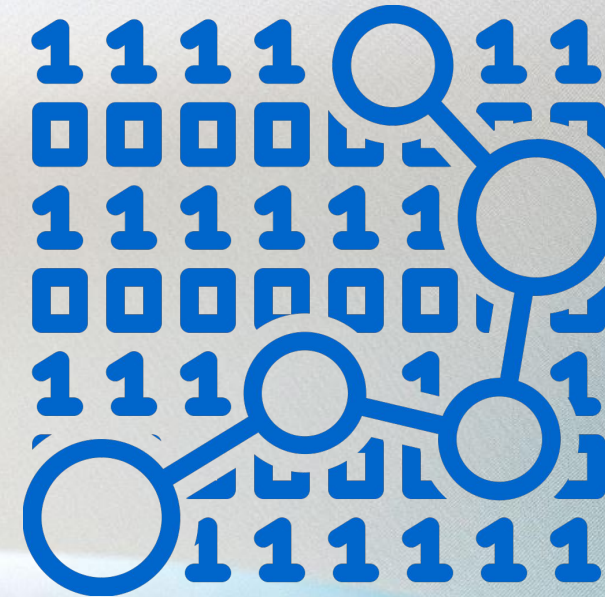


# ***CODERHOUSE***

## **Data Science - Comisión 49140 Daniel Fanara Trabajo Final**



***Análisis de Robos de vehículos en  
Argentina***



# Análisis de Robos de vehículos en Argentina

## Se pretende analizar

- Modelos más Robados
- Zonas de mayor frecuencia
- Frecuencia por fechas de tramite
- Incidencia en la frecuencia del año de la unidad siniestrada
- Incidencia en la frecuencia por algún patrón en la edad de los titulares

Contexto Empresarial



Problema Empresarial



Contexto Analítico







## Contexto empresarial

En el complejo y exhaustivo panorama económico que está atravesando la argentina, con el constante cambio de costos y precios todos los días, se necesita realizar un análisis y reformulación constante de los costos de las coberturas de seguros , de acuerdo con aquellas que generan gran volumen de erogación monetaria en pagos por siniestros. Siendo de esta forma, reformulada la base de aplicación de las nuevas tarifas , en función a la frecuencia siniestral de cada tipo de unidad suscripta



En este contexto se pretende comprender, en el caso de los robos de unidad , algún patrón o parámetro que permita identificar vehículos con una propensión frecuente a la siniestralidad para dicha cobertura que genera grandes montos de pagos por indemnizaciones





## Problema empresarial

Dentro de la revisión se determina que hubo un gran incremento de la cantidad de robos totales en el territorio argentino, el amparo de tal cobertura, al ser abonada ante el siniestro, genera automáticamente una desproporción entre el costo ingresado de la prima y el pago del siniestro. Se requiere una reestructuración de los costos, pero para ello se debe contar con datos acerca de si existe una relación directa entre la siniestralidad ( cantidad de casos registrados) y el tipo de unidad que se suscribe.



¿Existe algún tipo de inclinación en cuanto a la frecuencia de los casos de acuerdo a las fechas registradas?

¿Hay alguna marca que sea de incidencia en cuanto a las unidades robadas?

¿El año de la unidad es de interés para determinar una frecuencia estimada?





## Contexto analítico

Dentro de la revisión se determina que hubo un gran incremento de la cantidad de robos totales en el territorio argentino, el amparo de tal cobertura, al ser abonada ante el siniestro, genera automáticamente una desproporción entre el costo ingresado de la prima y el pago del siniestro. Se requiere una reestructuración de los costos, pero para ello se debe contar con datos acerca de si existe una relación directa entre la siniestralidad ( cantidad de casos registrados) y el tipo de unidad que se suscribe.





