

School AAA

PROCESO DE RETENCIÓN Y CONSECUCCIÓN DE ESTUDIANTES

AUTOR. SANDRA PATRICIA CANO OBANDO

AGENDA

01. Contexto y Audiencia
02. Hipótesis/Preguntas de Interés
03. Metadata
04. Análisis Exploratorio
05. Insights y Recomendaciones

01. Contexto y Audiencia

Contexto: La Institución educativa AAA School, es una institución de educación internacional que comercializa servicios de turismo para estudiantes de diferentes regiones. Brinda programas de formación en sus instalaciones y a la vez da a conocer los principales sitios turísticos de la ciudad en donde está ubicada, cuenta con un aplicativo de inscripciones en un CRM que les permite guardar la información de los estudiantes que han tomado los paquetes de formación y los planes de viaje con ellos directamente.

Audiencia: Este análisis pretende ayudar al área de marketing de la institución y a las principales directivas a encontrar en donde focalizar sus esfuerzos de promoción y asignación presupuestal, para retener estudiantes, fomentar la renovación de las matrículas y permanencia de los estudiantes en los cursos de la institución.

► Contexto Analítico

Para este análisis, se cuenta con la base de datos de las inscripciones del año anterior.

La información contiene las fechas en que fueron registrados los estudiantes, si estos tomaron el paquete de viaje o no, también si los estudiantes cuentan con los ingresos familiares no solo que les permita costear el programa inicial sino, matricularse en programas adicionales con la institución.

► Problemática

El departamento de marketing de la institución requiere, conocer los principales estudiantes que se matriculan en sus programas y si renuevan matrícula para tomar nuevos cursos al año siguiente, por otra parte, requieren conocer de donde provienen los alumnos frecuentes y en que estados se debería hacer mayor promoción de su oferta académica, para lograr la consecución de mayores programas contratados.

► Objetivo

El departamento de marketing podrá con esta información, realizar una estrategia de mercadeo más eficiente y focalizada, que les permita orientarse a los estudiantes con la capacidad de tomar no solo uno sino varios programas en la institución, reteniendo a los estudiantes objetivo.



02. Preguntas de Interés

- ▶ 1. Podemos predecir la cantidad de estudiantes que se van a mantener en la institución en el futuro?
- ▶ 2. Podemos saber cual es el Estado en donde mas participacion de estudiantes tenemos matriculados en el instituto, para enfocar planes de accion?
- ▶ 3. Podemos saber el tipo de viaje que toman los estudiantes que estan matriculados en la institution.

03. Metadata

una vez analizada la información, se puede evidenciar que la data cuenta con información de fechas de viaje, fechas de pago, fechas de inscripción, tipo de programa, numero de matrícula, nivel de ingresos, grado de los estudiantes actuales que van a tomar la inducción o los programas, fechas de reuniones con los padres de los estudiantes, procedencia del estudiante, tipo de viaje a contratar, formas de pago, entre otras características, para lo cual adjuntamos el Dataset y sus características.



2389 Registros

55 variables

4 Tipos de viaje

3 Formas de pago

54 Estados donde provienen los estudiantes

28 Programas académicos

6 Regiones

04. Análisis Exploratorio

- ▶ Utilizaremos, diferentes gráficos y modelos, que nos permitieron evaluar diferentes variables dentro del dataset, para poder llegar a unas conclusiones para el equipo de marketing y los directivos.

Gráfico Bivariada Barplot

- Teniendo en cuenta el tipo de viaje que realizan los estudiantes, se quiere evidenciar con un gráfico el número promedio de días de alojamiento que toman en su paquete para la inducción.

