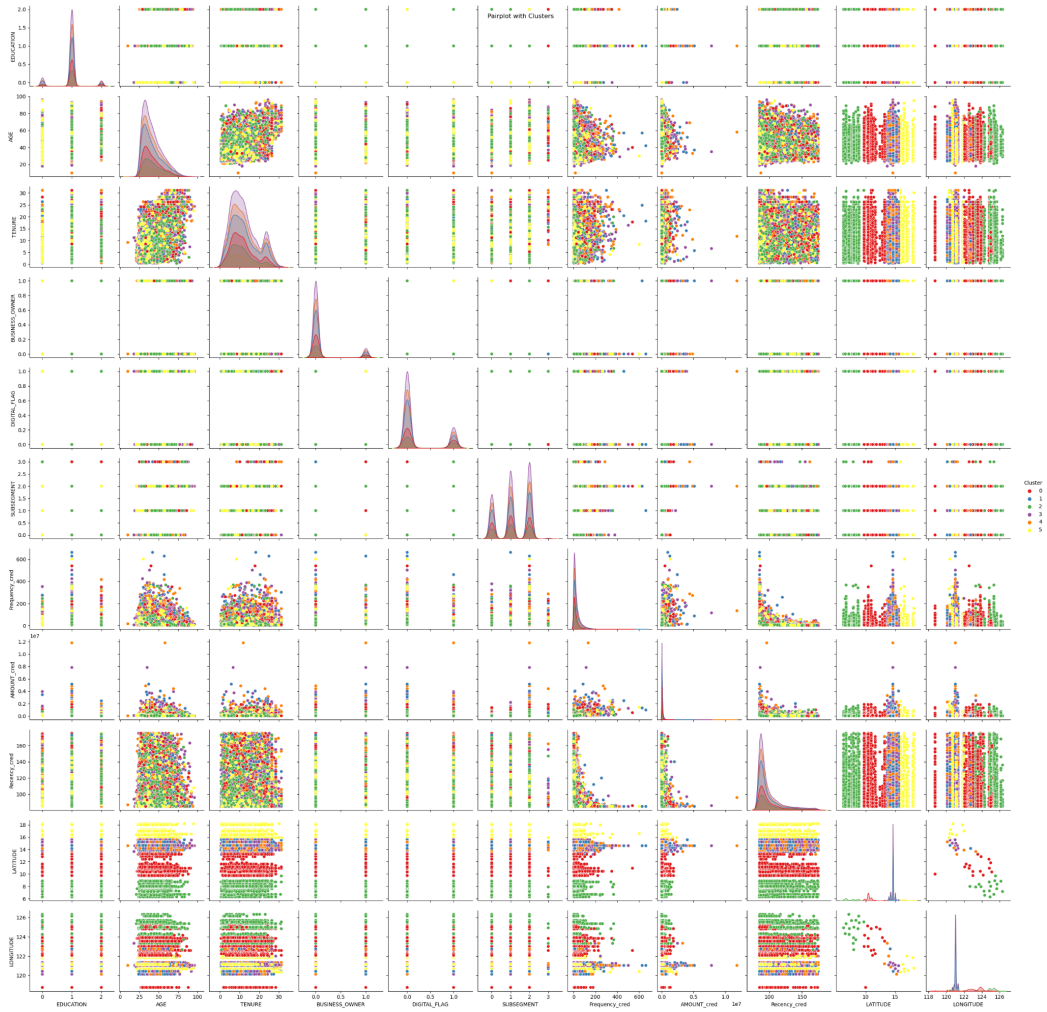


- Silhouette Value for n_clusters = 2 is high but it does not appear to be the elbow. **We try n_clusters = 6**.

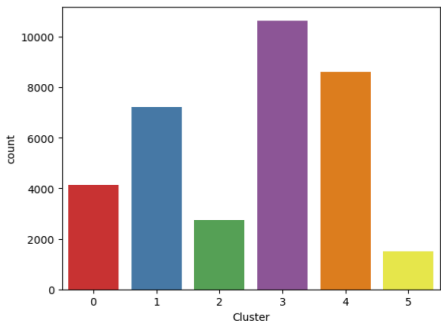
- **Cluster EDA:**
  - **PCA Pairplot:**



- **Pairplot (Original Merged Dataset):**



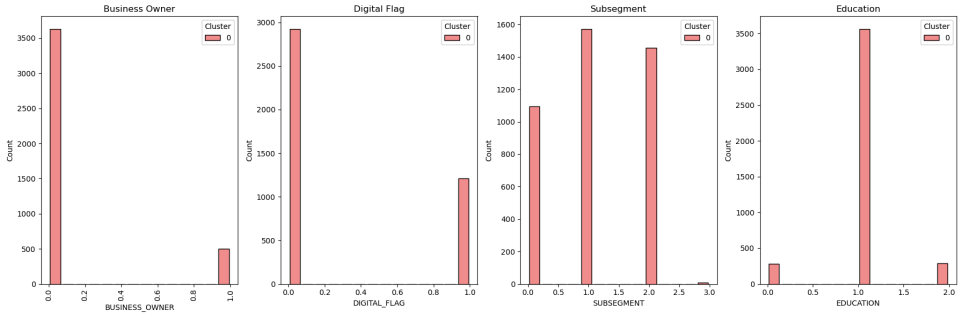
-
- Clustering became based on Longitude and Latitude (maybe due to them not being scaled)
- Cluster 5 = Northern Luzon, Cluster 1, 3, 4 = Southern Luzon + NCR, Cluster 0 = Visayas, Cluster 2 = Mindanao
- Same distribution of Ages, Tenure, Recency for all clusters
- Cluster 1, 3, 4 tends to spend more (Southern Luzon + NCR)
- Cluster 2 and 5 are less frequent



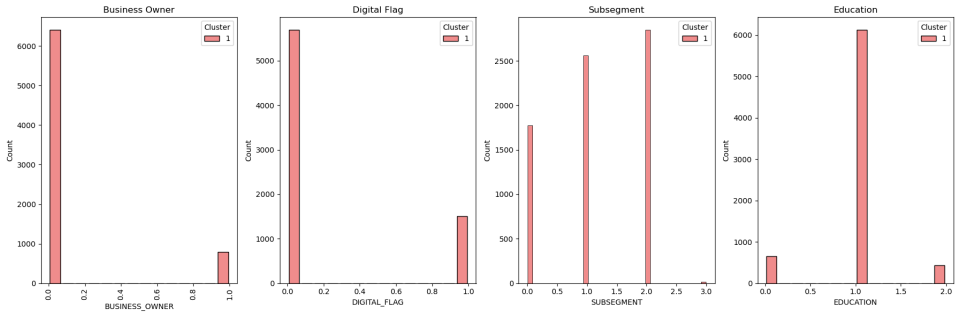
- **Histograms:**

Label Encoded

- **Cluster 0 and 2**



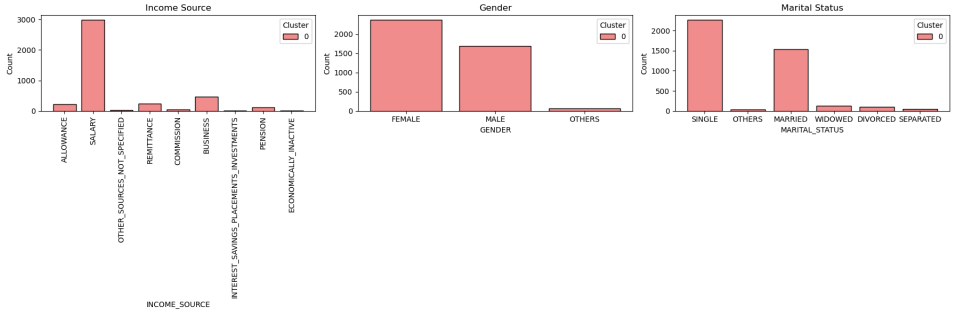
- Slightly more lower middle class
- **Cluster 1, 3, 4, and 5**



- Mostly No Business, Mostly Digital, Mostly lower to upper mid class (slightly more upper mid), mostly mid educated

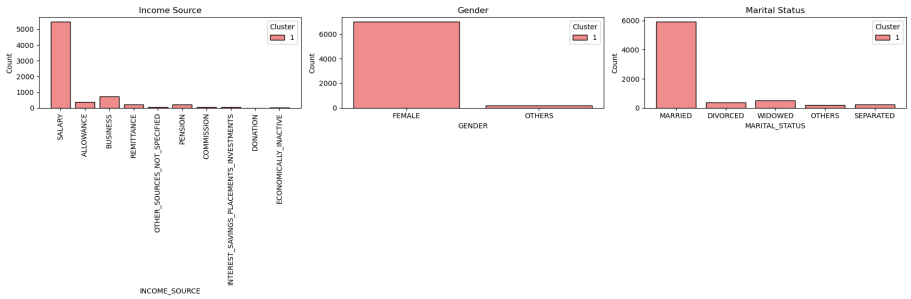
Income Source, Gender, Marital Status

- **Cluster 0:**



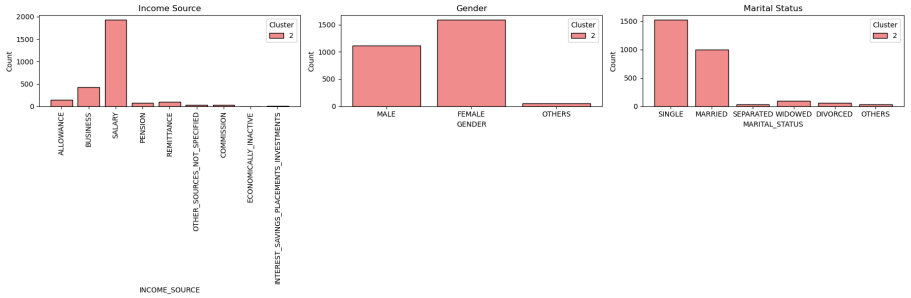
- Mostly Salary, Mostly both male and female (slightly more female), Mostly both single and married (slightly more single)

- **Cluster 1:**



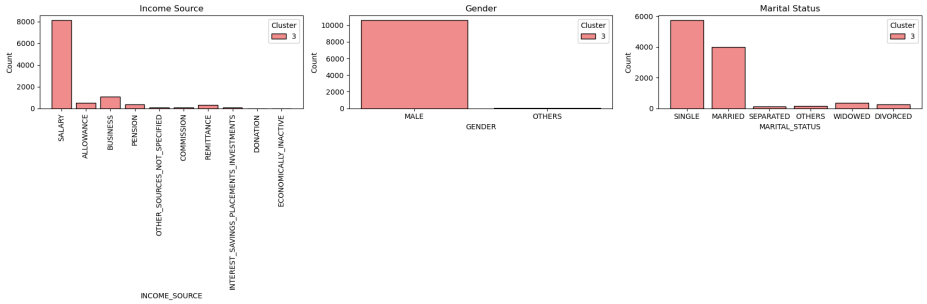
- Mostly Salary, Mostly female, Mostly married

- **Cluster 2:**



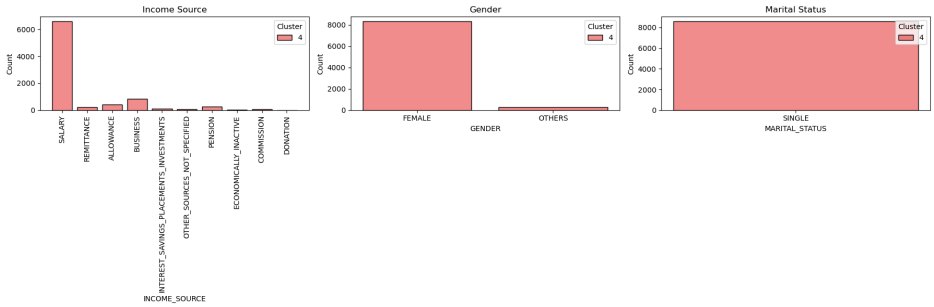
- Mostly salary, Mostly both male and female (slightly more female), Mostly both single and married (slightly more single)

