



**Universidad Autónoma del Estado de México**

**Facultad de Economía**

**Licenciatura en Actuaría**

## **Cómputo Científico**

### **Reporte de la base de datos "dataCar"**

Alumno: Diego Isau Barranco Herrera

No.Cuenta: 1810088

Maestra: Martha Paola Hernández Soto

**Octavo semestre**

**21-02-25**

## Reporte de la base de datos dataCar

### 1. Sobre la base “dataCar”

La base de datos contiene información sobre pólizas de seguros de autos del 2004 y 2005

Hay 67886 pólizas, con 11 variables:

|           |  |
|-----------|--|
| veh_value | Valor del vehículo, en \$10,000                                      |
| exposure  | Valores 0 o 1, indica el tiempo que permaneció asegurado el vehículo |
| clm       | Ocurrencia del reclamo (si=1, no=0)                                  |
| numclaims | Número de reclamos   |
| claimcst0 | Monto del reclamo  |
| veh_body  | Tipo de vehículo   |
| veh_age   | Edad del vehículo  |
| gender    | Genero del asegurado (F=mujer, M=hombre)                             |
| area      | Variable categórica con valores: A, B, C, D, E, F                    |
| age_cat   | Valores de 0 a 6, categoría para las edades de los asegurados        |
| X_OBSTAT_ | Variable categórica con valores: 01101, 0 0 0                        |

No existen valores faltantes en la base, por lo que no necesitamos realizar ninguna transformación de normalización

### 2. Explorando los datos

Una vez que sabemos el contexto e información de la base dataCar, es momento de analizar y realizar cuestionamientos de interés sobre la información recabada.

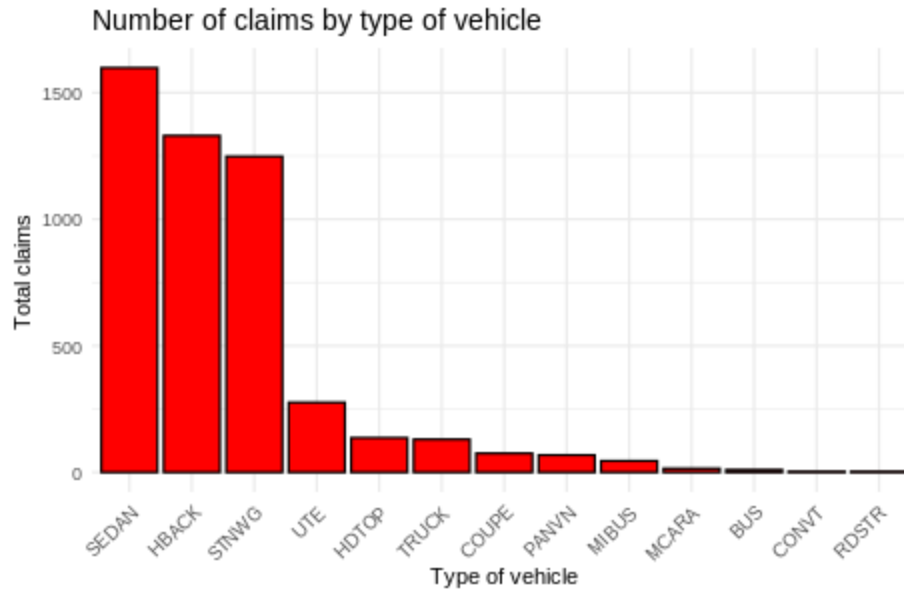
#### **¿Qué porcentaje de pólizas incurren en al menos una reclamación?**

Existen 4624 pólizas con al menos una reclamación, haciendo una comparación con el total, el 6.814431% incurre en al menos una reclamación

#### **¿Cuáles son los 5 vehículos con mayor número de reclamaciones?**

Necesitamos agrupar por tipo de vehículo, y después realizar la suma de la variable `numclaims`, dándonos como top 5: SEDAN (1598), HBACK (1330), STNWX (1248), UTE (276), HDTOP (136)

