

Descripción de clusters

Utilizamos scaled_df (robust scaler) para crear las características con las que luego se realiza el pca + imp de esas características en el componente principal (df_with_pca)

description_df = df_with_pca + importance_df_filt (importancias en el componente)

- Filtramos solo por importancias con thresh + 0.05
- Se calcula la media del resto de características por cluster

| | 3 | 0 | 1 | 2 |
|----------------------------------|-----------|----------|----------|-----------|
| MAX(sales.raw_earn) | -0.143413 | 1.530048 | 6.398908 | 25.567111 |
| MEAN(sales.raw_earn) | -0.162812 | 1.464448 | 6.035324 | 26.933683 |
| SUM(sales.raw_earn) | -0.142242 | 1.413018 | 5.961999 | 19.909549 |
| MAX(sales.PERCENTILE(raw_earn)) | -0.722129 | 0.409907 | 0.518868 | 0.527169 |
| MEAN(sales.PERCENTILE(raw_earn)) | -0.270417 | 0.811930 | 1.432882 | 1.743683 |
| SUM(sales.PERCENTILE(raw_earn)) | -0.181266 | 0.812113 | 1.482631 | 1.298388 |
| PERCENTILE(MAX(sales.raw_earn)) | -0.269619 | 0.645787 | 0.943105 | 0.996227 |
| PERCENTILE(SUM(sales.raw_earn)) | -0.264532 | 0.630299 | 0.944029 | 0.997129 |
| MIN(sales.raw_earn) | 0.111351 | 0.761282 | 1.197889 | 2.866333 |

Robust Scaler

El `RobustScaler` es una técnica de escalado que no se ve afectada por outliers, a diferencia de otras técnicas como el `StandardScaler`. Este escalador utiliza el **mediano** y el **rango intercuartílico (IQR)**, que es la diferencia entre el percentil 75 y el percentil 25 de los datos.

El escalado se realiza con la siguiente fórmula:

$$X_{\text{scaled}} = \frac{X - \text{mediana}(X)}{\text{IQR}(X)}$$



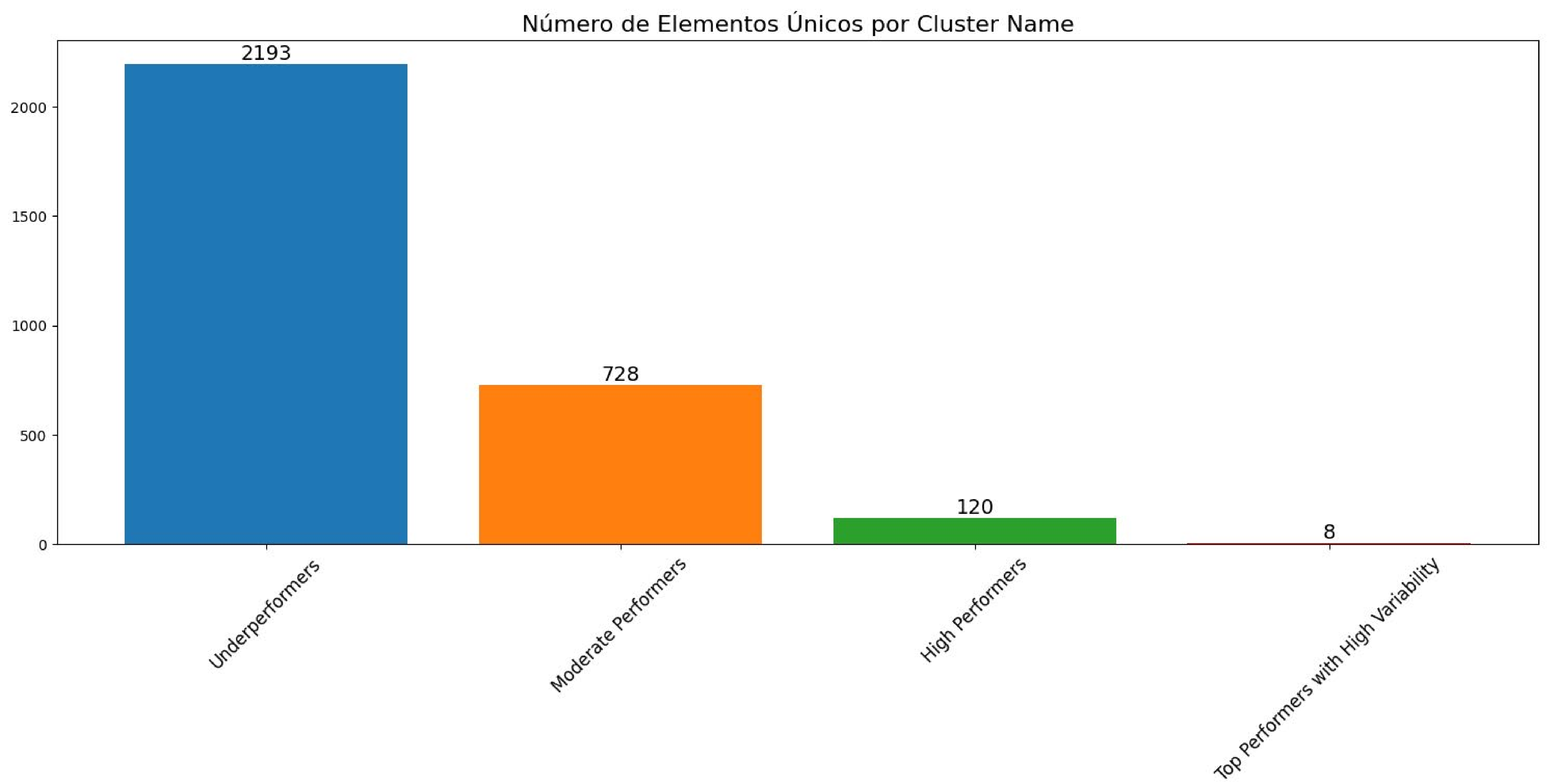
CLUSTERING
PRODUCTOS

1

PERFORMANCE EN VENTAS

file/

items_clusters_raw_earn.csv



- 0: { 'cluster_name': 'Moderate Performers',
 'description': 'Productos con ingresos medios y estables, con un rendimiento aceptable, pero sin alcanzar los niveles más altos. Presentan un crecimiento moderado, con baja variabilidad en sus ingresos.'
 },
- 1: { 'cluster_name': 'High Performers',
 'description': 'Productos con ingresos elevados en todas las métricas (SUM, MEAN, y MAX), lo que indica un buen desempeño continuo. Son productos que consistentemente generan buenas ventas con poca variabilidad.'
 },
- 2: { 'cluster_name': 'Top Performers with High Variability',
 'description': 'Productos con los ingresos más altos, pero con una notable variabilidad. Estos productos tienden a generar altos ingresos, pero sus ventas pueden ser impulsadas por promociones o eventos especiales.'
 },
- 3: { 'cluster_name': 'Underperformers',
 'description': 'Productos con ingresos bajos o incluso negativos en algunas métricas. Estos productos están rezagados en términos de rendimiento y podrían estar enfrentando dificultades para generar ingresos de manera constante.'
 }

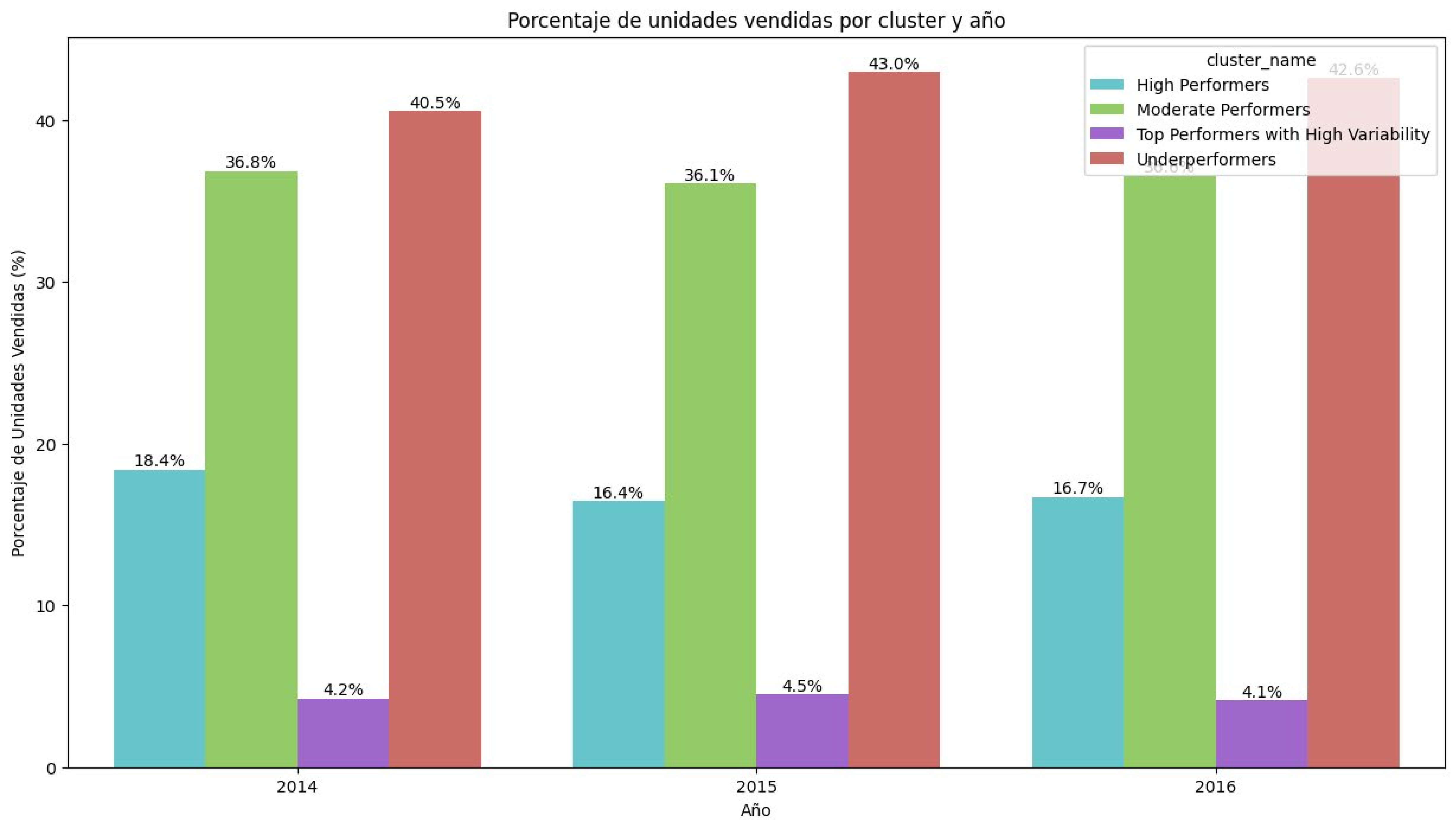
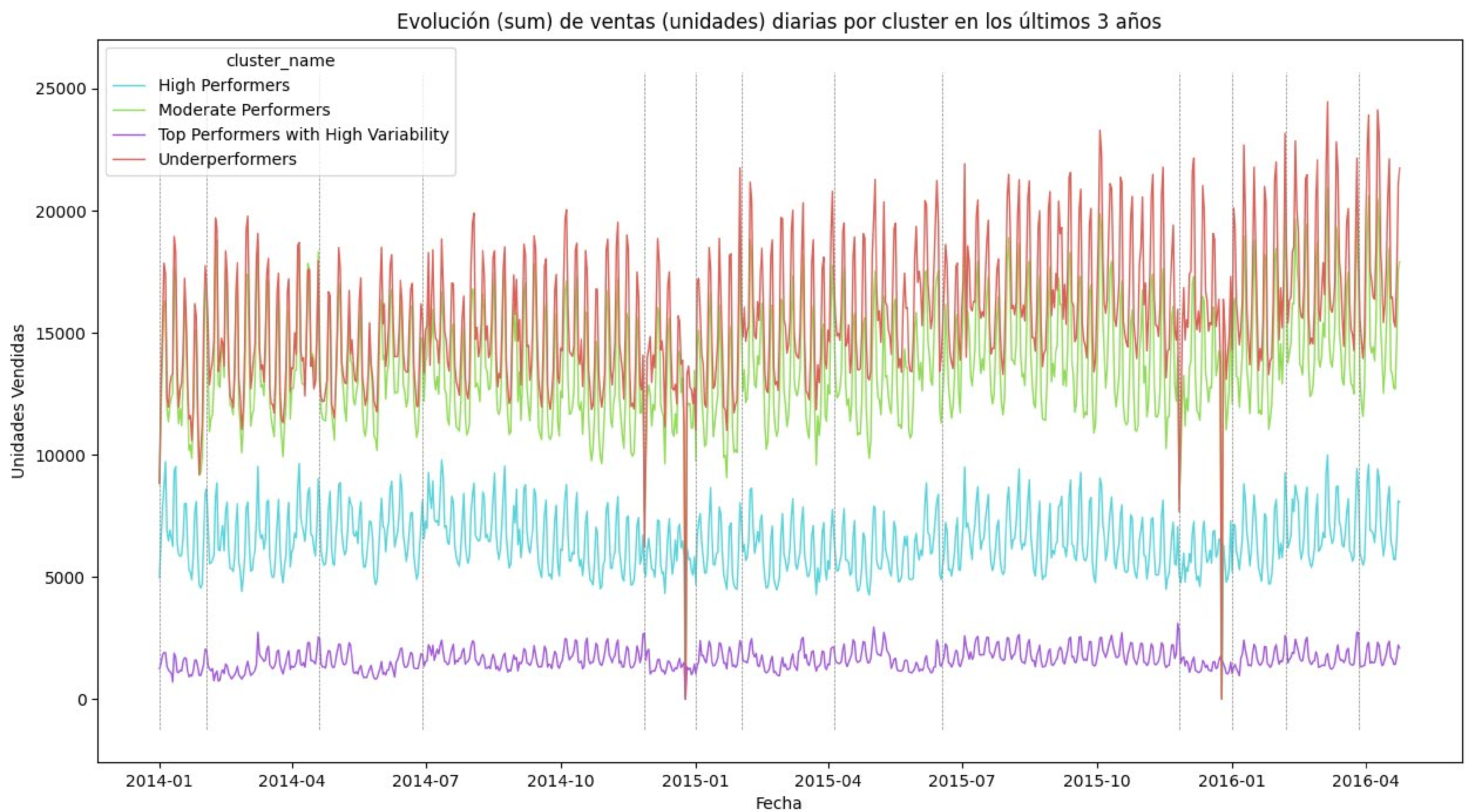
1