## Hipótesis de aproximación

***En el presente trabajo se dispone consensuar si, existe un tipo de unidad del ramo automotor, más frecuentada o pasible de sufrir un siniestro por robo total, dentro del total de población de unidades robadas. Determinando según el muestreo obtenido, un porcentaje de propensión al riesgo***

***Con la información recabada se pretende obtener una estimación de que esa propensión se traduzca en una proyección de gastos por siniestros abonados por unidades que se contratasen.***

***Es decir se intentará obtener una estimación de propensión al riesgo que permita visualizar si un tipo de unidad dada su siniestralidad, sea pasible de reevaluar su costo de seguro según qué tan frecuente tenga siniestros de robo total.***

### **Potencial Usuario de la info:**

Una empresa de seguros

Hipótesis 0: Existe una relacion entre la cantidad de robos y el año de la unidad



# Metadata Utilizada

**Fuente:** DNRPA, datos oficiales robo de unidades 202310

[https://datos.gob.ar/dataset/justicia-robos-recuperos-autos/archivo/justicia\\_707d3749-fbdc-4b7c-a015-60e7c2912ecc](https://datos.gob.ar/dataset/justicia-robos-recuperos-autos/archivo/justicia_707d3749-fbdc-4b7c-a015-60e7c2912ecc)

## Variables contenidas en el Dataset

- tramite\_tipo: Tipo de trámite (**object**)
- tramite\_fecha: Fecha del trámite (**datetime64[ns]**)
- dia: Día del mes del trámite (**int64**)
- fecha\_inscripcion\_inicial: Fecha de inscripción inicial (**datetime64[ns]**)
- q\_casos: Cantidad de casos (**int64**)
- registro\_seccional\_codigo: Código del registro seccional (**int64**)
- registro\_seccional\_descripcion: Descripción del registro seccional (**object**)
- registro\_seccional\_provincia: Provincia del registro seccional (**object**)
- automotor\_origen: Origen del automotor (**object**)
- automotor\_anio\_modelo: Año modelo del automotor (**float64**)
- EDAD: Edad del titular del automotor (**int64**)
- automotor\_tipo\_codigo: Código del tipo de automotor (**float64**)
- automotor\_tipo\_descripcion: Descripción del tipo de automotor (**object**)
- automotor\_modelo\_descripcion: Descripción del modelo del automotor (**object**)