Podemos visualizarlo en la siguiente gráfica

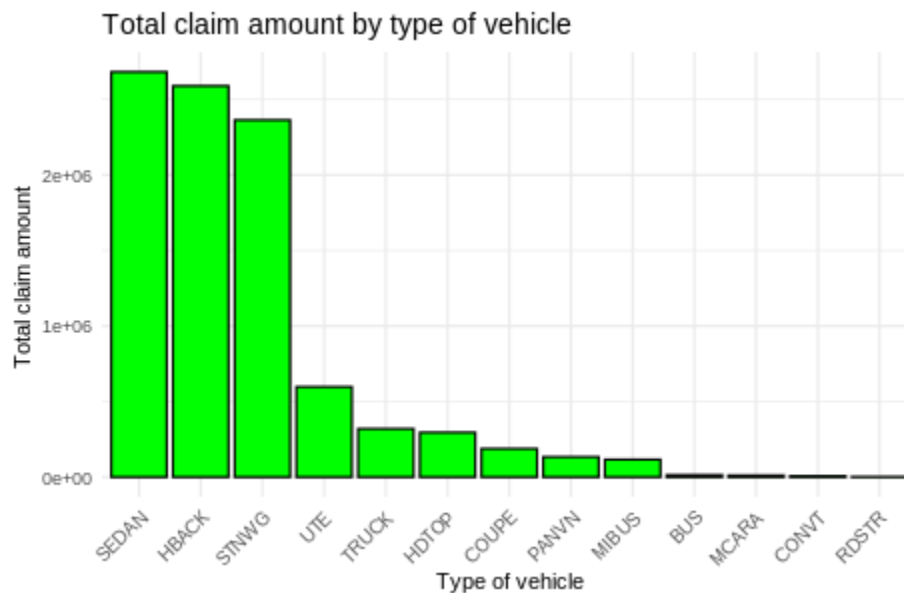


### ¿Cuáles son los 10 vehículos con el mayor monto de reclamaciones?

Realizamos la misma operación que la anterior, solo que con la variable `claimcst0`

Dentro de este top, los primeros 4 tipos de vehículos son los mismos que tienen el mayor número de reclamaciones, sin embargo, el tipo “TRUCK” tiene más monto en la posición 5, por lo que podemos interpretar que el monto de reclamación es más alto, aunque presenten solamente 130 reclamaciones

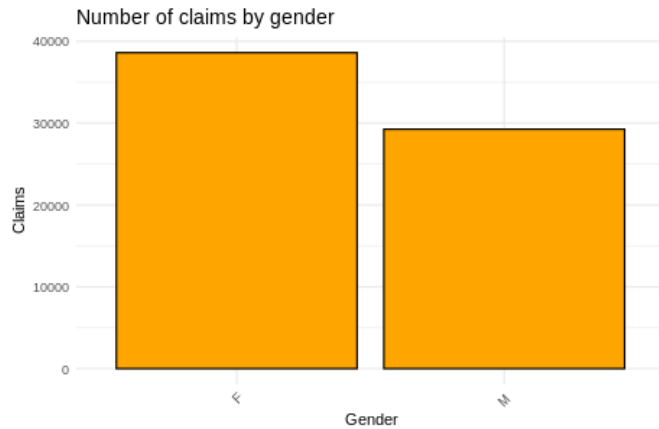
Se observa que los primeros 3 vehículos son similares a el monto de reclamación



### ¿Cuántas reclamaciones presentaron las mujeres y cuántas los hombres?

Agrupamos ahora por la variable `genre`, contando las observaciones de cada categoría

