



TRABAJO FINAL

Analítica de Negocios

Simón Vélez, Sofía Villegas y María
Arango

TABLA DE *Contenidos*

- 01 Descripción de la Base de Datos
- 02 Objetivo
- 03 Modelos implementados
- 04 Código Modelo Uno
- 05 Código Modelo Dos
- 06 Resultados y Conclusiones



MOBILE DEVICE USAGE AND USER BEHAVIOR DATASET

- Análisis de patrones de uso de dispositivos móviles y clasificación del comportamiento de los usuarios.
- 700 muestras de datos de usuario.
- Cada usuario se clasifica en una de cinco clases de comportamiento del usuario, que van desde uso ligero hasta uso extremo.

App Usage
Time (min/day)

Screen On
Time
(hours/day)

Battery Drain
(mAh/day)

Number of
Apps
Installed

Data Usage
(MB/day)

Age

Descripción BASE DE DATOS





OBJETIVO



modelos



Usar los modelos seleccionados para entender la base de datos y conocer la relación que tienen las variables. Además, como objetivo se quiere analizar la clasificación que se le da a los usuarios y qué los hace tener ese pronóstico.

KNN y
KMeans

Árbol de
Decisión

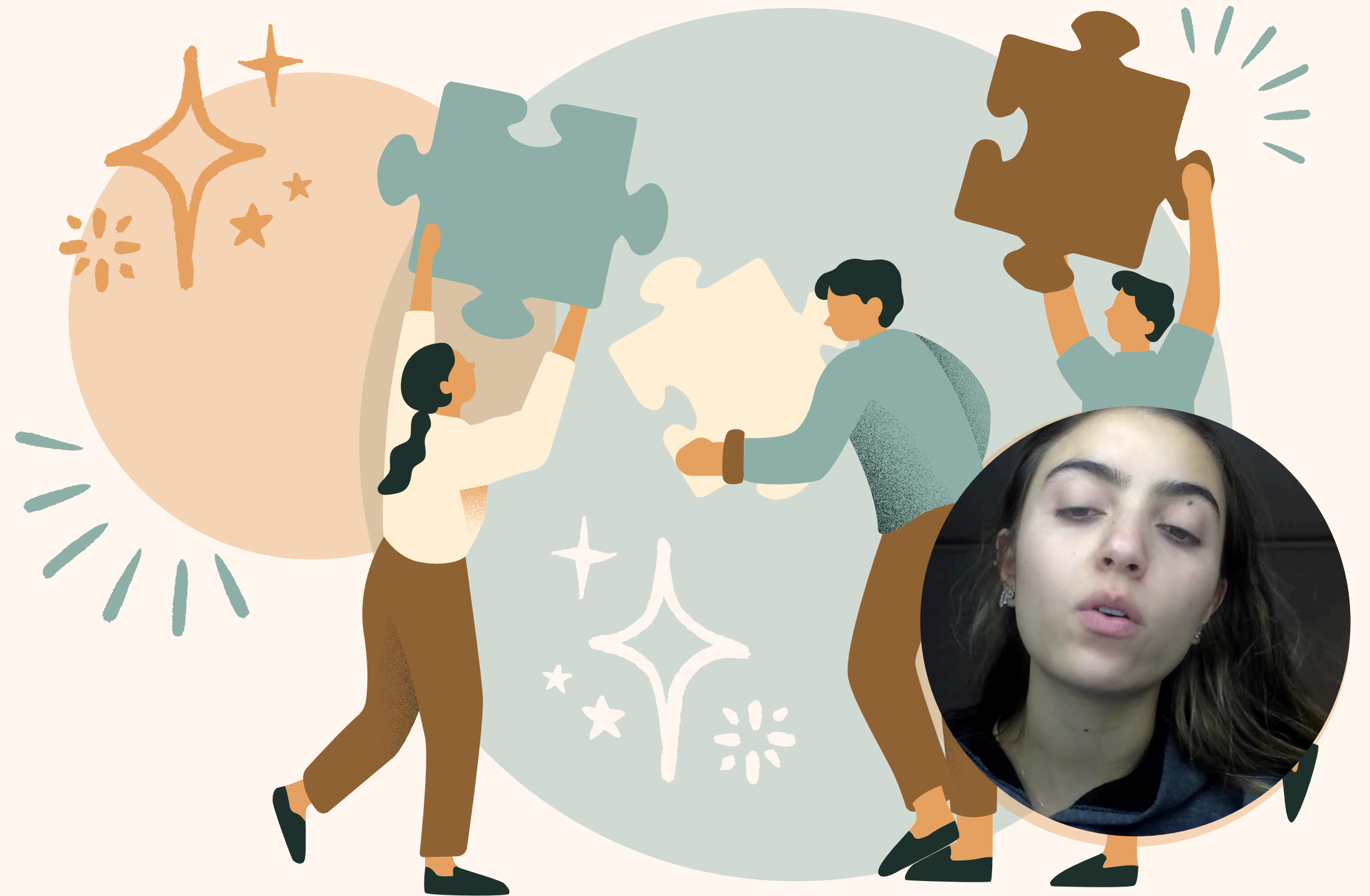


CÓDIGO

No

KNN y KMeans

<https://colab.research.google.com/drive/1xc995ZB6ZrQ-q73sl3T0123XxSZZlati?usp=sharing>

