***Predicción de insolvencia de las 10000 empresas colombianas más importantes a partir de sus perfiles financieros.***

**Introducción**

El presente informe constituye el segundo avance del proyecto destinado a estudiar la salud financiera de las empresas más relevantes de Colombia y a identificar señales tempranas de riesgo de insolvencia a partir de información contable y financiera, para esto partimos del reporte anual de la Superintendencia de Sociedades (corte 31-12-2024) y de dos bases consolidadas utilizadas en el análisis, esto con el propósito de construir indicadores financieros, evaluar su distribución y relacionarlos con características sectoriales y territoriales.

Este avance se concentra en la preparación rigurosa de los datos siendo esta la estandarización, imputación y control de outliers y en un análisis exploratorio exhaustivo que incluye estadísticas descriptivas, histogramas de distribución, matriz de correlaciones y agregados por sector y departamento. A partir de estos procedimientos se generan insumos robustos como ratios financieros estandarizados y síntesis por sector, que nos permitirán, en fases posteriores, desarrollar modelos de diagnóstico o alerta temprana.

La estructura del documento es la siguiente: descripción y limpieza de los datos; construcción de indicadores; resultados del análisis exploratorio; análisis comparativo por sector y territorio; y finalmente, interpretación económica y conclusiones.

**Preparación y Construcción de Indicadores**

# **Integración y limpieza de las bases de datos**

# El análisis se realizó a partir de la unión de dos bases: la de las 1000 empresas más grandes y la de 9000 empresas adicionales, ambas publicadas por la Superintendencia de Sociedades. El procedimiento incluyó:

# Unificación de bases en un único DataFrame de 10.000 registros y 20 variables.

# Renombramiento de columnas para asegurar consistencia en los nombres de las variables financieras.

# Eliminación de duplicados y corrección de inconsistencias en categorías como región y macrosector.

# Tratamiento de valores faltantes: las cifras de 2024 ausentes fueron imputadas a partir de los valores de 2023 ajustados por inflación (9,28%), garantizando así continuidad temporal.

# Con este proceso aseguramos que la información esté depurada y sea comparable entre empresas

**Construcción de indicadores financieros**

A partir de las variables contables básicas (activos, pasivos, patrimonio, ingresos y utilidades), se construyeron ocho indicadores clave que sintetizan la posición financiera de cada empresa:

1. **Liquidez General = Activos / Pasivos**  
   Evalúa la capacidad de la empresa para responder con sus activos al total de pasivos.
2. **Apalancamiento = Pasivos / Activos**  
   Mide qué proporción de los activos está financiada con deuda.
3. **Razón de Endeudamiento = Pasivos / Patrimonio**  
   Refleja la dependencia de la financiación externa frente a los recursos propios.
4. **Autonomía Financiera = Patrimonio / Activos**  
   Complementa al apalancamiento, mostrando la independencia de los acreedores.
5. **ROA (Return on Assets) = Utilidad / Activos**  
   Mide la eficiencia en la utilización de activos para generar beneficios.
6. **ROE (Return on Equity) = Utilidad / Patrimonio**  
   Evalúa la rentabilidad para los accionistas.
7. **Margen Neto = Utilidad / Ingresos**  
   Indica qué porcentaje de las ventas se convierte en ganancia.
8. **Rotación de Activos = Ingresos / Activos**  
   Mide la eficiencia en el uso de activos para generar ingresos operacionales.

