Explorando sitios de comercio electrónico

Final AID

Leonardo Aranda

Introducción

- Crecimiento sostenido del comercio electrónico
- ¿Qué características tiene un sitio de elevado tráfico?
- Estimaciones de tráfico: Alexa + SimilarWeb.

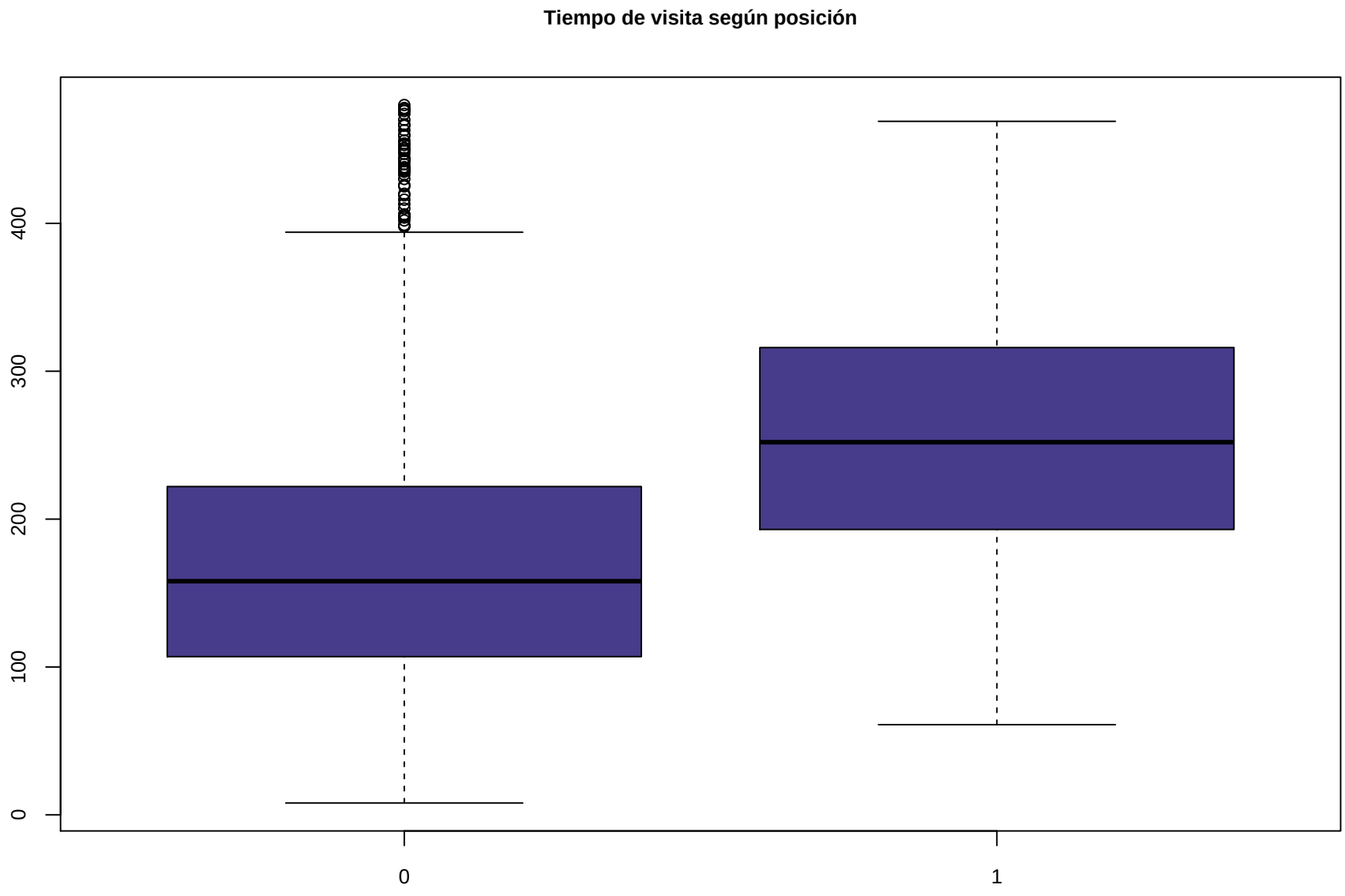
Variables

Variable	Descripción
url	Dirección del sitio web
position	Pertenece al Top 50.000 mundial
bounce	Porcentaje de visitas con una sola página vista
ppv	Páginas vistas por visita
time	Tiempo promedio de la visita
paid_search	Tráfico pago
traffic_direct	Tráfico directo
traffic_display	Tráfico de redes de publicidad
traffic_search	Tráfico de buscadores
traffic_mail	Tráfico de correo electrónico
traffic_social	Tráfico de redes sociales
traffic_referrals	Tráfico referido desde otras páginas

Vector de medias

	V1	V2
Group.1	0	1
bounce	40.48016	43.02896
paid_search	13.92329	18.45615
ppv	4.517698	5.803474
time	172.9875	255.8759
traffic_direct	24.45626	28.58323
traffic_display	0.8779869	1.7365757
traffic_search	54.47384	45.90362
traffic_mail	4.214623	5.198238
traffic_social	3.117430	3.837295
traffic_referrals	12.86008	14.74079

