

# PROYECTO 1

PATRONES DE COMPORTAMIENTO EN EL  
USO DE BICICLETAS DEL SISTEMA DE  
TRANSPORTE MEXICANO ECOBICI EN LOS  
AÑOS 2018-2019

Wilfredo Gallegos - 20399  
Javier Aguilar - 20611



# Tabla de contenido

01

**Descripción de los datos**

02

**Planteamiento de preguntas**

03

**VARIABLES a considerar**

04

**Metodología y herramientas**

05

**Resultados**

06

**Conclusiones**

# Contexto

## ¿Que es ECOBICI?



Ecobici es un sistema de bicicletas públicas en la Ciudad de México que promueve la movilidad sostenible. Ofrece a los usuarios la posibilidad de tomar y devolver bicicletas en diferentes estaciones, con trayectos de hasta 45 minutos. Publica datos en tiempo real sobre la disponibilidad de bicicletas y estaciones en su página web.

## Descripción de los Datos

- **Genero\_Usuario:** Género de la persona(M/F/O)
- **Edad\_Usuario:** Edad del usuario
- **Bici/BiciID:** Número de identificación de la bicicleta()
- **Ciclo\_Estacion\_Retiro:** Numero de estación donde se solicita una bicicleta
- **Fecha\_Retiro:** Fecha de Solicitud
- **Hora\_Retiro:** Hora de solicitud
- **Ciclo\_EstacionArribo:** Numero de estación donde se regresa una bicicleta
- **Fecha Arribo:** Fecha de devolución
- **Hora\_Arribo:** Hora de devolución

Ademas, fue necesario agregar los siguientes datos al conjunto,

- **estacion:** Número de la estación
- **duracion:** Duracion del viaje en bici

# Planteamiento de Preguntas

## Pregunta o Hipotesis principal

¿De qué manera se vincula el tiempo de uso de las bicicletas y la frecuencia de las solicitudes por hora con las diferentes estaciones de forma individual?

## Preguntas secundarias



¿Como cambia la relacion de las variables en la pandemia(2020-2021) y despues de la pantemia(2022-2023)?



¿De que manera se pueden agrupar las estaciones?



¿Que varialbes pueden tomarse en cosideracion para poder realizar alguna proyeccion lineal o clustering?

# Variables a Considerar



Estaciones



Ciclo\_estacion\_Retir

o



Hora\_Retiro



Edad\_Usuario

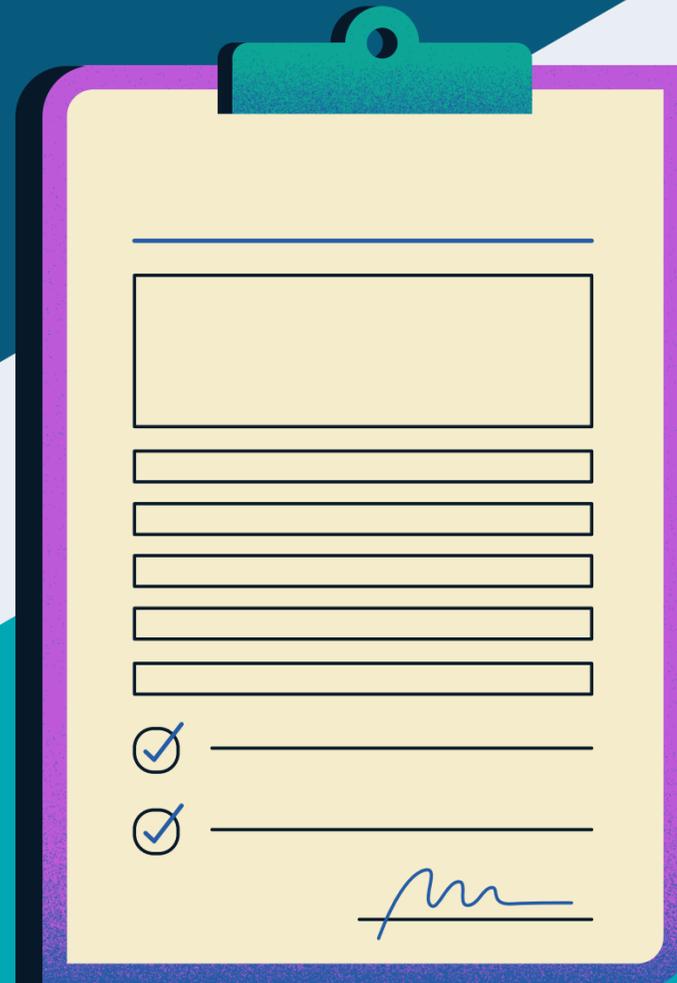


Hora\_Arribo

# Metodologia

## Analisis por año

1. Se procede a juntar la base de datos de los años 2018-2019, 2020-2021 y 2022-2023
2. Se hace limpieza, correccion y categorizacion de los datos, para luego hacer un analisis exploratorio y ver comportamiento y correlacion de las variables.
3. Se procede a ejecutar las herramientas para obtencion de informacion, tales como Clustering y PCA



# Herramientas

## PCA



El análisis de componentes principales es una técnica utilizada para describir un conjunto de datos en términos de nuevas variables no correlacionadas. con el objetivo de simplificar la complejidad

## Proyeccion lineal

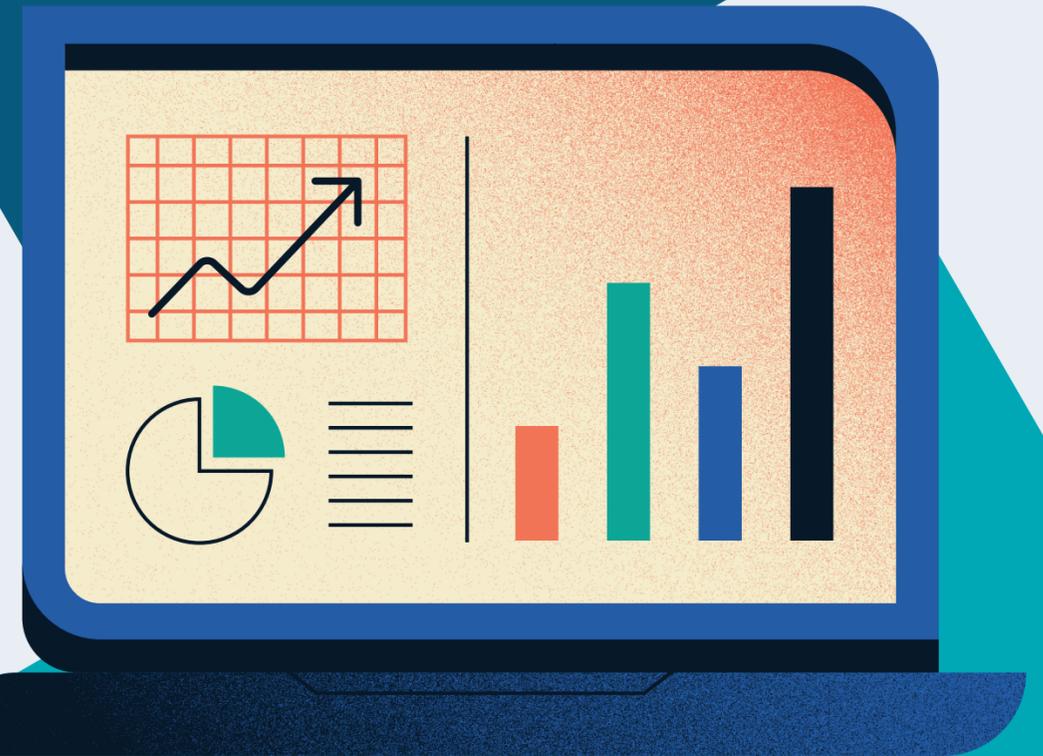


Es un proceso matemático que transforma los datos a un espacio de menor dimensión, manteniendo la estructura esencial de los datos. Es similar al PCA en que ambos buscan simplificar la complejidad de los datos.

