



Colegio de Estudios
Superiores de Administración

**Evaluación de variables determinantes del capital de trabajo en medianas empresas
de Bogotá**

Daniel Sebastián Aguilera Garzón

Maestría en Finanzas Corporativas

Colegio de Estudios Superiores de Administración

Bogotá, D.C., Colombia

2024

**Evaluación de variables determinantes del capital de trabajo en medianas empresas
de Bogotá**

Daniel Sebastián Aguilera Garzón

Director

Julio Alejandro Sarmiento Sabogal

Maestría en Finanzas Corporativas

Colegio de Estudios Superiores de Administración

Bogotá, D.C., Colombia

2024

Tabla de Contenido

Resumen.....	6
1. Introducción.....	7
1.1. Objetivo General.....	10
1.2. Objetivos Específicos.....	10
2. Estado del Arte.....	12
3. Base de Datos.....	18
3.1. Definición de variables.....	18
3.2. Recopilación de Datos.....	19
3.3. Preparación de datos.....	24
3.3.1. Pruebas de Distribución.....	26
3.3.1.1. Histogramas.....	26
3.3.1.2. Sesgo y Curtosis.....	29
3.3.1.3. Shapiro-Wilk.....	31
3.3.1.4. Kolmogorov-Smirnov.....	34
4. Metodología.....	38
4.1.1. Pruebas de Multicolinealidad.....	39
4.1.1.1. Correlación de Variables.....	40
4.1.1.2. VIF (Factor de Inflación de la Varianza) y Tolerancia.....	45
4.1.1.3. Número de Condición.....	50

4.1.2. Transformación de Datos.....	51
5. Variable dependiente y selección de variables independientes (transf.).....	56
5.1.1. Presentación de la variable dependiente.....	56
5.1.2. Presentación de las variables independientes.....	57
5.1.2.1. Análisis luego de transformación.....	57
5.1.2.2. Selección de las variables independientes.....	58
5.1.2.3. Estadística Descriptiva.....	60
6. Resultados.....	63
6.1.1. Cálculo Iterativo de selección de variables.....	63
6.1.2. Regresión Lineal Múltiple.....	64
6.1.3. Bondad de ajuste del modelo.....	67
6.1.4. Pruebas de hipótesis.....	69
7. Conclusiones.....	74
Referencias.....	78

Lista de Tablas

Gráfica 1 - Fenómenos Financieros en Pequeñas y Medianas empresas.....	15
Tabla 1 - Limpieza e imputación de los datos.....	25
Gráfico 2 - Histogramas Datos Recopilados.....	26
Tabla 2 - Cálculos y Pruebas de Distribución.....	28
Tabla 3 - Análisis de sesgo y curtosis.....	31
Tabla 4 - Prueba de Shapiro.....	33
Tabla 5 - Kolmogorov-Smirnov.....	36
Gráfico 3 - Mapa de Calor Matriz de Correlación.....	41
Tabla 6 - Matriz de Correlación con la Dependiente.....	43
Tabla 7 - Vif y Tolerancia.....	48
Tabla 8 - Transformación Yeo-Johnson.....	53
Tabla 9 - Sesgo y Curtosis (transformados).....	54
Tabla 11 - Kolmogorov-Smirnov (transformados).....	54
Tabla 11 - Variables Seleccionadas.....	59
Tabla 12 - Estadísticas Descriptivas.....	60
Tabla 13 - Regresión Lineal Múltiple.....	65
Tabla 16 - Bondad de ajuste Ridge - Lasso.....	68
Tabla 18.1 - Prueba de Hipótesis mediante Prueba T y F.....	70
Tabla 18.2 - Prueba de Hipótesis mediante Prueba T y F.....	72

Resumen

El presente trabajo de grado, aborda el problema de la evaluación de variables determinantes del capital de trabajo en medianas empresas de Bogotá, esto, a través de un modelo estadístico detallado.

Este documento analiza la relación entre la gestión efectiva del capital de trabajo y la rentabilidad y liquidez de las empresas, empleando variables como los ciclos de conversión de efectivo y las políticas de manejo de inventarios. Así mismo, el modelo planteado examina el impacto de la gestión del capital de trabajo en la creación de valor para los accionistas y la estabilidad financiera de las empresas.

Para desarrollar este estudio, se utilizaron datos de fuentes oficiales como la Superintendencia de Sociedades y la Cámara de Comercio de Bogotá, así como el cálculo de ratios relevantes para la medición del desempeño de las empresas; todas evaluadas a un corte del cierre fiscal año 2022.

Esta investigación, mediante paquetes y bibliotecas estadísticas, desplegadas en Python, empleó técnicas avanzadas de regresión lineal múltiple y modelado predictivo para evaluar el impacto de la gestión del capital de trabajo basado en indicadores financieros.

A través del modelo propuesto, basado en datos empíricos, se demostró con un soporte sólido que una correcta evaluación de variables determinantes en el capital de trabajo conduce a mejoras significativas en la eficiencia; y que existen indicadores que pudieran ser implementados en planes de mejora para aumentar la misma gestión.

1. Introducción

En América Latina más del 90% de las empresas son Pymes (Fajardo, 2017), las cuales, a pesar de ser principal fuente de trabajo en estas economías, difícilmente alcanzan una etapa de madurez y/o eficiencia productiva; lo que a su vez se traduce en el tamaño al que logran llegar, en donde respecto a su capacidad instalada, solo el 30% se convierte en mediana empresa, independientemente del sector al que pertenezcan (ANIF, 2021).

Esta ineficiencia no es ajena para Colombia, pues tal como lo ha venido reflejando la Gran Encuesta Pyme, realizada por ANIF, en los últimos años se ha demostrado que este subdesarrollo industrial es consecuencia de las débiles estructuras financieras o administrativas al interior de las empresas, las que con dificultad logran hacer un correcto aprovechamiento de los recursos que se tienen disponibles, y que al mismo tiempo estas firmas carecen de alternativas económicas para su correcta operación (2019).

Así mismo, ANIF (2019) también ha identificado en las últimas encuestas que el aporte al producto interno bruto por parte de las Pymes en el país podría ser mayor, pero debido a que gran parte este tejido empresarial, que está formado por empresas que han nacido como negocios familiares, carecen de procesos eficientes. Tanto así, que para su último volumen, ANIF (2021) identificó que menos del 10% de las Pymes de la región realizan actividades de exportación, reflejo de la dificultad productiva, industrial y comercial a las que se ven enfrentadas.

De hecho, si se mira desde una perspectiva de región internacional, “en Brasil, más del 51% de las 200 empresas más grandes que cotizan en bolsa, son empresas gobernadas por familias” (Corporación Financiera Internacional, 2010). Este es un claro ejemplo de cómo

se encuentra conformada la industria en América latina, la cual tiene una estructura corporativa basada en la familiaridad.

Situación que, como ya ha mencionado ANIF en encuestas previas, aquellas firmas que cumplan con las características de trayectoria, ventas y tamaño, suelen ser empresas familiares, con la intención de llevar a cabo una expansión de su negocio, liderada por una gerencia de familia.

Lo anterior es relevante debido a que la mayoría de compañías en esta estructura carecen de estructuras o procedimientos que faciliten una expansión en el largo plazo. A su vez, para este ejercicio investigativo, tal como presenta Damodaran, A. (2009) en su documento *Valuing Young, Start-up and Growth Companies: Estimation Issues and Valuation Challenges*, uno de los problemas existentes en la valoración de compañías de este tipo reside principalmente en la obtención de información histórica suficiente y pertinente, con el fin de establecer resultados significativos.

Como consecuencia de este fenómeno, tal como evidencia el termómetro de la ANIF, la Gran Encuesta Pyme o GEP, uno de los factores más críticos que enfrentan las empresas en el país es la gestión del capital para atender sus obligaciones, tanto financieras como operativas; las cuales, no logran ser suplidas por la generación de recursos económicos obtenidos por las compañías (2019), y que en muchas ocasiones y a pesar de haber presentado encontrado mejoras y reducciones en los costos operativos, sus márgenes resultantes son deficitarios y con poca generación de valor. (ANIF, 2019)

Por lo cual, con el fin de abordar data significativa para la investigación y presentación de este documento, se utilizaron fuentes oficiales para el estudio de la situación empresarial en

el país, tales como superintendencias, cámaras de comercio y entidades de control y supervisión¹

Con ello, se expone la necesidad de tener estadísticas para establecer oportunamente mecanismos que optimicen los recursos de las empresas sin destruir el potencial valor de una firma, puesto que como lo plantea Jensen, M.C. (2002) en su artículo Value Maximization, Stakeholder Theory, and the Corporate Objective Function, las necesidades corporativas deben encontrar un balance que maximice esfuerzos y resultados.

Ante la situación presentada, tal como lo define Correa, J.S., & Murillo, J. (2015), manteniendo como eje central la actividad investigativa, este documento tiene como fin establecer si es posible determinar un modelo estadístico que permita evaluar un conjunto de variables determinantes sobre el comportamiento del capital de trabajo de las empresas, siguiendo las mejores prácticas establecidas en las teorías planteadas por Jensen, M.C. (2002), Dyreng, Mayew, & Schipper (2017) y Coase (1937); y usando las metodologías planteadas por Damodaran, A. (2007), (2009), Chen & Dodd (1997).

1.1. Objetivo General

El objetivo general del presente documento es, establecer si es posible determinar un modelo estadístico que evaluar un conjunto de variables que sean determinantes para pronosticar un valor racional del capital de trabajo en medianas empresas de Bogotá, partiendo de la selección de varias variables, siguiendo las mejores prácticas establecidas

¹ Fuentes oficiales de consulta a través del sistema EMIS. Superintendencia de Sociedades, Superintendencia Financiera, Superintendencia de Salud, Superintendencia de Transporte, Cámara de Comercio de Bogotá, y Contraloría General de la Nación.

en las teorías planteadas por diversos autores y utilizando metodologías específicas para la evaluación y gestión del capital de trabajo.

1.2. Objetivos Específicos

- 1.2.1. Identificar los principales elementos financieros, de la data recopilada, que influyen en la gestión del capital de trabajo en medianas empresas. Este objetivo busca determinar cuáles son las variables específicas tienen un impacto significativo en el capital de trabajo.
- 1.2.2. Analizar cómo la gestión del capital de trabajo afecta la rentabilidad y la liquidez de las medianas empresas en Bogotá. Este documento de investigación intenta establecer una correlación clara entre prácticas eficientes de gestión del capital de trabajo y mejoras en indicadores financieros.
- 1.2.3. Desarrollar un modelo estadístico capaz de predecir y mejorar la gestión del capital de trabajo. El modelo estimado deberá evaluar y estimar las variables determinantes, cuya significancia necesaria, basado en el análisis estadístico, puedan responder y ser utilizadas por los gerentes financieros para pronosticar un valor estimado del capital de trabajo, basándose en datos históricos y tendencias del sector.

2. Estado del Arte

En Bogotá, existen más de 3000 empresas que están en la capacidad de generar ingresos anuales entre 10.000 y 30.000 millones de pesos anuales, donde el percentil 70 de la población puede generar más de 19.000 millones al año, de las cuales tienen una constitución de hasta 23 años. Sin embargo, y explicado por los costos de agencia de Jensen M. y Meckling (1976), solo aquellas que superan el percentil 70 generan un aprovechamiento económico de los recursos por actividades propiamente operativas superior al 41%.

Los autores Jensen M. y Meckling (1976) explican mediante el costo de agencia como las asunciones temporales y permanentes de los gerentes y/o accionistas afectan esta estructura de capital, la cual puede darse en el origen de recursos propios o en la financiación a través de deuda. Esto puede verse explicado en que muchas de estas compañías se han visto expuestas en al menos un cambio generacional o un cambio de gerencia general; lo que conlleva a que la gestión y seguimiento de objetivos de crecimiento anteriormente planteados respondan idóneamente al core del negocio.,

Esto, tal como presenta ANIF (2019) en la GEP, trae como consecuencia costos temporales a los que se ven expuestas las medianas empresas, los cuales determinan en un muy corto periodo de tiempo grandes esfuerzos que atiendan a las cambiantes demandas de las diferentes industrias.

Las principales causas de una gestión pobre en el capital de trabajo se pueden evaluar bajo un ciclo de conversión del efectivo negativo (Jensen M. y Meckling, 1976), en el cual, es

necesario evaluar los tres elementos que componen esta gestión, rotación de inventarios, cuentas por cobrar y cuentas por pagar.

El ideal esperado para todas las empresas, sería un ciclo de conversión donde los días de cuentas por cobrar excedan a las cuentas por pagar, para que de esta manera, la operación estaría soportada mediante sus propios proveedores, o variantes de este, como los productos de banca como el factoring.

También, se deberían buscar eficiencias o integraciones verticales que impacten los costos, con el objetivo de disminuir la rotación de las cuentas por cobrar. Así como, definir un control de inventarios que permita gestionar de forma eficiente; disminuyendo la generación de desperdicios, pérdidas de inventario, inventarios vencidos, entre otros.

Con esto en mente, se evidencia que la capacidad de financiación de estas empresas requiere niveles de liquidez suficientemente robustos que puedan responder a la actividad establecida en el modelo de negocio, la cual pudiese ser de intermediación o de distribución entre la oferta y la demanda. Es por esto que, las fuentes de capitales para estas compañías deben permitir mantener niveles de rotación de acuerdo a las demandas establecidas, y al mismo tiempo, permitir el fondeo de la operación.

En consecuencia, lo que buscan muchas veces es recurrir a fuentes de fondeo, como capitalización propia, con el objetivo de extraer al máximo las utilidades de la inversión, y no llevar a cabo la debida diligencia de identificación de nuevas fuentes que mejoren la estructura de capital de la firma, esto puede evidenciarse en que el percentil 75 de la población tiene estructuras donde sus necesidades operativas representan más del 50% del capital utilizado, que a su vez representan el 70% de los flujos de capital generados.

Lo que genera como consecuencia que se deje en riesgo el capital de los asociados, enfrentándose a costos muy altos y se subutilizan las capacidades de tecnificación y crecimiento. (Baker, 2009).

Adicionalmente, las gerencias no buscan ofrecer evaluar nuevas alternativas de financiación puesto que, la gran mayoría de estas compañías se encuentran en etapas muy tempranas, lo que significa que la incidencia de los stakeholders suele ser excesiva y no buscan tampoco ofrecer medidas o métodos de recompensa por la buena gestión; en el caso de investigación se encuentra que solo una de las más de 2.900 empresas estudiadas es cotizante en el mercado local.