- **Cluster 3:**



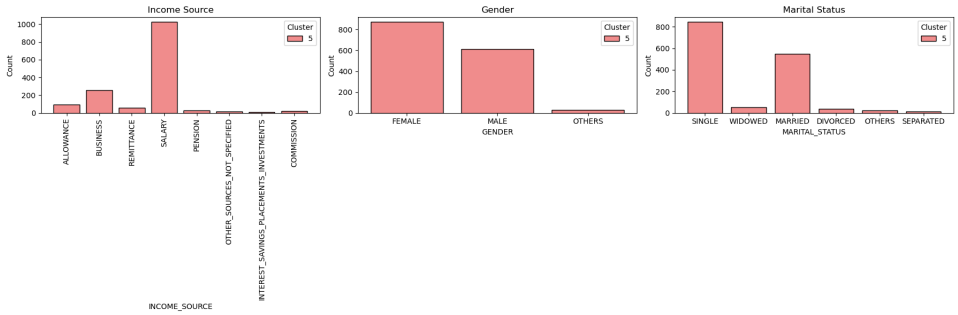
- Mostly salary, Mostly male, Mostly both single and married (slightly more single)

- **Cluster 4:**



- Mostly salary, Mostly female, Single only

- **Cluster 5:**

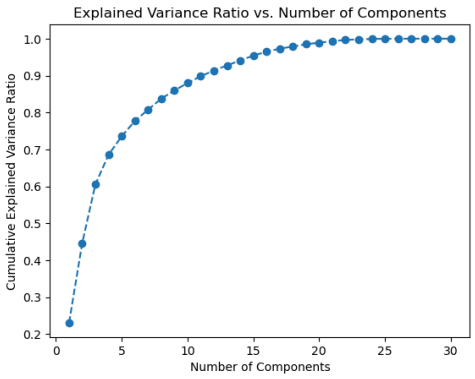


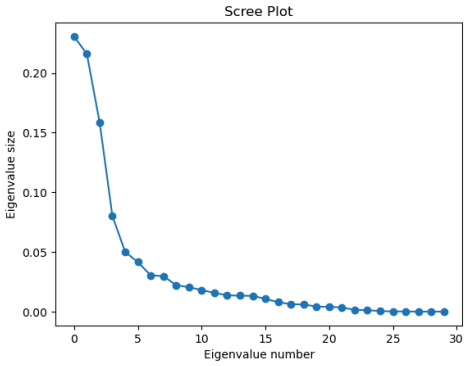
- Mostly salary, Mostly both male and female (slightly more female), Mostly both single and married (slightly more single)

TRIAL 3

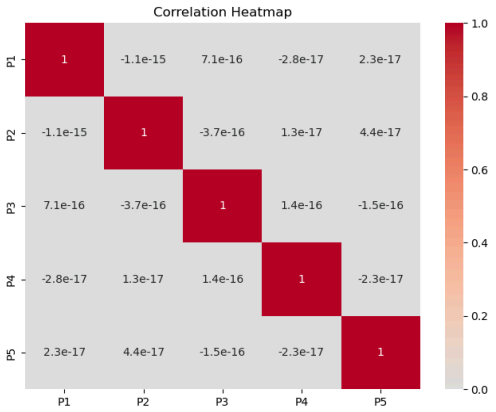
- **File Name:** clustering_trial_3.ipynb
- **Dataset:** minmax_scaled_all_df.parquet
-

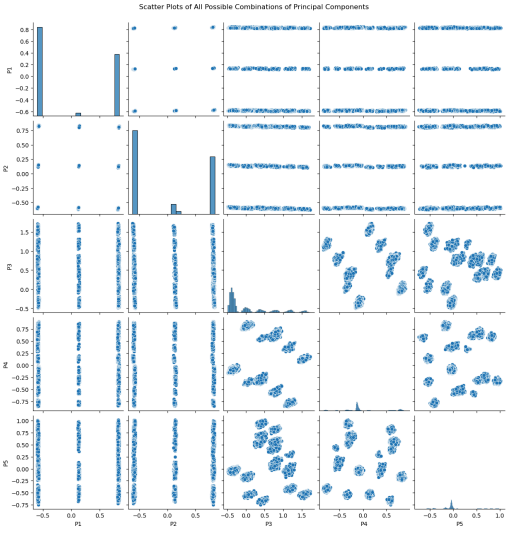
CUST_NUM	EDUCATION	AGE	TENURE	BUSINESS_OWNER	DIGITAL_FLAG	SUBSEGMENT	Frequency_cred	AMOUNT_cred	Recency_cred	LATITUDE	LONGITUDE	INCOME_SOURCE_ALLOWANCE	INCOME_SOURCE_BUSINESS	INCOME_SOURCE_COMMISSION	INCOME_SOU
13401.256807	0.5	0.151163	0.031125	0.0	0.0	0.000000	0.024169	0.000418	0.000000	0.777889	0.220123	0.0	0.0	0.0	
4230.004965	0.0	0.139535	0.255402	0.0	0.0	0.666667	0.123867	0.011176	0.000000	0.700531	0.300916	1.0	0.0	0.0	
4481.937304	0.0	0.151163	0.218675	0.0	0.0	0.666667	0.061934	0.001719	0.077778	0.700531	0.300916	0.0	0.0	0.0	
4734.959768	0.0	0.151163	0.101645	0.0	0.0	0.000000	0.015106	0.004222	0.077778	0.700531	0.300916	0.0	0.0	0.0	
4828.128416	0.5	0.151163	0.146732	0.0	0.0	0.333333	0.000000	0.000176	0.466667	0.700531	0.300916	0.0	0.0	0.0	

- **Characteristics:**
  - CUST_INFO and CREDIT_TRANSACTIONS dataset
  - Scaled all
  - MinMaxScaler()
  - Tenure (has 0.64 corr with Age) and Frequency (has 0.43 corr with Amount) included
- **PCA:**
  - Cumulative Explained Variance Ratio
  - 

  - Scree Plot (Eigenvalue size vs Eigenvalue number)
  - 

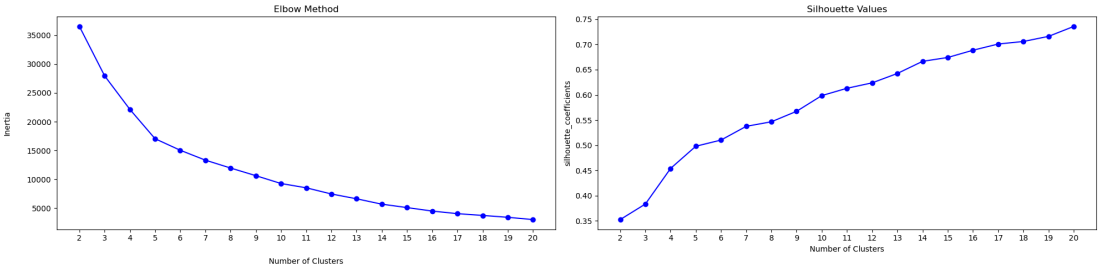
  - Chose **5** as the number of components (ok?)

- Correlation Matrix
- 

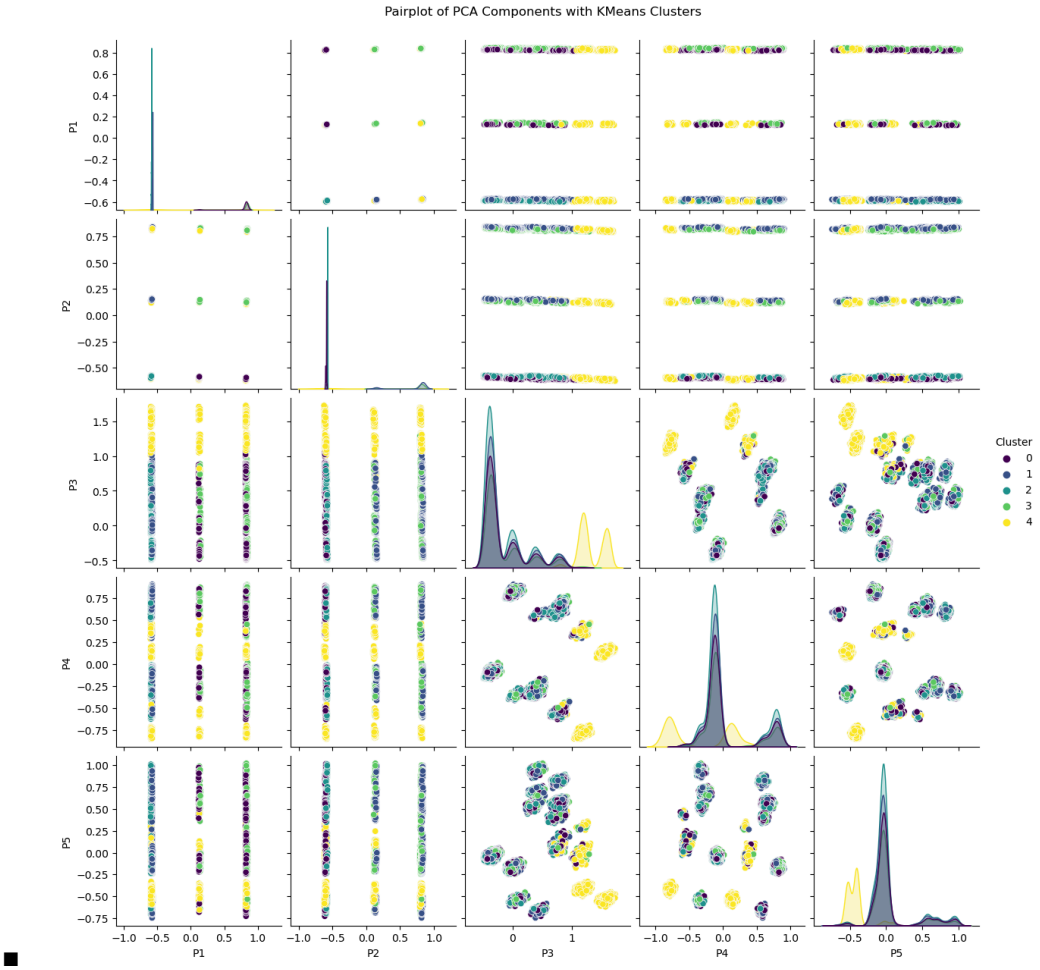
- Pairplot
- 

- **Hyperparameter Tuning:**
  - For this trial, we performed hyperparameter tuning
  - We set the range of numbers of clusters from 2 - 10 (ok?)
  - We use the **silhouette coefficient** as the metric
  - random_state = 42
  - Paramer Grid:
    -
```
param_grid = {
    'n_clusters': range(2, 11),
    'init': ['k-means++'],
    'n_init': [10, 20, 30],
    'algorithm': ['lloyd', 'elkan']
}
```
(ok?)
  - Best Parameters:
    - {'algorithm': 'lloyd', 'init': 'k-means++', 'n_clusters': 10, 'n_init': 30}
    - Best number of clusters: 10
    - Best silhouette score: 0.5992802537438063

