viu
.es

2024 - 2025

ACTIVIDAD 1

Máster en Big Data y Data Science

06MBID – Estadística Avanzada

Nombre: **Gonzalo Antonio Delgado Rubio**

Fecha: **08/09/2024**

Curso 2024 – Ed. Abril

viu | Universidad Internacional de Valencia

TABLA DE CONTENIDO

1 INTRODUCCIÓN

1.1 CONTEXTO Y MOTIVACIÓN

En el mundo actual, el sector de seguros de viaje es crucial para proporcionar protección financiera ante eventos imprevistos que pueden afectar a los viajeros, como enfermedades, cancelaciones de vuelos o pérdidas de equipaje. Las compañías de seguros se enfrentan al desafío de gestionar eficientemente las reclamaciones, optimizando su toma de decisiones y sus políticas de suscripción. Por otro lado, las ventas netas también son un aspecto fundamental para la sostenibilidad financiera de estas compañías, ya que reflejan la efectividad de sus estrategias de mercado y la aceptación de sus productos.

Este estudio tiene como objetivo aplicar tres modelos de regresión distintos para analizar dos aspectos clave del negocio de seguros de viaje: el resultado de las reclamaciones (aceptación o rechazo) y las ventas netas. Utilizaremos un conjunto de datos público de Kaggle, que proporciona información detallada sobre ambos aspectos. La comprensión de los factores que influyen tanto en las ventas como en el manejo de reclamaciones permitirá a las compañías de seguros mejorar sus modelos de riesgo, optimizar sus políticas de precios y aumentar la eficiencia operativa.

Es así que **para poder analizar las ventas netas de las pólizas de seguros se utilizarán modelos de regresión lineal y multilineal**. El objetivo aquí será identificar las variables que influyen en el volumen de ventas, como la edad del cliente, el tipo de seguro contratado, y la duración del viaje. De este análisis podrémos entender:

- Dinámicas del mercado: Evaluar como diversas características del cliente y producto afectarán las ventas netas, de esta forma ajustar la estrategia de marketing y precios.
- Optimizar los ingresos: Al permitirnos identificar segmentos del mercado más rentables.

Posteriormente, **se empleará la regresión logística para predecir la aceptación o rechazo de las reclamaciones de seguros de viaje**. Comprender qué factores influyen en el resultado de una reclamación es fundamental para las aseguradoras, ya que les permite:

- Reducir los costos operativos: Al poder tomar decisiones basadas en datos se reducirán los costos y permitirá aumentar la eficiencia.
- Mejorar la experiencia al cliente: Los cuales podrán tener mayor claridad de los criterios establecidos para la aceptación o rechazo de algún reclamo

1.2 OBJETIVOS DEL ANÁLISIS

Este estudio tiene como propósito principal aplicar técnicas de regresión para analizar dos aspectos clave del negocio de seguros de viaje: el manejo de reclamaciones y el análisis de ventas netas.

Los objetivos específicos del análisis son los siguientes:

1. Describir y analizar estadísticamente el conjunto de datos de seguros de viaje.

2. Estimar y validar un modelo de regresión lineal para explorar la relación entre las variables numéricas y las ventas netas.
3. Estimar y validar un modelo de regresión multivariable para predecir las ventas netas de las pólizas de seguro.
4. Estimar y validar un modelo de regresión logística para predecir para analizar el resultado de las reclamaciones de seguros de viajes.
5. Extraer conclusiones significativas y proporcionar recomendaciones.

2 DESCRIPCIÓN DE LOS DATOS A ANALIZAR

El conjunto de datos utilizado en este análisis proviene de una fuente pública en Kaggle, llamada "Travel Insurance Dataset", y contiene información detallada sobre las reclamaciones de seguros de viaje y las ventas netas de pólizas. El conjunto de datos está compuesto por 12 atributos los cuales se describen a continuación:

- **Claim Status (Estado de la Reclamación):**
  La variable objetivo para el modelo de regresión logística, indica si una reclamación de seguro de viaje fue aceptada o rechazada. Es una variable categórica binaria.
- **Agency (Agencia):**
  Nombre de la agencia de seguros que gestionó la póliza. Esta variable es categórica y puede tener múltiples valores, representando diferentes agencias.
- **Agency Type (Tipo de Agencia):**
  Tipo de agencia de seguros (por ejemplo, 'Travel Agency' o 'Airlines'), categorizando las agencias en función de su modelo de negocio principal. Esta variable es categórica.
- **Distribution Channel (Canal de Distribución):**
  Canal de distribución utilizado por la agencia de seguros (por ejemplo, 'Offline' o 'Online'), que indica cómo se comercializó la póliza. Es una variable categórica.
- **Product Name (Nombre del Producto):**
  Nombre del producto de seguro de viaje adquirido (por ejemplo, 'Basic', 'Comprehensive'). Esta variable es categórica y representa diferentes tipos de pólizas de seguro de viaje.
- **Duration (Duración del Viaje):**
  Duración del viaje asegurado en días. Es una variable numérica que puede influir en la aceptación de la reclamación y en el valor de las ventas netas de las pólizas.
- **Destination (Destino del Viaje):**
  Destino del viaje asegurado, representando el país o región a la que se dirige el cliente. Es una variable categórica.
- **Net Sales (Ventas Netas):**
  Monto total de ventas de las pólizas de seguro de viaje. Esta es una variable numérica continua que servirá como variable dependiente en los modelos de regresión lineal y multilineal para analizar los factores que afectan las ventas.
- **Commission (Comisión):**
  Comisión recibida por la agencia de seguros por cada póliza vendida. Es una variable

numérica que puede estar relacionada con las ventas netas y las estrategias de distribución.

- **Gender (Género del Asegurado):**
  Género del cliente que ha contratado la póliza de seguro ('M' para masculino, 'F' para femenino). Es una variable categórica que puede influir en el comportamiento de compra y las reclamaciones.
- **Age (Edad del Asegurado):**
  Edad del cliente en el momento de la compra de la póliza. Es una variable numérica que podría afectar tanto la probabilidad de una reclamación como el tipo de producto adquirido.

En la fase inicial, se realizará un análisis exploratorio de datos (EDA) para comprender mejor las distribuciones y relaciones entre estas variables, identificar posibles valores atípicos y determinar la adecuación de cada variable para los modelos de regresión.

Al analizar tanto el **estado de las reclamaciones** como las **ventas netas** de las pólizas de seguro, este conjunto de datos proporciona una base sólida para entender los factores que influyen en la aceptación de las reclamaciones y en la rentabilidad de los productos de seguros de viaje.

2.1 Resumen Estadístico del Conjunto de Datos

El conjunto de datos inicial contiene alrededor de 63 000 registros. Sobre estos estamos considerando 10 variables que han de ser analizada, estas mismas las podemos dividir en 2 grupos los cuales son:

- **Variables Numéricas:** age, commision, duration, net_sales
- **Variables Categóricas:** agency, agency_type, destination, distribution_channel, product_name, gender

De las estadísticas iniciales de los datos, nos damos cuenta de que existen valores negativos para los campos duration y net_sales, los cuales fueron considerados como datos errones del dataset. Por lo cual fueron eliminados del mismo.

A psych: 12 × 13

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
id	1	62288	3.284495e+04	1.806542e+04	33446.50	3.308708e+04	22892.826600	0	63323.00	63323.00	-0.09030053	-1.1633701	7.238454e+01
age	2	62288	3.966632e+01	1.401465e+01	36.00	3.769584e+01	7.413000	0	118.00	118.00	2.86558337	12.4269866	5.615393e-02
agency*	3	62288	7.200729e+00	2.674967e+00	8.00	7.164513e+00	1.482600	1	16.00	15.00	-0.07415869	0.3235035	1.071806e-02
agency_type*	4	62288	1.655279e+00	4.752811e-01	2.00	1.694092e+00	0.000000	1	2.00	1.00	-0.65340786	-1.5730834	1.904357e-03
commision	5	62288	1.282970e+01	2.349874e+01	1.88	7.235987e+00	2.787288	0	262.76	262.76	3.36748063	16.5153677	9.415480e-02
destination*	6	62288	6.090452e+01	2.849731e+01	73.00	6.311366e+01	23.721600	1	102.00	101.00	-0.60548286	-0.8625374	1.141831e-01
distribution_channel*	7	62288	1.982083e+00	1.326501e-01	2.00	2.000000e+00	0.000000	1	2.00	1.00	-7.26837803	50.8301352	5.315025e-04
duration	8	62288	6.095880e+01	1.143253e+02	25.00	3.532459e+01	26.686800	-2	4881.00	4883.00	14.86221447	555.5211229	4.580789e-01
gender*	9	62288	1.544920e+00	7.800916e-01	1.00	1.431169e+00	0.000000	1	3.00	2.00	0.98997324	-0.6451385	3.125672e-03
net_sales	10	62288	5.071706e+01	6.316671e+01	29.70	3.709657e+01	21.201180	-389	682.00	1071.00	2.71121551	9.6263399	2.530965e-01
product_name*	11	62288	1.026384e+01	6.487542e+00	11.00	9.880258e+00	8.895600	1	25.00	24.00	0.33389501	-0.5538235	2.599429e-02
claim	12	62288	2.000064e-01	4.000080e-01	0.00	1.250201e-01	0.000000	0	1.00	1.00	1.49991371	0.2497451	1.602753e-03

Luego de una limpieza inicial, podemos seguir el estudio del dataset con un conjunto de datos de alrededor de 50 000 registros.