Con base en esto, y partiendo desde las teorías planteadas por Jensen, M.C. (2002), el objetivo corporativo de las compañías debería atender también a la maximización de la firma, estableciendo estructuras que mejoren la gerencia y toma de decisiones.

Tal como se mencionó anteriormente, parte de los fenómenos atípicos que se evidencian tanto en las pequeñas como medianas empresas es la dificultad que existe para generar valor, puesto que no solo depende de la disminución de sus costos (ANIF, 2019); esto mismo se presenta a continuación en la Gráfica 1:

Gráfica 1 - Fenómenos Financieros en Pequeñas y Medianas empresas

Gráfico 15. Costos por tamaño de empresa (Balance de respuestas)

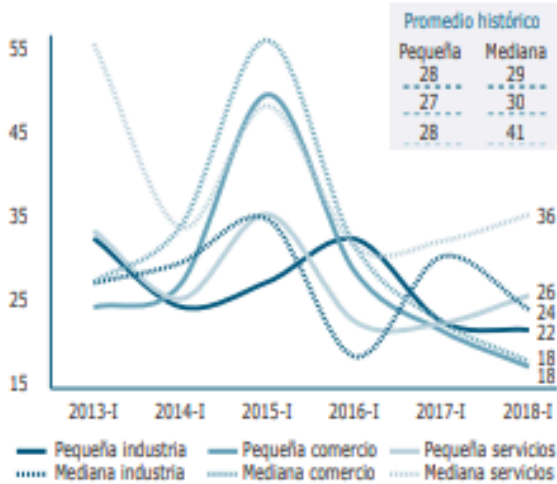


Gráfico 16. Márgenes (Balance de respuestas)

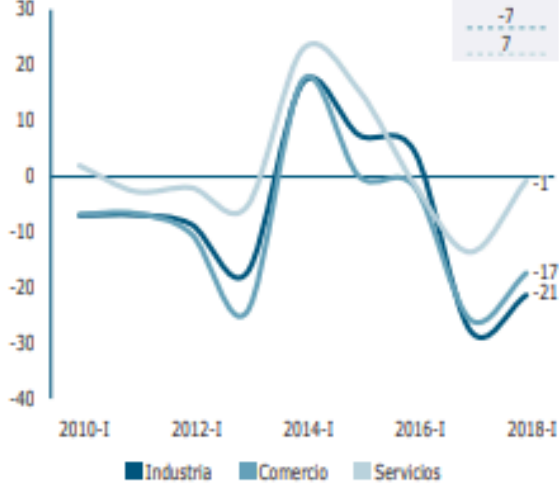


Gráfico 17. Márgenes por tamaño de empresa (Balance de respuestas)

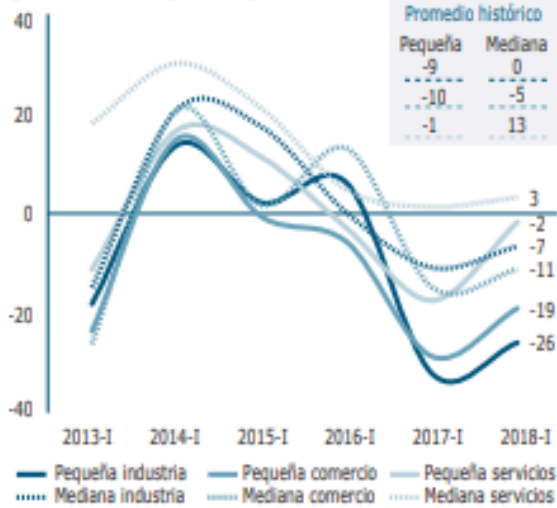
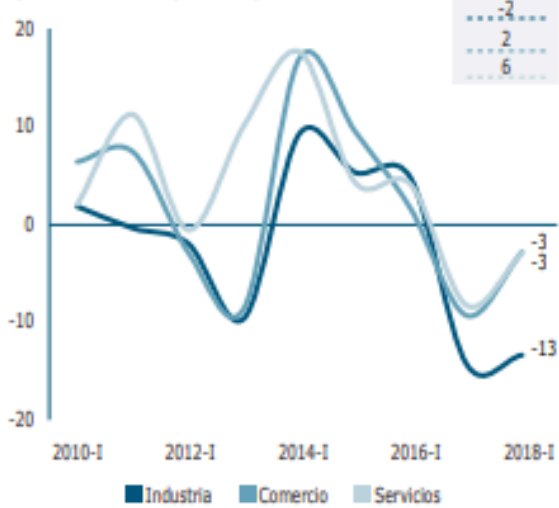


Gráfico 18. Empleo (Balance de respuestas)



Nota: Los gráficos 15-18 fueron adaptados de la revista ANIF (2019).

Para Jensen M. y Meckling, el uso de recursos será completamente diferente según su fuente y su gestor, en donde por una parte, para la fuente su interés principal es la maximización de utilidades mientras que para la gestión será la asegurar la operatividad del modelo de negocio (1976). Ya suponen bien los autores Trujillo, Guzmán y Prada al

referirse al escenario propuesto por Nguyen y Nielsen (2010) de suponer muertes súbitas para directores y gerentes de una compañía (Trujillo, Guzmán, & Prada, 2015), el cual, entre otras cosas, evidencia la sensibilidad que tiene la gestión y uso de recursos según los intereses de quienes dirigen las empresas.

Por otra parte, se estima que en América Latina menos del 30% de las medianas empresas tienen estructuras definidas de gobierno corporativo, que de una u otra manera eliminan el ruido ocasionado por la dualidad gerencial existente en estas empresas (Corporación Financiera Internacional, 2010); evidencia de esto se encuentra en que sólo 29 de las 2.958 empresas que hacen parte de la población de estudio tienen comités que evalúan la gobernabilidad de la firma, utilizando a su vez auditores expertos en el tema como Deloitte, KPMG, PwC o EY.

Es por esto que, según Jensen (2002) el objetivo que deberían establecer las compañías como fin último corporativo debería tanto propender la maximización del valor de la firma, como la debida consecución y utilización de los recursos disponibles.

Con base a lo anterior, la investigación continuará sobre los estudios realizados por el autor Damodaran, A. (2009) en cuanto se pretende explicar mediante el cálculo y medición de los ratios Return on Capital (ROC), Return on Invested Capital (ROIC) and Return on Equity (ROE), como, desde una perspectiva de gerencia financiera, pueden cometerse errores en la interpretación de lo que se considera como generación de valor; puesto que si bien puede no estarse viendo afectando el valor en libros de una firma, el valor de mercado de esta se puede ver perjudicado por errores en la clasificación de costos y gastos, disminuyendo las utilidades percibidas a costo de una equivocada disposición de los recursos.

Para esto, la investigación busca identificar aquellos elementos y decisiones contables o financieras de la industria que caigan en el error enunciado por Damodaran, A. (2009), con el objetivo de poder demostrar que bajo una estructura definida, y un modelo de gestión óptimo de capital de trabajo, se puede alcanzar la generación de valor de la firma.

Ejemplo de ello, ANIF ha identificado una mejora en el Indicador Pyme Anif o IPA, el cual mide los esfuerzos realizados por las empresas para obtener mejores resultados futuros; realizando actividades como capacitación y especialización de su personal en temas de gestión de recursos y decisión en el uso del capital de trabajo, adoptando medidas e indicadores al interior de las empresas para tal fin. (ANIF, 2017)

Lo anterior, es un punto de partida para evidenciar que existe un interés real en la implementación de nuevas y mejores prácticas que generen valor económico positivo para las firmas. Para esto, se propone junto con los estudios realizados por los autores Chen y Dodd, donde el objetivo principal busca demostrar que la generación de valor, medida por el ratio EVATM, resulta ser la manera más eficiente de comprobar el performance financiero de una empresa cuyas prácticas, estructuras decisorias y políticas de gestión de dinero favorecen a la firma (Economic Value Added (EVATM): An Empirical Examination Of A New Corporate Performance Measure, 1997).

3. Base de Datos

3.1. Definición de variables

Greene, W. H. (2018) plantea que la utilización de variables categóricas y continuas le otorga al modelo la posibilidad de estimar relaciones no lineales de forma flexible, lo cual no podrían si se usaran modelos tradicionales. Del mismo modo, se busca mediante la utilización de estas variables mejorar la interpretación de la data disponible, especialmente cuando se trata de explicar en un modelo de regresión lineal múltiple patrones o tendencias en poblaciones de estudio con interacciones complejas (Agresti, 2012).

Esto significa que, dentro del modelo a probar en este documento, para poder trabajar con variables continuas y categóricas, es necesario evaluar una a una las variables respecto a su relación con la variable dependiente y reducir mediante métodos no paramétricos o de aprendizaje automático la normalidad y linealidad de los mismos (Hastie, T., Tibshirani, R., & Friedman, J., 2017)

Con lo anterior, para la selección de las variables de estudio, se incluyeron los parámetros establecidos por Coase (1937), Damodaran (2009), Dyreng, Mayew, & Schipper (2017), y Walbert (1994). También se tuvo en cuenta las implicaciones identificadas por Nourayi y Daro (1996), en las cuales determina que para la debida estimación y evaluación de performance de las empresas se debe implementar el cálculo de la métrica EVATM de los estudios de Chen y Dodd (1997); así como los ratios utilizados comúnmente para el cálculo del mismo.

Con base en esto, y partiendo desde los supuestos y estrategias definidas por Sampieri (2018), para el desarrollo de la metodología cuantitativa se ha construido una base de datos,

se incluyen dentro de la base las variables de estudio elementos como: ítems del balance, estado de resultados, ratios calculados, elementos cuantitativos de la operación como número de empleados, CIU NAICS, entre otros; los cuales resultan ser 105 variables iniciales para el estudio.

3.2. Recopilación de Datos

En consecuencia, lo que se busca dentro de esta investigación es identificar si estas estructuras deberían ser adoptadas por las pequeñas y medianas empresas de Bogotá, y de esta manera medir eficientemente su actividad por indicadores relevantes para el crecimiento de la compañía.

Por ello, desde los planteamientos establecidos por Damodaran, A. (2009) y Dyreng, Mayew, & Schipper (2017) se requiere ver a grandes rasgos la información disponible de las compañías, evitando caer en los sesgos de información que estos pudiesen tener.

Para esto, dentro de las actividades investigativas a realizar se ahondó sobre las estructuras decisorias establecidas en las empresas de mediano tamaño, teniendo en cuenta únicamente las empresas ubicadas en la ciudad de Bogotá, que presentarán cifras fiscales oficiales, con cierre fiscal del año 2022; adicionalmente, que los estados financieros auditados estén consolidados, preferiblemente.

Junto con esto, se utilizaron dos criterios de observación que permitiera estudiar únicamente a medianas empresas. Para el primer criterio, de acuerdo con la Ley 1450 de 2011, Artículo 43, se le confirió la facultad de establecer las definiciones de tamaño empresarial al Gobierno de la República (Congreso de la República de Colombia, 2011), el cual, el Ministerio de Comercio, Industria y Turismo - MinTIC (s.f.) reglamentó que el

tamaño de una mediana empresa serían aquellas con activos totales entre 5.001 y 50.000 salarios mínimos mensuales legales vigentes (SMMLV), y con un número de empleados desde 51 hasta 200. Este primer universo observado arrojó que a cierre de 2022 existían 13.879 medianas empresas que podrían ser objeto de estudio.

El segundo criterio, con el fin de enfocar aún más la población de estudio, alineados con la visión de a2censo (Bolsa de Valores de Colombia, 2023) en la definición de empresas productivas, se implementó un criterio en términos de resultados, definiendo a que son empresas productivas aquellas que logran ingresos operativos totales anuales entre 900 y 50.000 millones de pesos. Aplicando los filtros anteriores mencionados, este universo observado presentó 8.717 medianas empresas que podrían ser objeto de estudio.

Finalmente, se estableció que el universo de estudio buscaría a aquellas empresas que tuviesen hasta 250 empleados contratados directamente; esta definición se hace partiendo desde lo definido por Thibeault, Buijink, & Joos, en cuanto al tamaño de nómina contratada internacionalmente utilizado para la categorización de medianas y pequeñas empresas (2009). Ampliar de 200 empleados, reglamentado por MinTIC(s.f.) a 250, como lo presenta Thibeault, Buijink, & Joos, parte de la intención de definir como criterio una práctica utilizada a nivel internacional, puesto que aplicar la definición propuesta por MinTIC (s.f.) cerraría el universo de estudio a menos de 1.000 empresas, mientras que el criterio escogido permite que una mediana empresa pueda tener una nómina inferior a los 51 recursos humanos siempre y cuando cumpla con un mínimo de activos totales y un mínimo de ingresos operativos.

Bajo estas definiciones, aplicando los criterios de empresas que presentaran cifras fiscales oficiales en fuentes validadas, con cierre fiscal del año 2022, con estados financieros

auditados, que preferiblemente están consolidados, que cuenten con un total activos entre 5.001 y 50.000 SMMLV, con ingresos operativos totales entre 900 y 50.000 millones, y con hasta un máximo de 250 empleados; el escenario de estudio acotó el universo observado a una población de 1.180 medianas empresas.

Las fuentes de consulta para la construcción de los datos iniciales, se realizó a través del sistema EMIS; el cual, recupera y consolida la información oficial que presentan las empresas en entidades como. Superintendencia de Sociedades, Superintendencia Financiera, Superintendencia de Salud, Superintendencia de Transporte y Cámara de Comercio de Bogotá.