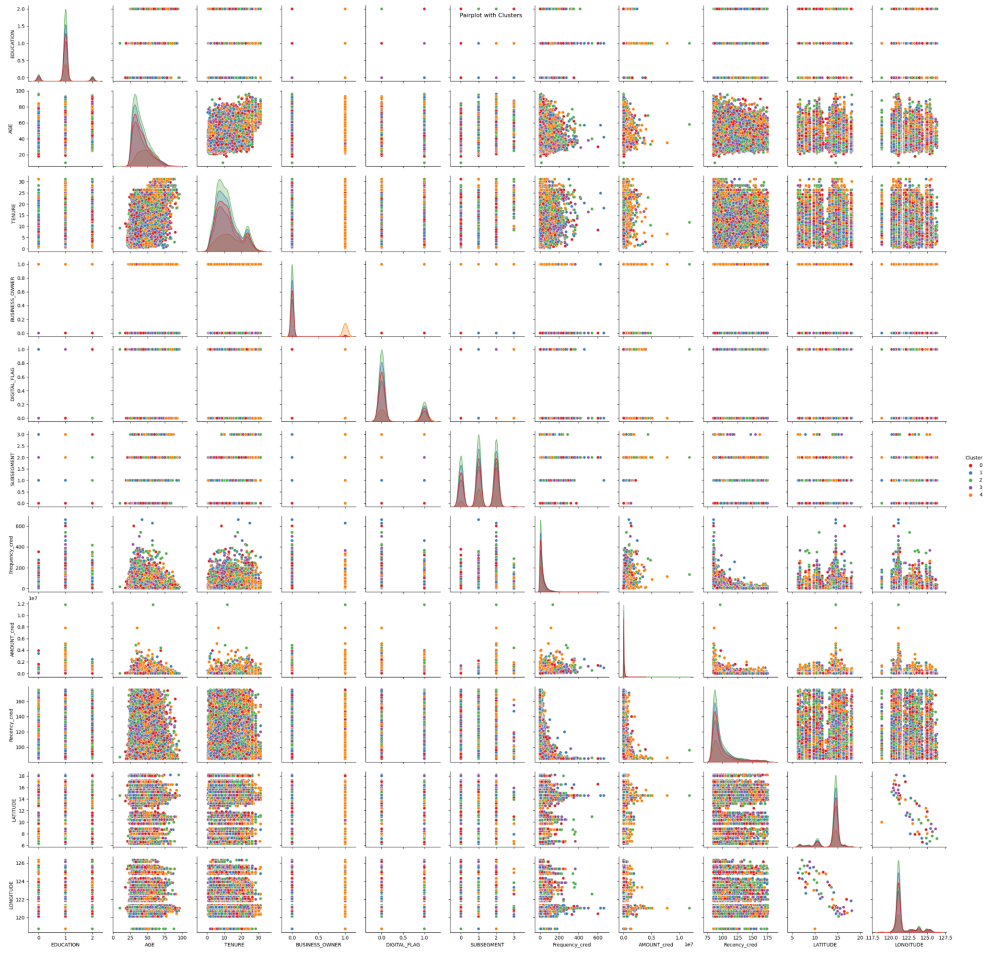
- **Elbow Method and Silhouette Coefficients:**
  - Used the best parameters in hyperparameter tuning and used elbow method to check



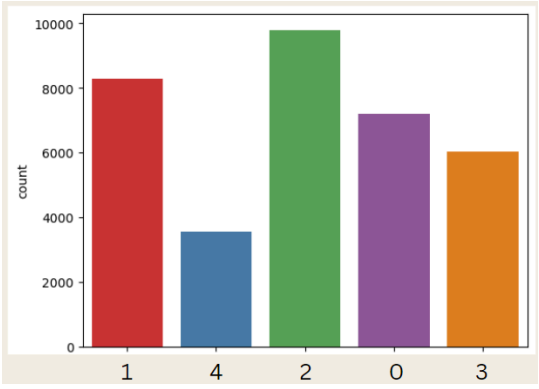
  - n_clusters = 5 appears to be the elbow (with > 0.5 silhouette coefficient). **We choose n_clusters = 5.**
- **Cluster EDA:**
  - **PCA Pairplot:**



- **Pairplot (Original Merged Dataset):**



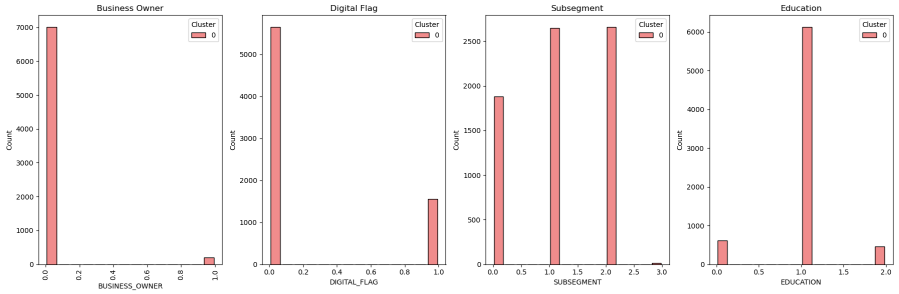
  - No apparent distinctions between clusters (RFM) except for cluster 4 and 2
  - Cluster 2 has higher frequency and amount
  - Cluster 4 has many older people than other clusters



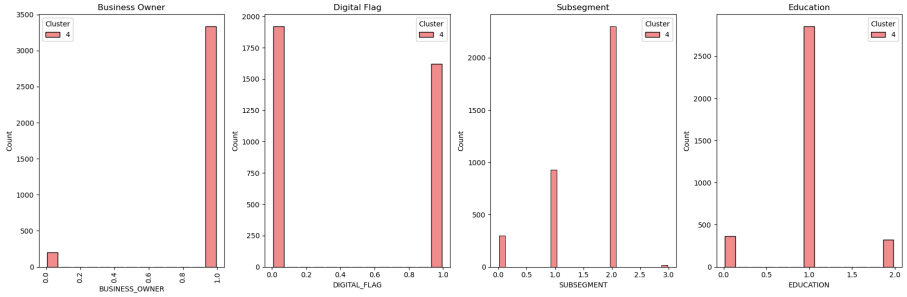
- **Histograms:**

Label Encoded

- **Clusters 0, 1, 2, 3**



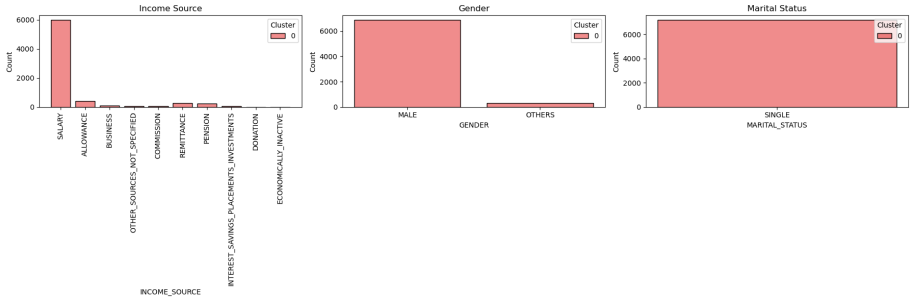
- Mostly no business, mostly digital, mostly lower to upper mid class, mostly mid educated
- **Cluster 4**



- Mostly business owners, both digital and traditional, mostly upper mid class, mostly mid educated

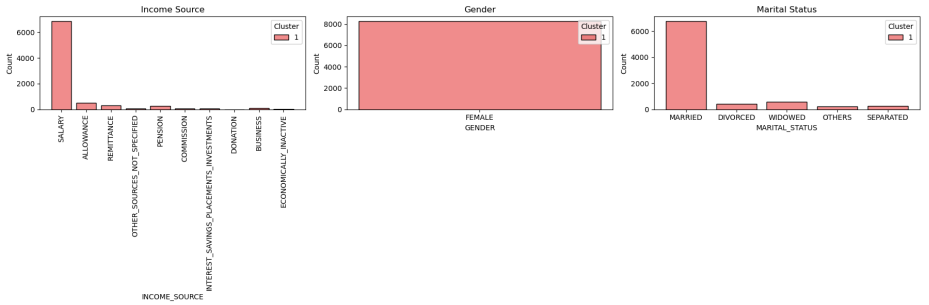
Income Source, Gender, Marital Status

- **Cluster 0 Histogram:**



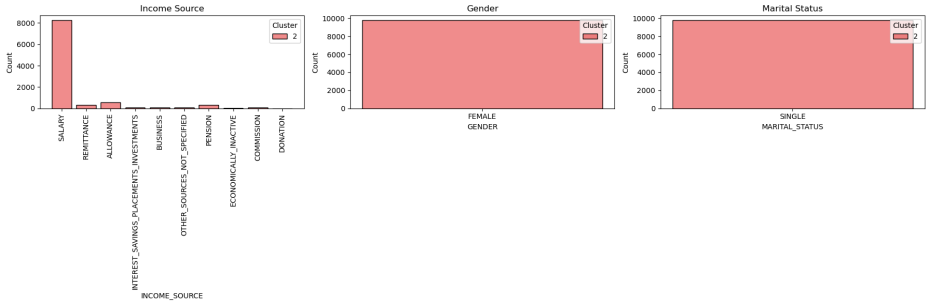
- Mostly salary, mostly male, single only

- **Cluster 1 Histogram:**



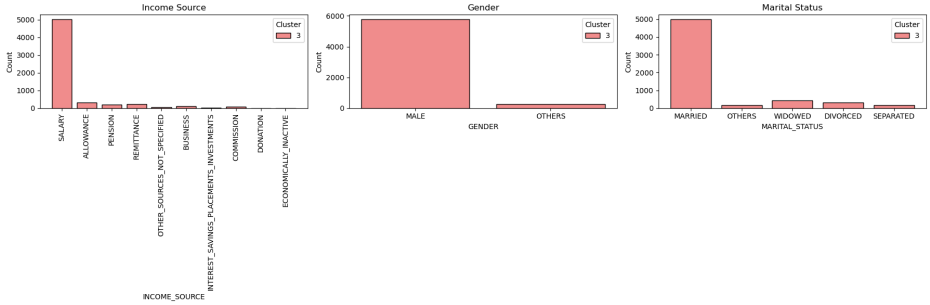
- Mostly salary, female only, mostly married

- **Cluster 2 Histogram:**



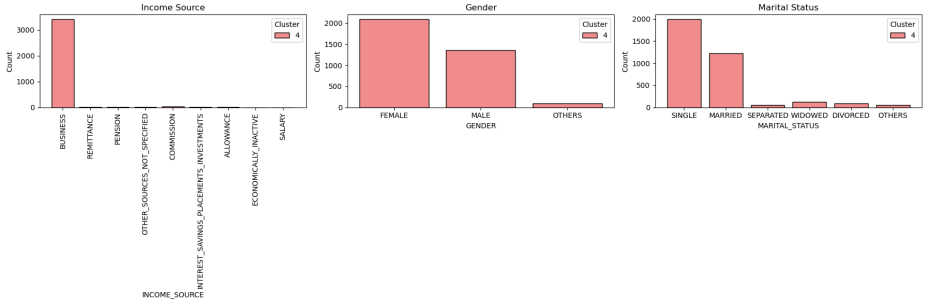
- Mostly salary, female only, single only

- **Cluster 3 Histogram:**



- Mostly salary, mostly male, mostly married

- **Cluster 4 Histogram:**



- **Mostly business**, both male and female (slightly more female), mostly single and married (slightly more single)

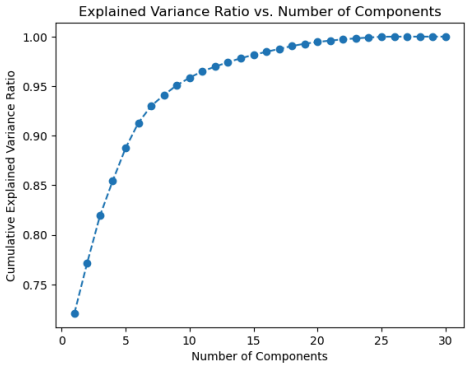
TRIAL 4

- **File Name:** clustering_trial_4.ipynb
- **Dataset:** minmax_binned_scaled_all_except_long_lat_df.parquet
-

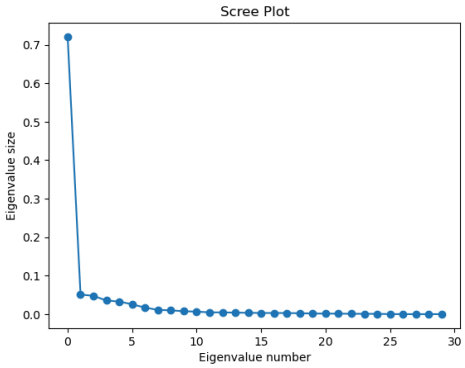
CUST_NUM	EDUCATION	AGE	TENURE	BUSINESS_OWNER	DIGITAL_FLAG	SUBSEGMENT	LATITUDE	LONGITUDE	INCOME_SOURCE_ALLOWANCE	I
13401.256807	0.5	0.151163	0.031125	0.0	0.0	0.000000	15.527737	120.419269	False	
4230.004965	0.0	0.139535	0.255402	0.0	0.0	0.666667	14.608637	121.031947	True	
4481.937304	0.0	0.151163	0.218675	0.0	0.0	0.666667	14.608637	121.031947	False	
4734.959768	0.0	0.151163	0.101645	0.0	0.0	0.000000	14.608637	121.031947	False	
4828.128416	0.5	0.151163	0.146732	0.0	0.0	0.333333	14.608637	121.031947	False	

