# Retail store Clustering

Benjamin Lee

# Data

```
data.head()
```

| | Population | AverageStoreOpenHours | Average Sales Conversion | rainingDaysPerMonth |
|---|---|---|---|---|
| 0 | 30720 | 9.000000 | 24.275787 | 25 |
| 1 | 1835 | 8.428571 | 21.696321 | 25 |
| 2 | 59041 | 9.285714 | 106.997143 | 21 |
| 3 | 46148 | 9.857143 | 21.797819 | 22 |
| 4 | 8730 | 8.714286 | 20.457215 | 23 |

# Scaled data

```
scaler = StandardScaler()
scaled_features = scaler.fit_transform(data)
```

```
scaled_features
```

```
array([[ 0.15867895, -0.24932989, -0.50517591,  1.2123422 ],
       [-0.39221696, -0.85268559, -0.60005917,  1.2123422 ],
       [ 0.69881824,  0.05234796,  2.53765183, -0.02102328],
       [ 0.45292241,  0.65570366, -0.59632566,  0.28731809],
       [-0.26071524, -0.55100774, -0.64563854,  0.59565946],
       [ 0.22991303,  0.12776743,  3.82829989,  0.28731809],
       [ 0.87086712, -1.38062182, -0.40939717, -0.94604738],
       [ 0.1216029 , -0.24932989, -0.43635359, -0.32936464],
       [ 2.39588662, -1.2297829 ,  1.32739791,  0.59565946],
       [-0.22382991, -0.24932989, -0.62641185,  1.2123422 ],
       [-0.34097039, -0.24932989, -0.22990242, -0.02102328],
       [-0.34890436,  1.6738664 , -0.62875384,  0.90400083],
       [-0.41802144, -0.70184666, -0.47203853, -0.32936464],
       [-0.39429582, -1.15436344, -0.29392028,  0.59565946],
       [-0.37182895, -0.70184666, -0.52264999, -0.02102328],
       [-0.22920823,  0.65570366, -0.30817681, -0.02102328],
       [-0.36229293, -1.45604129,  0.64106835, -0.02102328],
       [ 0.04283537, -1.60688021,  0.22212769, -2.48775422],
       [-0.33252148, -1.15436344, -0.23169958,  0.59565946],
       [-0.0688505 , -1.90855806,  0.19897838, -1.87107149],
       [-0.28091253, -0.24932989,  2.63047969, -0.02102328],
       [-0.41693433,  2.31493184,  0.45805332,  0.90400083],
       [-0.36660321,  0.65570366, -0.76765917, -0.32936464],
       [-0.35912697, -0.24932989, -0.59009754,  1.2123422 ],
       [ 5.74013094, -0.09849096, -0.38020642,  0.28731809],
...
       [-0.25297199,  0.80654259,  0.2344226 , -1.87107149],
       [-0.13619388,  0.80654259, -0.74061954,  0.28731809],
       [-0.38458815, -1.07894397, -0.95610193,  1.2123422 ],
       [-0.30889122,  1.56073721,  0.14793299, -0.32936464],
       [-0.1960038 ,  0.80654259, -0.07861423, -0.02102328]])
```

# PCA

```
pca = PCA()
pca.fit(scaled_features)
```

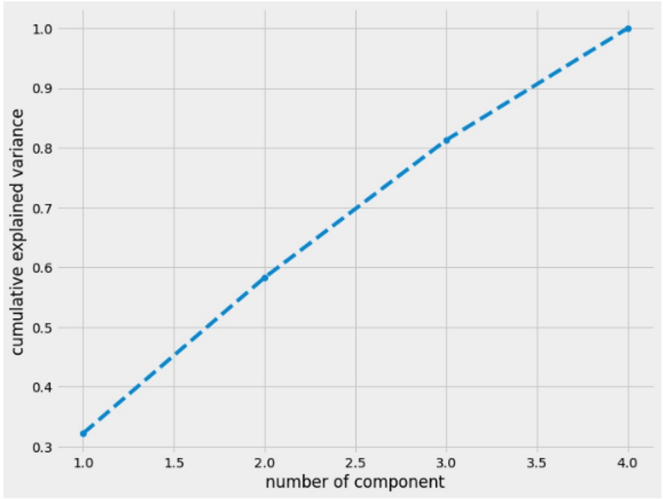
```
▾ PCA
PCA()
```

```
pca.explained_variance_ratio_
```

```
array([0.32212678, 0.26076482, 0.22988476, 0.18722364])
```

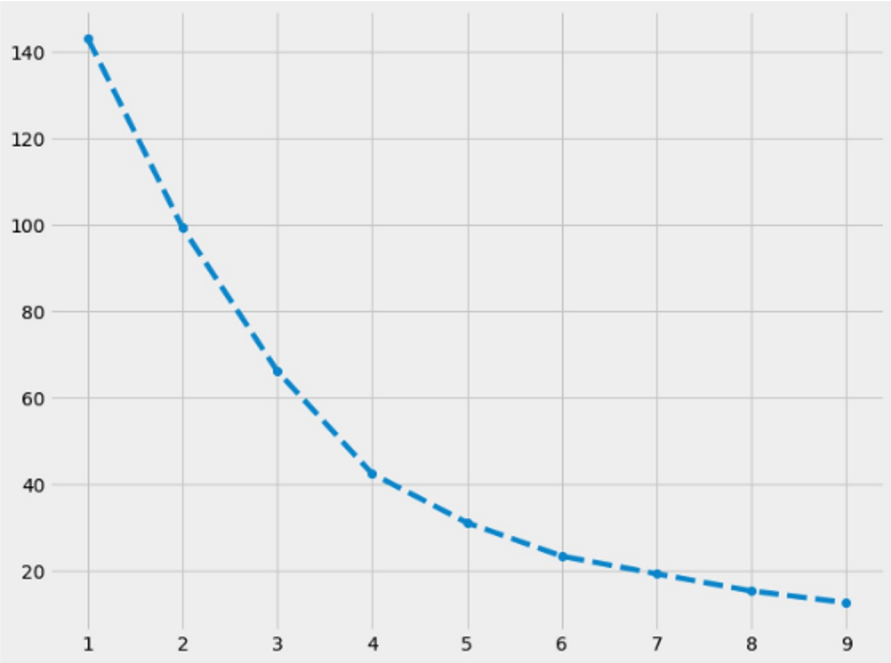
```
plt.figure(figsize = (10,8))
plt.plot(range(1,5),pca.explained_variance_ratio_.cumsum(), marker = 'o', linestyle = '--')
plt.xlabel('number of component')
plt.ylabel('cumulative explained variance')
```

```
Text(0, 0.5, 'cumulative explained variance')
```