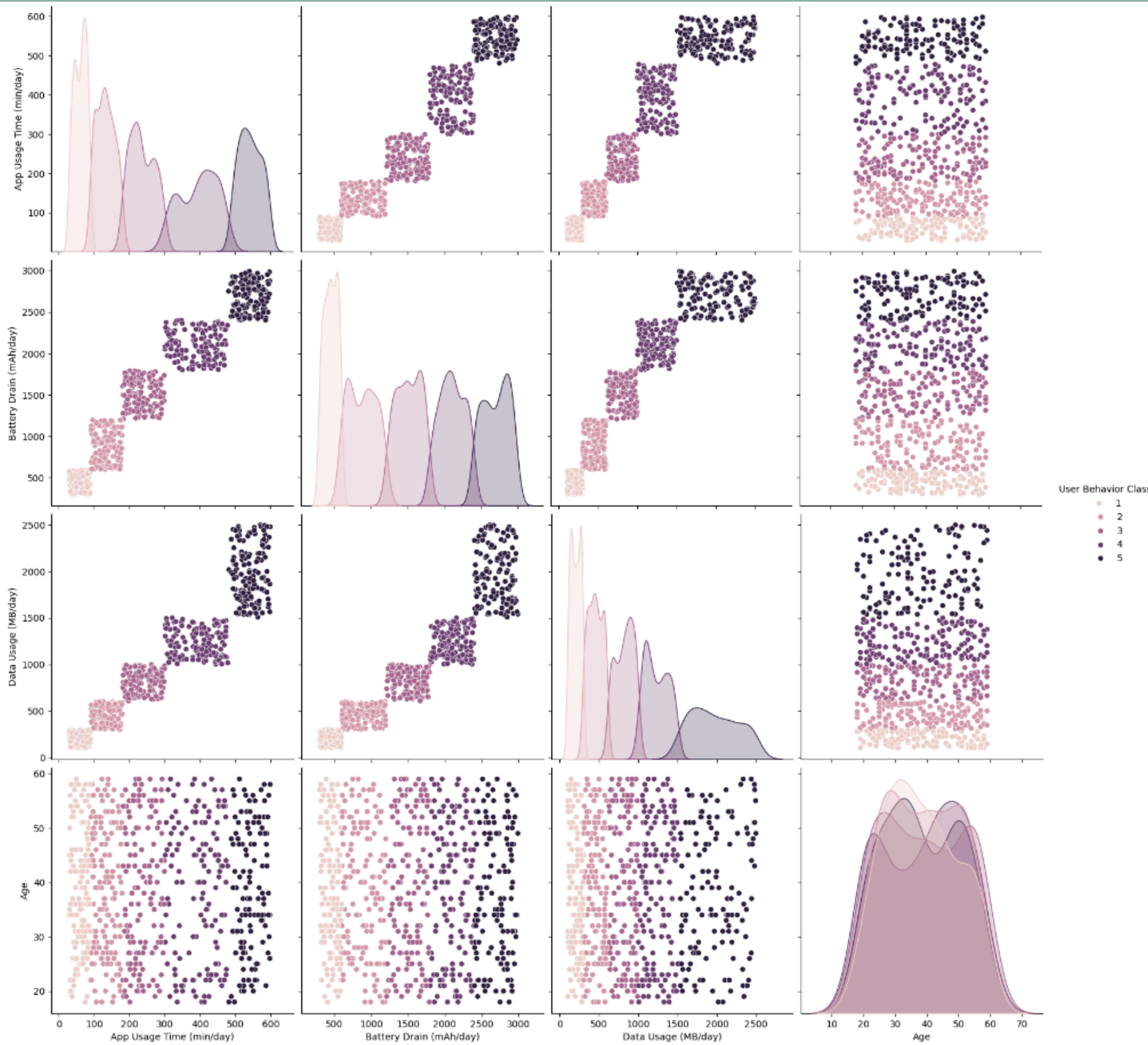
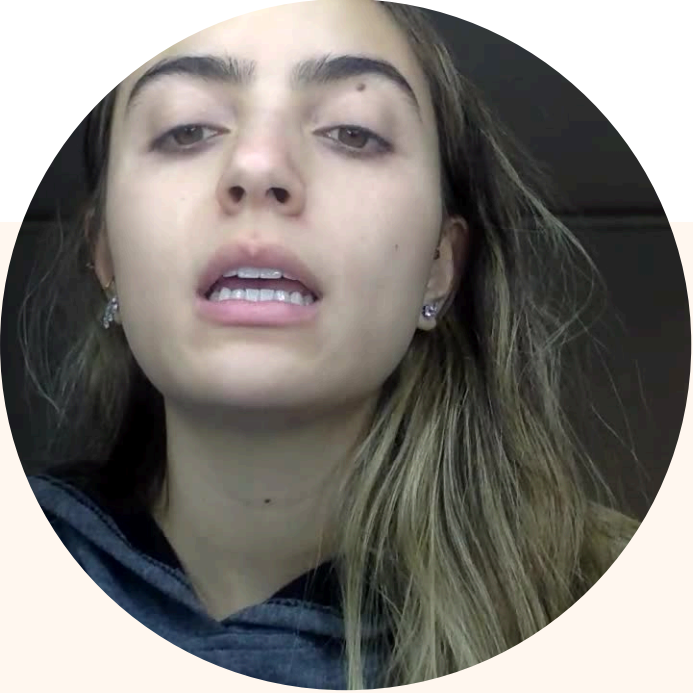