Frequency_bins	Recency_bins	Amount_bins
0.666667	0.000000	0.000000
1.000000	0.000000	1.000000
1.000000	0.666667	0.333333
0.333333	0.666667	0.666667
0.000000	1.000000	0.000000

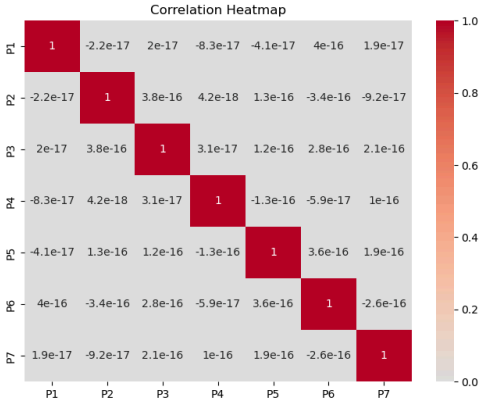
- **Characteristics:**
  - CUST_INFO and CREDIT_TRANSACTIONS dataset
  - Scaled all except longitude and latitude
  - MinMaxScaler
  - Tenure (has 0.64 corr with Age) and Frequency (has 0.43 corr with Amount) included
  - **Qcut RFM**
- **PCA:**
  - Cumulative Explained Variance Ratio



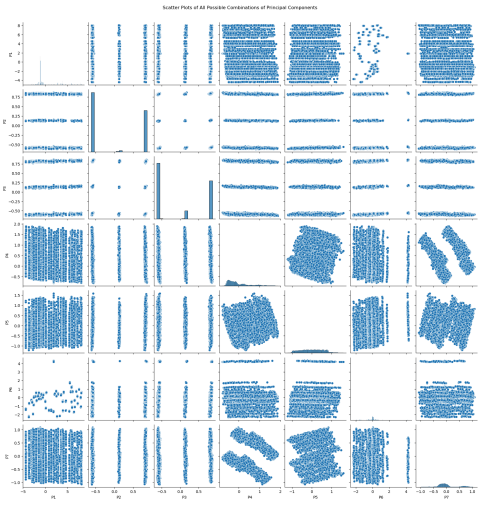
  - Scree Plot (Eigenvalue size vs Eigenvalue number)



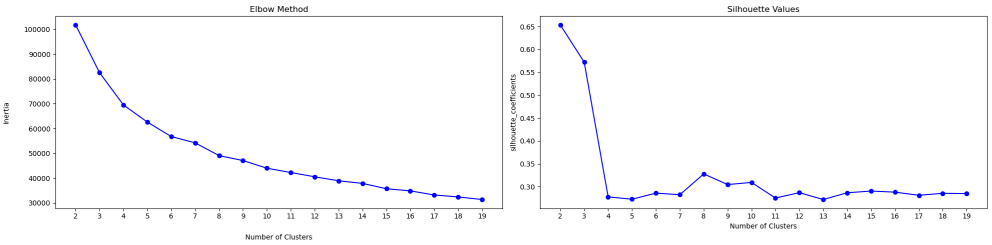
- Chose **7** as the number of components (ok?)
- Correlation Matrix



- Pairplot

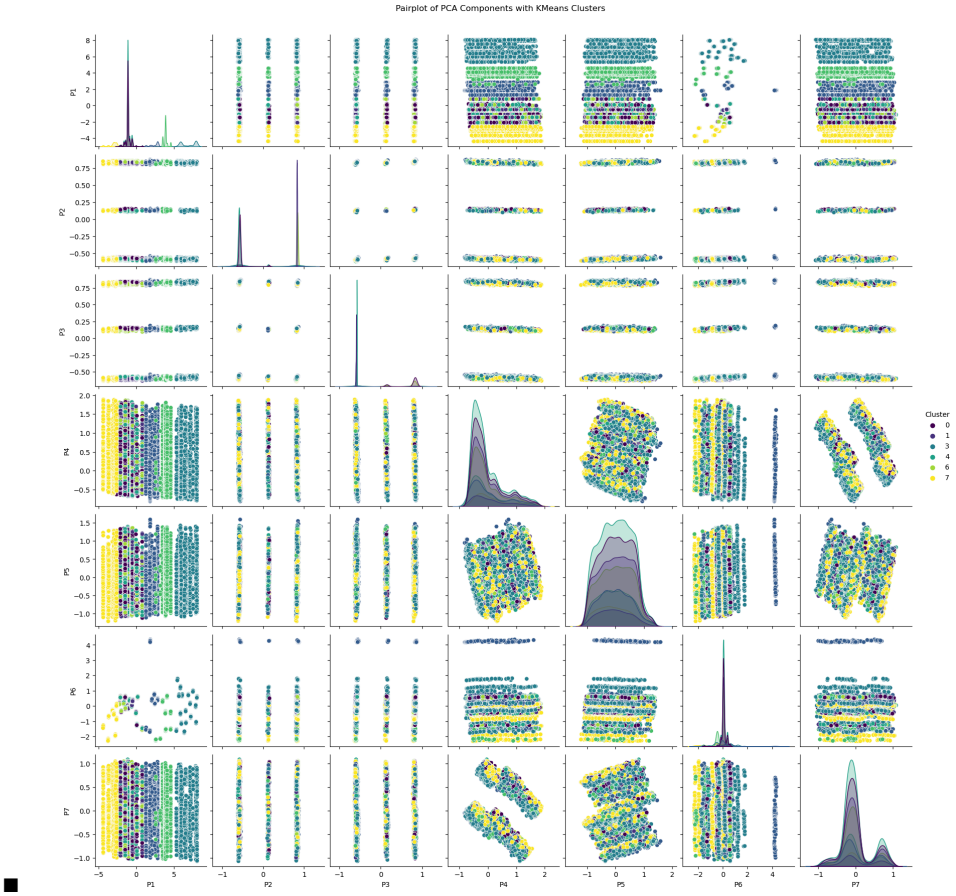


- **Elbow Method and Silhouette Coefficients:**
  - Used the best parameters in hyperparameter tuning and used elbow method to check

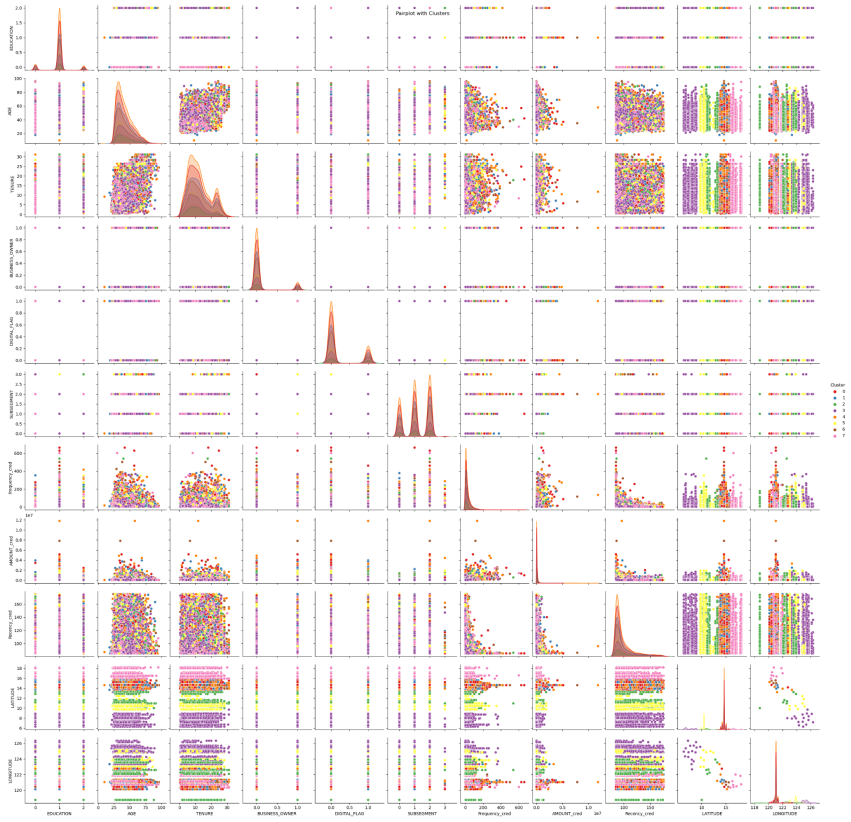


- n_clusters = 8 appears to be the elbow (with ~0.35 silhouette coefficient). **We choose n_clusters = 8.**

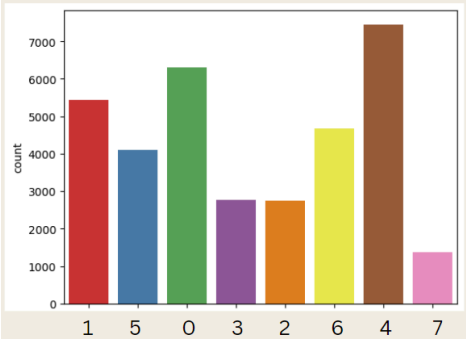
- **Cluster EDA:**
  - **PCA Pairplot:**



- **Pairplot (Original Merged Dataset):**



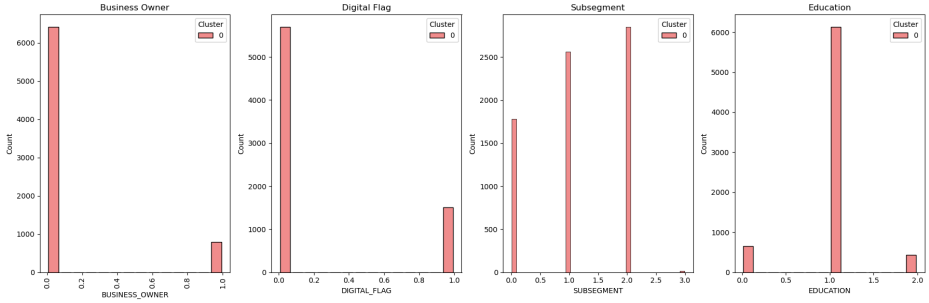
- Cluster is based on longitude and latitude
- The highest AMOUNT_cred in NCR
- Same distribution of AGE, Recency, Frequency, and AMOUNT
- Clusters 0, 6 and 4 have highest transaction AMOUNTs



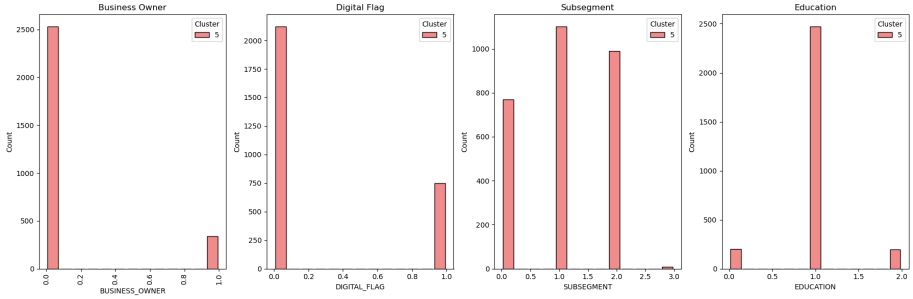
- **Histograms:**

Label Encoded

- **Clusters 0, 1, 4, 6, 7**



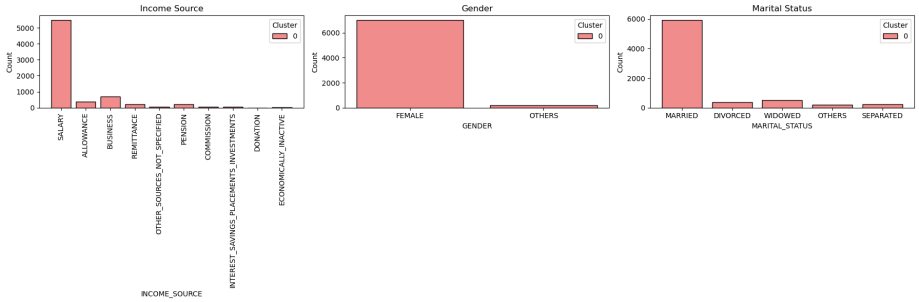
- Mostly no business, mostly digital, mostly lower to upper mid class, mostly mid educated
- **Cluster 2, 3, 5**