## Clustering(k-means)



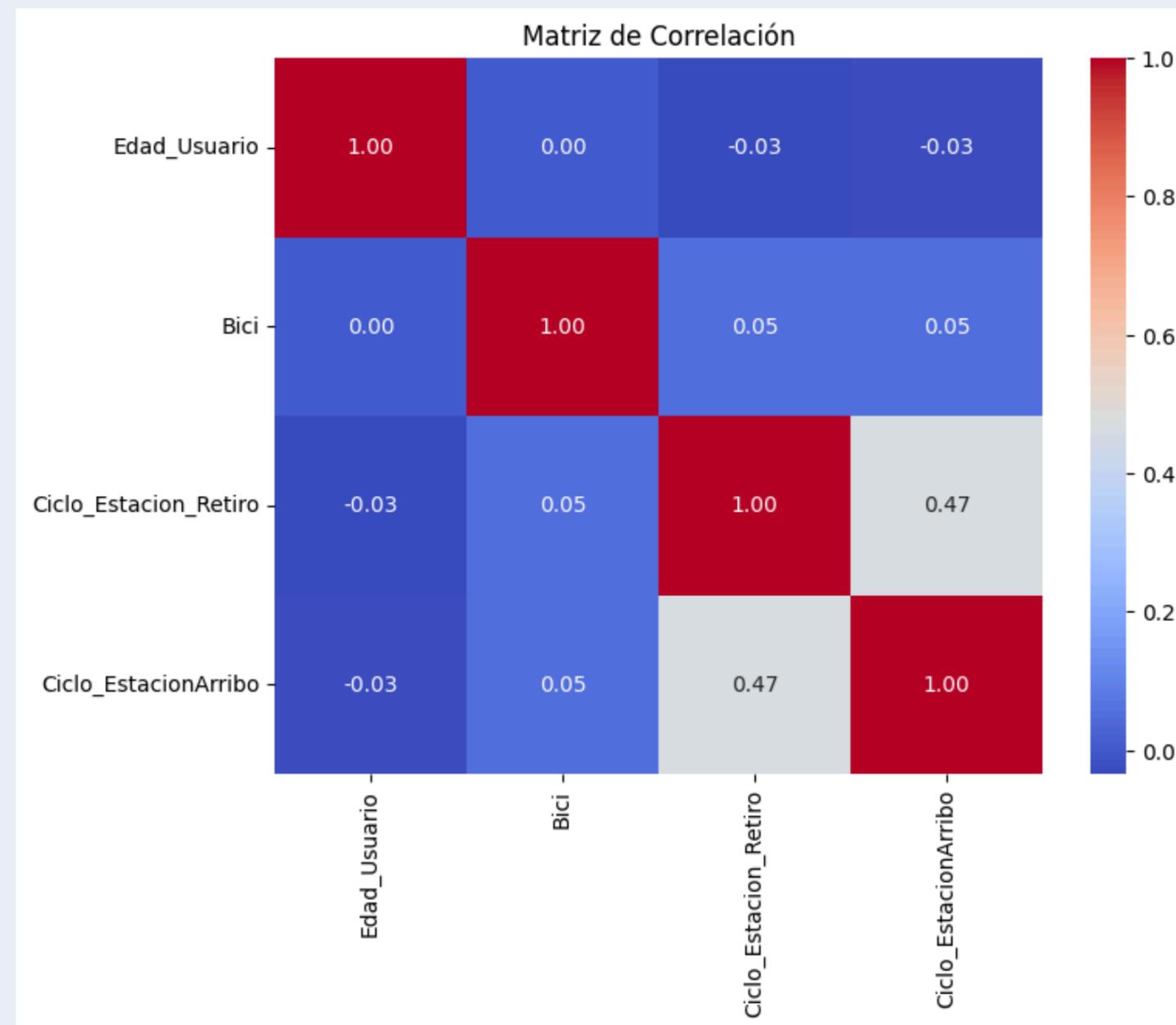
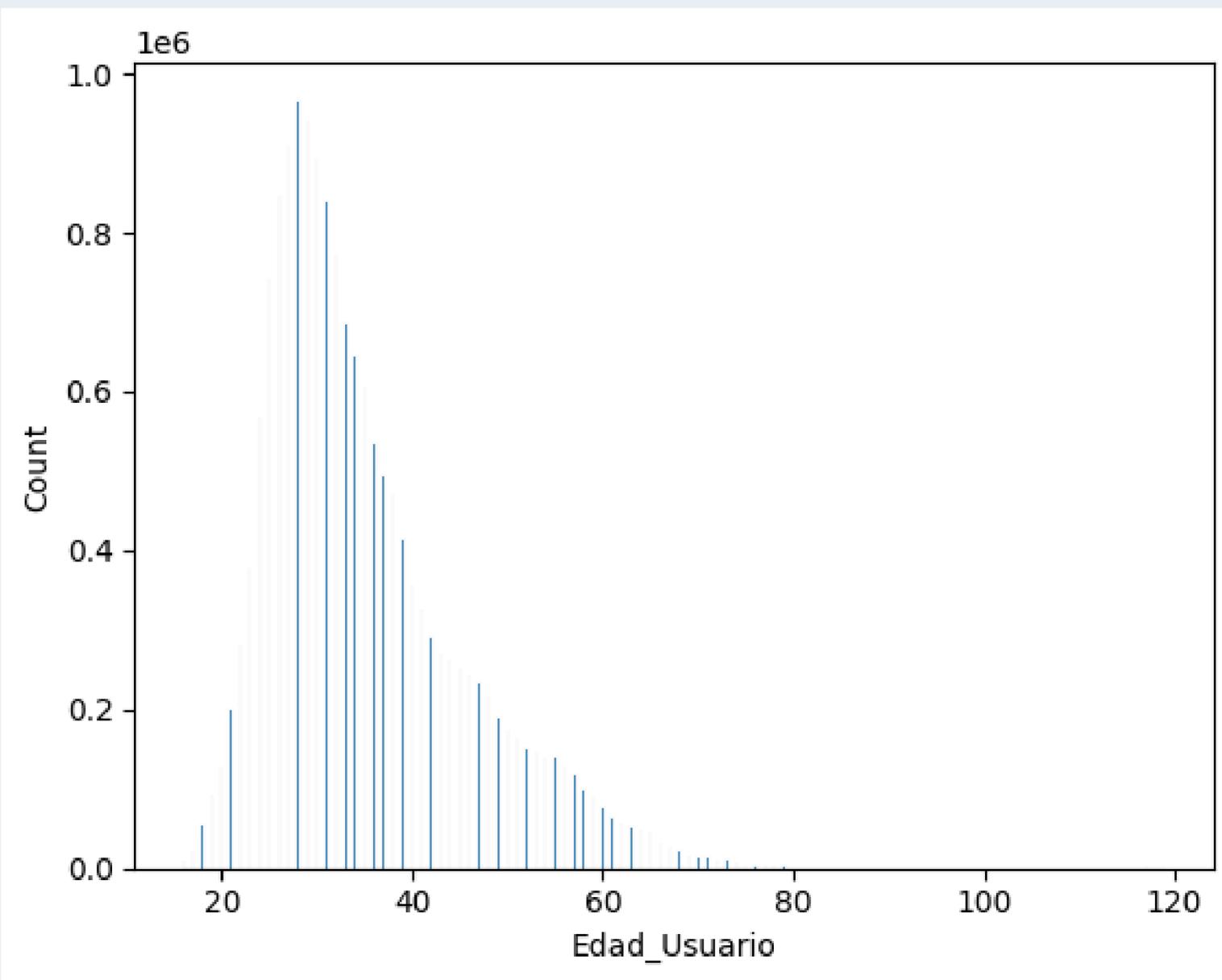
Es un método de agrupamiento, que tiene como objetivo la partición de un conjunto en grupos en el que cada observación pertenece al grupo.