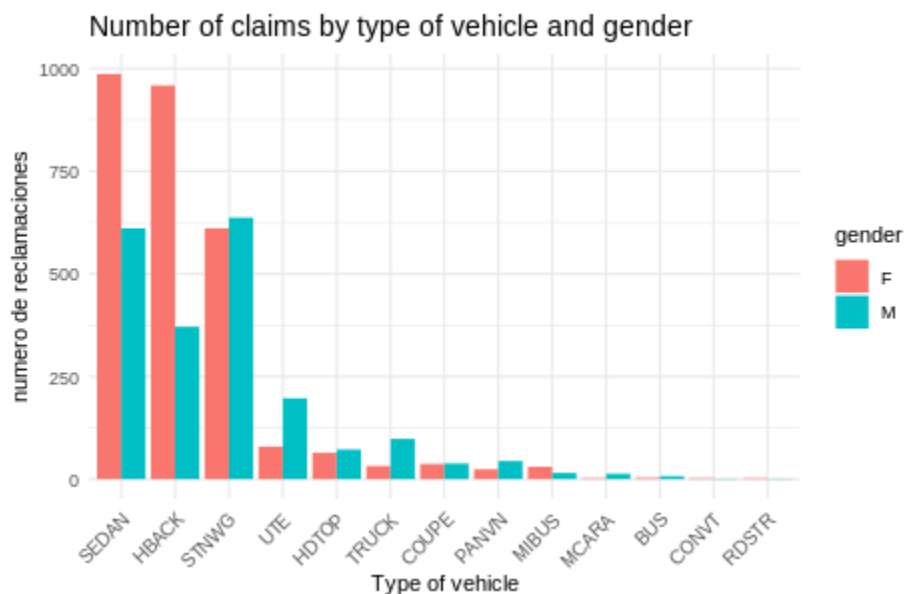
Las mujeres realizaron más reclamos que los hombres. Contando con 38603 para las mujeres y 29253 reclamos para los hombres.



### ¿Cuál es el monto de reclamaciones por tipo de vehículo y por género?

Agruparemos por las variables `veh\_body` y `gender`. Sumamos el número de reclamos

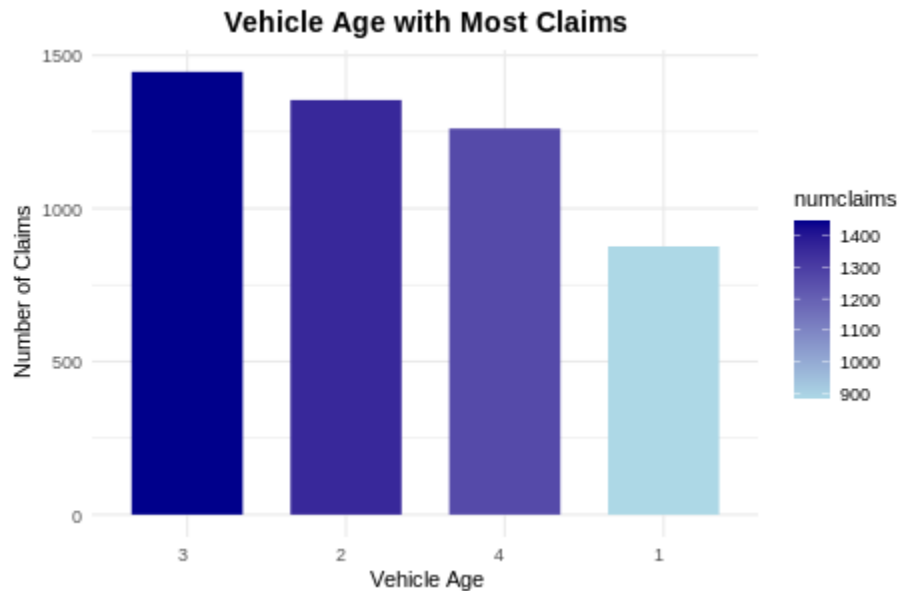
Predominan las reclamaciones de las mujeres en los tipos SEDAN y HBACK, mientras que, en el resto de los vehículos, los hombres presentan la mayoría de las reclamaciones



### ¿Cuál es la edad del vehículo donde se presentan más reclamos?

Esta vez utilizamos la variable `veh\_age` y sumamos el número de reclamos.