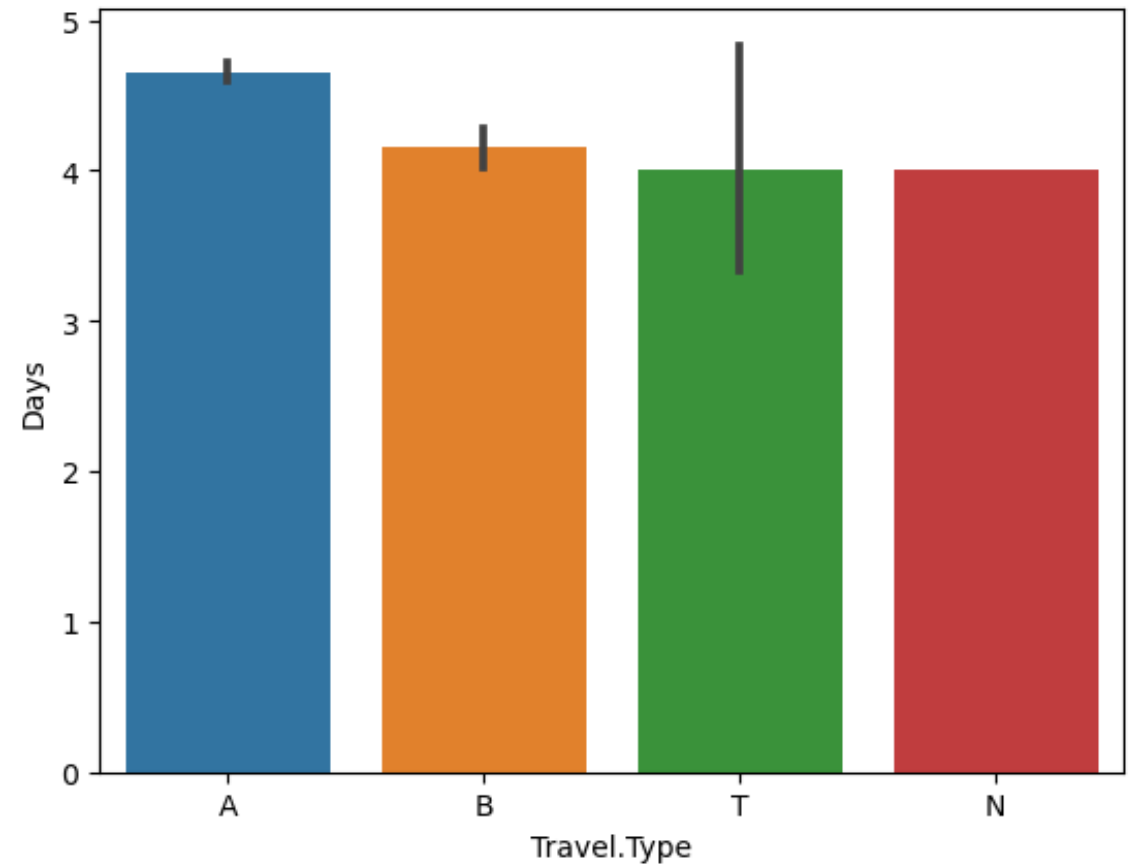
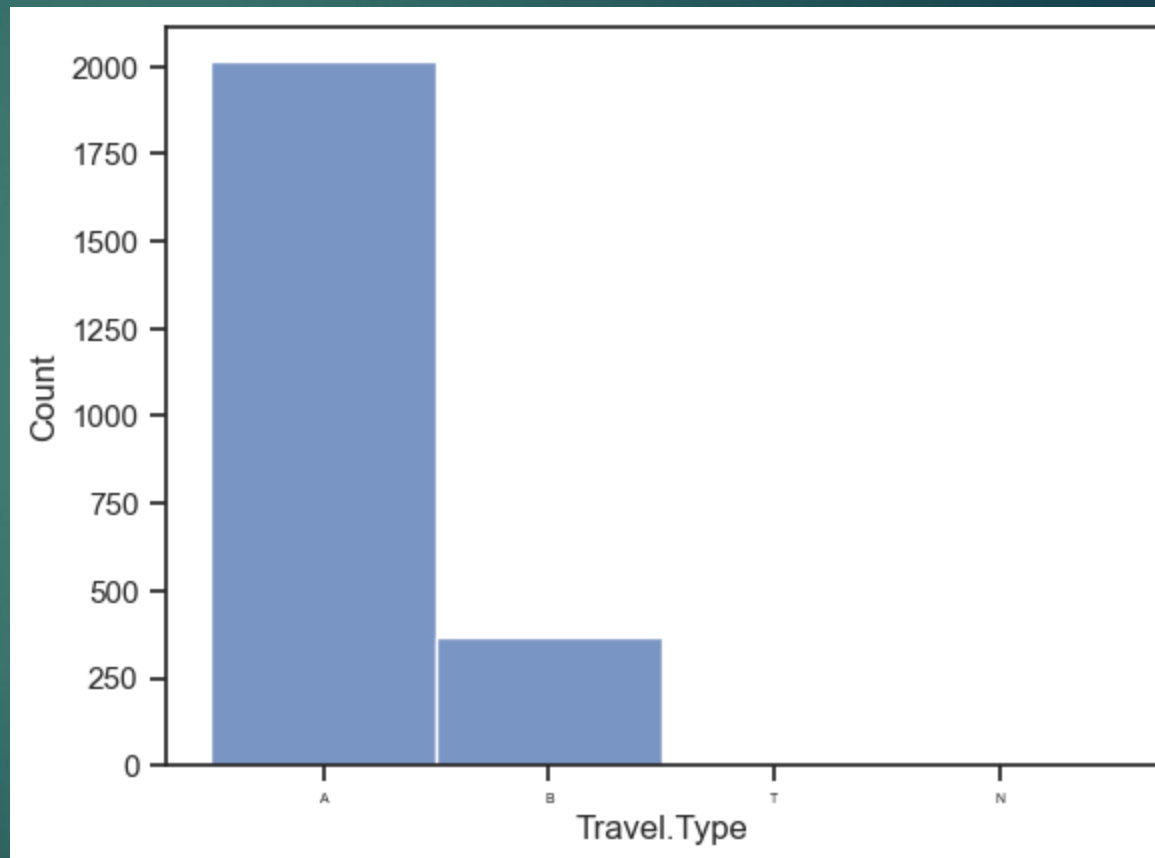