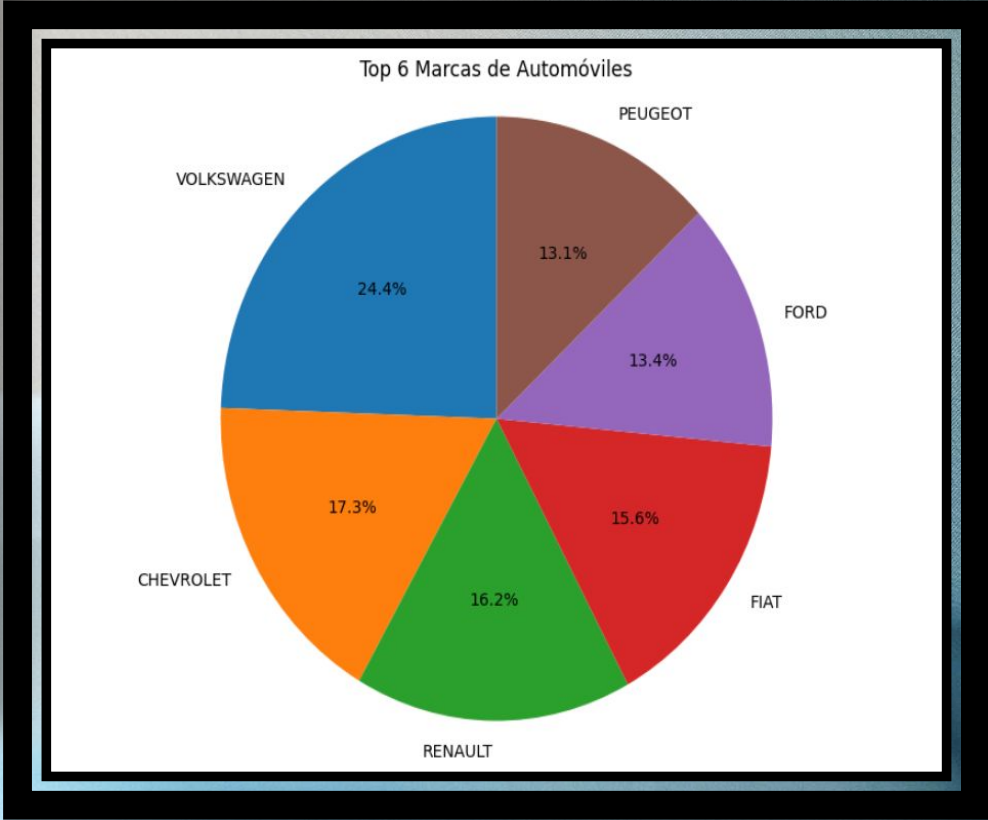
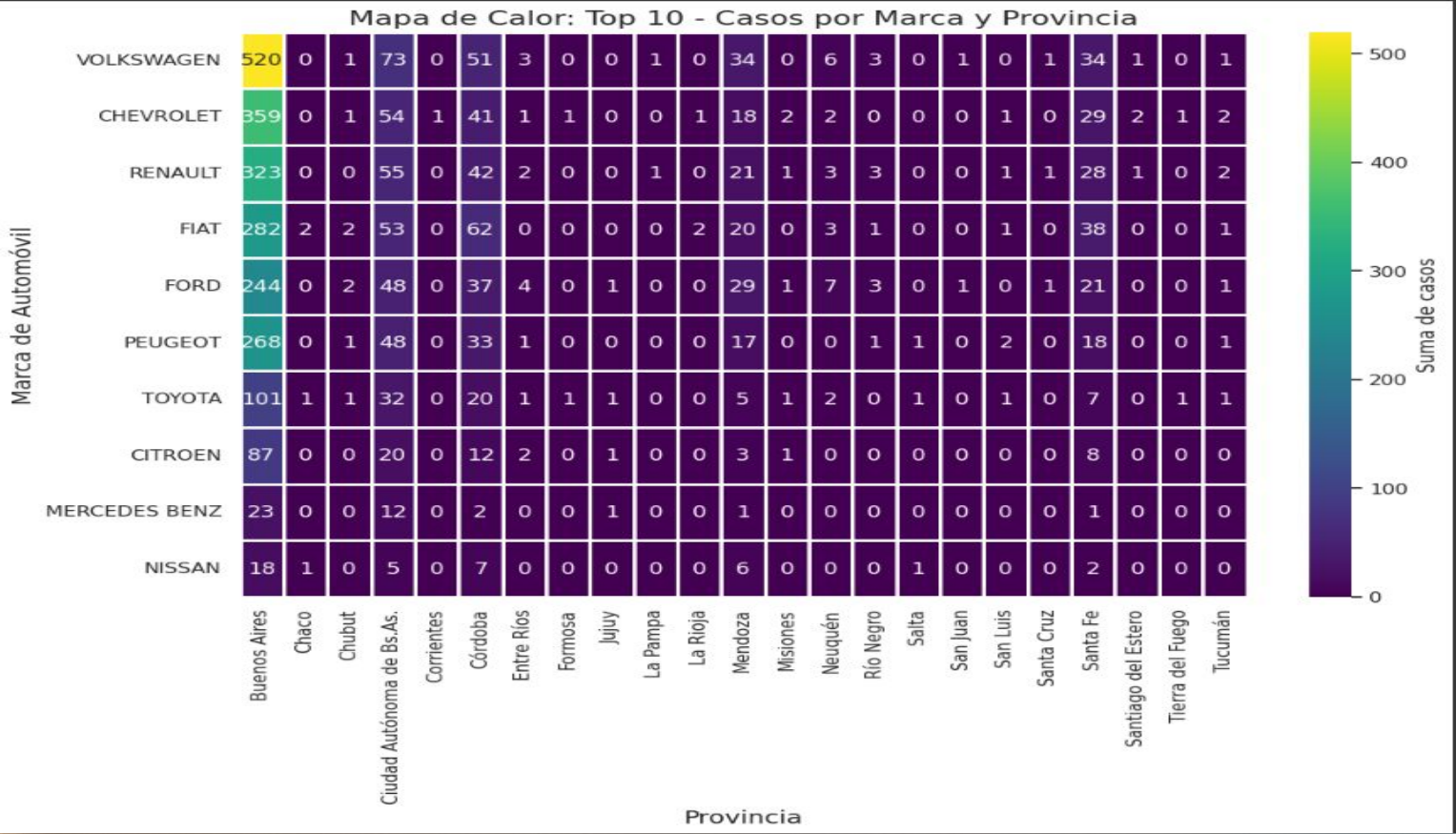
- automotor\_uso\_codigo: Código del uso del automotor (**int64**)
- automotor\_uso\_descripcion: Descripción del uso del automotor (**object**)
- titular\_tipo\_persona: Tipo de persona del titular (**object**)
- titular\_domicilio\_localidad: Localidad del domicilio del titular (**object**)
- titular\_domicilio\_provincia: Provincia del domicilio del titular (**object**)
- titular\_genero: Género del titular (**object**)
- titular\_anio\_nacimiento: Año de nacimiento del titular (**int64**)
- titular\_pais\_nacimiento: País de nacimiento del titular (**object**)
- titular\_porcentaje\_titularidad: Porcentaje de titularidad del automotor (**int64**)
- titular\_domicilio\_provincia\_id: ID de la provincia del domicilio del titular (**int64**)
- titular\_pais\_nacimiento\_id: ID del país de nacimiento del titular (**object**)
- automotor\_marca\_codigo: Código de la marca del automotor (**object**)
- automotor\_marca\_descripcion: Descripción de la marca del automotor (**object**)
- automotor\_modelo\_codigo: Código del modelo del automotor (**object**)