PERFORMANCE EN VENTAS

units

file/

items_clusters_raw_earn.csv



1 **PERFORMANCE EN VENTAS**

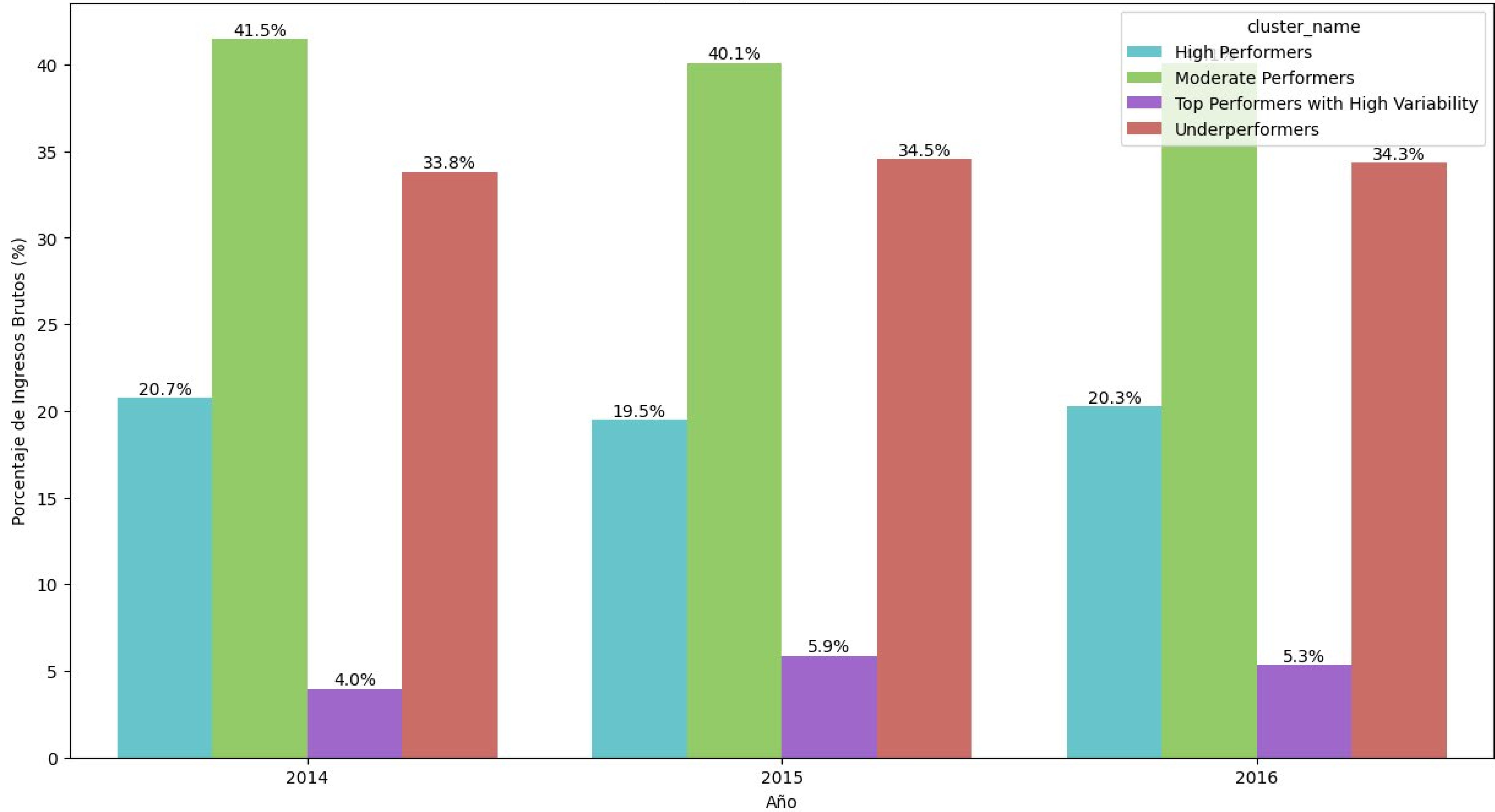
raw_earn

file/
items_clusters_raw_earn.csv

Evolución (sum) de ingresos brutos por cluster



Porcentaje de ingresos brutos por cluster



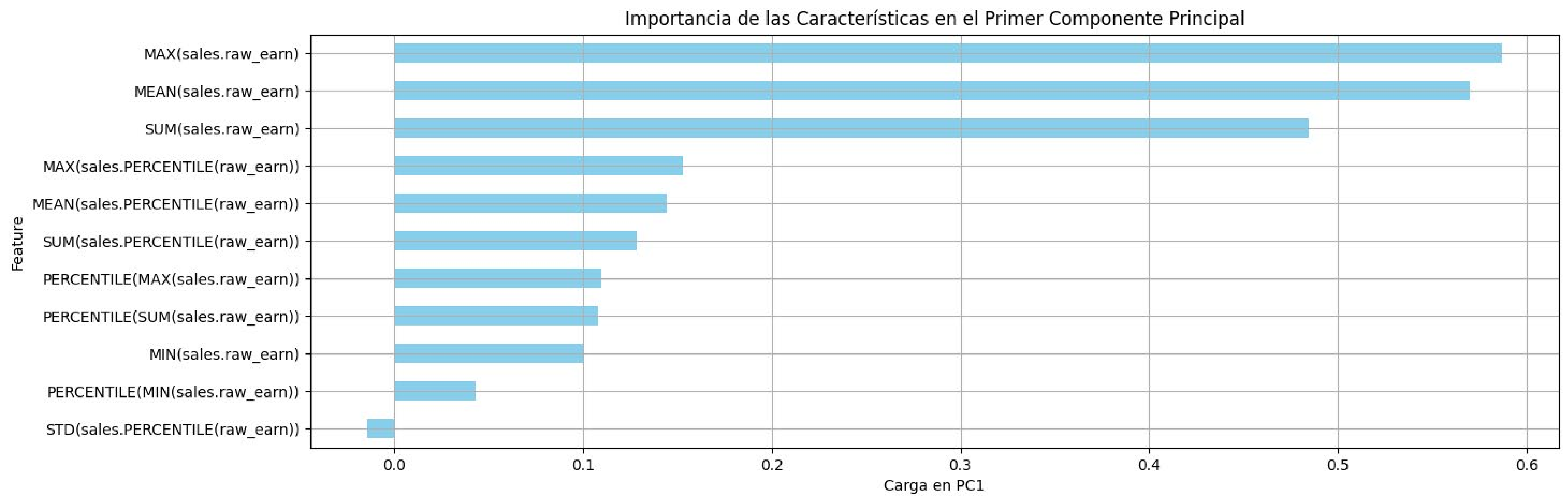
1

PERFORMANCE EN VENTAS

file/

items_clusters_sales_performance.csv

| | 3 | 0 | 1 | 2 |
|----------------------------------|-----------|----------|----------|-----------|
| MAX(sales.raw_earn) | -0.143413 | 1.530048 | 6.398908 | 25.567111 |
| MEAN(sales.raw_earn) | -0.162812 | 1.464448 | 6.035324 | 26.933683 |
| SUM(sales.raw_earn) | -0.142242 | 1.413018 | 5.961999 | 19.909549 |
| MAX(sales.PERCENTILE(raw_earn)) | -0.722129 | 0.409907 | 0.518868 | 0.527169 |
| MEAN(sales.PERCENTILE(raw_earn)) | -0.270417 | 0.811930 | 1.432882 | 1.743683 |
| SUM(sales.PERCENTILE(raw_earn)) | -0.181266 | 0.812113 | 1.482631 | 1.298388 |
| PERCENTILE(MAX(sales.raw_earn)) | -0.269619 | 0.645787 | 0.943105 | 0.996227 |
| PERCENTILE(SUM(sales.raw_earn)) | -0.264532 | 0.630299 | 0.944029 | 0.997129 |
| MIN(sales.raw_earn) | 0.111351 | 0.761282 | 1.197889 | 2.866333 |

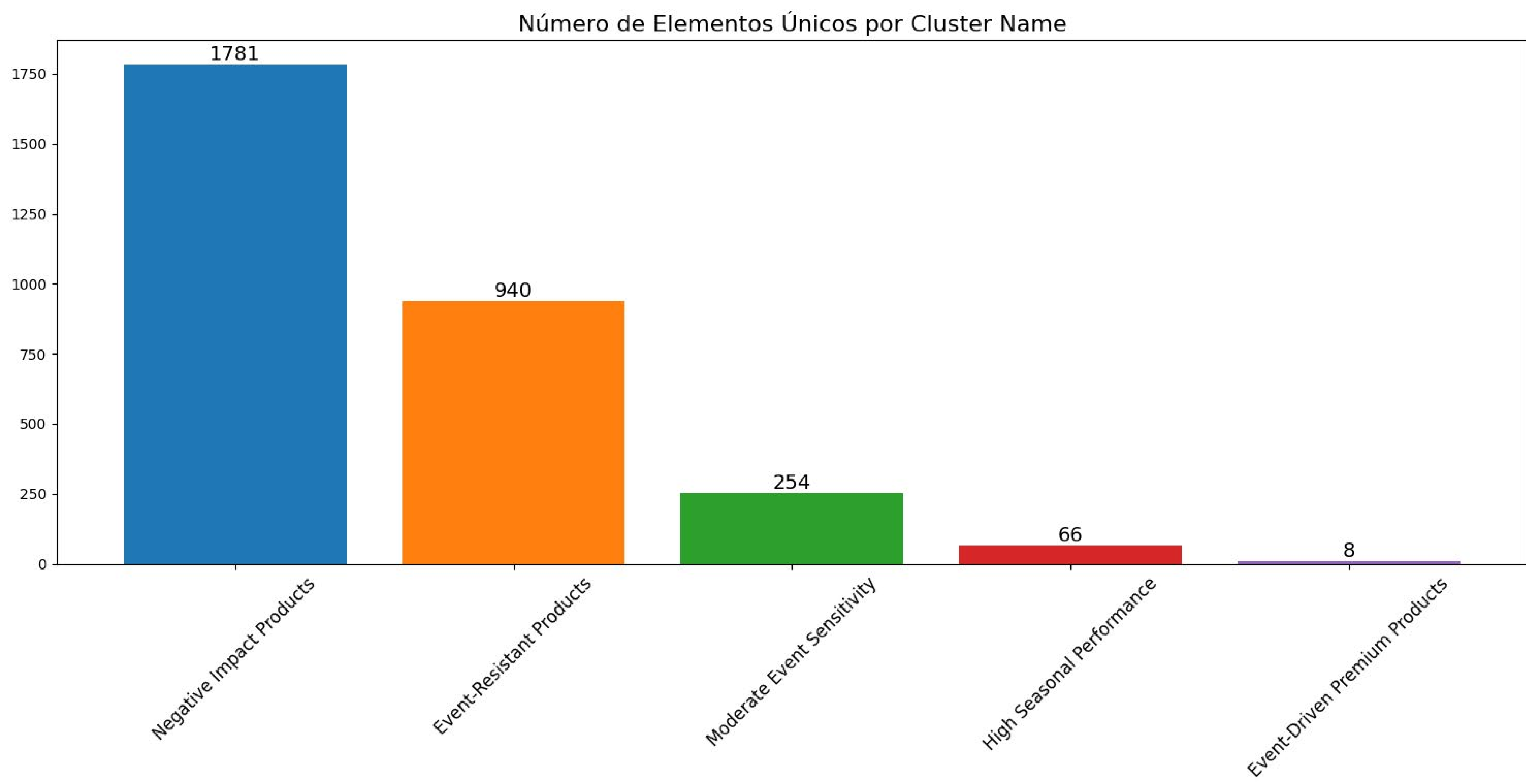


2

INFLUENCIA EVENTOS SOBRE VENTAS

file/

items_clusters_events_influence_sales.csv



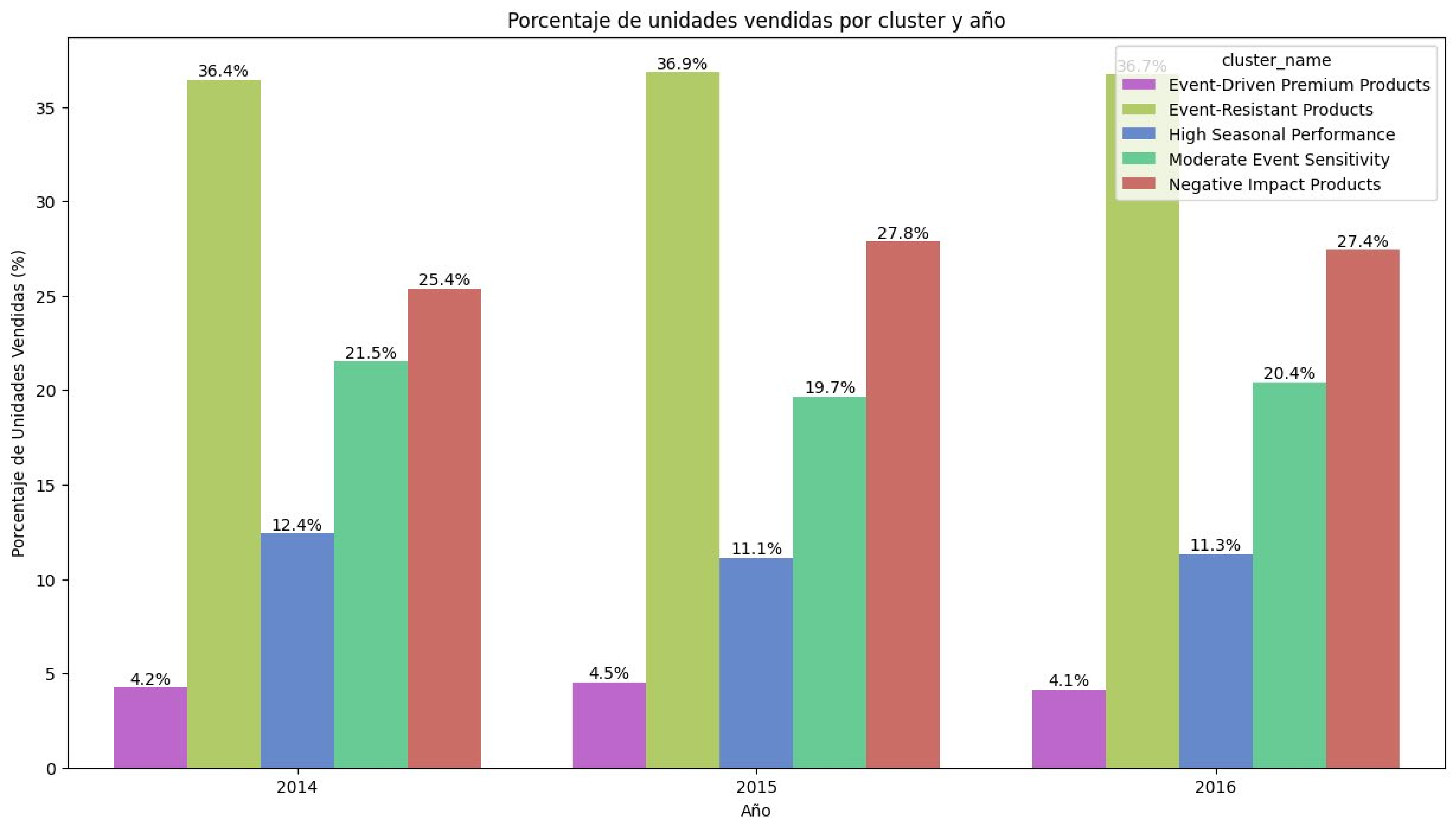
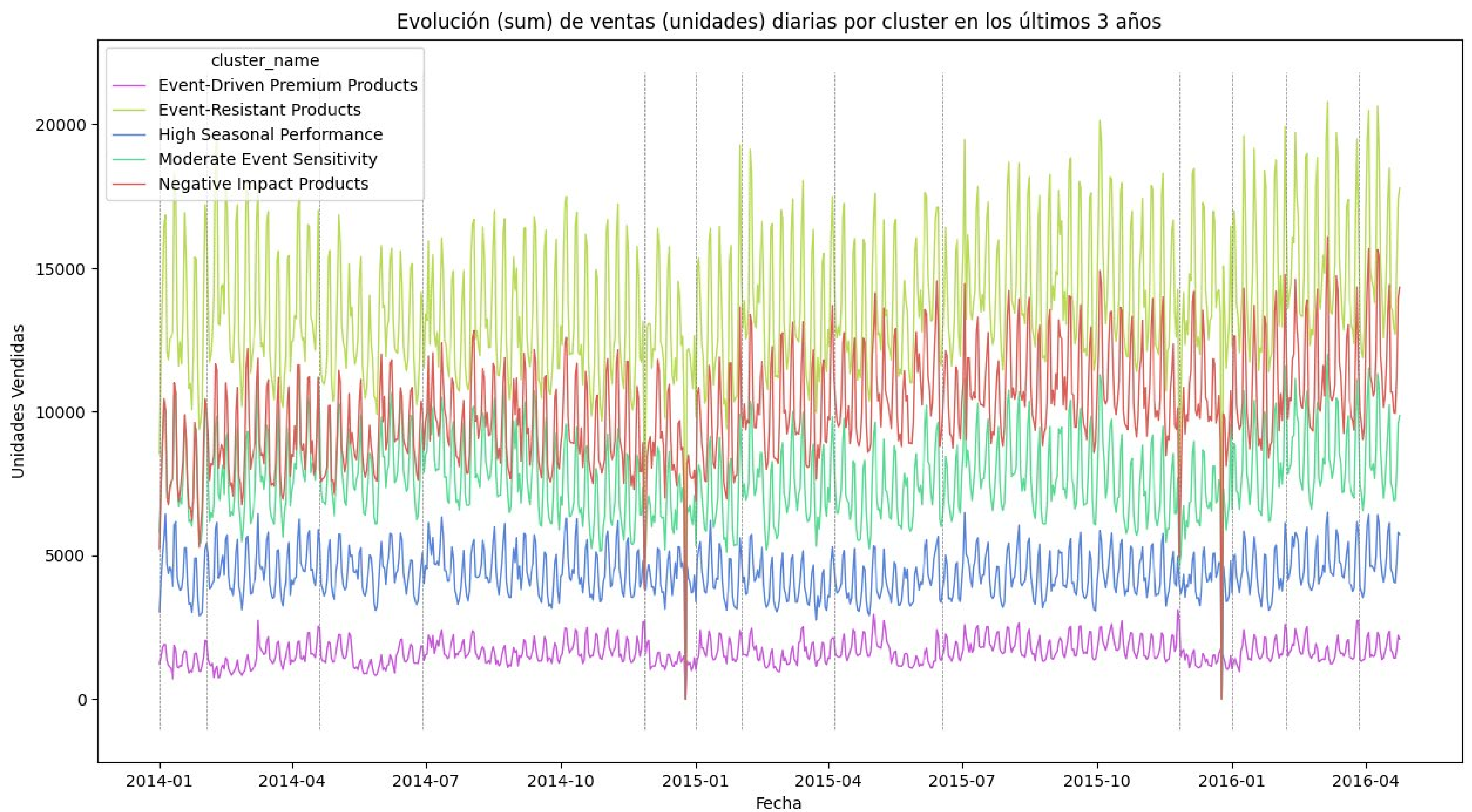
- 0: { 'cluster_name': 'Negative Impact Products',
 'description': 'Productos que presentan un impacto negativo en ingresos durante la mayoría de los eventos, con una ligera caída en las ventas durante festividades como el Año Nuevo y la Pascua.'
 },
- 1: { 'cluster_name': 'High Seasonal Performance',
 'description': 'Productos con ingresos muy elevados, especialmente durante eventos importantes como el Año Nuevo y el Super Bowl. Su desempeño está fuertemente ligado a eventos específicos.'
 },
- 2: { 'cluster_name': 'Event-Driven Premium Products',
 'description': 'Productos con ingresos muy altos y sensibilidad a eventos como el Año Nuevo y el Super Bowl. Estos productos tienden a rendir mejor en fechas especiales.'
 },
- 3: { 'cluster_name': 'Moderate Event Sensitivity',
 'description': 'Productos que muestran un desempeño positivo moderado durante eventos, sin grandes variaciones. Están ligeramente influenciados por eventos como el Año Nuevo, Pascua y el Super Bowl.'
 },
- 4: { 'cluster_name': 'Event-Resistant Products',
 'description': 'Productos con un impacto marginalmente positivo durante eventos, pero que no dependen fuertemente de los mismos. Muestran estabilidad en ventas, independientemente de festividades.'
 }