- Mostly no business, mostly digital, mostly lower to upper mid class, mostly mid educated

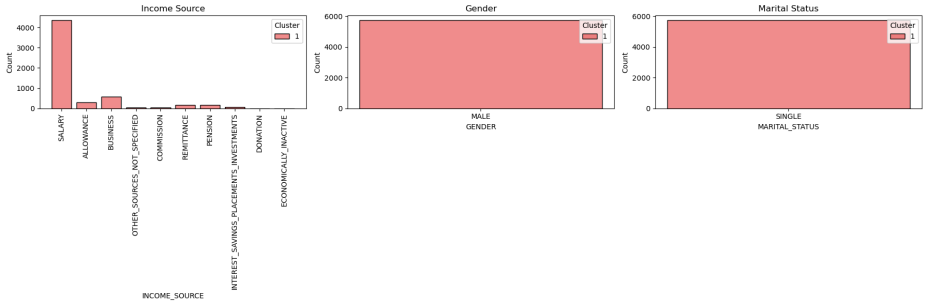
Income Source, Gender, Marital Status

- **Cluster 0 Histogram:**



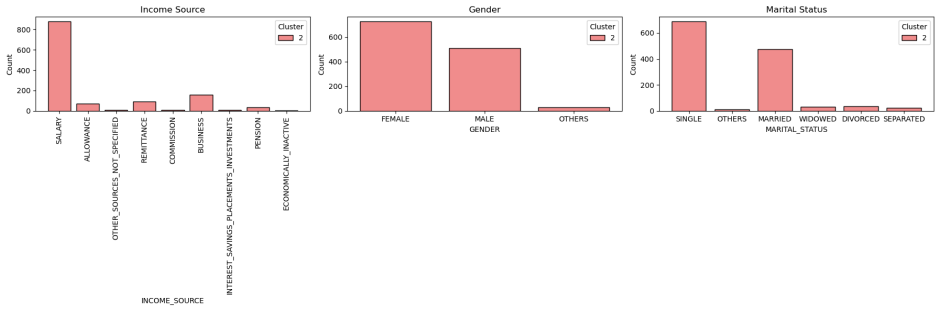
- Mostly salary, mostly female, mostly married

- **Cluster 1 Histogram:**



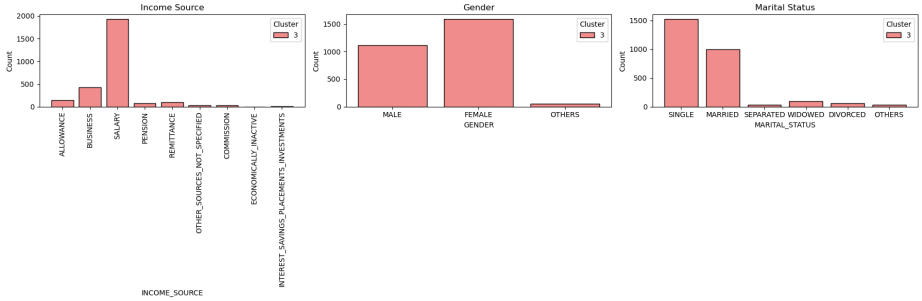
- Mostly salary, male only, single only

- **Cluster 2 Histogram:**



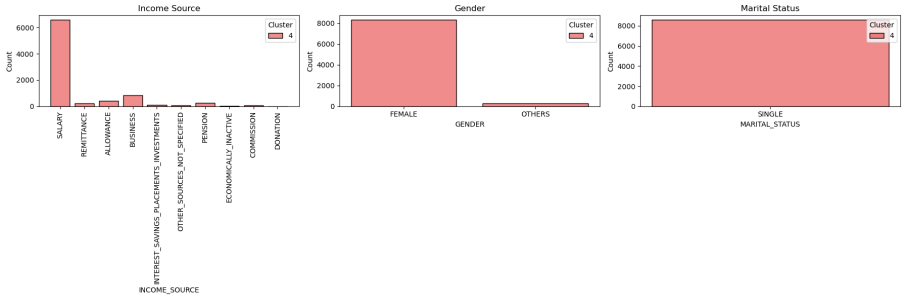
- Mostly salary, both male and female (slightly more female), both single and married (slightly more single)

- **Cluster 3 Histogram:**



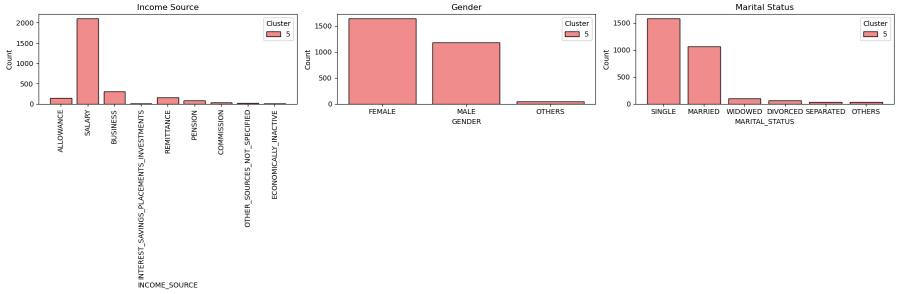
- Mostly salary, both male and female (slightly more female), both single and married (slightly more single)

- **Cluster 4 Histogram:**



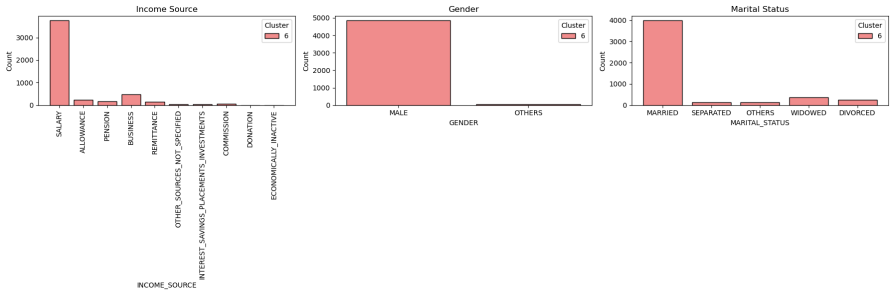
- Mostly salary, mostly female, single only

- **Cluster 5 Histogram:**



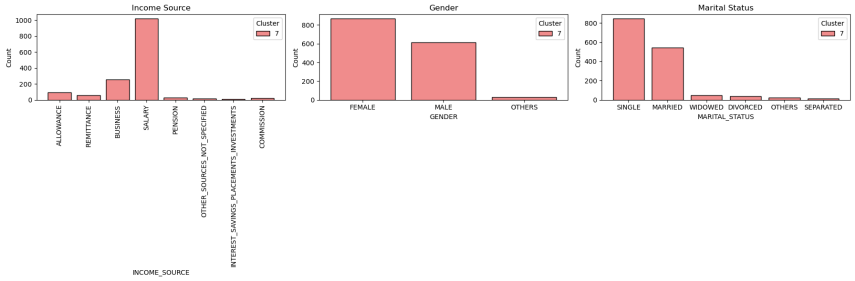
- Mostly salary, both male and female (slightly more female), both single and married (slightly more single)

- **Cluster 6 Histogram:**



- Mostly salary, mostly male, mostly married

- **Cluster 7 Histogram:**



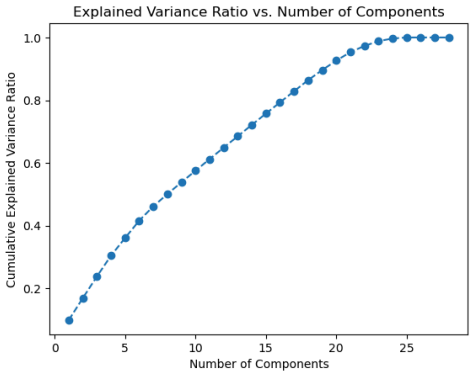
- Mostly salary, both male and female (slightly more female), both single and married (slightly more single)

TRIAL 5

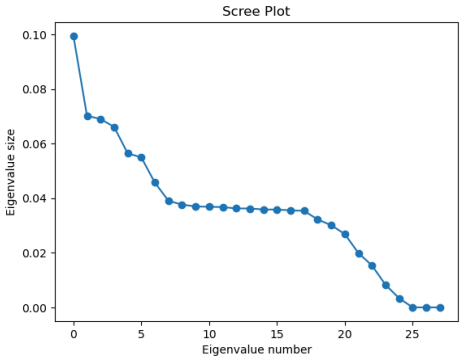
- **File Name:** clustering_trial_5.ipynb
- **Dataset:** stdscale_binned_scaled_all_notenfreq_df.parquet
-

CUST_NUM	EDUCATION	AGE	TENURE	BUSINESS_OWNER	DIGITAL_FLAG	SUBSEGMENT	LATITUDE	LONGITUDE	INCOME_SOURCE_ALLOWANCE	I
13401.256807	0.5	0.151163	0.031125	0.0	0.0	0.000000	15.527737	120.419269	False	
4230.004965	0.0	0.139535	0.255402	0.0	0.0	0.666667	14.608637	121.031947	True	
4481.937304	0.0	0.151163	0.218675	0.0	0.0	0.666667	14.608637	121.031947	False	
4734.959768	0.0	0.151163	0.101645	0.0	0.0	0.000000	14.608637	121.031947	False	
4828.128416	0.5	0.151163	0.146732	0.0	0.0	0.333333	14.608637	121.031947	False	

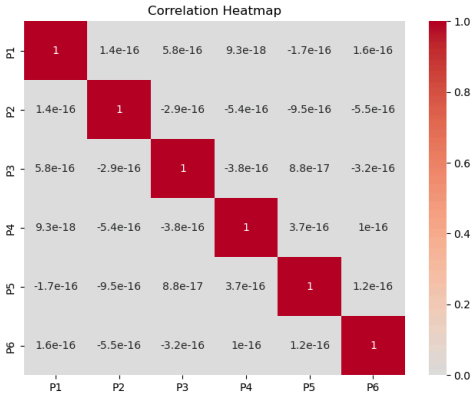
Frequency_bins	Recency_bins	Amount_bins
0.666667	0.000000	0.000000
1.000000	0.000000	1.000000
1.000000	0.666667	0.333333
0.333333	0.666667	0.666667
0.000000	1.000000	0.000000

- **Characteristics:**
  - CUST_INFO and CREDIT_TRANSACTIONS dataset
  - Scaled all
  - StandardScaler
  - Tenure (has 0.64 corr with Age) and Frequency (has 0.43 corr with Amount) **NOT INCLUDED**
  - **Qcut RFM**
- **PCA:**
  - Cumulative Explained Variance Ratio
  - 

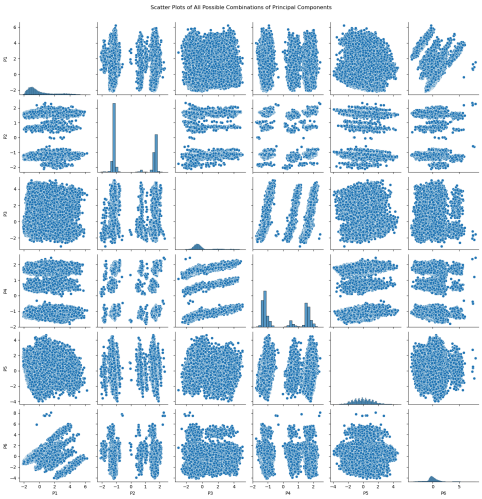
  - Scree Plot (Eigenvalue size vs Eigenvalue number)