# Insights Obtenidos

¿Hay alguna marca que sea de incidencia en cuanto a las unidades robadas?

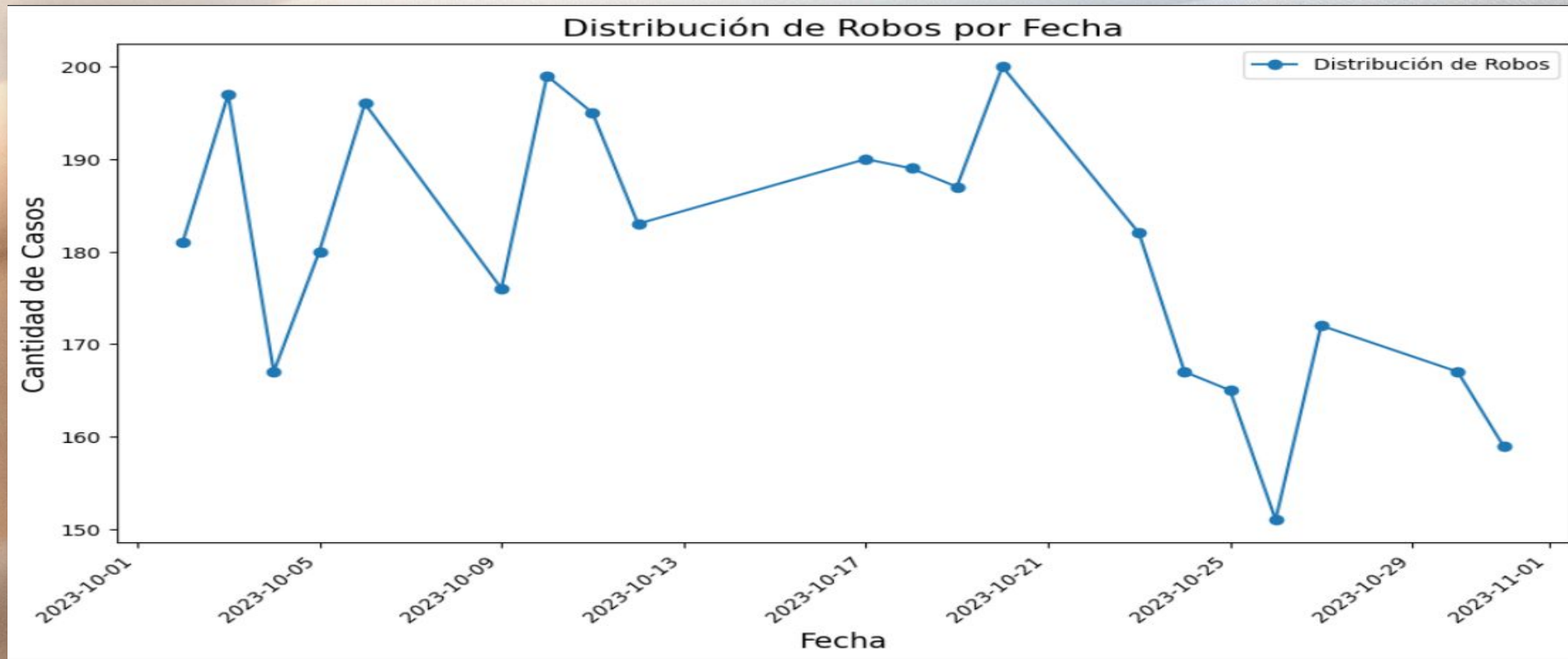
En un primer acercamiento, se puede apreciar una tendencia marcada , entre todas las zonas , por el robo de una marca en particular que se distancia por más del 7% del resto y sobre todo una tendencia de mayor frecuencia en una provincia en específico.