A psych: 12 × 13

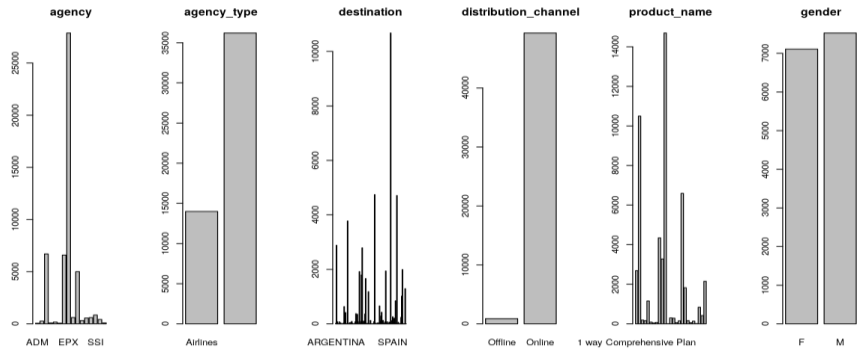
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
id	1	50207	3.161462e+04	1.828197e+04	31557	31597.750392	23478.4536	21	63323.00	63302.00	0.007738696	-1.1985683	8.159074e+01
age	2	50207	3.990673e+01	1.396943e+01	36	37.912690	5.9304	0	118.00	118.00	2.990224554	13.1659366	6.234426e-02
agency*	3	50207	7.649212e+00	2.387566e+00	8	7.717554	0.0000	1	16.00	15.00	-0.144376246	1.5352747	1.065549e-02
agency_type*	4	50207	1.721294e+00	4.483671e-01	2	1.776608	0.0000	1	2.00	1.00	-0.987090469	-1.0256728	2.001021e-03
commision	5	50207	9.729067e+00	1.976992e+01	0	4.962594	0.0000	0	262.76	262.76	4.060193400	25.0858310	8.823131e-02
destination*	6	50207	5.902613e+01	2.888012e+01	64	60.780218	37.0650	1	102.00	101.00	-0.448271308	-1.0125404	1.288893e-01
distribution_channel*	7	50207	1.982373e+00	1.315928e-01	2	2.000000	0.0000	1	2.00	1.00	-7.331153059	51.7468358	5.872863e-04
duration	8	50207	4.934947e+01	1.041308e+02	22	30.819155	23.7216	0	4881.00	4881.00	23.397450277	1010.1142402	4.647259e-01
gender*	9	14637	1.514245e+00	4.998141e-01	2	1.517804	0.0000	1	2.00	1.00	-0.056996185	-1.9968878	4.131260e-03
net_sales	10	50207	4.191868e+01	4.790829e+01	27	32.512494	17.7912	0	682.00	682.00	3.698666296	19.6287403	2.138103e-01
product_name*	11	50207	1.020384e+01	6.546333e+00	11	9.761396	8.8956	1	25.00	24.00	0.355222772	-0.4629375	2.921568e-02
claim	12	50207	1.838389e-02	1.343365e-01	0	0.000000	0.0000	0	1.00	1.00	7.170153815	49.4120899	5.995311e-04

Además de esto, también se vio que existía una gran cantidad de valores nulos para el campo gender, alrededor de 64% de nulos, sobre los cuales se tomarán acciones más adelante.

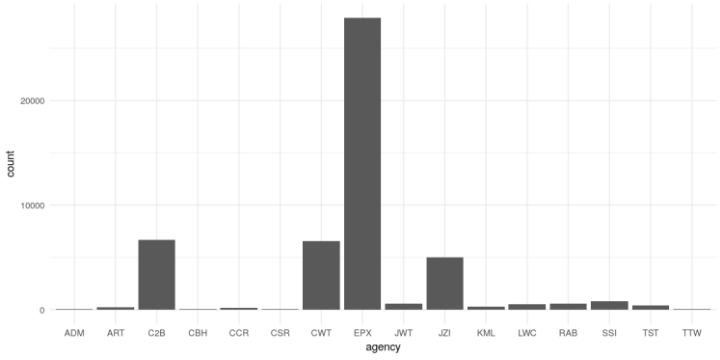
A continuación, se presentan las visualizaciones de las distribuciones para cada una de las variables mencionadas.

