

DESCRIPCIÓN DE DATOS

Para comenzar con la descripción de los datos, hemos dividido la descripción de los datos en dos nodos, uno analítico y otro gráfico.

Nodo analítico

En el nodo analítico hemos realizado en primer lugar, descripciones estadísticas y medias, para ver la importancia y la relación de las variables. A continuación hemos incluido un resumen de 4 de las variables: empleados, ventas, resultados y activos.

EMPLEADOS	
Estadísticas	
Recuento	21562
Media	15.544
Suma	335158.000
Mín.	0.000
Máx.	82470.000
Rango	82470.000
Varianza	335473.675
Desviación estándar	579.201
Error estándar de la media	3.944
VENTAS	
Estadísticas	
Recuento	21562
Media	2927955.718
Suma	63132581197.440
Mín.	-4569000.000
Máx.	11722958280.000
Rango	11727527280.000
Varianza	10507793641879358.000
Desviación estándar	102507529.684
Error estándar de la media	698089.727
RESULTADOS	
Estadísticas	
Recuento	21562
Media	126068.317
Suma	2718285047.480
Mín.	-1062761000.000
Máx.	1373621000.000
Rango	2436382000.000
Varianza	169134780192381.500
Desviación estándar	13005182.820
Error estándar de la media	88567.001

EMPLEADOS

En el análisis de empleados observamos que dentro de las 21.562 empresas hay un total de 335.158 empleados, lo que aproximadamente corresponde a unos 15/16 empleados por empresa. Adicionalmente, hay empresas que no cuentan con trabajadores, mientras que otras cuentan con hasta 82.470 trabajadores.

VENTAS

Mediante este análisis de ventas podemos observar que los datos de las empresas registradas son muy dispersos, la diferencia entre el máximo y el mínimo es descomunal. Adicionalmente, esta variable cuenta con una varianza de 10.507.793.641.879.358

RESULTADOS

Al igual que la variable ventas, obtenemos valores muy dispersos y una varianza muy elevada, además de una desviación de error muy significativa.


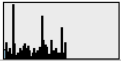
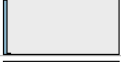


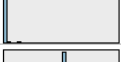


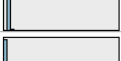


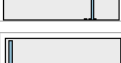


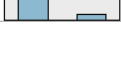
	EMPLEADOS	NUM_INCIDENCIAS	VENTAS	RESULTADOS	ACTIVO	PATRIMONIO_NETO
EMPLEADOS	1.000/Perfecto	-0.001/Débil	0.120/Fuerte	0.021/Fuerte	0.063/Fuerte	0.041/Fuerte
VENTAS	0.120/Fuerte	0.067/Débil	1.000/Perfecto	0.085/Fuerte	0.277/Fuerte	0.150/Fuerte
RESULTADOS	0.021/Fuerte	0.134/Débil	0.085/Fuerte	1.000/Perfecto	0.311/Fuerte	0.584/Fuerte
ACTIVO	0.063/Fuerte	0.351/Fuerte	0.277/Fuerte	0.311/Fuerte	1.000/Perfecto	0.788/Fuerte
PATRIMONIO_NETO	0.041/Fuerte	0.247/Fuerte	0.150/Fuerte	0.584/Fuerte	0.788/Fuerte	1.000/Perfecto
CAPITAL_SOCIAL	0.039/Fuerte	0.217/Fuerte	0.130/Fuerte	0.708/Fuerte	0.804/Fuerte	0.947/Fuerte
ENDEUDAMIENTO_CORTO_PLAZO	0.100/Fuerte	0.338/Fuerte	0.521/Fuerte	0.039/Fuerte	0.638/Fuerte	0.212/Fuerte
FONDO_MANIOBRA	-0.012/Medio	-0.222/Fuerte	-0.145/Fuerte	0.569/Fuerte	-0.364/Fuerte	0.069/Fuerte
FONDOS_PROPIOS	0.041/Fuerte	0.257/Fuerte	0.148/Fuerte	0.582/Fuerte	0.789/Fuerte	1.000/Fuerte

Adicionalmente, mediante la descripción estadística de las variables en la matriz de correlación de Pearson, observamos que hay variables que tienen una correlación perfecta como es “Patrimonio neto” y “Fondo Propios”, detalle que será importante y que tendremos en cuenta a la hora de realizar la preparación de datos.

En segundo lugar, hemos realizado un análisis de medias, para ver la importancia de las variables, frente a que una empresa se encuentre en riesgo o no. Como podemos observar en la tabla, que una empresa sea denominada en riesgo o no depende de una combinación de muchas variables. Lo que sí que podemos determinar, es que la cantidad de empleados y las ventas generadas por esa empresa no son significativas a la hora de determinar si son empresas en riesgo.