Esta es la matriz de correlación:

	App Usage Time (min/day)	Battery Drain (mAh/day)	\
App Usage Time (min/day)	1.000000	0.956385	
Battery Drain (mAh/day)	0.956385	1.000000	
Data Usage (MB/day)	0.942308	0.932276	
Age	0.004382	-0.002722	
User Behavior Class	0.970498	0.978587	

	Data Usage (MB/day)	Age	User Behavior Class
App Usage Time (min/day)	0.942308	0.004382	0.970498
Battery Drain (mAh/day)	0.932276	-0.002722	0.978587
Data Usage (MB/day)	1.000000	0.003999	0.946734
Age	0.003999	1.000000	-0.000563
User Behavior Class	0.946734	-0.000563	1.000000

[[393 1872 1122 40]
[268 1331 944 47]
[154 761 322 42]
...
[99 942 457 50]
[62 431 224 44]
[212 1306 828 23]]



Esta es la matriz de confusión:

```
[[136  0  0  0  0]
```

```
[ 0 146  0  0  0]
```

```
[ 0  0 143  0  0]
```

```
[ 0  0  0 139  0]
```

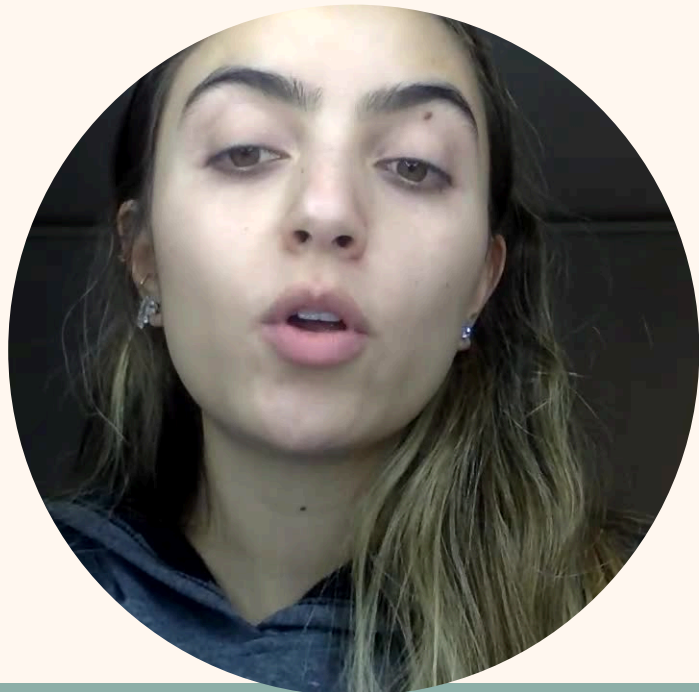
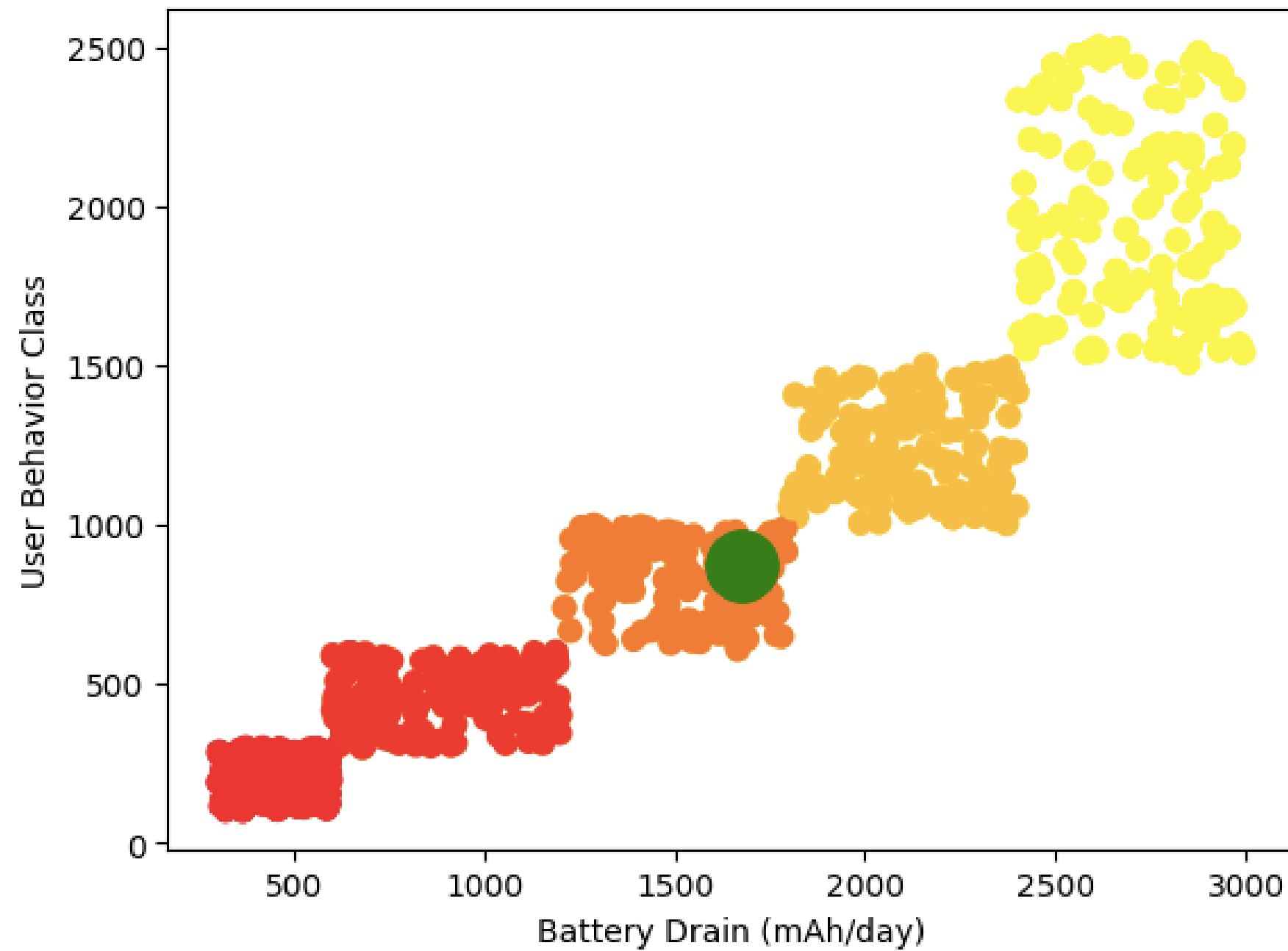
```
[ 0  0  0  0 136]]
```

