



Universidad Nacional de Córdoba  
Facultad de Matemática, Astronomía y Física  
Diplomatura en Ciencia de Datos

## **Presentación de la Mentoría**

### **Predicción de ausentismo de pacientes en turnos clínicos**

**Sandra Mónica Olariaga**

**Mentor: Nahuel Almeira**

# Descripción de la problemática

En esta mentoría trabajaremos con datos de una institución de salud de la Ciudad de Córdoba y la problemática que abordaremos será el ausentismo de los pacientes.

El ausentismo de los pacientes genera pérdidas económicas en la institución, ya que los turnos asignados que no se efectivizan suponen una distribución ineficiente de recursos. La relevancia económica es tal, que la problemática es ampliamente estudiada en todo el mundo.

Buscaremos identificar las variables que conllevan al ausentismo, y construir modelos que permitan predecir la probabilidad de que los turnos no sean atendidos. Evaluaremos la relevancia de cada conjunto de variables (administrativas, de pacientes, meteorológicas).

Con esta información, la institución podría implementar medidas específicas para intentar reducir la tasa de ausentismo, como por ejemplo, enviar recordatorios a los pacientes, redistribuir los horarios según los servicios, etc.



# Análisis y Visualización

El dataset contiene en su mayoría datos categóricos

## **Referidos al paciente:**

Sexo, Edad, Obra Social

## **Referidos al turno:**

Centro de Atención

Servicio

Fecha del Turno

Estado del Turno

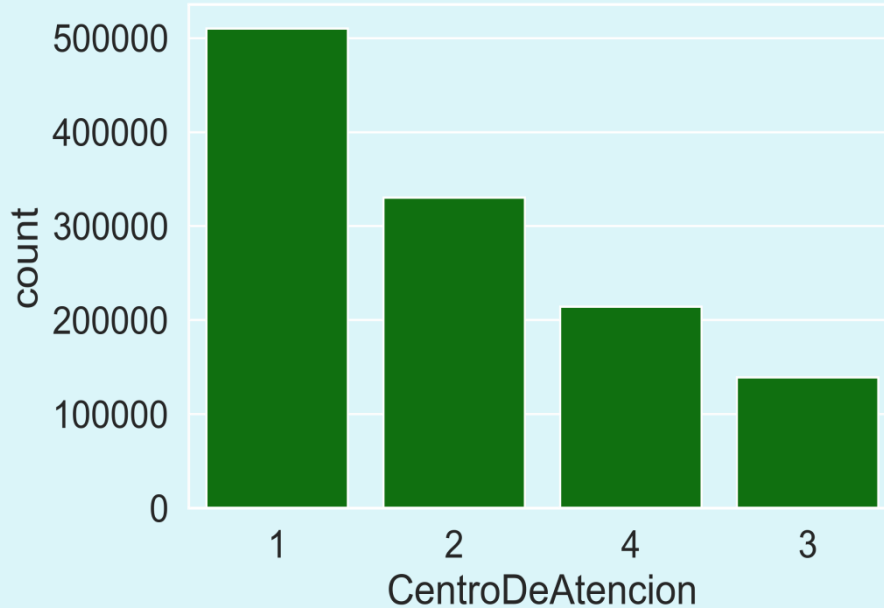
Tipo de Turno

Fecha de Otorgamiento del Turno

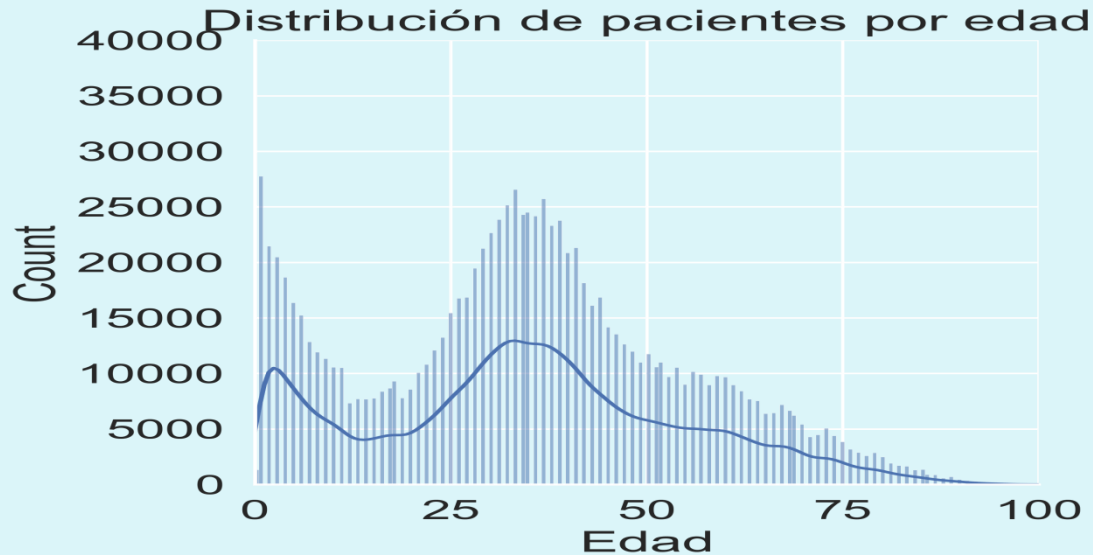
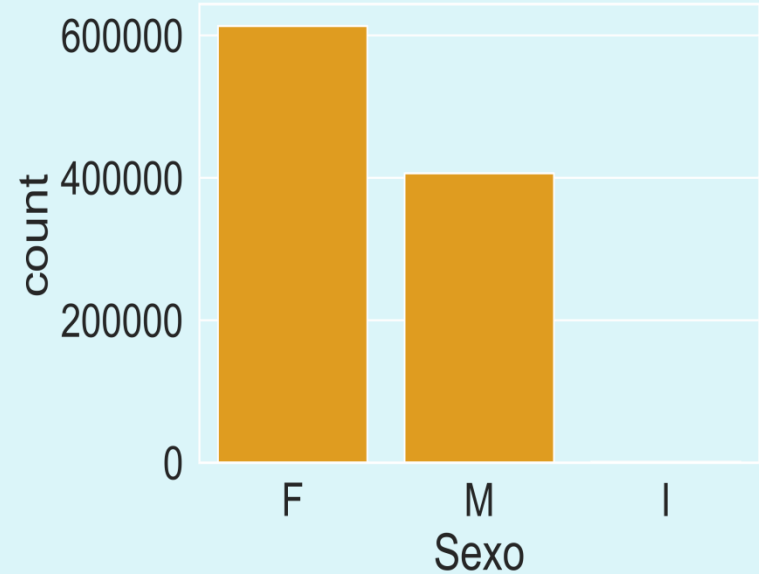
Sobre turno

