Es importante aclarar que todos los montos de dinero presentados están divido entre 1000 con el fin de hacer más fácil la interpretación de los resultados.

Punto 1. ($/1000)

Interfaz de usuario gráfica, Tabla

El contenido generado por IA puede ser incorrecto.

Punto 2. ($/1000)\*\*\*NO TENGO FACTOR DE GASTOS NI UTILIDADES PARA CALCULAR LA PRIMA DE RIESGO

Texto

El contenido generado por IA puede ser incorrecto.

Punto 3.



Punto 4.

* 1. El fitted averge al ser un promedio de las predicciones del modelo no captura el efecto que si captura el Model Prediction at base por medio de las relatividades (exp(B1))
  2. .1 a) Es posible que exista datos atípicos dentro de la muestra y que deben ser limpiados, también es posible que en Arauca mi frecuencia es explicada por otro tipo de variable no incluidas dentro del modelo y que me generan ruido en mi predicted. Además, la muestra de este departamento es muy pequeña, es posible que el modelo no logré captar el efecto correcto gracias a esta característica muestral.

b) Podría cobrar una prima más elevada de la correcta para los residentes de Arauca, convirtiendo a la empresa en una opción poco competitiva vs otras aseguradoras de la zona

1.3.1 a) Al igual que en el caso de Arauca, el modelo puede estar detectando casos atípicos que empuja la frecuencia hacia arriba, sin embargo, es posible que estos casos deban ser excluidos porque representan un outliers para mi modelo. Además, podemos estar excluyendo otra variable diferente al departamento/ciudad que explica la alta frecuencia en Tuluá.

b) Es primordial revisar la calidad de la data del modelo y evaluar otras interacciones diferentes al departamento que me ayuden entender le porqué de la frecuencia alta en el predicted. Antes de realizar una campaña agresiva sobre este municipio, es necesario realizar un análisis de siniestralidad dentro del mismo, si Tuluá mantiene una baja siniestralidad, no es recomendable subir los precios de forma agresiva ya que perderemos competitividad y mercado dentro de la zona.

1.3 Hacer una análisis univariado y bivariado (en el caso de que dentro del modelo se desee incluir mas variables), volviendo al caso de Arauca o Amazonas, al tener una muestra tan pequeña, puede que mi modelo no capture la información correcta, y si es bivariado, es posible que mi muestra se centre aún mas en grupo de población especifico (e.g. personas entre 40 y 60 años). Por otro lado, se debe validar si el monto pagado ya está completamente desarrollado, si no es así se debería incluir factores de IBNR por medio offsets.

2. La baja tasa observada en los vehículos de alta cilindrada puede ser explicada por medio de la baja muestra de este tipo de población, en comparación con los otros dos segmentos, esta categoría tiene muy pocos asegurados. Por otro lado, los asegurados con vehículos de alta cilindrada se concentran en la edades mayores a 43 años y credit score mayores a 700, es posible que al ser personas mayores tengan hábitos de conducción mas tranquilos y prudentes, el mismo efecto se puede ver a través de sus CS, buenos hábitos financieros pueden significar buenos hábitos de conducción.

3. En primer lugar buscaría limpiar la data lo mejor posible, definir/excluir/incluir outliers, siniestros en ceros o negativos, siniestros fueras de vigencias o exposición. Intentaría incluir dentro del modelo diferentes variables que puedan explicar mi frecuencia y severidad. Adicionalmente, haría un análisis univariado y bivariado con el fin de determinar la calidad de la muestra. También es importante asegurarse que los pagos estén completamente desarrollados (se puede ver gráficamente) si no es así es recomendable incluir un factor de IBNR. Finalmente, intentaría realizar fowards que me permitan encontrar el mejor modelo posible, esto se puede determinar a través de la desvianza y criterios AIC y BIC.

Código SAS:

**PROC** **IMPORT** DATAFILE="\\NASCO\Colombia\PIPELINE\CARPETAS PERSONALES\Sergio Motta\Prueba Tecnica.xlsx"

OUT=Prueba

dbms=excelcs

replace;

SHEET="BASE\_TOTAL";

server\_name=NTSASPCFILES;

port\_number=**9621**;

**RUN**;

**PROC** **SQL**;

CREATE TABLE PRUEBA\_2 AS

SELECT \*, COUNT(id\_poliza) as dup, MIN(FECHA\_INICIO) AS MIN\_F FORMAT DDMMYY10.,

MAX(FECHA\_FIN) AS MAX\_F FORMAT DDMMYY10.