Tiempo de visita - Distribución



Tiempo de visita - Normalidad

- Prueba de Shapiro-Wilk

position	p-value
0	0.0000000
1	0.0240839

- Transformación de Box-Cox

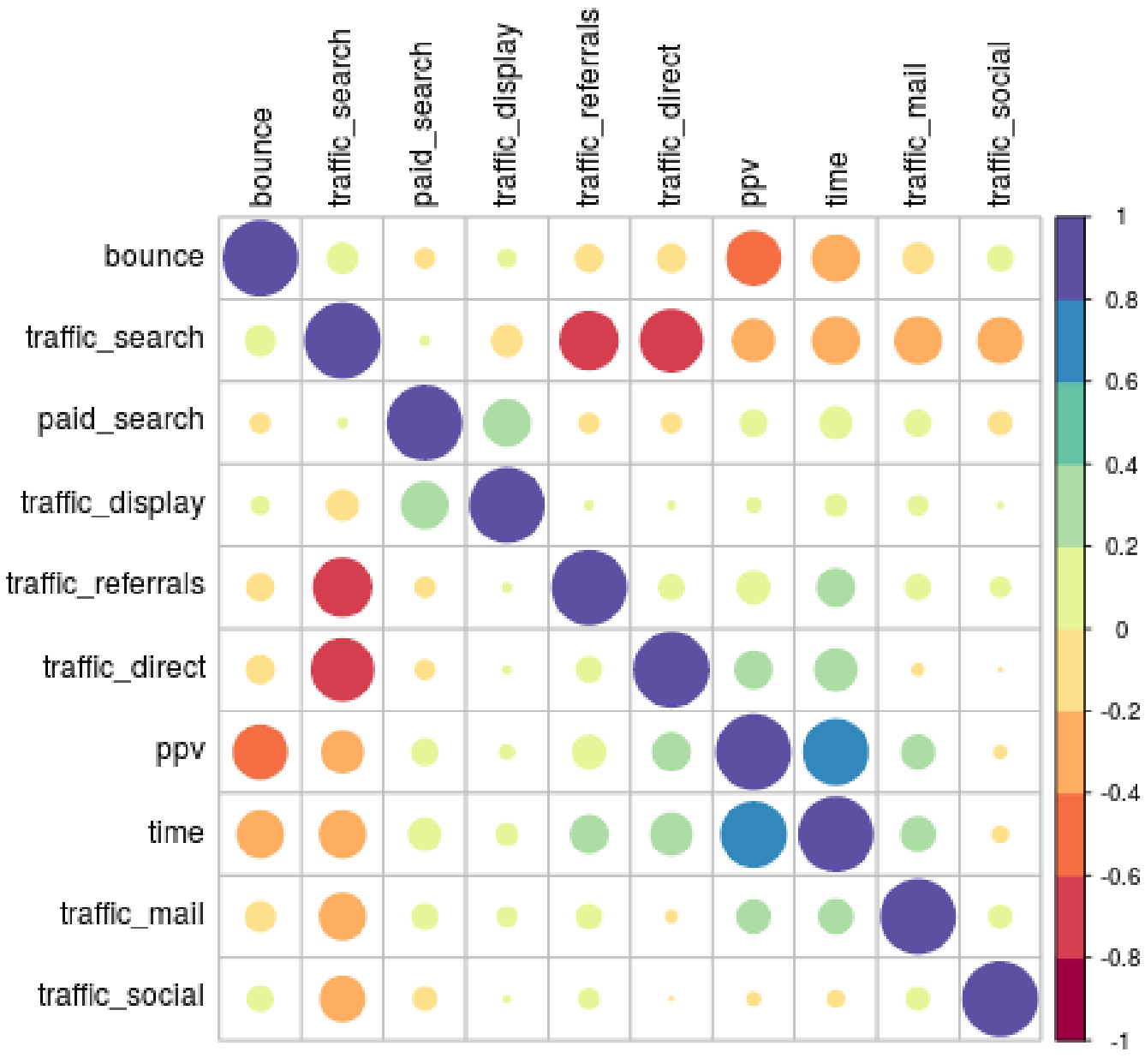
position	p-value
0	0.0002660
1	0.0019657

Tiempo de visita - Prueba de Mann-Whitney-Wilcoxon

- H_0: Los grupos pertenecen a la misma distribución.
- H_1: Existen diferencias significativas en la distribución.