2

INFLUENCIA EVENTOS SOBRE VENTAS

units

file/
items_clusters_events_influence_sales.csv



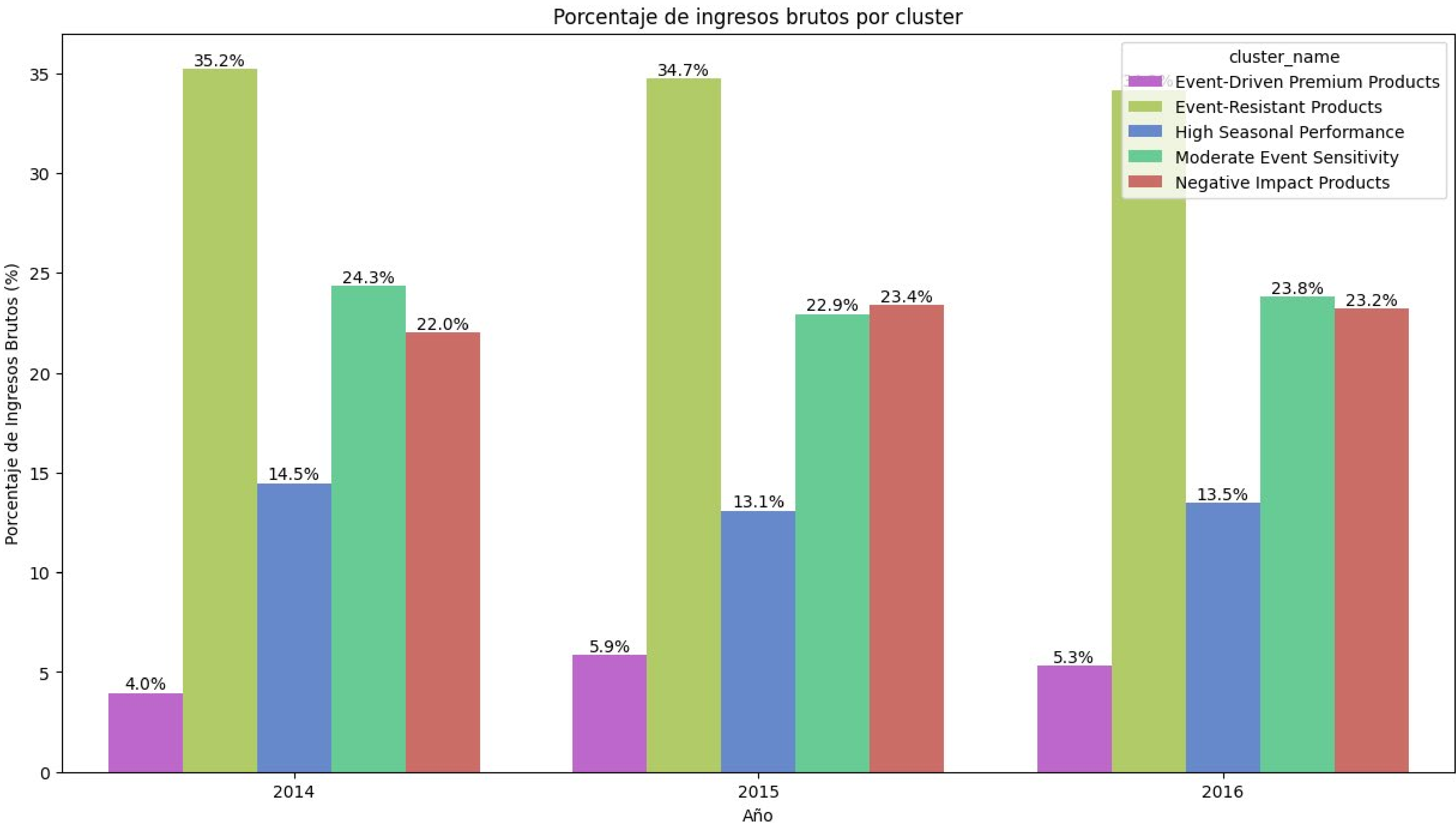
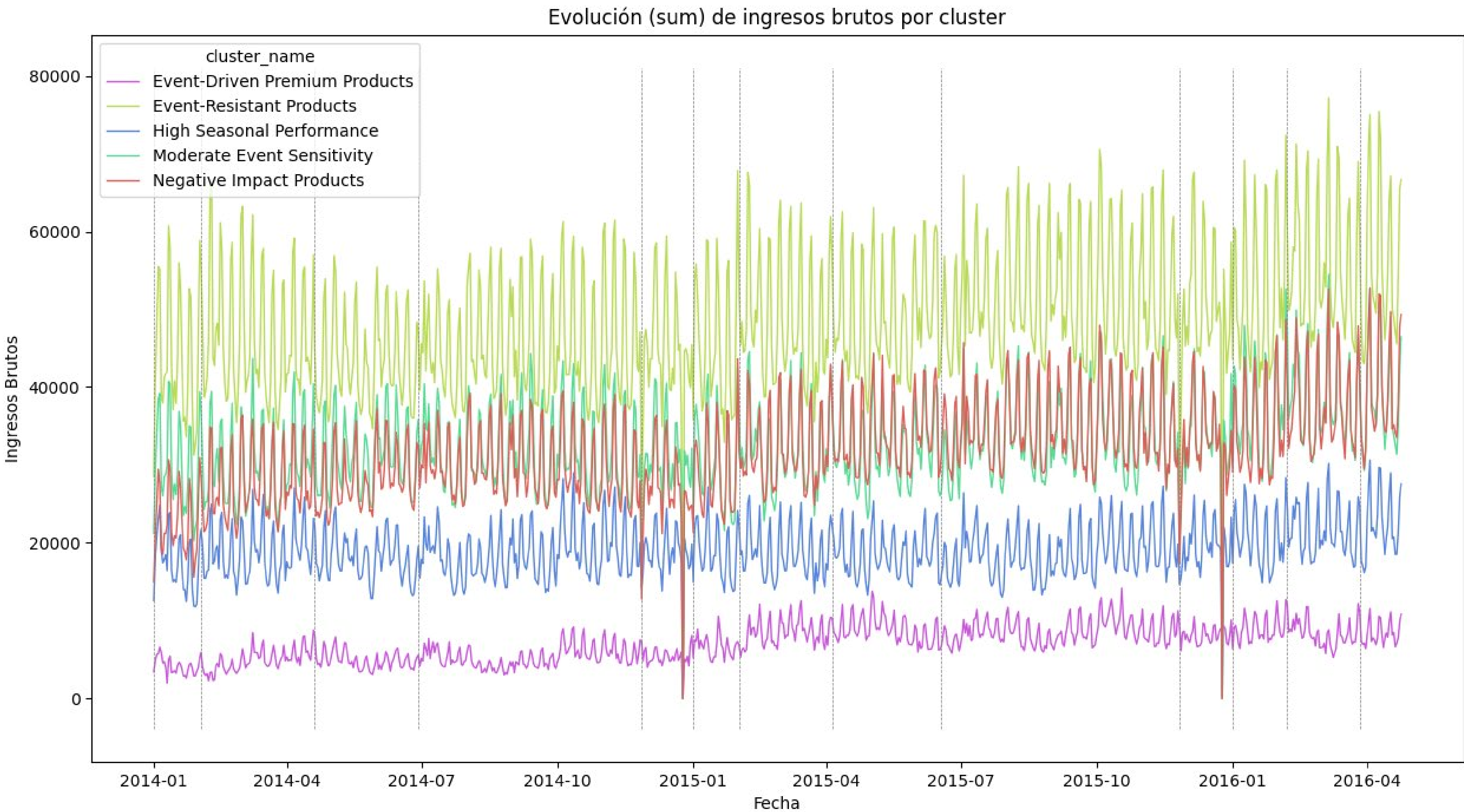
2

INFLUENCIA EVENTOS SOBRE VENTAS

raw_earn

file/

items_clusters_events_influence_sales.csv



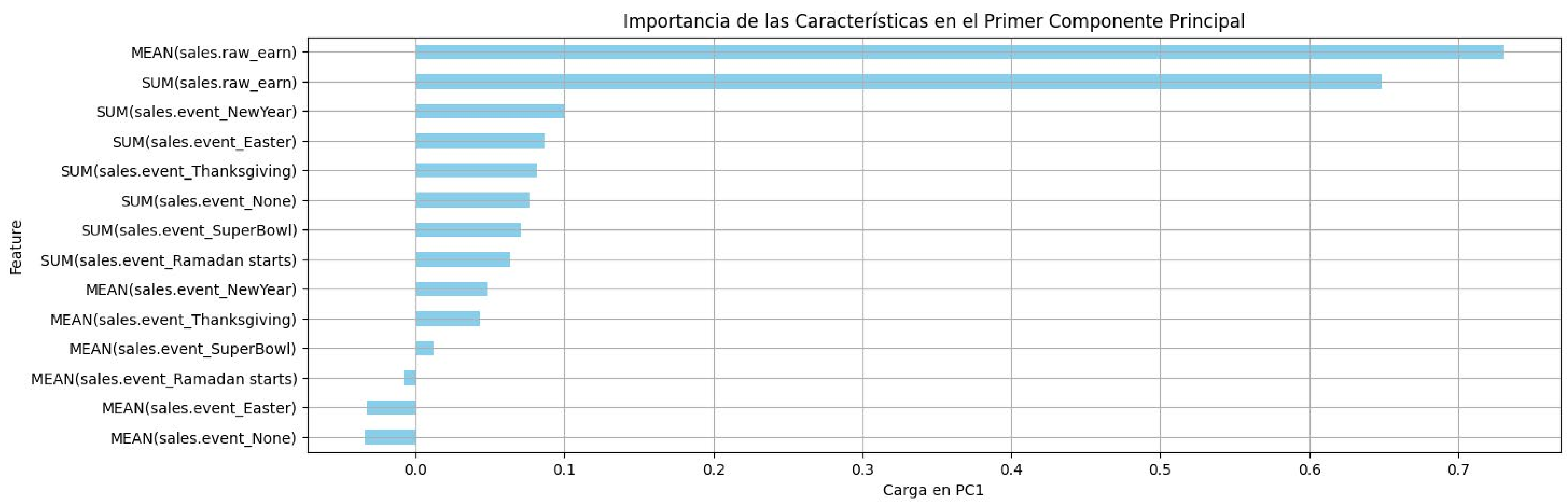
2

INFLUENCIA EVENTOS SOBRE VENTAS

file/

items_clusters_events_influence_sales.csv

| | 0 | 1 | 2 | 3 | 4 |
|---------------------------------|-----------|----------|-----------|----------|----------|
| MEAN(sales.raw_earn) | -0.248575 | 7.559069 | 26.933683 | 2.940950 | 0.670343 |
| SUM(sales.raw_earn) | -0.273698 | 7.616904 | 19.909549 | 2.918931 | 0.718632 |
| SUM(sales.event_NewYear) | -0.381483 | 0.607656 | 0.210526 | 0.479072 | 0.336842 |
| SUM(sales.event_Easter) | -0.372263 | 0.444129 | 0.156250 | 0.384843 | 0.293085 |
| SUM(sales.event_Thanksgiving) | -0.290707 | 0.566667 | 0.087500 | 0.405709 | 0.330532 |
| SUM(sales.event_None) | -0.302265 | 0.466429 | 0.043164 | 0.358349 | 0.296623 |
| SUM(sales.event_SuperBowl) | -0.273195 | 0.429697 | 0.085000 | 0.331811 | 0.280723 |
| SUM(sales.event_Ramadan starts) | -0.244619 | 0.348485 | -0.011905 | 0.323397 | 0.266261 |

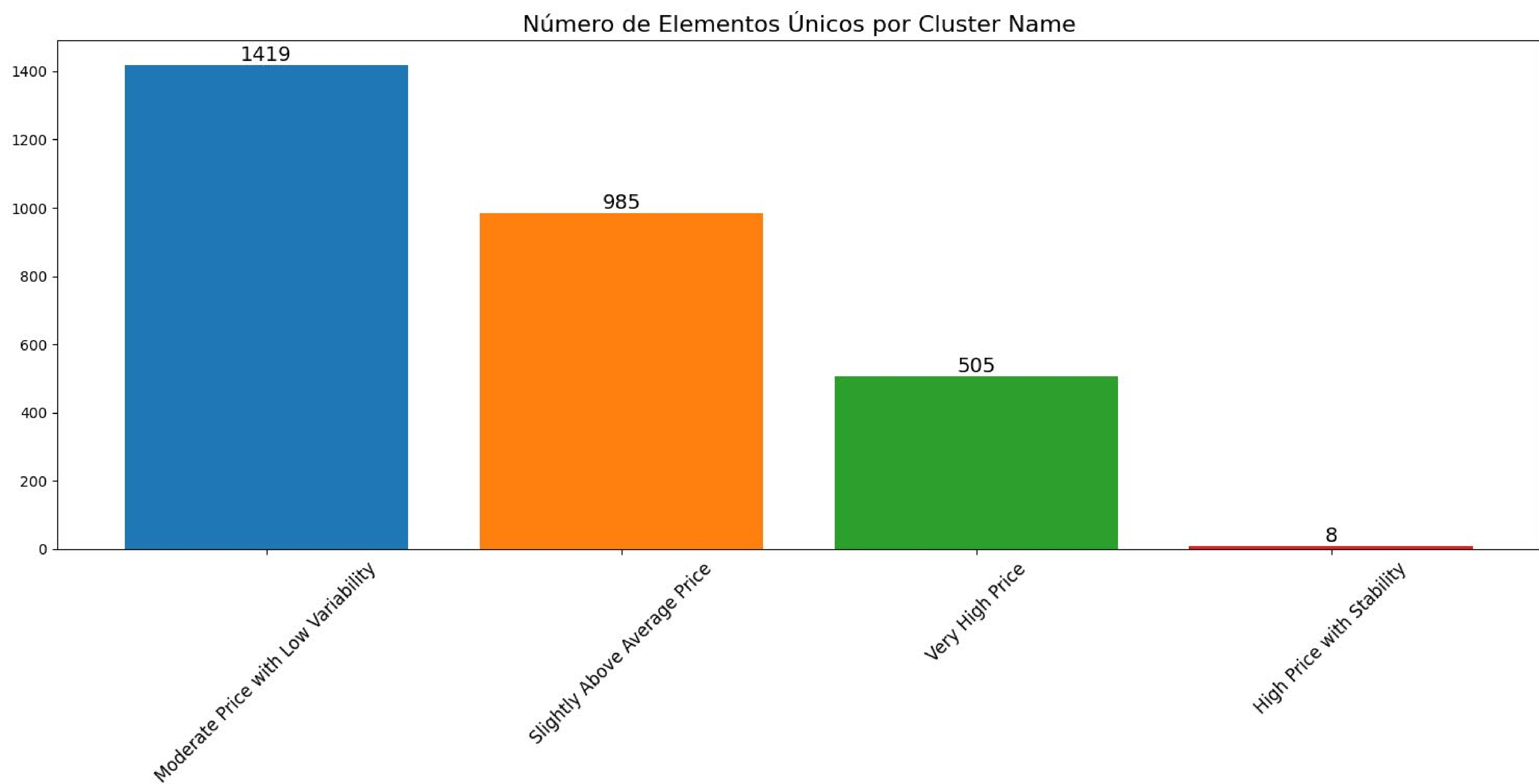


3

RANGO DE PRECIOS

file/

items_clusters_price_range.csv



```
0: {
  'cluster_name': 'Very High Price',
  'description': 'Productos con precios extremadamente altos en todas las métricas, con una gran variabilidad y dispersion de precios.'
},
1: {
  'cluster_name': 'Moderate Price with Low Variability',
  'description': 'Productos de precio moderado, con variación de precios limitada. Los productos parecen estar en un rango estable de precios medios.'
},
2: {
  'cluster_name': 'High Price with Stability',
  'description': 'Productos de precio alto con cierta estabilidad en los precios, pero manteniendo variación considerable en el precio máximo y promedio.'
},
3: {
  'cluster_name': 'Slightly Above Average Price',
  'description': 'Productos con precios ligeramente superiores al promedio, pero con una variación significativa en los precios máximos y mínimos.'
}
```