# Análisis y Visualización

Cantidad por centro de atención

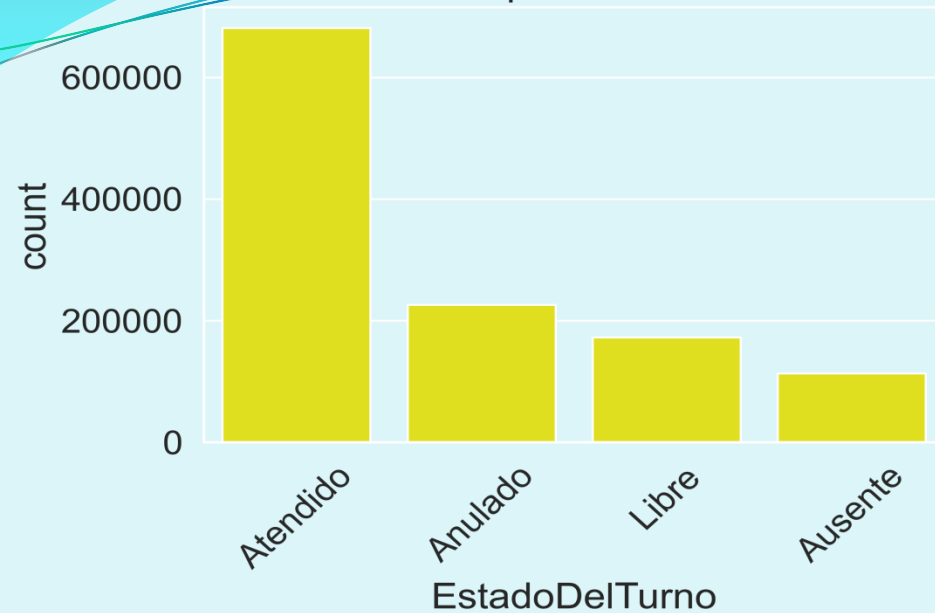


Distribucion por sexo

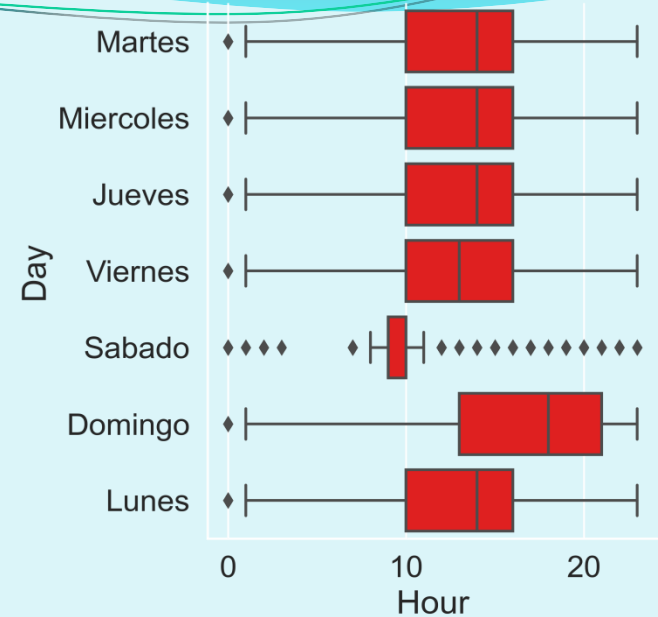


# Análisis y Visualización

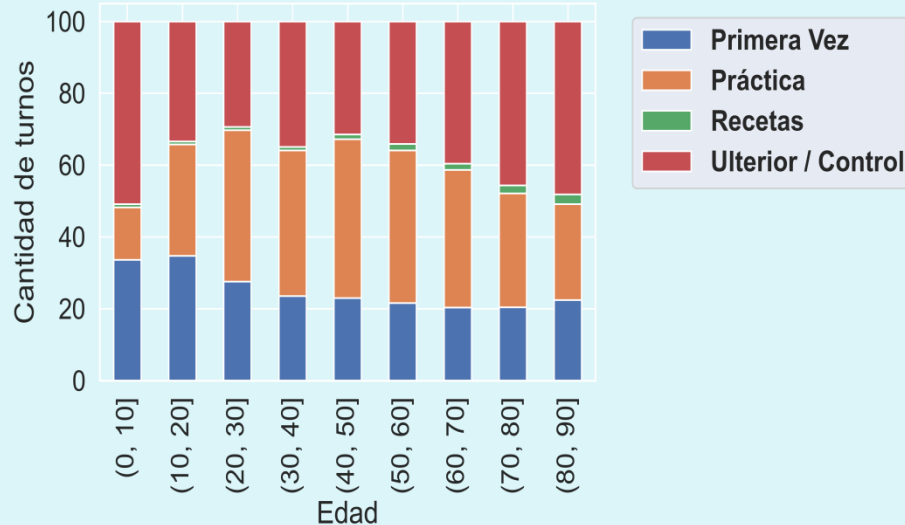
## Distribución por estado del turno



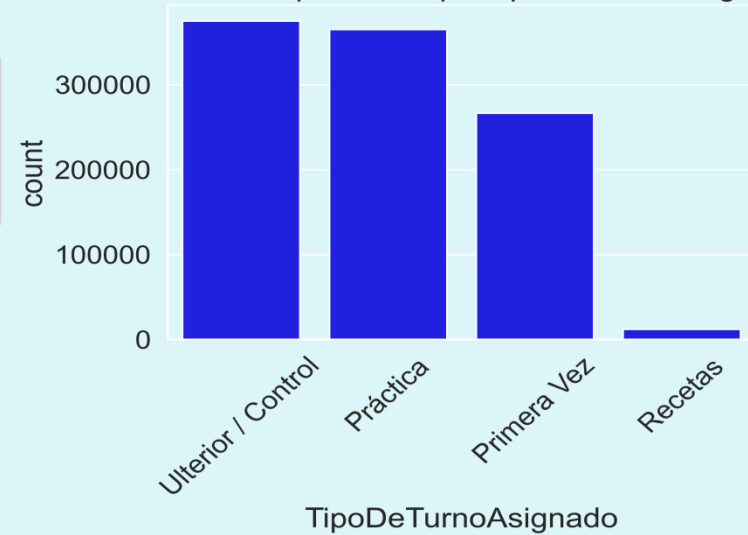
## Turnos por día de la semana y por hora



## Cantidad de turnos por intervalo de edad



## Cantidad de pacientes por tipo de turno asignado



# Análisis y Curación de Datos



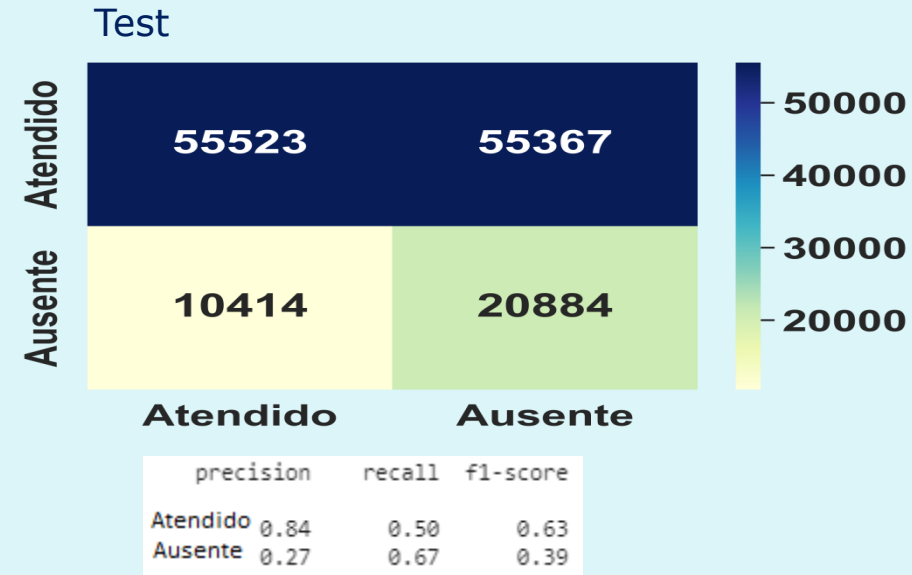
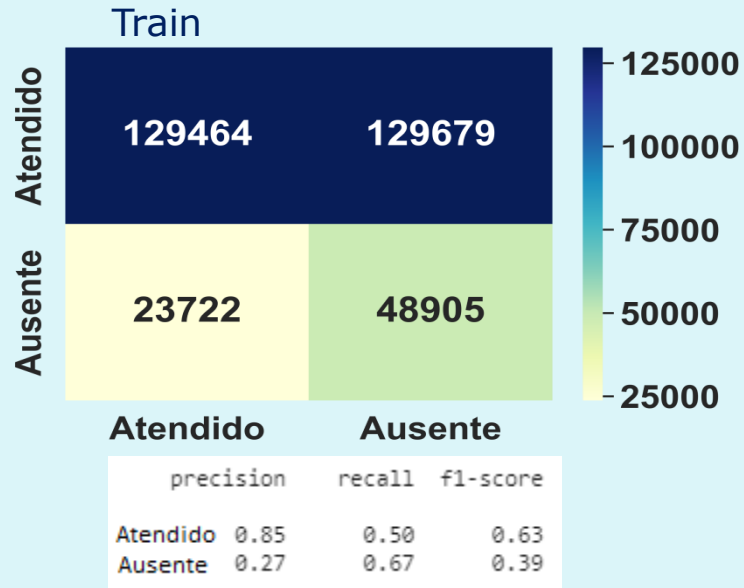
El dataset contenía datos cargados manualmente por lo que debimos corregir palabras mal escritas por ejemplo el nombre de la localidad.