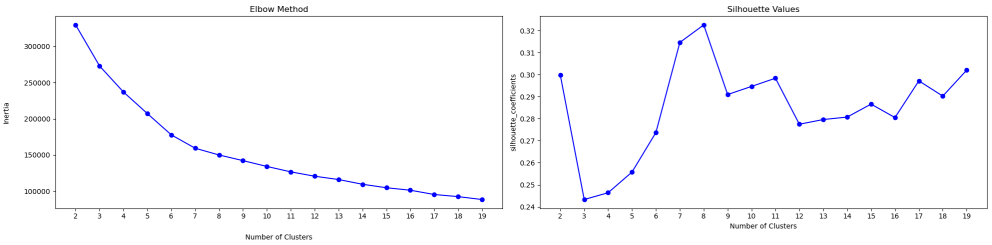
- Chose **6** as the number of components (ok?)
- Correlation Matrix



- Pairplot

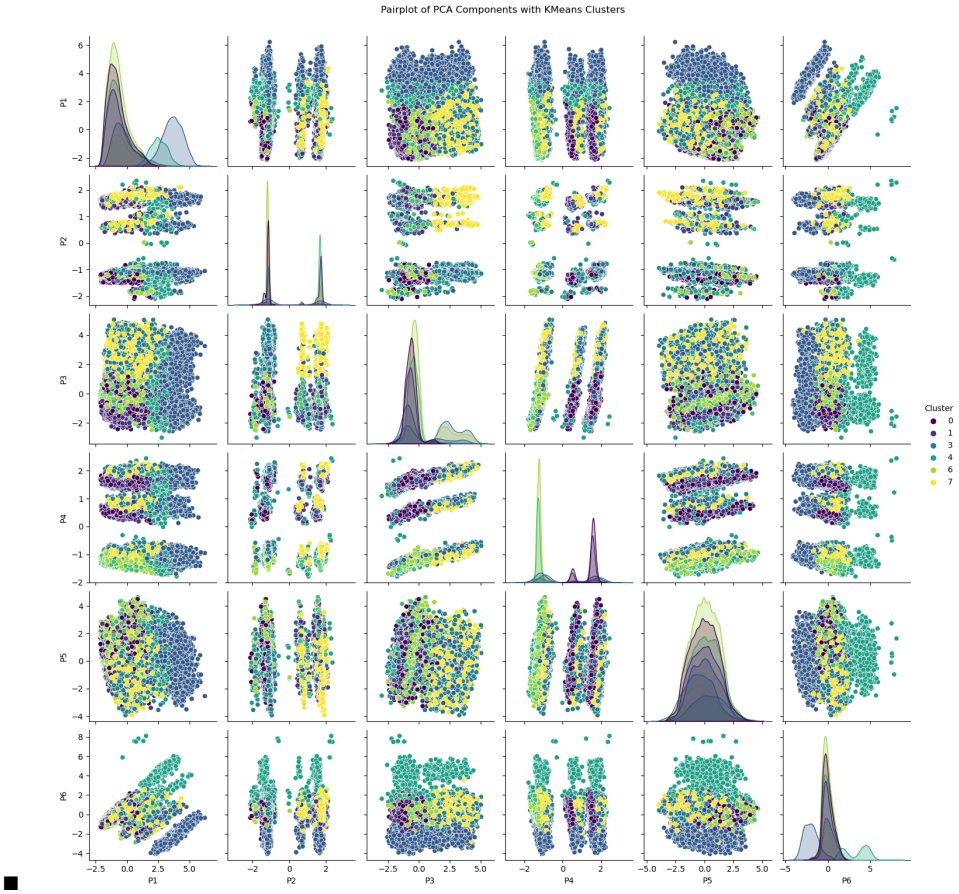


- **Elbow Method and Silhouette Coefficients:**
  - Used the best parameters in hyperparameter tuning and used elbow method to check

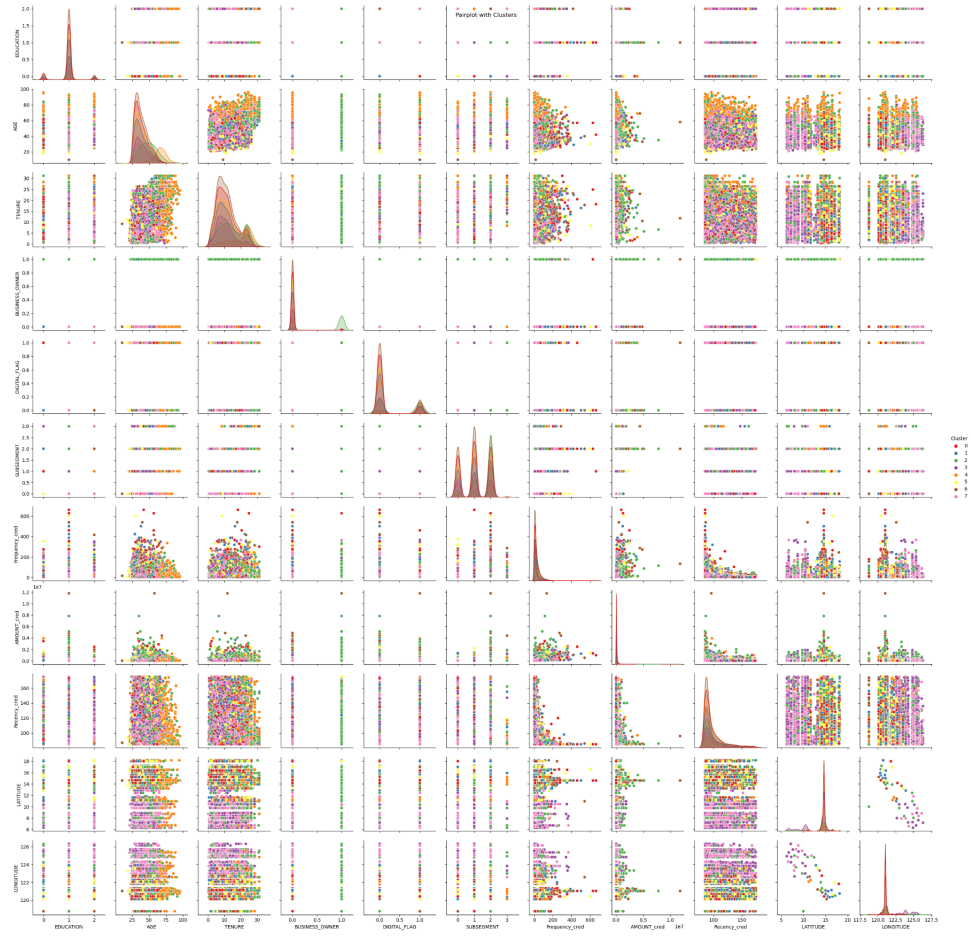


- n_clusters = 8 appears to be the elbow (with ~0.32 silhouette coefficient). **We choose n_clusters = 8.**

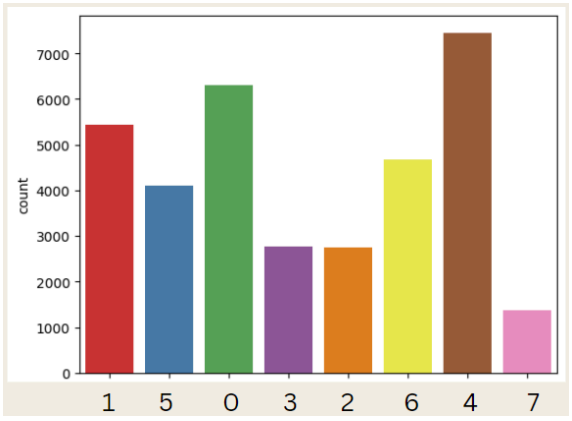
- **Cluster EDA:**
  - **PCA Pairplot:**



- **Pairplot (Original Merged Dataset):**



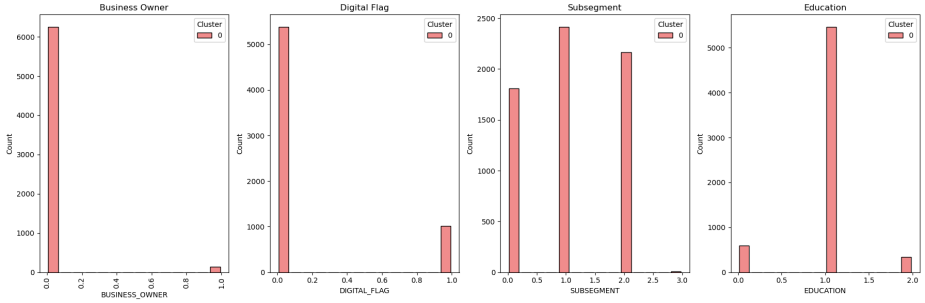
- Clusters 2 and 4 have many older people (especially 4)
- Clusters 6 and 2 have highest transaction amounts



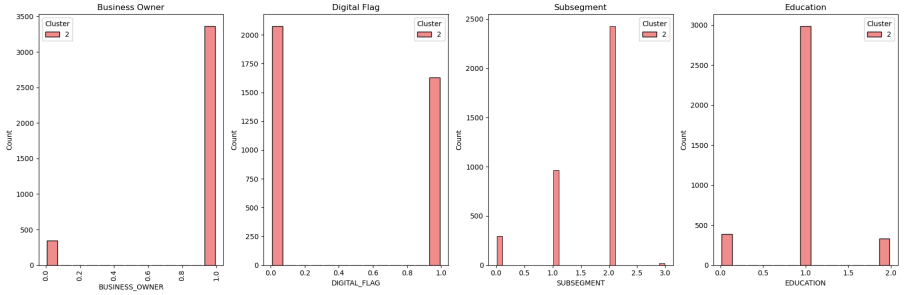
- **Histograms:**

Label Encoded

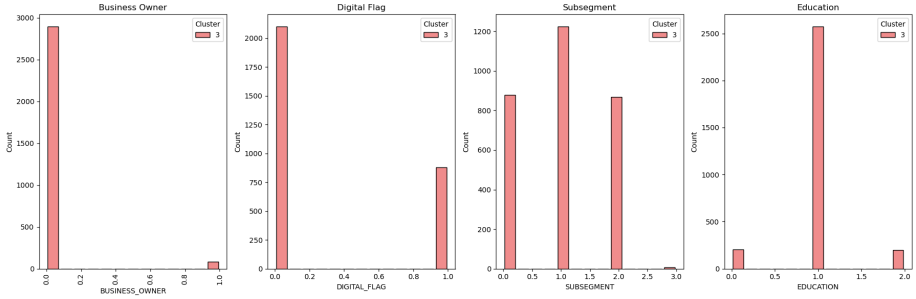
- **Clusters 0, 1, 5, 6**



- Mostly no business, mostly digital, mostly lower to upper mid class, mostly mid educated
- **Cluster 2,**

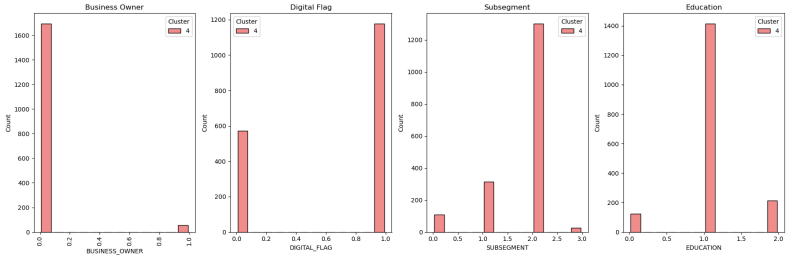


- Mostly business, both digital and traditional, mostly upper middle class, mostly mid educated
- **Cluster 3, 7**



- Mostly no business, mostly digital, mostly lower to upper mid class, mostly mid educated