**Tratamiento de valores extremos**

Al calcular los indicadores, se detectaron problemas comunes en análisis financiero siendo lo mas común divisiones por cero y distribuciones fuertemente sesgadas y para abordarlos se implementaron dos estrategias:

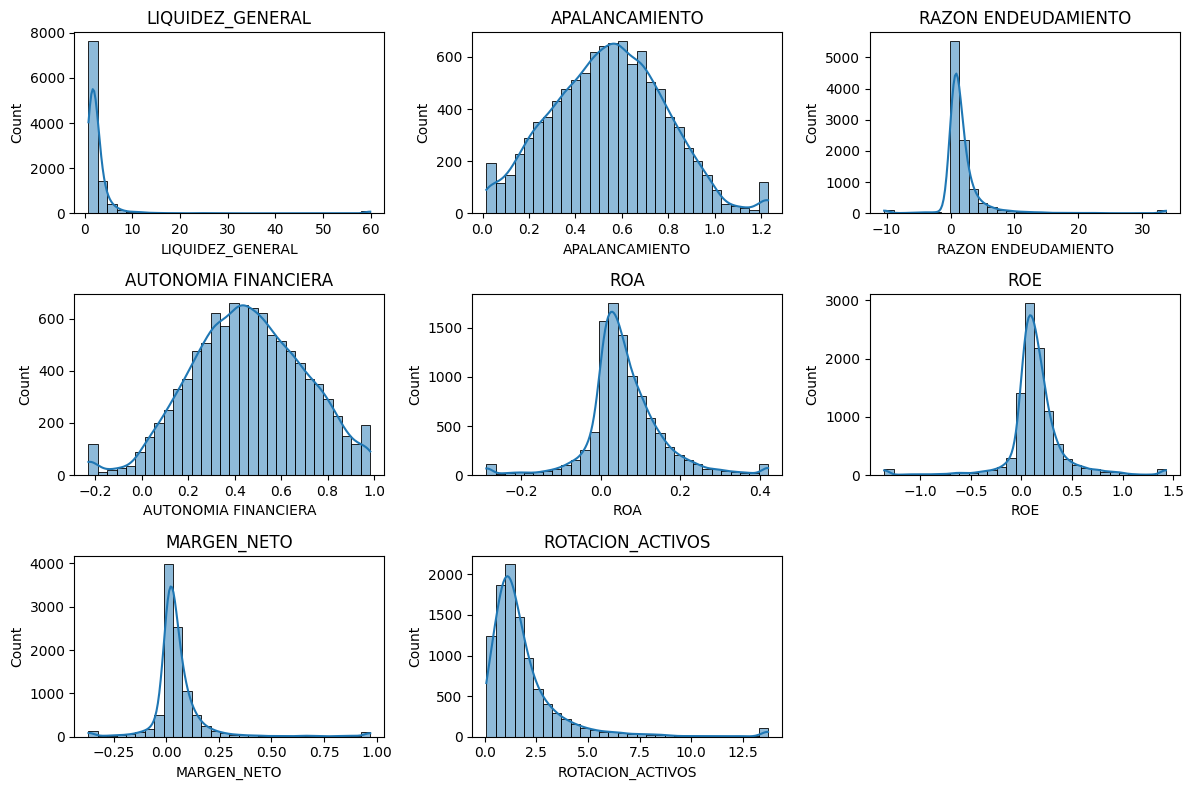
* **Reemplazo de infinitos por valores nulos (NaN)** en casos de denominadores cercanos a cero.
* **Winsorización en percentiles 1% y 99%** con el fin de recortar los valores extremos sin eliminar información valiosa.

Con estas medidas aplicadas, los indicadores representan mejor la distribución típica del mundo empresarial y evitan que unos pocos casos extremos distorsionen las conclusiones generales.

**Tratamiento de valores extremos**

Al calcular los indicadores, se detectaron problemas comunes en análisis financiero: divisiones por cero, valores infinitos y distribuciones fuertemente sesgadas. Para abordarlos se implementaron dos estrategias:

* **Reemplazo de infinitos por valores nulos (NaN)** en casos de denominadores cercanos a cero.
* **Winsorización en percentiles 1% y 99%**, que recorta los valores extremos sin eliminar información valiosa.