Además:

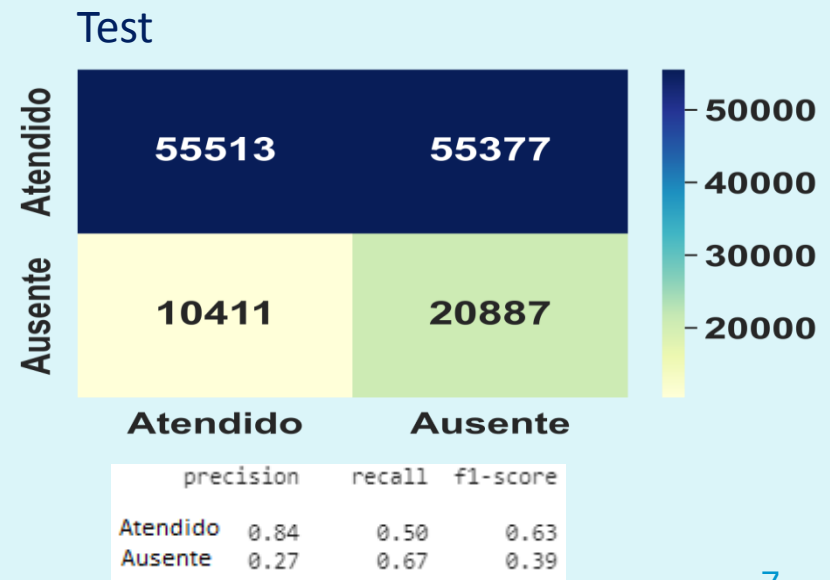
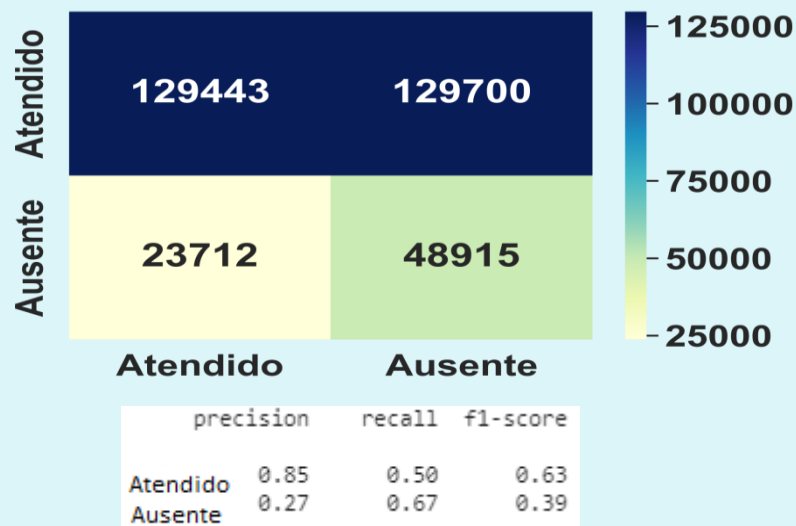
- \* Corregimos los tipos de datos para que sean los adecuados
- \* Restringimos los valores posibles de campos como por ejemplo el estado del turno en atendido (había varios subtipos) y ausente (que es el que nos interesa).
- \* Agregamos columnas con los nombres de los días de la semana.
- \* Descartamos los servicios que sean pediátricos
- \* Descartamos los dos centros de atención menos importantes
- \* Consideramos consumidor final o no consumidor final (obras sociales)
- \* Restringimos sexo a femenino y masculino
- \* Eliminamos registros duplicados

# Aprendizaje Automático

## \*Modelo lineal Regresión Logística :

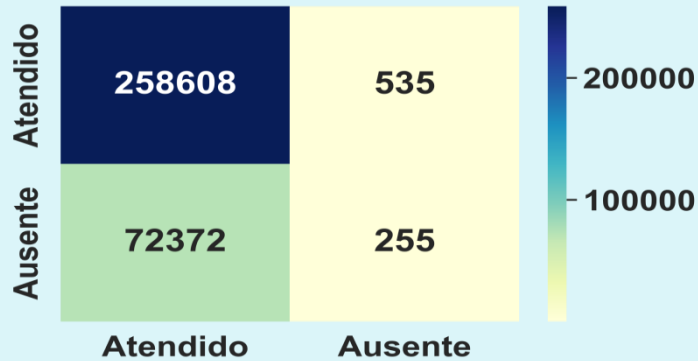


## Con MinMaxScaler: Train



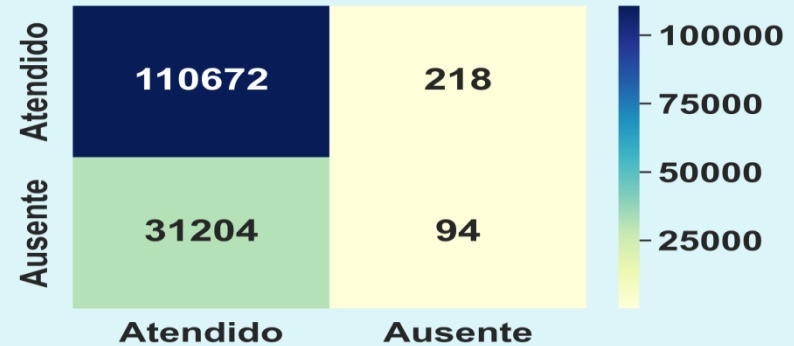
# Aprendizaje Automático

\* SGD Classifier: Train



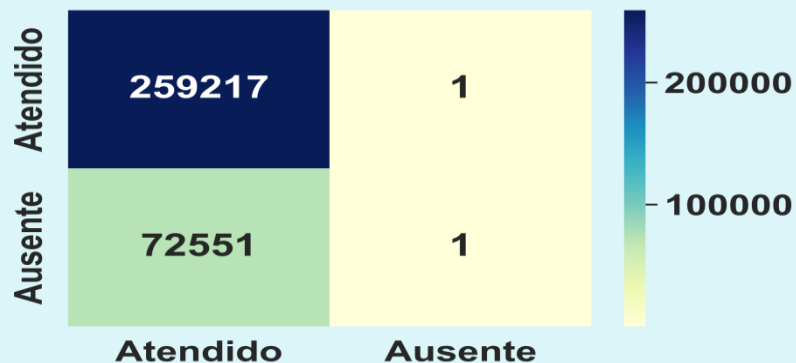
	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.32	0.00	0.01