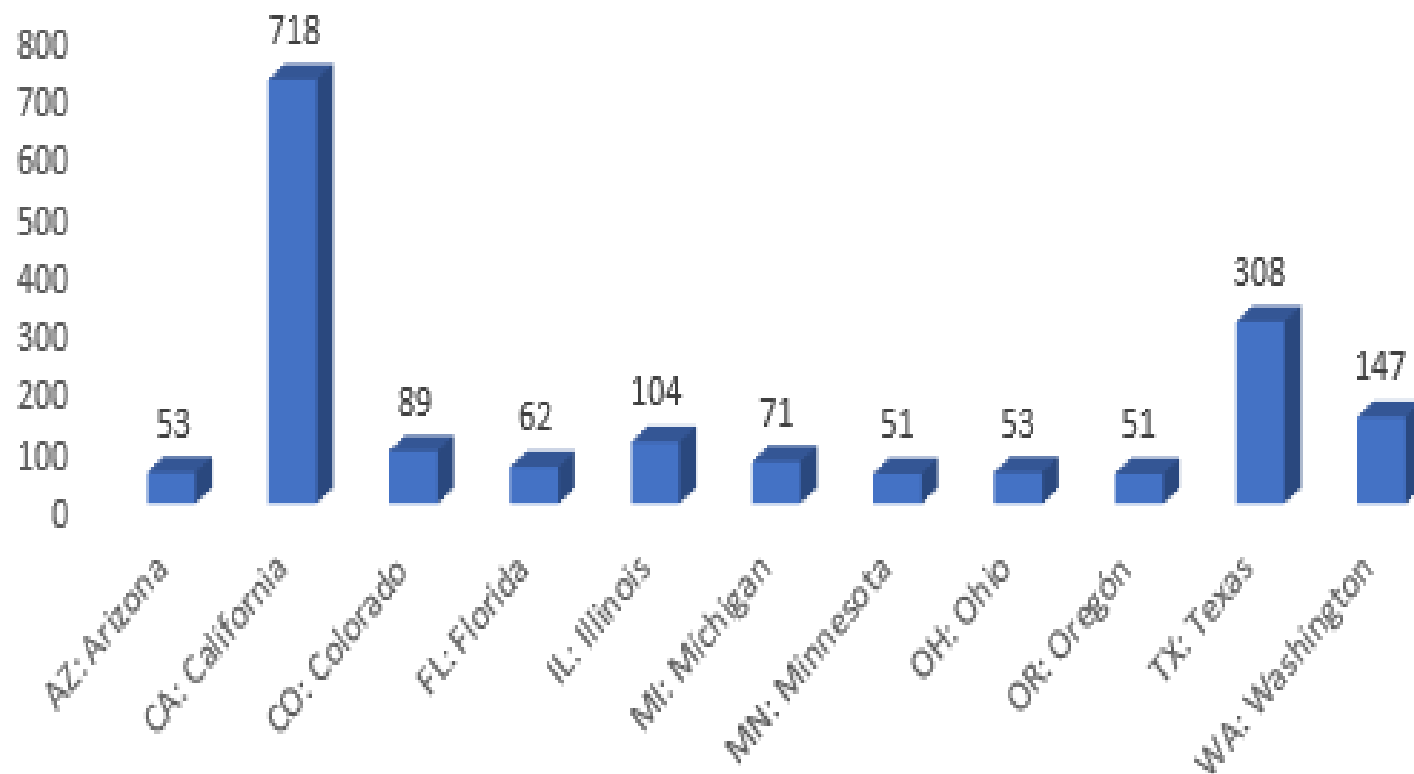
Gráfico Univariado

- Vemos interesante también ir analizando que tipo de paquete toman los estudiantes, cual es el de mayor frecuencia y así, ir determinando si hay o no una preferencia notable

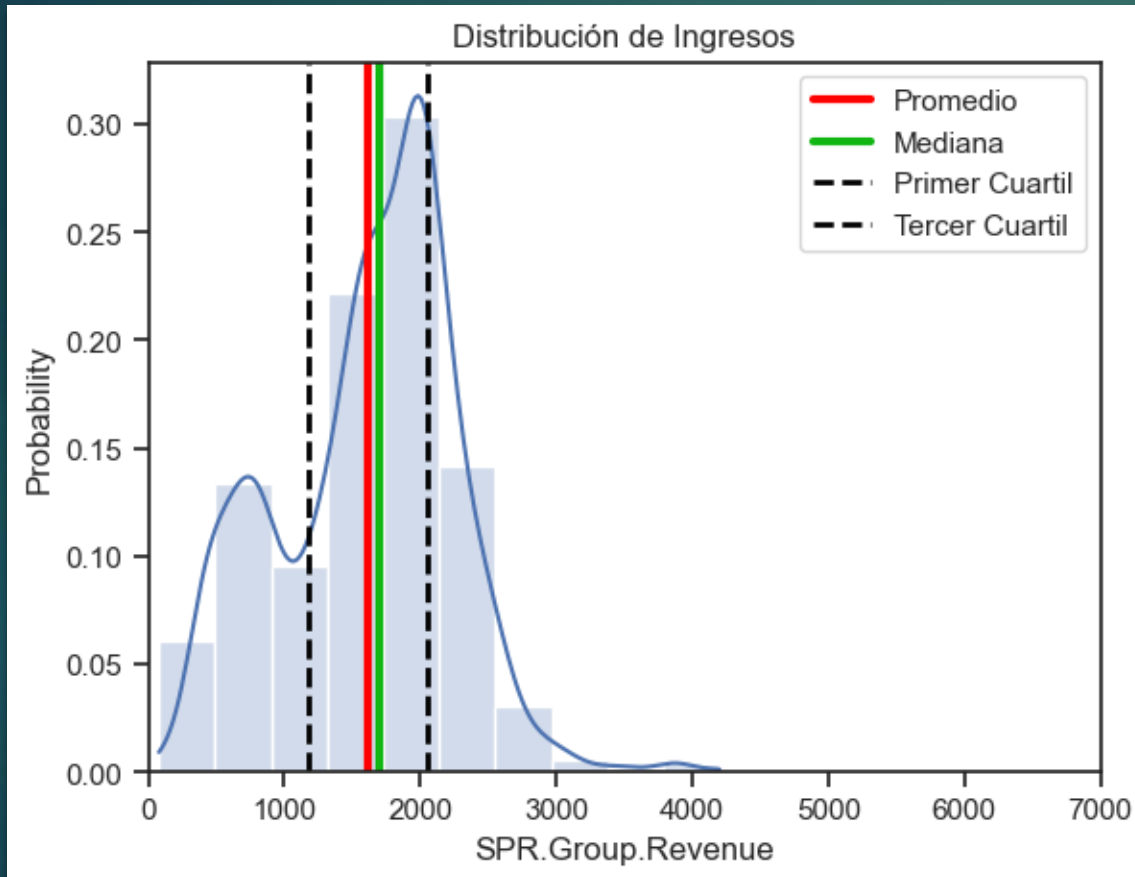
En este caso los estudiantes que asisten a la institución toman el paquete tipo A, con todo incluido en su gran mayoría, esto es entendible, porque son de estados diferentes a donde está la universidad y desean garantizar que tiene todo cubierto.



podemos observar que la mayoría de los estudiantes vienen del estado de California, y corresponde al 30% de los inscritos, seguido de Texas con un 12%, los cuales al sumarse representan el 42% de nuestros registros.