Sobre estas 1180 empresas, se recopilaron 106 indicadores, razones financieras, y elementos de los estados financieros reportados; los cuales serán referenciados en adelante como las variables de estudio. Estas fueron:

- CapitalDeTrabajo
- NAICS
- CIU
- TotalIngresoOperativo
- ActivosTotales
- GananciaoperativaEBIT
- Gananciasantesdeimpuestos
- Impuestoalarenta
- Gananciasdespuesdeimpuestos
- Gananciadeoperacionesdiscontinuas
- GananciaPerdidaNeta
- Propiedadplantayequipo
- Activosintangiblesyvalorllave

- Efectivo o Equivalentes
- Total de patrimonio
- Patrimonio neto atribuible a los propietarios de la controladora
- Capital suscrito
- Resultados acumulados
- Pasivos Totales
- Flujo neto de efectivo por utilizados en actividades de explotación
- Flujo neto de efectivo de utilizadas en actividades de inversión
- Flujo neto de efectivo de utilizados en actividades de financiación
- Aumento o disminución neto de efectivo y equivalentes de efectivo
- Efectivo al inicio del periodo
- Efectivo al final del periodo
- Rendimiento sobre Activos Anualizado
- Rendimiento sobre el Patrimonio Anualizado
- Rendimiento sobre el Capital Empleado
- Margen Neto
- Margen de Ganancia Bruta
- Margen Operacional
- Margen Ebitda
- ROA Operativo
- Rotación de Inventario
- Rotación de Cuentas por Cobrar
- Rotación del Activo Corriente
- Rotación del Activo No Corriente
- Rotación de Activos
- Rotación de Cuentas por Pagar
- Rotación del Capital de Trabajo
- Valor Contable

- EfectivoNeto
- DeudaEMIS
- DeudaALargoPlazo
- DeudaACortoPlazo
- DeudaNeta
- CapitalEmpleado
- RazonDeLiquidez
- Pruebaacida
- RazonDeEfectivo
- CoeficienteDeEfectivo
- RelacionDeFlujoDeEfectivoOperativo
- RelacionDeudaActivosTotales
- RelacionDeudaCapital
- RelacionDeudaLargoPlazoCapitalEmpleado
- RelacionDeudaEbitda
- RelacionFlujoDeCajaDeuda
- RelacionActivosPatrimonio
- TendenciaDeLosIngresosNetos
- TendenciaDeLosIngresosOperacionales
- TendenciaDeLaUtilidadBruta
- TendenciaDelEbitda
- TendenciaDeLaUtilidadOperacional
- TendenciaDeLaUtilidadNeta
- TendenciaCuentasPorCobrar
- TendenciaDelInventario
- TendenciaPropiedadPlantaYEquipoNetoPP&E
- TendenciaDeLosActivosTotales
- TendenciaDelValorContable

- TendenciaDelPatrimonioNeto
- TendenciaDelFlujoDeCajaOperativo
- TendenciaGastosDeCapital
- CoberturaDeLaDeudax
- RazonFlujoDeCajaOperativoIngresos
- RazonFlujoDeCajaOperativoActivos
- RazonFlujoDeCajaOperativoPatrimonio
- RazonFlujoDeCajaOperativoUtilidadOperacional
- RelacionEfectivoTotalActivos
- RelacionCuentasPorCobrarActivosTotales
- RelacionInventarioTotalActivos
- RelacionActivofijoTotalactivo
- RelacionPasivoCorrientePasivototal
- RelacionGastosAdministrativosVentas
- RelacionImpuestoALaRentaVentas
- RelacionFlujoDeCajaOperativoFlujoDeCajaTotal
- RelacionFlujoDeEfectivoDeInversionFlujoDeEfectivoTotal
- RelacionFlujoDeEfectivoDeFinanciacionFlujoDeEfectivoTotal
- ModeloZScoredeAltman
- Numerodeempleados
- ActivosCorrientes
- PasivosCorrientes
- ValordeMercadodelEquity
- CostodeCapitalPropio
- EBIT
- GAIMP
- InteresesPagados
- CostoDeuda

- TasadeImpuesto
- DeudaTotal
- WACC
- NOPAT
- CapitalInvertido
- EVA
- ROA
- ROE

En este sentido, el objeto de investigación de este documento utilizará estas variables como aquellas a evaluar para validar si es posible determinar un modelo estadístico que permita pronosticar y superar barreras deficitarias en la gestión del capital de trabajo; las cuales, estas deberían estar alineadas con los intereses de los stakeholders de las firmas (Dyrenge, Mayew, & Schipper, 2017).

3.3. Preparación de datos.

Ahora bien, de la base construida, se requiere la preparación de la data, enfocándonos en la depuración de aquellas variables significativas y relevantes para el correcto análisis financiero (Vance, D. E., 2002)., para luego ser transformada y poder aplicar correctamente técnicas de estadística descriptiva para la evaluación del desempeño del modelo.

Para esto, partiremos de la metodología de limpieza de datos de Rahm, Erhard & Do, Hong. (2000), en la cual buscamos eliminar la mayor cantidad de inconsistencias y errores presentes en la data, los cuales puedan comprometer y sesgar los resultados del modelo. Siguiendo el uso de la biblioteca propuesta por McKinney (2022), se procedió a utilizar la biblioteca de Pandas para mostrar todas las columnas sin truncar; seguido por la

identificación de variables faltantes para poderlas imputar y tener certeza de la idoneidad de la fuente.

En la tabla 1 se presenta el resultado de la limpieza e imputación de los datos, que mediante el uso de clases como `clean_and_convert`, se verifica si el tipo de datos por columna del Dataframe es un objeto o valor. En el caso de identificar faltantes, el código imputa estos vacíos con la media, iterada por todas las columnas del DataFrame.

Tabla 1 - Limpieza e imputación de los datos.

```
Primeras filas de los datos:
   NOPAT  CapitalInvertido ...      EVA   ROA   ROE
0  2175.37      47938.33 ...    -4053.71  3.52  22.41
1  3579.22      17008.61 ...    -4298.84  21.24  42.56
2  3946.08      25171.55 ...    -2075.40  15.68  29.07
3  1440.34      16068.38 ...     -456.58  2.67  15.43
4  3984.49      35901.13 ...    -5588.78  12.48  16.63

Valores faltantes en cada columna:
CapitalDeTrabajo          0
NAICS                     0
TotalIngresoOperativo    0
...
WACC                      0
NOPAT                     0
CapitalInvertido          0
EVA                       0
ROA                       0
ROE                       0
dtype: int64

|Información del DataFrame después de la imputación:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1180 entries, 0 to 1180
Columns: 107 entries, Num to NIT
dtypes: float64(102), int64(5)
memory usage: 986.5 KB
None

Información del DataFrame después de la imputación:
None

Descripción:
- Se limpiaron y convirtieron las columnas numéricas.
- Se imputaron los valores faltantes utilizando la media para las columnas numéricas.
```

Nota. Extracto de resultado de elaboración y ejecución propia.

El uso de estos Pandas en este ejercicio logra, como menciona McKinney (2022), “*sus objetivos principales es buscar trabajar con la data faltante y hacer el ejercicio estadístico menos tortuoso*”. De este modo se elimina la posibilidad de trabajar con data faltante y que a lo largo del trabajo de programación y análisis estadístico, se excluyan objetos o se imposibilite el debido continuar por el no reconocimiento de la data faltante en distintas bibliotecas estadísticas.

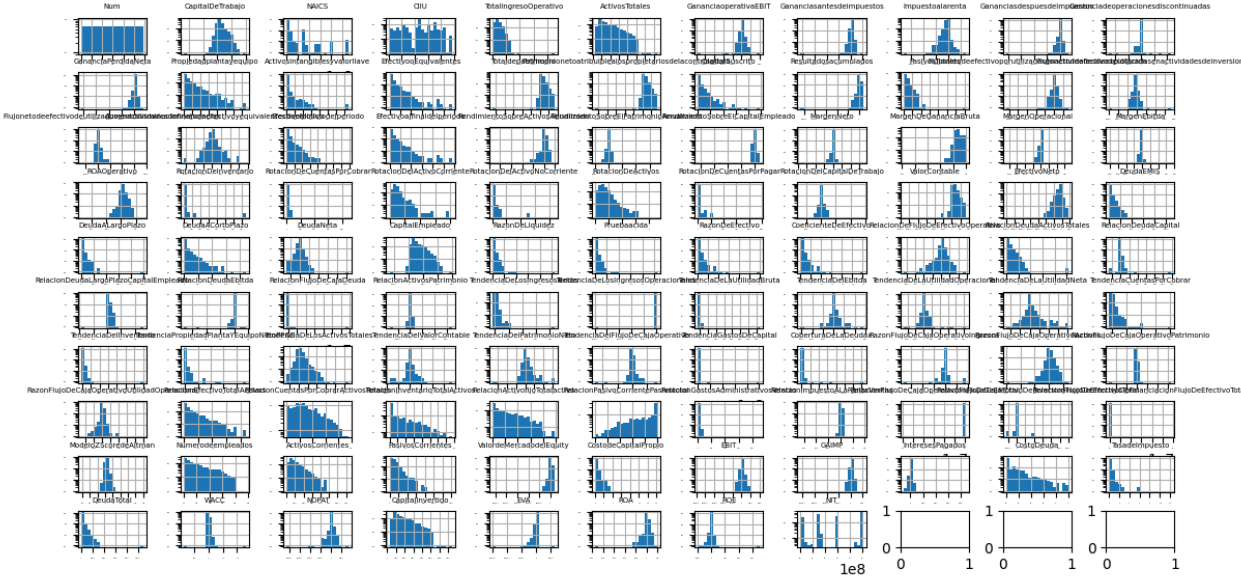
A su vez, podemos validar que la idoneidad de la data extraída y organizada para este estudio tiene las características necesarias para desarrollar correctamente el modelo de regresión. En cuanto a los valores presentados en la Tabla 1 - Limpieza e imputación de los datos, evidenciamos además que la data ya se encontraba bien recopilada y que no fue necesario imputar ningún valor como NaN.

3.3.1. Pruebas de Distribución

3.3.1.1. Histogramas

Para las pruebas de distribución se graficaron los histogramas de la data mediante el uso de la biblioteca Matplotlib.pyplot, las cuales, siguiendo las prácticas de Hunter, J. D. (2007) se evidenciaron las siguientes distribuciones:

Gráfico 2 - Histogramas Datos Recopilados



Nota. Extracto de resultado de elaboración y ejecución propia.

Los histogramas generados a partir del uso de la biblioteca Matplotlib.pyplot para las variables numéricas del conjunto de datos nos otorgan una representación gráfica de las distintas distribuciones atribuibles a las variables independientes y la dependiente. Estos resultados reflejan, presuntamente, que la mayoría de la data recopilada carece de condiciones de normalidad, simetría, sesgos en la forma de la distribución, y cuenta con la presencia de datos atípicos.

En primera instancia, como ya se mencionó, se evidencian diferentes grados de asimetría en la data de las variables, en donde predominan distribuciones con sesgos positivos, como en los valores de *ActivosTotales*, *PasivosTotales*, *PruebaAcida*, *RotaciondelosActivos*, *RotacionCuentasporPagar*, *DeudaTotal*, *ROE*, *TasadeImpuesto*, *WACC*, entre otros. Mientras que aquellas con presunto sesgo negativo se evidencian, *NOPAT*, *EVA*, *RelaciónDeudaEbitda*, *Ebit*, entre otros.

Ya sea el sesgo que se presente, derecha o izquierda, podría afectar el análisis posterior que asumen normalidad, razón por la cual es necesario aplicar métodos estadísticos que presuponen normalidad. Como se presentará más adelante en este texto, se realizó la transformación de datos mediante el modelo de transformación de potencias (Yeo, In-Kwon & Johnson, Richard., 2000), a través de la biblioteca de scikit-learn (Pedregosa et al., 2011) utilizando la clase PowerTransformer, para hacer que los datos sigan una distribución más gaussiana o normal. Este paso es crucial para mejorar la validez de los tests estadísticos que se apliquen y del modelo predictivo.

Para profundizar más sobre el análisis de la distribución, se realizaron las pruebas de Simetría, Shapiro-Wilk, Kolmogorov-Smirnov, y análisis de Curtosis sobre la data recopilada. Estas proporcionan información más detallada sobre la normalidad y las

características de la distribución de las variables dentro del conjunto de datos. Para ello, se utilizaron las bibliotecas de Numpy (McKinney, W., 2022), así como las proporcionadas por Scipy.stats, usando las clases kurtosis, skew, shapiro, ks_2samp, utilizando la versión 1.13.0 de SciPy (Virtanen et al., 2020). En la Tabla 2 se pueden ver los resultados de forma general.

Tabla 2 - Cálculos y Pruebas de Distribución

Cálculos Estadísticos:

CapitalDeTrabajo:

Sesgo: 0.5834440947838461

Curtosis: 8.446400464060622

Prueba de Shapiro-Wilk (p-valor): 6.765589933158646e-29

Prueba de Kolmogorov-Smirnov:

KS estadística: 0.0

p-valor KS: 1.0

...

EVA:

Sesgo: -17.98873459054208

Curtosis: 528.2470354576278

Prueba de Shapiro-Wilk (p-valor): 1.994266577657746e-58

Prueba de Kolmogorov-Smirnov:

KS estadística: 0.8347457627118644

p-valor KS: 0.0

ROA:

Sesgo: -14.04757894293821

Curtosis: 340.5666461253592

Prueba de Shapiro-Wilk (p-valor): 3.31267775968688e-51

Prueba de Kolmogorov-Smirnov:

KS estadística: 0.8694915254237288

p-valor KS: 0.0

ROE:

Sesgo: 23.585162302309406

Curtosis: 741.9625708649739

Prueba de Shapiro-Wilk (p-valor): 4.7301455994791246e-58

Prueba de Kolmogorov-Smirnov:

KS estadística: 0.8559322033898306

p-valor KS: 0.0

Descripción:

- Se realizaron sesgo, pruebas de Shapiro-Wilk y Kolmogorov-Smirnov, y curtosis.

Nota. Extracto de resultado de elaboración y ejecución propia.

3.3.1.2.Sesgo y Curtosis

Primero, se decidió realizar prueba de sesgo de Fisher (1920), y de curtosis en los datos para medir la distribución, con el objetivo de identificar los diferentes sesgos de la distribución, las variables con más alta asimetría, y el empinamiento o peso en las colas (Bai, J., & Ng, S., 2005)., para poder seleccionar aquellas que se comporten mejor en el modelo; esto con el fin de evitar problemáticas en ciertos análisis estadísticos. Chesher, A., & Peters, S. (1994).

En cuanto a la simetría de las distribuciones, bajo el modelo de Fisher (1920), se indicó si los valores están distribuidos uniformemente alrededor de la media. Una simetría alrededor de 0 indica que está es perfectamente simétrica y que los valores están igualmente distribuidos a ambos lados de la media.

Mientras que si se obtienen valores mayor o menores a 0.0, significa que la distribución está sesgada, siendo valores > 0.0 o positivos tienen una cola positiva; o valores negativos, marcan habría presencia de una cola negativa, o que hay más valores hacia la izquierda desde la media que los que hay hacia la derecha.

Se buscó también medir la "colas" de las distribuciones mediante el análisis de la curtosis (Pearson, K., 1892)., es decir, medir que tan qué tan pesadas son las colas en comparación con una distribución normal y que tanto estas se parecen o diferencian de nuestra variable dependiente.

En términos generales, a una curtosis positiva se estaría indicando la presencia de una distribución con colas más pesadas que la distribución normal, o lo que es lo mismo, que

hay más valores más extremos en los datos. Mientras que una curtosis negativa marcaría la presencia de colas más ligeras que una normal, o que hay menos valores extremos.