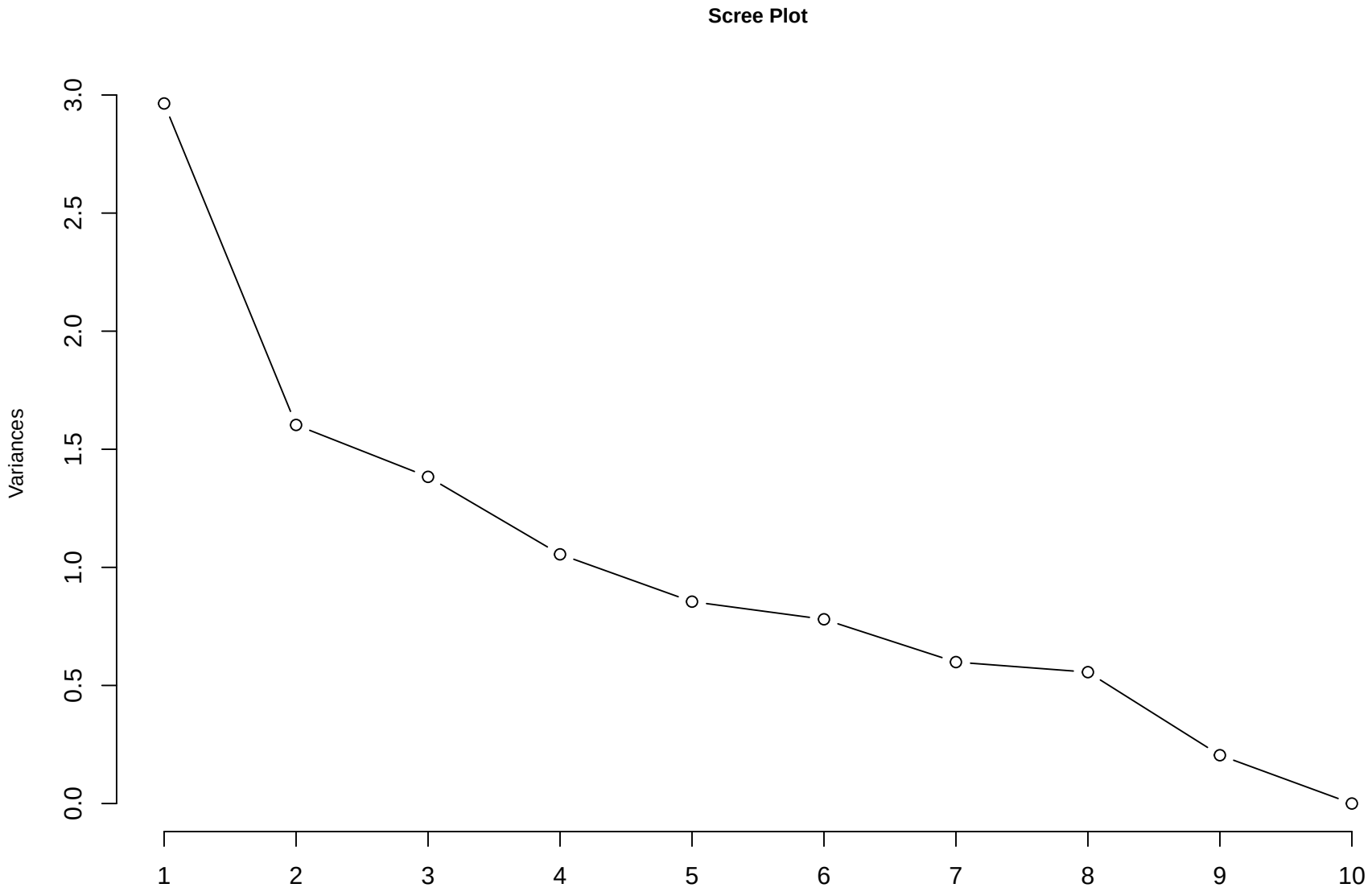
```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  time by position
## W = 148840, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal
```