2.2 Visualización de Distribuciones

2.2.1 Distribución de Variables Categóricas

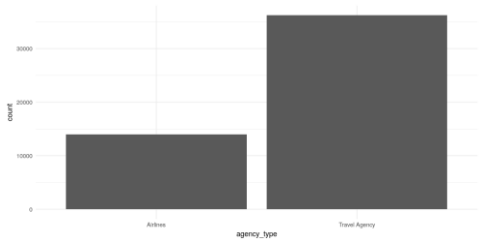


- **Agencia:**
  - EPX es la agencia con mayor cantidad reclamaciones
  - En tanto ADM, TTW, CBH y CSR contiene menor cantidad de reclamaciones



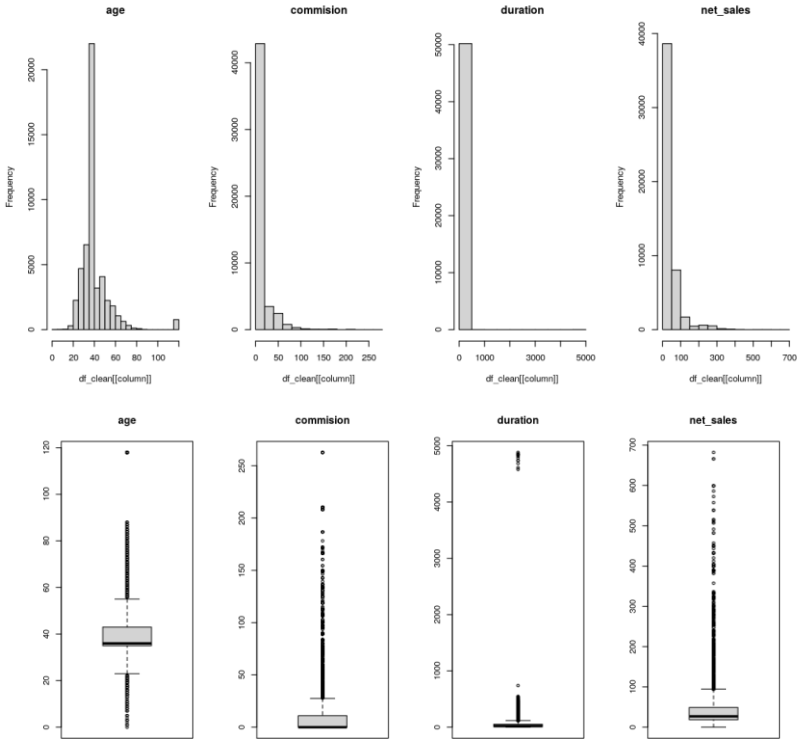
- **Tipo de Agencia**
  - Agencias de viajes son las que se llevan la mayor parte de las reclamaciones

- Aerolíneas poseen poca cantidad de reclamaciones



- **Canal de Distribución**
  - El canal de distribución online es el preferido para los reclamos
- **Género**
  - Se observa un balance entre la cantidad de reclamos según el sexo.

2.2.2 Distribución de Variables Numéricas

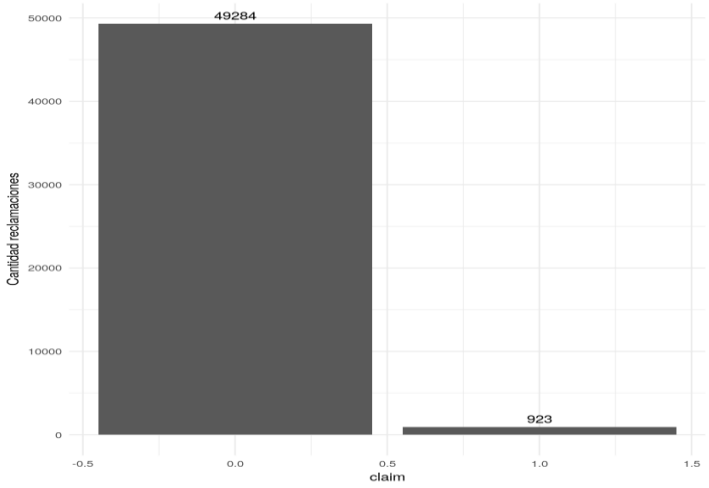


- **Age:** La edad promedio de las personas que realizan un reclamo esta alrededor de 36 años.
- **Commision:** Las comisiones suelen ser bajas y están por los 10 dolares. Para esta variable observamos muchos datos outliers, por lo que estos deberán ser tratados antes de realizar la regresión logística.
- **Duration:** En general vemos que la duración de los seguros de viajes adquiridos por los clientes es de 22 dias. Aunque existen muchos casos atípicos que generan ruido en nuestro conjunto de datos.