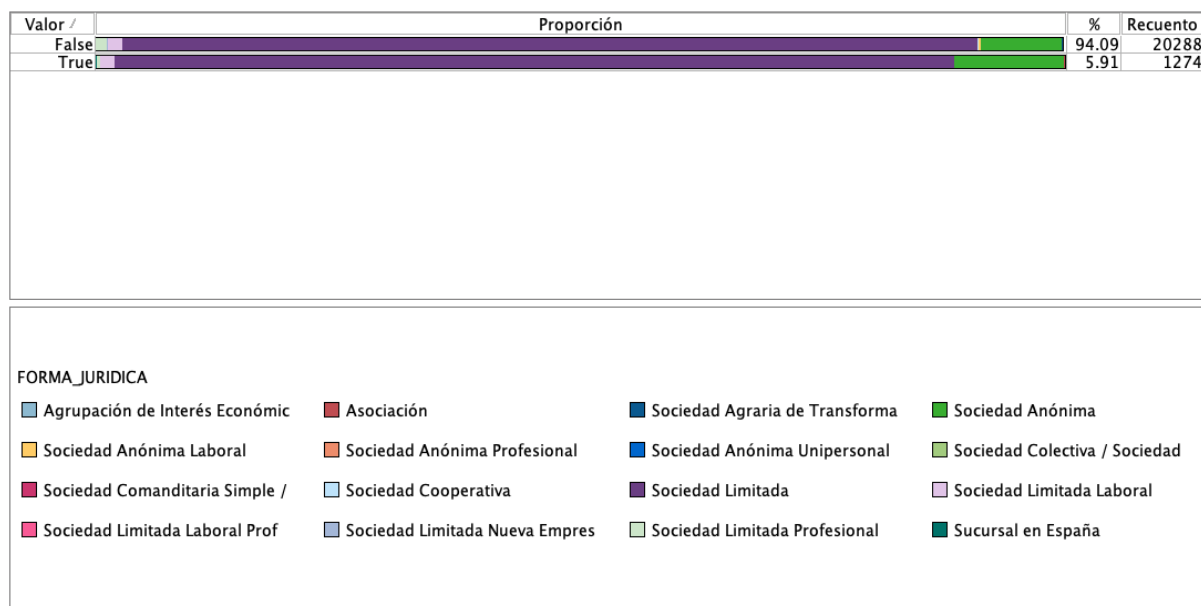
Campo /	False*	True*	Importancia
ACTIVO	3538144.308	29036165.945	1.000 ★ Importante
CAPITAL_SOCIAL	577436.940	3622786.661	0.999 ★ Importante
EMPLEADOS	15.811	11.290	0.213 □ Sin importancia
ENDEUDAMIENTO_CORTO_PLAZO	994846.202	6337354.566	1.000 ★ Importante
FONDO_MANIOBRA	278291.725	-2663492.196	1.000 ★ Importante
FONDOS_PROPIOS	1547379.923	13339164.941	1.000 ★ Importante
PATRIMONIO_NETO	1563930.072	13318039.299	1.000 ★ Importante
RESULTADOS	165161.672	-496479.559	0.922 + Marginal
VENTAS	2850711.987	4158034.852	0.341 □ Sin importancia

Para concluir con el nodo analítico, hemos realizado una descripción estadística de todas las variables, tanto continuas como nominales.

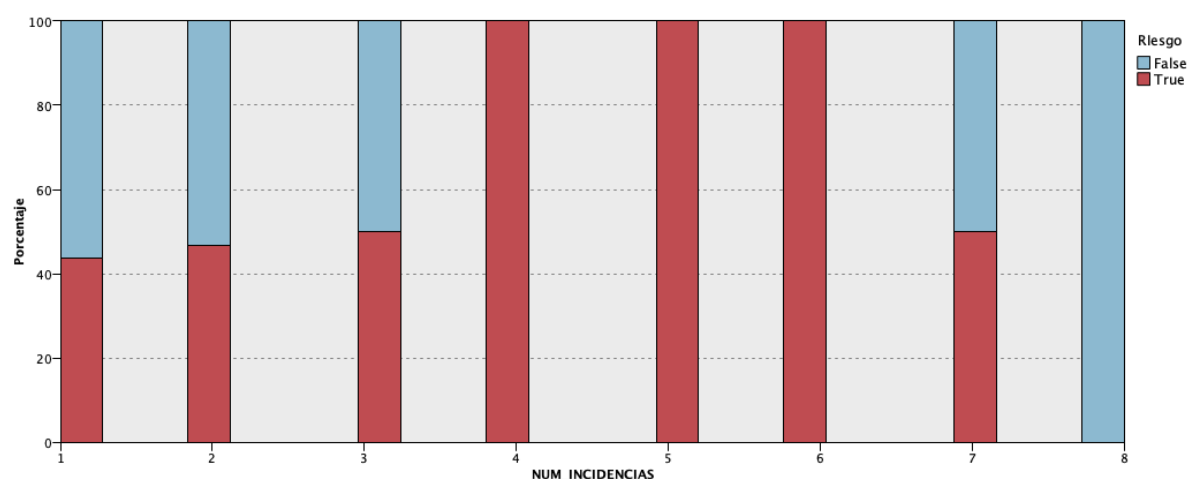
Campo	Gráfico de m...	Medida	Mín.	Máx.	Media	Desv. estándar	Sesgo	Exclusivo	Válido
FORMA_JURIDICA		Nominal	--	--	--	--	--	16	21562
PROVINCIA		Nominal	--	--	--	--	--	53	21562
EMPLEADOS		Continuo	0.000	82470.000	15.544	579.201	134.642	--	21562
NUM_RECLAMACIONES		Continuo	1.000	32.000	2.716	3.479	3.557	--	536
NUM_INCIDENCIAS		Continuo	1.000	8.000	1.514	1.294	3.413	--	105
VENTAS		Continuo	-45690...	1172295828...	2927955....	102507529....	89.558	--	21562
RESULTADOS		Continuo	-10627...	1373621000....	126068.3...	13005182.8...	35.205	--	21562
ACTIVO		Continuo	-48693...	1299975400...	5044705....	160760680....	69.671	--	21562
PATRIMONIO_NETO		Continuo	-43004...	1283765900...	2258426....	98829173.1...	112.459	--	21562
CAPITAL_SOCIAL		Continuo	-33035....	3245816000....	757372.7...	30522118.8...	93.914	--	21562
ENDEUDAMIENTO_C...		Continuo	-27864...	3948421000....	1310510....	39965896.4...	69.428	--	21562
FONDO_MANIOBRA		Continuo	-24166...	1072102000....	104475.1...	22339634.6...	-69.906	--	21562
FONDOS_PROPIOS		Continuo	-43017...	1283725400...	2244102....	98788994.7...	112.574	--	21562
EJERCICIO		Continuo	2015.000	2015.000	2015.000	0.000	--	--	21562
Riesgo		Marca	--	--	--	--	--	2	21562