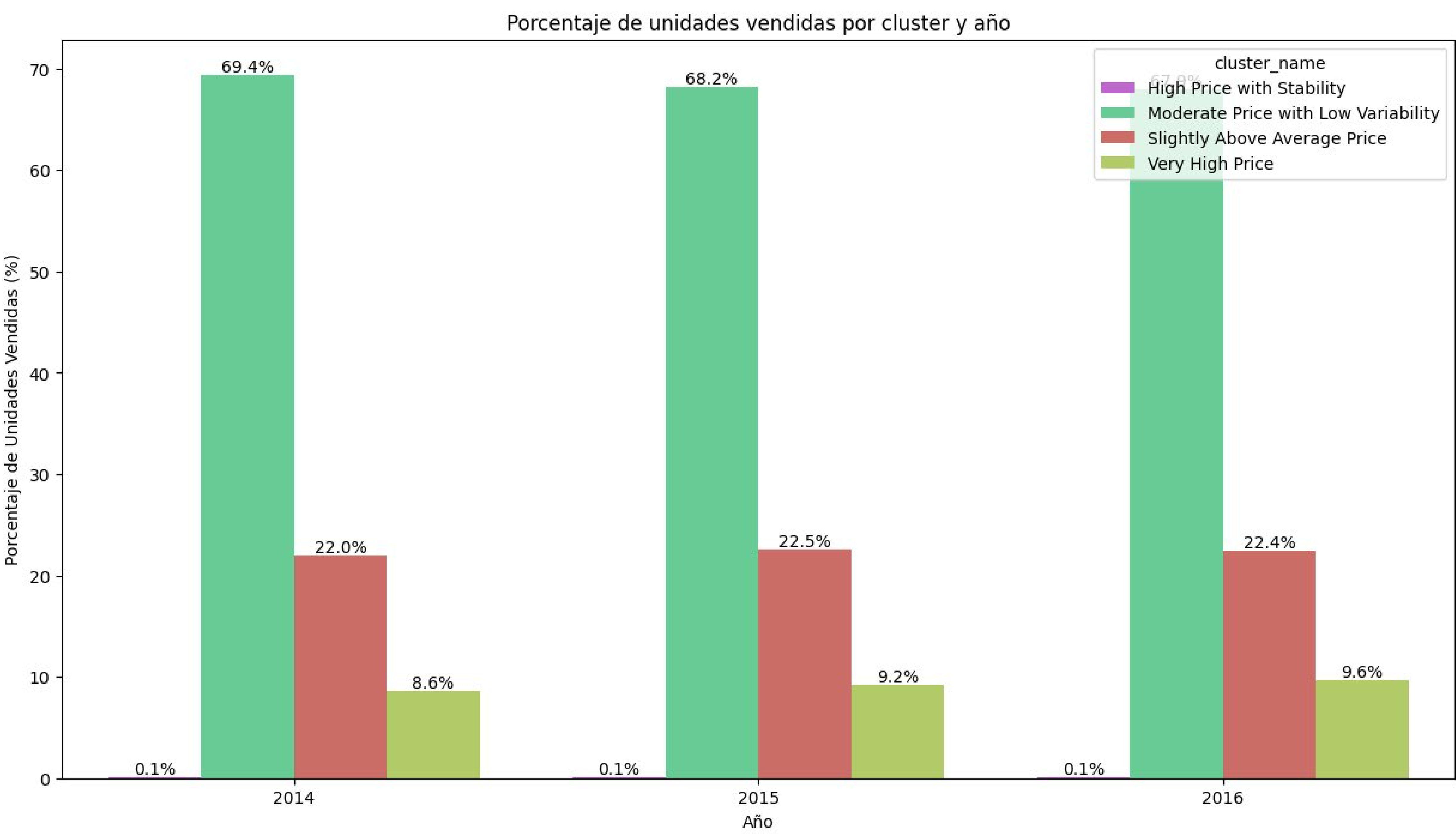
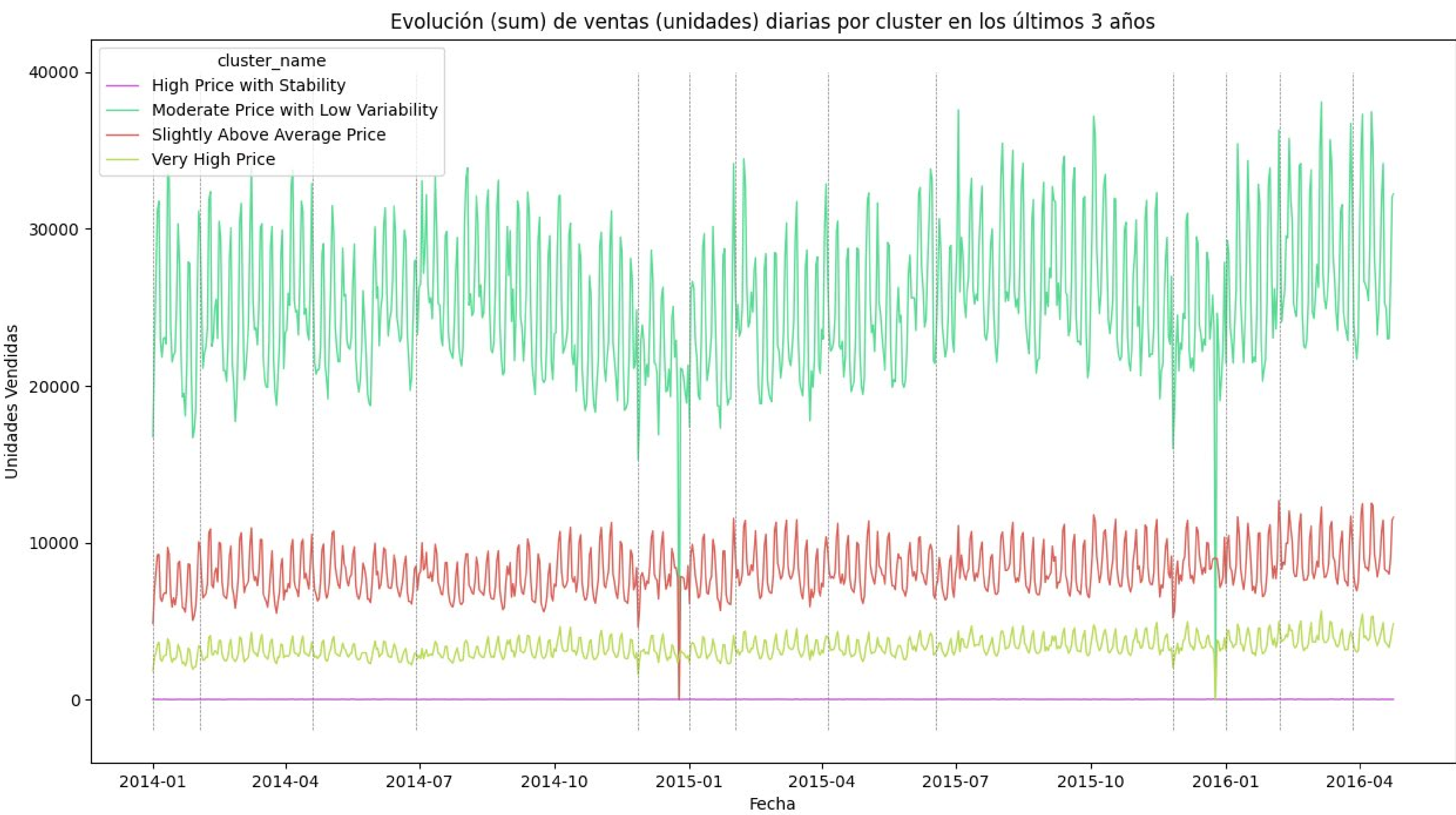

3

RANGO DE PRECIOS

units

file/

items_clusters_price_range.csv



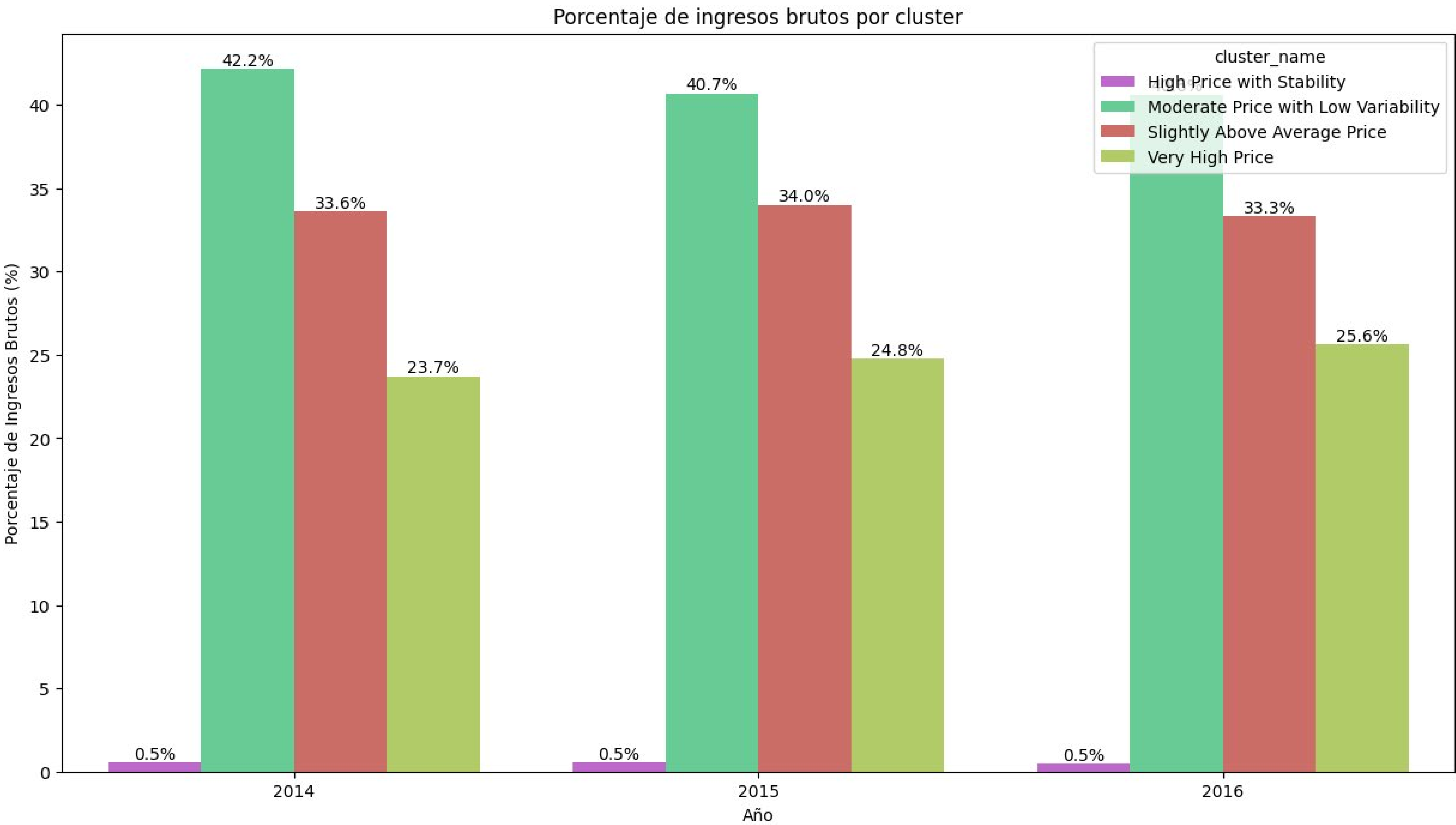
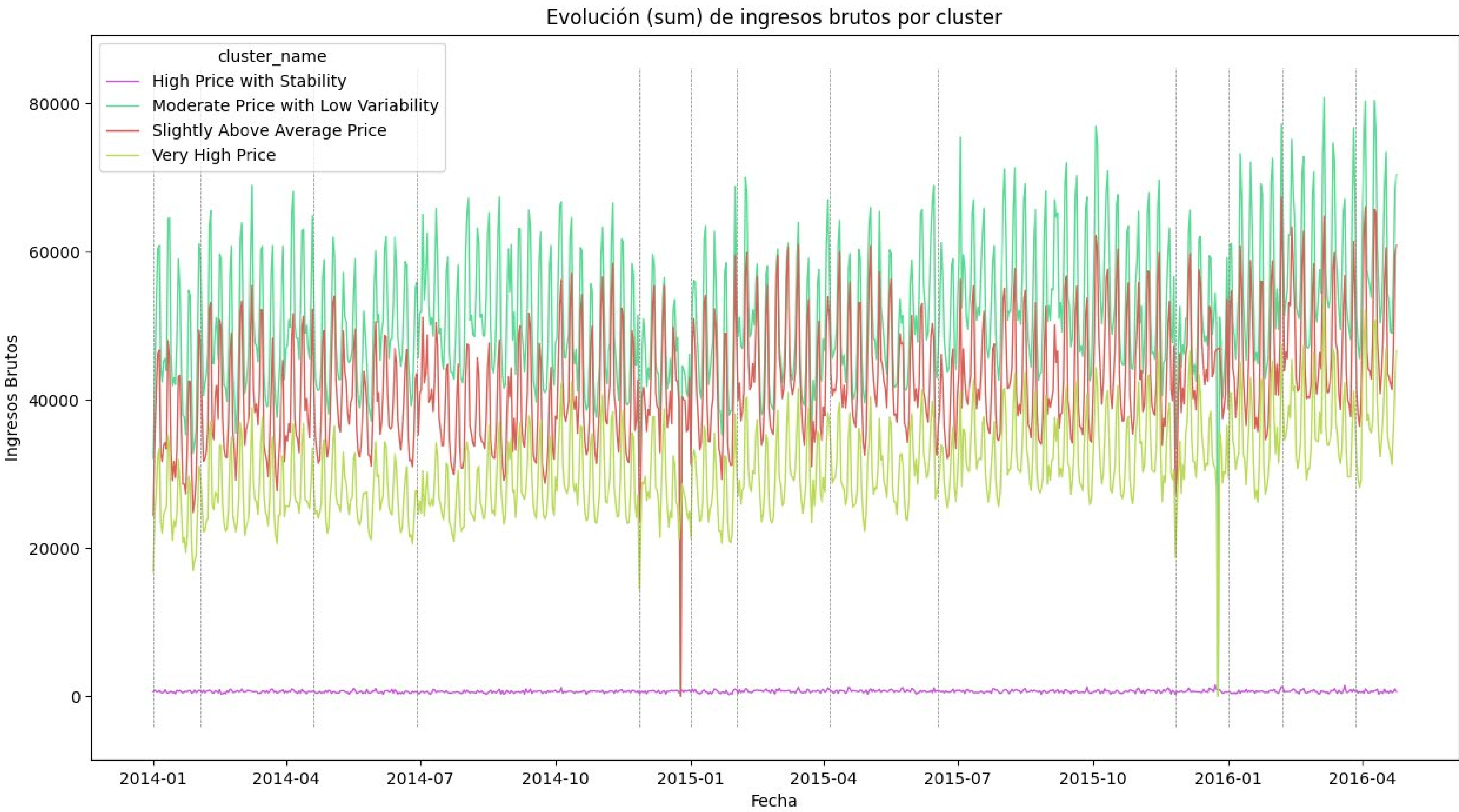
3

RANGO DE PRECIOS

raw_earn

file/

items_clusters_price_range.csv



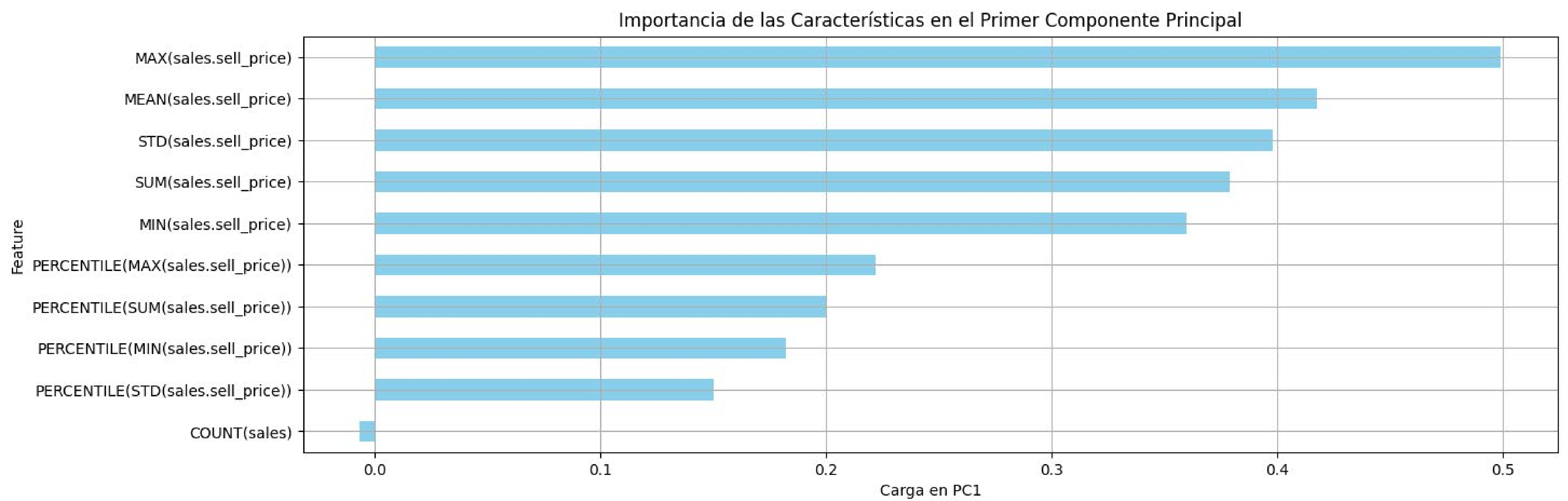
3

RANGO DE PRECIOS

file/

items_clusters_price_range.csv

| | 3 | 1 | 0 | 2 |
|-----------------------------------|----------|-----------|----------|----------|
| MAX(sales.sell_price) | 0.609491 | -0.321900 | 2.289315 | 7.288280 |
| MEAN(sales.sell_price) | 0.580943 | -0.303939 | 2.142619 | 5.013687 |
| STD(sales.sell_price) | 0.439989 | -0.180353 | 1.854049 | 6.662186 |
| SUM(sales.sell_price) | 0.583073 | -0.296628 | 2.093163 | 3.764488 |
| MIN(sales.sell_price) | 0.557055 | -0.155979 | 1.942822 | 4.749228 |
| SUM(sales.PERCENTILE(sell_price)) | 0.573927 | -0.328042 | 1.013351 | 0.741601 |
| MAX(sales.PERCENTILE(sell_price)) | 0.412539 | -0.439736 | 0.801154 | 0.898683 |
| PERCENTILE(SUM(sales.sell_price)) | 0.389426 | -0.420222 | 0.810597 | 0.929562 |
| MIN(sales.PERCENTILE(sell_price)) | 0.467907 | -0.163330 | 1.070628 | 1.285097 |
| PERCENTILE(STD(sales.sell_price)) | 0.196571 | -0.262381 | 0.626702 | 0.970073 |