Test



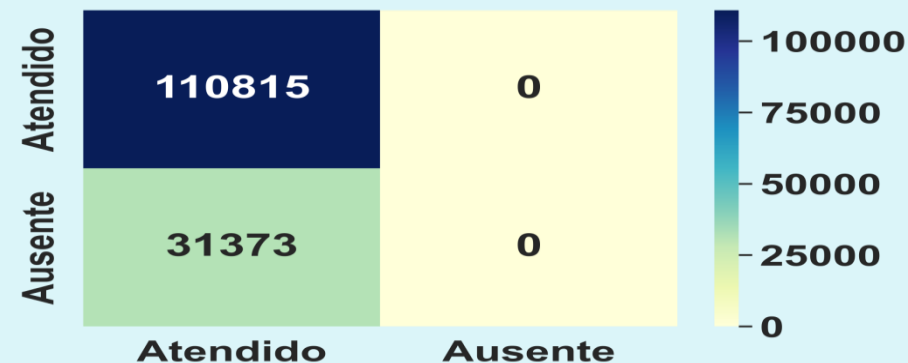
	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.30	0.00	0.01

Con MinMaxScaler: Train



	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.50	0.00	0.00

Test



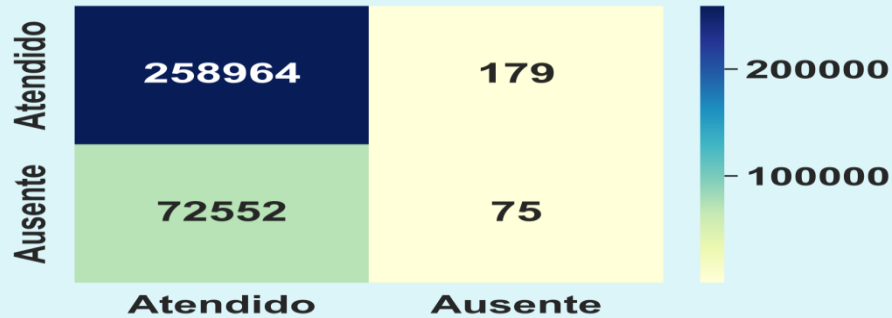
	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.00	0.00	0.00



# Aprendizaje Automático

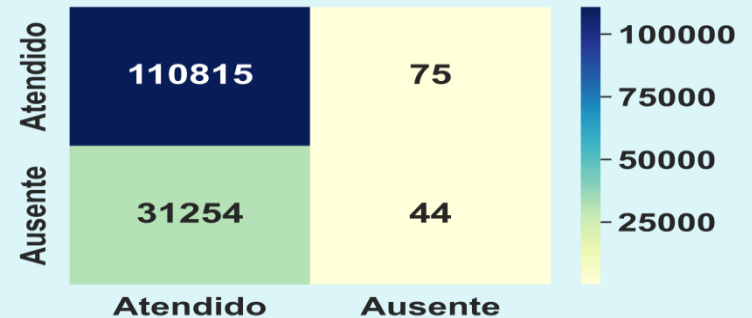
## \* Clasificador Polinomial:

Train



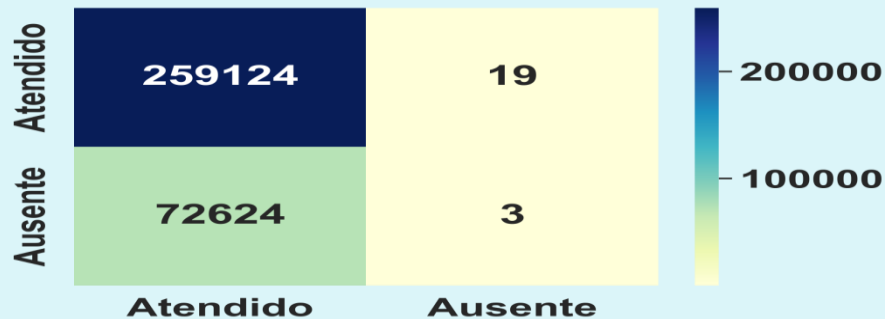
	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.30	0.00	0.00

Test



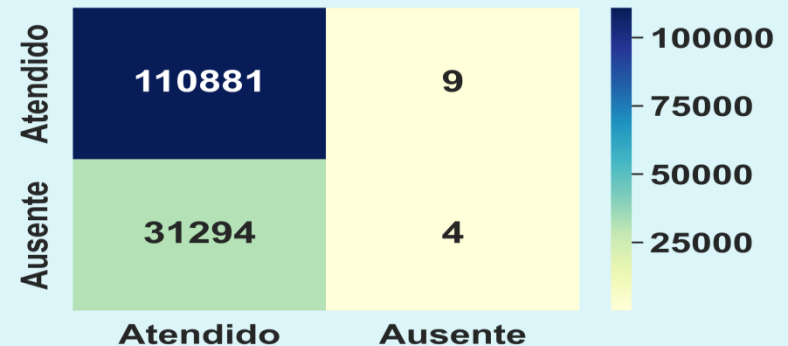
	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.37	0.00	0.00

## Con MinMaxScaler: Train



	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.14	0.00	0.00

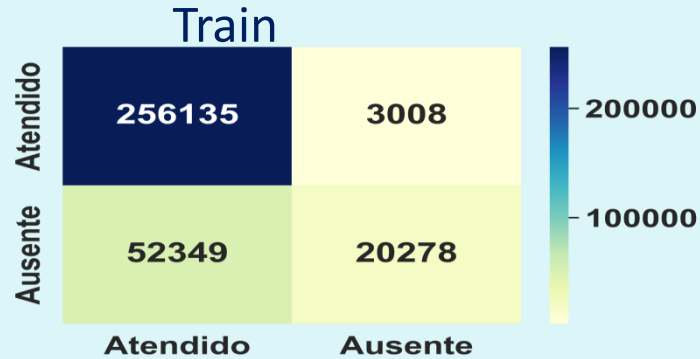
Test



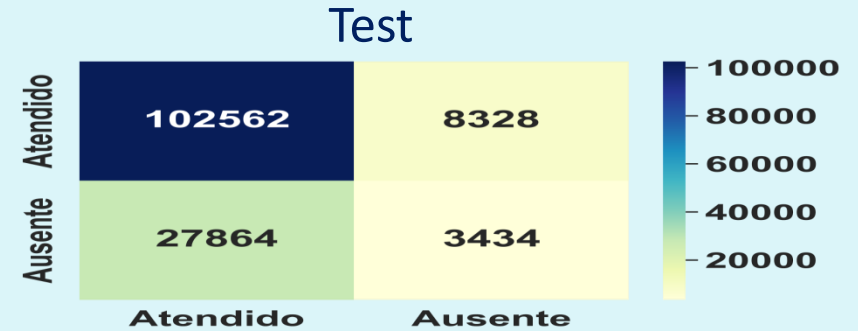
	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.31	0.00	0.00