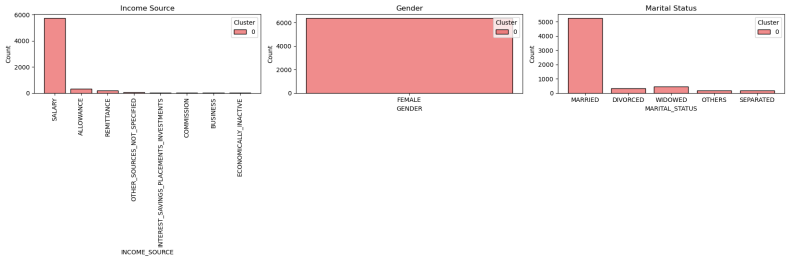
- **Cluster 4**



- Mostly no business, mostly traditional, mostly upper middle class, mostly mid educated

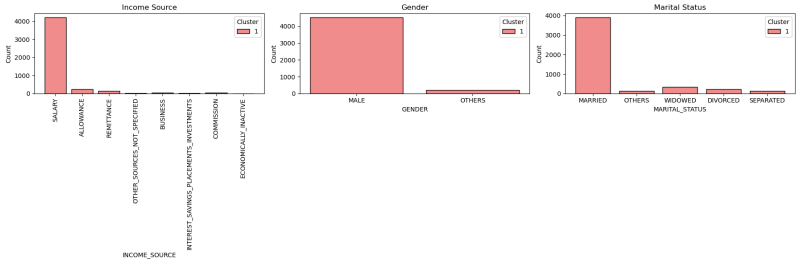
Income Source, Gender, Marital Status

- **Cluster 0 Histogram:**



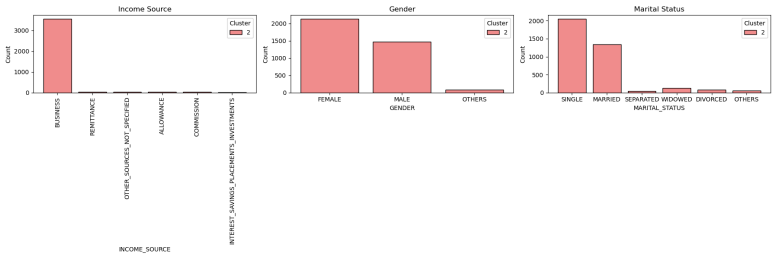
- Mostly salary, female only, mostly married

- **Cluster 1 Histogram:**

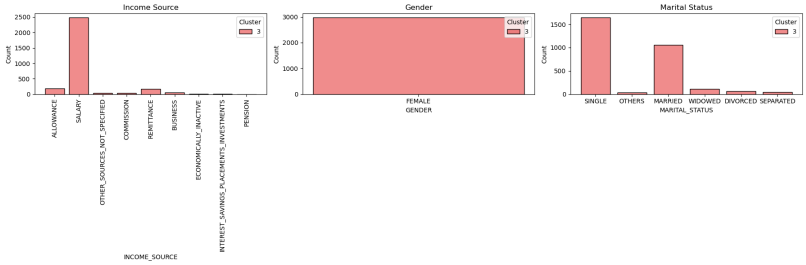


- Mostly salary, mostly male, mostly married

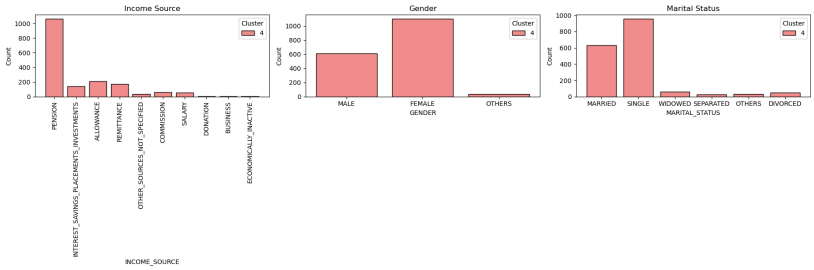
- **Cluster 2 Histogram:**



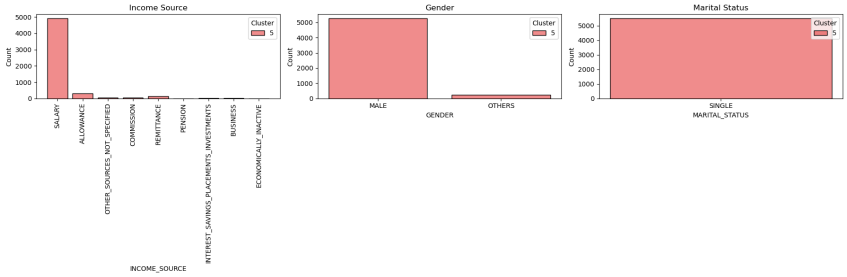
- **Mostly business**, both male and female (slightly more female), both single and married (slightly more single)

- **Cluster 3 Histogram:**
- 

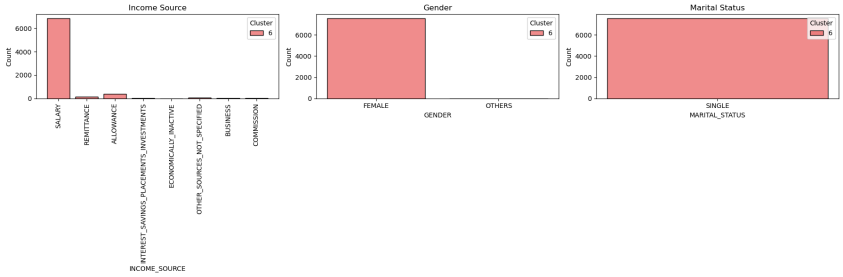
- Mostly salary, female only, both single and married (slightly more single)

- **Cluster 4 Histogram:**
- 

- **Mostly pension**, both male and female (slightly more female), both single and married (slightly more single)

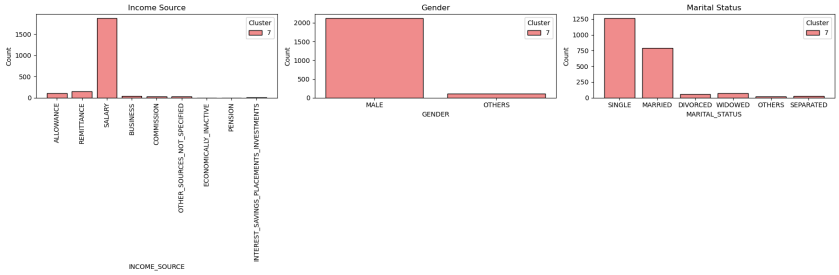
- **Cluster 5 Histogram:**
- 

- Mostly salary, mostly male, single only

- **Cluster 6 Histogram:**
- 

- Mostly salary, mostly female, single only

- **Cluster 7 Histogram:**



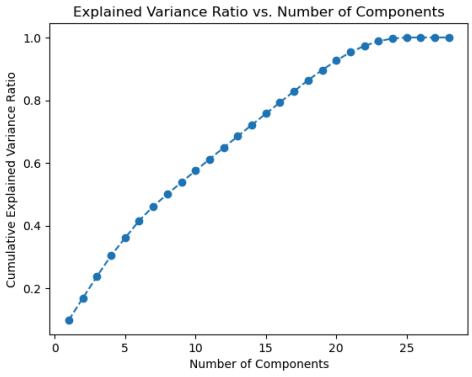
- Mostly salary, mostly male, both single and married (slightly more single)

TRIAL 6 (Other Clustering Algorithms)

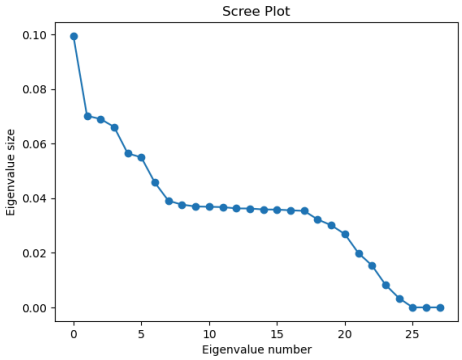
- **File Name:** clustering_trial_6.ipynb
- **Dataset:** stdscale_binned_scaled_all_notenfreq_df.parquet
-

CUST_NUM	EDUCATION	AGE	TENURE	BUSINESS_OWNER	DIGITAL_FLAG	SUBSEGMENT	LATITUDE	LONGITUDE	INCOME_SOURCE_ALLOWANCE	I
13401.256807	0.5	0.151163	0.031125	0.0	0.0	0.000000	15.527737	120.419269	False	
4230.004965	0.0	0.139535	0.255402	0.0	0.0	0.666667	14.608637	121.031947	True	
4481.937304	0.0	0.151163	0.218675	0.0	0.0	0.666667	14.608637	121.031947	False	
4734.959768	0.0	0.151163	0.101645	0.0	0.0	0.000000	14.608637	121.031947	False	
4828.128416	0.5	0.151163	0.146732	0.0	0.0	0.333333	14.608637	121.031947	False	

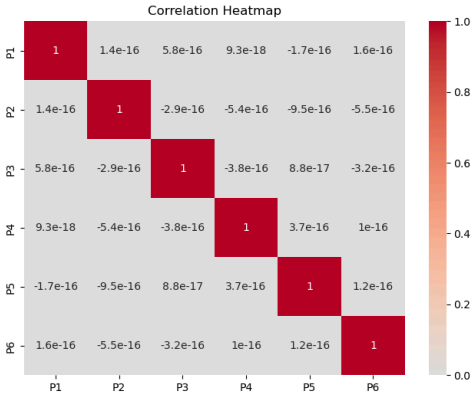
Frequency_bins	Recency_bins	Amount_bins
0.666667	0.000000	0.000000
1.000000	0.000000	1.000000
1.000000	0.666667	0.333333
0.333333	0.666667	0.666667
0.000000	1.000000	0.000000

- **Characteristics:**
  - CUST_INFO and CREDIT_TRANSACTIONS dataset
  - Scaled all
  - StandardScaler
  - Tenure (has 0.64 corr with Age) and Frequency (has 0.43 corr with Amount) **NOT INCLUDED**
  - **Qcut RFM**
- **PCA:**
  - Cumulative Explained Variance Ratio
  - 

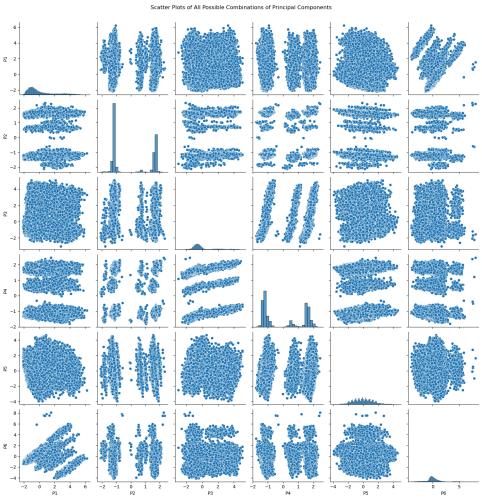
  - Scree Plot (Eigenvalue size vs Eigenvalue number)