N_component = 3

# Elbow method to find optimal number of cluster

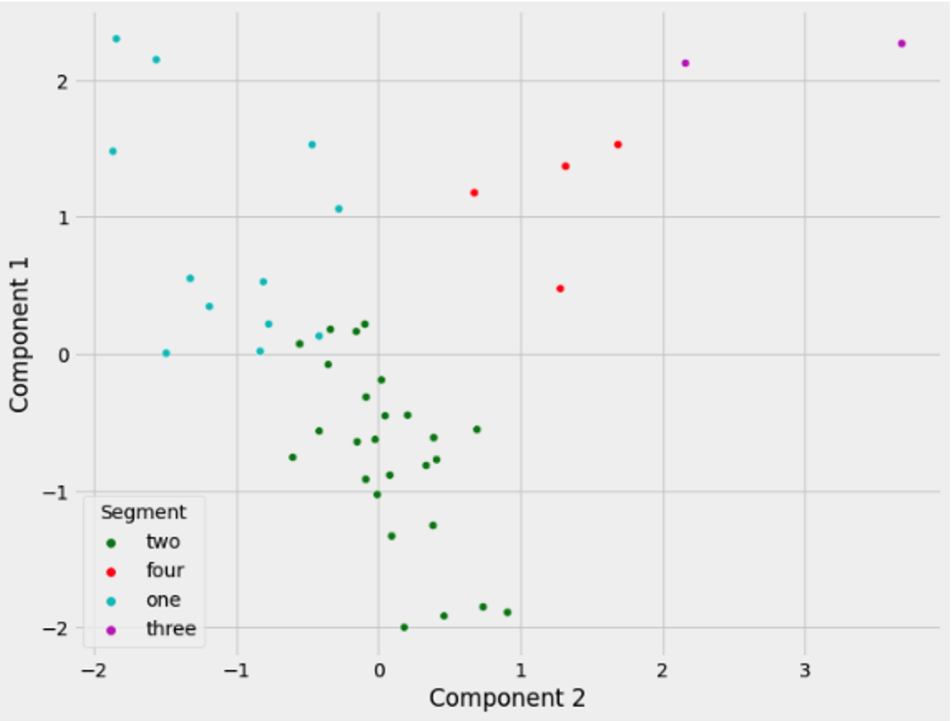



N_cluster = 4

# Cluster

| | Population | AverageStoreOpenHours | Average Sales Conversion | rainingDaysPerMonth | Component 1 | Component 2 | Component 3 | K-means PCA |
|---|---|---|---|---|---|---|---|---|
| 0 | 30720 | 9.000000 | 24.275787 | 25 | -0.552790 | 0.693042 | -0.715533 | 1 |
| 1 | 1835 | 8.428571 | 21.696321 | 25 | -0.447385 | 0.204778 | -0.772393 | 1 |
| 2 | 59041 | 9.285714 | 106.997143 | 21 | 1.369899 | 1.315872 | 1.820918 | 3 |
| 3 | 46148 | 9.857143 | 21.797819 | 22 | -0.612059 | 0.388543 | -0.488150 | 1 |
| 4 | 8730 | 8.714286 | 20.457215 | 23 | -0.315529 | -0.087128 | -0.686904 | 1 |
| 5 | 34455 | 9.357143 | 142.084294 | 22 | 1.527921 | 1.684903 | 3.076155 | 3 |
| 6 | 68062 | 7.928571 | 26.879598 | 18 | 1.527269 | -0.466311 | -1.081291 | 0 |
| 7 | 28776 | 9.000000 | 26.146769 | 20 | 0.178524 | -0.338266 | -0.460825 | 1 |
| 8 | 148023 | 8.071429 | 74.095561 | 23 | 2.122419 | 2.159049 | -0.422412 | 2 |
| 9 | 10664 | 9.000000 | 20.979905 | 25 | -0.771385 | 0.408736 | -0.653443 | 1 |
| 10 | 4522 | 9.000000 | 31.759286 | 21 | -0.077852 | -0.353610 | -0.130486 | 1 |
| 11 | 4106 | 10.821429 | 20.916237 | 24 | -1.913663 | 0.461444 | 0.110314 | 1 |
| 12 | 482 | 8.571429 | 25.176650 | 20 | 0.217747 | -0.772312 | -0.420525 | 0 |
| 13 | 1726 | 8.142857 | 30.018916 | 23 | 0.163293 | -0.156468 | -0.551822 | 1 |
| 14 | 2904 | 8.571429 | 23.800743 | 21 | 0.072551 | -0.554228 | -0.518355 | 1 |
| 15 | 10382 | 9.857143 | 29.631343 | 21 | -0.641798 | -0.150303 | 0.075662 | 1 |
| 16 | 3404 | 7.857143 | 55.437222 | 21 | 1.058753 | -0.278562 | 0.173807 | 0 |
| 17 | 24646 | 7.714286 | 44.048034 | 13 | 2.300483 | -1.844001 | -0.105773 | 0 |
| 18 | 4965 | 8.142857 | 31.710429 | 23 | 0.216900 | -0.095801 | -0.526622 | 1 |
| 19 | 18790 | 7.428571 | 43.418704 | 15 | 2.147753 | -1.563316 | -0.256447 | 0 |
| 20 | 7671 | 9.000000 | 109.520732 | 21 | 1.176154 | 0.674024 | 2.207102 | 3 |
| 21 | 539 | 11.428571 | 50.461833 | 24 | -1.886891 | 0.907363 | 1.262424 | 1 |
| 22 | 3178 | 9.857143 | 17.140000 | 20 | -0.755279 | -0.602646 | -0.209230 | 1 |
| 23 | 3570 | 9.000000 | 21.967135 | 25 | -0.814707 | 0.335573 | -0.566054 | 1 |
| 24 | 323371 | 9.142857 | 27.673168 | 22 | 2.265383 | 3.680389 | -2.817440 | 2 |
| 25 | 17 | 11.000000 | 16.371927 | 23 | -1.996419 | 0.181120 | 0.108017 | 1 |
| 26 | 31016 | 9.000000 | 19.748723 | 22 | -0.191066 | 0.020033 | -0.746224 | 1 |
| 27 | 374 | 10.711429 | 31.423410 | 25 | -1.848023 | 0.735904 | 0.382540 | 1 |
| 28 | 2199 | 10.000000 | 28.651676 | 24 | -1.252472 | 0.383915 | 0.055851 | 1 |
| 29 | 3713 | 10.000000 | 25.616316 | 15 | 0.004418 | -1.493343 | 0.279187 | 0 |
| 30 | 2041 | 9.142857 | 43.275915 | 20 | 0.130474 | -0.416762 | 0.329020 | 0 |
| 31 | 7345 | 9.428571 | 17.852459 | 21 | -0.563502 | -0.416686 | -0.416823 | 1 |
| 32 | 3001 | 9.571429 | 35.897727 | 18 | 0.019606 | -0.831919 | 0.329000 | 0 |
| 33 | 13259 | 10.000000 | 46.069663 | 14 | 0.550234 | -1.324025 | 0.859936 | 0 |
| 34 | 5444 | 8.142857 | 23.777085 | 20 | 0.526359 | -0.809947 | -0.662306 | 0 |
| 35 | 5922 | 8.285714 | 35.408158 | 14 | 1.478979 | -1.867510 | -0.041135 | 0 |
| 36 | 5052 | 8.857143 | 96.812051 | 25 | 0.475905 | 1.279351 | 1.643354 | 3 |
| 37 | 247 | 10.142857 | 21.769444 | 23 | -1.330272 | 0.093015 | -0.048038 | 1 |
| 38 | 10304 | 10.000000 | 29.467990 | 22 | -0.885110 | 0.079597 | 0.087940 | 1 |
| 39 | 9136 | 10.000000 | 44.382279 | 15 | 0.345768 | -1.188997 | 0.805588 | 0 |
| 40 | 15259 | 10.000000 | 17.875091 | 22 | -1.027014 | -0.008055 | -0.304437 | 1 |
| 41 | 2235 | 8.214286 | 12.017055 | 25 | -0.452147 | 0.046542 | -1.149323 | 1 |
| 42 | 6204 | 10.714286 | 42.031000 | 20 | -0.916130 | -0.089747 | 0.840912 | 1 |
| 43 | 12123 | 10.000000 | 35.872159 | 21 | -0.625299 | -0.023249 | 0.304264 | 1 |

# Cluster plot




- Purple (cluster 3) -> very high population (ireland, cork)

- Red (cluster 4) -> very high average sales conversion

- Light blue (cluster 1) -> short operating hours (average 7 - 8 hours per day)

- Green (cluster 2) -> relatively long operating hours (average 10 - 11 hours per day)

# Clustering 1

Feature preprocessing:

▼ Features included

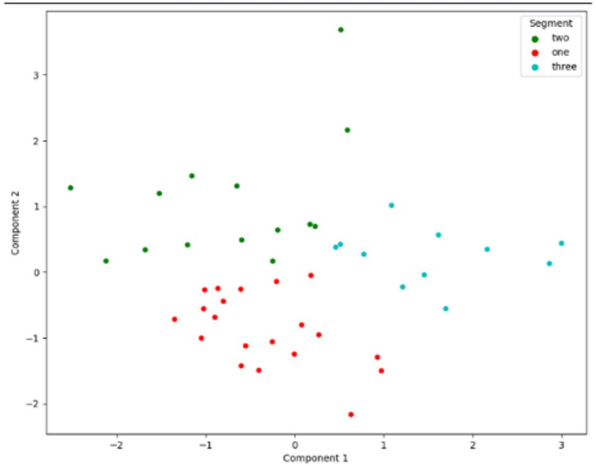
['AvgOhPerDay', 'AvgTcPerday', 'MaletoFemaleRatio', 'KidstoAdultRatio', 'RainingDaysPerMonth']