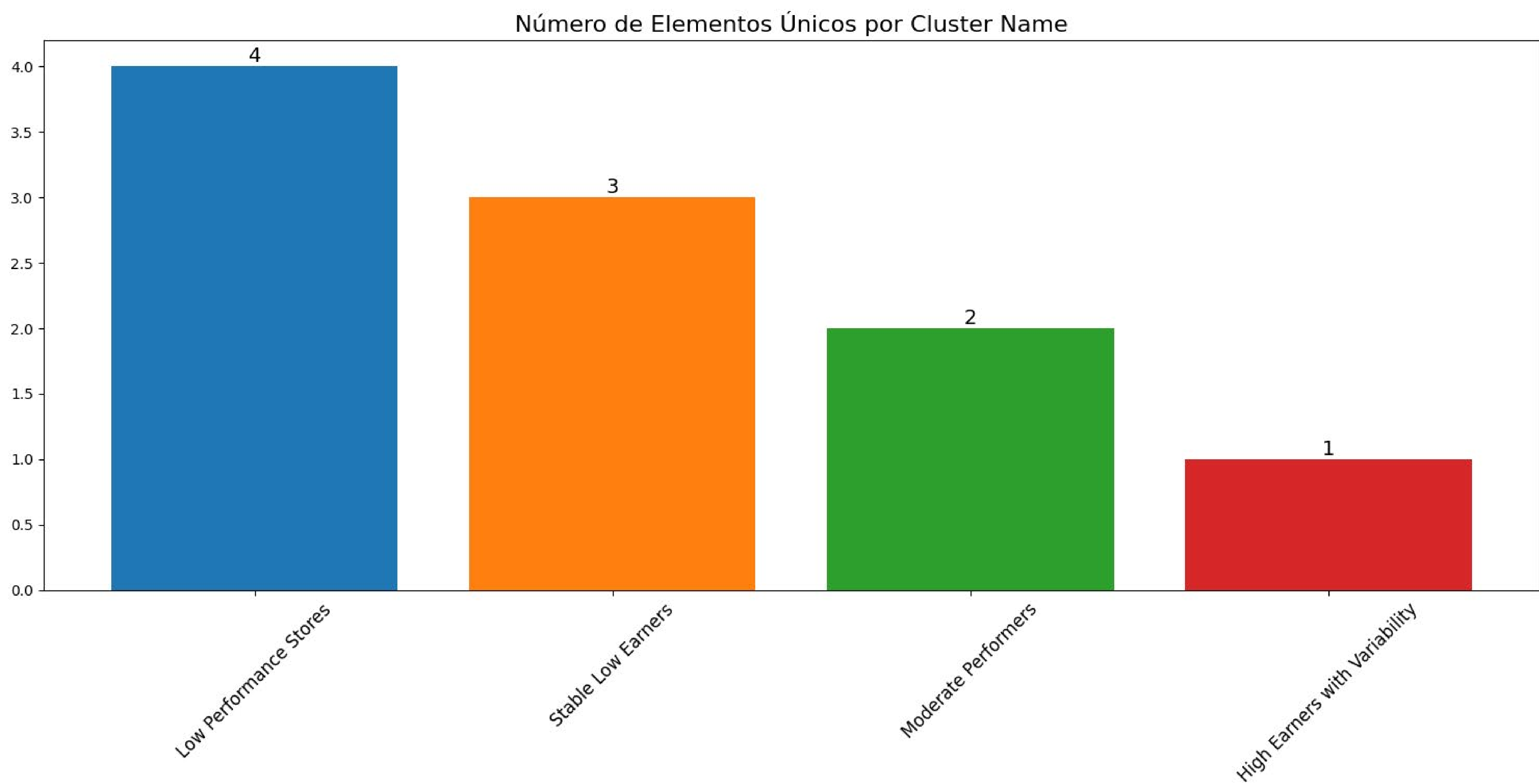
CLUSTERING
TIENDAS

4

PERFORMANCE GENERAL VENTAS

file/

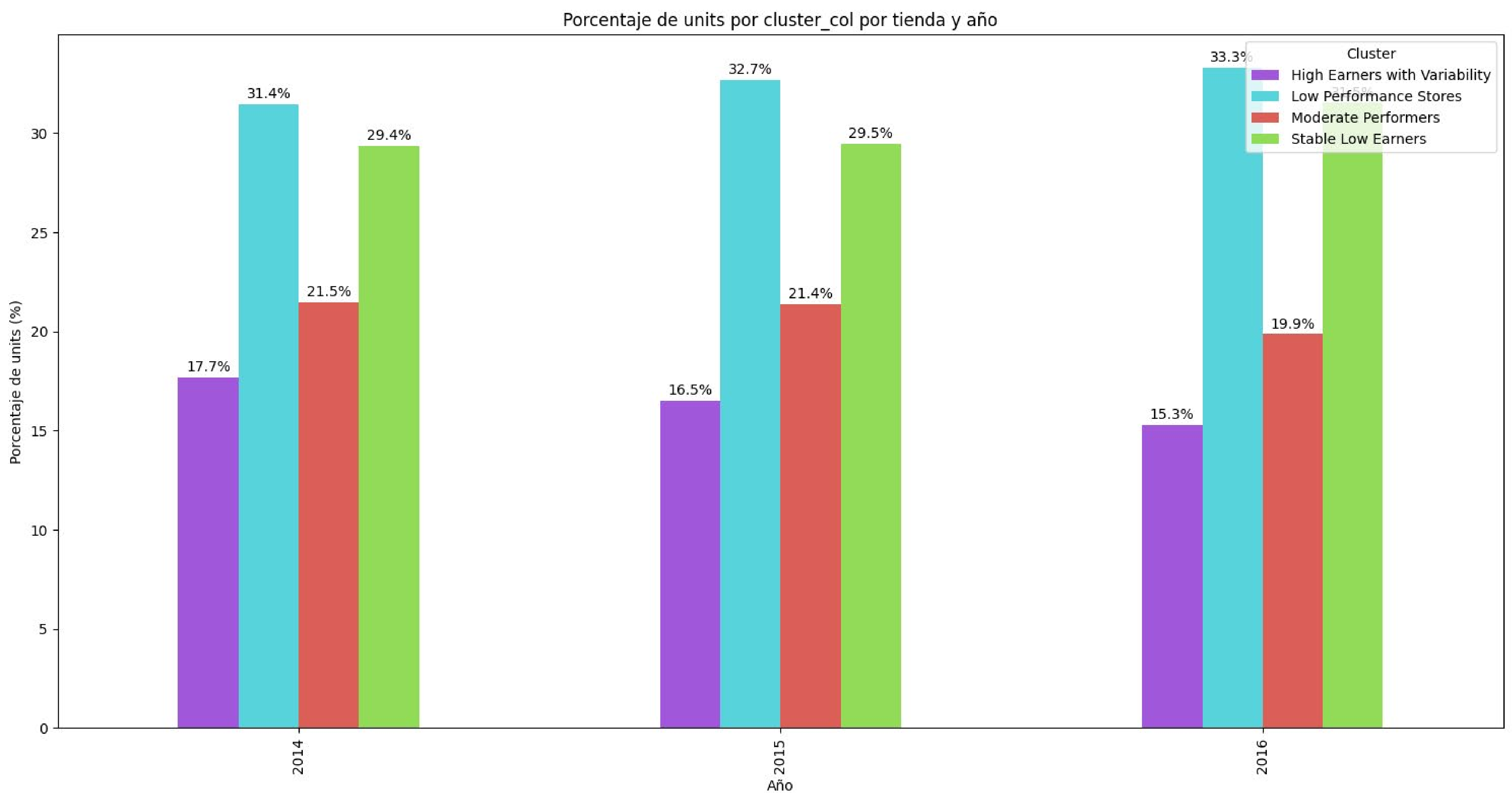
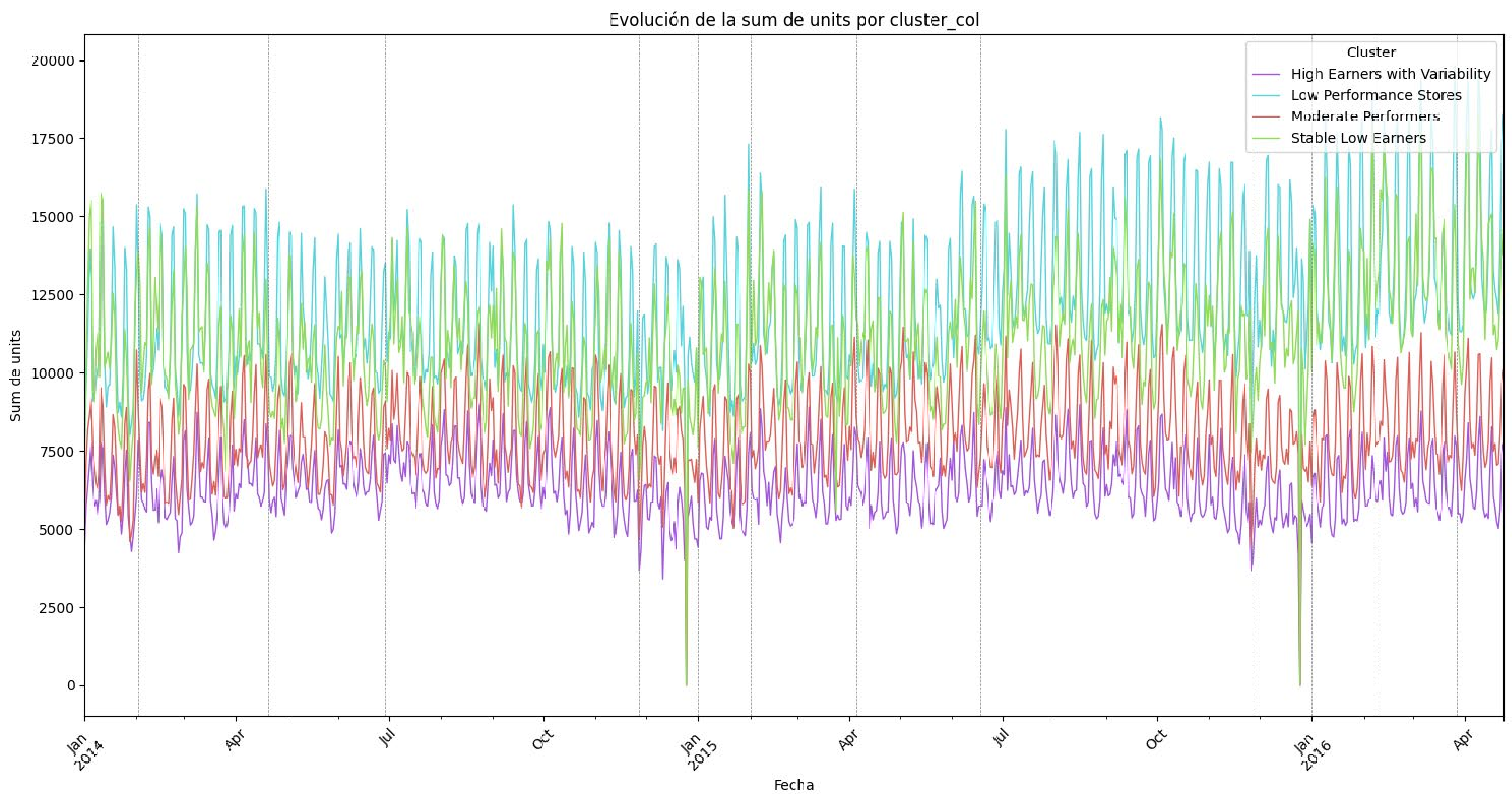
shops_global_sales_performance.csv



```
0: {
  'cluster_name': 'Low Performance Stores',
  'description': 'Tiendas con ingresos consistentemente bajos y una desviación estándar relativamente alta, lo que indica variabilidad en sus ventas. Estas tiendas tienen ingresos máximos por debajo del promedio, lo que sugiere que no tienen picos de ventas significativos.'
},
1: {
  'cluster_name': 'Stable Low Earners',
  'description': 'Tiendas con ingresos bajos pero estables, como lo muestra la baja desviación estándar. Aunque los ingresos máximos y promedio son bajos, la estabilidad indica una clientela constante, sin grandes fluctuaciones.'
},
2: {
  'cluster_name': 'High Earners with Variability',
  'description': 'Tiendas con ingresos promedios altos y picos significativos de ventas, pero con una considerable variabilidad en sus ingresos, lo que sugiere dependencia de ciertos eventos o productos de alto rendimiento.'
},
3: {
  'cluster_name': 'Moderate Performers',
  'description': 'Tiendas con ingresos promedio moderados, junto con una desviación estándar media. Estas tiendas muestran una variabilidad limitada en sus ventas, lo que sugiere que tienen un rendimiento más predecible.'
}
```

4 **PERFORMANCE GENERAL VENTAS**
unidades

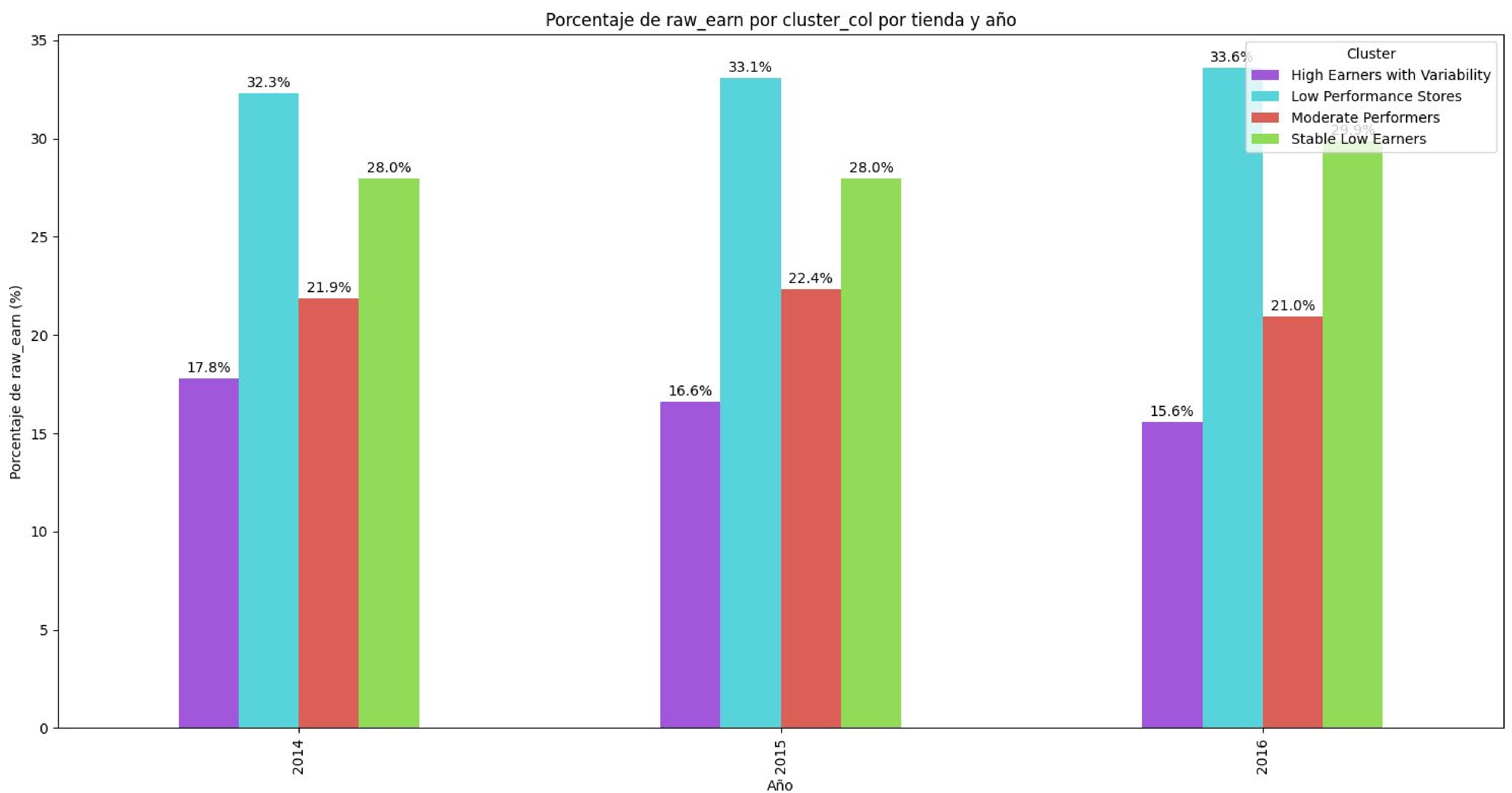
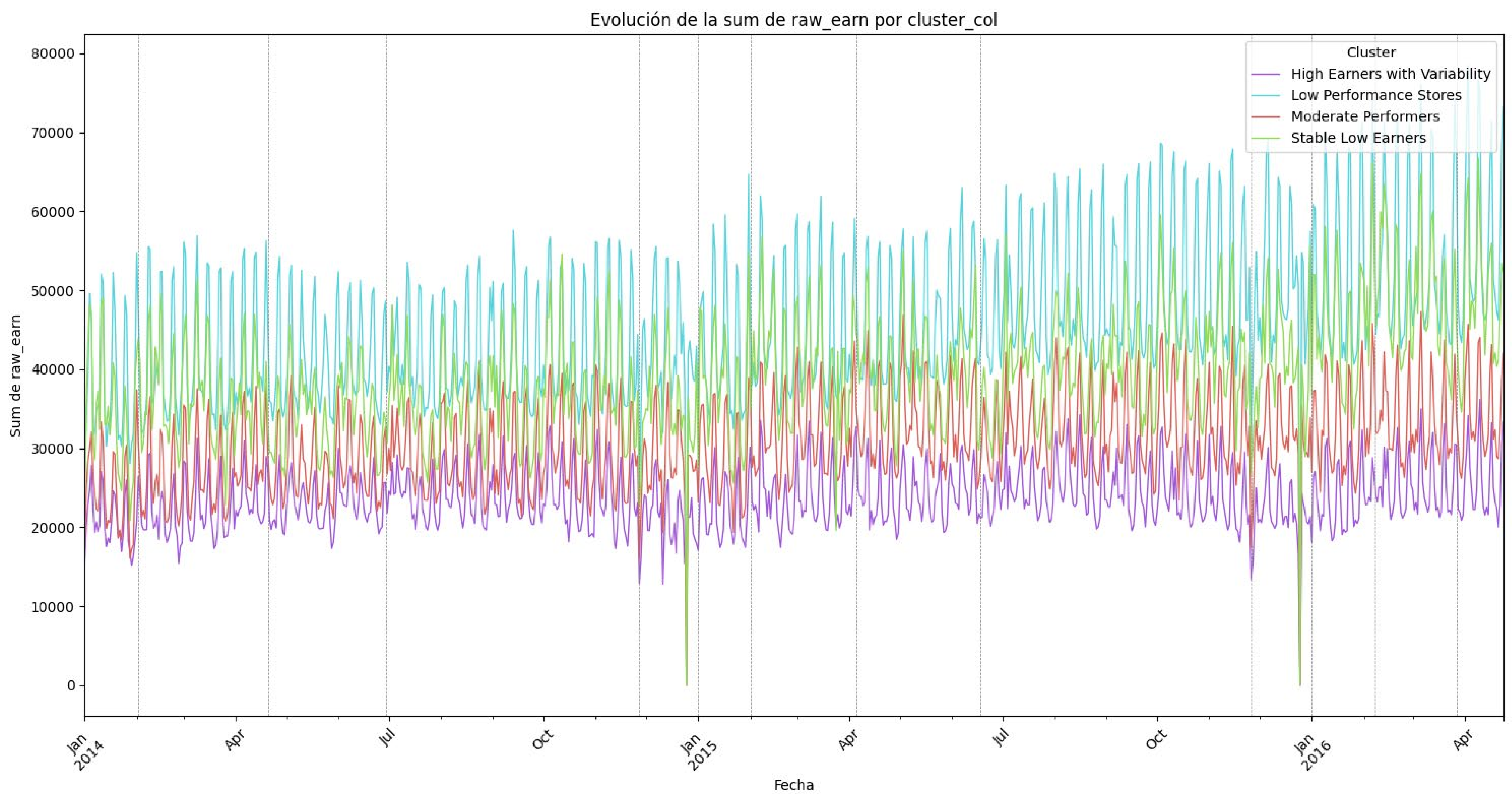
file/
shops_global_sales_performance.csv