- **Net_Sales:** Las ventas netas promedio están alrededor de 29.7. En este caso existen demasiados datos outliers que también deberán ser tratados antes de realizar la regresión.

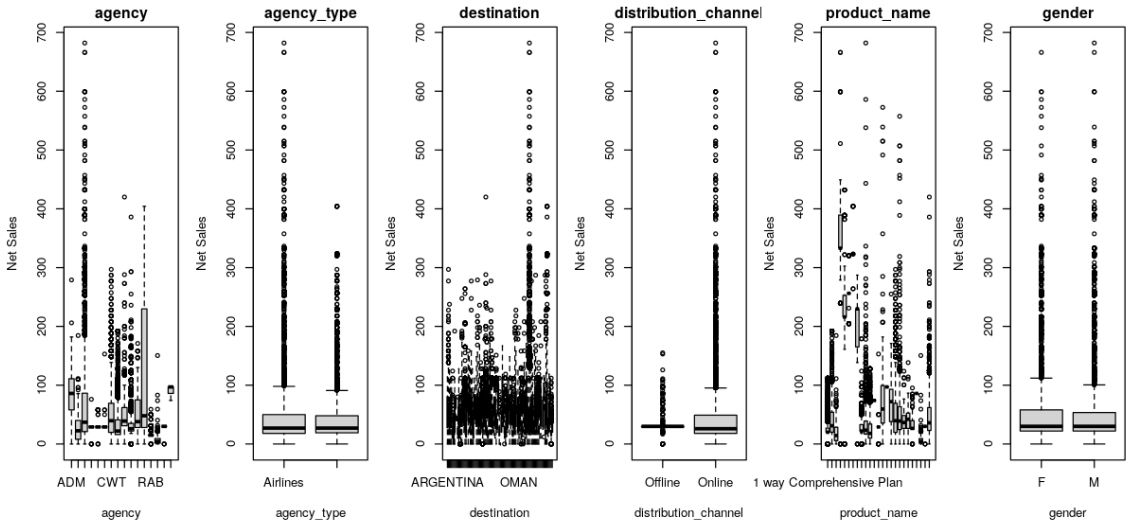
2.2.3 Distribución de Variable Objetivo Claim

Al dar un vistazo a nuestro variable objetivo "claim" podemos observar que la gran mayoría de las reclamaciones han sido rechazadas. Esto a futuro origina un desbalance a los datos que tendrá implicaciones importantes en el desarrollo del modelo predictivo de regresión logística, para esto se deberá aplicar técnicas de balanceo en dicho apartado.



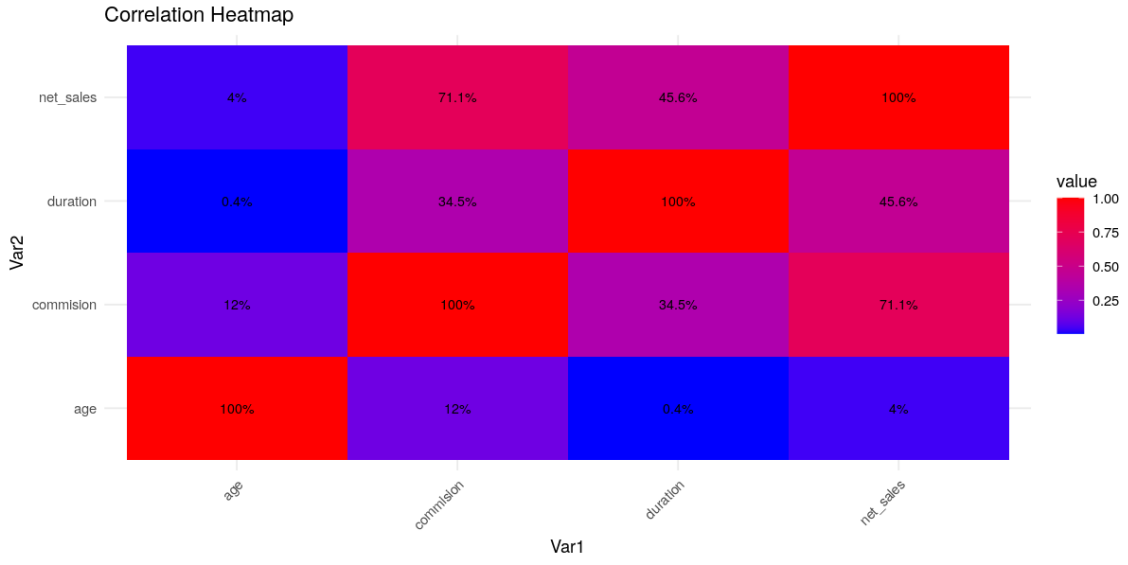
2.2.4 Relación Variables Categóricas con Variable Objetivo Net Sales