# Aprendizaje Automático

## \* Decision Tree Classifier:

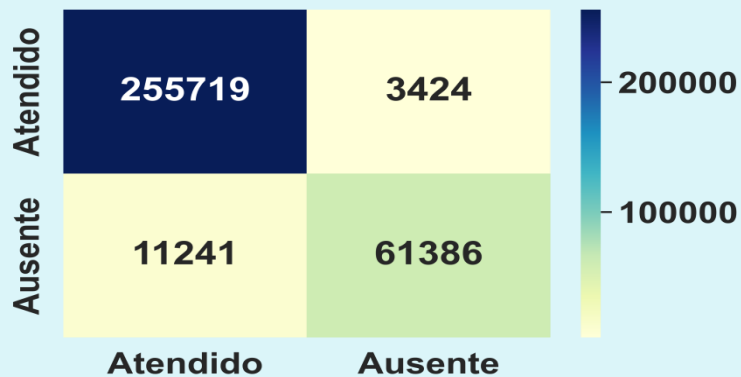


	precision	recall	f1-score
Atendido	0.83	0.99	0.90
Ausente	0.87	0.28	0.42

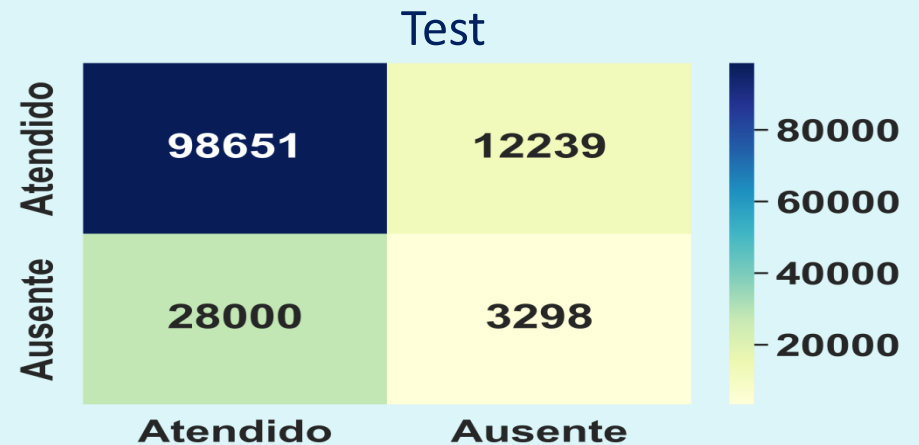


	precision	recall	f1-score
Atendido	0.79	0.92	0.85
Ausente	0.29	0.11	0.16

## \* Random Forest: Train



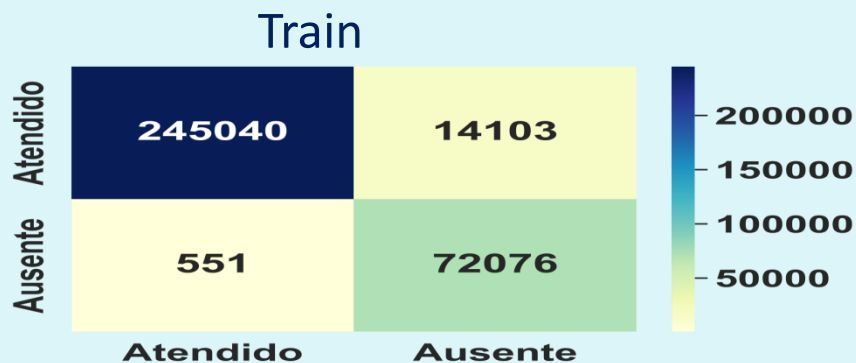
	precision	recall	f1-score
Atendido	0.96	0.99	0.97
Ausente	0.95	0.84	0.89



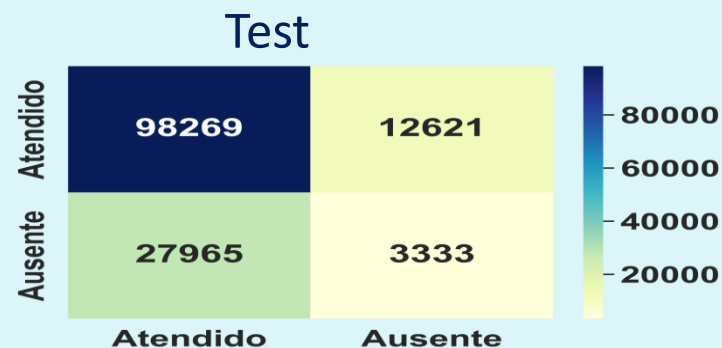
	precision	recall	f1-score
Atendido	0.78	0.89	0.83
Ausente	0.21	0.10	0.14

# Aprendizaje supervisado

## \* Random Forest con GridSearchCV :

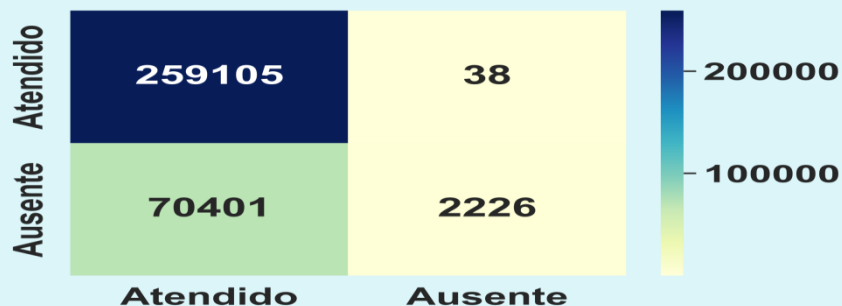


	precision	recall	f1-score
Atendido	1.00	0.95	0.97
Ausente	0.84	0.99	0.91

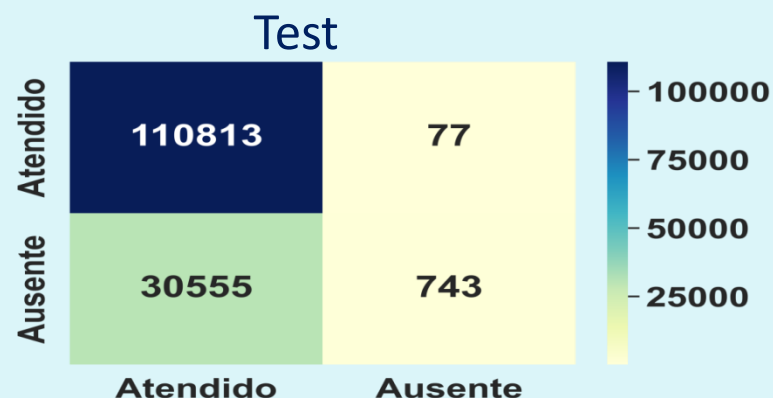


	precision	recall	f1-score
Atendido	0.78	0.89	0.83
Ausente	0.21	0.11	0.14

## \* Bayes SearchCV: Train



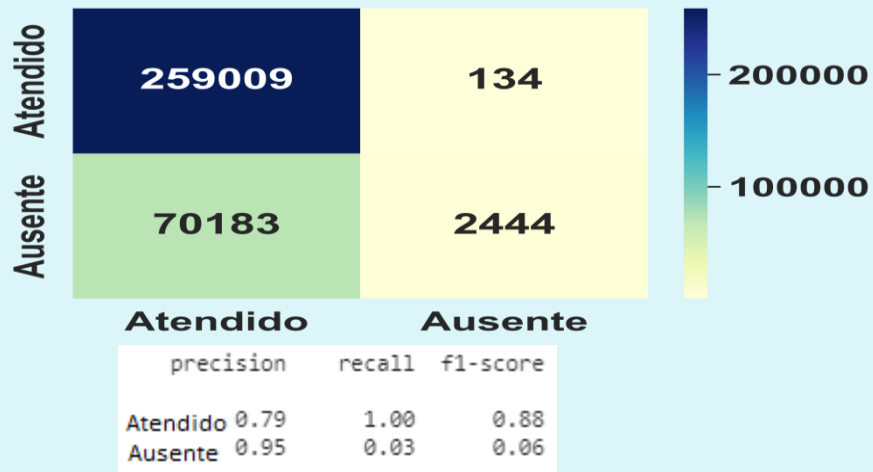
	precision	recall	f1-score
Atendido	0.79	1.00	0.88
Ausente	0.98	0.03	0.06