La Exactitud Alcanzada por el Modelo es: 1.0

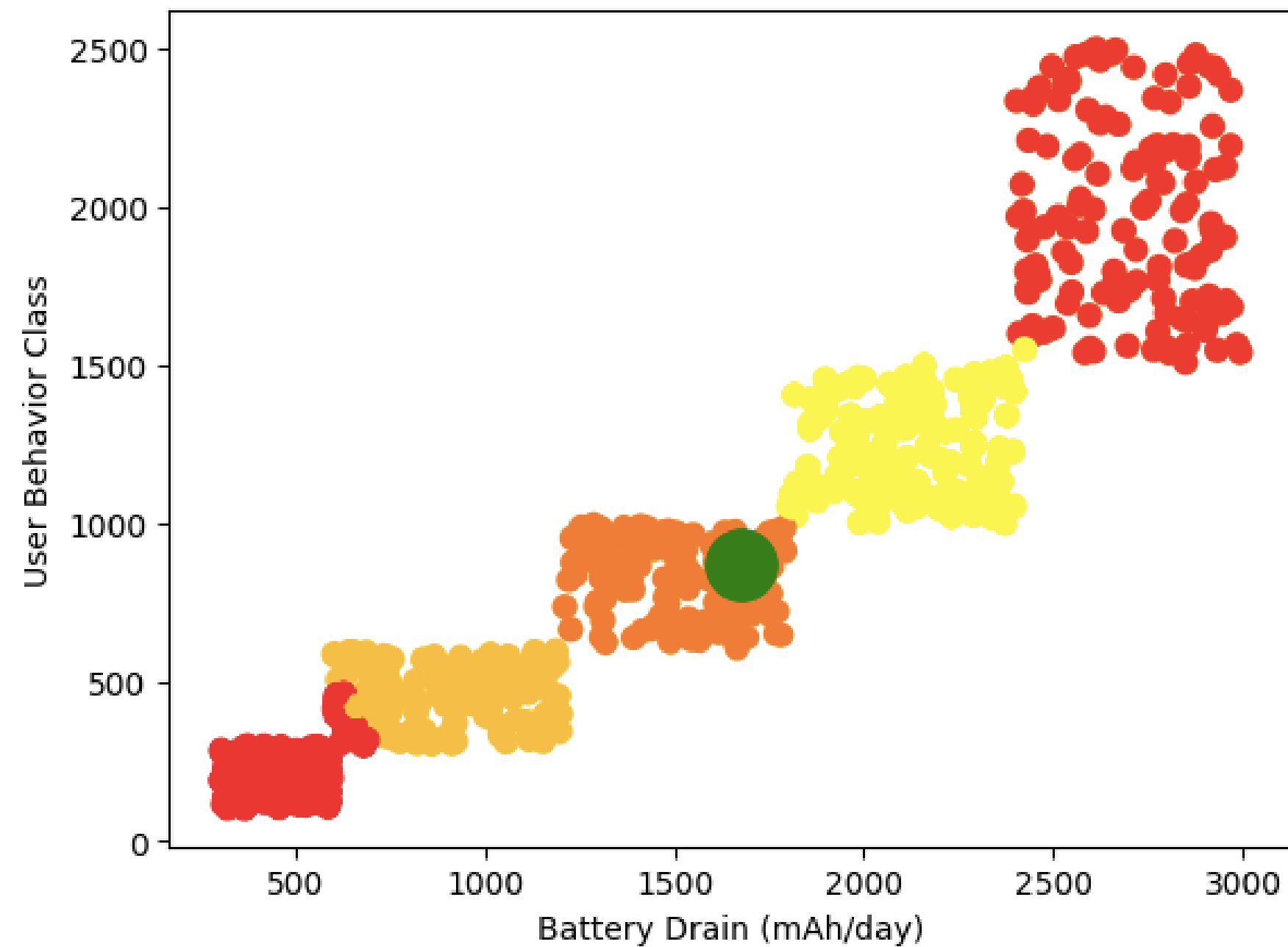
La Sensibilidad Alcanzada por el Modelo es: 1.0