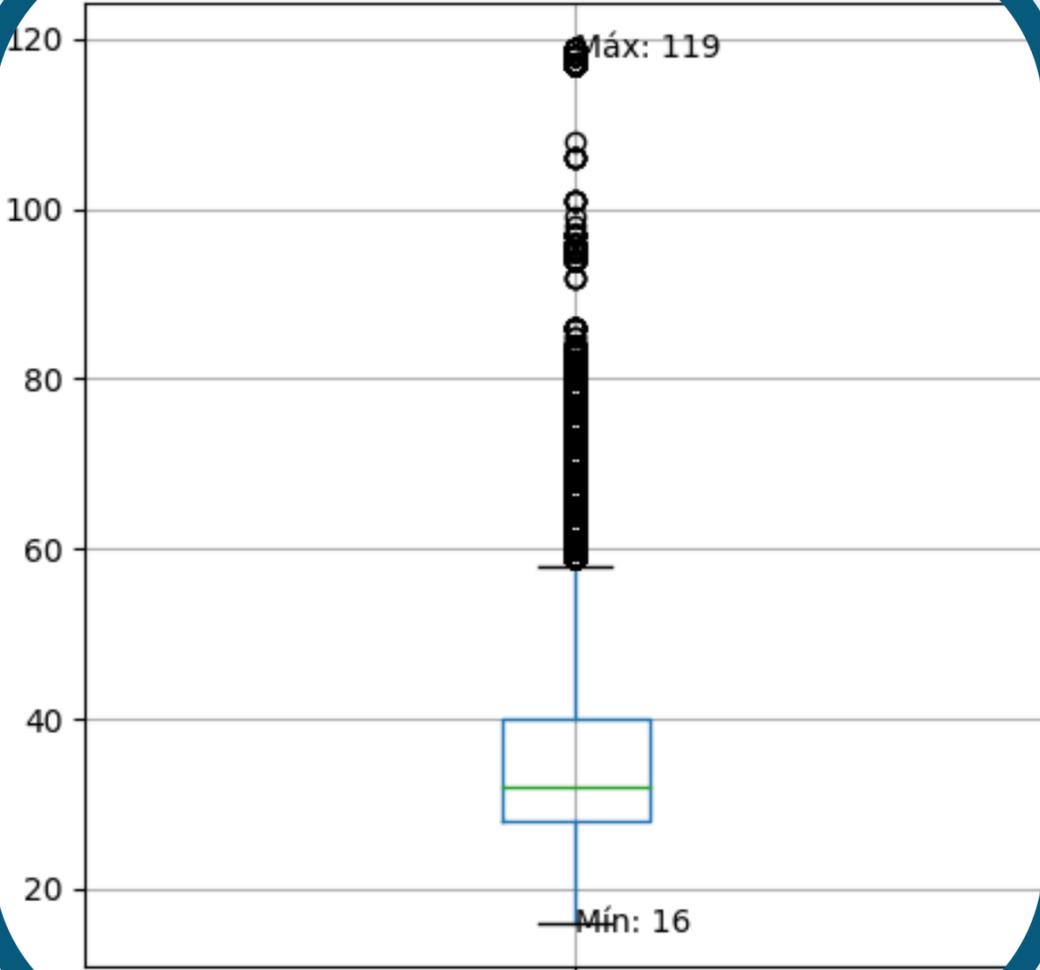
# RESULTADOS

2018-2019

# EXPLORACIÓN Y ANÁLISIS



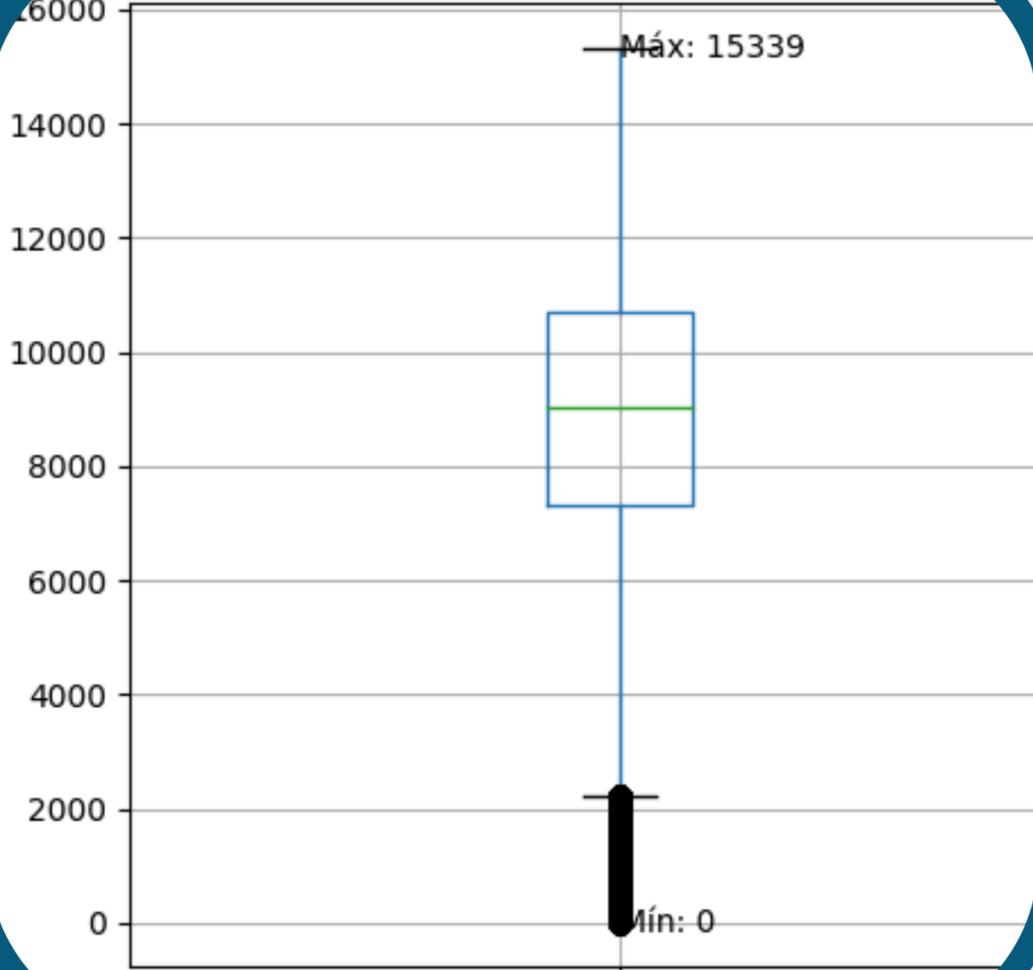
Edad\_Usuario



Edad\_Usuario

Edad de usuarios

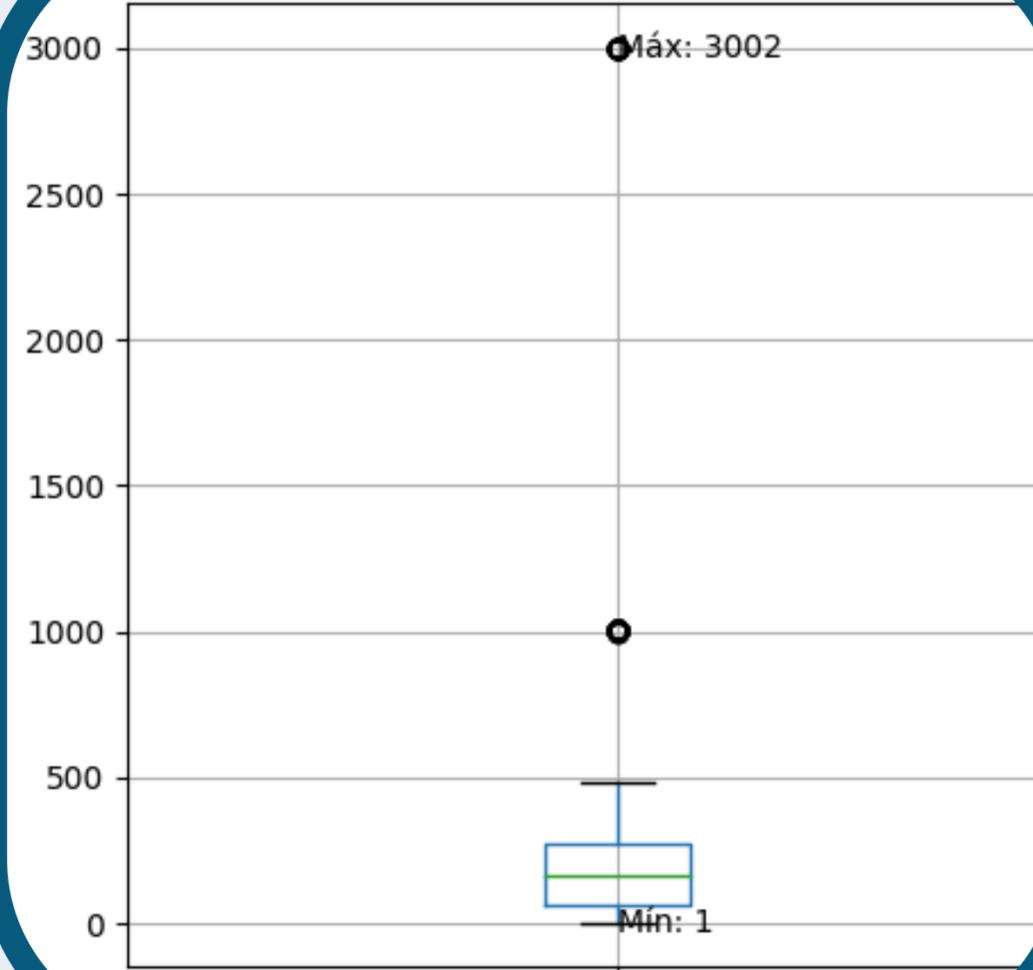
Bici



Bici

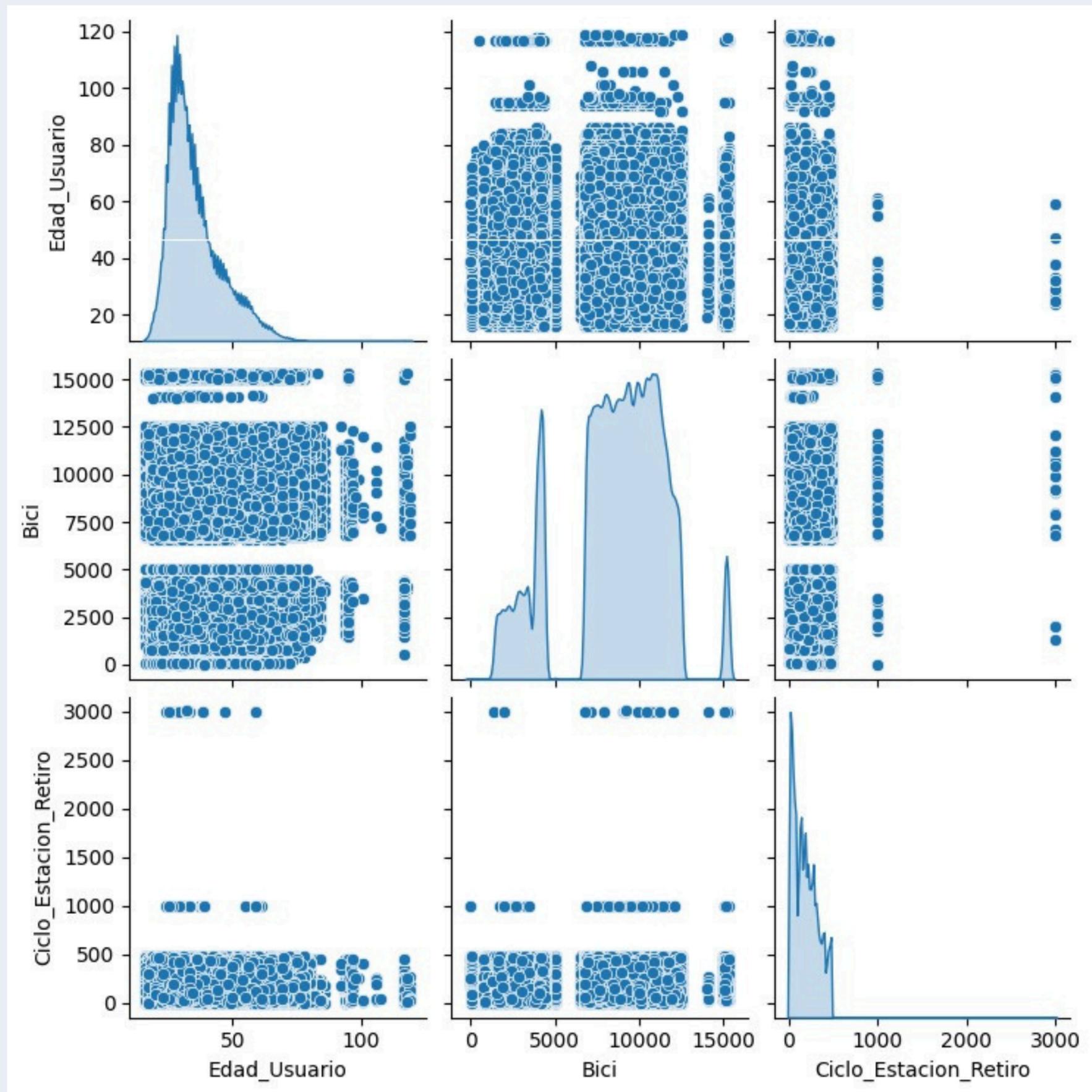
ID de la Bici

Ciclo\_Estacion\_Retiro



Ciclo\_Estacion\_Retiro

Estación de Retiro



Pair Plot