La interpretación realizada para las variables independientes evidenció que, por ejemplo, para variables como *"NAICS"*, *"TotalIngresoOperativo"*, *"CapitalSuscrito"*, *"DeudaALargoPlazo"*, *"ModeloZScoredeAltman"*, posiblemente existe presencia de asimetría y curtosis positiva. Esto sugiere que hay valores extremadamente altos que pueden influir en la distribución de los datos. Para aquellas variables que presentan estas mismas características, la asimetría y curtosis positiva es bastante alta.

Mientras que, para algunas variables como *"Resultadosacumulados"*, la simetría y la curtosis pueden estar más cerca de cero, lo que indica una distribución más cercana a la normalidad.

Del mismo análisis realizado, en general, la presencia de asimetría y curtosis en las variables es presente con frecuencia, lo que indicaría que los datos no se ajustan perfectamente a una distribución normal, por lo que se presume que estos estarían causados por la diferencia significativa de las empresas del scope, lo que puede requerir técnicas de ajuste para el análisis estadístico alternativas.

Buscando analizar las variables de mayor interés, en la investigación, en la **Tabla 3** se presentan los análisis de sesgo y curtosis. En esta vista generada, como se mencionó anteriormente, es claro que los valores requieren un ajuste para tener características de distribución normal.

Tabla 3 - Análisis de sesgo y curtosis

Cálculos Estadísticos:
Tabla 3

Sesgo de las columnas (ordenadas de menor a mayor):		
RelacionFlujoDeCajaOperativoFlujoDeCajaTotal:	Sesgo: -34.29229641142469	Curtosis: 1174.306583270332
TendenciaDeLosIngresosOperacionales:	Sesgo: -34.21419917571042	Curtosis: 1170.7807285741064
...		
EVA:	Sesgo: -17.98873459054208	Curtosis: 528.2470354576278
ROA:	Sesgo: -14.04757894293821	Curtosis: 340.5666461253592
...		
EBIT:	Sesgo: -11.045724408790727	Curtosis: 268.96821981918214
NOPAT:	Sesgo: -10.59193448512451	Curtosis: 224.48365886437026
...		
MargenOperacional:	Sesgo: -7.118473074809034	Curtosis: 462.17236219253493
EfectivoNeto:	Sesgo: -6.209621827776102	Curtosis: 98.95790046825448
...		
MargenEbitda:	Sesgo: -4.318419139906433	Curtosis: 511.9623862123175
...		
CIIU:	Sesgo: -0.18768657501959257	Curtosis: -0.9269857168436917
CapitalDeTrabajo:	Sesgo: 0.5834440947838461	Curtosis: 8.446400464068622
...		
RelacionInventarioTotalActivos:	Sesgo: 1.2979295004535887	Curtosis: 0.9240284469405866
ActivosCorrientes:	Sesgo: 1.7200856123046637	Curtosis: 4.66881488070183
RelacionEfectivoTotalActivos:	Sesgo: 1.9799261088869307	Curtosis: 4.243717681620614
CostoDeuda:	Sesgo: 2.19747892668569	Curtosis: 5.284026845397969
PasivosCorrientes:	Sesgo: 3.4580468531165045	Curtosis: 19.964032432160174
...		
NAICS:	Sesgo: 3.8016717224176326	Curtosis: 17.29756834742042
RotacionDelActivoCorriente:	Sesgo: 3.865598952529343	Curtosis: 22.293536201370777
...		
DeudaNeta:	Sesgo: 4.545690353647196	Curtosis: 85.97764221998764
ModeloZScoredeAltman:	Sesgo: 4.9261564556807445	Curtosis: 135.60619944343836
PasivosTotales:	Sesgo: 6.397646932973096	Curtosis: 97.42183935796157
DeudaTotal:	Sesgo: 10.742881003845456	Curtosis: 212.90772672517065
WACC:	Sesgo: 11.230096652153524	Curtosis: 483.87281068676657
CostodeCapitalPropio:	Sesgo: 11.975238344209606	Curtosis: 224.71570726812453
RazonDeEfectivo:	Sesgo: 12.95313225909561	Curtosis: 249.8779158812862
TasadeImpuesto:	Sesgo: 13.306317438209687	Curtosis: 220.70335163530612
...		
RazonDeLiquidez:	Sesgo: 17.03176987566808	Curtosis: 410.15576035359686
Pruebaacida:	Sesgo: 18.38007031925696	Curtosis: 456.7678438839751
RotacionDeInventario:	Sesgo: 19.088282396314415	Curtosis: 386.08128014316765
...		
ROE:	Sesgo: 23.585162302309406	Curtosis: 741.9625708649739
RelacionDeudaCapital:	Sesgo: 24.47055500833354	Curtosis: 755.1967280845258
TendenciaDeLaUtilidadBruta:	Sesgo: 33.98448936213548	Curtosis: 1160.1925963766116
RelacionFlujoDeEfectivoDeFinanciacionFlujoDeEfectivoTotal:	Sesgo: 34.29730958868406	Curtosis: 1174.5384384206448

Nota. Extracto de resultado de elaboración y ejecución propia.

3.3.1.3.Shapiro-Wilk

Por otra parte, se realizaron las pruebas de Shapiro-Wilk (1965) y Kolmogorov-Smirnov. (1933), presentadas por Razali, N. M., & Wah, Y. B. (2011), busca tanto evaluar si la muestra proviene de una distribución normal y si al comparar una a una las variables independientes con la dependiente se ajustan a una misma distribución.

Utilizando la versión 1.13.0 de SciPy (Virtanen et al., 2020), se realizó la prueba de Shapiro-Wilk donde la hipótesis nula está definida en que datos de la muestra observada siguen una distribución normal; para tal efecto, el estadístico del valor p resultante de la prueba es aquel que valida o no la hipótesis.

Utilizando la clase shapiro de la biblioteca SciPy, si el p valor es menor que el nivel de significancia predeterminado, es decir 0.05, se rechaza la hipótesis nula (Virtanen et al., 2020), esto indicaría que los datos recopilados no están distribuidos de forma normal. En cambio, si p -valor es mayor que el nivel de significancia, no existe suficiente evidencia para rechazar la hipótesis nula.

Tabla 4 - Prueba de Shapiro

Cálculos Estadísticos:

Tabla 4.

Resultados de la prueba de Shapiro-Wilk (ordenados de menor a mayor):

RelacionFlujoDeEfectivoDeFinanciacionFlujoDeEfectivoTotal:	p-valor:	6.951696597586707e-61
RelacionFlujoDeCajaOperativoFlujoDeCajaTotal:	p-valor:	7.009237057002548e-61
TendenciaGastosDeCapital:	p-valor:	8.136302971416227e-61
...		
EVA:	p-valor:	1.994266577657746e-58
ROE:	p-valor:	4.7301455994791246e-58
WACC:	p-valor:	1.0199043868255145e-57
TendenciaDelFlujoDeCajaOperativo:	p-valor:	1.056117405047164e-57
GananciaPerdidaNeta:	p-valor:	2.0423522408072233e-57
...		
NOPAT:	p-valor:	4.6745764508142075e-52
GananciaoperativaEBIT:	p-valor:	1.0412193849837811e-51
EBIT:	p-valor:	1.0412193849837811e-51
ROA:	p-valor:	3.31267775968688e-51
NAICS:	p-valor:	1.856637444666618e-49
PasivosTotales:	p-valor:	2.346986187306956e-43
CostoDeuda:	p-valor:	1.09360218175826e-42
PasivosCorrientes:	p-valor:	4.613154546398092e-41
ROAOperativo:	p-valor:	1.4446056646384846e-40
DeudaNeta:	p-valor:	1.68564424772524e-40
RelacionInventarioTotalActivos:	p-valor:	1.2904606809652393e-35
MargenDeGananciaBruta:	p-valor:	1.6006130278384585e-35
ActivosCorrientes:	p-valor:	7.231206037934134e-31
Numerodeempleados:	p-valor:	2.355211924474698e-28
ActivosTotales:	p-valor:	7.392127215096044e-27
CapitalInvertido:	p-valor:	7.392127215096044e-27
RelacionPasivoCorrientePasivototal:	p-valor:	2.3713556607362574e-26

Descripción:

- Se realizó prueba de Shapiro-Wilk (ordenados de menor a mayor)
- Se calcularon 107 entradas junto con la variable dependiente.

Nota. Extracto de resultado de elaboración y ejecución propia.

Analizando la Tabla 4 - Prueba de Shapiro, se evidencia que todas las variables no siguen una distribución normal, lo que justifica necesaria la transformación de la data. Donde en cada línea presentada da como resultado qué tan probable es observar los datos con distribución normal. Por ejemplo, para la variable "EVA", el $p\text{-valor}=1.994266577657746e-58$; en otras palabras, la probabilidad de que estos datos se comporten como una población con una distribución normal es extremadamente baja, prácticamente cero.

Los datos se presentan ordenados de menor a mayor, y en todos los casos la prueba da valores inferiores a 0.05, lo que sugiere que las distribuciones de todas las variables observadas difieren significativamente de ser una distribución normal.

En consecuencia, para la Prueba de Shapiro-Wilk (1965) realizada para todos los datos, hay evidencia significativa para rechazar la hipótesis nula de que los datos provienen de una distribución normal; es decir, que los valores recolectados no sigan una distribución normal.

3.3.1.4. Kolmogorov-Smirnov

Según Razali, N. M., & Wah, Y. B. (2011)., el uso de esta prueba no suele ser la más fuerte, si se compara con la prueba de Shapiro-Wilk, sin embargo, es ampliamente utilizada cuando se trata de analizar muestras no simétricas y recobra significancia en su uso cuando se utiliza para determinar si dos muestras provienen de la misma distribución o son consistentes (Hodges Jr, J. L., 1958).

Siguiendo las pautas de la biblioteca de SciPy en su blog de comunidad (2022) se implementó la prueba de Kolmogorov-Smirnov (KS) para diferentes variables mediante el uso de la biblioteca SciPy.Stats con la clase `ks_2samp`; arrojando como resultado como resultado el valor del estadístico KS y el p-valor de KS.

El estadístico KS presenta la máxima discrepancia absoluta entre las dos distribuciones de dos muestras; y para el ejercicio, se planteó comparar todas las variables independientes con la variable dependiente. Para el estadístico KS, cuanto mayor sea el resultado, mayor será la discrepancia de la muestra de la variable independiente de la dependiente.

A su vez, el estadístico p-valor KS presenta la probabilidad de observar una diferencia tan grande entre las distribuciones, es decir, que rechace la hipótesis nula. Para entender este estadístico, en la mayoría de los casos, se define que una probabilidad grande es aquella en donde no hay suficiente evidencia para rechazar la hipótesis nula (H_0); y generalmente se interpreta como cualquier p valor mayor que un umbral de significancia utilizado, comúnmente 0.05 o 0.01 (Hodges Jr, J. L., 1958).

Dentro de esta misma prueba de hipótesis, se establece que si *p valor* establece que no hay diferencia entre las variables, no existe suficiente evidencia para concluir que hay una diferencia significativa entre las muestras o variables, por lo que no se rechaza la hipótesis nula. Mientras que un valor de p bajo indica que hay evidencia suficiente para rechazar la hipótesis nula y que las dos muestras provienen de la misma distribución.(Hodges Jr, J. L., 1958)

Tabla 5 - Kolmogorov-Smirnov

Cálculos Estadísticos
Tabla 5.

Resultados de la prueba de Kolmogorov-Smirnov (ordenados de menor a mayor):		
Impuestoalarenta:	KS estadística: 0.8398305084745763	p-valor KS: 0.0
Gananciadeoperacionesdiscontinuas:	KS estadística: 0.8694915254237288	p-valor KS: 0.0
RendimientoSobreElCapitalEmpleado:	KS estadística: 0.8542372881355932	p-valor KS: 0.0
MargenNeto:	KS estadística: 0.8661016949152542	p-valor KS: 0.0
MargenDeGananciaBruta:	KS estadística: 0.8661016949152542	p-valor KS: 0.0
MargenOperacional:	KS estadística: 0.8669491525423729	p-valor KS: 0.0
MargenEbitda:	KS estadística: 0.8669491525423729	p-valor KS: 0.0
...		
RotacionDelCapitalDeTrabajo:	KS estadística: 0.8550847457627119	p-valor KS: 0.0
EfectivoNeto:	KS estadística: 0.8144067796610169	p-valor KS: 0.0
RazonDeLiquidez:	KS estadística: 0.8711864406779661	p-valor KS: 0.0
Pruebaacida:	KS estadística: 0.8711864406779661	p-valor KS: 0.0
RazonDeEfectivo:	KS estadística: 0.8728813559322034	p-valor KS: 0.0
CoficienteDeEfectivo:	KS estadística: 0.8728813559322034	p-valor KS: 0.0
TendenciaDeLaUtilidadBruta:	KS estadística: 0.8449152542372881	p-valor KS: 0.0
TendenciaDelEbitda:	KS estadística: 0.8152542372881356	p-valor KS: 0.0
...		
Numerodeempleados:	KS estadística: 0.8567796610169491	p-valor KS: 0.0
CostodeCapitalPropio:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
CostoDeuda:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
TasadeImpuesto:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
WACC:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
EVA:	KS estadística: 0.8347457627118644	p-valor KS: 0.0
ROA:	KS estadística: 0.8694915254237288	p-valor KS: 0.0
ROE:	KS estadística: 0.8559322033898306	p-valor KS: 0.0
DeudaACortoPlazo:	KS estadística: 0.5491525423728814	p-valor KS: 8.77026248619982e-164
DeudaALargoPlazo:	KS estadística: 0.5135593220338983	p-valor KS: 3.414495392470852e-142
NOPAT:	KS estadística: 0.4364406779661017	p-valor KS: 2.340988108306036e-101
...		
Propiedadplantayequipo:	KS estadística: 0.13050847457627118	p-valor KS: 3.558178826064034e-09
ValordeMercadodelEquity:	KS estadística: 0.1059322033898305	p-valor KS: 3.483007543759245e-06
Resultadosacumulados:	KS estadística: 0.06610169491525424	p-valor KS: 0.011512088177701149
CapitalDeTrabajo:	KS estadística: 0.0	p-valor KS: 1.0

Descripción:
- Se realizó prueba de Kolmogorov-Smirnov (ordenados de menor a mayor).
- Se calcularon 107 entradas junto con la variable dependiente.]

Nota. Extracto de resultado de elaboración y ejecución propia

En el análisis realizado sobre las muestras presentadas en la **Tabla 4**, se encontró que en todos los p-valores KS son extremadamente bajos o cercanos a cero, lo que sugiere que hay evidencia significativa para rechazar la hipótesis nula en todos los casos; y en consecuencia, se concluye que las distribuciones de las variables independientes son diferentes en comparación con la variable dependiente.