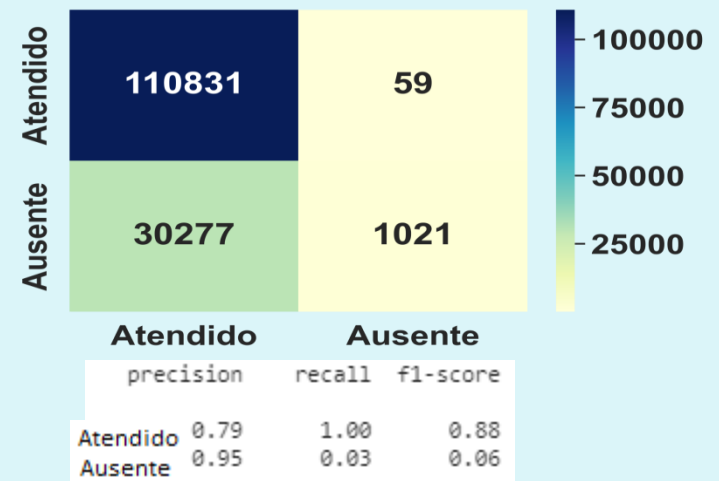
	precision	recall	f1-score
Atendido	0.78	1.00	0.88
Ausente	0.90	0.02	0.04