# Insights Obtenidos

**¿Existe algún tipo de inclinación en cuanto a la frecuencia de los casos de acuerdo con las fechas registradas?**

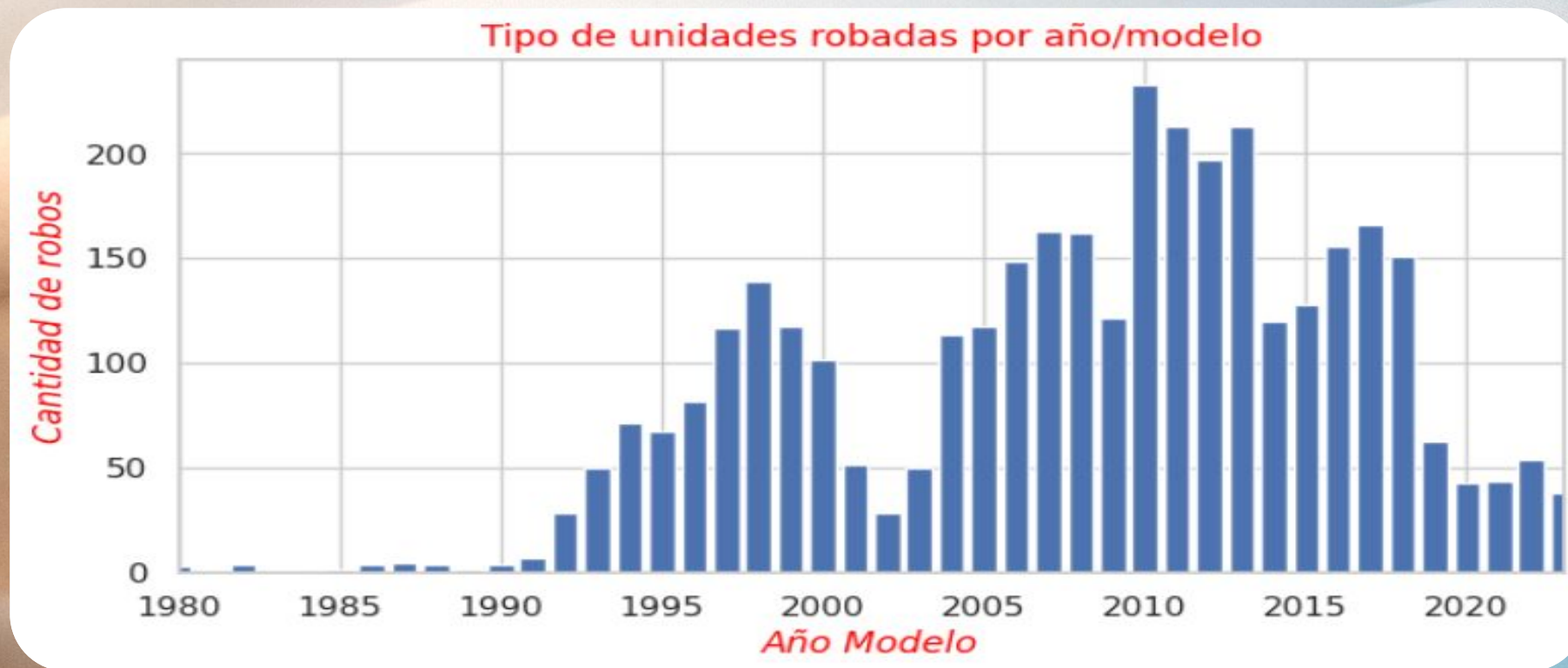


Se pudo obtener que, en cuanto a frecuencia de casos, hay una marcada tendencia en ciertos días de la semana donde se elevan abruptamente la cantidad de casos analizados. Luego del último pico, en la semana final, comienzan a decaer los casos-



# Insights Obtenidos

¿El año de la unidad es de interés para determinar una frecuencia estimada?