4 **PERFORMANCE GENERAL VENTAS**

raw_earn

file/
shops_global_sales_performance.csv



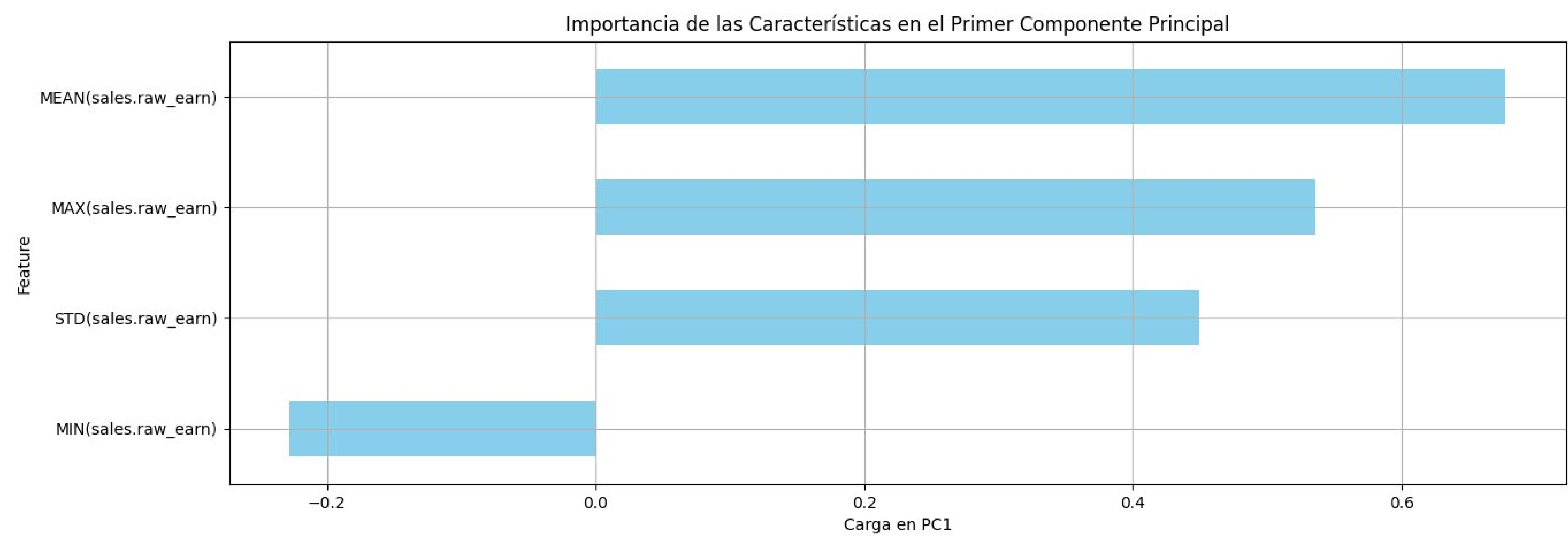
4

PERFORMANCE GENERAL VENTAS

file/

shops_global_sales_performance.csv

| | 0 | 1 | 2 | 3 |
|----------------------|-----------|----------|----------|----------|
| MEAN(sales.raw_earn) | -0.904585 | 0.060299 | 3.465769 | 0.431925 |
| MAX(sales.raw_earn) | -0.531259 | 0.145151 | 2.622752 | 1.234718 |
| STD(sales.raw_earn) | -0.948582 | 0.096116 | 1.627404 | 0.574433 |



5

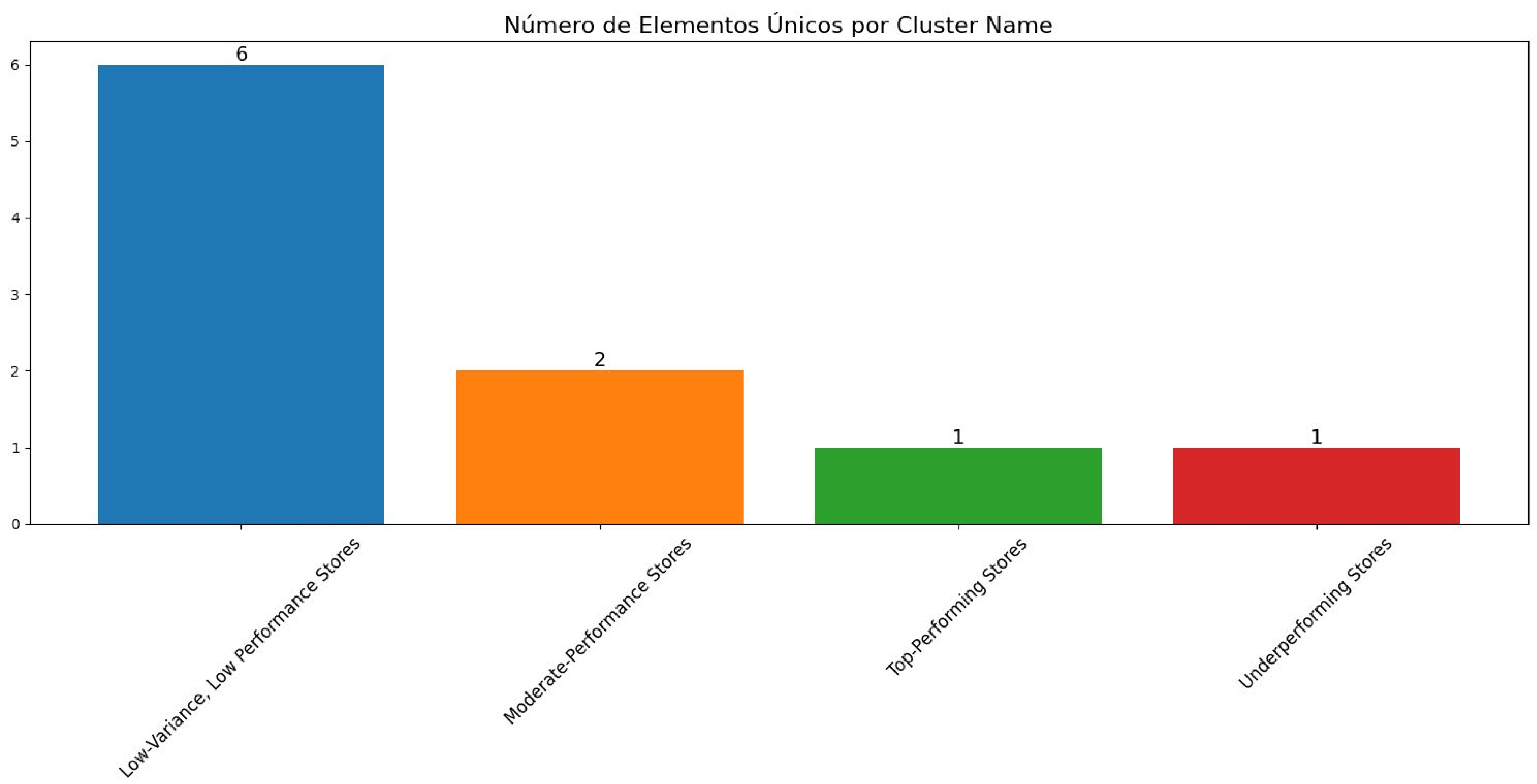
MIX DE PRODUCTOS

file/

shops_products_mix.csv

Esta configuración se centra en entender la variedad y popularidad de los productos en diferentes tiendas.

Problema dimensionalidad.
Utilizamos los clusters creados en la segmentación de productos por performance en ventas globales para reducir la dimensionalidad de los datos y ver como funciona esa segmentación a las tiendas.

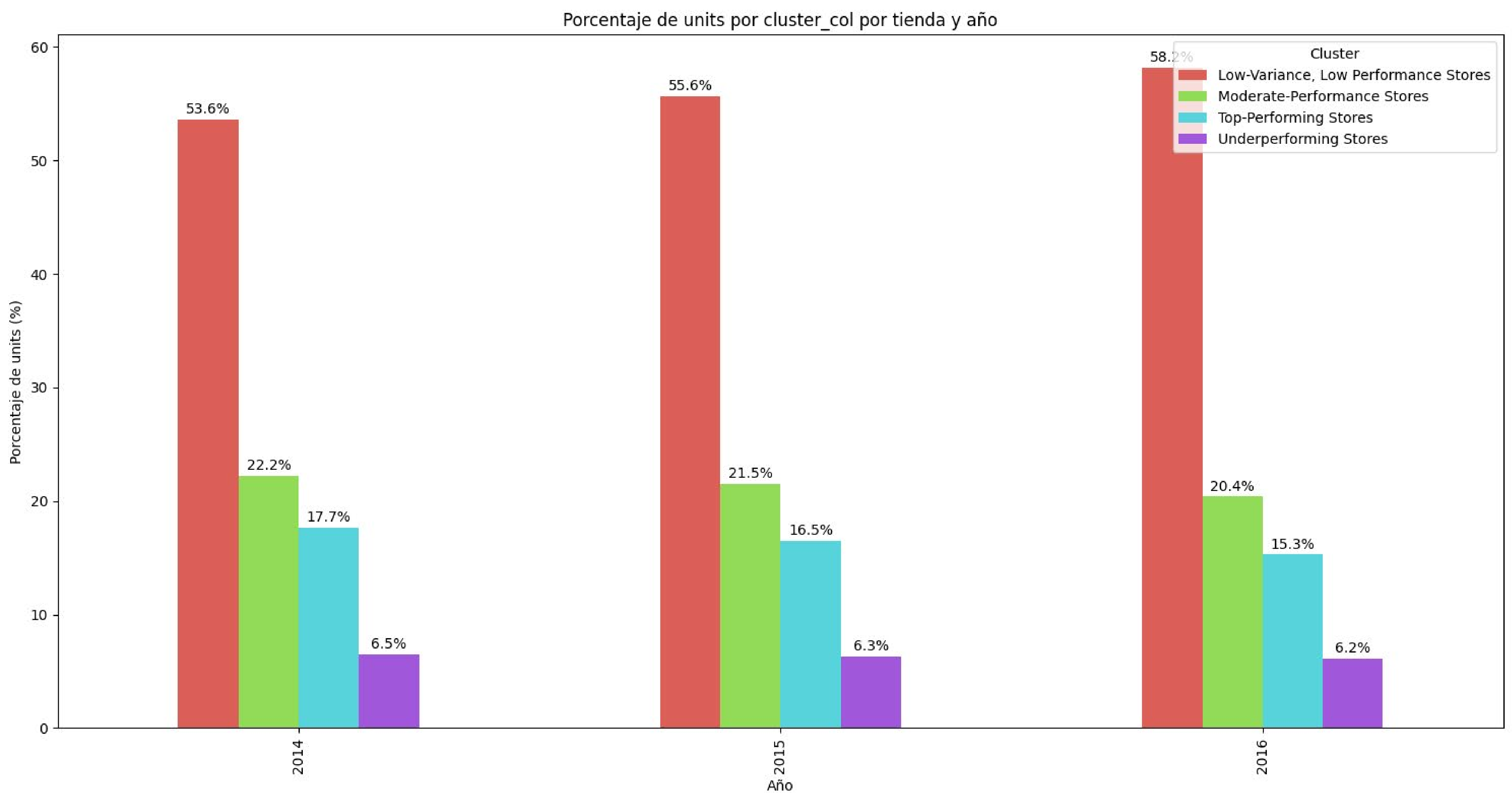
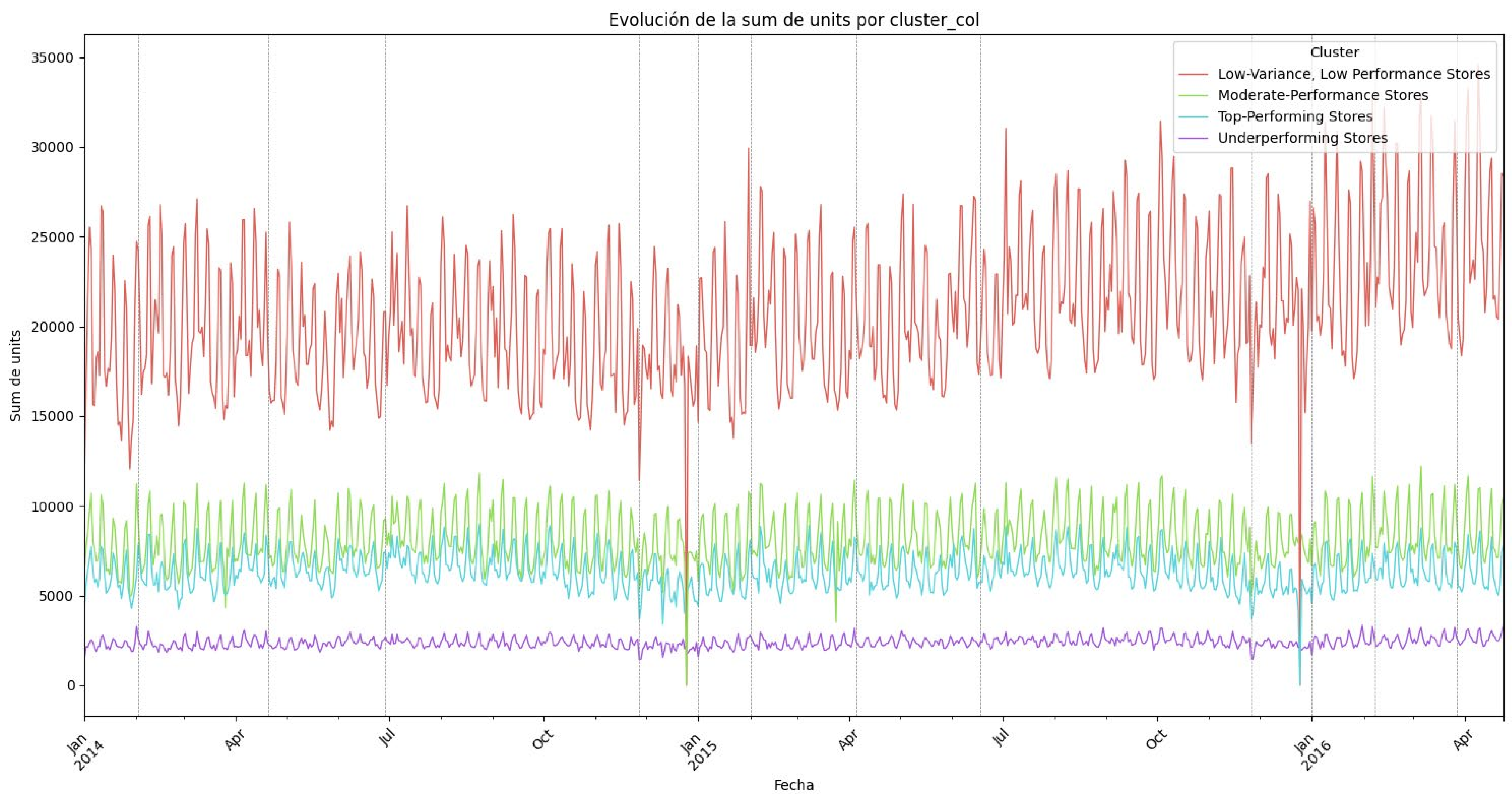