Matriz de correlación



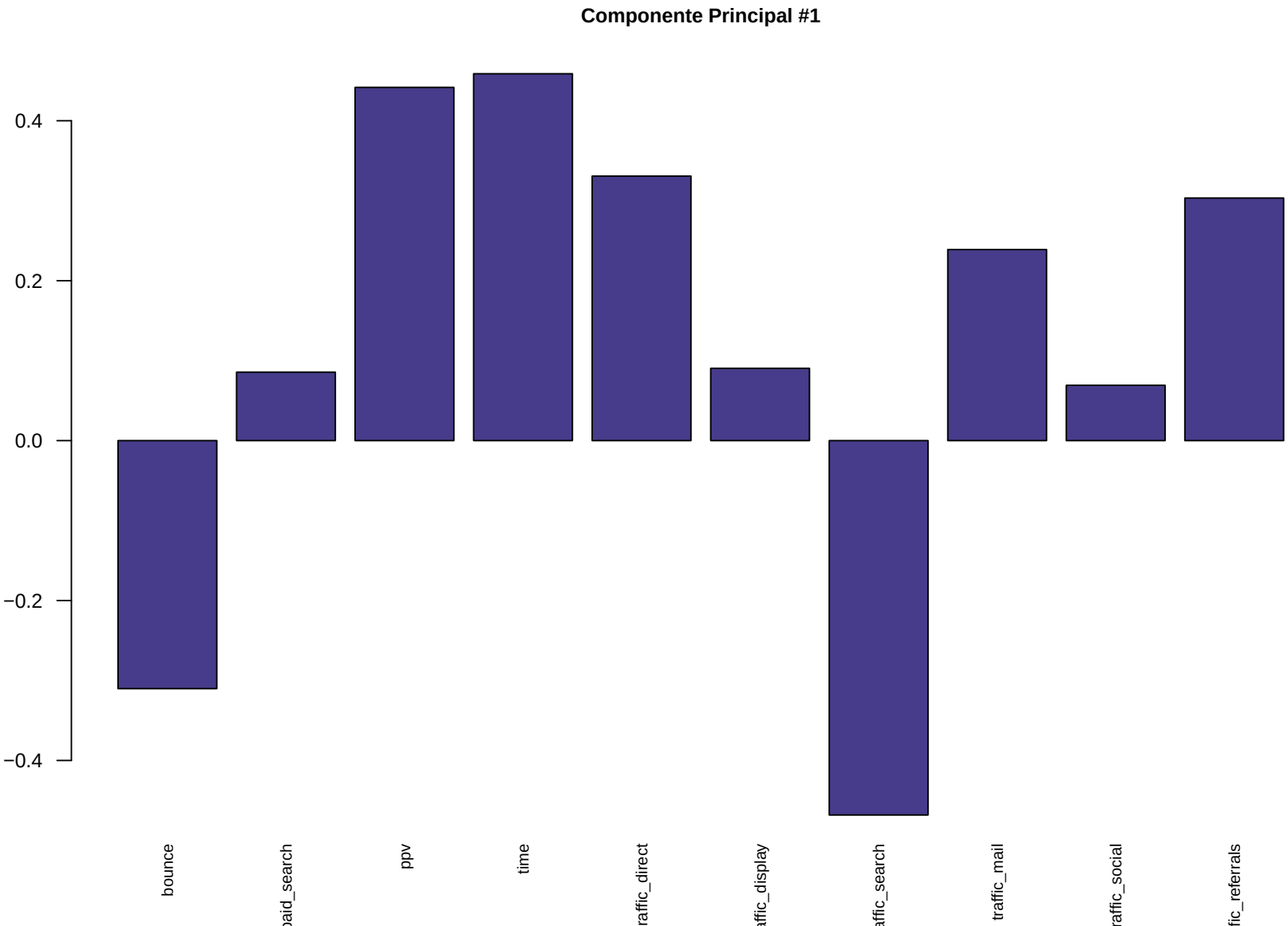
Componentes principales - Scree Plot

- Tres componentes explican el 60% de la variabilidad.