Para los estadísticos resultantes de la prueba KS en la variable *CapitaldeTrabajo* se obtuvo que *KS estadístico=0.0* y *p-valor KS=1.0*. Cuando se compara una variable consigo misma utilizando esta prueba indica que no hay discrepancia alguna entre la función de distribución empírica de la variable y sí misma, ya que la variable es idéntica a sí misma. Y

en cuanto a la prueba de hipótesis, el p-valor KS de 1.0 significa que, bajo la hipótesis nula de que la variable es idéntica a sí misma, la probabilidad de que exista una diferencia tan grande del 100%, es decir, no hay evidencia suficiente para rechazar esta hipótesis nula y dado el alto *p-valor*, no hay diferencias entre la variable dependiente y sí misma, lo cual es un resultado esperado. (Hodges Jr, J. L., 1958)

4. Metodología

Primero, se buscó recopilar la información mediante la consulta en fuentes oficiales con información estandarizada (Sampieri, 2018). Segundo, se seleccionó el software que permitiera realizar los cálculos estadísticos necesarios para la investigación, y que al mismo tiempo fuera escalable y ágil al momento de procesar la data (McKinney, 2022). Tercero, se validó que la información recopilada fuera consistente mediante la limpieza de la data y la imputación de datos (si fuera necesario), pues según McKinney, la construcción y preparación de la información para cualquier modelo estadístico, “estas tareas ocupan hasta el 80% de la tarea investigativa” (2022).

Cuarto, se realizaron pruebas de distribución y se graficaron los histogramas de la data. En este paso, se realizaron las pruebas de Shapiro-Wilk, Kolmogorov-Smirnov, Simetría, y análisis de Curtosis. Luego se visualizaron los histogramas para contrastar las pruebas realizadas y evaluar si se ajustan a una distribución normal.

Quinto, se aplicaron pruebas de multicolinealidad sobre las variables independientes. Más adelante, el resultado del cuarto paso evidenció que era también necesario realizar transformaciones de potencia para ajustar la data a una distribución más adecuada. Se calculó entonces el coeficiente de correlación entre las variables independientes contra la dependiente, así como el factor de inflación de la varianza (VIF) y el número de condición con respecto a la variable dependiente.

Sexto, se analizaron y seleccionaron las variables independientes que serían incluidas dentro de la regresión lineal múltiple mediante la eliminación hacia atrás Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005).

Séptimo, se procedió a ajustar el modelo de regresión lineal múltiple utilizando las variables predictoras seleccionadas. Se estimaron los coeficientes de regresión y sus correspondientes errores estándar, así como también se evaluó la bondad de ajuste del modelo mediante estadísticos como el coeficiente de determinación R^2 y el análisis de residuos.

Finalmente, se realizaron las pruebas de hipótesis para determinar si estas variables tienen un efecto significativo en la variable dependiente; en la cual, H_0 implica que no hay relación entre la variable predictora y la variable dependiente; para ello, H_1 , se probó mediante el estadístico valor-t o el valor F, determinando si hay o no evidencia estadística para rechazar la hipótesis nula y concluir que existe una relación significativa entre las variables predictoras y la variable dependiente.

4.1.1. Pruebas de Multicolinealidad.

Haitovsky, Y. (1969). plantea que la multicolinealidad existente sobre las muestras a estudiar en modelos estadísticos, especialmente en regresiones distorsiona los coeficientes y conduce a incorrectas estimaciones entre las variables dependientes e independientes

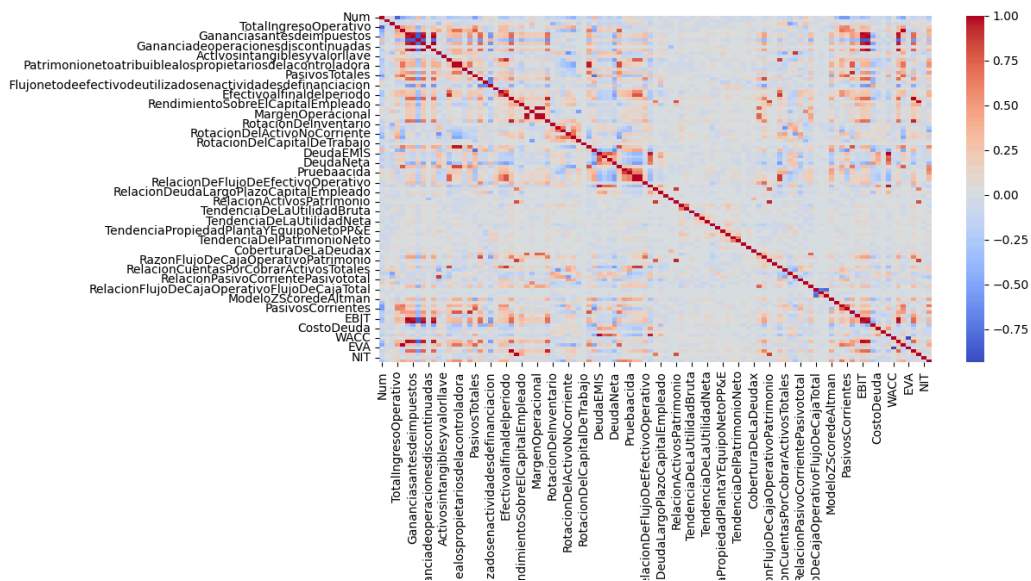
Desde la perspectiva de esta investigación se busca probar la existencia de multicolinealidad, la cual y dados los resultados de las pruebas de distribución realizadas, se encontró evidencia suficiente para realizar pruebas de multicolinealidad. Se tuvieron en cuenta las correlaciones, Factor de Inflación de la Varianza, en adelante VIF; así como la tolerancia. Estos estadísticos son utilizados en el contexto de análisis de regresión múltiple y eliminar posibles afectaciones en la precisión e interpretación de los resultados

4.1.1.1. Correlación de Variables

Para el análisis de la correlación se utilizó la biblioteca Pandas y NumPy (McKinney, 2022) en la clase `corr()`: para el cálculo de la correlación entre las columnas del dataframe con la variable dependiente y `np.linalg.cond()` para calcular el número de condición de la matriz, lo evalúa la estabilidad numérica de los cálculos en sistemas de ecuaciones lineales.

Sobre el análisis de la matriz de correlaciones, Steiger, J. H. (1980). plantea que cada matriz suele ser simétrica donde su diagonal resulta ser siempre la correlación entre las variables y si mismas, estos valores suelen presentarse con valores entre menos uno y uno para correlaciones perfectas o cero para la ausencia de correlación, para el caso de investigación y utilizando las bibliotecas Matplotlib Hunter, J. D. (2007), se representó gráficamente en un mapa de calor la matriz de correlación mediante la clase `sns.heatmap()` de tal modo el gráfico tres presenta la matriz de correlación entre las variables independientes.

Gráfico 3 - Mapa de Calor Matriz de Correlación



Nota. Extracto de resultado de elaboración y ejecución propia.

En esta mapa de calor se representan los valores con correlación positiva perfecta o valores iguales a 1.0 con color rojo, los valores para las correlaciones negativas perfectas serán -1.0 con color azul, y en la ausencia de color o valores iguales a 0 se representan valores sin correlación. Con ellos se proporciona de forma integral la relación existente entre las variables independientes y se exige de esta representación dependiente.

Tal como lo presentó Steiger, J. H. (1980), la diagonal de este mapa de calor representa las relaciones perfectas que hay entre las variables con sí mismas; el uso de este mapa pretende proporcionar a la investigación herramientas gráficas e indicativas que evidencien posibles conflictos de multicolinealidad.

Es claro esperar que existan altas correlaciones positivas y negativas debido a la relación directa que presentan algunas de las variables con otras, como por ejemplo, la presencia de indicadores financieros, razones financieras con elementos del balance o del estado de resultados. Sin embargo, la ausencia de calor sugiere que no existe una relación lineal directa entre ellas debido a la naturaleza de los datos.

Un claro ejemplo de la correlación positiva existente está presente entre *CapitalDeTrabajo* y *TotalIngresoOperativo*, los cuales, podrían estar relacionados al efecto que tienen las empresas, en cuanto a que, frente a la presencia de liquidez y recursos operativos aumentan los ingresos del negocio por la mejor y eficiente gestión de las empresas.

Otro ejemplo de esto se evidencia entre el EVA y el WACC, cuya correlación negativa sugiere que el impacto en los costos de capital de la empresa destruyen el valor de la misma. También podemos analizar la correlación existente entre ROA y el ROE cuya

relación positiva se ve reflejada desde la perspectiva de los accionistas quienes esperan un mayor rendimiento sobre su posición respecto a la rentabilidad de la firma.

Sobre la correlación existente entre la variable dependiente con las independientes, mediante el uso de la clase **corr.()** de la biblioteca Matplotlib (Hunter, J. D., 2007), se construyó una matriz de correlaciones para entender y evaluar si existe multicolinealidad con la variable que se busca explicar, la cual se presenta en la siguiente tabla.

Tabla 6 - Matriz de Correlación con la Dependiente

Correlación de todas las columnas con la variable dependiente (CapitalDeTrabajo):

ActivosCorrientes	0.650367
RazonDeLiquidez	0.643163
CapitalEmpleado	0.585514
...	...
EfectivoNeto	0.181068
Flujonetodeefectivoporutilizadosenactividadesdeexplotacion	0.178755
RendimientoSobreActivosAnualizado	0.156518
ROA	0.156434
Flujonetodeefectivodeutilizadasenactividadesdeinversion	0.146163
RelacionInventarioTotalActivos	0.128297
PasivosTotales	0.121470
...	...
WACC	0.023022
TendenciaDeLosIngresosOperacionales	0.022388
Gananciadeoperacionesdiscontinuas	0.018340
RelacionFlujoDeEfectivoDeFinanciacionFlujoDeEfectivoTotal	0.018070
...	...
RelacionActivosPatrimonio	-0.013410
ROE	-0.021648
CoberturaDeLaDeudax	-0.023197
EVA	-0.024365
TendenciaDeLaUtilidadBruta	-0.024963
...	...
NAICS	-0.041411
TasadeImpuesto	-0.044495
DeudaALargoPlazo	-0.048230
RazonFlujoDeCajaOperativoUtilidadOperacional	-0.051737
Activosintangiblesyvalorllave	-0.064748
DeudaACortoPlazo	-0.068514
DeudaTotal	-0.070601
...	...
InteresesPagados	-0.278913
Impuestoalarenta	-0.304637
DeudaNeta	-0.362679
RotacionDelActivoCorriente	-0.468914

Name: CapitalDeTrabajo, dtype: float64

Nota. Extracto de resultado de elaboración y ejecución propia

La tabla 6 provee a la investigación una base de estudio necesaria, donde el estadístico resuelto permite evaluar aquellas variables independientes, que posiblemente, sean seleccionadas para la construcción del modelo robusto y explicativo gracias a su poder predictivo.

Dentro de los hallazgos encontrados existen algunas correlaciones positivas significativas como *ActivosCorrientes* y *RazonDeLiquidez* con fuertes correlaciones positivas de 0.650367 y 0.643163, respectivamente. Esto sugiere la necesidad dentro de las empresas la presencia de buena gestión frente a estos indicadores para soportar y entender sus operaciones corrientes.

Por otra parte, indicadores como *ActivosTotales* y *CapitalInvertido* presentan correlaciones de 0.423333 y 0.423333 respectivamente. Esto sugiere que dependiendo del tamaño de la empresa, aquellas con mayores activos tienden a tener mayores recursos disponibles para financiar sus operaciones diarias. Caso similar sucede para *Efectivo o Equivalentes* y *Efectivo al Final del Periodo*; en donde, con correlaciones de 0.366970 y 0.393297, se destaca la relevancia de la liquidez inmediata para mantener un *Capital de Trabajo*.

También, para *GananciasAntesdeImpuestos* y *GananciasDespuésdeImpuestos*, con estadísticos de correlación de 0.359480 y 0.353443, respectivamente, se evidencia la relación que existe en cómo la rentabilidad operativa y neta afecta directamente la capacidad de la empresa para gestionar eficazmente su *Capital de Trabajo*.

Para los resultados de las variables con resultados negativos, existe por ejemplo la relación entre el *Capital de Trabajo* con la *Deuda Neta* (*corr: -0.362679*), planteando que frente a un aumento en esta variable se disminuye el Capital de Trabajo disponible. O para el caso de *Rotación del Activo Corriente* (*corr: -0.468914*), donde se podría indicar que ante una alta rotación de activos corrientes, a pesar de los *ActivosCorrientes*, no necesariamente se contribuye a un mayor Capital de Trabajo pues por la presencia de ciclos de conversión de efectivo cortos, la liquidez no siempre es inmediata.

Finalmente, existen correlaciones con variables menos explicativas, como *NAICS*, *CIU*, o *Número de Empleados*, las cuales tienen correlaciones muy cercanas a 0; lo que podría sugerir que, estos elementos podrían no ser predictores del *Capital de Trabajo* y por tanto, podrían hacer el modelo predictivo.

4.1.1.2. VIF (Factor de Inflación de la Varianza) y Tolerancia

Akinwande, M.O., Dikko, H.G. and Samson, A. (2015) sugieren que los conceptos y pruebas de VIF (Factor de Inflación de la Varianza) y Tolerancia “*son más comunes en los procesos de investigación de lo que las personas piensan debido a su aplicabilidad*”.

En un modelo que se enfrenta a multicolinealidad, dos o más variables explicativas de la variable a predecir suelen ser supresoras entre sí; y en un modelo de regresión múltiple también afecta el valor a predecir.

Este efecto aumenta de manera exponencial el error estándar de los coeficientes estudiados motivo por el cual, dentro del análisis de regresión múltiple el uso de de VIF para estimar la multicolinealidad resulta ser crucial para determinar cuando los valores predictores están altamente correlacionados.

Para Akinwande, M.O., Dikko, H.G. and Samson, A. (2015), la interpretación de VIF se evalúa cuando se aumentó la varianza de un coeficiente de regresión estimado, cuando existe correlación entre ellos. Si las variables no están correlacionadas el estadístico será inferior a 1.0; si el VIF se encuentra en 1.0 y 5.0 indica que los valores están moderadamente correlacionados; si VIF está entre 5.0 y 10.0 que las variables predictoras presentan una alta correlación y si VIF supera 10.0 se puede interpretar que los coeficientes de regresión resultantes de la regresión lineal múltiple están mal estimados como

consecuencia de la multicolinealidad y evitaría garantizar la precisión de los resultados del análisis de regresión.