- 0: { 'cluster_name': 'Underperforming Stores',
 'description': 'Tiendas con bajo desempeño en ventas, tanto en ingresos como en la venta de productos de alto rendimiento. Estas tiendas tienen un precio de venta promedio cercano al promedio, pero muestran una fuerte caída en la venta de productos premium y de alto rendimiento.' }
- 1: { 'cluster_name': 'Top-Performing Stores',
 'description': 'Tiendas con el mejor desempeño en ventas, generando ingresos y vendiendo productos de alto rendimiento en grandes cantidades. Estas tiendas también muestran un excelente desempeño en la venta de productos con alta variabilidad, lo que sugiere un mix diverso de productos.' },
- 2: { 'cluster_name': 'Moderate-Performance Stores',
 'description': 'Tiendas con un desempeño moderado en ventas, con ingresos y precios de venta cercanos al promedio. Estas tiendas venden tanto productos de alto rendimiento como productos con variabilidad moderada, lo que indica un mix de productos equilibrado.' },
- 3: { 'cluster_name': 'Low-Variance, Low Performance Stores',
 'description': 'Tiendas con bajo rendimiento en ventas y con una baja variabilidad en la venta de productos. Estas tiendas no dependen de productos de alto rendimiento y muestran una estabilidad general en sus ventas, aunque a un nivel bajo.' }

5 MIX DE PRODUCTOS

units

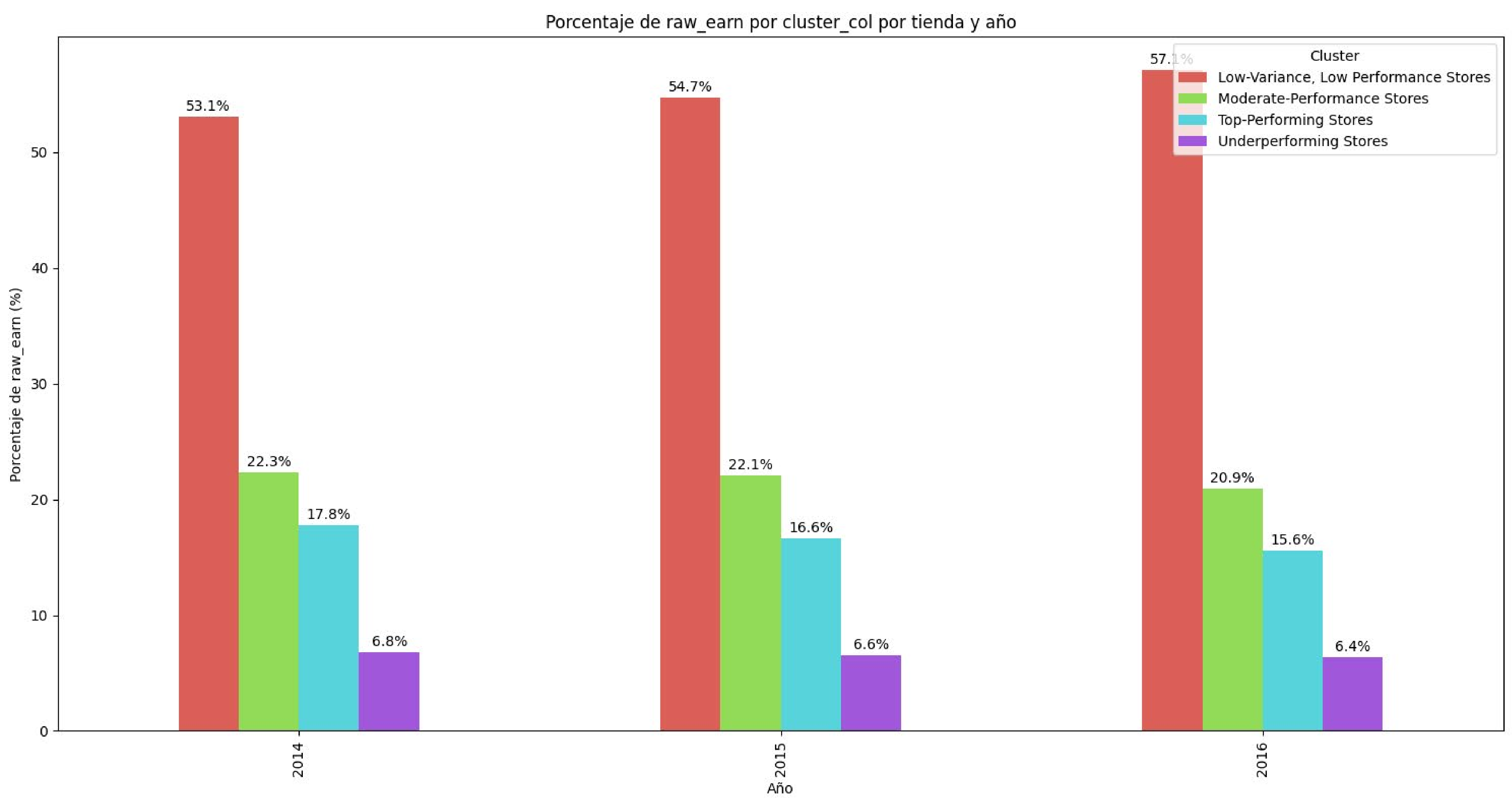
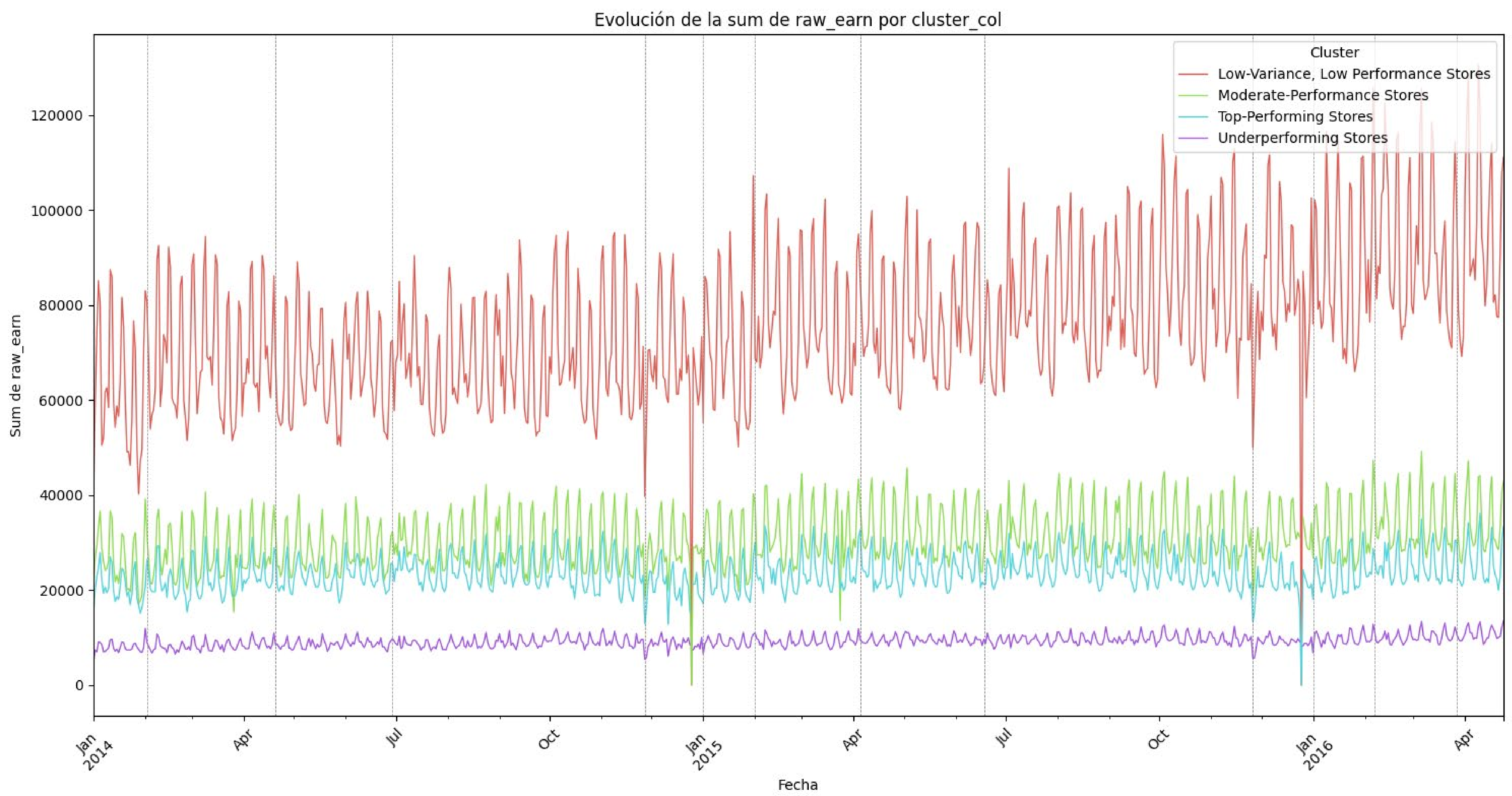
file/
shops_products_mix.csv



5 MIX DE PRODUCTOS

raw_earn

file/
shops_products_mix.csv



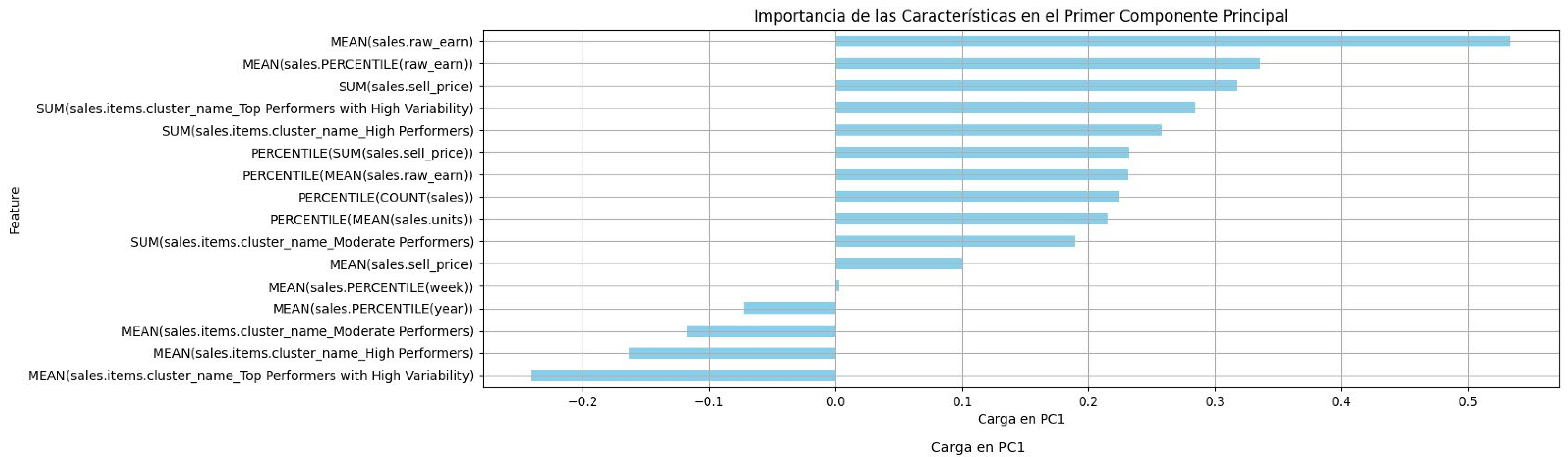
5

MIX DE PRODUCTOS

file/

shops_products_mix.csv

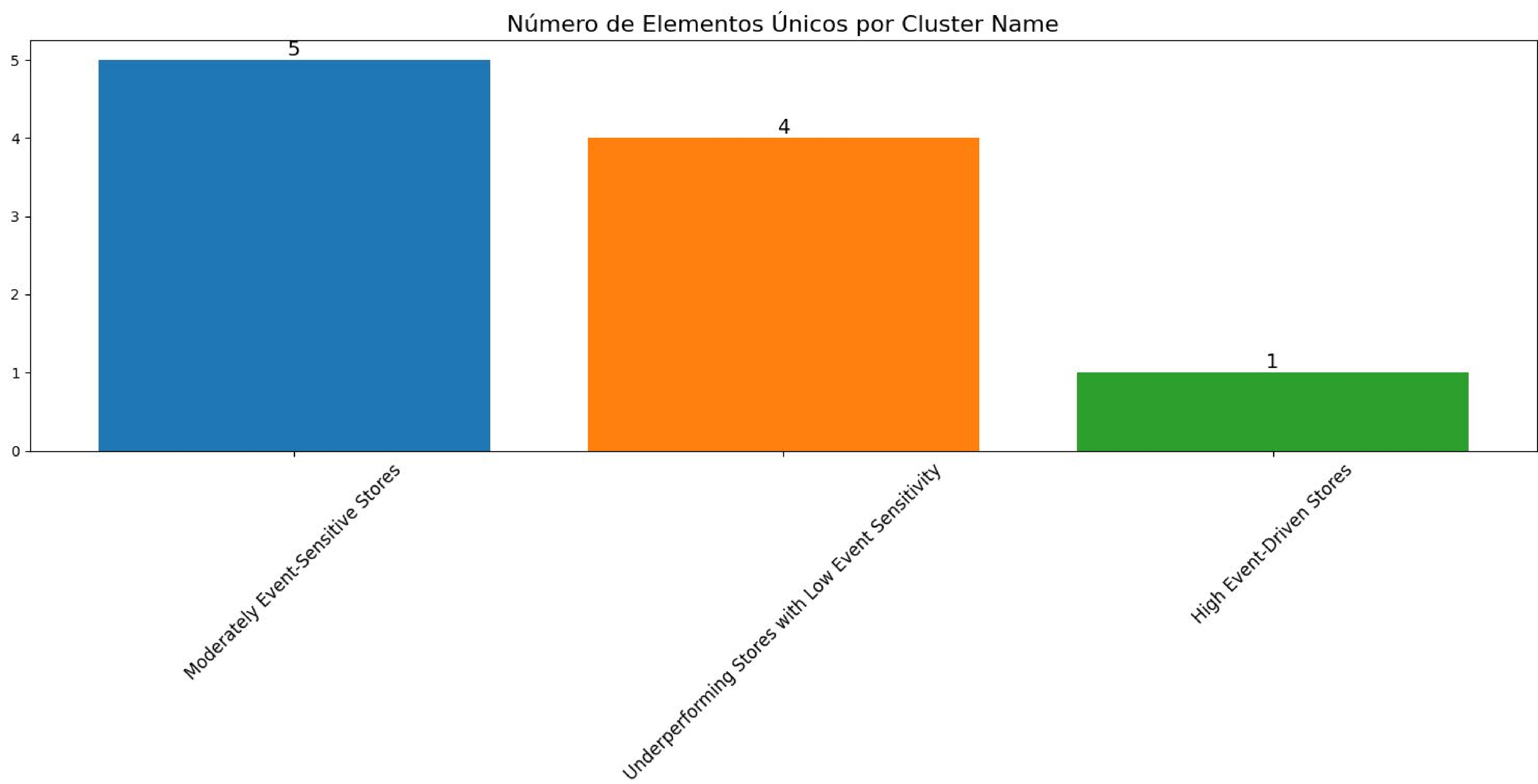
| | 0 | 1 | 2 | 3 |
|--|-----------|----------|----------|-----------|
| MEAN(sales.raw_earn) | -1.800360 | 3.465769 | 0.650335 | -0.345651 |
| MEAN(sales.PERCENTILE(raw_earn)) | -0.951202 | 2.150999 | 0.686163 | -0.080033 |
| SUM(sales.sell_price) | -0.201956 | 1.890619 | 1.236033 | -0.095482 |
| SUM(sales.items.cluster_name_Top Performers with High Variability) | -0.217742 | 2.927419 | 0.250000 | 0.139785 |
| SUM(sales.items.cluster_name_High Performers) | -0.387300 | 1.464535 | 0.306688 | -0.556932 |
| PERCENTILE(SUM(sales.sell_price)) | -0.555556 | 1.000000 | 0.666667 | -0.296296 |
| PERCENTILE(MEAN(sales.raw_earn)) | -1.000000 | 1.000000 | 0.666667 | -0.222222 |
| PERCENTILE(COUNT(sales)) | -0.555556 | 1.000000 | 0.666667 | -0.296296 |
| PERCENTILE(MEAN(sales.units)) | -1.000000 | 1.000000 | 0.555556 | -0.185185 |
| SUM(sales.items.cluster_name_Moderate Performers) | -0.171159 | 0.908577 | 0.540434 | -0.368662 |
| MEAN(sales.sell_price) | 0.069526 | 0.540119 | 0.407123 | 0.030906 |