# Aprendizaje supervisado

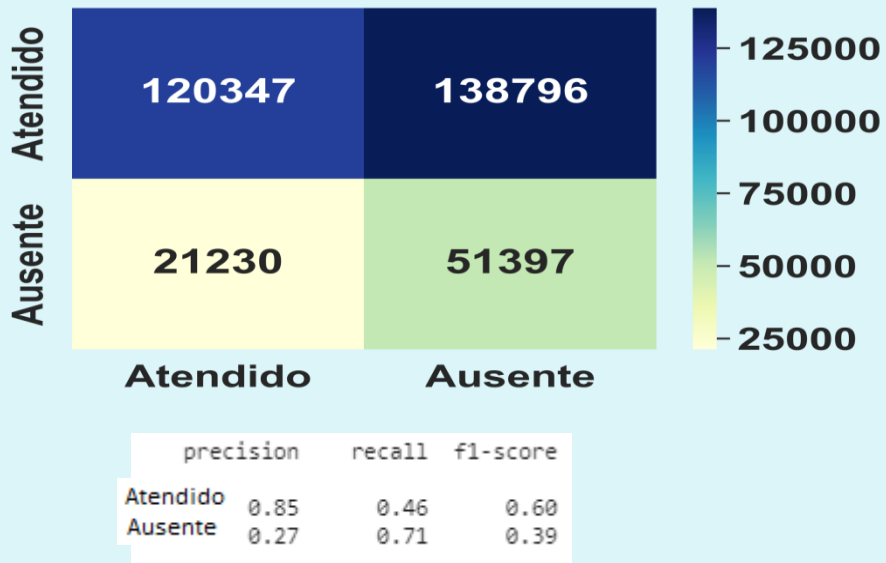
## \* Boosting con xgboost: Train



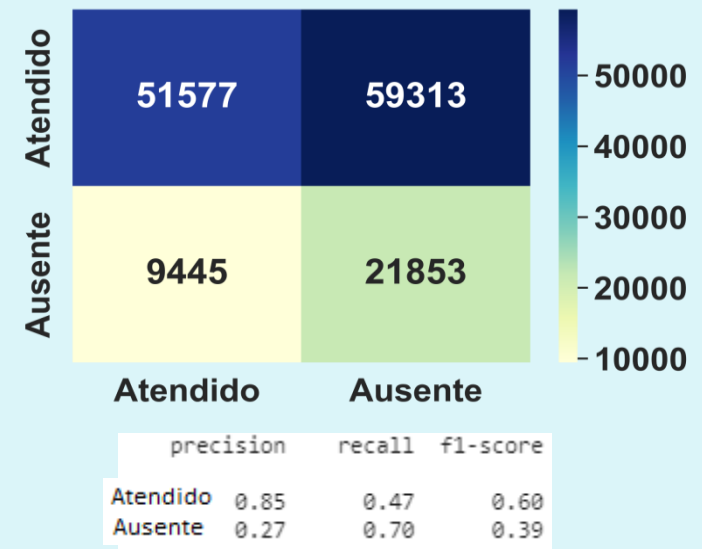
## Test



## \* LinearSVC: Train



## Test



# Aprendizaje No supervisado

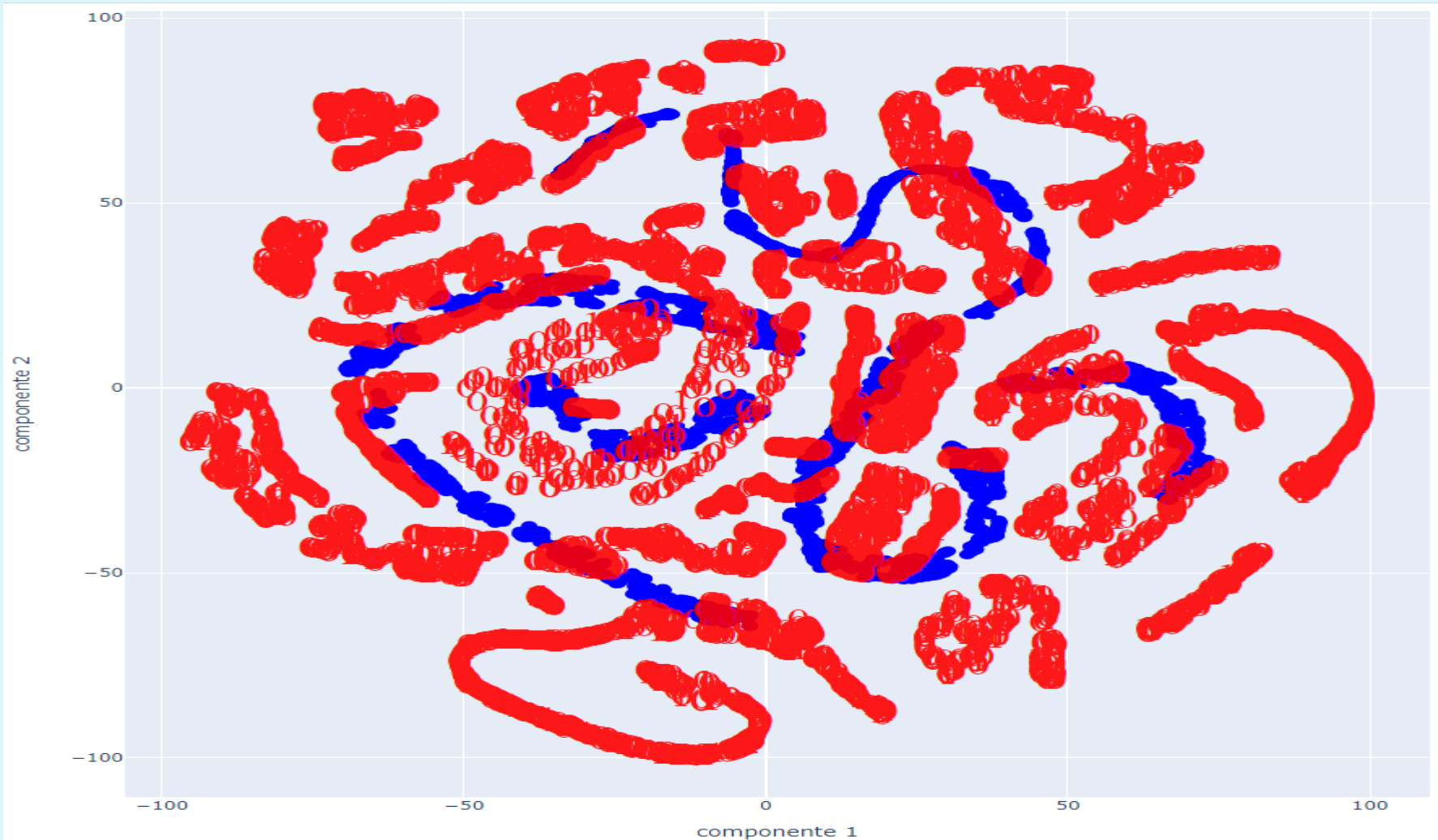
Visualizacion-con-embeddings:tSNE



Ausente

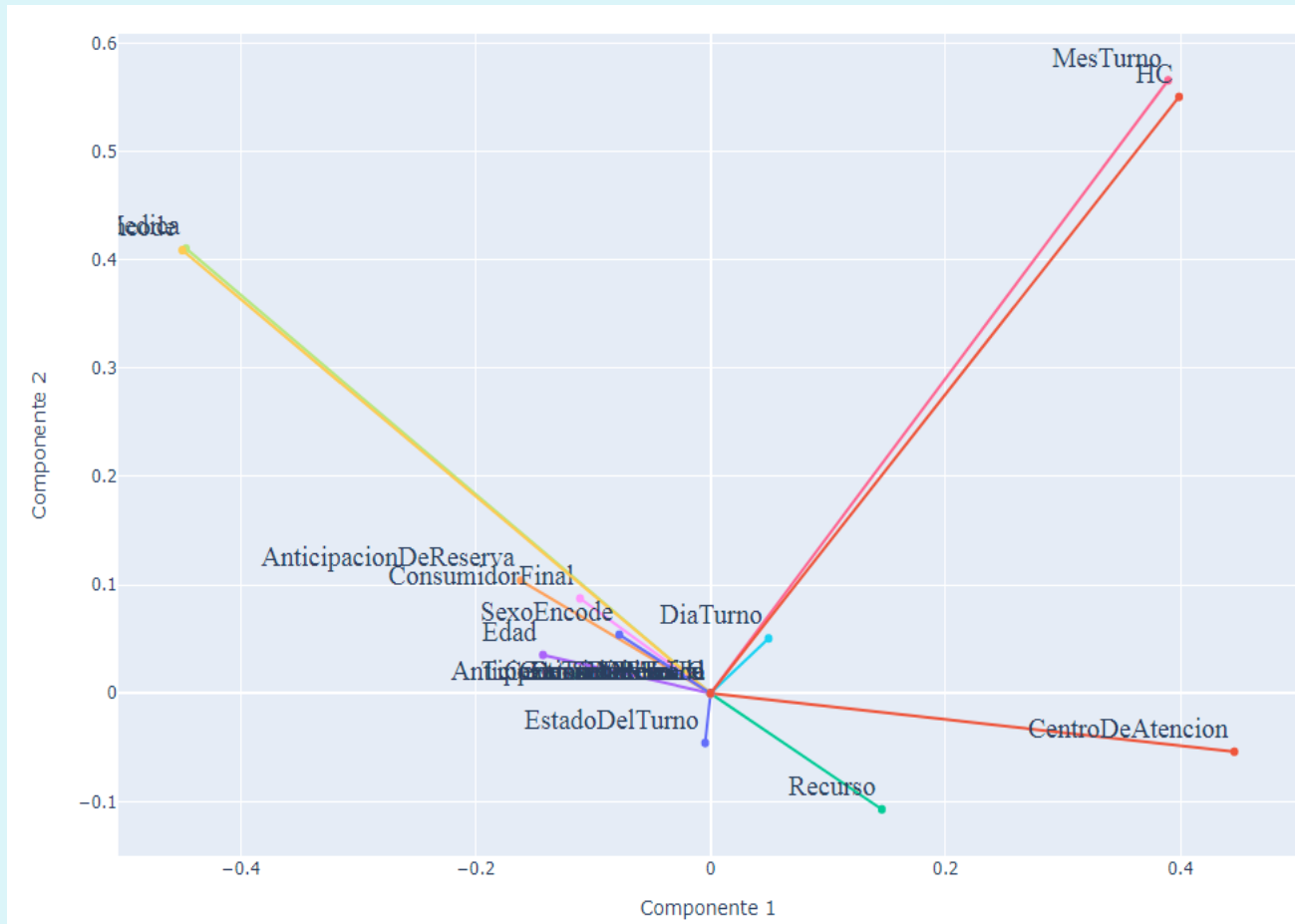


Atendido



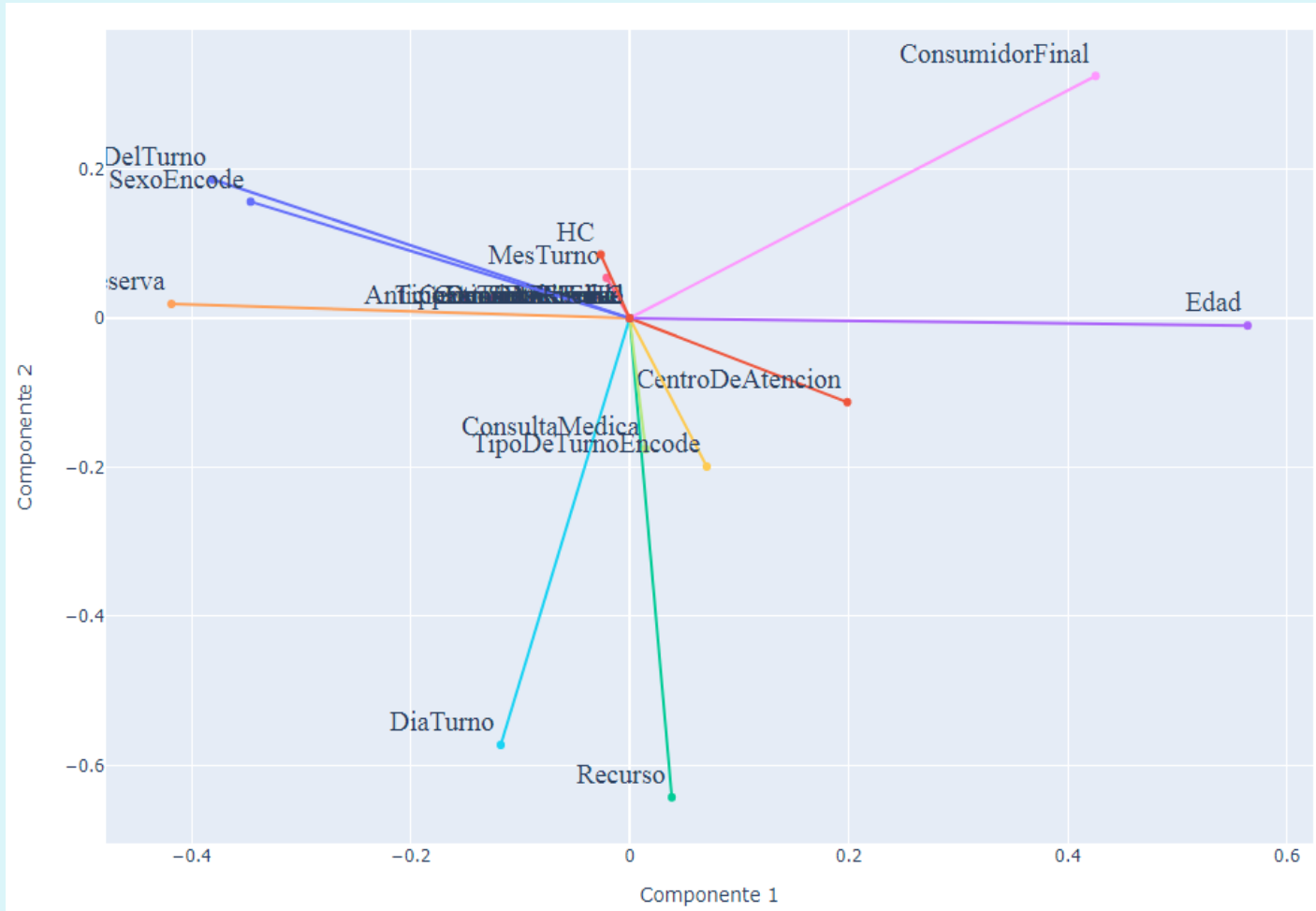
# Aprendizaje No Supervisado

## Grafico dos primeras componentes principales



# Aprendizaje No Supervisado

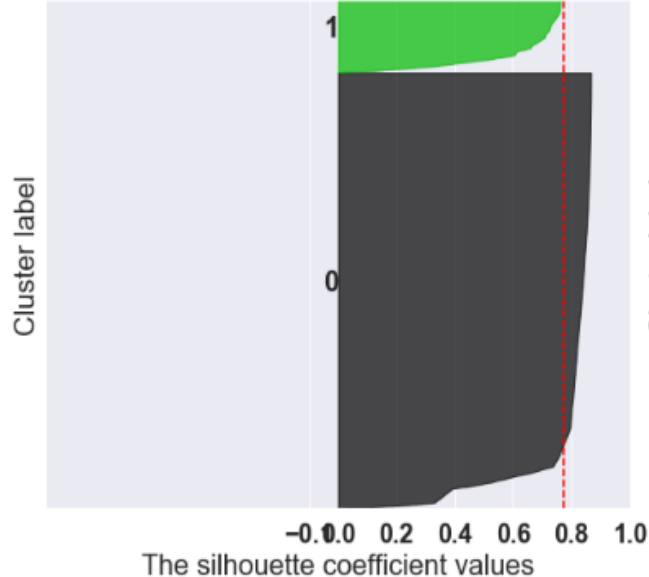
## Grafico otras dos componentes principales



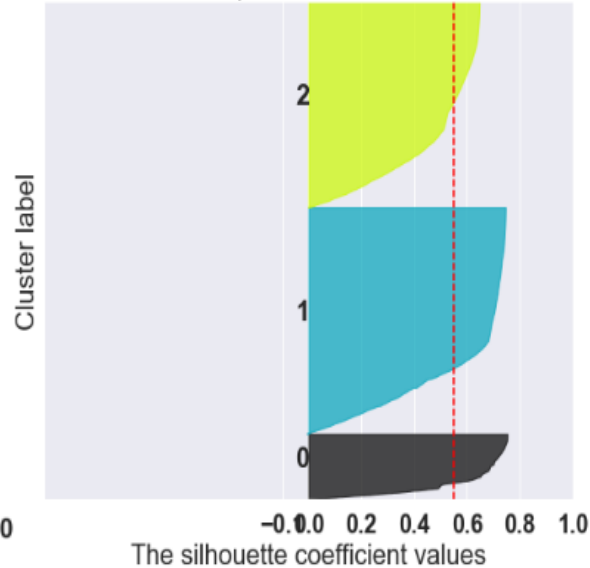
# Aprendizaje No Supervisado

## Análisis de Silhouette La más homogenea sería la de 6 clusters

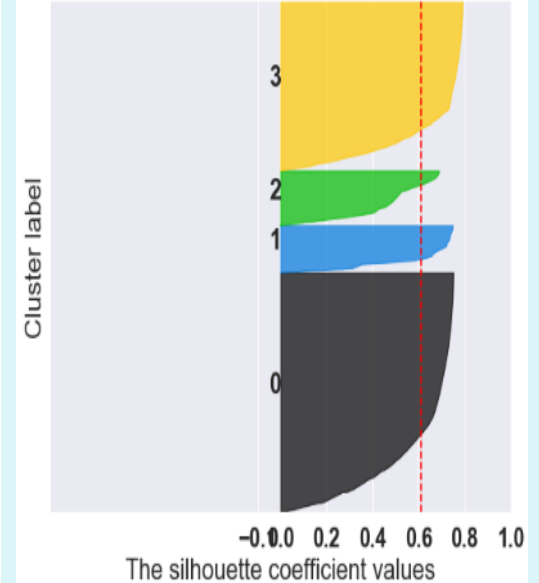
The silhouette plot for the various clusters.



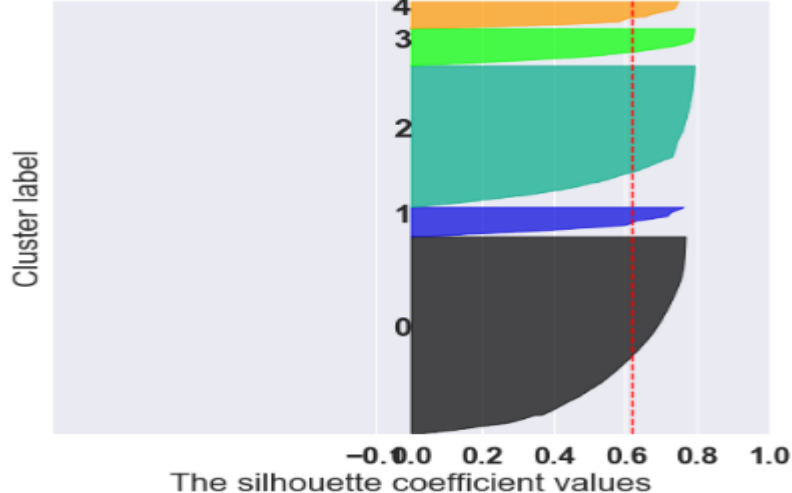