Con estas medidas, los indicadores representan mejor la distribución típica del mundo empresarial y evitan que unos pocos casos extremos distorsionen las conclusiones generales.  
Así mismo para homogenizar las métricas y prepararlas para análisis comparativos, los indicadores fueron estandarizados mediante z-score (StandardScaler), de forma que cada variable tiene media cero y desviación estándar uno. Esto facilita la interpretación de correlaciones y comparaciones entre ratios de distinta magnitud.  
**Análisis exploratorio de datos**Los indicadores financieros fueron modelados en forma de histograma y heatmap con el fin de analizar de forma general como se relacionan los indicadores con el fin de hallar cuales son relevantes para el caso de estudio.  
Histograma  


El análisis de las distribuciones de los indicadores financieros revela la gran heterogeneidad de la estructura empresarial colombiana.

La Liquidez General presenta una distribución fuertemente sesgada hacia la derecha: la mayoría de empresas se ubican en valores entre 1 y 3, lo que refleja una capacidad razonable de cubrir obligaciones de corto plazo; sin embargo, existen empresas con liquidez extremadamente alta, que más que indicar fortaleza pueden estar reflejando una ineficiencia en la utilización de activos corrientes, por el contrario, las compañías con liquidez por debajo de 1 constituyen un grupo vulnerable, pues carecen de recursos suficientes para responder a sus pasivos inmediatos.

El Apalancamiento, medido como la proporción de deuda sobre activos, muestra una concentración en valores intermedios, lo que denota una estructura de financiamiento relativamente equilibrada; sin embargo, aquellas empresas con niveles superiores al 0.7 exhiben un riesgo incrementado de insolvencia al depender excesivamente del crédito.

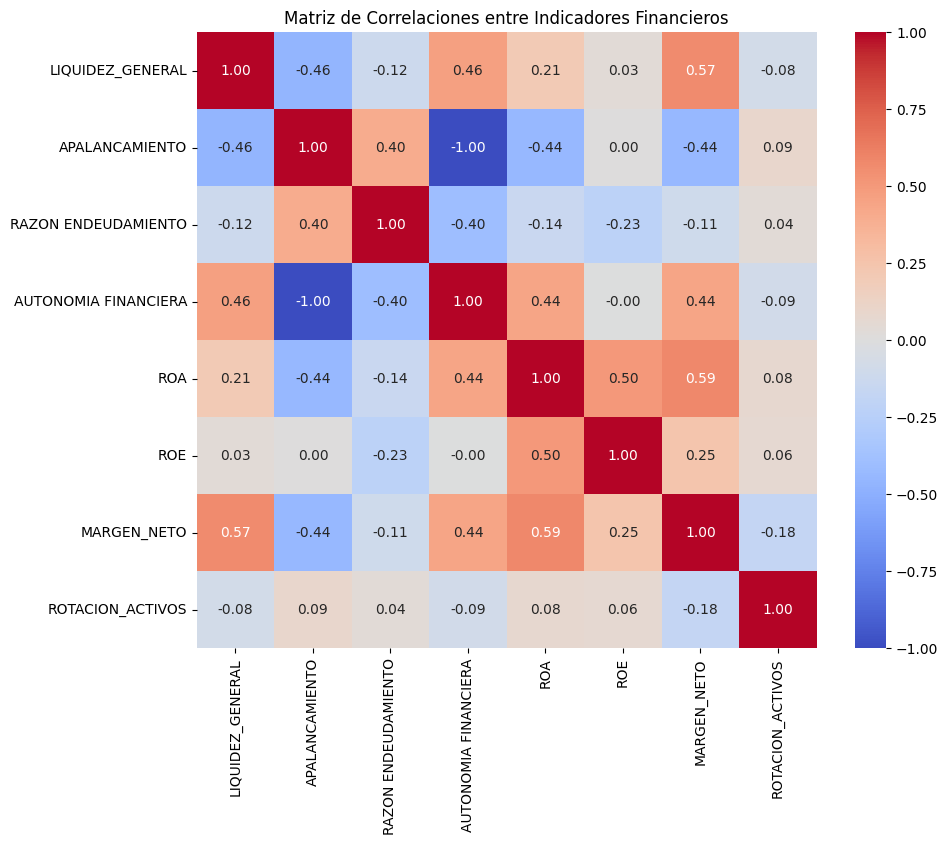
La Razón de Endeudamiento, definida como pasivos sobre patrimonio, evidencia casos extremos de compañías con valores muy altos e incluso negativos, estos últimos asociados a patrimonios negativos, ambos escenarios constituyen señales críticas de fragilidad y riesgo de quiebra.