Los gráficos de cajas realizados muestran como se comporta las variables categóricas respecto a las ventas realizadas por las diferentes compañías de seguros. Apreciamos que, de acuerdo con la agencia, las ventas obtenidas por estas pueden variar mucho. Mientras que, para las variables de tipo de agencia, canal de distribución y genero parece encontrarse más estables los valores obtenidos. Esto nos servirá al momento de realizar el modelo de regresión multilineal.



2.3 Matriz de Correlación

Se generó el siguiente mapa de calor para las variables numéricas comentadas donde se puede ver que existe una **relación más fuerte entre la variable comisión y net_sales**. Esta **se empleará para poder realizar el primer modelo de regresión lineal**.



Para los demás casos, al no tratarse de la variable objetivo, se puede decir que es adecuado que no exista una correlación entre las mismas variables para que cada una pueda aportar cierto valor al target. Se disminuye la multicolinealidad.

2.4 Valores Atípicos

El siguiente cuadro muestra la cantidad de valores atípicos existentes en el dataset y el porcentaje que estos representan en el conjunto total.

A data.frame: 4 × 4

	Variable	Outliers	Record_Count	Percentage_Outliers
	<chr>	<int>	<int>	<dbl>
age	age	5844	50207	11.639811
commision	commision	5589	50207	11.131914
duration	duration	4373	50207	8.709941
net_sales	net_sales	4028	50207	8.022786

2.4.1 Estandarización de los Datos

Considerando la presencia de valores atípicos importantes en nuestros datos y su potencial relevancia para nuestro análisis, se hizo uso de diversos métodos de estandarización para finalizar la etapa de preprocesamiento y construcción de los features finales. Es así que, según el modelo a realizar, se emplearon diversas técnicas descritas a continuación.

- **Estandarización Modelo Regresión Lineal**:
  - Empleo de rango intercuartil para tratamiento de los valores atípicos en el modelo a construir.
- **Estandarización Modelo Regresión Lineal Múltiple:**
  - Aplicación de transformación logarítmica sobre las variables numéricas.
  - Aplicación de rango intercuartil para tratar valores atípicos.
- **Estandarización Modelo Regresión Logística**:
  - Aplicación de transformación logarítmica sobre las variables numéricas
  - Aplicación de rango intercuartil para tratar valores atípicos.
  - Balanceo de los datos por técnica SMOTE

3 Análisis

Según lo comentado en el apartado de introducción, se esta sección se estará realizando modelos de regresión lineal, multivariable y logarítmica para las variables objetivo-identificadas como ventas netas y reclamaciones de viajes realizados.

Además, hay que comentar que para cada modelo se han empleado diversas formas de estandarización de los datos tal como se menciono en el apartado anterior.

3.1 Regresión Lineal

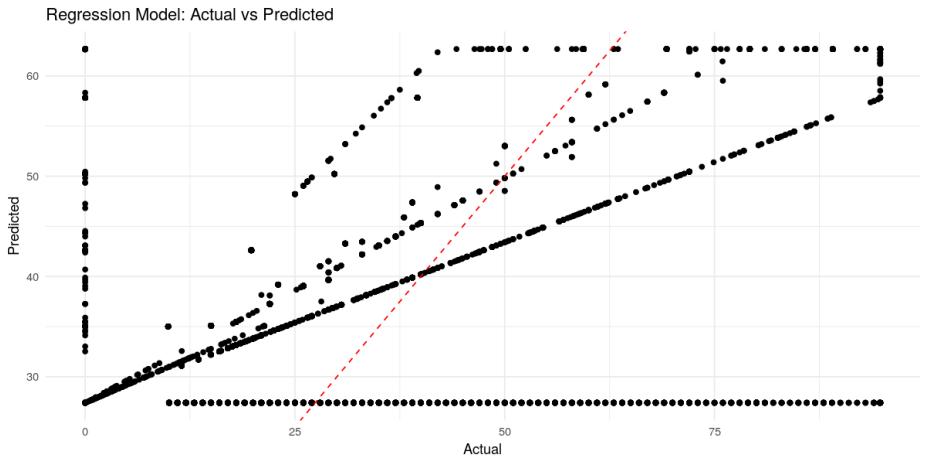
Para este primero modelo a construir, se tomará en cuenta la variable dependiente NetSale y su comportamiento en relación con la variable independiente commision.

```
Call:
lm(formula = net_sales ~ ., data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-62.682 -15.406  -7.406   9.716  67.344

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 27.40576    0.14212   192.8   <2e-16 ***
commision    1.28044    0.01216   105.3   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.35 on 40165 degrees of freedom
Multiple R-squared:  0.2164,    Adjusted R-squared:  0.2164
F-statistic: 1.109e+04 on 1 and 40165 DF,  p-value: < 2.2e-16
```

```
Mean Squared Error (MSE): 530.903
Root Mean Squared Error (RMSE): 23.04133
R-squared: 0.2183604
```