▶ PCA: 4 components

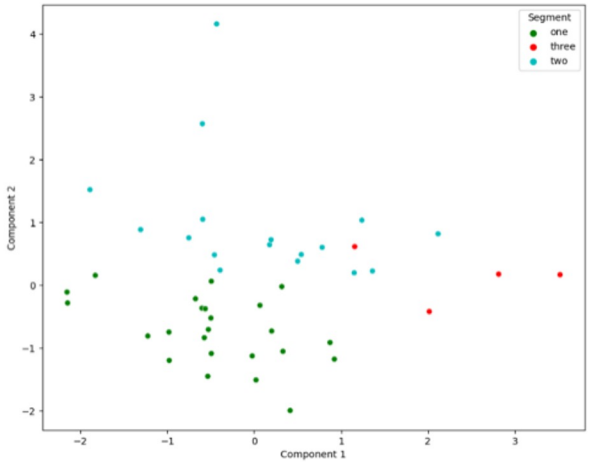
▶ 3 clusters

▶ PCA components as inputs to k-means

## Clustered: First 3 weeks

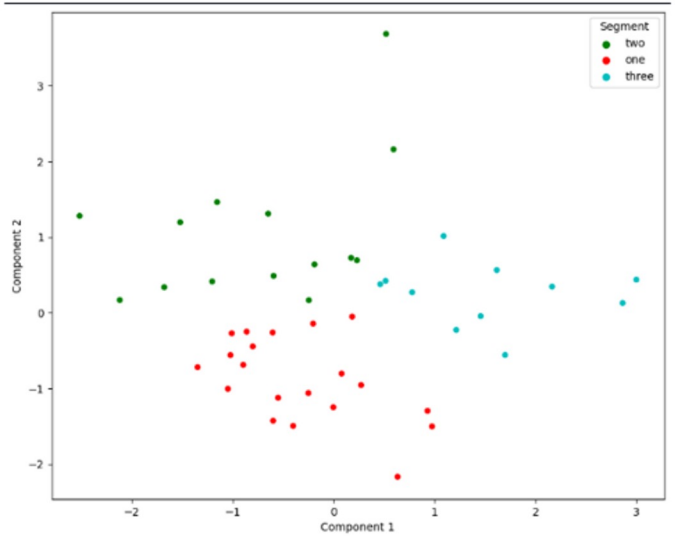



Taking PCA components as inputs
Silhouette coefficient: 0.275875112092234

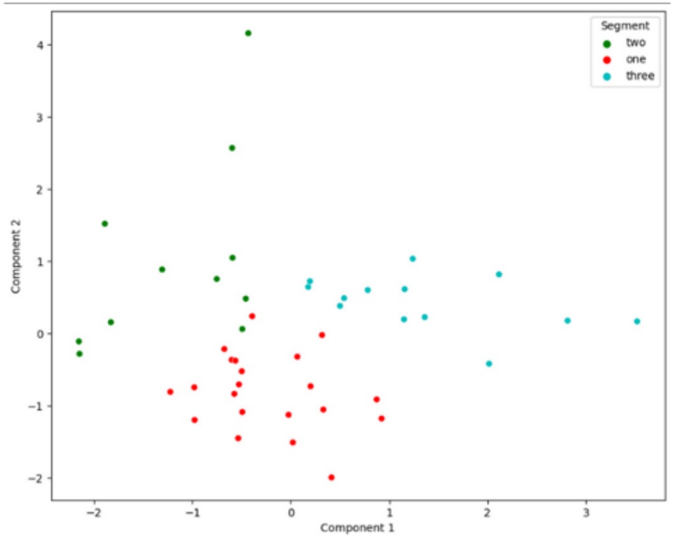



Taking original features as inputs
Silhouette coefficient: 0.23353861209473806

# Validation: 4 weeks data




First 3 weeks data
Silhouette coefficient: 0.275875112092234




Four weeks data
Silhouette coefficient: 0.27257463142079835

## Explained:

▼ Red: Cluster 1

21 counts

Short Operating hour Per Day: 7 - 9 hours

Low Transaction Count Per Day: 10 - 40+

High Male to Female Ratio: 0.90 - 1.00+

| | AvgOhPerDay | AvgTcPerday | MaletoFemaleRatio | KidstoAdultRatio | RainingDaysPerMonth | Component 1 | Component 2 | Component 3 | Component 4 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 8.43 | 37.50000 | 0.947983 | 0.367454 | 7 | 0.200449 | -0.729520 | -0.341850 | 0.584502 | 0 | one |
| | 8.71 | 40.06667 | 0.952583 | 0.323974 | 11 | -0.600757 | -0.364397 | -0.108325 | 0.029915 | 0 | one |
| | 9.36 | 34.40000 | 0.950467 | 0.382090 | 10 | -0.391940 | 0.239679 | -0.354022 | -0.122944 | 0 | one |
| | 7.93 | 32.30000 | 0.903300 | 0.273412 | 10 | -0.977528 | -1.196117 | 0.184548 | 0.536903 | 0 | one |
| | 9.00 | 28.43333 | 0.964098 | 0.277561 | 10 | -0.497580 | -0.522473 | -0.209963 | -0.485941 | 0 | one |
| | 8.07 | 29.33333 | 0.936891 | 0.331919 | 12 | -1.224305 | -0.808066 | -0.350469 | 0.352282 | 0 | one |
| | 9.00 | 32.16667 | 0.942795 | 0.295605 | 10 | -0.562908 | -0.374179 | -0.044918 | -0.242825 | 0 | one |
| 1 | 9.00 | 33.36667 | 0.895222 | 0.359150 | 6 | 0.066283 | -0.320349 | 0.081681 | 0.523380 | 0 | one |
| 1 | 8.57 | 31.53333 | 0.991736 | 0.305031 | 11 | -0.527414 | -0.704091 | -0.526283 | -0.295704 | 0 | one |
| 1 | 8.14 | 35.03333 | 1.023447 | 0.445854 | 6 | 0.872644 | -0.911430 | -1.317820 | 0.659314 | 0 | one |
| 1 | 8.57 | 29.16667 | 0.899281 | 0.267454 | 10 | -0.980695 | -0.745028 | 0.265987 | 0.116990 | 0 | one |
| 1 | 7.86 | 23.66667 | 0.974478 | 0.325517 | 7 | 0.020726 | -1.507090 | -0.749005 | 0.431526 | 0 | one |
| 1 | 7.71 | 35.26667 | 0.964764 | 0.279827 | 10 | -0.534301 | -1.448725 | -0.267876 | 0.370476 | 0 | one |
| 1 | 8.14 | 33.06667 | 0.967895 | 0.294951 | 10 | -0.492939 | -1.085214 | -0.336692 | 0.131022 | 0 | one |
| 1 | 7.43 | 31.86667 | 0.985838 | 0.293499 | 6 | 0.414003 | -1.992121 | -0.588679 | 0.643441 | 0 | one |
| 2 | 9.14 | 29.10000 | 0.938152 | 0.442315 | 6 | 0.317915 | -0.021334 | -0.648332 | 0.476256 | 0 | one |
| 2 | 9.00 | 25.80000 | 0.941777 | 0.350297 | 10 | -0.675523 | -0.215437 | -0.397347 | -0.090854 | 0 | one |
| 3 | 8.14 | 28.26667 | 0.966053 | 0.369479 | 10 | -0.574471 | -0.834486 | -0.730334 | 0.387524 | 0 | one |
| 3 | 8.29 | 28.46429 | 0.982591 | 0.378808 | 4 | 0.922475 | -1.175787 | -0.855818 | 0.563822 | 0 | one |
| 3 | 8.86 | 11.58621 | 0.901393 | 0.309124 | 3 | 0.330859 | -1.051909 | -0.243693 | 0.265237 | 0 | one |
| 4 | 8.21 | 16.06667 | 0.946864 | 0.375902 | 6 | -0.022352 | -1.124905 | -0.859168 | 0.551719 | 0 | one |

▼ Green: Cluster 2

11 counts

Long Operating hour Per Day: 9 - 11 hours

Low Transaction Count Per Day: 10 - 40+

Low Male to Female Ratio: 0.80 - 0.95

| | AvgOhPerDay | AvgTcPerday | MaletoFemaleRatio | KidstoAdultRatio | RainingDaysPerMonth | Component 1 | Component 2 | Component 3 | Component 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 9.00 | 56.83333 | 0.914854 | 0.317935 | 11 | -0.492932 | 0.062025 | 0.599973 | 0.255363 | 1 | two |
| 23 | 9.00 | 27.73913 | 0.866179 | 0.340281 | 13 | -1.829077 | 0.156283 | 0.243182 | 0.172809 | 1 | two |
| 25 | 11.00 | 24.13333 | 0.888889 | 0.272727 | 9 | -0.591840 | 1.047710 | 0.573994 | -1.255747 | 1 | two |
| 28 | 10.00 | 23.86667 | 0.959893 | 0.346324 | 10 | -0.457333 | 0.482484 | -0.409849 | -0.832162 | 1 | two |
| 29 | 10.00 | 41.43333 | 0.933854 | 0.323232 | 12 | -0.752846 | 0.754163 | 0.252326 | -0.669432 | 1 | two |
| 30 | 9.14 | 24.66667 | 0.832136 | 0.186799 | 13 | -2.147521 | -0.281752 | 1.077885 | -0.373859 | 1 | two |
| 32 | 9.57 | 47.80000 | 0.593733 | 0.059045 | 8 | -2.154142 | -0.108726 | 3.941811 | 0.790935 | 1 | two |
| 33 | 10.00 | 58.10000 | 0.827315 | 1.145080 | 10 | -0.432426 | 4.153853 | -1.949227 | 3.610009 | 1 | two |
| 37 | 10.14 | 37.20000 | 0.871212 | 0.408163 | 15 | -1.889532 | 1.519225 | 0.296905 | -0.252598 | 1 | two |
| 42 | 10.71 | 13.73333 | 0.885714 | 0.270154 | 11 | -1.307271 | 0.884712 | 0.338901 | -1.307631 | 1 | two |
| 44 | 11.00 | 46.53333 | 0.910814 | 0.569019 | 12 | -0.595886 | 2.567772 | -0.322568 | -0.067754 | 1 | two |

▼ Light Blue: Cluster 3

13 counts

Long Operating Hour Per Day:  9 - 11 hours

High Transaction Count Per Day: 35 - 100+

High Male to Female Ratio: 0.90 - 1.0+

| | AvgOhPerDay | AvgTcPerday | MaletoFemaleRatio | KidstoAdultRatio | RainingDaysPerMonth | Component 1 | Component 2 | Component 3 | Component 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 9.29 | 50.46667 | 0.994696 | 0.373009 | 2 | 2.015648 | -0.418298 | -0.299058 | 0.246245 | 2 | three |
| 3 | 9.86 | 46.83333 | 0.976529 | 0.373357 | 8 | 0.541548 | 0.488275 | -0.168451 | -0.351567 | 2 | three |
| 1 | 10.82 | 23.63333 | 1.022660 | 0.460483 | 5 | 1.239655 | 1.034023 | -1.219300 | -0.957232 | 2 | three |
| 15 | 9.86 | 36.36667 | 1.062786 | 0.389642 | 7 | 1.150923 | 0.196903 | -1.100114 | -0.862204 | 2 | three |
| 20 | 9.00 | 167.30000 | 0.926903 | 0.263792 | 3 | 3.519790 | 0.167640 | 3.115774 | 1.725961 | 2 | three |
| 21 | 11.43 | 47.63333 | 1.081081 | 0.262431 | 5 | 2.116296 | 0.817552 | -0.240104 | -2.179720 | 2 | three |
| 22 | 9.86 | 108.17390 | 1.008850 | 0.253253 | 4 | 2.813343 | 0.176890 | 1.404712 | -0.075387 | 2 | three |
| 27 | 10.71 | 26.76923 | 1.021622 | 0.323529 | 7 | 0.781993 | 0.602099 | -0.604419 | -1.513053 | 2 | three |
| 31 | 9.43 | 56.50000 | 0.975524 | 0.428305 | 5 | 1.360605 | 0.226148 | -0.236573 | 0.411966 | 2 | three |
| 38 | 10.00 | 54.36667 | 0.916310 | 0.287126 | 7 | 0.500080 | 0.382191 | 0.815787 | -0.289612 | 2 | three |
| 39 | 10.00 | 97.23333 | 0.926613 | 0.237496 | 8 | 1.156556 | 0.613937 | 1.860643 | -0.088716 | 2 | three |
| 40 | 10.00 | 54.03333 | 1.035891 | 0.324129 | 12 | 0.177651 | 0.644273 | -0.240204 | -1.110503 | 2 | three |
| 43 | 10.00 | 62.93333 | 0.945907 | 0.313690 | 10 | 0.194074 | 0.723597 | 0.666347 | -0.412152 | 2 | three |

# Clustering 2

Feature preprocessing:

▼ Features included

['AvgOhPerDay', 'AvgTcPerDay', 'AvgAmtPerTran', 'MaletoFemaleRatio']

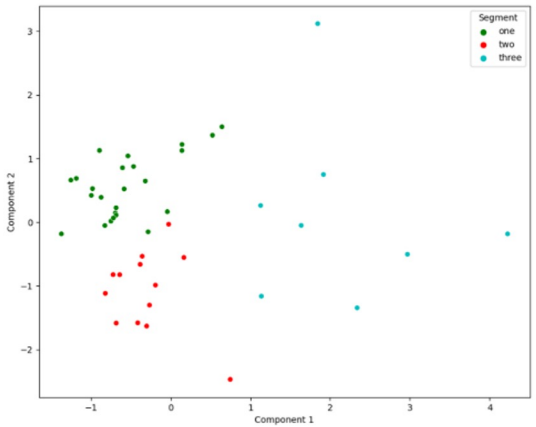
- Average amount per transaction might depict some picture of wealth index?
- Did not directly take AverageSalesConversion, instead take AverageTcPerDay and AvgAmtPerTran
- Retain MaletoFemaleRatio, Regatta clothing's product mainly is on outdoor garments, male is more active ?