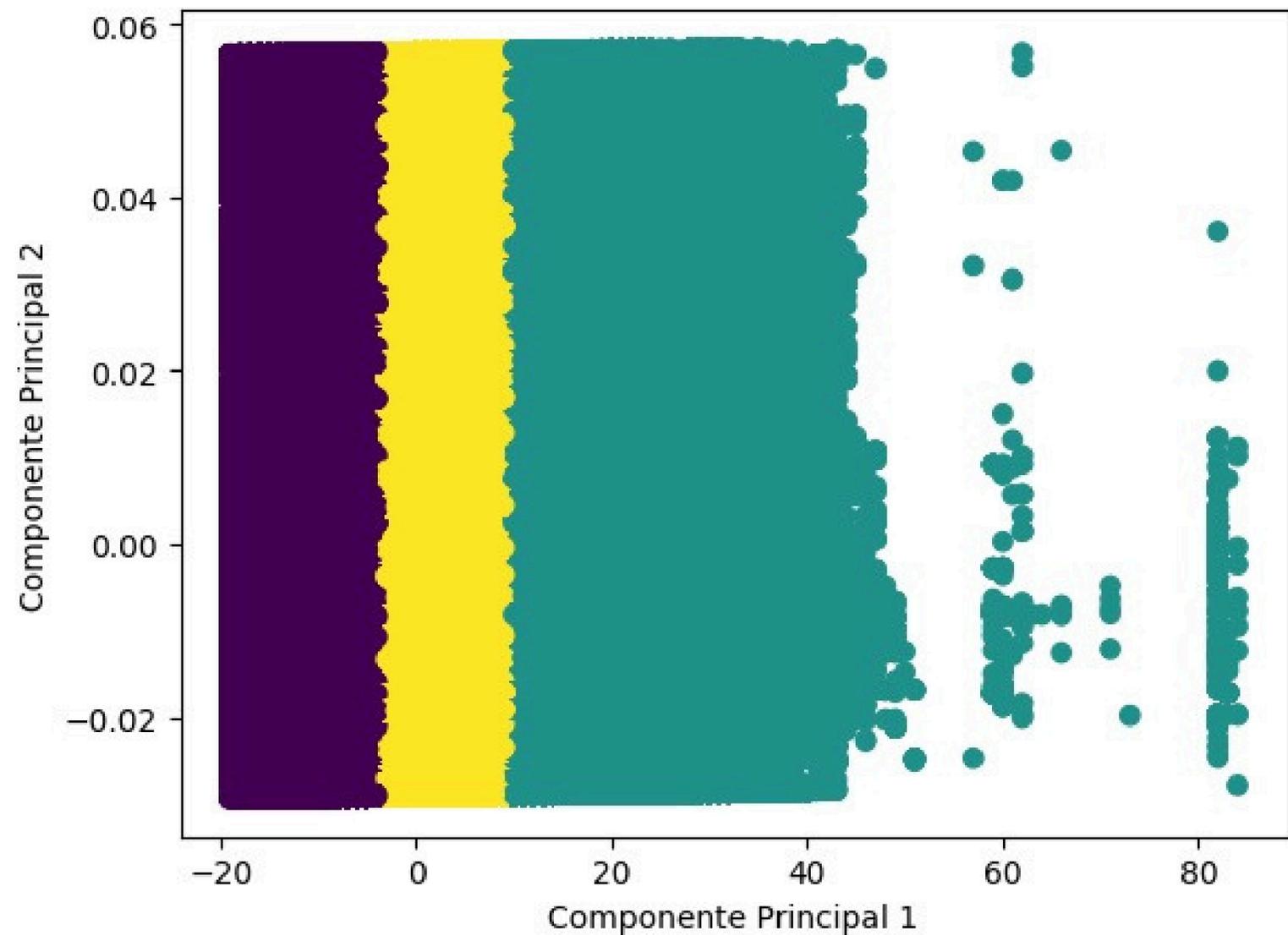
Genero_Usuario	F	M
name		
1 RIO SENA-RIO BALSAS	36427	116834
10 REFORMA-RAMIREZ	15172	59530
100 20 DE NOVIEMBRE-MESONES	1927	7512
101 1ER. CALLEJÓN DE MESONES-MESONES	1417	5616
102 ECHEVESTE-BOLIVAR	3547	12083
...	...	...
E464 MATEO ALEMÁN-PROLONGACIÓN MOLIERE	2918	10686
E467 LAGO MAYOR-LAGO ALBERTO	2247	8590
E469 BAHÍA DE LA CONCEPCIÓN-BAHÍA DE TODOS LOS SAN	1400	4876
E472 BAHÍA SANTA BÁRBARA-BAHÍA DE BALLENAS	2077	9644
E479 LAGO MURITZ-AV. MARINA NACIONAL	2342	5628

**Usuarios de estacion por generos**

Hora	
0	54920
1	5880
2	2946
3	57
4	22
5	84510
6	351029
7	909652
8	1666815
9	1269379
10	829273
11	743495
12	790418
13	964947
14	1180214
15	1153624
16	968132
17	1212299
18	1660283
19	1303192
20	788186
21	519832
22	324800
23	171796