A partir de los resultados obtenidos podemos decir que para este primer modelo generado que:

- Aunque la commission es un predictor estadísticamente significativo de net_sales, el bajo valor de R-squared (21.83%) indica que hay otros factores importantes que afectan las ventas netas que no se han incluido en este modelo.
- Como se verá más adelante, será beneficioso agregar más variables a nuestro modelo

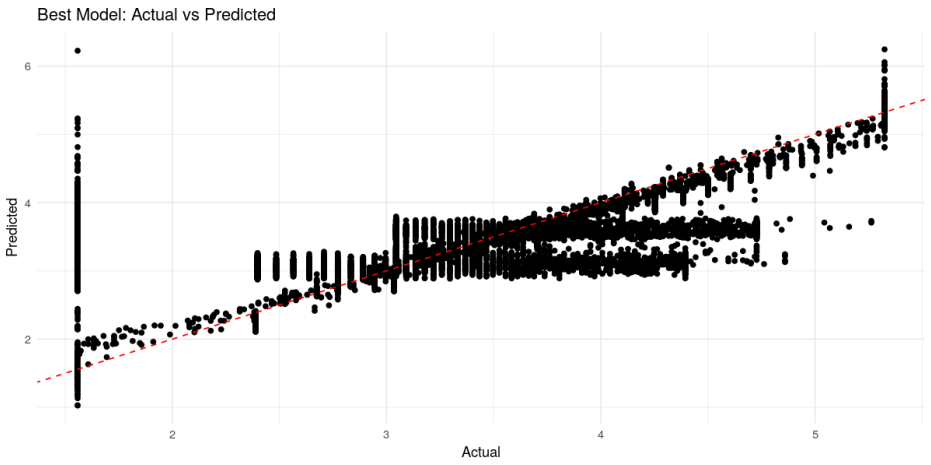


3.2 Regresión Multilineal

Con las nuevas variables generadas durante el procesamiento de los datos, se ha empleado el método forward para que en r se pueda ir ingresando variables al modelo y finalmente se escoja el modelo con el mejor performance. Para esto se obtuvieron los siguiente resultados.

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.4575 -0.2515  0.0394  0.2364  1.7447

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                           0.718657   0.056754  12.663  < 2e-16
commision                             0.908043   0.007131 127.342  < 2e-16
agencyEPX                             1.907463   0.018983 100.484  < 2e-16
agencyCWT                            -0.626169   0.012860 -48.692  < 2e-16
product_name2wayComprehensivePlan     0.504922   0.008138  62.047  < 2e-16
agencyTTW                             3.167254   0.071750  44.143  < 2e-16
agencyC2B                             0.341751   0.016122  21.198  < 2e-16
...
Residual standard error: 0.5659 on 40145 degrees of freedom
Multiple R-squared:  0.4953,    Adjusted R-squared:  0.495
F-statistic:  1876 on 21 and 40145 DF,  p-value: < 2.2e-16
```

```
Mean Squared Error (MSE): 0.3374363
Root Mean Squared Error (RMSE): 0.5808927
R-squared: 0.473162
```



- El modelo de regresión lineal múltiple ha identificado 23 variables significativas que afectan las ventas netas (net_sales), incluyendo la commission, ciertas agencies, y product_name.
- Aunque el modelo explica aproximadamente el 47.31% de la variabilidad en las ventas netas, todavía hay una cantidad considerable de variabilidad que no se explica. Esto

sugiere que puede haber otras variables no incluidas en el modelo que influyen en net_sales.

- El modelo es estadísticamente significativo y tiene un rendimiento decente según las métricas de error, pero puede beneficiarse de ajustes adicionales o de la inclusión de más variables predictoras relevantes para mejorar su capacidad explicativa.

3.3 Regresión Logística

Se empleo regresión logística para predecir los valores de reclamaciones realizadas los por los usuarios. De aquí se obtuvo los siguientes resultados.