La Especificidad Alcanzada por el Modelo es: 1.0

La clase del comportamiento del usuario: [3]



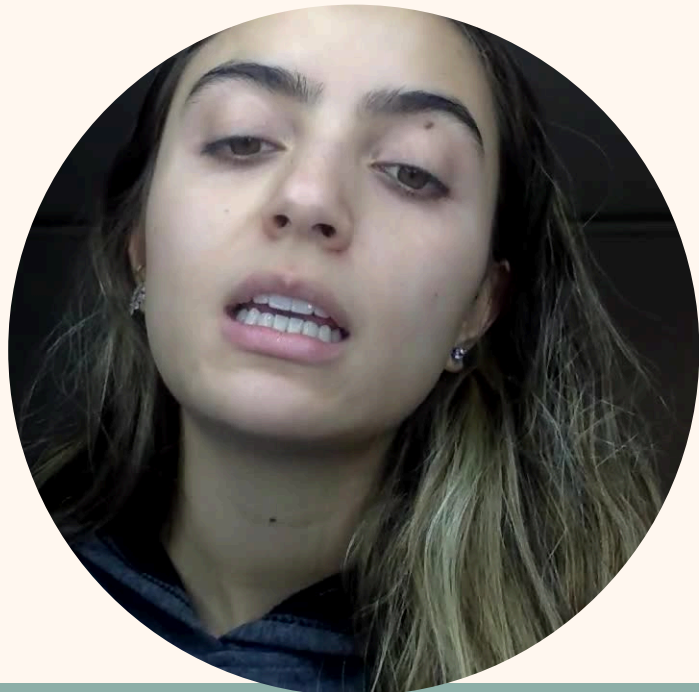
Donde quedó el primer usuario 4
Donde quedó el segundo usuario 2
Donde quedó el tercer usuario 3