Por otra parte, podemos entender la tolerancia según Oke, J., Akinkunmi, W. B., & Etebefia, S. O. (2019) desde la perspectiva de estimación de VIF en donde busca determinar dentro de la varianza de los coeficientes estimados presentan al menos una pequeña parte característica de multicolinealidad.

Entonces se puede entender que para los valores estimados de la tolerancia cercanos o iguales a cero existe la probabilidad de presencia de multicolinealidad en los datos estimados. Para todos los valores superiores a 0.1, se puede interpretar que no necesariamente exista una multicolinealidad significativa.

Como consecuencia de lo anterior buscando reducir el impacto negativo dentro de los coeficientes de regresión por efecto de multicolinealidad se decide dentro de esta investigación aplicar estas dos pruebas.

Adicionalmente, frente a la existencia de multicolinealidad podría presentarse también la pérdida de la significancia de las variables predictoras, la inestabilidad de los coeficientes y la hipersensibilidad frente a cambios en los coeficientes estimados.

Como se verá más adelante en este documento y utilizando el modelo de transformación de potencias propuesto por Yeo, In-Kwon & Johnson, Richard., (2000), se busca que los datos de las variables predictoras, se aborden de manera más efectiva.

Utilizando las bibliotecas Pandas y NumPy (McKinney, 2022) y las bibliotecas statsmodels (Seabold, S., & Perktold, J. 2010) se realizó el análisis de VIF y tolerancia presentadas en la Tabla 7.

Tabla 7 - Vif y Tolerancia

Tabla 7 - Vif y Tolerancia

Para las variables independientes con respecto a la dependiente (CapitalDeTrabajo):

Variable	VIF	Tolerancia
Totaldepatrimonio	inf	0.000000
EBIT	inf	0.000000
ROA	inf	0.000000
...
CapitalInvertido	inf	0.000000
DeudaTotal	inf	0.000000
RazonDeEfectivo	inf	0.000000
Resultadosacumulados	inf	0.000000
...
Gananciasantesdeimpuestos	inf	0.000000
GananciaoperativaEBIT	inf	0.000000
ActivosTotales	inf	0.000000
ROE	inf	0.000000
...
Impuestoalarenta	81611344466.561493	0.000000
Gananciadeoperacionesdiscontinuadas	39809593757.280403	0.000000
...
NOPAT	59.080066	0.016926
EVA	26.218979	0.038140
WACC	23.556320	0.042451
...
RelacionImpuestoALaRentaVentas	11.471696	0.087171
CIIU	8.107689	0.123340
RelacionInventarioTotalActivos	7.221727	0.138471
Numerodeempleados	7.177736	0.139320
...
CostodeCapitalPropio	1.769772	0.565045
CostoDeuda	1.685406	0.593329
NAICS	1.556148	0.642612
TendenciaDelFlujoDeCajaOperativo	1.422427	0.703024
...
RotacionDeCuentasPorCobrar	1.056313	0.946689
RotacionDeInventario	1.049442	0.952888
TendenciaGastosDeCapital	1.042027	0.959668

Número de Condición:
2.2401775785864096e+23

Descripción:

- Se calcularon la correlación, el VIF, la tolerancia y el número de condición con respecto a la variable dependiente.

Nota. Extracto de resultado de elaboración y ejecución propia

Los hallazgos encontrados dentro de la Tabla 7 se pueden ver desde la visión de VIF y de tolerancia; en el cual, empezando con VIF, puede verse valores con VIF infinito y VIF con valores finitos. Para aquellas variables con VIF infinito se puede interpretar que están

altamente correlacionadas con otras variables en el modelo, lo que indica multicolinealidad severa (Oke, J., Akinkunmi, W. B., & Etebefia, S. O. 2019). Cuando el VIF es infinito, se evidencia que la muestra presenta una varianza de un coeficiente de regresión inflada de forma considerablemente, y que en consecuencia, existe multicolinealidad.

Por otra parte, existen las variables con VIF finito, las cuales se pueden dividir en dos, las que están en valores superiores a 10 y las que no. Para aquellas que se encuentran con VIF >10 , se puede inferir que también tienen una alta correlación con otras variables en el modelo, pero no tan alto como las que tienen a infinito; de todos modos, son variables que requieren un ajuste por transformación para poder ser tenidas en cuenta dentro del modelo.

Y para las variables con VIF finito < 10 , se puede interpretar que están correlacionados con otros predictores y que en, menor medida que las anteriores, podrían necesitar un un ajuste por transformación para poder ser tenidas en cuenta dentro del modelo de regresión.

En todo caso, para la prueba realizada, 71 de las 107 variables recolectadas requieren un ajuste por transformación, pues variables de interés como *EBIT*, *ROA*, *CapitalInvertido*, *DeudaTotal*, *RazonDeEfectivo*, *Resultadosacumulados*, *ActivosTotales*, *ROE*, *NOPAT*, *EVA*, *WACC*, entre otros, se encuentran bajo esta condición de multicolinealidad.

Para la interpretación de la Tolerancia estimada, podemos inducir que los resultados de Tolerancia cercana a cero presentan multicolinealidad y una gran proporción de la varianza de la variable independiente está asociada con otras variables independientes en el modelo. Esto soportará más los valores por VIF, pues todos los valores con VIF infinito y la mayoría de los valores con VIF > 10 tienen tolerancia igual a cero.

Aunque no existen valores con tolerancia =1, existen unas pocas variables con resultados cercano a uno. Esto Indica que aunque hay poca multicolinealidad, no es nula para ningún de los datos.

4.1.1.3. Número de Condición

Finalmente, se incorporó el análisis del Numero de Condición dentro del estudio, donde Trefethen, L. N., & Bau, D. (2022) aborda este concepto como una medida a la sensibilidad en problemas de álgebra lineal frente a pequeñas perturbaciones o errores en los datos de entrada. Concepto que es completamente pertinente para el análisis de VIF si se quiere comprobar que tan sensible resultan los coeficientes de regresiones de las variables predictoras.

El número de condición, cuando su cálculo es grande, supone que el problema es mal condicionado pues se ve altamente afectado a pequeñas variaciones en los datos de entrada, y resultar en grandes significativos en la regresión estimada. Mientras que a valores de número de condición pequeños, el problema estaría bien condicionado, y no se vería afectado a cambios o perturbaciones en la data de las variables independientes.

En el contexto de la prueba realizada, utilizando la clase **np.linalg.cond()** de la biblioteca Numpy de McKinney, W. (2022), el valor calculado para el Número de Condición es igual a $2.24e+23$. Un valor extremadamente grande, para sugerir que la matriz de los datos evaluados, es decir, todas las variables independientes y la variable dependiente es casi singular o completamente singular; y que para Trefethen, L. N., & Bau, D. (2022), la singularidad o descomposición en valores singulares (SVD), hace referencia a que una o

más de sus columnas pueden ser expresadas como combinaciones lineales de otras columnas, es decir, que hay presencia de multicolinealidad.

4.1.2. Transformación de Datos

La data recolectada, tras las pruebas realizadas anteriormente, evidencian una necesidad de transformación para eliminar la multicolinealidad, la asimetría, alta curtosis, y que tengan una distribución más gaussiana o normal. Para ello, se definió que el método a utilizar sería el propuesto por Yeo-Johnson (2000), el cual es una técnica utilizada para estabilizar la varianza y hacer que los datos se aproximen más a una distribución normal.

Modificado por Yeo (2000), este modelo es una mejora, por así decirlo, de la transformación de Box, G. E. P., & Cox, D. R. (1964) propuesta por Johnson, N. L.(1949), plantea como principal diferencia que la transformación de modelo Yeo-Johnson en comparación a el modelo Box-Cox acepta en su cálculo tanto valores positivos como negativos, mientras que la segunda acepta estrictamente datos positivos.

Como se mencionó anteriormente, a través de la biblioteca de scikit-learn (Pedregosa et al., 2011) y utilizando la clase PowerTransformer, se pudo aplicar el modelo de Transformación de Datos de Yeo, In-Kwon & Johnson, Richard. (2000), con el objetivo de permitir que los datos sigan una distribución más gaussiana o normal, y a su vez, se elimine la multicolinealidad. Este paso es crucial para mejorar la validez de los tests estadísticos que se apliquen de forma adecuada al modelo predictivo; este modelo consiste principalmente en estabilizar la varianza y hacer que los datos se aproximen más a una distribución normal.

Donde:

$$y_i^{(\lambda)} = \begin{cases} (y_i + 1)^\lambda - 1, & \text{si } \lambda \neq 0, y_i \geq 0 \\ \ln(y_i + 1), & \text{si } \lambda = 0, y_i \geq 0 \\ -(-y_i + 1)^{2-\lambda} + 1, & \text{si } \lambda \neq 2, y_i < 0 \\ -\ln(-y_i + 1), & \text{si } \lambda = 2, y_i < 0 \end{cases}$$

y_i : *valor original de la observación*

λ : *es el parámetro de transformación*

La transformación de Yeo, In-Kwon & Johnson, Richard. (2000) es bastante flexible y dependiendo del valor de la variable original se puede aplicar un camino u otro. Motivo por el cual se utiliza comúnmente en estadística y analítica de datos cuando estos no cumplen los criterios de normalidad o presentan multicolinealidad.

Tabla 8 - Transformación Yeo-Johnson

|Tabla 8 - Transformación Yeo-Johnson

Primeras filas de los datos transformados:

	NAICS	CIIU	TotalIngresoOperativo	ActivosTotales	\		
0	0.195477	-1.152019	2.117304	1.873766			
1	0.428739	0.269351	2.097511	0.202594			
2	-2.099388	0.082765	0.627047	0.839465			
3	-0.925633	0.023679	1.261987	0.109728			
4	1.534226	-1.011229	2.353414	1.411411			
...							
	WACC	NOPAT	CapitalInvertido	EVA	ROA	ROE	\
0	-0.115197	0.296773	1.873766	-0.031562	-0.377186	0.076458	
1	0.230603	0.914177	0.202594	-0.042821	1.236027	0.254514	
2	0.001029	1.078946	0.839465	0.059843	0.686668	0.136058	
3	-0.125816	-0.015369	0.109728	0.135941	-0.440049	0.012873	
4	0.032552	1.096266	1.411411	-0.101887	0.385066	0.023904	
	NIT	CapitalDeTrabajo					
0	-0.680087	1965.35					
1	1.198103	6500.78					
2	-0.692478	10467.32					
3	1.197904	7094.02					
4	0.105892	12553.36					

Información del DataFrame después de la transformación:

None

Descripción:

- Se realizó una transformación de potencia en las variables independientes.

Nota. Extracto de resultado de elaboración y ejecución propia

La Tabla 8 presenta los datos transformados mediante la metodología de Yeo, In-Kwon & Johnson, Richard. (2000) mediante Python, usando la biblioteca scikit-learn (Pedregosa et al., 2011) y utilizando la clase **PowerTransformer()**. A primera vista, se evidencia que para todos los valores de las variables independientes han sido transformados.

Para analizar el impacto de la transformación, se corrieron nuevamente las pruebas anteriores y estos fueron los resultados.

Tabla 9 - Sesgo y Curtosis (transformados)

Tabla 9 - Sesgo_t de las columnas (ordenadas de menor a mayor) y Curtosis:

EVA:	sesgo_t: -11.389701857607724	Curtosis: 433.504657316402
CoberturaDeLaDeudax:	sesgo_t: -6.593763375702745	Curtosis: 421.22357895376587
Flujonetodeefectivodeutilizadasenactividadesdeinversion:	sesgo_t: -4.4665715175217136	Curtosis: 214.42665720633084
...
NAICS:	sesgo_t: -0.00881978430377713	Curtosis: -0.6381984444149928
RelacionGastosAdministrativosVentas:	sesgo_t: -0.00823078676320411	Curtosis: 1.6072913460144873
EfectivooEquivalentes:	sesgo_t: -0.00222302458889313	Curtosis: -0.00832930668466414
ActivosTotales:	sesgo_t: 0.004136528818896239	Curtosis: -0.8857691319984542
CapitalInvertido:	sesgo_t: 0.004136528818896239	Curtosis: -0.8857691319984542
RelacionEfectivoTotalActivos:	sesgo_t: 0.004648728152861353	Curtosis: -0.8553624653279575
ROA:	sesgo_t: 0.9155929975403917	Curtosis: 25.828078944370663
...
NOPAT:	sesgo_t: 3.902820308926022	Curtosis: 70.08787955805423
RazonFlujoDeCajaOperativoIngresos:	sesgo_t: 3.940162777066265	Curtosis: 266.4928488855823
GananciaoperativaEBIT:	sesgo_t: 4.229046609482915	Curtosis: 78.95021429759683
EBIT:	sesgo_t: 4.229046609482915	Curtosis: 78.95021429759683
...
ROE:	sesgo_t: 9.74761960252138	Curtosis: 388.53118780606617
MargenEbitda:	sesgo_t: 10.801410426055138	Curtosis: 569.2996759225524
MargenOperacional:	sesgo_t: 11.540636315884594	Curtosis: 546.6306622413412
RelacionActivosPatrimonio:	sesgo_t: 17.573219382156662	Curtosis: 693.7215025806337

Nota. Extracto de resultado de elaboración y ejecución propia

Tabla 11 - Kolmogorov-Smirnov (transformados)

Tabla 11 - Resultados de la prueba de Kolmogorov-Smirnov (ordenados de menor a mayor):

NAICS:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
TotalIngresoOperativo:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
ActivosTotales:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
GananciaoperativaEBIT:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
...
WACC:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
NOPAT:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
CapitalInvertido:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
EVA:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
ROA:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
ROE:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
NIT:	KS estadística: 0.8737288135593221	p-valor KS: 0.0
CapitalDeTrabajo:	KS estadística: 0.0	p-valor KS: 1.0

Nota. Extracto de resultado de elaboración y ejecución propia

Como se puede evaluar en la Tabla 9, las variables pasaron a tener un menor sesgo, presentando en la gran mayoría valores cercanos a cero, y evaluaciones del estadístico curtosis cercanos a cero en los cálculos realizados post-transformación; lo que asegura que la influencia de valores atípicos extremos es mínimo y que la distribución de la data es lo suficientemente moderada para no distorsionar los resultados del análisis, llegando en algunos casos a comportamientos casi normales.

Por su parte, la Tabla 11 presenta las pruebas corridas de Kolmogorov-Smirnov para las variables post-transformación, los cuales, para KS todos los p-valor en este escenario transformado son igual a cero, es decir, existe evidencia significativa para rechazar la hipótesis nula de que los datos siguen una distribución normal.(Hodges Jr, J. L., 1958)

5. Variable dependiente y selección de variables independientes (transf.)

Como se presentó anteriormente, la investigación trabajó con una base de datos transversales, para un único punto en el tiempo, específicamente con cifras de cierre de 2022, de 1180 medianas empresas de la ciudad de Bogotá, con 106 variables continuas representadas por indicadores financieros, limpios de error, validados e imputados como valor (si fuese necesario), transformadas bajo el modelo Transformación Yeo-Johnson (2000), de la siguiente manera.