6 INFLUENCIA EVENTOS EN VENTAS

file/

shops_clusters_events_influence_sales.csv

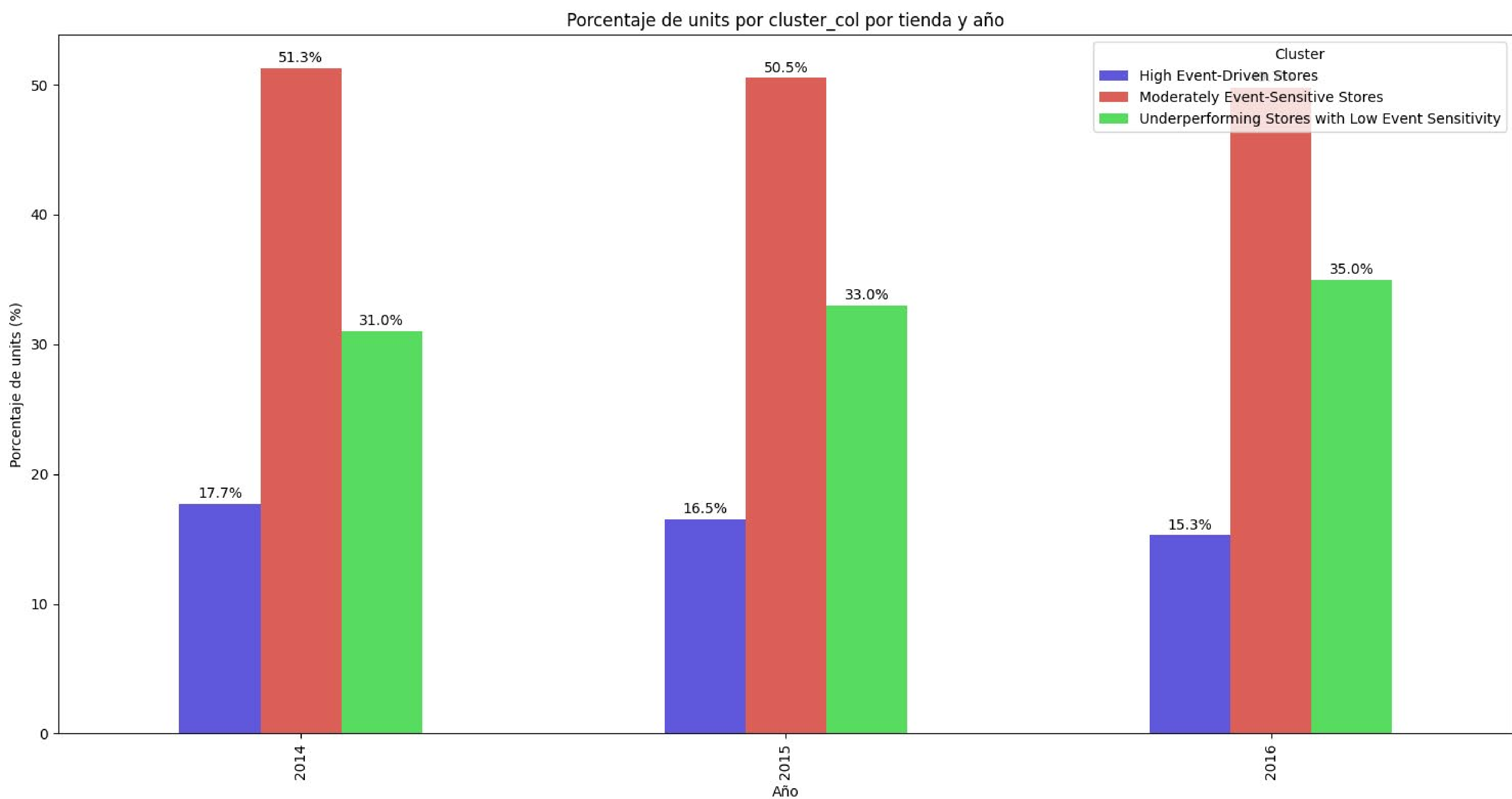
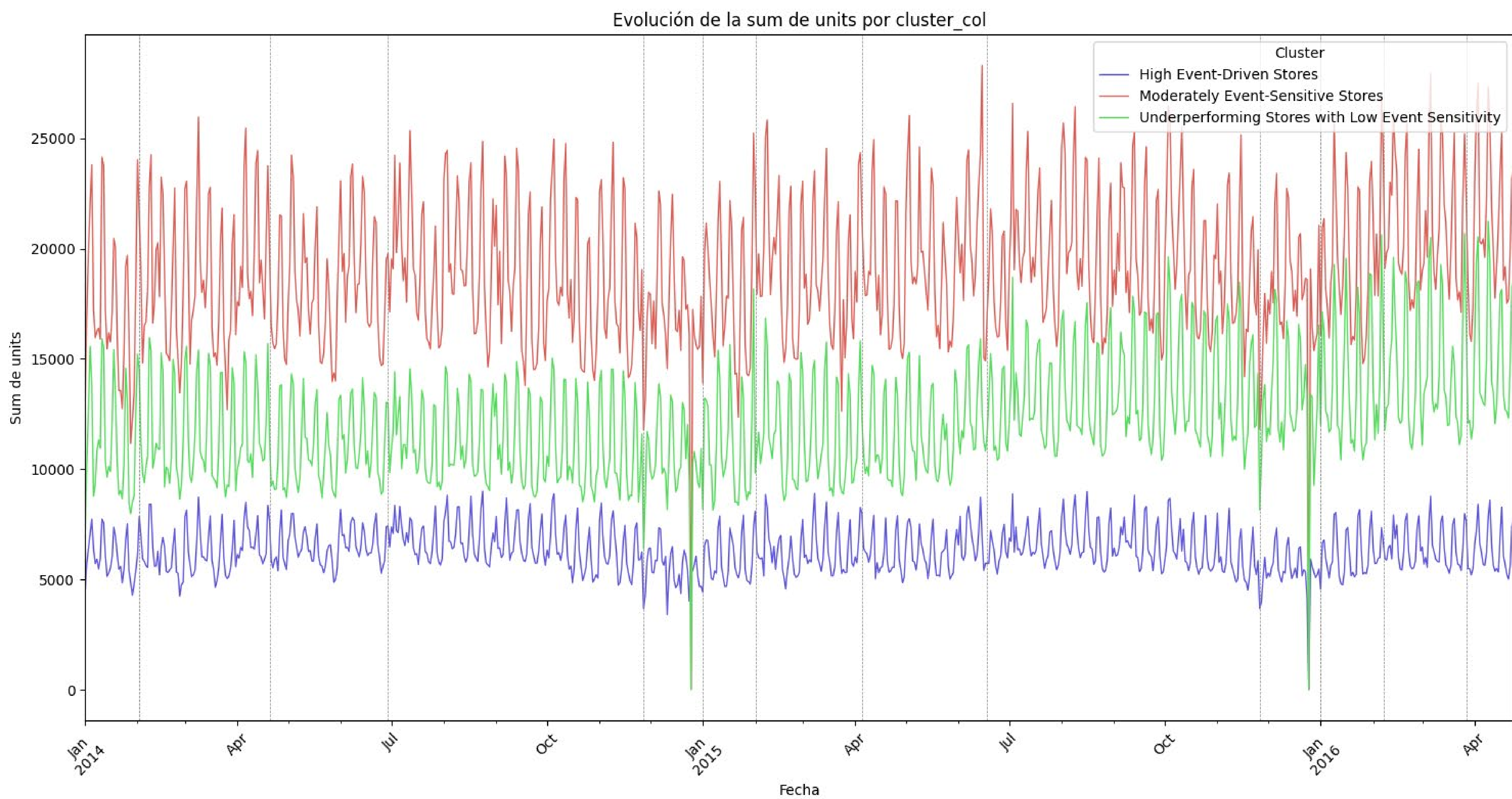


```
0: {
  'cluster_name': 'Underperforming Stores with Low Event Sensitivity',
  'description': 'Tiendas con ingresos bajos que no se benefician significativamente de los eventos. Estas
tiendas presentan una baja varianza en las ventas y no muestran un impulso notable durante eventos impor-
tantes como el Año Nuevo o la Pascua. Son especialmente poco sensibles a eventos no festivos como el inicio
del Ramadán.'
},
1: {
  'cluster_name': 'Moderately Event-Sensitive Stores',
  'description': 'Tiendas con ingresos moderados y una ligera influencia de eventos en las ventas. Estas tien-
das tienen una respuesta decente durante el Año Nuevo y la Pascua, y muestran una estabilidad razonable en
las ventas sin depender exclusivamente de eventos. Aunque no son muy sensibles, reaccionan a algunos even-
tos como el inicio del Ramadán.'
},
2: {
  'cluster_name': 'High Event-Driven Stores',
  'description': 'Tiendas con ingresos altos y una fuerte dependencia de eventos para incrementar sus ven-
tas. Estas tiendas se destacan en eventos como el Año Nuevo y la Pascua, mostrando una alta varianza en las
ventas y un impulso significativo durante estos momentos clave. Sin embargo, no parecen ser sensibles a even-
tos como el inicio del Ramadán.'
}
```

6 INFLUENCIA EVENTOS EN VENTAS

units

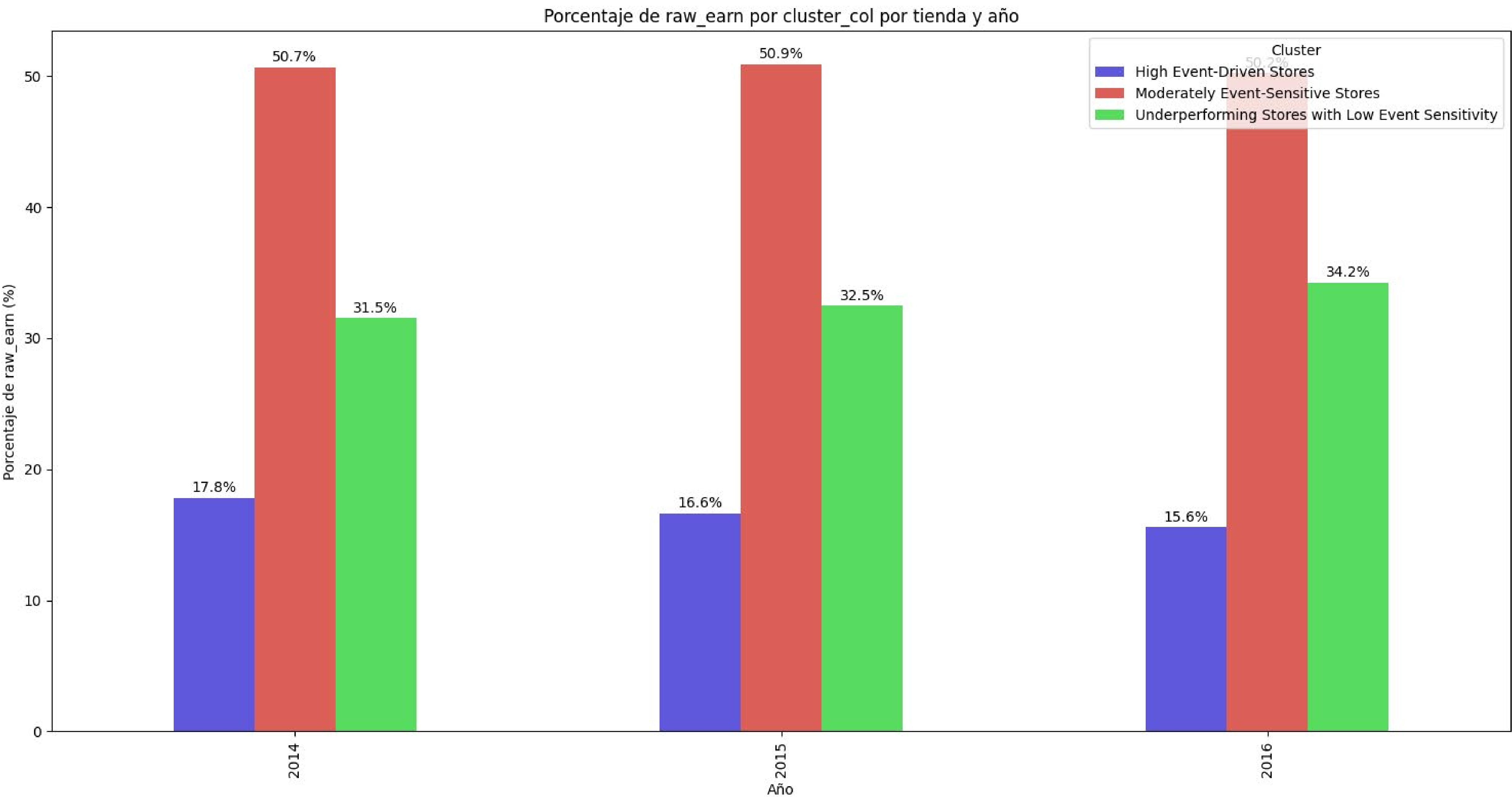
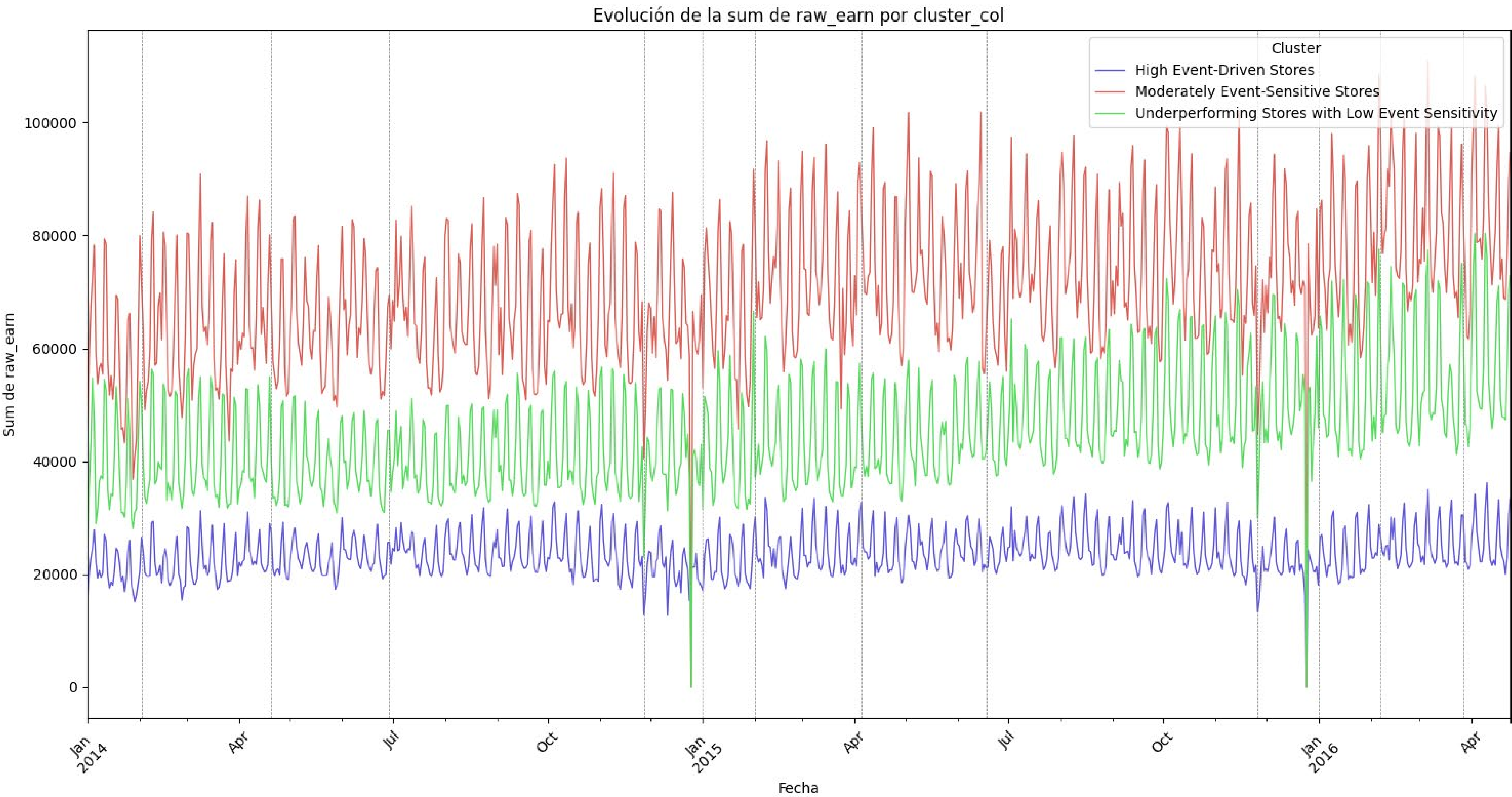
file/
shops_clusters_events_influence_sales.csv



6 INFLUENCIA EVENTOS EN VENTAS

raw_earn

file/
shops_clusters_events_influence_sales.csv



6 INFLUENCIA EVENTOS EN VENTAS

file/
shops_global_sales_performance.csv

| | 0 | 1 | 2 |
|----------------------------------|-----------|----------|-----------|
| MEAN(sales.raw_earn) | -0.796113 | 0.122171 | 3.465769 |
| MAX(sales.raw_earn) | -0.532542 | 0.582004 | 2.622752 |
| MEAN(sales.event_None) | -1.166692 | 0.515101 | 0.118106 |
| STD(sales.raw_earn) | -0.855723 | 0.213156 | 1.627404 |
| MEAN(sales.event_Ramadan starts) | -0.538886 | 0.394197 | -0.290920 |
| SUM(sales.event_Easter) | -0.162817 | 0.186234 | 1.430597 |
| SUM(sales.event_NewYear) | 0.119640 | 0.283925 | 2.015096 |