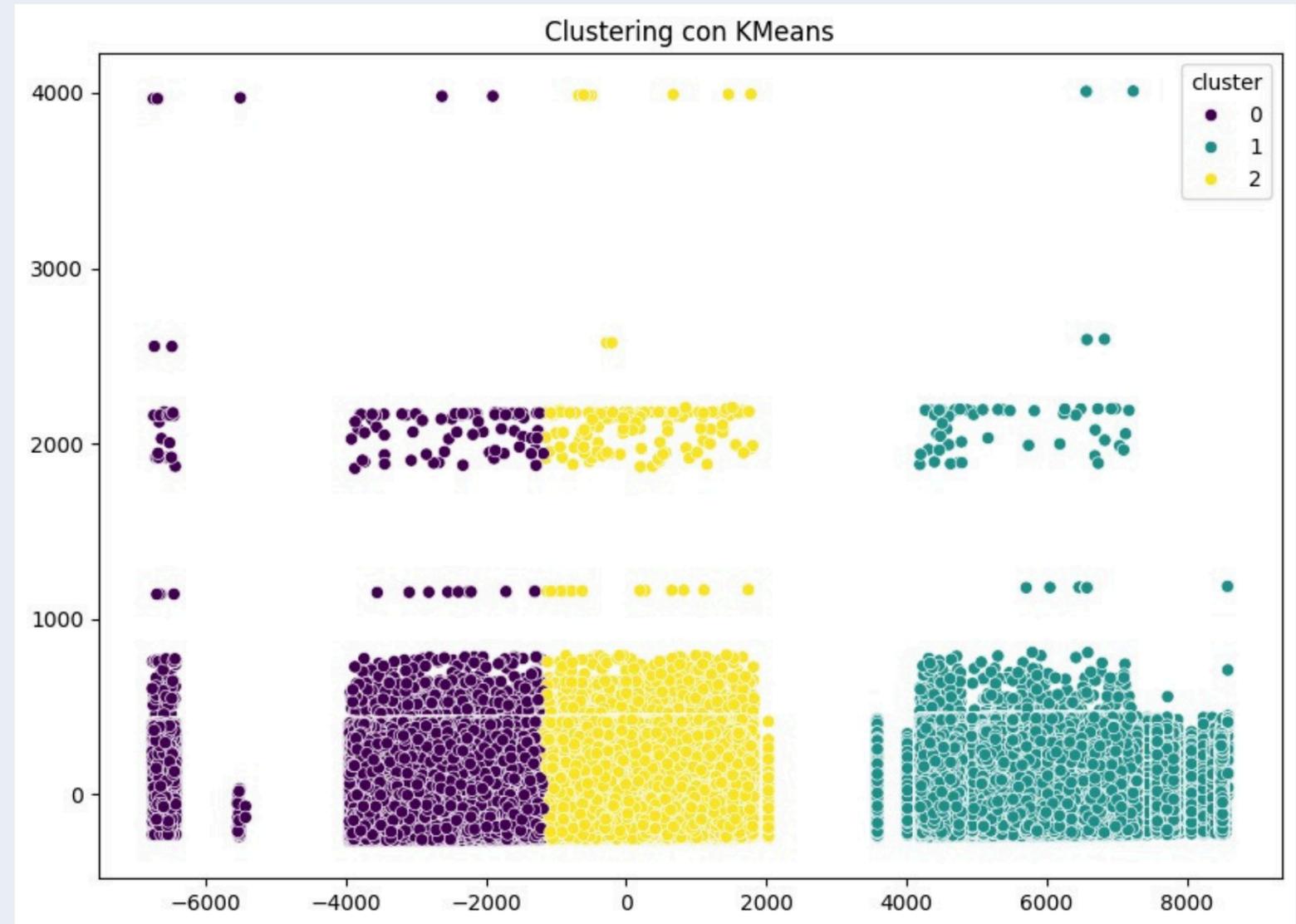
**reccurrencia en horas**

Edad de usuarios

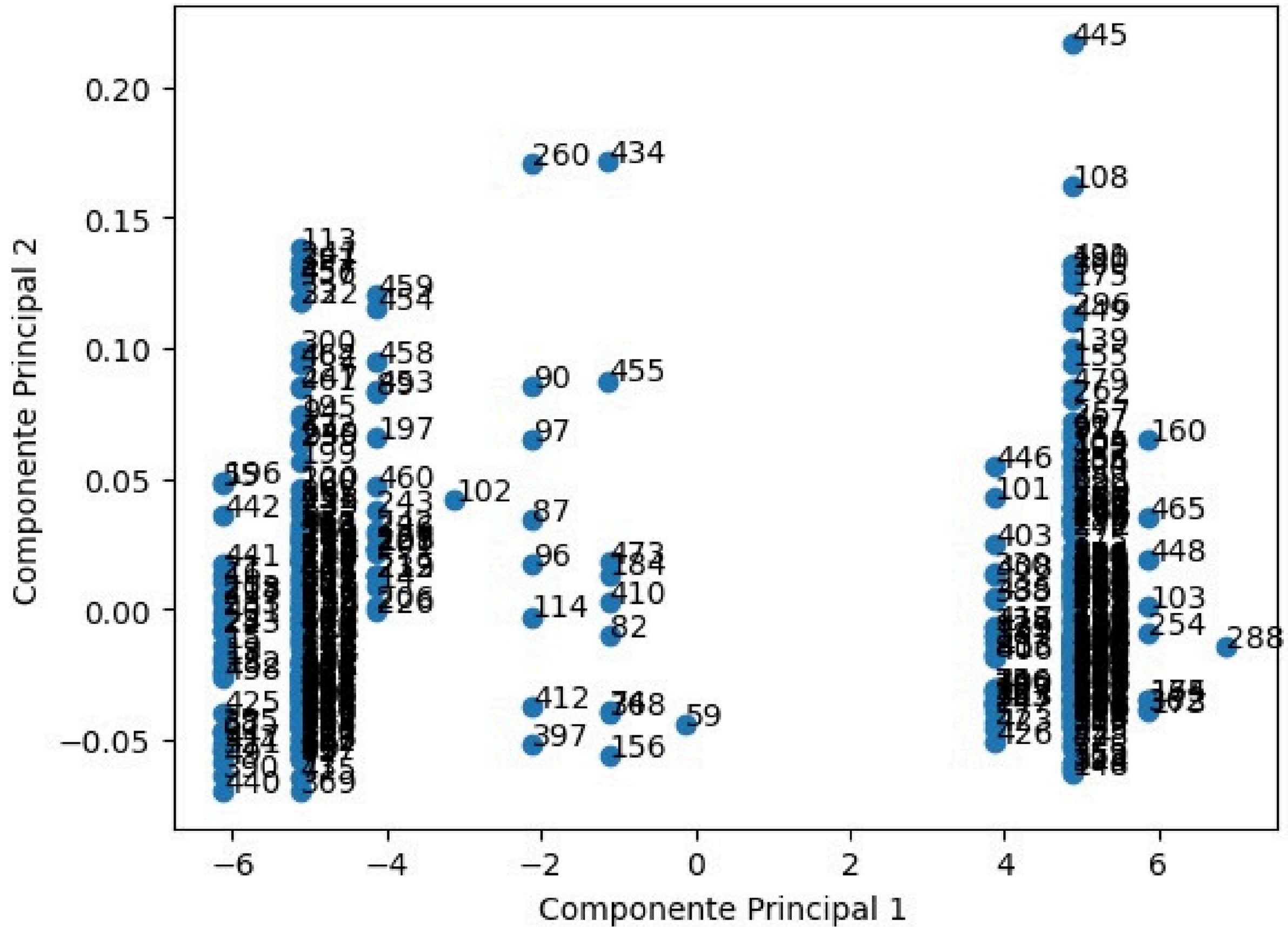


Clustering con PCA

Todas las columnas numéricas del DF



Proyección Lineal con Clustering

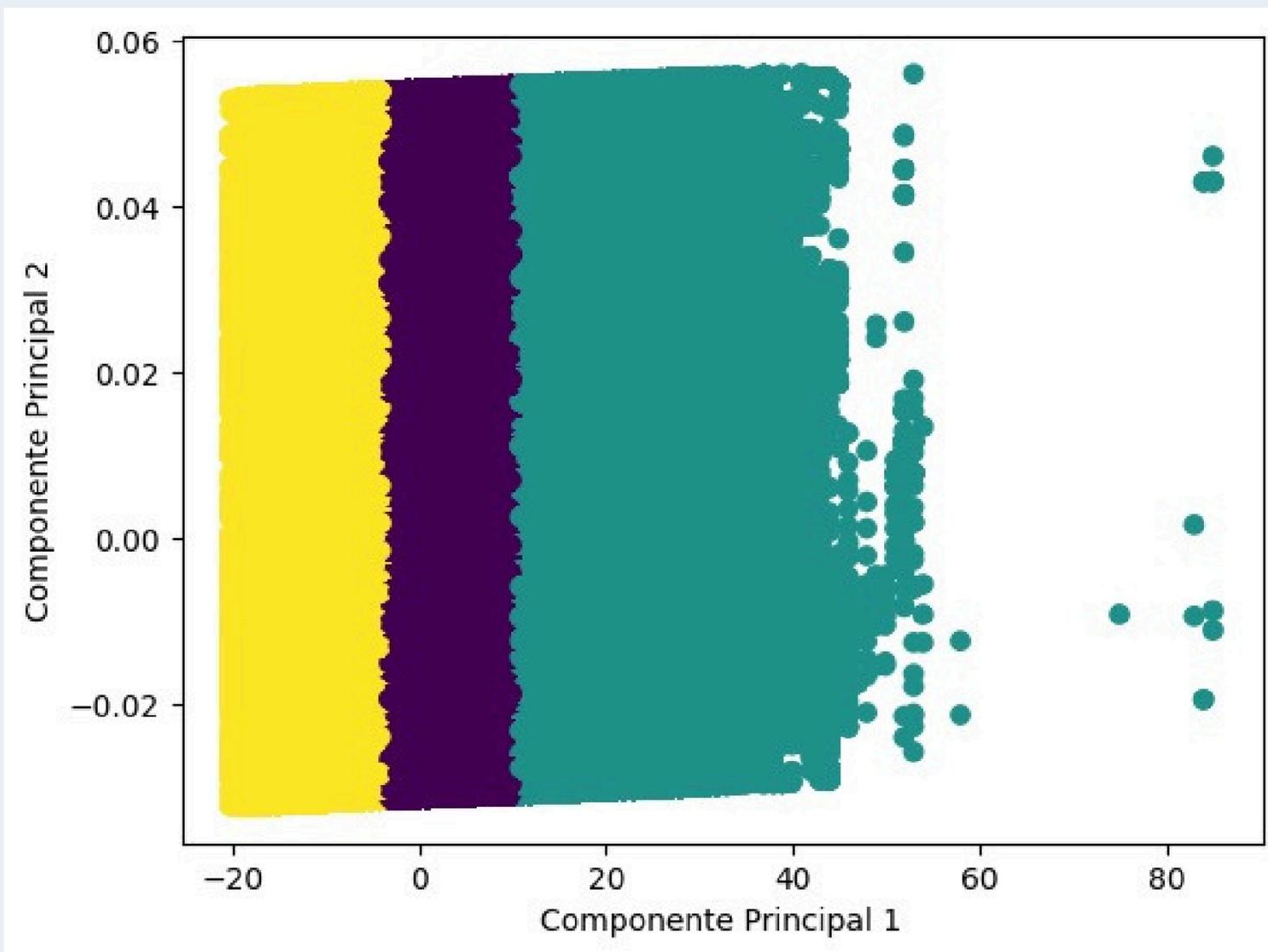


PCA 18-19

# RESULTADOS

2020-2021