Cuando los vehículos tienen alrededor de 3 años, es cuando se presentaron más reclamos. Vehículos nuevos, presentaron una cantidad de reclamos mucho menor



### ¿Cuál es el promedio de edad de los vehículos SEDAN?

Filtrando las observaciones con ``veh_body` == "SEDAN"`, y utilizando la función `mean()`, obtenemos una edad promedio de 2.6868 años

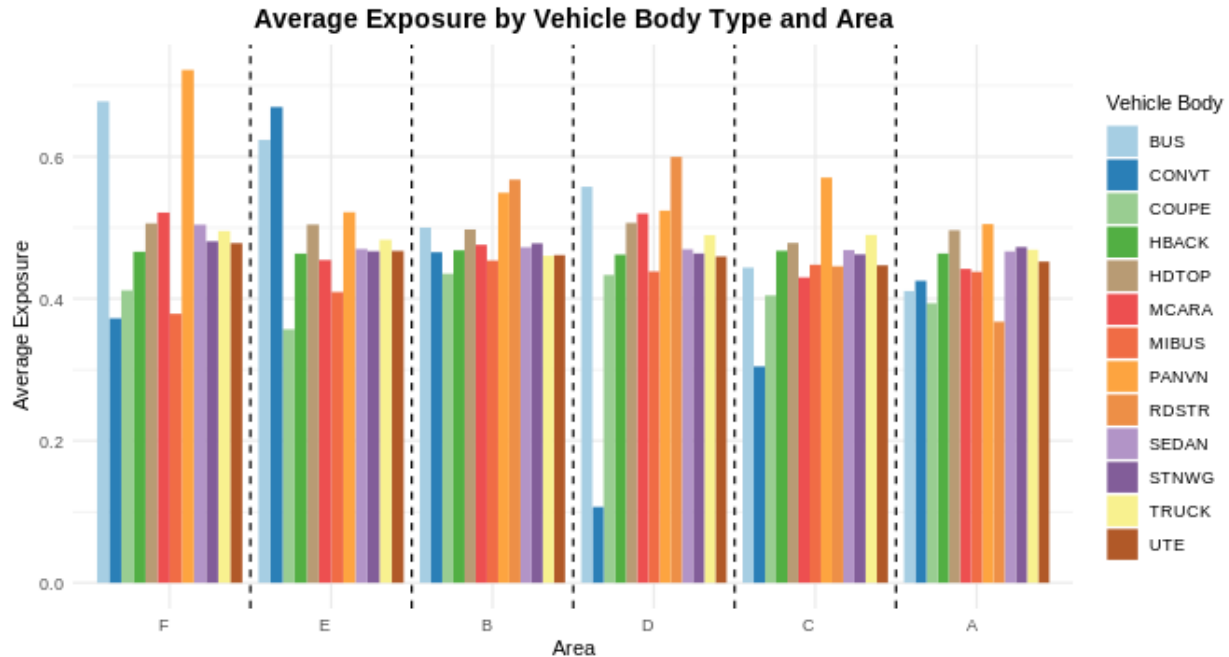
### ¿Cuál es el área con mayor promedio de duración del seguro, clasificado en tipo de vehículo?

La variable ``exposure`` mide cuánto tiempo el vehículo permaneció asegurado durante un año, es decir si es 1, estuvo asegurado un año, 0.5, medio año.

Agrupamos las variables ``veh_body`` y ``area``. Realizamos un promedio de la variable ``exposure``

Tenemos los siguientes resultados:

- Los promedios son similares entre las seis áreas, sin embargo, existen desviaciones grandes en las áreas F, E
- La duración promedio del seguro más corta entre las áreas y vehículos se da en D, con el tipo de vehículo CONV. La duración promedio más larga está en el área F con el tipo PANVN
- En general, existe un promedio de duración del seguro de medio año



### ¿Dónde se encuentran los vehículos más caros y baratos?

Las variables `veh\_value` y `area` se agruparán para responder a la pregunta. Como `veh\_value` puede ser cualquier número entre 0 y 34.56, creamos intervalos para realizar el conteo.

