**EDUVIAJES (A)**

Ud. ha sido recientemente contratado como analista de datos en Eduviajes, una empresa de turismo educativo. Durante su primera semana en la empresa, ud. se reunió con los gerentes de diferentes áreas funcionales y asistió al programa de inducción. Ahora está ansioso por comenzar su primer proyecto.

Unas horas antes, en la reunión semanal de estrategia de marketing, su jefe, destacó la urgencia de una nueva iniciativa de datos centrada en la retención de clientes. En menos de un mes, las oportunidades de renovación de contrato comenzarían para los clientes que habían realizado un viaje en Eduviajes en 2020. Durante la reunión, su jefe presentó un conjunto de datos con toda la información conocida sobre la base de clientes del año anterior (ver Anexos).

A partir de su experiencia pasada, su jefe confiaba en que se podrían construir modelos para predecir si un cliente volvería a reservar en 2021. Con tal modelo, esperaba diseñar una estrategia de marketing más matizada que se dirigiera a ciertos subconjuntos de la población de clientes

Con los conocimientos sobre árboles de decisión, ud. sabía que necesitaba ponerse a trabajar de inmediato para poder darle a su jefe un modelo de predicción preciso cuanto antes.

Anexo 1

Eduviajes (A)

Descripción de los datos

|  |  |  |
| --- | --- | --- |
| **Nombre de la Variable** | **Ejemplo** | **Descripción** |
| ID | 1 | Identificación del registro |
| CodigoPrograma | HD | Identificación de los programas ofrecidos por Eduviajes |
| GradoMinimo | 8 | Menor grado de escolaridad de un participante del programa |
| GradoMaximo | 8 | Menor grado de escolaridad de un participante del programa |
| Estado | IN | Código del estado donde se encuentra la escuela |
| NoAnual | 1 | Variable binaria 1/0 que indica (1) si la escuela se salta un año entre programas. |
| Dias | 3 | Numero de dias que el grupo estuvo en el programa |
| Transporte | A | Modo de transporte utilizado por el grupo desde la escuela hasta el programa  (A = Aire, B = Bus, T = Tren). |
| TipoPago | NA | Código del tipo de pago |
| Precio | 1174 | Precio pagado por participante para asisitir al programa |
| Seguro | 72 | Numero de participantes en el programa que pagaron el seguro de cancelación |

|  |  |  |
| --- | --- | --- |
| SeguroCancelado | 13 | El número de participantes que solicitó el seguro pero no lo terminó adquiriendo |
| PorcentajeSeguro | 0.6857 | Porcentaje de los participantes que solicitaron el seguro y lo terminó pagando |
| PagoCancelado | 15 | Número de participantes que pagaron el depósito del programa pero no asistieron. |
| Chaperones | 7 | Numero de participantes que no asistieron al programa en calidad de acompañantes y no pagaron el precio completo |
| LineaPobreza | A | Codigo del area donde se ubica la escuela con base al porcentaje estimado de residents por debajo de linea de pobreza. Va desde la A (menos del 6% por debajo de linea de pobreza) hata le D (más de 30%). E es para las areas no clasificadas. |
| Region | Otros | Region al que pertenece el estado donde se encuentra la escuela |
| SegmentoCRM | 1 | Codigo del segmento asignado a la escuela de acuerdo al Sistema de administración de la relación de clients (CRM) de Eduviajes. |
| TipoEscuela | PUBLICA | Publica o Privada |
| ReunionPadres | 1 | Variable binaria que indica si hubo una reunion de padres de familia previo al viaje (1). |
| MenorGradoEscuela | 7 | Menor grado académico impartido en la escuela |
| MayorGradoEscuela | 8 | Mayor grado académico impartido en la escuela |
| TotalAlumnosEscuela | 955 | Número total de alumnos inscritos en la escuela |

|  |  |  |
| --- | --- | --- |
| NivelIngresos | P | Similar a LineaPobreza, es un Código que indica la capacidad de pago de los padres. A es la menor, Q es la mayor. Z indica datos no clasificados. |
| PagoFacil | 0.2286 | % de participantes de pago completo que pagaron a través de cuotas asociadas al Sistema de banca en línea |
| ApoyoOficial | 0 | Variable binaria que indica si la escuela ha autorizado oficialmente la participación del programa (1) o si los alumnos van de forma independiente (0). |
| TipoPrograma | Costa Este | Código interno de agregación de los progamas individuales |
| NuevoExistente | Existente | EXISTENTE significa que el el grupo ya ha realizado programas con Eduviajes. Nuevo significa que era su primer programa. |
| ParticipantesPago | 105 | Número de participantes que realizaron el pago completo |
| TotalParticipantes | 112 | Número total de participantes incluyendo los Chaperones |
| TotalIngreso | 125735.4 | Ingreso total obtenido por la participanción del grupo en el programa |
| Reuniones | 0 | Número total de reuniones sostenidas con los padres antes de la participación en el programa |
| DiasPrimeraReunion | 93 | Número de días desde la primera reunion hasta el inicio del programa |
| DiasUltimaReunion | 103 | Número de días desde la última reunion hasta el inicio del programa |
| TipoGradoMinimo | Primaria | El tipo de grado minimo impartido en la escuela |
| TipoGradoMaximo | Bachillerato | El tipo de grado maximo impartido en la escuela |
| MesPartida | Febrero | Mes de Partida |

|  |  |  |
| --- | --- | --- |
| CodigoAgregacionPrograma | H | La primera letra del codigo del programa |
| MismoGrado | 1 | Binario indicando si todos los participantes son del mismo grado |
| ParticipantesEstudiantado | 0.06364617 | Razón del número de participantes al número total de alumnos en la escuela |
| RatioPago | 0.93650794 | Razón del número de participantes pagando precio complete al número total de participantes incluyendo chaperones |
| Tamano | L | Categoría del tamaño de la escuela establecido en quintiles (S, M, L, S-M, M-L) |
| Retenidos2020 | 1 | Variable a Predecir indicando si el grupo regreso al siguiente año |