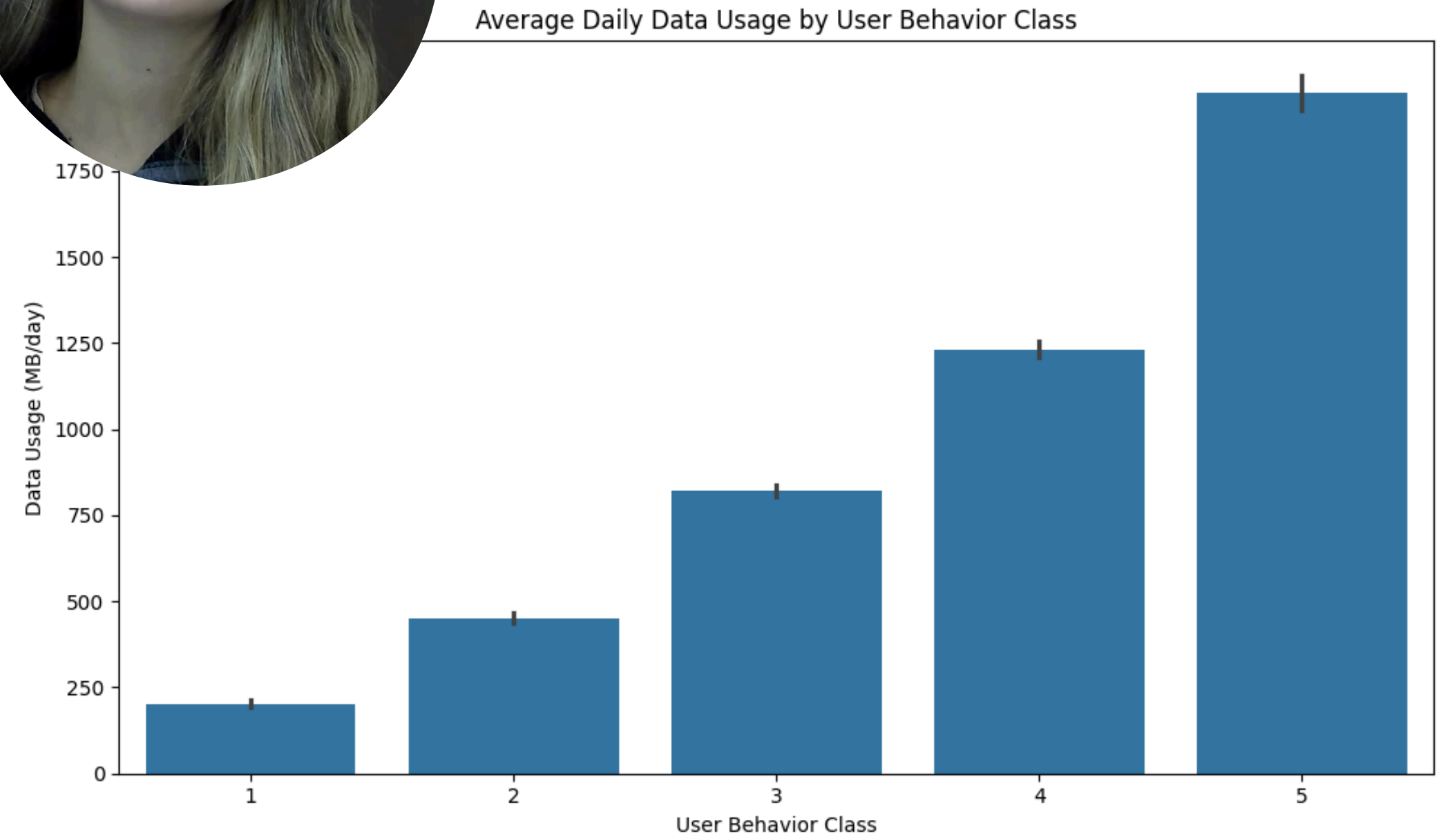
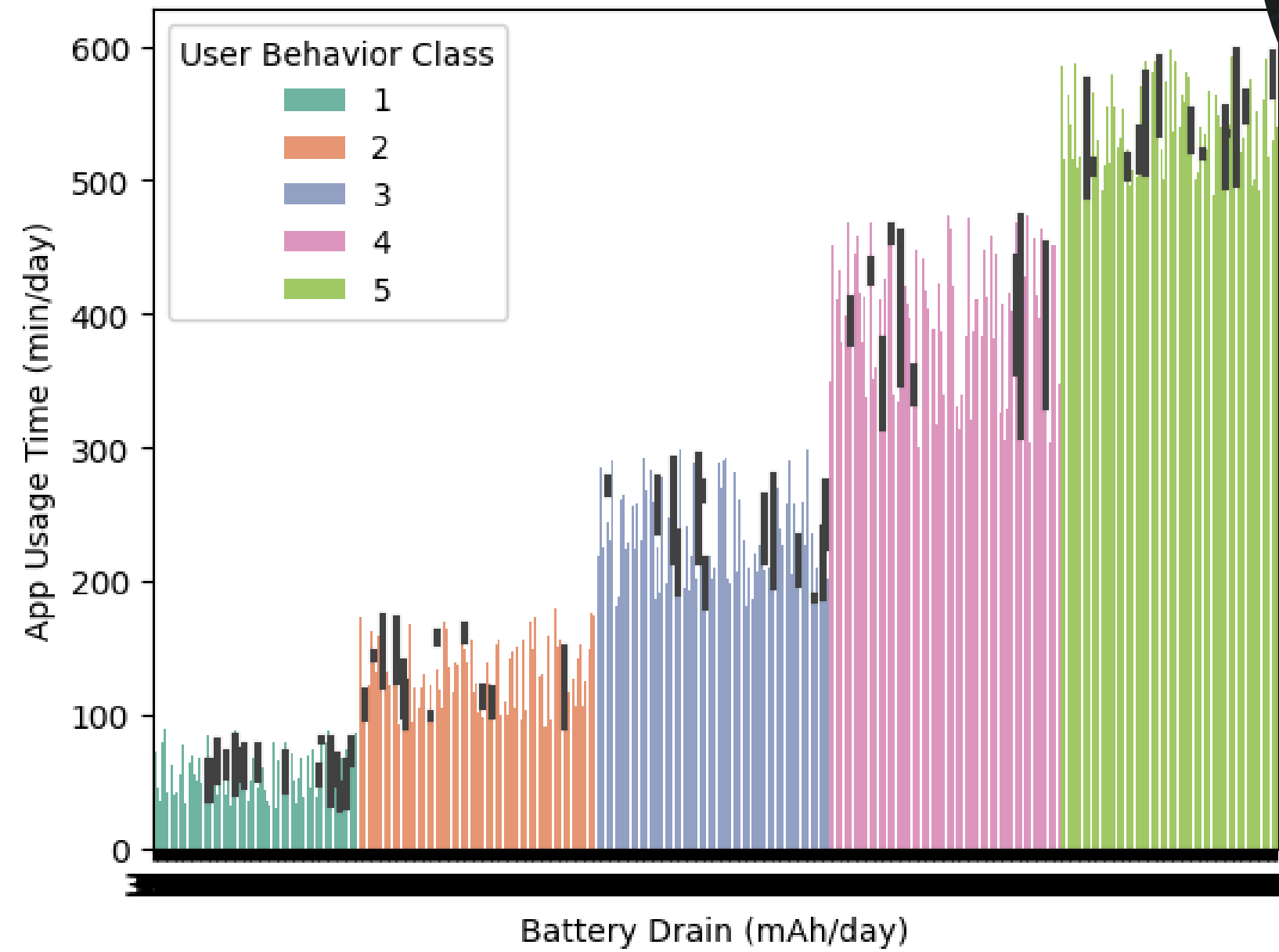
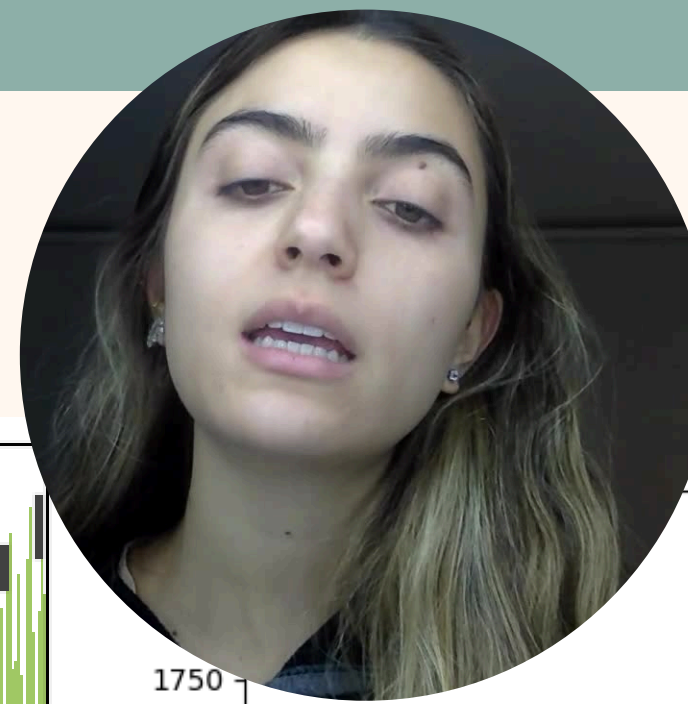