## Edad de usuarios

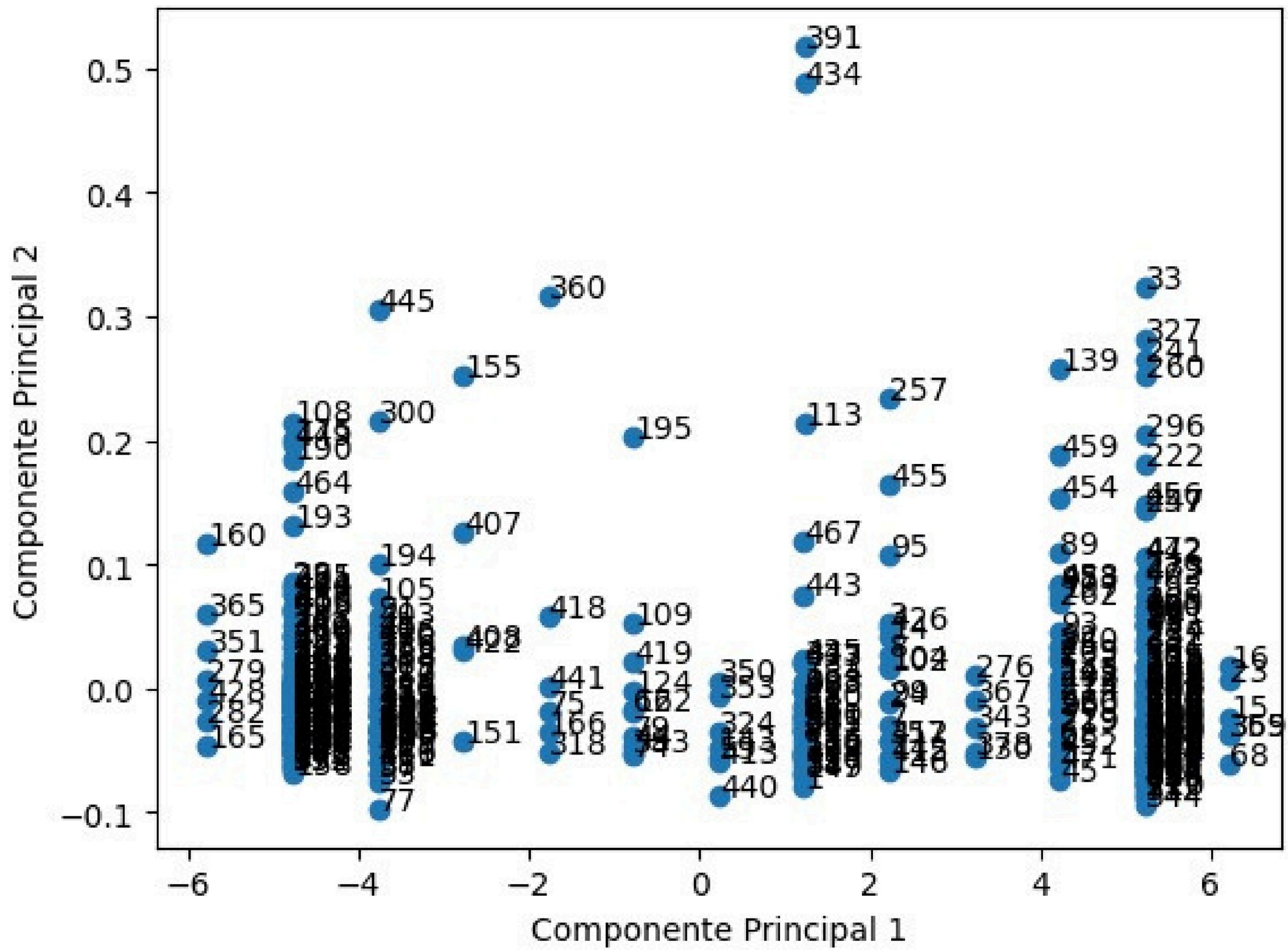


## Clustering con PCA

Genero_Usuario	F	M	Other
id			
1	18164	51752	2053
2	6488	15619	653
3	5681	18787	1500
4	6668	14311	453
5	5350	16558	577
...	...	...	...
476	5068	14687	475
477	2227	8261	592
478	1975	8328	273
479	2062	3768	257
480	4761	11626	1173