```
Best Model Selected by Forward Selection:

Call:
glm(formula = claim ~ net_sales + age + duration + commision,
    family = binomial, data = train_data)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.6471026  0.0403651 -40.805  < 2e-16 ***
net_sales    0.0115571  0.0003688  31.340  < 2e-16 ***
age         -0.0109741  0.0009649 -11.374  < 2e-16 ***
duration    -0.0004681  0.0001786  -2.621  0.00878 **
commision    0.0019584  0.0007405   2.645  0.00818 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 49538  on 49359  degrees of freedom
Residual deviance: 44493  on 49355  degrees of freedom
AIC: 44503

Number of Fisher Scoring iterations: 5
```

```
                    y_test
binary_predictions     0    1
                   0 9586 1992
                   1  260  502
```

- El modelo inicial mostró una baja precisión pero un buen recall. Esto indica que, aunque el modelo identifica muchas reclamaciones aceptadas correctamente, tiene una alta tasa de falsos positivos.
- El modelo ajustado con forward selection muestra coeficientes significativos y mejoró el ajuste del modelo en comparación con el modelo nulo. Sin embargo, el modelo aún podría tener un desempeño limitado en términos de precisión y F1-Score.

4 CONCLUSIONES

El análisis realizado mediante diferentes modelos de regresión ha permitido obtener importantes insights sobre las ventas netas y el comportamiento de las reclamaciones en el negocio de seguros de viaje. Los principales hallazgos incluyen:

- **Regresión lineal**: Aunque la variable "comisión" se mostró como un predictor significativo de las ventas netas, el bajo valor de R^2 (21.83%) indica que otros factores tienen una influencia considerable en las ventas, lo que sugiere que el modelo es limitado si no se incluyen más variables relevantes.
- **Regresión multilineal**: Este modelo identificó 23 variables significativas, lo que mejoró la capacidad explicativa del modelo, alcanzando un R^2 de 47.31%. A pesar de ello, aún queda un margen importante de variabilidad que no se explica, lo que implica que podrían existir factores adicionales que influyen en las ventas netas y que no fueron considerados en el análisis actual.
- **Regresión logística**: El modelo inicial mostró una alta tasa de falsos positivos y, aunque la precisión fue baja, el recall fue alto. Esto significa que el modelo puede predecir correctamente muchas reclamaciones aceptadas, pero a costa de generar errores de predicción. El ajuste mediante selección "forward" mejoró el rendimiento del modelo, aunque persisten limitaciones en términos de precisión.

En resumen, los modelos empleados proporcionaron información valiosa para la comprensión de las ventas y reclamaciones en seguros de viaje, pero aún hay margen para mejorar la predicción y la identificación de variables adicionales que puedan contribuir a una mayor capacidad explicativa.

5 LIMITACIONES Y TRABAJO FUTURO

Como limitaciones y trabajo futuro quedan lo siguiente:

- **Bajo valor de R^2 en algunos modelos**: A pesar de identificar variables significativas, el bajo R^2 en el modelo de regresión lineal sugiere que una gran parte de la variabilidad no se explica. Esto puede deberse a la ausencia de otras variables predictoras o a posibles relaciones no lineales no exploradas.
- **Desbalance en los datos**: El desbalance en los datos de las reclamaciones, donde la mayoría fueron rechazadas, puede haber afectado la precisión del modelo de regresión logística, dando lugar a un alto número de falsos positivos.
- **Datos atípicos**: Aunque se utilizaron técnicas para tratar valores atípicos, su presencia en variables clave como "ventas netas" y "duración del viaje" pudo haber afectado la precisión y estabilidad de los modelos.
- **Falta de información adicional**: Algunas variables potencialmente influyentes, como el comportamiento de mercado, cambios en políticas de precios, o factores externos (por ejemplo, condiciones económicas), no fueron incluidas en el análisis.

Estas acciones permitirán avanzar hacia un análisis más profundo y una mejora en la capacidad predictiva de los modelos para la toma de decisiones en el ámbito de los seguros de viaje.

6 ANEXOS

6.1 ANEXO 1: DATASET

Se Brinda link de donde se obtuvieron los datos tratados en el estudio

https://www.kaggle.com/datasets/mhdzahier/travel-insurance/data

6.2 CÓDIGO REALIZADO PARA EL TRABAJO

Se brinda link con el desarrollo del caso además del cuaderno descargado

https://colab.research.google.com/drive/16vDOo7s1dcGucAKvXoDo4g8GekgpF_g6#scrollTo=g-SAbL0HSfuA

actividad1_GonzaloDelgado.ipynb