- En la zona E y F se encuentran solamente vehículos con un factor entre 0 y 12. La zona con más vehículos en ese intervalo es la C
- Existe un vehículo con un valor entre 24 y 35, se encuentra en la zona A
- Hay muy pocos vehículos con un valor entre 13 y 23, sólo 29, dispersos en las zonas A (18), B (1), C (8), D (2)
- Podemos decir que esta aseguradora prefiere un perfil de autos baratos. El área con más vehículos de este tipo es la C

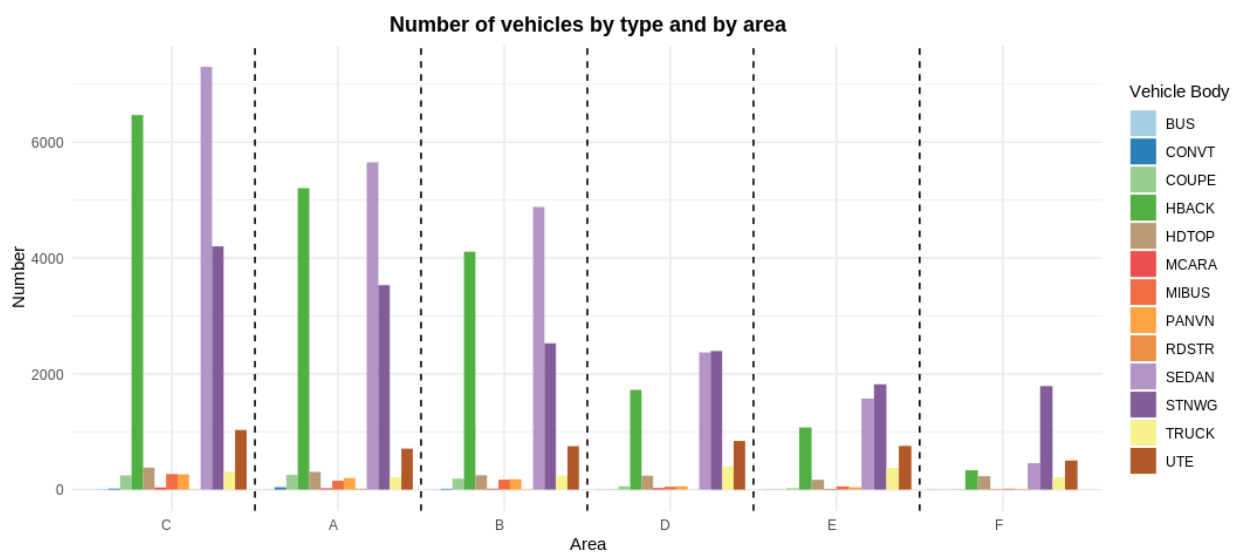


### ¿Cuántos vehículos hay por área?

Usamos las variables `area` y `veh\_body` para agruparlas y contamos las observaciones con `n()`

Nuevamente, el SEDAN predomina en 3 áreas: C, A, B

En el área F no existen muchos vehículos asegurados



### 3. Conclusiones (insights)

El perfil de esta cartera se centra en vehículos con un valor de hasta \$120, 000. Aunque existen autos con un valor mayor a este, son pocos. Aunado a este hecho, el SEDAN, con un valor entre \$2900 y \$226,900 fue el vehículo más frecuente en las pólizas, pero también el que más reclamos cometió. Se encuentra entre la edad más frecuente (2 y 3), resultando un promedio de edad de 2.68 años

En general, las mujeres realizaron más reclamos que los hombres, sin embargo, específicamente sólo en los autos SEDAN y HBACK. En los demás tipos de auto, los hombres reclamaron más

Considerando la duración del seguro, no hay grandes diferencias entre las áreas; en promedio existe una duración de medio año para cualquier tipo de auto

GitHub:

[https://github.com/Diego195200/CC\\_course\\_2025A/blob/main/R%20code/Reporte\\_dataCar.R](https://github.com/Diego195200/CC_course_2025A/blob/main/R%20code/Reporte_dataCar.R)