Nodo Gráfico

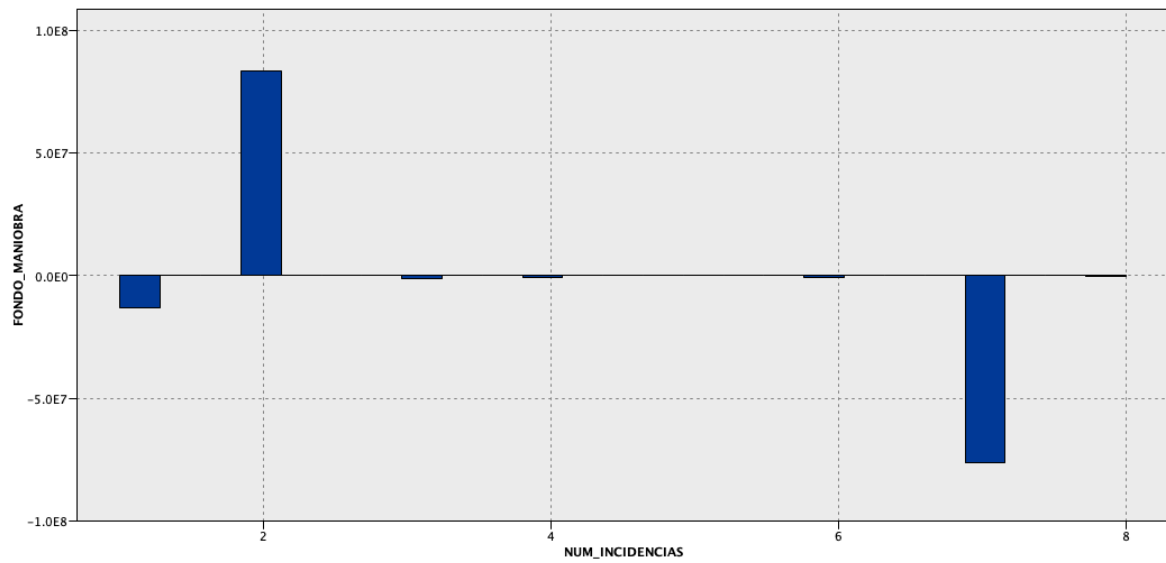
A continuación, mediante esta gráfica, podemos afirmar que la forma jurídica de las empresas no es relevante a la hora de determinar el riesgo de ella. Así mismo, observamos que los datos de nuestro modelo contiene un 6% de empresas que son evaluadas como riesgo.



Así mismo, hemos realizado un análisis del comportamiento de las incidencias.



En segundo lugar, mediante esta gráfica, podemos observar cómo se comporta el fondo de maniobra respecto a las incidencias. En este caso, aquellas empresas con un fondo de maniobra positivo tienen un número de incidencias menor, mientras que, cuando el número de incidencias crece, el fondo de maniobra disminuye. Es por ello, que podemos afirmar que están negativamente correlacionados.



En tercer lugar, con el objetivo de analizar el comportamiento de diferentes variables de nuestro dataset, nos hemos percatado de la necesidad de tratar los outliers del modelo.