Histograma



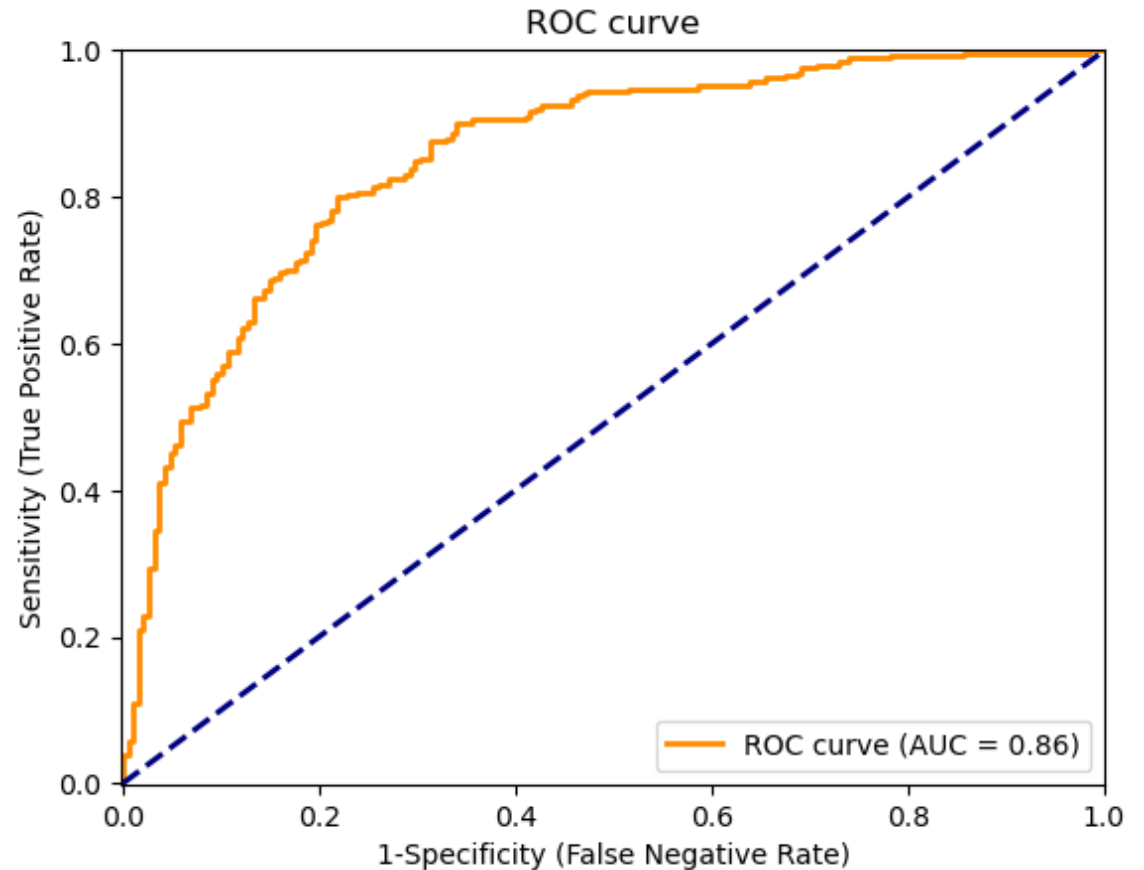
- ▶ Nos pareció importante analizar, la distribución de ingresos de las familias que envían a sus hijos al Instituto School AAA, para ver qué tan dispersos estaba los datos.
- ▶ La línea del promedio y la mediana están demasiado cerca lo que sugiere que la distribución es simétrica y no se encuentra mucho sesgo en la distribución. La gran mayoría están en el rango de 1000 a 2000, con una probabilidad de ocurrencia del 50%, lo que sugiere que tiene la capacidad de asumir los gastos de los programas, la renovación de los mismos y los planes de viaje.