5.1.1. Presentación de la variable dependiente

Como se ha mencionado anteriormente, la definición y elección del Capital de Trabajo como la variable dependiente a utilizar en este documento de investigación radica, como se ha presentado anteriormente, en poder resolver la pregunta de investigación, la cual pretende validar si es posible definir un modelo estadístico que le permita a las medianas empresas de Bogotá estimar y mejorar su operatividad y sostenibilidad financiera:

A lo largo de la evaluación y pruebas realizadas, hemos encontrado que los datos recopilados para el Capital de Trabajo buscan explicar las condiciones y decisiones de las empresas. Por su parte, si se analiza como un indicador de liquidez o solvencia, este es un indicador crítico que refleja en buena medida las condiciones de la empresa, mostrando su capacidad para cubrir obligaciones en el corto plazo Van Horne, (J. C., & Wachowicz, J. M. 2004).

Para Vance, D. E. (2002), un adecuado nivel de gestión sobre el Capital de Trabajo permite a las empresas gestionar mejor los riesgos asociados con incertidumbres económicas y financieras, formando una base sólida para la toma de decisiones estratégicas e informadas.

Sharma, A. K., & Kumar, S. (2011). señalan que un Capital de Trabajo saludable ayuda a las empresas a ser más resilientes en condiciones de choque, lo que le permite a la firma estar preparada en situaciones de estrés.

Adicionalmente, este indicador financiero escogido como la variable dependiente, tiene una relación directa con la gestión de los recursos en la operación de las empresas, puesto que, según Smith, K.G. (1980) “la capacidad de las empresas para gestionar sus fuentes de conversión de efectivo determina la rentabilidad de la firma”. Deloof, M. (2003), por ejemplo, halló que existe una correlación significativa entre la gestión del Capital de Trabajo y la rentabilidad de las empresas, elemento que sería significativo para esta investigación en la que se busca las empresas gestionen prudentemente sus recursos apuntando a mejorar la sostenibilidad en el largo plazo.

5.1.2. Presentación de las variables independientes

5.1.2.1. Análisis luego de transformación

Solo las variables independientes fueron transformadas bajo el modelo Transformación Yeo-Johnson (2000). “La selección de las variables es un paso crucial para cualquier modelo” (Núñez, E., Steyerberg, E. W., & Núñez, J. 2011), razón por la cual es necesario basarse en normas estadísticas típicas y seguir las etapas y pautas de la investigación (Sampieri, R. 2018).

Partiendo desde la relevancia teórica en la explicación del Capital de Trabajo, así como los resultados y comportamientos estadísticos presentados, los cuales se probaron anteriormente se realizó la selección de las variables que sean más aptas para una regresión lineal múltiple.

Para ello, a pesar que existieron variables que mejoraron luego de la transformación en aspectos de simetría y curtosis, como lo es el *ROE*; donde se mostraron mejoras significativas, no deberían ser incluídas dentro del estudio por los efectos de multicolinealidad que podrían representar.

En cambio, para variables como el EVA, que a pesar de haber transformado su varianza e intentado normalizar su distribución pero seguir manteniendo una no normalidad, deben incluirse dada su importancia en la evaluación del desempeño financiero.

5.1.2.2. Selección de las variables independientes

En la tabla 11 se presentan las 16 variables independientes seleccionadas, junto con sus respectivos resultados de las pruebas de distribución con la data ya transformada, junto con las pruebas de multicolinealidad de la data original.

Estas variables, fueron definidas por los análisis realizados, buscando capturar efectos positivos y negativos en la dependiente. Como se menciona más adelante en este documento, también se implementó un entrenamiento del modelo en Python que iterara el conjunto de variables utilizadas en la modelo de regresión lineal múltiple mejorando el coeficiente de determinación R^2 .

Tabla 11 - Variables Seleccionadas

Variable Dependiente				
	CapitalDeTrabajo			
sesgo_t	0.5834440947838461			
Curtosis	8.446400464060622			
p-valor SW	6.765589933158646e-29			
KS estadística	0.0			
p-valor KS	1.0			
Variables Independientes				
	RelacionDeudaEbitda	RelacionDeudaCapital	RelacionEfectivoTotalActivos	
sesgo_t	5.77103666028502	7.7577587177043	0.004648728152861353	
Curtosis	319.6782066029444	393.00010854395845	-0.8553624653279575	
p-valor SW	1.4705838699528173e-58	7.986914393482898e-59	4.297873288014945e-11	
KS estadística	0.8737288135593221	0.8737288135593221	0.8737288135593221	
p-valor KS	0.0	0.0	0.0	
	WACC	EVA	GananciaPerdidaNeta	CostoDeuda
sesgo_t	-0.34944683541184607	-11.389701857607724	3.305167254602142	0.599202737527855
Curtosis	423.3177469588479	433.504657316402	84.98360423177908	-1.3458957137253664
p-valor SW	1.4292659237220295e-57	2.4547090562374046e-58	1.672213113684245e-47	5.288641765921803e-39
KS estadística	0.8737288135593221	0.8737288135593221	0.8737288135593221	0.8737288135593221
p-valor KS	0.0	0.0	0.0	0.0
	NAICS	RelacionDeFlujoDeEfectivoOper	RotacionDeCuentasPorCobrar	RotacionDeCuentasPorPagar
sesgo_t	-0.00881978430377713	0.6494588970148814	-0.0830546523565241	-0.14388167973101348
Curtosis	-0.6381984444149928	21.905770936997342	0.9822525910536988	2.4096850587268204
p-valor SW	9.555059412388367e-22	1.2453291794697578e-38	3.5450576884821487e-10	3.692834749725391e-17
KS estadística	0.8737288135593221	0.8737288135593221	0.8737288135593221	0.8737288135593221
p-valor KS	0.0	0.0	0.0	0.0
	RotacionDeInventario	RotacionDelCapitalDeTrabajo	ValordeMercadodelEquity	CoberturaDeLaDeudax
sesgo_t	0.17716186854241378	2.3460159650527896	0.9306465949449643	-6.593763375702745
Curtosis	-1.1198618760352073	146.90144359845812	9.637631557876638	421.22357895376587
p-valor SW	9.055479961168504e-27	2.7571879610565597e-56	4.888646470997845e-34	5.664433252184349e-60
KS estadística	0.8737288135593221	0.8737288135593221	0.8737288135593221	0.8737288135593221
p-valor KS	0.0	0.0	0.0	0.0

Nota. Extracto de resultado de elaboración y ejecución propia

5.1.2.3. Estadística Descriptiva

Tabla 12 - Estadísticas Descriptivas

Número	Variable								
1	CapitalDeTrabajo								
2	RelacionDeudaEbitda								
3	RelacionDeudaCapital								
4	RelacionEfectivoTotalActivos								
5	WACC								
6	EVA								
7	GananciaPerdidaNeta								
8	CostoDeuda								
9	NAICS								
10	RelacionDeFlujoDeEfectivoOperativo								
11	RotacionDeCuentasPorCobrar								
12	RotacionDeCuentasPorPagar								
13	RotacionDeInventario								
14	RotacionDelCapitalDeTrabajo								
15	ValordeMercadodelEquity								
16	CoberturaDeLaDeudax								

	1	2	3	4	5	6	7	8
count	1180.000000	1.180000e+03	1.180000e+03	1.180000e+03	1180.000000	1.180000e+03	1180.000000	1.180000e+03
mean	4799.315703	-1.204310e-17	3.010774e-18	7.828013e-17	0.000000	9.032323e-18	0.000000	9.032323e-17
std	5684.968028	1.000424e+00	1.000424e+00	1.000424e+00	1.000424	1.000424e+00	1.000424	1.000424e+00
min	-47270.910000	-1.626928e+01	-1.499052e+01	-1.900958e+00	-22.632213	-2.565359e+01	-12.916570	-8.192985e-01
25%	1431.145000	-9.697335e-02	-1.071324e-01	-7.598951e-01	-0.179047	2.339072e-02	-0.413416	-8.192985e-01
50%	3638.835000	-7.446002e-02	-6.816053e-02	-3.441354e-02	-0.083395	1.085056e-01	-0.234242	-8.192985e-01
75%	6935.805000	7.498935e-04	3.035182e-02	7.915377e-01	0.053527	1.399889e-01	0.188421	1.013871e+00
max	40739.360000	2.278173e+01	2.468633e+01	2.232455e+00	22.111915	1.654833e+01	15.995641	1.882176e+00

	9	10	11	12	13	14	15	16
count	1.180000e+03	1.180000e+03	1.180000e+03	1.180000e+03	1.180000e+03	1.180000e+03	1.180000e+03	1.180000e+03
mean	-2.739805e-16	3.010774e-18	-2.408619e-17	3.010774e-17	9.032323e-17	6.021549e-18	-8.430168e-17	3.763468e-18
std	1.000424e+00	1.000424e+00	1.000424e+00	1.000424e+00	1.000424e+00	1.000424e+00	1.000424e+00	1.000424e+00
min	-2.287247e+00	-8.610835e+00	-2.959424e+00	-6.883933e+00	-1.198001e+00	-1.531603e+01	-9.292498e+00	-2.474989e+01
25%	-9.249987e-01	-3.378760e-01	-5.644894e-01	-5.062887e-01	-1.198001e+00	-4.943385e-02	-6.061539e-01	-1.139474e-02
50%	3.473749e-01	-1.241077e-01	-8.892370e-03	5.621018e-02	6.397876e-02	-2.403064e-02	-2.750310e-01	-1.126567e-02
75%	4.872633e-01	2.779795e-01	5.628291e-01	5.534587e-01	7.830321e-01	2.837300e-02	2.936510e-01	-9.626734e-03
max	2.275541e+00	1.060856e+01	3.473235e+00	4.425543e+00	2.434847e+00	1.717389e+01	5.896229e+00	1.820175e+01

Nota. Extracto de resultado de elaboración y ejecución propia

La Tabla 12 presenta las estadísticas descriptivas detalladas para las variables transformadas del conjunto seleccionado. Estas estadísticas reflejan un detallado análisis de las principales y más comunes métricas usadas (media, desviación estándar, mínimos, máximos y cuartiles), para comprender mejor la distribución y características de las variables mencionadas.

Mediante el uso de la clase **describe()** de la biblioteca Numpy de McKinney, W. (2022) se presentan calcularon todas las estadísticas descriptivas. Empezando con **count**, o También

interpretada como contar. la cual refleja que para todas las muestras existen la misma cantidad de datos; siendo 1180 observaciones.

Luego, se presentó la **mean()** o media, la cual representa el promedio de las observaciones para cada variable. Dentro de los valores arrojados, se presenta una diferencia significativa entre las variables independientes de la variable dependiente. Condición que aplica para todas las estadísticas como consecuencia de la transformación de Yeo, K. & Johnson, R. (2000). Esto se interpreta también a la diferencia en magnitudes y escalas.

Para la desviación estándar o std, se muestra la variabilidad o dispersión de los datos respecto a la media. Una desviación estándar más alta indica mayor dispersión. Ahora bien, recordemos que los datos se encuentran transformados (Yeo, K. & Johnson, R. 2000) y que al aplicar la transformación a los datos, especialmente cuando estos presentan heterocedasticidad, es probable que se logre una mayor estabilización de la varianza. Esto significa que la magnitud de la dispersión de los datos se reduce o se hace más uniforme a lo largo de los diferentes niveles de la variable. Por tal motivo, la varianza para los datos de las variables independientes se encuentran estabilizados, lo cual, mejora la validez de estos en el modelo y hace que su análisis sea más confiable.

También se exponen los valores de mínimos y máximos, los cuales ayudan a entender el rango dentro del cual se encuentran cada variable. Los cuartiles que se presentan son para los valores 25%, 50% y 75%, estos describen la distribución subyacente y proporcionan una visión de la distribución de los datos sin ser demasiado influenciados por valores extremos.

Respecto al *CapitaldeTrabajo*, el estadístico más relevante pudiera ser la Media=4,799.3175. lo que indica que en promedio, las medianas empresas de Bogotá mantienen un capital de trabajo positivo. Por otra parte, se evidencia una alta desviación estándar, la cual refleja la significativa variabilidad en el capital de trabajo entre las empresas que, a pesar de estar categorizadas como medianas, tienen condiciones muy diferentes de liquidez.

RotacionDeCuentasPorPagar: media = cercana a 0, lo cual es esperado en esta métrica, pues sugiere que las empresas pagan a sus proveedores dentro del mismo período de tiempo que lo generado, generalmente un año.

WACC: media y desviación estándar = media cercana a cero y desviaciones estándar bajas, lo que puede sugerir que se utilizó una buena y consistente metodología para calcular el indicador.

Estos análisis sobre las estadísticas descriptivas revelan a groso modo una variabilidad significativa entre las empresas en términos de su estructura financiera y operacional. También, las diferencias presentadas en los rangos de los cuartiles sugieren que mientras algunas empresas están financieramente más estables otras pueden estar enfrentando desafíos financieros significativos.

6. Resultados

6.1.1. Cálculo Iterativo de selección de variables

De las 106 variables disponibles para la construcción de esta regresión, se implementó en Python, mediante el uso de la biblioteca *itertools* (Van Rossum, G. ,2020), un cálculo iterativo que escogiera las variables que aportaran mejor al R^2 , y que al mismo tiempo tuviera en cuenta las consideraciones de multicolinealidad calculadas anteriormente.

Esta estrategia, poco usada en modelos e investigaciones en finanzas, donde usualmente se basan en datos o en selección de variables anteriormente usadas en otros trabajos, o en la determinación de variables con metodologías tipo regresiones *stepwise* para seleccionarlas (Pope, P. T., & Webster, J. T., 1972). Sin embargo, el uso de estas metodologías finalmente terminan siendo un problema de ordenación, lo que se ve afectado de manera significativa en el proceso cuando se cambia su ordenación, afectando así la continuidad del modelo.

Considerando el resultado obtenido, la combinación con mayor R^2 posible, que hiciera sentido al modelo fue la presentada en esta investigación.

Las independientes que le aportaron significativamente positivamente, están destacadas *RelacionEfectivoTotalActivos*, *GananciaPerdidaNeta*, *RotacionDeCuentasPorPagar*, y *ValordeMercadodelEquity*, pues son aquellas que tienen coeficientes positivos significativos. Estas indican que existe una relación directa con la variable dependiente, y que frente incrementos en estas métricas explicarían incrementos en la variable dependiente.