FROM PRUEBA

GROUP BY ID\_POLIZA;

**QUIT**;

/\*EXTRAEMOS DUPLICADOS\*/

**DATA** PRUEBA\_DUP;

SET PRUEBA\_2;

WHERE DUP EQ **2** AND FECHA\_INICIO EQ MIN\_F AND FECHA\_FIN EQ MAX\_F;

**RUN**;

/\*ELIMINAMOS DUP\*/

**PROC** **SORT**

DATA= PRUEBA\_DUP NODUPKEY out=PRUEBA\_DUP;

by \_ALL\_

;

**run**;

/\*POLIZA = 33165 SON DOS VEHÍCULOS DISTINTOS?\*/

**DATA** PRUEBA\_2;

SET PRUEBA\_2;

IF DUP EQ **2** AND FECHA\_INICIO EQ MIN\_F AND FECHA\_FIN EQ MAX\_F THEN DELETE;

**RUN**;

**DATA** PRUEBA\_2;

SET PRUEBA\_2

PRUEBA\_DUP;

**RUN**;

**DATA** PRUEBA\_3;

SET PRUEBA\_2;

DAYS\_EYEAR = INPUT(PUT(INPUT ('3112'||PUT(**2022**, **4.**), DDMMYY8.), JULDAY.), **3.**);

EXPY\_DAYS = INTCK('DAY', FECHA\_INICIO, FECHA\_FIN);

EXPY = EXPY\_DAYS / DAYS\_EYEAR;

ANNUAL\_PREMIUM\_RC = PRIMA\_RCD\_DEV/EXPY;

ANNUAL\_PREMIUM\_PTD = PRIMA\_PTD\_DEV/EXPY;

**RUN**;

/\*ASUMO VALOR ASEG DE RC = VALOR ASEG PTD (NO ESPECIFIFCA RC, 4000 MILL?)\*/

**PROC** **SQL**;

CREATE TABLE REV AS

SELECT DISTINCT SUM(SUMA\_ASEG\_VEHI)/**1000** AS VALOR\_ASEGURDO\_TOTAL, SUM(IMP\_INCURRIDO\_\_RCD)/**1000** AS

MONTO\_RC,

SUM(IMP\_INCURRIDO\_PTD)/**1000** AS MONTO\_PTD, SUM(EXPY) AS EXPO\_TOTAL,

SUM(PRIMA\_PTD\_DEV)/**1000** AS PRIMA\_PTD, SUM(PRIMA\_RCD\_DEV)/**1000** AS PRIMA\_RC,

SUM(NUM\_SINI\_PTD) AS NUM\_PTD, SUM(NUM\_SINI\_RCD) AS NUM\_RCD, SUM(ANNUAL\_PREMIUM\_RC)/**1000** AS

ANNUAL\_PREMIUM\_RC, SUM(ANNUAL\_PREMIUM\_PTD)/**1000** AS ANNUAL\_PREMIUM\_PTD

FROM PRUEBA\_3;

**QUIT**;

/\*NO TENGO FACTOR DE GASTOS NI UTILIDADES PARA CALCULAR LA PRIMA DE RIESGO\*/

**DATA** REV2(KEEP= FRECUENCIA\_RC FRECUENCIA\_PTD COSTO\_MEDIO\_RC COSTO\_MEDIO\_PTD

ANNUAL\_PREMIUM\_RC ANNUAL\_PREMIUM\_PTD);

SET REV;

FRECUENCIA\_RC = NUM\_RCD/EXPO\_TOTAL;

FRECUENCIA\_PTD = NUM\_PTD/EXPO\_TOTAL;

COSTO\_MEDIO\_RC = MONTO\_RC/NUM\_RCD;

COSTO\_MEDIO\_PTD = MONTO\_PTD/NUM\_PTD;

**RUN**;

**DATA** CILINDRADA;

SET PRUEBA\_3;

LENGTH CILIN $ **5**;

IF CILINDRAJE <= **1500** THEN CILIN = "BAJA";

ELSE IF CILINDRAJE > **1500** AND CILINDRAJE <= **5000** THEN CILIN = "MEDIA";

ELSE CILIN = "ALTA";

**RUN**;

**DATA** alta;

SET CILINDRADA;

where CILIN = "ALTA";

**run**;

**PROC** **FREQ** DATA=alta;

TABLES EDAD GENERO MARCA SCORE\_FINAN;

**RUN**;

**PROC** **SQL**;

CREATE TABLE CILINDRADA\_2 AS

SELECT DISTINCT CILIN AS CILINDRAJE, SUM(EXPY) AS EXPO\_TOTAL,

SUM(NUM\_SINI\_PTD) AS NUM\_PTD, SUM(NUM\_SINI\_RCD) AS NUM\_RCD

FROM CILINDRADA

GROUP BY CILIN;

**QUIT**;

**DATA** CILINDRADA\_2 (KEEP= CILINDRAJE TASA\_RC TASA\_PTD);

SET CILINDRADA\_2;

TASA\_RC = NUM\_RCD/EXPO\_TOTAL;

TASA\_PTD = NUM\_PTD/EXPO\_TOTAL;

**RUN**;