La clase del comportamiento del usuario: [2]

Las características de cada cluster son:

```
[[ 68.17532468  476.0974026   222.94805195   38.32467532]
 [ 541.42222222 2703.06666667 1977.91851852   38.22222222]
 [ 235.3986014  1515.05594406   822.01398601   38.67832168]
 [ 132.7109375   918.703125    461.6328125    38.5703125 ]
 [ 396.78571429 2108.07857143 1234.5          38.62857143]]
```



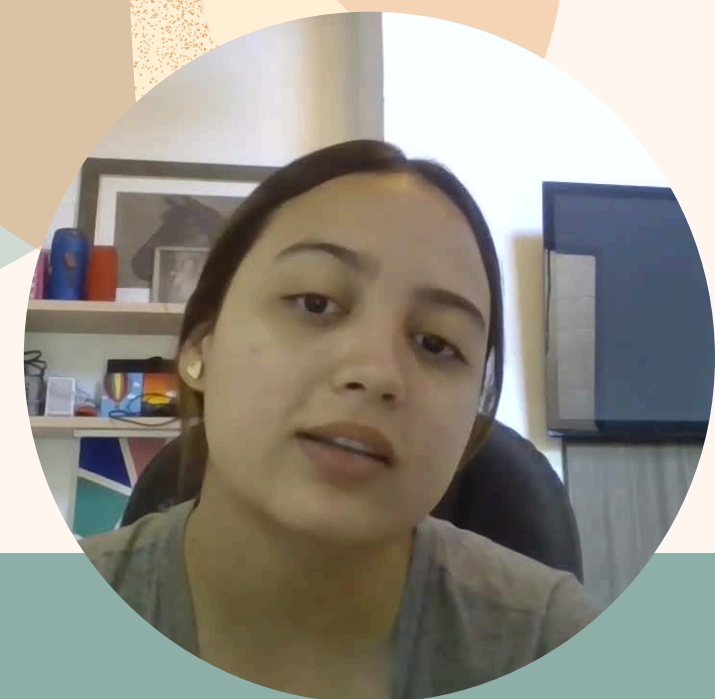


CÓDIGO



Árbol de Decisión

<https://colab.research.google.com/drive/1-VpsTmo7a3McEKGIWk0tYImQbsSEl3Zb?usp=sharing>



ÁRBOL DE DECISIÓN

```
[[136    0    0    0    0]
 [   0 146    0    0    0]
 [   0    0 143    0    0]
 [   0    0    0 139    0]
 [   0    0    0    0 136]]
```