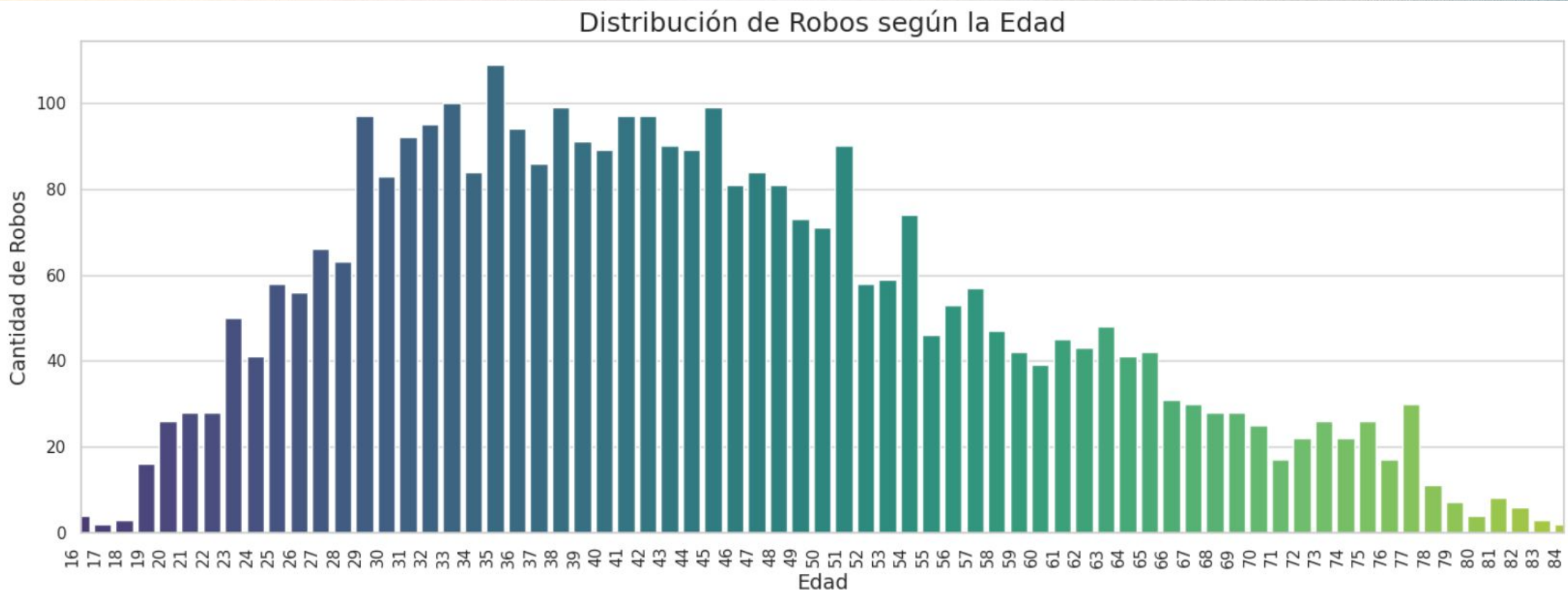


Se advierte que la mayor cantidad de robos se dan en unidades modelos 2010-2013, con una tendencia un poco menor hacia las unidades modelo 1996-2000 pero con menor magnitud en cuanto a cantidades con respecto a las de 2010-2023 Siendo Volkswagen y Chevrolet las unidades más robadas ( sin determinar por año) las mismas.