The silhouette plot for the various clusters.



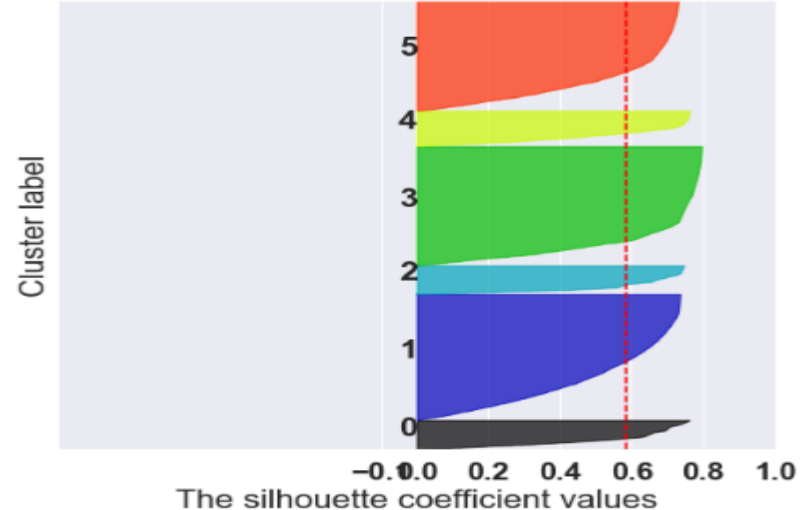
The silhouette plot for the various clusters.



The silhouette plot for the various clusters.



The silhouette plot for the various clusters.





# Conclusiones

En primer lugar nos tomamos su tiempo analizar y comprender el dataset con el que debíamos trabajar y de allí comenzar a implementar los modelos de aprendizaje supervisado y luego hacer análisis en el aprendizaje no supervisado.

En los resultados obtenidos se observa que el dataset provisto no contiene toda la información que sería necesaria para poder hacer una mejor clasificación de los datos de entrada, quizás con algún otro tipo de datos, tales como si el paciente está bajo un tratamiento específico o no, o si tiene alguna enfermedad crónica o no, etc.

La incorporación de estos otros datos quedará para un futuro análisis.