- Chose **6** as the number of components (ok?)
- Correlation Matrix



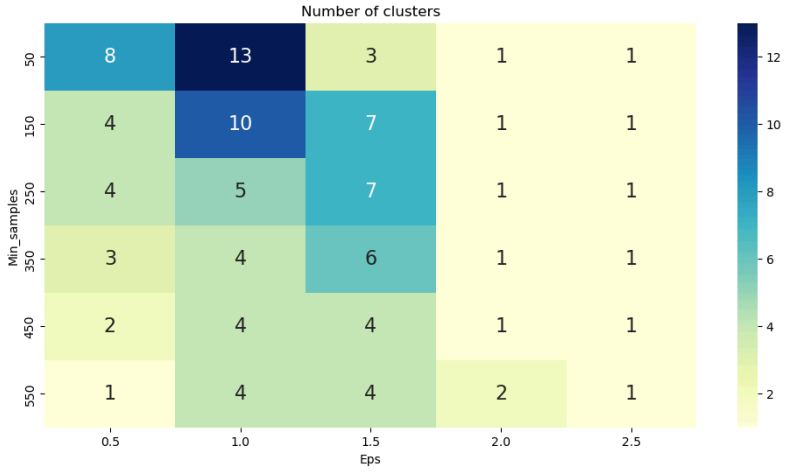
- Pairplot

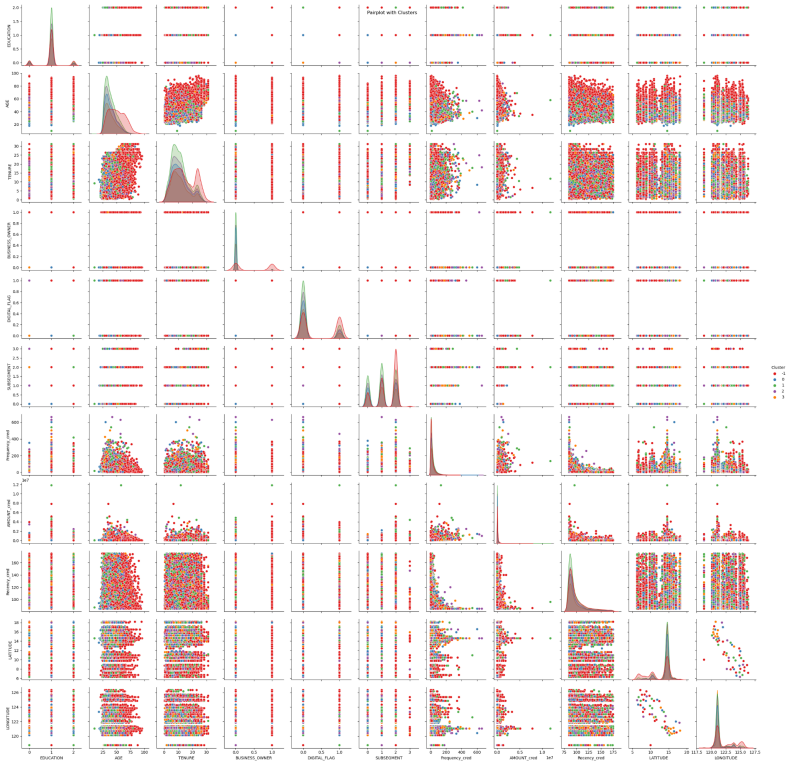


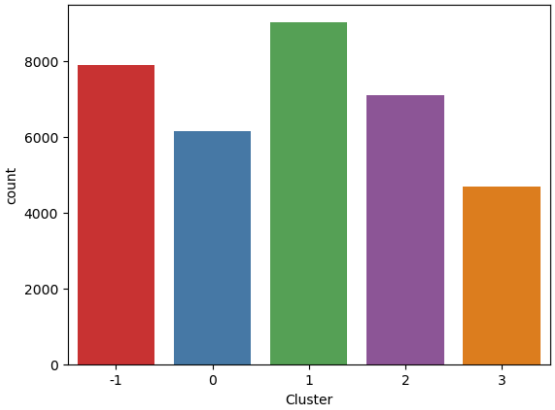
- **DBSCAN**

```
from itertools import product

eps_values = np.arange(0.5,3,0.5) # eps values to be investigated
min_samples = np.arange(50, 650, 100) # min_samples values to be investigated

DBSCAN_params = list(product(eps_values, min_samples))
```

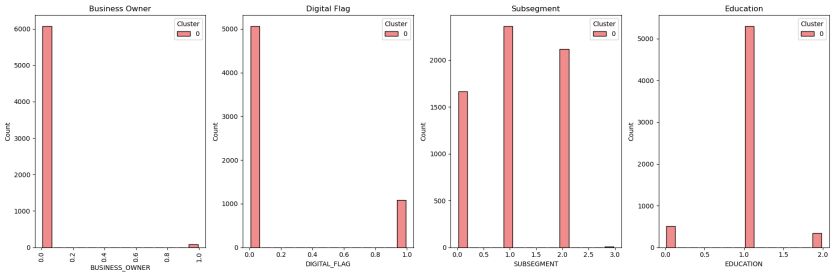
- Went through each eps and min_samples to get the number of clusters

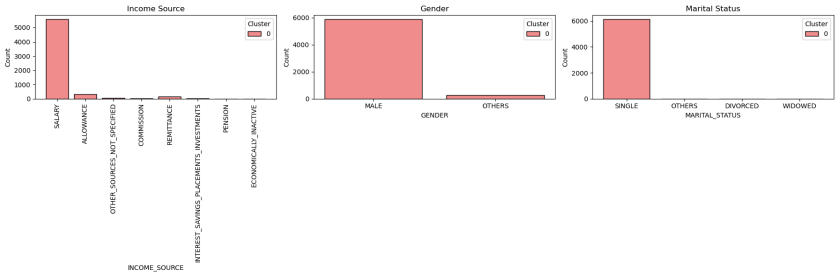
-
```
Number of clusters for each parameter set: [ 8  4  4  3  2  1 13 10  5  4  4  4  3  7  7  6  4  4  1  1  1  1  1  2
  1  1  1  1  1  1]
Silhouette scores for valid parameter sets: [-0.05193456  0.07776386  0.01808764 -0.05993859 -0.14155058  0.01487537
  0.18457262  0.17313031  0.17425535  0.16737846  0.15809517  0.17888718
  0.21477682  0.25470364  0.23383903  0.23825477  0.22590629  0.3062136 ]
```
- 

- We chose 4 number of clusters. min_samples = 450 and eps = 1.5 (silhouette = 0.24)
- **Pairplot**
- 

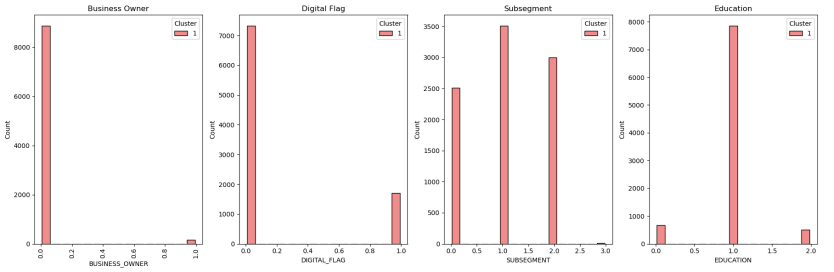


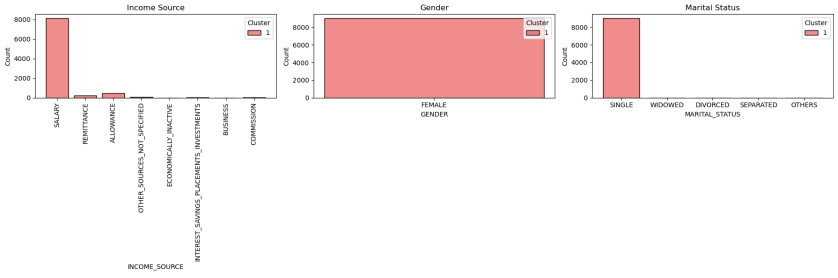
- ## Cluster 0



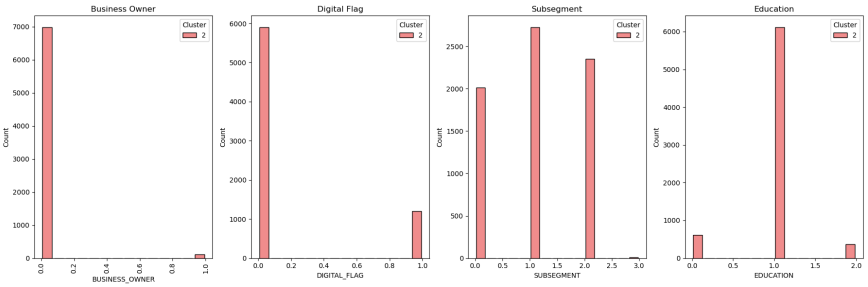


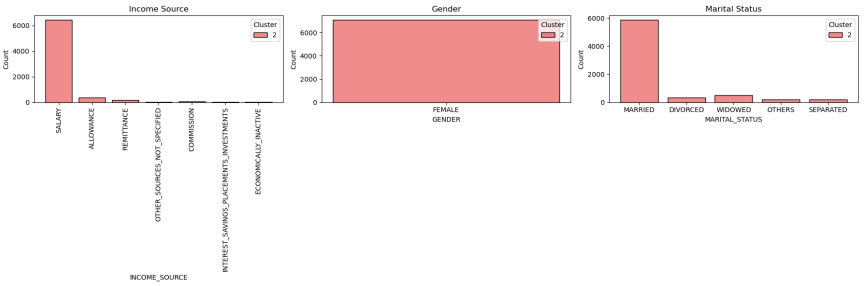
- ## Cluster 1



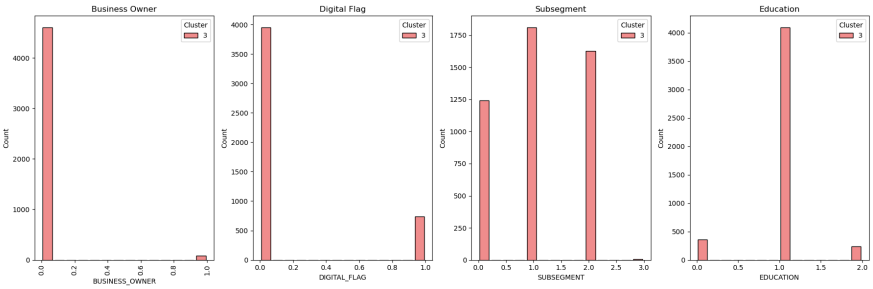


- ## Cluster 2

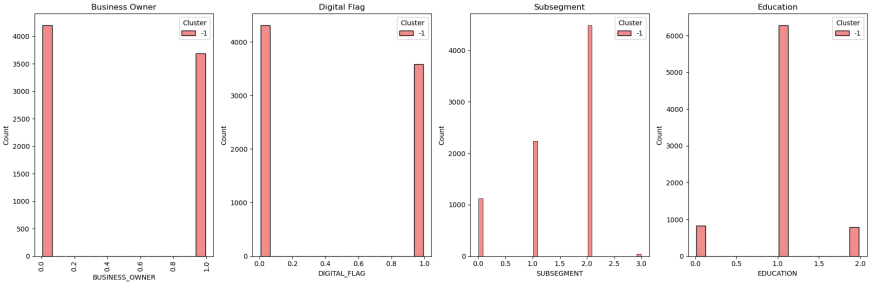


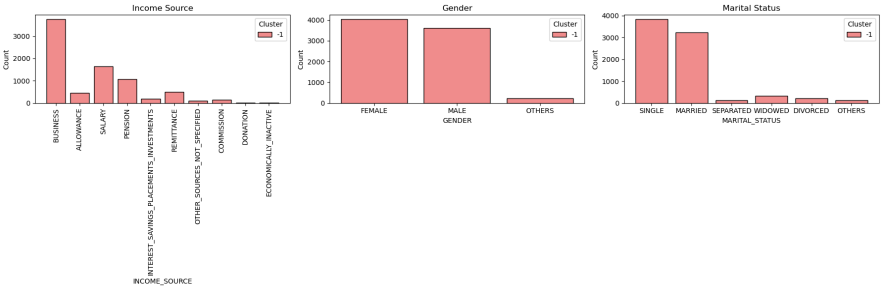


Cluster 3



Cluster -1




Income Source
Cluster
-1
Count
0
1000
2000
3000
BUSINESS
ALLOWANCE
SALARY
PENSION
INTEREST_SAVINGS_PLACEMENTS_INVESTMENTS
REMITTANCE
OTHER_SOURCES_NOT_SPECIFIED
COMMISSION
DONATION
ECONOMICALLY_INACTIVE
INCOME_SOURCE
Gender
Cluster
-1
Count
0
1000
2000
3000
4000
FEMALE
MALE
OTHERS
GENDER
Marital Status
Cluster
-1
Count
0
1000
2000
3000
4000
SINGLE
MARRIED
SEPARATED
WIDOWED
DIVORCED
OTHERS
MARITAL_STATUS

○