▶ PCA: 4 components

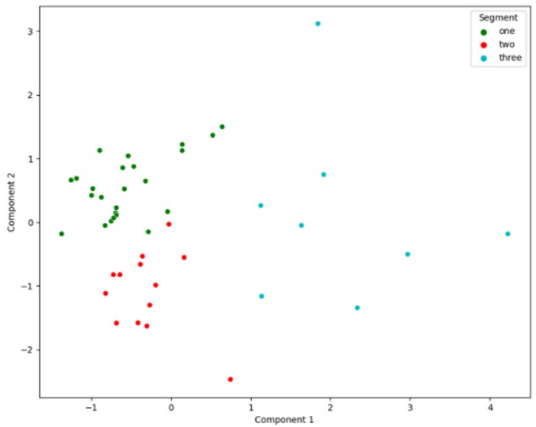
▶ 3 clusters

▶ PCA components as inputs to k-means

## Clustered: First 3 weeks

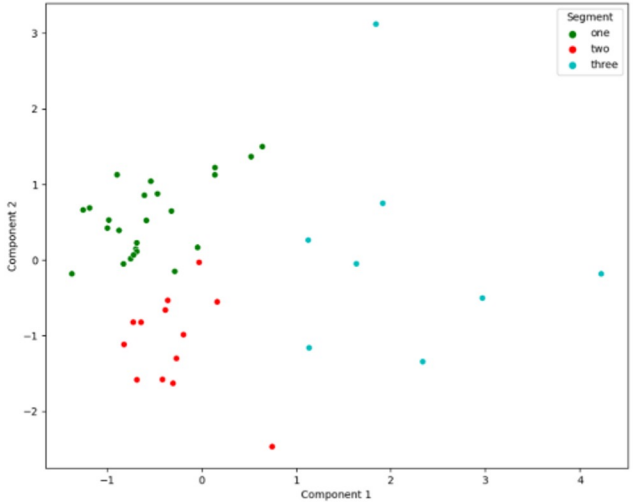



Silhouette Coefficient: 0.3608930164363887
Taking PCA components as inputs

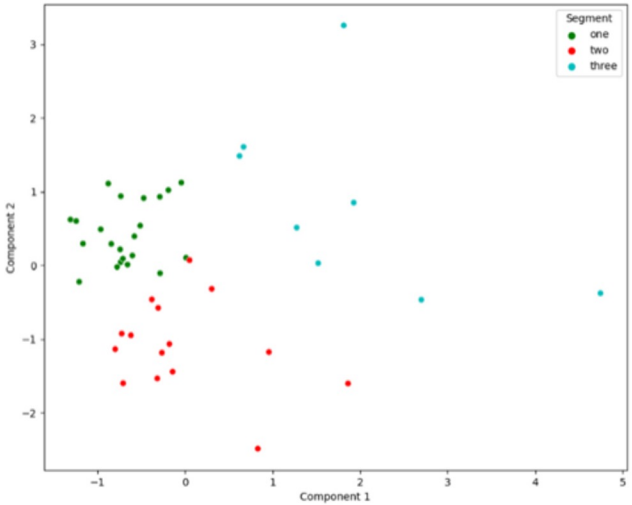



Silhouette Coefficient: 0.3608930164363887
Taking original features as input, output is exactly the same as left one

# Validation: Four weeks data




First 3 weeks of data
Silhouette Coefficient: 0.3608930164363887




Four weeks of data
Silhouette Coefficient: 0.36024408841137945

## Explained:

- Green: Cluster 1
  - Low Average Transaction Count Per Day: 16 - 50+
  - Relatively Low Average Operating Hour Per day: 7.0 - 9.0
  - Relatively High Average Amount Per Transaction: 300 - 400+
  - High Male to Female Ratio: 0.9 - 1.0+

| | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | MaletoFemaleRatio | Component 1 | Component 2 | Component 3 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 9.00 | 56.83333 | 382.8841 | 0.914854 | 0.013327 | 0.102391 | -0.052305 | 0 | one |
| 1 | 8.43 | 37.50000 | 393.9200 | 0.947983 | -0.840541 | 0.290407 | -0.448818 | 0 | one |
| 4 | 8.71 | 40.06667 | 398.2033 | 0.952583 | -0.733539 | 0.043074 | -0.288598 | 0 | one |
| 5 | 9.36 | 34.40000 | 348.6317 | 0.950467 | -0.283783 | -0.106897 | 0.130159 | 0 | one |
| 6 | 7.93 | 32.30000 | 380.6533 | 0.903300 | -0.874612 | 1.108883 | -0.485373 | 0 | one |
| 7 | 9.00 | 28.43333 | 356.7680 | 0.964098 | -0.655295 | 0.010975 | -0.140292 | 0 | one |
| 8 | 8.07 | 29.33333 | 419.9037 | 0.936891 | -1.306761 | 0.620229 | -0.461245 | 0 | one |
| 9 | 9.00 | 32.16667 | 371.9291 | 0.942795 | -0.599951 | 0.132736 | 0.013256 | 0 | one |
| 10 | 9.00 | 33.36667 | 386.9057 | 0.895222 | -0.509982 | 0.537519 | 0.383840 | 0 | one |
| 12 | 8.57 | 31.53333 | 350.0329 | 0.991736 | -0.773573 | -0.022279 | -0.719009 | 0 | one |
| 13 | 8.14 | 35.03333 | 385.7702 | 1.023447 | -1.207291 | -0.222922 | -1.199601 | 0 | one |
| 14 | 8.57 | 29.16667 | 346.0510 | 0.899281 | -0.469856 | 0.911954 | -0.044744 | 0 | one |
| 16 | 7.86 | 23.66667 | 364.2565 | 0.974478 | -1.240867 | 0.601030 | -0.991109 | 0 | one |
| 17 | 7.71 | 35.26667 | 269.7790 | 0.964764 | -0.286369 | 0.930141 | -1.468737 | 0 | one |
| 18 | 8.14 | 33.06667 | 400.5496 | 0.967895 | -1.163420 | 0.294426 | -0.734530 | 0 | one |
| 19 | 7.43 | 31.86667 | 201.8838 | 0.985838 | -0.040499 | 1.122520 | -1.999232 | 0 | one |
| 23 | 9.00 | 27.73913 | 338.6968 | 0.866179 | -0.188781 | 1.020907 | 0.508853 | 0 | one |
| 24 | 9.14 | 29.10000 | 385.1471 | 0.938152 | -0.707123 | 0.091731 | 0.229822 | 0 | one |
| 26 | 9.00 | 25.80000 | 370.5473 | 0.941777 | -0.739301 | 0.216190 | 0.093240 | 0 | one |
| 34 | 8.14 | 28.26667 | 358.0334 | 0.966053 | -0.961466 | 0.488378 | -0.795431 | 0 | one |
| 35 | 8.29 | 28.46429 | 304.7540 | 0.982591 | -0.576703 | 0.393830 | -0.968179 | 0 | one |
| 41 | 8.21 | 16.06667 | 299.0389 | 0.946864 | -0.731543 | 0.938799 | -0.641864 | 0 | one |

- Light blue: Cluster 2
  - Relatively long Average Operating Hour Per Day: 9 - 10+
  - Low Average Amount Per Transaction: 100 - 200+
  - Low Male to Female Ratio: 0.8 - 0.90+

| | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | MaletoFemaleRatio | Component 1 | Component 2 | Component 3 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|
| 20 | 9.00 | 167.30000 | 93.1460 | 0.926903 | 4.744384 | -0.378918 | -2.374815 | 1 | two |
| 30 | 9.14 | 24.66667 | 234.1572 | 0.832136 | 0.670080 | 1.607567 | 0.568871 | 1 | two |
| 32 | 9.57 | 47.80000 | 290.3248 | 0.593733 | 1.813624 | 3.254834 | 2.483776 | 1 | two |
| 33 | 10.00 | 58.10000 | 213.6329 | 0.827315 | 1.927272 | 0.850497 | 0.782068 | 1 | two |
| 36 | 8.86 | 11.58621 | 150.6857 | 0.901393 | 0.624247 | 1.483652 | -0.235802 | 1 | two |
| 38 | 10.00 | 54.36667 | 213.2482 | 0.916310 | 1.521193 | 0.029172 | 0.192668 | 1 | two |
| 39 | 10.00 | 97.23333 | 188.0632 | 0.926613 | 2.699162 | -0.465893 | -0.477566 | 1 | two |
| 42 | 10.71 | 13.73333 | 160.3067 | 0.885714 | 1.276800 | 0.512536 | 1.271871 | 1 | two |