La Autonomía Financiera, por su parte, mantiene un comportamiento complementario al apalancamiento, situándose alrededor del 0.45 en promedio, valores bajos que reflejan dependencia de acreedores, lo cual aumenta la vulnerabilidad frente a crisis financieras.

En términos de rentabilidad, tanto el ROA como el ROE muestran distribuciones sesgadas hacia la derecha, donde la mayoría de las empresas obtienen utilidades modestas en relación con sus activos y patrimonio, y solo una minoría logra retornos muy altos, los casos negativos en ambos indicadores son preocupantes, pues evidencian pérdidas operativas que, sumadas a altos niveles de deuda, elevan de manera significativa la probabilidad de insolvencia.

El Margen Neto refuerza esta visión ya que la mayoría de las empresas operan con márgenes reducidos, lo cual implica que cualquier choque de costos o caída en ingresos podría empujarlas rápidamente a terreno deficitario.

La Rotación de Activos refleja fuertes diferencias sectoriales: mientras que muchas compañías presentan niveles bajos a moderados de eficiencia en el uso de sus activos, otras exhiben rotaciones muy altas, típicas del comercio minorista, más, sin embargo, una alta rotación no siempre se traduce en salud financiera si los márgenes son estrechos, por lo que este indicador debe analizarse en conjunto con el margen y la rentabilidad.

Este primer análisis exploratorio muestra que, si bien la mayoría de las empresas operan en rangos financieros relativamente saludables, existe un subconjunto con valores extremos y combinaciones peligrosas como baja liquidez, endeudamiento elevado, patrimonio negativo o rentabilidad negativa que constituyen claros focos de riesgo de insolvencia.  
**Heatmap**  


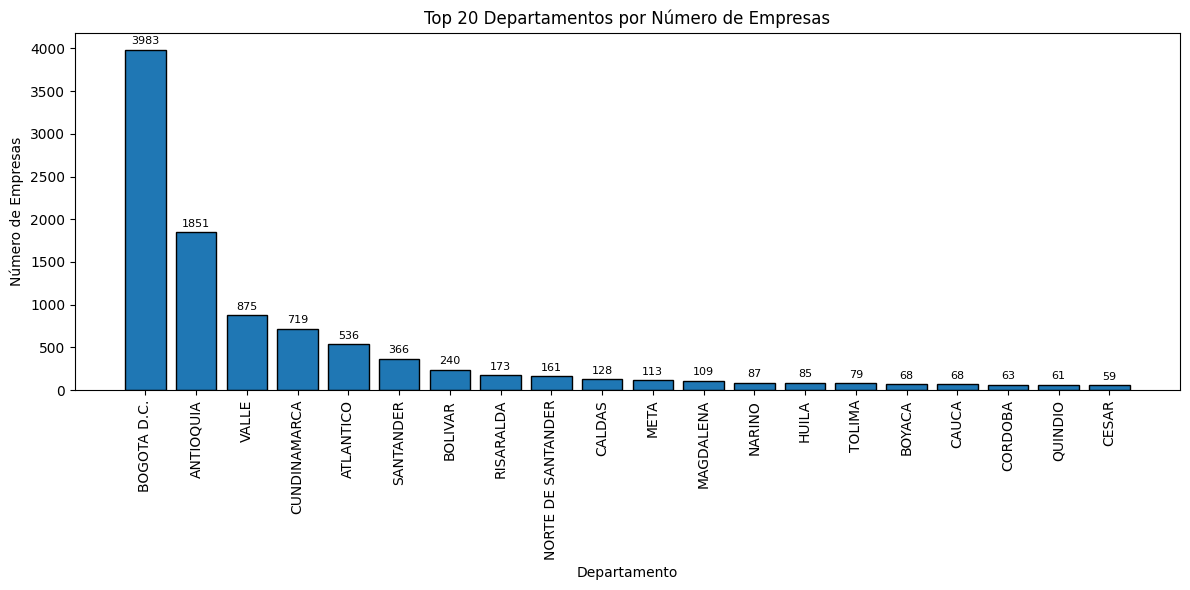
La matriz de correlaciones entre indicadores financieros nos ofrece una visión clara de cómo interactúan los distintos componentes de la estructura empresarial.