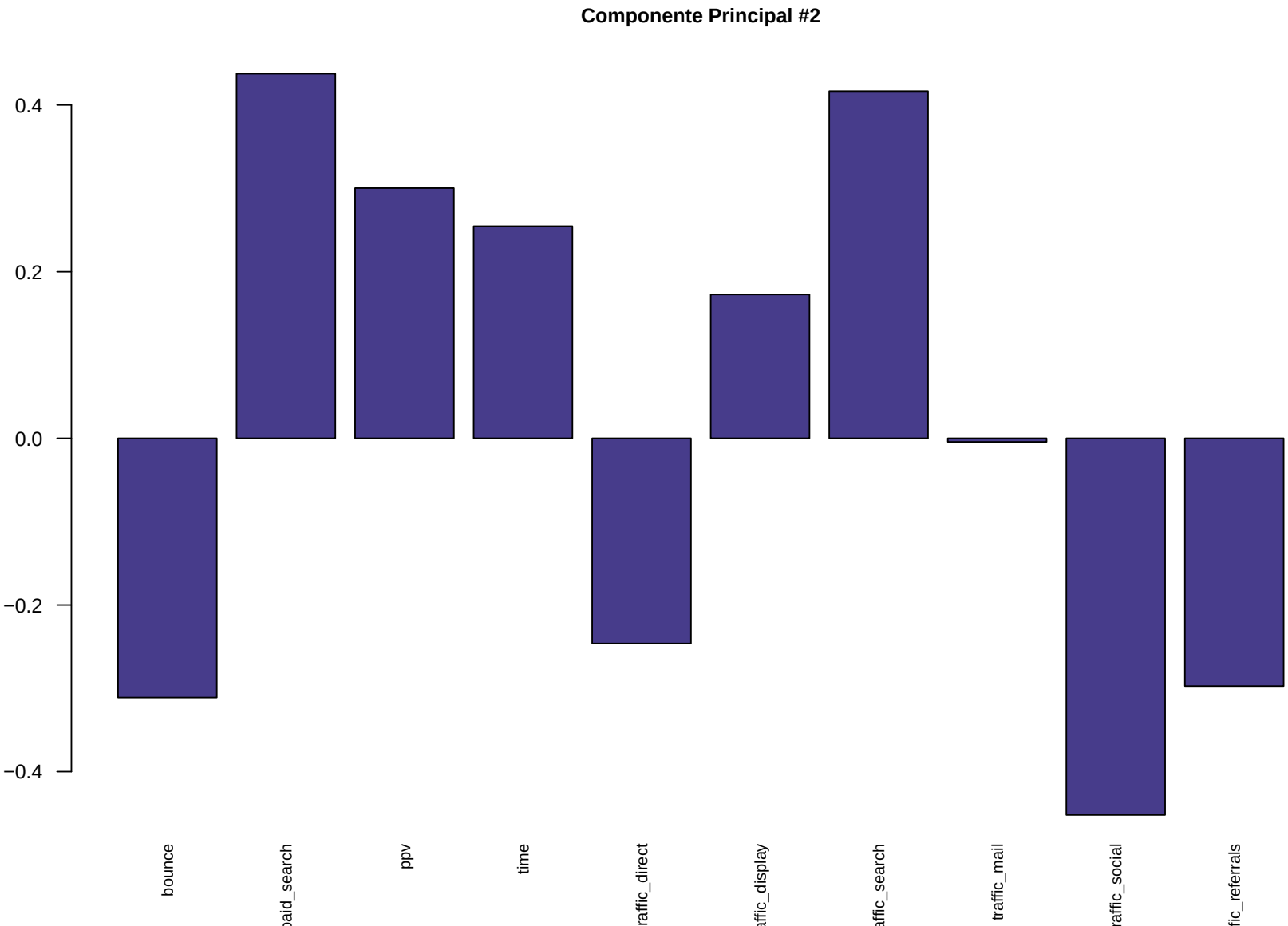
Componentes principales - Componente 1

- Nivel de interacción
- Componente de forma



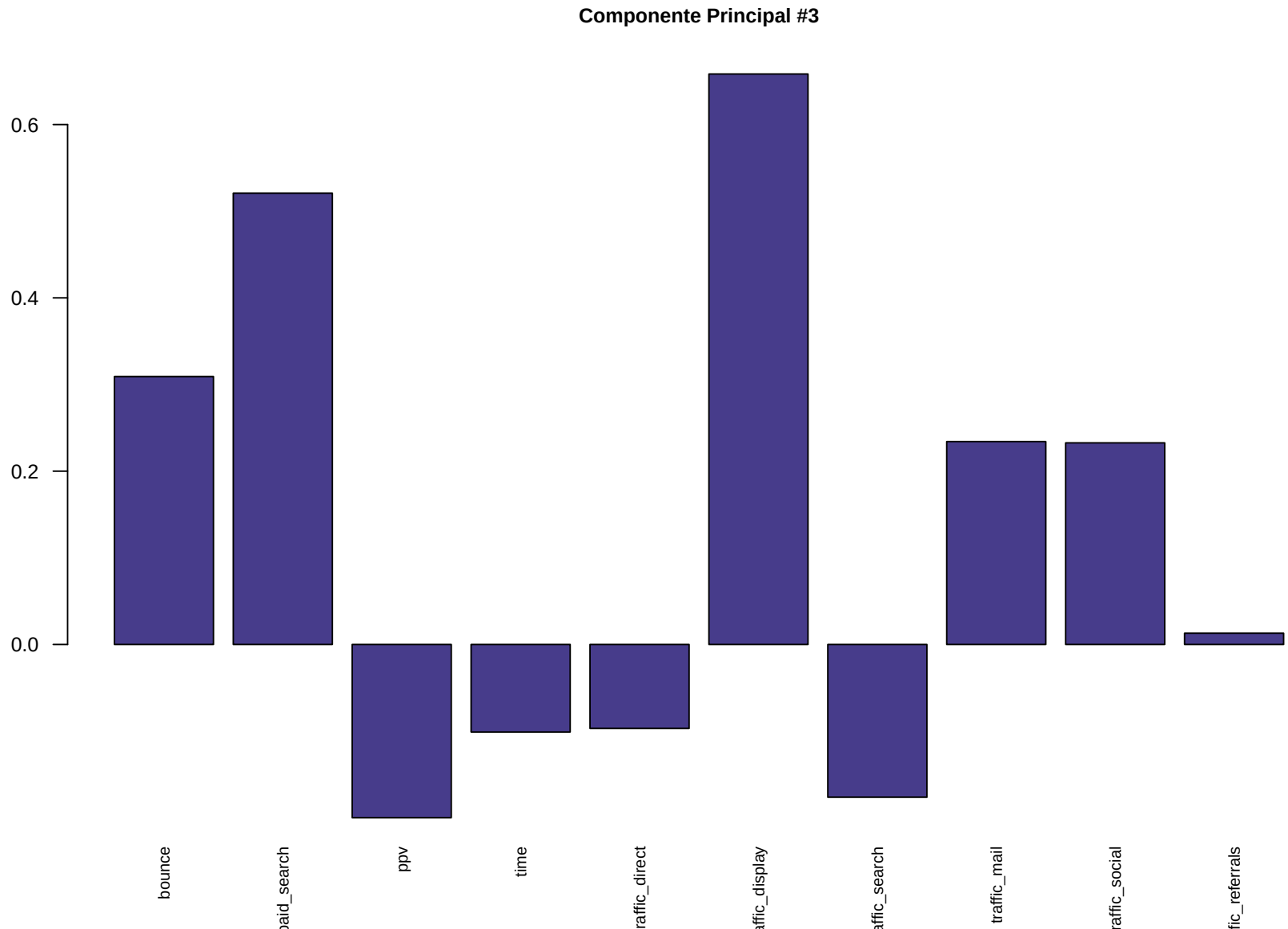
Componentes principales - Componente 2

- Conocimiento de marca
- Componente de forma



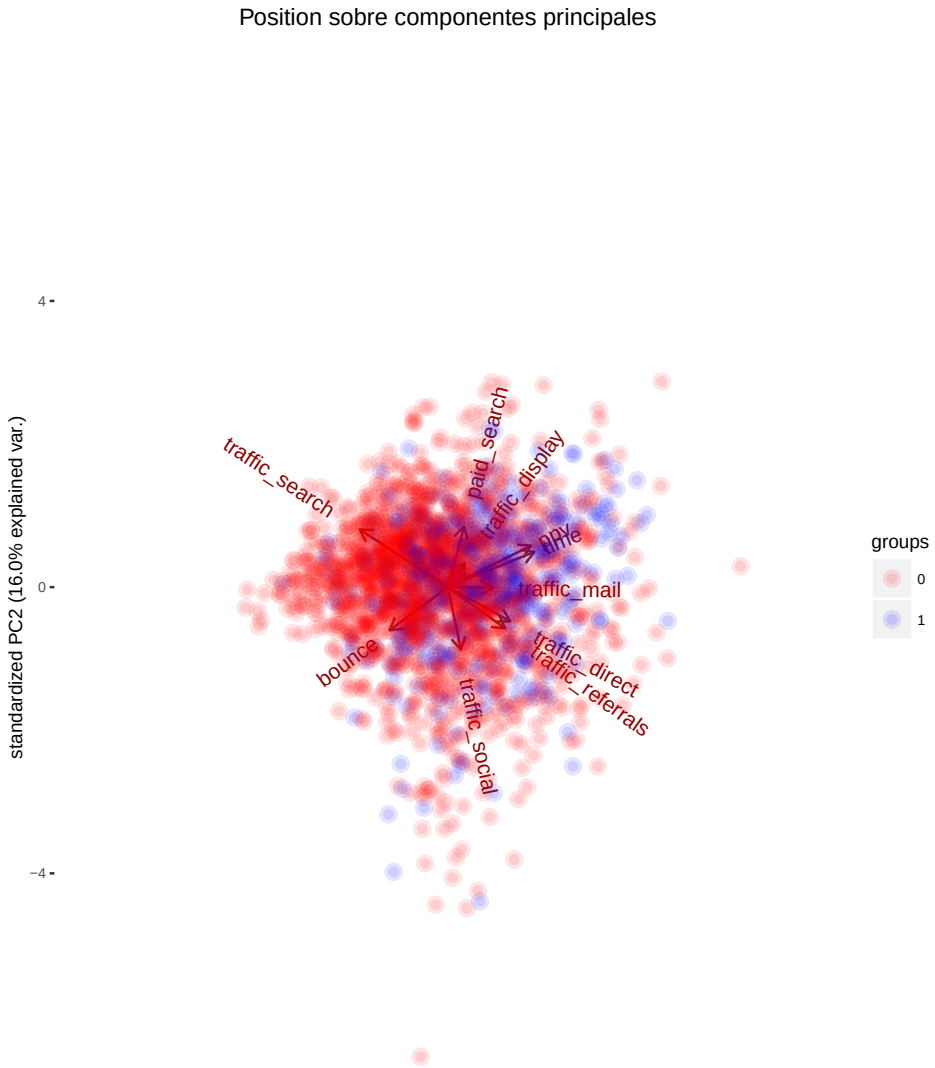
Componentes principales - Componente 3

- Tráfico pago