- Red: Cluster 3
  - Long Operating Hour Per Day: 9 - 10+
  - High Male to Female Ratio: 0.9 - 1.0+
  - Relatively High Average Amount Per Transaction: 300 - 400+

| | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | MaletoFemaleRatio | Component 1 | Component 2 | Component 3 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 9.29 | 50.46667 | 436.4372 | 0.994696 | -0.721106 | -0.926230 | -0.158891 | 2 | three |
| 3 | 9.86 | 46.83333 | 469.5500 | 0.976529 | -0.795881 | -1.136824 | 0.546181 | 2 | three |
| 11 | 10.82 | 23.63333 | 348.4369 | 1.022660 | -0.314763 | -1.534771 | 0.844718 | 2 | three |
| 15 | 9.86 | 36.36667 | 380.4095 | 1.062786 | -0.705296 | -1.599925 | -0.218014 | 2 | three |
| 21 | 11.43 | 47.63333 | 269.5119 | 1.081081 | 0.832582 | -2.487553 | 0.352384 | 2 | three |
| 22 | 9.86 | 108.17390 | 291.7931 | 1.008850 | 1.862613 | -1.603638 | -0.978295 | 2 | three |
| 25 | 11.00 | 24.13333 | 338.8502 | 0.888889 | 0.306848 | -0.319080 | 1.896388 | 2 | three |
| 27 | 10.71 | 26.76923 | 330.3883 | 1.021622 | -0.140272 | -1.442751 | 0.675314 | 2 | three |
| 28 | 10.00 | 23.86667 | 351.2724 | 0.959893 | -0.377416 | -0.461689 | 0.680343 | 2 | three |
| 29 | 10.00 | 41.43333 | 411.8261 | 0.933854 | -0.304236 | -0.577135 | 0.841248 | 2 | three |
| 31 | 9.43 | 56.50000 | 457.6493 | 0.975524 | -0.615786 | -0.948982 | 0.075808 | 2 | three |
| 37 | 10.14 | 37.20000 | 386.0250 | 0.871212 | 0.053720 | 0.070656 | 1.363485 | 2 | three |
| 40 | 10.00 | 54.03333 | 230.3441 | 1.035891 | 0.959426 | -1.176806 | -0.601237 | 2 | three |
| 43 | 10.00 | 62.93333 | 459.2811 | 0.945907 | -0.177313 | -1.066950 | 0.642768 | 2 | three |
| 44 | 11.00 | 46.53333 | 479.6039 | 0.910814 | -0.261955 | -1.185791 | 1.906624 | 2 | three |

# Clustering 3

Feature preprocessing:

▼ Features included

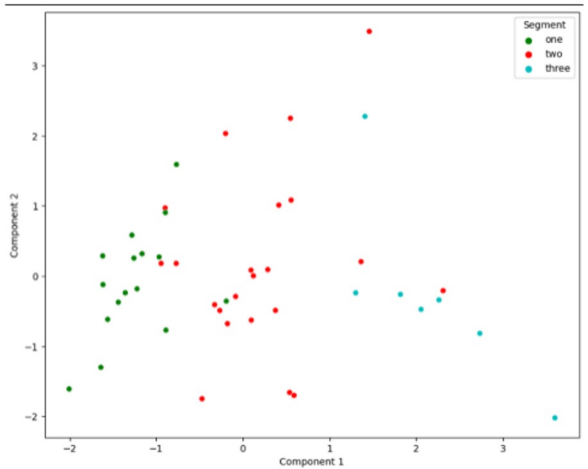
['Population_lg', 'IsCity', 'AvgOhPerDay', 'AvgTcPerday', 'AvgAmtPerTran', 'MaletoFemaleRatio']

- Take log(10) of population to bring closer the scale
- IsCity returns 0 or 1

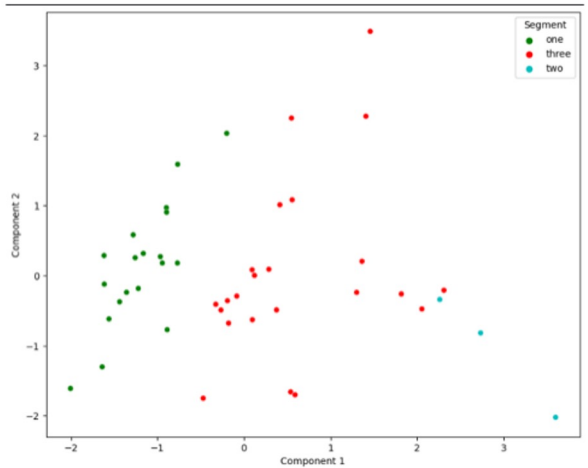
▶ PCA: 4 components

▶ 3 clusters

## Clustered: First 3 Weeks




Using PCA component as input
Silhouette coefficient: 0.2395744738250792




Using original features as input
Silhouette coefficient: 0.22709149191674818

## Explained (input from original features):

- Green: Cluster 1
  - IsCity = 1
  - High Average Amount Per Transaction: 300 - 400+

| | Population_Ig | IsCity | vgOhPerDay | AvgTcPerday | AvgAmtPerTran | MaletoFemaleRatio | Component 1 | Component 2 | Component 3 | Component 4 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 4.487421 | 1 | 9.00 | 55.95238 | 367.0523 | 0.914854 | -0.886372 | -0.772846 | -0.258430 | 0.875171 | 0 | one |
| 1 | 3.263636 | 1 | 8.43 | 34.23810 | 372.0652 | 0.947983 | -1.162274 | 0.317321 | 0.385748 | -0.373395 | 0 | one |
| 2 | 4.771154 | 1 | 9.29 | 48.14286 | 418.2994 | 0.994696 | -1.558956 | -0.618435 | -1.085106 | 0.614186 | 0 | one |
| 3 | 4.664153 | 1 | 9.86 | 45.85714 | 463.8836 | 0.976529 | -1.613378 | -0.123091 | -1.100705 | 1.178597 | 0 | one |
| 4 | 3.941014 | 1 | 8.71 | 37.80952 | 368.3986 | 0.952583 | -1.219912 | -0.183685 | -0.009567 | 0.030853 | 0 | one |
| 7 | 4.459030 | 1 | 9.00 | 27.90476 | 360.1021 | 0.964098 | -1.436763 | -0.373372 | -0.170049 | 0.180417 | 0 | one |
| 8 | 5.170329 | 1 | 8.07 | 31.38095 | 410.7235 | 0.936891 | -2.004344 | -1.611205 | 0.169572 | 0.432720 | 0 | one |
| 11 | 3.613419 | 1 | 10.82 | 22.42857 | 313.6511 | 1.022660 | -0.765401 | 1.590017 | -1.088362 | 0.082459 | 0 | one |
| 12 | 2.683047 | 1 | 8.57 | 32.14286 | 327.4585 | 0.991736 | -0.895676 | 0.970598 | 0.139216 | -1.090598 | 0 | one |
| 13 | 3.237041 | 1 | 8.14 | 34.61905 | 398.5717 | 1.023447 | -1.616270 | 0.286224 | -0.229759 | -1.054851 | 0 | one |
| 14 | 3.462997 | 1 | 8.57 | 29.42857 | 344.0882 | 0.899281 | -0.942531 | 0.179676 | 0.850354 | 0.033208 | 0 | one |
| 15 | 4.016281 | 1 | 9.86 | 35.90476 | 356.6930 | 1.062786 | -1.279262 | 0.582250 | -1.479423 | -0.304539 | 0 | one |
| 24 | 5.509701 | 1 | 9.14 | 29.14286 | 368.8822 | 0.938152 | -1.637853 | -1.303285 | -0.281862 | 0.943569 | 0 | one |
| 29 | 3.569725 | 1 | 10.00 | 42.57143 | 411.0637 | 0.933854 | -0.892770 | 0.905589 | -0.371296 | 0.923314 | 0 | one |
| 31 | 3.865992 | 1 | 9.43 | 52.71429 | 439.7713 | 0.975524 | -1.256635 | 0.255317 | -0.829917 | 0.600597 | 0 | one |
| 34 | 3.735918 | 1 | 8.14 | 28.42857 | 343.1107 | 0.966053 | -1.356108 | -0.239296 | 0.341192 | -0.646991 | 0 | one |
| 37 | 2.392697 | 1 | 10.14 | 36.47619 | 371.1970 | 0.871212 | -0.198589 | 2.032593 | 0.612618 | 0.824753 | 0 | one |
| 41 | 3.349278 | 1 | 8.21 | 14.00000 | 230.2717 | 0.946864 | -0.768572 | 0.179195 | 0.951657 | -1.176312 | 0 | one |
| 43 | 4.083610 | 1 | 10.00 | 60.42857 | 441.3509 | 0.945907 | -0.965773 | 0.272237 | -0.982809 | 1.310082 | 0 | one |

- Light blue: Cluster 2
  - IsCity = 0
  - Low Average Amount Per Transaction: 100+
  - High Average Transaction Count Per Day: 110+

| | Population_lg | IsCity | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | MaletoFemaleRatio | Component 1 | Component 2 | Component 3 | Component 4 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 3.884852 | 0 | 9.00 | 167.3000 | 93.1460 | 0.926903 | 3.602276 | -2.023433 | -1.935871 | 0.152052 | 1 | two |
| 22 | 3.502154 | 0 | 9.86 | 115.3571 | 199.3733 | 1.008850 | 2.263725 | -0.340921 | -2.025436 | -0.275838 | 1 | two |
| 39 | 3.960756 | 0 | 10.00 | 112.2857 | 137.4153 | 0.926613 | 2.736150 | -0.818386 | -1.320990 | 0.397299 | 1 | two |