La curva ROC

► Analizamos la variable target, para verificar si los alumnos de un año, fueron retenidos al año siguiente, separamos las muestras en test y train para enseñar el modelo y aplicamos luego los códigos que queremos probar.

► En este caso usaremos la función `roc_curve` para calcular las tasas de falsos positivos (FPR) y verdaderos positivos (TPR) y luego utiliza la función `plot_roc` para visualizar la curva ROC. La variable `y_pred_proba` contiene las probabilidades predichas por el modelo para la clase positiva.

► La curva ROC y el área bajo la curva (AUC) proporcionan información sobre el rendimiento del modelo en términos de sensibilidad y especificidad. Una AUC cercana a 1 indica un buen rendimiento, como es el caso del grafico



1. Verdaderos positivos (TP): 228
2. Verdaderos negativos (TN): 147
3. Falsos positivos (FP): 62
4. Falsos negativos (FN): 41

1.Accuracy (Exactitud):

1. $\text{Exactitud} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$
2. Accuracy: 0.7845 (o 78.45%)

2.Sensibilidad (Recall):

1. $\text{Sensibilidad} = \text{TP} / (\text{TP} + \text{FN})$
2. Sensibilidad: 0.7862 (o 78.62%)
3. También conocido como Recall o True Positive Rate.

3.Especificidad (Fall-out):

1. $\text{Especificidad} = \text{TN} / (\text{TN} + \text{FP})$
2. Especificidad: 0.7819 (o 78.19%)
3. Mide la capacidad del modelo para evitar falsos positivos.

4. Valor Predictivo Positivo (Precisión):

1. $\text{Precisión} = \text{TP} / (\text{TP} + \text{FP})$
2. Precisión: 0.8476 (o 84.76%)
3. Proporción de casos positivos predichos correctamente entre todos los casos positivos predichos.

5. Valor Predictivo Negativo:

1. $\text{Negativo Predictivo Valor} = \text{TN} / (\text{TN} + \text{FN})$
2. Valor Predictivo Negativo: 0.7033 (o 70.33%)
3. Proporción de casos negativos predichos correctamente entre todos los casos negativos predichos.

Estas métricas proporcionan una visión detallada del rendimiento del modelo en términos de su capacidad para predecir correctamente las clases positivas y negativas. En general, una alta sensibilidad y precisión indican un buen rendimiento.



05. Insights y Recomendaciones

- ▶ Los estudiantes del estado de California son los que más se inscriben en el Instituto SchoolAAA
- ▶ Seguido de los del estado de Texas.
- ▶ El tipo de viaje que escogen los estudiantes es el tipo A, que es el todo incluido, tienen
 - ▶ pasajes, alojamiento, tures, inscripción a los cursos, materiales y sostenimiento, para los estudiantes
- ▶ Más del 50% de las familias que envían a sus hijos o hermanos al Instituto tiene un buen nivel de
- ▶ Ingresos, lo que les permite cubrir todos los gastos de los estudiantes y realizar las renovaciones de matrícula año a año.



05. Insights y Recomendaciones

► Del % de alumnos que no continúan año a año en la institución, la curva de Roc nos mostró un 14% de falsos positivos a los que les debemos hacer un trabajo de retención importante. Se debe dar acompañamiento desde antes de que el estudiante se retire, para identificar los factores clave que le hacen tomar esta decisión, si bien el porcentaje es pequeño, la gestión social de ayudarlos a culminar es relevante en este aspecto.

► El Departamento de Marketing, debe empezar a profundizar su gestión en Texas que tiene 308 alumnos y en Washington con 147 alumnos, para dar a conocer los testimonios de todos los Servicios del instituto School AAA, programas académicos, Servicios turísticos, con el ánimo de sensibilizar a otras personas de estas zonas quienes al ver personas de su misma región hablando bien de la institución, sientan mayor grado de confianza al hacer la inversión.

► Estas zonas tienen buen volumen de estudiantes no obstante tienen también alta capacidad para matricular un mayor número.

► Cuentan con un poder adquisitivo bastante bueno lo que facilita la inscripción y el mantenimiento del programa año a año.

► California sin duda sigue siendo el estado de mayor afluencia en el instituto, así que se debe seguir con el plan de medios y el presupuesto, el programa de testimonios es muy apropiado y un plan de referidos.