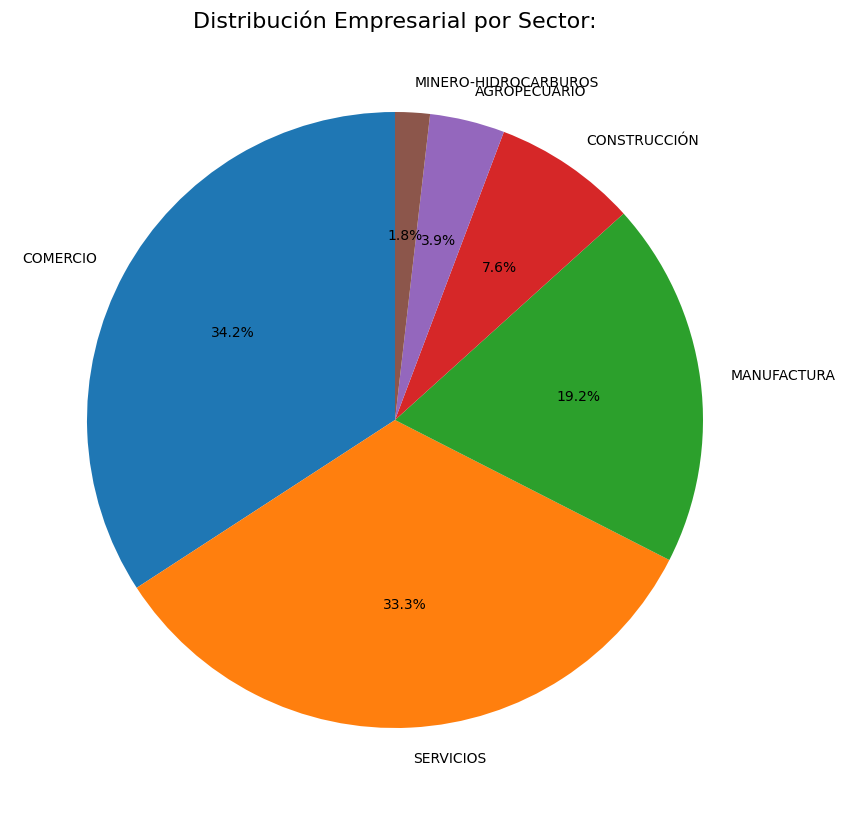
El hallazgo más evidente es la correlación negativa casi perfecta entre Apalancamiento y Autonomía Financiera, lo que responde a su relación contable directa ya que a mayor peso de la deuda, menor participación del patrimonio en los activos, este resultado es importante metodológicamente, pues sugiere que no conviene incluir ambas variables de manera simultánea en modelos predictivos debido a problemas de colinealidad, otro resultado relevante es la correlación positiva entre Liquidez y Margen Neto ya que las empresas que generan mayores márgenes suelen mantener posiciones de liquidez más holgadas, lo que las hace más resilientes frente a shocks, en contraste, compañías con liquidez ajustada tienden a operar con rentabilidades limitadas, aumentando su vulnerabilidad, asimismo, la correlación negativa entre Apalancamiento y ROA refleja que el mayor uso de deuda no siempre se traduce en mayor eficiencia en el uso de activos y por el contrario, muchas veces el peso del servicio financiero erosiona las utilidades. En la misma línea, la relación positiva entre ROA, ROE y Margen Neto confirma que las empresas con operaciones eficientes logran convertir ingresos en utilidades y, a la vez, generar retornos atractivos para los accionistas. Finalmente, la Rotación de Activos muestra correlaciones débiles con el resto de indicadores, lo que sugiere que su efecto depende más de la estructura sectorial que de la solvencia financiera en sí misma.