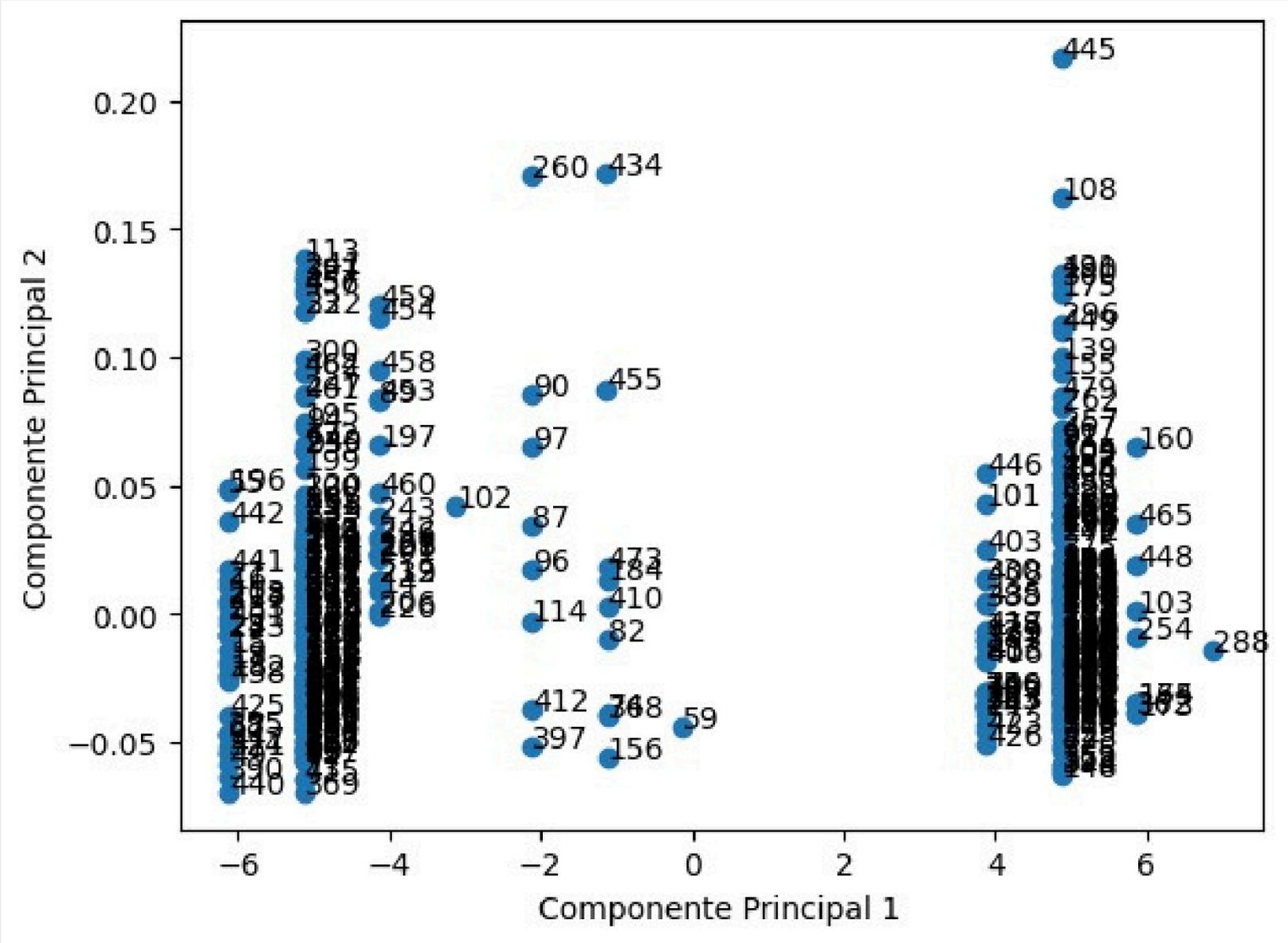
479 rows × 3 columns

## Usuarios de estacion (por ID) por generos

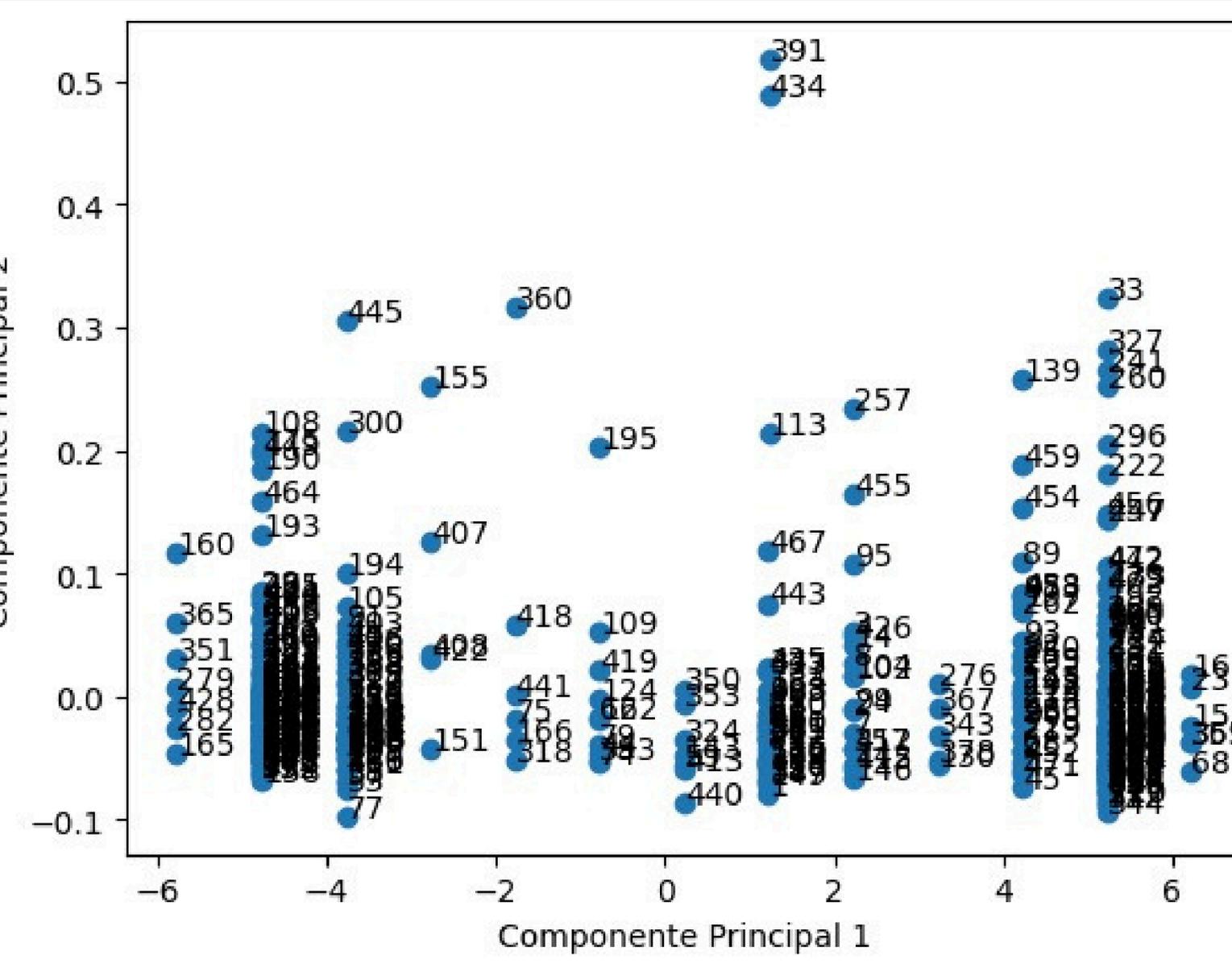


PCA 20-21

# Comparación PCA's



PCA 18-19



PCA 20-21

**GRACIAS**



# Pricing Package

You can use presentation media in the form of a combination of images, videos, graphics, tables and others.

## BASIC

- Add Description Here
- Add Description Here
- Add Description Here
- Add Description Here

**\$500**

Get Started

## PREMIUM

- Add Description Here
- Add Description Here
- Add Description Here
- Add Description Here

**\$750**

Get Started