- Componente de forma

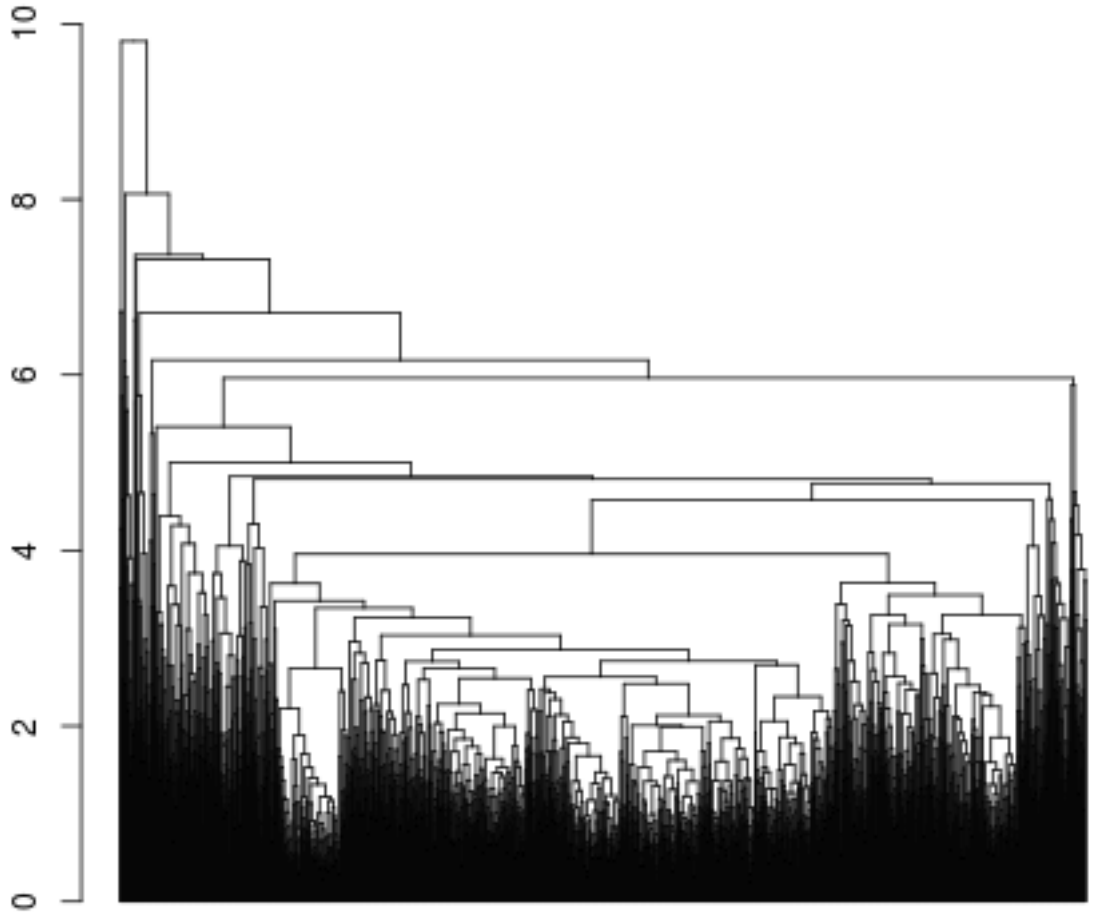


Componentes principales - Biplot

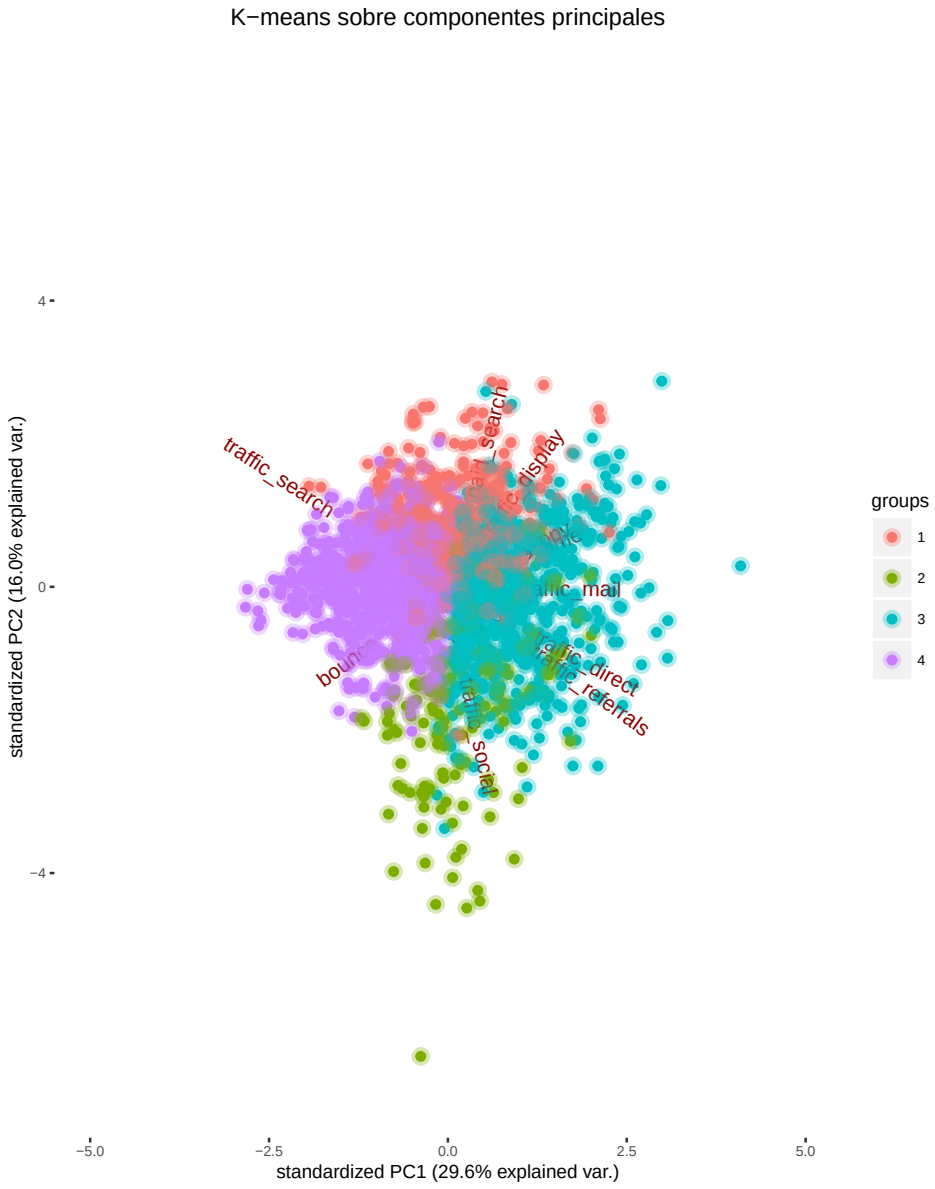
- Agrupamiento basado en el nivel de tráfico
- Grupo 1 proyecta valores positivos sobre la primer componente



Clustering - Jerárquico



Clustering - K-means



Clustering - Grupos según posición global

- Cluster 3 concentra a los primeros sitios del ranking.
- Cluster 4 concentra a los últimos sitios del ranking.

cluster	0	1
1	259	106
2	122	21
3	440	187
4	704	89

Clustering - Grupos identificados

- Cluster 3: Baja tasa de rebote, elevado tiempo de visita, elevado tráfico directo.
- Cluster 4: Bajo nivel de páginas por visita, elevado tráfico pago.
- ¿ Hub vs Authority ?