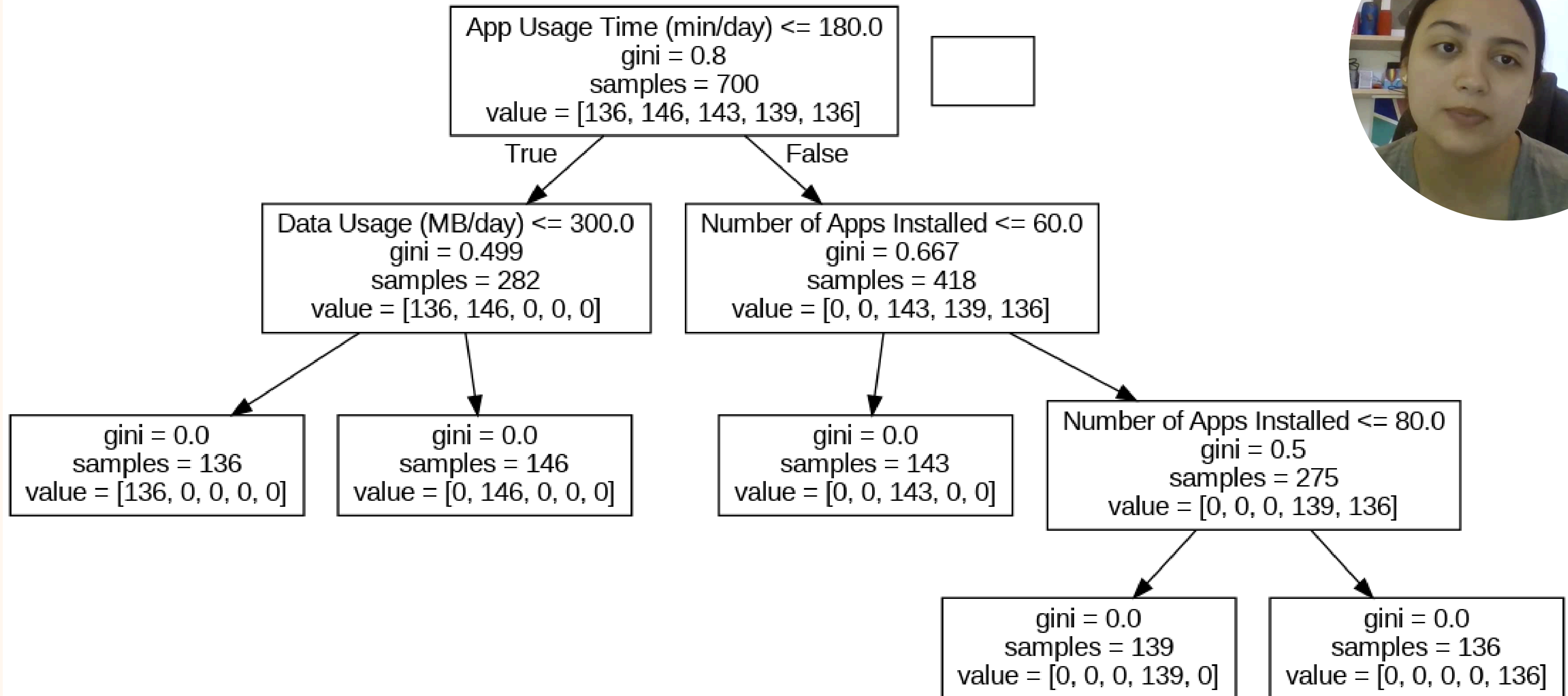
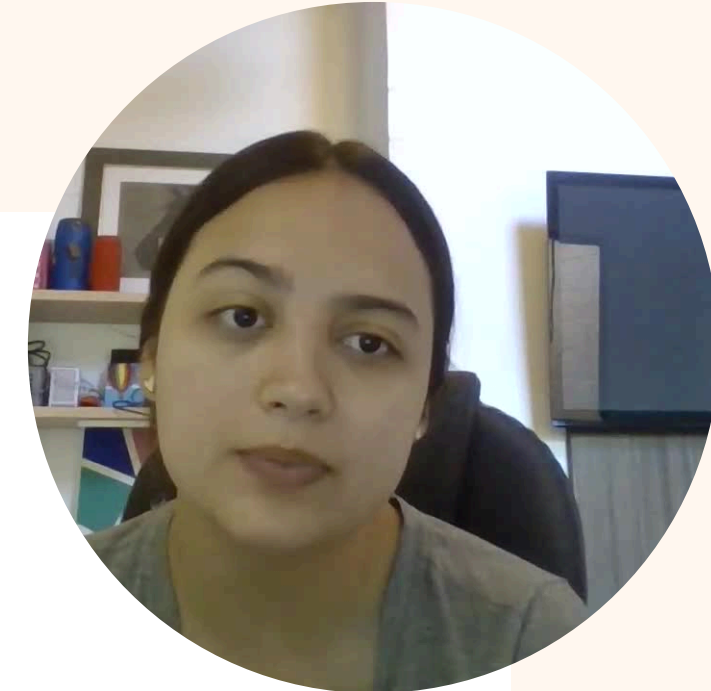
La exactitud alcanzada por el modelo es: 1.0

La sensibilidad alcanzada por el modelo es: 1.0

La especificidad alcanzada por el modelo es: 1.0



ÁRBOL DE DECISIÓN



RESULTADOS

Árbol de decisión

Este árbol clasifica a los usuarios en función de cómo usan las apps, cuántos datos consumen y cuántas apps tienen instaladas.

Las divisiones están bien estructuradas, logrando nodos puros (Gini = 0.0) en cada hoja.

APP USAGE TIME

La variable App Usage Time (min/day) tiene el mayor impacto inicial en la clasificación, ya que es el criterio de división del nodo raíz.



RESULTADOS

Y Conclusiones



MODELOS

100% de precisión, exactitud y especificidad en los dos modelos

CORRELACION

Mayor correlacion entre battery drain y user behavior class.

Sin embargo todas tienen una alta correlacion.

APP USAGE TIME

La variable "App usage time" influye significativamente en la clasificación de los usuarios

EDAD Y GENERO

Baja relevancia en el modelo