****

**Prueba Técnica**

**Analista Científico de Datos**

**Generalidades**

Bienvenidos a la Prueba Técnica para aspirar al rol de Analista Científico de Datos, corresponde a una base de datos del sector automotriz colombiano en donde deberán, por un lado, hacer un análisis exploratorio de la información y, por otro lado, generar un modelo para predecir el **valor de la prima anual** que se cobra por cada tipo de vehículo.

Los datos recopilan todas las pólizas emitidas del seguro de automóviles para el periodo 01/10/2011 a 31/12/2013, y cada una de las 121,039 entradas representan datos agregados de 22 características o coberturas de vehículos en diferentes departamentos de Colombia.

Las características y coberturas son:

* ID: Identificador fila
* Tipo póliza: pólizas Nuevas o renovadas
* Valor Asegurado: Suma total de los valores asegurados de cada una de las coberturas de la póliza.
* Fecha Emisión: Fecha en la que se emite la póliza.
* Fecha Inicio: Fecha en que inicia la cobertura del vehículo.
* Fecha fin: Fecha en que termina la cobertura del vehículo.
* Valor prima Anual: Costo total anualizado de las coberturas otorgadas por la aseguradora de cada póliza expedida.
* Valor asegurado Vehículo: Valor asegurado del casco del vehículo de acuerdo con la Guía de valores de Fasecolda.
* Ciudad: Ciudad de circulación del vehículo.
* Departamento: Departamento de circulación del vehículo.
* Ocupación: Ocupación del asegurado.
* Edad: Edad del Asegurado.
* Características guía de valores Fasecolda:
  + MARCA
  + REF1
  + REF2
  + REF3
  + CLASE
* Modelo del vehículo: Año del modelo del vehículo de acuerdo con la tarjeta de propiedad.
* Color: Color del vehículo de acuerdo con la tarjeta de propiedad.
* Años de no reclamación: Número de años en que el asegurado no ha presentado siniestros.
* Género: Genero del asegurado.
* Deducibles: Plan de deducible escogido por el asegurado en caso de sufrir un accidente o una pérdida.

¿Qué evaluamos?

El desafío busca evaluar distintos aspectos como:

* Capacidad analítica y de exploración de datos
* Visualización de resultados
* Conocimientos de técnicas de generación de features y modelado
* Análisis de performance
* Buenas prácticas de desarrollo

**Exploración y análisis**

En la primera parte del ejercicio queremos ver cómo le dan valor a los datos mediante el uso de minería de datos, es importante que tengas en cuenta los siguientes puntos:

1. Piensa en la forma en la que vas a presentar los resultados, es posible que entre el público usuario haya personas con conocimientos técnicos, así como también personas del negocio sin ningún background técnico.
2. En esta parte del desafío, las preguntas son abiertas, pero a modo de ayuda estas son algunas de las que se podrían responder:

* ¿Cómo es el comportamiento de la Prima Anual?
* ¿Cómo se encuentra compuesta la base de datos en términos de edad y ocupación?
* ¿Existen valores faltantes o valores atípicos? ¿Qué método propones para corregir esto?

**Modelo**

Utilizando los que encontraste en la exploración de los datos, ahora tienes como objetivo crear un modelo predictivo que permita predecir el **valor de la prima anual**. El modelo por utilizar es libre; sin embargo, ten presente que valoramos la comparación entre distintos caminos.

Es importante que nos compartas tu código utilizado, así como la interpretación de los resultados.

Bienvenida toda tu creatividad.

¡Éxitos!