- Red: Cluster 3
  - IsCity = 0
  - Can't find obvious characteristics, perhaps that's the reason why clustering is not very obvious

| | Population_lg | IsCity | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | MaletoFemaleRatio | Component 1 | Component 2 | Component 3 | Component 4 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 4.537252 | 0 | 9.36 | 35.95238 | 328.5731 | 0.950467 | 0.098100 | -0.629577 | -0.104711 | 0.163384 | 2 | three |
| 6 | 4.832905 | 0 | 7.93 | 34.14286 | 374.5494 | 0.903300 | -0.470092 | -1.750586 | 0.833806 | 0.081713 | 2 | three |
| 9 | 4.027920 | 0 | 9.00 | 31.42857 | 362.3668 | 0.942795 | -0.082086 | -0.292808 | 0.305080 | -0.080322 | 2 | three |
| 10 | 3.655331 | 0 | 9.00 | 31.80952 | 373.5470 | 0.895222 | 0.122166 | 0.002741 | 0.834346 | 0.207485 | 2 | three |
| 16 | 3.531990 | 0 | 7.86 | 23.14286 | 337.3016 | 0.974478 | -0.324681 | -0.409791 | 0.764101 | -1.318091 | 2 | three |
| 17 | 4.391748 | 0 | 7.71 | 38.66667 | 179.7334 | 0.964764 | 0.540916 | -1.661517 | 0.505864 | -1.440311 | 2 | three |
| 18 | 3.695919 | 0 | 8.14 | 33.52381 | 363.9960 | 0.967895 | -0.264780 | -0.491328 | 0.453888 | -0.852818 | 2 | three |
| 19 | 4.273927 | 0 | 7.43 | 35.85714 | 143.3167 | 0.985838 | 0.591570 | -1.701871 | 0.528567 | -1.977118 | 2 | three |
| 21 | 2.731589 | 0 | 11.43 | 40.23810 | 210.3304 | 1.081081 | 1.410067 | 2.276219 | -1.691462 | -0.983227 | 2 | three |
| 23 | 3.552668 | 0 | 9.00 | 27.28571 | 353.1899 | 0.866179 | 0.288486 | 0.091453 | 1.236146 | 0.286303 | 2 | three |
| 25 | 1.230449 | 0 | 11.00 | 24.14286 | 326.7199 | 0.888889 | 1.459572 | 3.487369 | 0.905844 | 0.074470 | 2 | three |
| 26 | 4.491586 | 0 | 9.00 | 25.09524 | 338.8939 | 0.941777 | -0.176306 | -0.678958 | 0.333787 | -0.029824 | 2 | three |
| 27 | 2.572872 | 0 | 10.71 | 22.38095 | 326.6503 | 1.021622 | 0.549933 | 2.249039 | -0.542318 | -0.643055 | 2 | three |
| 28 | 3.342225 | 0 | 10.00 | 23.38095 | 319.5205 | 0.959893 | 0.415410 | 1.011342 | 0.108957 | -0.211969 | 2 | three |
| 30 | 3.309843 | 0 | 9.14 | 25.23810 | 191.4179 | 0.832136 | 1.364486 | 0.204971 | 1.717814 | -0.104792 | 2 | three |
| 32 | 3.477266 | 0 | 9.57 | 49.33333 | 240.5864 | 0.593733 | 2.311827 | -0.208722 | 3.252484 | 2.566781 | 2 | three |
| 33 | 4.122511 | 0 | 10.00 | 61.23810 | 168.0302 | 0.827315 | 2.057388 | -0.476542 | 0.535338 | 0.950188 | 2 | three |
| 35 | 3.772468 | 0 | 8.29 | 21.71429 | 208.8178 | 0.982591 | 0.376920 | -0.489987 | 0.573233 | -1.572009 | 2 | three |
| 36 | 3.703463 | 1 | 8.86 | 10.47619 | 102.6965 | 0.901393 | 0.095350 | 0.083781 | 1.189038 | -0.839625 | 2 | three |
| 38 | 4.013008 | 0 | 10.00 | 60.28571 | 160.9141 | 0.916310 | 1.819755 | -0.261329 | -0.252044 | 0.133548 | 2 | three |
| 40 | 4.183526 | 0 | 10.00 | 58.33333 | 169.0734 | 1.035891 | 1.303826 | -0.240523 | -1.391457 | -0.770939 | 2 | three |
| 42 | 3.792672 | 1 | 10.71 | 12.33333 | 113.1687 | 0.885714 | 0.557711 | 1.082490 | 0.536272 | 0.360676 | 2 | three |
| 44 | 5.275848 | 0 | 11.00 | 48.23810 | 476.3094 | 0.910814 | -0.190251 | -0.358930 | -0.913353 | 2.342799 | 2 | three |

Might be data issue here, high population areas are usually city area

# Clustering 4

Feature preprocessing:

▼ Features included

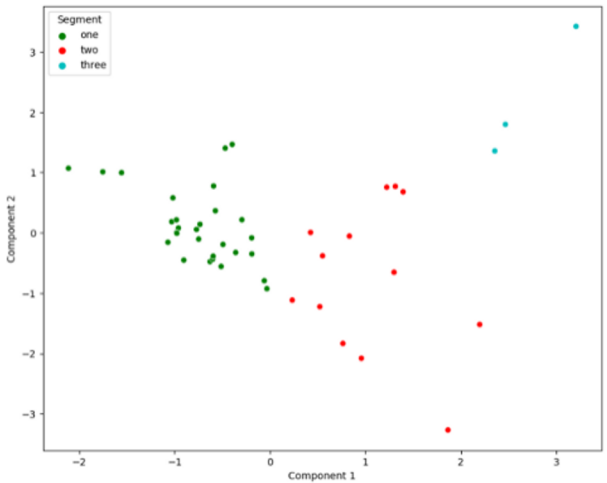
['Population_lg', 'AvgOhPerDay', 'AvgTcPerday', 'AvgAmtPerTran']

- Take log(10) of population to bring closer the scale
- IsCity is removed

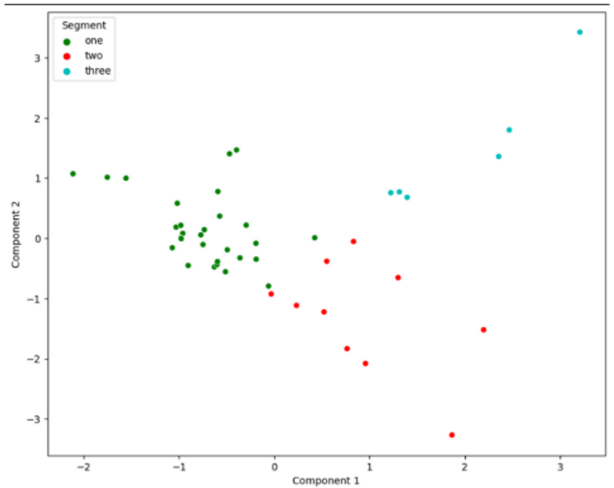
▶ PCA: 3 components

▶ 3 clusters

## Clustered: First 3 Weeks




Using original features as input
Silhouette coefficient: 0.3188649174410176




Using PCA component as input
Silhouette coefficient: 0.34978719542889736

## Explained:

- ▼ Green: Cluster 1
  - ▶ High Population: 3.0 - 5.0+
  - ▶ Higher Average Amount Per Transaction: 200 - 400+
  - ▶ Shorter Operating Hour Per Day: 8 - 10

| | Population_lg | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | Component 1 | Component 2 | Component 3 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 4.487421 | 9.00 | 55.95238 | 367.0523 | -0.592501 | 0.776851 | 0.647270 | 0 | one |
| 1 | 3.263636 | 8.43 | 34.23810 | 372.0652 | -0.630471 | -0.476591 | -0.544286 | 0 | one |
| 2 | 4.771154 | 9.29 | 48.14286 | 418.2994 | -1.018594 | 0.580277 | 1.164591 | 0 | one |
| 3 | 4.664153 | 9.86 | 45.85714 | 463.8836 | -0.960512 | 0.080932 | 1.719076 | 0 | one |
| 4 | 3.941014 | 8.71 | 37.80952 | 368.3986 | -0.771274 | 0.056705 | -0.014614 | 0 | one |
| 5 | 4.537252 | 9.36 | 35.95238 | 328.5731 | -0.573519 | 0.366737 | 0.458893 | 0 | one |
| 6 | 4.832905 | 7.93 | 34.14286 | 374.5494 | -1.754014 | 1.012047 | -0.123779 | 0 | one |
| 7 | 4.459030 | 9.00 | 27.90476 | 360.1021 | -1.031802 | 0.184914 | 0.259328 | 0 | one |
| 8 | 5.170329 | 8.07 | 31.38095 | 410.7235 | -2.114040 | 1.070390 | 0.295973 | 0 | one |
| 9 | 4.027920 | 9.00 | 31.42857 | 362.3668 | -0.748438 | -0.104126 | 0.110294 | 0 | one |
| 10 | 3.655331 | 9.00 | 31.80952 | 373.5470 | -0.601433 | -0.437000 | 0.000520 | 0 | one |
| 13 | 3.237041 | 8.14 | 34.61905 | 398.5717 | -0.904850 | -0.452100 | -0.599455 | 0 | one |
| 14 | 3.462997 | 8.57 | 29.42857 | 344.0882 | -0.596544 | -0.386709 | -0.568802 | 0 | one |
| 15 | 4.016281 | 9.86 | 35.90476 | 356.6930 | -0.190712 | -0.348350 | 0.701286 | 0 | one |
| 16 | 3.531990 | 7.86 | 23.14286 | 337.3016 | -1.071278 | -0.156646 | -1.123069 | 0 | one |
| 17 | 4.391746 | 7.71 | 38.66667 | 179.7334 | -0.469782 | 1.403614 | -1.501447 | 0 | one |
| 18 | 3.695919 | 8.14 | 33.52381 | 363.9960 | -0.981758 | 0.005909 | -0.586728 | 0 | one |
| 19 | 4.273927 | 7.43 | 35.85714 | 143.3167 | -0.397385 | 1.466742 | -1.978978 | 0 | one |
| 23 | 3.552668 | 9.00 | 27.28571 | 353.1899 | -0.512604 | -0.556981 | -0.214703 | 0 | one |
| 24 | 5.509701 | 9.14 | 29.14286 | 368.8822 | -1.556921 | 0.997396 | 0.912785 | 0 | one |
| 26 | 4.491586 | 9.00 | 25.09524 | 338.8939 | -0.981494 | 0.215205 | 0.123861 | 0 | one |
| 29 | 3.569725 | 10.00 | 42.57143 | 411.0637 | -0.060239 | -0.793614 | 0.964004 | 0 | one |
| 31 | 3.865992 | 9.43 | 52.71429 | 439.7713 | -0.494030 | -0.191325 | 1.003642 | 0 | one |
| 34 | 3.735918 | 8.14 | 28.42857 | 343.1107 | -0.977467 | -0.006589 | -0.744285 | 0 | one |
| 35 | 3.772468 | 8.29 | 21.71429 | 208.8178 | -0.295159 | 0.217363 | -1.449361 | 0 | one |
| 36 | 3.703463 | 8.86 | 10.47619 | 102.6965 | 0.424491 | 0.007604 | -1.822455 | 0 | one |
| 41 | 3.349278 | 8.21 | 14.00000 | 230.2717 | -0.360948 | -0.325408 | -1.676711 | 0 | one |
| 43 | 4.083610 | 10.00 | 60.42857 | 441.3509 | -0.192726 | -0.082951 | 1.588257 | 0 | one |
| 44 | 5.275848 | 11.00 | 48.23810 | 476.3094 | -0.734364 | 0.142337 | 2.866557 | 0 | one |