En conjunto, la matriz de correlaciones refuerza la idea de que la insolvencia no se explica por un solo indicador, sino por la combinación de varios factores como el alto apalancamiento junto con baja rentabilidad y márgenes negativos constituye la señal más clara de fragilidad empresarial.  
Con ambos gráficos encontramos variables candidatas a formar señales de alerta como el apalancamiento (o autonomía), Margen Neto, Liquidez, ROA, esto dado que la matriz respalda que estas variables se relacionan con el estatus financiero y entre sí de manera interpretable. Así mismo, dada su alta correlación se evitará usar apalancamiento y autonomía simultáneamente.

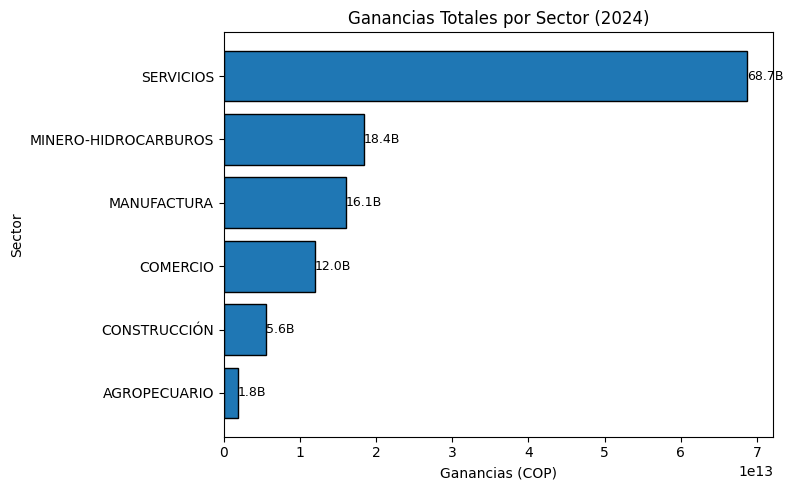
Por otro lado, la simultaneidad de apalancamiento alto y baja rentabilidad (correlación negativa) nos sugiere que las firmas con deuda elevada no están generando retorno suficiente, por tanto, probabilidad de insolvencia mayor.

Ahora para el análisis de variables no numéricas:

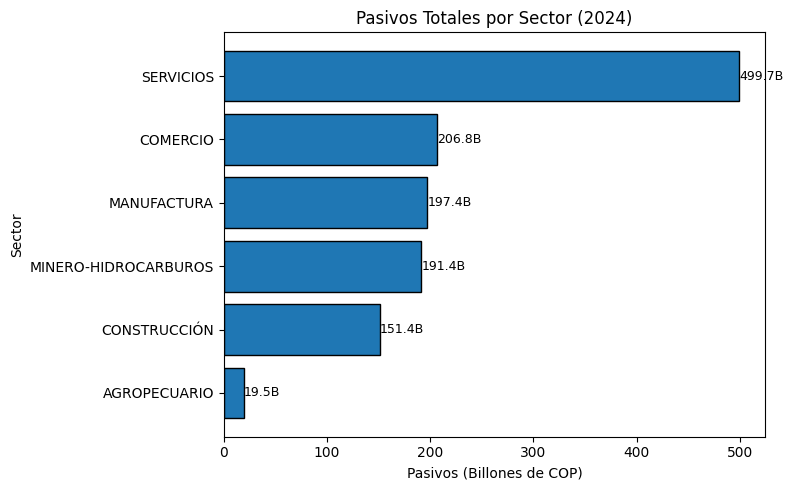
  
La distribución territorial de las empresas evidencia una alta concentración geográfica: Bogotá D.C. reúne casi el 40% del total de firmas analizadas (3.983), seguida de Antioquia (18,5%) y Valle del Cauca (8,7%). En conjunto, estos tres departamentos concentran cerca de dos tercios del tejido empresarial del país mientras el resto de departamentos muestran participaciones mucho menores, lo que refleja la centralización de la actividad económica y empresarial en unos pocos polos regionales, esto implica que los riesgos de insolvencia no solo tienen un impacto empresarial, sino también regional es decir que una crisis localizada en Bogotá o Antioquia podría generar efectos desproporcionados en el sistema económico nacional.



El análisis sectorial revela que Comercio (34%) y Servicios (33%) agrupan dos tercios de las empresas más importantes del país, les siguen la Manufactura (19%), la Construcción (7,6%), el sector Agropecuario (3,9%) y Minería e Hidrocarburos (1,8%). Esta composición muestra un modelo productivo dominado por actividades de intermediación y servicios, con un peso relativamente bajo de la industria transformadora y del agro, esto es relevante dado que esta estructura radica en que sectores como el comercio suelen operar con márgenes reducidos, mientras que otros, como minería y manufactura, generan mayor rentabilidad relativa.



En términos de utilidades, la mayor concentración se observa en el sector Servicios, que acumula ganancias por más de 68,7 billones de COP en 2024, equivalente a la mayor parte del total nacional. Le siguen Minería e Hidrocarburos (18,4B), Manufactura (16,1B) y Comercio (12B). Construcción (5,6B) y Agropecuario (1,8B) aportan en menor medida. Este panorama revela que, aunque el comercio concentra la mayor cantidad de empresas, su contribución a la rentabilidad total es relativamente baja, confirmando que el tamaño no siempre se traduce en eficiencia, en contraste, sectores intensivos en capital como minería y servicios logran transformar mejor sus ingresos en utilidades.



El análisis de pasivos muestra una realidad complementaria: el sector Servicios concentra también el mayor nivel de endeudamiento (499,7B), seguido por Comercio (206,8B), Manufactura (197,4B), Minería e Hidrocarburos (191,4B) y Construcción (151,4B) y el Agropecuario presenta un nivel bajo de deuda (19,5B). Este resultado implica que, aunque los servicios generan grandes utilidades, también cargan con la mayor exposición financiera, por su parte, la construcción muestra un alto nivel de pasivos en comparación con sus utilidades, lo que la convierte en un sector especialmente vulnerable a crisis de liquidez.

La evidencia territorial y sectorial manifiesta que la estructura empresarial colombiana está altamente concentrada, tanto en unos pocos departamentos como en sectores específicos. Bogotá, Antioquia y Valle funcionan como polos de atracción de empresas, mientras que comercio y servicios dominan la composición sectorial, más sin embargo, al comparar ganancias y pasivos, emergen riesgos diferenciados ya que el comercio, pese a ser el sector con más empresas, opera con márgenes bajos, la construcción exhibe una alta carga de deuda con utilidades reducidas, y el agro mantiene un rol marginal en tamaño y rentabilidad, en contraste, servicios y minería destacan por su elevada generación de utilidades, aunque acompañadas de altos niveles de endeudamiento, por lo que en conjunto, estos resultados sugieren que los riesgos de insolvencia se concentran en sectores de baja rentabilidad o alto apalancamiento, mientras que la estabilidad del sistema depende en gran medida de la solidez de los servicios y de las actividades minero–industriales.

**BIBLIOGRAFIA  
  
Superintendencia de Sociedades. (2025, 26 de junio).** *Las 1 000 empresas más grandes del país consolidan su peso en la economía nacional* [Informe]. Recuperado de <https://www.supersociedades.gov.co/noticias-supersociedades/-/asset_publisher/atwl/content/las-1.000-empresas-m%C3%A1s-grandes-del-pa%C3%ADs-consolidan-su-peso-en-la-econom%C3%ADa-nacional>