# Insights Obtenidos

No existe una tendencia hacia la distribución normalizada dentro del rango de edad donde se presentan la mayor cantidad de robos, siendo el promedio de edad de quienes sufren el siniestro de entre 43 y 44 años-





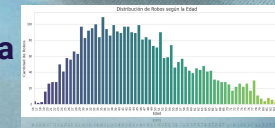
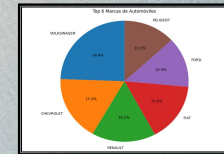
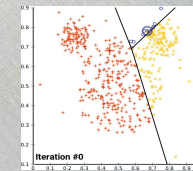
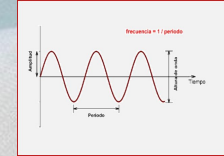
# Análisis de la exploración preliminar de datos

- Se verifica q la frecuencia posee picos de mayor cantidad de sucesos registrados la primer semana y media de mes , se mantiene constante a continuación, con su pico máximo el día 20
- Existen dos grandes grupos de unidades robada según el año de la misma, grupo1 = 1994-2000 y grupo2= 2006 a 2013. Es decir, hay dos grupos de tendencia con mayor frecuencia de robo, los más buscados a la hora de cometer el delito.
- Dentro del Top 6 de marcas más robadas, se desprende con un % por encima del resto de las del top Volkswagen , como la marca más robada.
- A su vez , ésta marca más robada , tiene su epicentro delictivo en la ciudad de Buenos Aires , registrando el mayor volumen por marca y ciudad de toda la muestra. Siguiendo la coherencia de mayor volumen de acuerdo a la densidad poblacional.
- El promedio de edad de quienes sufren robos es de 43/44 años, pero la distribución de robos por edad , no llega a tener una distribución normal, sino que tiene una inclinación un poco más para la izquierda la campana.



# Conclusiones del trabajo

1. **Frecuencia de robos a lo largo del mes:** Se observa que la frecuencia de robos presenta picos durante la primera semana y media del mes, manteniéndose constante posteriormente, con un pico máximo el día 20. Esto sugiere patrones temporales que podrían ser útiles para la asignación de recursos en la prevención del delito.
2. **Agrupación de unidades robadas por año:** Se identifican dos grandes grupos de unidades robadas según el año de fabricación, destacando los años 1994-2000 y 2006-2013. Esto indica la existencia de tendencias específicas en la selección de vehículos para cometer robos.
3. **Marcas más robadas:** Volkswagen sobresale como la marca más robada, con un porcentaje significativamente mayor que otras marcas en el top 6. Esto puede ser útil para la aplicación de medidas de rediseño de políticas de suscripción o bien de reevaluación de primas específicas por parte del equipo de actuarios para vehículos de esta marca, dada la tendencia a sufrir delitos de gran impacto siniestral.
4. **Epicentro delictivo por marca y ciudad:** La ciudad de Buenos Aires registra el mayor volumen de robos para la marca Volkswagen, lo que puede estar relacionado con la densidad poblacional y la disponibilidad de estos vehículos en el área urbana.
5. **Perfil de las víctimas de robos:** El promedio de edad de las personas que sufren robos es de aproximadamente 43-44 años, pero la distribución de los robos por edad no sigue una distribución normal, lo que sugiere un sesgo hacia la izquierda en la distribución.





# Conclusiones del Modelo

## Conclusión del uso del modelo aplicado

Dado el rango de valores de la variable objetivo y el MSE de 65.00665, se observa que las predicciones del modelo tienen una precisión aceptable en relación con la variabilidad de los datos. Aunque el MSE no es muy bajo, lo cual indicaría una precisión alta, tampoco es excepcionalmente alto, lo que sugiere que las predicciones del modelo son razonablemente cercanas a los valores reales en promedio.

Por lo tanto, se concluye que el modelo de regresión lineal actual sigue siendo útil y representativo para predecir la variable objetivo con las variables evaluadas. Sin embargo, es importante tener en cuenta que, a medida que el volumen de robos aumenta, las predicciones del modelo pueden ser menos precisas debido a la complejidad del fenómeno y a otros factores no capturados por las variables en el modelo.

En resumen, el análisis de regresión muestra que el modelo de predicción es representativo en el rango de valores observados, pero es posible que se requieran ajustes adicionales o consideración de otros factores para mejorar la precisión de las predicciones, especialmente en situaciones de mayor volumen de robos. Se concluye que dados los resultados del modelo, las variables seleccionadas son representativas y afectan al resultado final para poder estimar si un vehículo, según su marca, zona donde se encuentre, su año, y la edad del propietario registral son factores preponderantes al momento de entender si existe un tipo específico de combinaciones de esas variables desencadenen en un riesgo mayor a ser pasible de robo. A mayor cantidad de casos, más se nota la tendencia de ciertos modelos, años y lugares que son pasibles de generar siniestros por robo.