- Red: Cluster 2
  - Low Population: 1.2 - 3.3+
  - Low Average Transaction Count Per Day: 10 - 40+
  - Relatively Low Average Amount Per Transaction: 100 - 300+

| | Population_lg | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | Component 1 | Component 2 | Component 3 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3.613419 | 10.82 | 22.42857 | 313.6511 | 0.521768 | -1.223165 | 0.752506 | 1 | two |
| 1 | 2.683047 | 8.57 | 32.14286 | 327.4585 | -0.033551 | -0.925541 | -0.996829 | 1 | two |
| 2 | 2.731589 | 11.43 | 40.23810 | 210.3304 | 2.198272 | -1.519221 | 0.385491 | 1 | two |
| 2 | 1.230449 | 11.00 | 24.14286 | 326.7199 | 1.865045 | -3.267703 | -0.160936 | 1 | two |
| 2 | 2.572872 | 10.71 | 22.38095 | 326.6503 | 0.957186 | -2.078378 | 0.258474 | 1 | two |
| 2 | 3.342225 | 10.00 | 23.38095 | 319.5205 | 0.232415 | -1.115391 | 0.120668 | 1 | two |
| 3 | 3.309843 | 9.14 | 25.23810 | 191.4179 | 0.550310 | -0.381157 | -1.154423 | 1 | two |
| 3 | 3.477266 | 9.57 | 49.33333 | 240.5864 | 0.831668 | -0.054075 | -0.227748 | 1 | two |
| 3 | 2.392697 | 10.14 | 36.47619 | 371.1970 | 0.763767 | -1.833523 | 0.207793 | 1 | two |
| 4 | 3.792672 | 10.71 | 12.33333 | 113.1687 | 1.300094 | -0.653199 | -0.464964 | 1 | two |

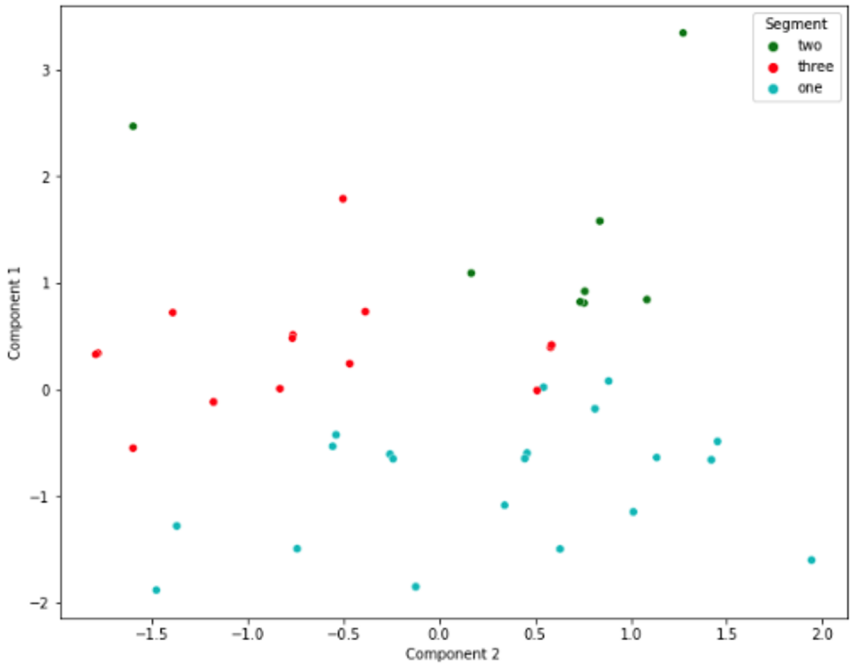
- Light Blue: Cluster 3
  - Higher Population: 3.5 - 4.0+
  - Higher Average Transaction Count Per Day: 60 - 150+
  - Long Operating Hour Per Day: 9 - 10

| | Population_lg | AvgOhPerDay | AvgTcPerday | AvgAmtPerTran | Component 1 | Component 2 | Component 3 | K-means PCA | Segment |
|---|---|---|---|---|---|---|---|---|---|
| 20 | 3.884852 | 9.00 | 167.30000 | 93.1460 | 3.208252 | 3.422206 | 0.183340 | 2 | three |
| 22 | 3.502154 | 9.86 | 115.35710 | 199.3733 | 2.355986 | 1.357552 | 0.541137 | 2 | three |
| 33 | 4.122511 | 10.00 | 61.23810 | 168.0302 | 1.313065 | 0.770005 | 0.106100 | 2 | three |
| 38 | 4.013006 | 10.00 | 60.28571 | 160.9141 | 1.395124 | 0.680416 | 0.003606 | 2 | three |
| 39 | 3.960756 | 10.00 | 112.28570 | 137.4153 | 2.467864 | 1.798160 | 0.472225 | 2 | three |
| 40 | 4.183526 | 10.00 | 58.33333 | 169.0734 | 1.223106 | 0.756384 | 0.105895 | 2 | three |