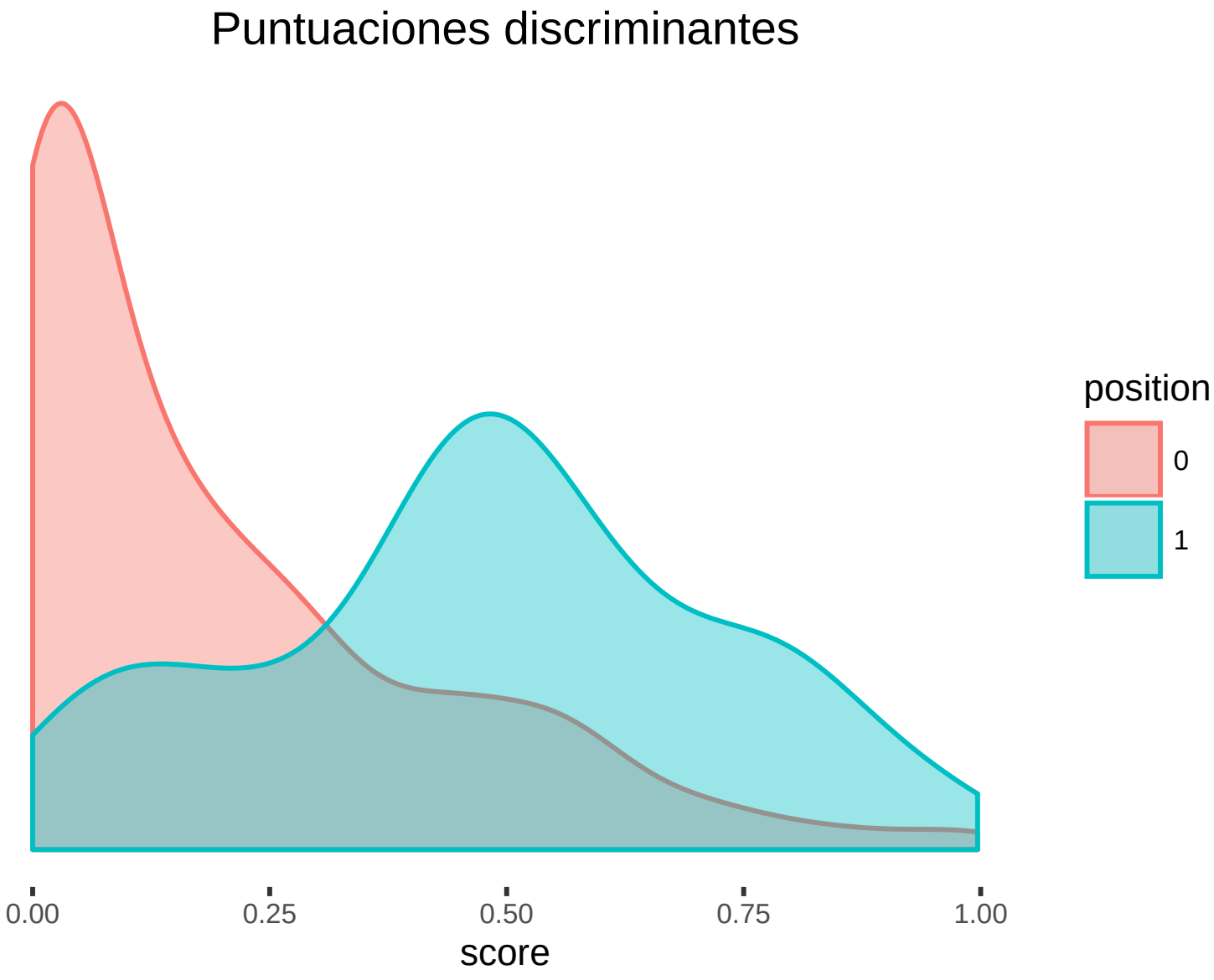
	V1	V2	V3	V4
Group.1	1	2	3	4
bounce	40.45811	46.58832	34.98708	45.02733
paid_search	38.133342	7.369371	11.379027	8.277062
ppv	5.019808	3.825105	6.575694	3.437718
time	212.4137	142.6154	263.6587	130.7503
traffic_direct	22.65986	21.16154	34.77341	19.81710
traffic_display	3.6406575	0.5546154	0.6288038	0.2980580
traffic_search	54.04164	33.90420	37.90638	67.12605
traffic_mail	5.600247	12.359231	5.123636	1.889294
traffic_social	2.165479	15.557063	3.006268	1.766103
traffic_referrals	11.891781	16.463007	18.562041	9.103468

Análisis discriminante - Resumen

- No se satisfacen supuestos de normalidad y homocedasticidad
- Análisis discriminante cuadrático
- 70% training / 30% testing.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 333  57
##          1  48  44
##
##                Accuracy : 0.7822
##                  95% CI : (0.7426, 0.8182)
##     No Information Rate : 0.7905
##     P-Value [Acc > NIR] : 0.6956
##
##                   Kappa : 0.3201
##  Mcnemar's Test P-Value : 0.4350
```

Análisis discriminiate - Puntuaciones



Software

- Python + Jupyter: Preparación de datos.
- Rstudio: Análisis y documentación.
- GitHub: Almacenamiento y versionado del código.

Conclusiones

- Diferencias significativas según el nivel de tráfico