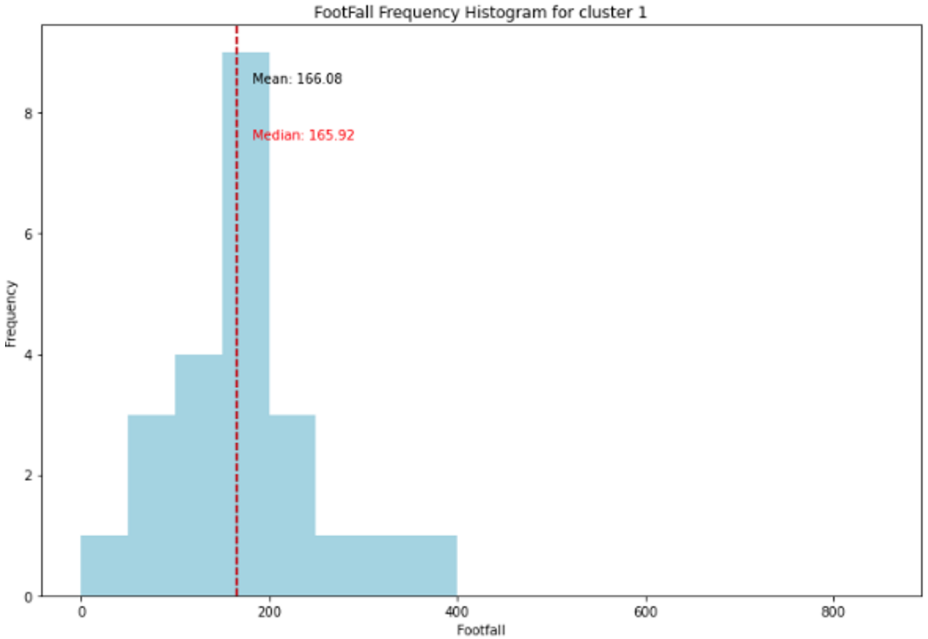
# Data

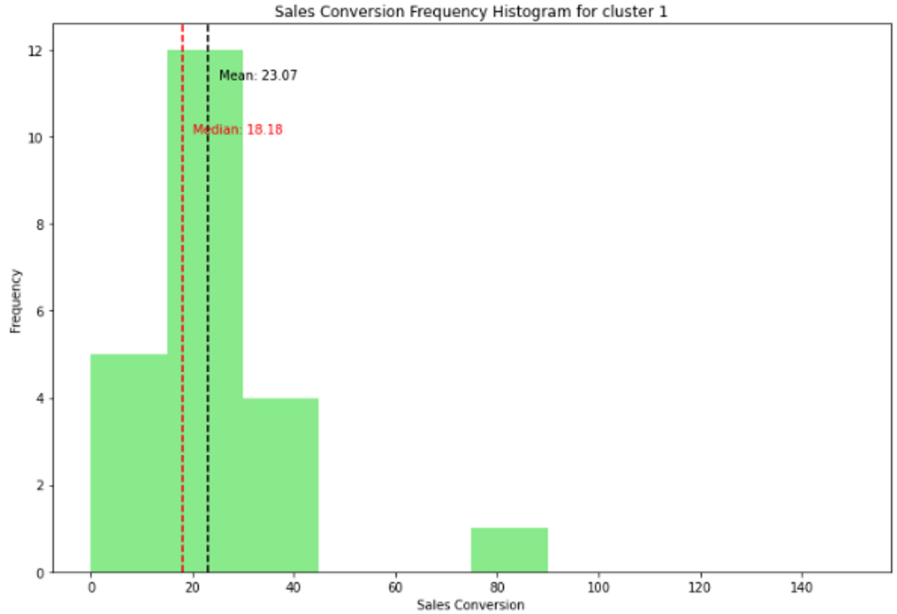
| | Population | AverageStoreOpenHours | RainingDaysPerMonth | Component 1 | Component 2 | Component 3 | K-means PCA |
|---|---|---|---|---|---|---|---|
| 0 | 30720 | 9.000000 | 11 | 0.842228 | 1.083763 | 0.597679 | 1 |
| 1 | 1835 | 8.428571 | 7 | 0.241832 | -0.467618 | -1.073796 | 2 |
| 2 | 59041 | 9.285714 | 2 | 2.466388 | -1.598279 | 2.207061 | 1 |
| 3 | 46148 | 9.857143 | 8 | 1.090903 | 0.166685 | 1.971386 | 1 |
| 4 | 8730 | 8.714286 | 11 | 0.079229 | 0.883624 | -0.604811 | 0 |
| 5 | 34455 | 9.357143 | 10 | 0.810642 | 0.755087 | 1.045466 | 1 |
| 6 | 68062 | 7.928571 | 10 | 3.341256 | 1.272426 | 1.493186 | 1 |
| 7 | 28776 | 9.000000 | 10 | 0.820889 | 0.735526 | 0.523752 | 1 |
| 8 | 10664 | 9.000000 | 10 | 0.020289 | 0.544001 | -0.291301 | 0 |
| 9 | 4522 | 9.000000 | 6 | 0.007158 | -0.831671 | -0.513476 | 2 |
| 10 | 4106 | 10.821429 | 5 | -1.279134 | -1.370254 | 0.838747 | 0 |
| 11 | 482 | 8.571429 | 11 | -0.180846 | 0.812602 | -1.082439 | 0 |
| 12 | 1726 | 8.142857 | 6 | 0.510623 | -0.764059 | -1.278070 | 2 |
| 13 | 2904 | 8.571429 | 10 | -0.009196 | 0.510533 | -0.959892 | 2 |
| 14 | 10382 | 9.857143 | 7 | -0.425461 | -0.539202 | 0.375444 | 0 |
| 15 | 3404 | 7.857143 | 7 | 0.729223 | -0.386241 | -1.429037 | 2 |
| 16 | 24646 | 7.714286 | 10 | 1.578915 | 0.837620 | -0.620258 | 1 |
| 17 | 4965 | 8.142857 | 10 | 0.395434 | 0.580916 | -1.186532 | 2 |
| 18 | 18790 | 7.428571 | 6 | 1.787444 | -0.502634 | -1.042487 | 2 |
| 19 | 7671 | 9.000000 | 3 | 0.340123 | -1.781414 | -0.331105 | 2 |
| 20 | 539 | 11.428571 | 5 | -1.880970 | -1.476808 | 1.130693 | 0 |
| 21 | 3178 | 9.857143 | 4 | -0.550126 | -1.598423 | 0.091923 | 2 |
| 22 | 3570 | 9.000000 | 13 | -0.487056 | 1.452029 | -0.651200 | 0 |
| 23 | 17 | 11.000000 | 9 | -1.848877 | -0.123015 | 0.733598 | 0 |
| 24 | 31016 | 9.000000 | 10 | 0.919903 | 0.759212 | 0.624554 | 1 |
| 25 | 374 | 10.711429 | 7 | -1.492807 | -0.741885 | 0.561719 | 0 |
| 26 | 2199 | 10.000000 | 10 | -1.085452 | 0.341114 | 0.073001 | 0 |
| 27 | 3713 | 10.000000 | 12 | -1.147710 | 1.012486 | 0.114023 | 0 |
| 28 | 2041 | 9.142857 | 13 | -0.659152 | 1.419664 | -0.613544 | 0 |
| 29 | 7345 | 9.428571 | 5 | -0.116996 | -1.178089 | -0.053499 | 2 |
| 30 | 3001 | 9.571429 | 8 | -0.607293 | -0.257178 | -0.183184 | 0 |
| 31 | 13259 | 10.000000 | 10 | -0.596570 | 0.458068 | 0.570709 | 0 |
| 32 | 5444 | 8.142857 | 10 | 0.416607 | 0.585981 | -1.164976 | 2 |
| 33 | 5922 | 8.285714 | 4 | 0.720769 | -1.391246 | -0.955676 | 2 |
| 34 | 5052 | 8.857143 | 3 | 0.328865 | -1.792912 | -0.555424 | 2 |
| 35 | 247 | 10.142857 | 15 | -1.599197 | 1.942681 | 0.023848 | 0 |
| 36 | 10304 | 10.000000 | 7 | -0.533418 | -0.556223 | 0.478396 | 0 |
| 37 | 9136 | 10.000000 | 8 | -0.649637 | -0.240893 | 0.412281 | 0 |
| 38 | 15259 | 10.000000 | 12 | -0.637346 | 1.134578 | 0.633601 | 0 |
| 39 | 2235 | 8.214286 | 6 | 0.480867 | -0.766774 | -1.201934 | 2 |
| 40 | 6204 | 10.714286 | 11 | -1.495557 | 0.630164 | 0.771984 | 0 |
| 41 | 12123 | 10.000000 | 10 | -0.646784 | 0.446055 | 0.519588 | 0 |

# Cluster plot

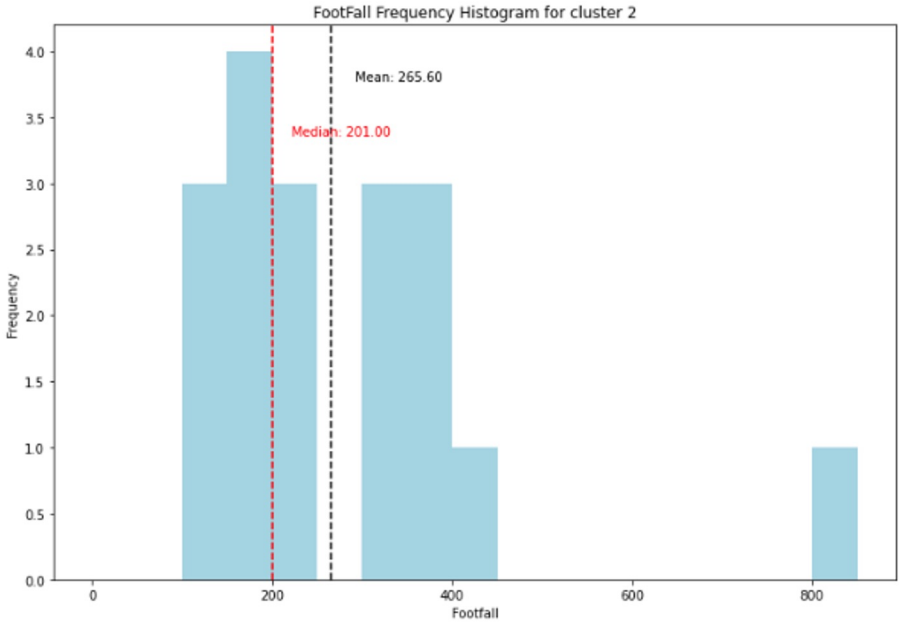


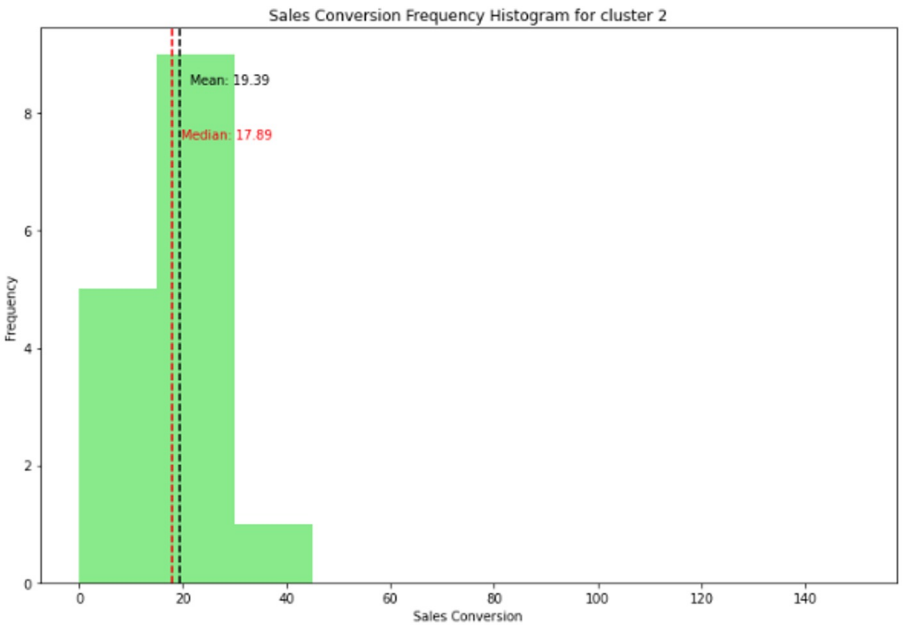
# Histogram for cluster 1



# Histogram for cluster 2



# Histogram for cluster 3