En contraparte, las variables significativas negativas, se obtuvieron *WACC*, *EVA*, *NAICS*, *RelacionDeFlujoDeEfectivoOperativo*, *RotacionDeCuentasPorCobrar*,

RotacionDeInventario, *RotacionDelCapitalDeTrabajo*, y *CoberturaDeLaDeudax*. Estas exponen, mediante sus coeficientes negativos, que ante un aumento en estos valores está asociado con una disminución en la variable dependiente.

Por ejemplo, para el análisis del *EVA* respecto al *CapitaldeTrabajo*, se evidencia una relación negativa entre ambas variables; esto podría interpretarse como la relación inversa entre ambas, donde, si EVA (Chen, S., & Dodd, J., 1997) aumenta se estaría sugiriendo una eficiencia operativa y uso de recursos financieros, y en consecuencia se requeriría menor inversión en capital de trabajo para mantener las operaciones de la firma.

6.1.2. Regresión Lineal Múltiple

Luego de haber corrido las diferentes iteraciones, se procedió con la ejecución de una regresión lineal múltiple, donde la variable dependiente es *CapitaldeTrabajo* y las independientes son las *variables de interés*, cuya matriz de datos contiene los valores transformados de la selección *variables_interes*.

Utilizando la biblioteca `sklearn.linear_model`, clase `LinearRegression` Pedregosa, F., et al. (2011), junto con la biblioteca `Numpy` (McKinney. 2022), se graficó la regresión para todas las variables de interés.

Tabla 13 - Regresión Lineal Múltiple

Método utilizado: Regresión Lineal Múltiple

Tamaño de las muestras: 1180

Total de observaciones: 1180

Variable	Coefficiente	Error estándar	Estadístico t	Valor p
RelacionDeudaEbitda	297.320271	0.029257	10162.340353	0.000000
RelacionDeudaCapital	39.455145	0.029257	1348.553080	0.000000
RelacionEfectivoTotalActivos	1363.545154	0.031335	43514.813143	0.000000
WACC	-730.360175	0.085867	-8505.740842	0.000000
EVA	-835.933426	0.085897	-9731.802022	0.000000
GananciaPerdidaNeta	658.047493	0.036439	18058.698666	0.000000
CostoDeuda	76.622679	0.030479	2513.946672	0.000000
NAICS	-259.498741	0.029363	-8837.621725	0.000000
RelacionDeFlujoDeEfectivoOperativo	-527.143955	0.032093	-16425.459279	0.000000
RotacionDeCuentasPorCobrar	-1614.169044	0.032424	-49783.210423	0.000000
RotacionDeCuentasPorPagar	863.919477	0.031582	27354.731586	0.000000
RotacionDeInventario	-75.548356	0.031237	-2418.538867	0.000000
RotacionDelCapitalDeTrabajo	-108.438076	0.029149	-3720.176091	0.000000
ValordeMercadodelEquity	2461.671328	0.035067	70198.922518	0.000000
CoberturaDeLaDeudax	-21.269230	0.029590	-718.798652	0.000000

Intercepto	4799.315703			

Error estándar:	4393.070430			
Coefficiente de determinación (R ²):	0.410452			

Nota. Extracto de resultado de elaboración y ejecución propia

La tabla 13 presenta los resultados de la regresión lineal múltiple corrida entre diferentes variables financieras y operacionales sobre el *CapitaldeTrabajo*. Como primer análisis, el tamaño de las muestras, el cual es para todas las variables de 1180 observaciones, demuestra ser adecuado para realizar inferencias estadísticas robustas.

Para los resultados presentados en el intercepto, se obtuvo un valor de 4799.315703, el cual, el modelo sugiere sería el valor a obtener en *CapitaldeTrabajo* cuando todas las variables independientes son iguales a cero.

Para los resultados obtenidos en los valores p de las variables seleccionadas, se puede determinar que las variables independientes en relación con la variable dependiente, son

estadísticamente significativas pues todas son mercanas a 0 o interiores a un nivel de significancia de 0.05. En otras palabras, se puede rechazar la hipótesis nula en el que cada variable independiente no tiene efecto o relación significativa con la dependiente.

Para el coeficiente de determinación o R^2 , se obtuvo un valor de 0.410452. El modelo sugiere que aproximadamente el 41% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes.

Para los valores del estadístico-t, sus resultados son extremadamente altos para casi todas las variables, indicando que los coeficientes son estadísticamente significativos y no son cero. En otras palabras, las variables seleccionadas refuerzan la confianza en la robustez dentro del modelo.

En ese sentido, la tabla 13 definió el modelo de regresión lineal múltiple de la siguiente manera:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

- *Y es CapitaldeTrabajo (dependiente);*
- *β_0 es el intercepto estimado igual a 4799.315703;*
- *$\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de las variables independientes;*
- *X_1, X_2, \dots, X_n son los valores que pueden tomar las variables independientes;*
- *ε es el valor esperado del error, igual a 4393.070430.*

Sin embargo, frente a la presencia de multicolinealidad, la cual disminuyó luego de la transformación de los datos mediante Yeo, In-Kwon & Johnson, Richard. (2000), para

Haitovsky, Y. (1969), podría seguir existiendo evidencia para suponer que la multicolinealidad sigue en las independientes. Ante esta situación, se entiende que si se mantiene una alta correlación entre dos o más variables independientes, es altamente probable que existan problemas en la estimación de los coeficientes de regresión y en la interpretación de los resultados.

Para ello, Haitovsky, Y. (1969) plantea que el uso de estrategias estadísticas para la regularización de los datos, pueden usarse las regresiones de Ridge o la regresión de Lasso; los cuales penalizan los coeficientes de regresión para reducir el efecto de la multicolinealidad. En general, este planteamiento enfatiza la importancia de identificar y tratar la multicolinealidad en el análisis de regresión para obtener resultados más confiables y precisos.

Motivo por el cual, como se presenta más adelante, se realizaron las pruebas de bondad de ajuste de los modelos de regresión de Ridge y Lasso a la data trabajada.

6.1.3. Bondad de ajuste del modelo

Usando las clases `from Ridge`, `Lasso`, `import train_test_split`, y `StandardScaler`, de la biblioteca `Sklearn` (Pedregosa, F., et al., 2011), y la biblioteca `Numpy` (McKinney. 2022), junto con otras; se estimaron ambas scores para determinar cuál de las dos regresiones se adapta mejor.

Estas puntuaciones representan la bondad de ajuste de los modelos de regresión de Ridge y Lasso en el conjunto de datos de prueba, es decir, califican cuánta varianza en la variable dependiente es explicada por el modelo (Melkumova, L. E., & Shatskikh, S. Y., 2017)

Para Melkumova, L. E., & Shatskikh, S. Y. (2017), Estos valores están en el rango de 0 a 1, donde 1 indica un ajuste perfecto del modelo a los datos y 0 indica que el modelo no explica nada de la varianza en los datos. Siguiendo las prácticas de Melkumova, L. E., & Shatskikh, S. Y. (2017) pero implementadas en Python y no en R, se utilizaron las bibliotecas de entrenamiento statsmodels.api con la clase sm, las bibliotecas de sklearn con las clases train_test_split, Ridge, Lasso, PolynomialFeatures, StandardScaler, entre otras; se corrió una prueba de bondad y estos fueron los resultados.

Tabla 14 - Bondad de ajuste Ridge - Lasso

```
Ridge Score: 0.39242808775103777
Coeficientes de Ridge: [ 360.68137614    5.13899522  1496.19023295 -1453.37093201
-1742.71009587   173.63731466   168.8527128   -273.10351707
-378.7869121   -1656.19447777   930.0809885   -125.66195802
-186.85457004   2329.4759682    -9.93032348]
Intercepción de Ridge: 4975.727531779661

Lasso Score: 0.3912409473483678
Coeficientes de Lasso: [ 360.15612134    4.28191165  1497.07669643 -1462.72047408
-1752.86123117   170.55828457   168.29596723   -272.26310193
-377.08515625   -1656.90292579   930.20537458   -124.84372248
-186.0278719    2331.92351776    -8.82999412]
Intercepción de Lasso: 4975.727531779661
```

Nota. Extracto de resultado de elaboración y ejecución propia

Para el Ridge Score se obtuvo 0.3924, indicando la capacidad predictiva que el modelo bajo Ridge sugiere tener una capacidad predictiva moderada. Para el Lasso Score, el cálculo obtenido es de 0.3912. Estos resultados sugieren que ambos modelos tienen un rendimiento similar si se usaran como factores de castigo dentro de los datos.

Estos resultados indican también que, al ser regresión regularizadas que penalizan los coeficientes grandes, estarían posiblemente castigando el sobreajuste de la transformación de Yeo-Johnson(2000) aplicada.

6.1.4. Pruebas de hipótesis.

Se realizaron pruebas de hipótesis a los datos, para probar la significancia de los datos de las variables independientes y calcular la significancia estadística de los coeficientes del modelo. Para ello se llevaron a cabo las pruebas t de los coeficientes individuales y la prueba F para estimar la significancia del modelo.

Para la realización de las pruebas se utilizaron las bibliotecas scikit-learn (Pedregosa et al., 2011), Numpy (McKinney. 2022), y SciPy (Virtanen et al., 2020); siguiendo los parámetros de prueba de Fisher, R. A., et al. (1966) para las pruebas t de Student y la prueba F de Fisher.

Adicionalmente, se utilizó como nivel de significancia α para ambas pruebas el valor de 0.05. Lo que significaría que si la probabilidad a obtener en el modelo observado es menor que 0.05, se rechaza la hipótesis nula (Fisher, R. A., et al., 1966).

Mediante el uso de las clases `X_train`, `y_train` se entrenó un modelo regresión lineal para probar las variables independientes, una a una, con la variable dependiente, utilizando como base del entrenamiento las muestras de los datos de la investigación. Utilizando la clase `LinearRegression` de la Biblioteca scikit-learn (Pedregosa et al., 2011) se corrieron los datos; los cuales están presentados en la tabla 18.1 y 18.2.

Tabla 15.1 - Prueba de Hipótesis mediante Prueba T y F

Conjunto de entrenamiento:				
	RelacionDeudaEbitda	RelacionDeudaCapital	...	CoberturaDeLaDeudax
296	-0.069552	-0.039882	...	-0.010486
599	0.121167	0.101110	...	-0.01138
711	-0.076085	-0.063904	...	-0.014328
...
1126	-0.096973	-0.107450	...	-0.011395

Conjunto de prueba:				
	RelacionDeudaEbitda	RelacionDeudaCapital	...	CoberturaDeLaDeudax
210	0.015193	0.149647	...	-0.010799
58	-0.096973	-0.107450	...	-0.011395
943	-0.096357	-0.105774	...	-0.000433
...
649	0.076416	0.439115	...	-0.011631

Valores reales:					
[4.729070e+03	2.069233e+04	4.106220e+03	2.643730e+03	4.832950e+03
	6.825490e+03	5.138200e+03	1.986910e+03	8.445600e+02	1.551420e+03

	1.849698e+04	6.001700e+02	2.548770e+03	4.802550e+03	4.955910e+03
	4.482900e+02]				

Valores predichos:					
[1974.1952833	7597.63762082	5761.99735524	5761.59052161	
	735.76920236	2790.01883347	5455.13827817	-427.31286206	
	3771.65115687	3279.9201899	6527.00270798	4449.72868022	
	
	5544.74077229	6715.43296722	7665.0013063	618.98969772]	

Error cuadrático medio:
22206213.38426181

Nota. Extracto de resultado de elaboración y ejecución propia

El resultado del entrenamiento, presentado en la tabla 18.1, obtuvo que en error cuadrático medio entre las predicciones con los valores reales del conjunto de prueba fue de 22206213.3842. Este resultado sugiere que en un modelo de regresión lineal, las variables independientes evaluadas con CapitaldeTrabajo no permitiría una buena estimación por la posible falta de valores o características relevantes en la prueba. También podría explicar que los valores de las muestras obtenidas presentan una amplia diferencia entre sí, y que se

podrían obtener predicciones no precisas. Sin embargo, gracias a los resultados de la prueba de ANOVA de la tabla 17, podemos decir que las variables seleccionadas para la predicción de Capital de Trabajo son significativas en el modelo gracias a la suma de cuadrados arroja un valor grande de $1.563982e+10$.

En la prueba F de Fisher (Fisher, R. A., et. al.,1966), se buscó probar la significancia global o total del modelo. Para ello, con el uso de la clase `f.cdf` y de los valores entrenados, se realizó el cálculo del estadístico F.

Y para la prueba de t de Student (Fisher, R. A., et. al.,1966), se buscó probar los coeficientes individuales con la variable dependiente *CapitaldeTrabajo*. Se estimaron valores de t valores p utilizando la clase `t.cdf`.

Tabla 15.2 - Prueba de Hipótesis mediante Prueba T y F

Hipótesis nula (H0):	No hay relación entre las variables independientes (variables_interes) y la variable dependiente (capital de trabajo).								
Hipótesis alternativa (H1):	Existe una relación significativa entre al menos una de las variables independientes y la variable dependiente.								
Prueba de significancia global (F-test):									
Estadístico F:	9.05289421592136								
Valor p:	0.999999999999997								
Resultado:	No se puede rechazar la hipótesis nula (H0), no hay suficiente evidencia para afirmar una relación significativa entre las variables independientes y la variable dependiente.								
Pruebas de significancia para coeficientes individuales (t-test):									
Número	Variable								
1	CapitalDeTrabajo								
2	RelacionDeudaEbitda								
3	RelacionDeudaCapital								
4	RelacionEfectivoTotalActivos								
5	WACC								
6	EVA								
7	GananciaPerdidaNeta								
8	CostoDeuda								
9	NAICS								
10	RelacionDeFlujoDeEfectivoOperativo								
11	RotacionDeCuentasPorCobrar								
12	RotacionDeCuentasPorPagar								
13	RotacionDeInventario								
14	RotacionDelCapitalDeTrabajo								
15	ValordeMercadodelEquity								
16	CoberturaDeLaDeudax								
Variable:	2	3	4	5	6	7	8	9	
Valor t:	34.6161354	0.5213449	133.1578042	-54.89073217	-64.6323258	13.5113913	15.6338088	-26.2019953	
Valor p:	0.0	0.6026508	0.0	0.0	0.0	0.0	0.0	0.0	
Resultado:	Se rechaza H0	No rechaza H0	Se rechaza H0	Se rechaza H0	Se rechaza H0	Se rechaza H0	Se rechaza H0	Se rechaza H0	
Variable:	10	11	12	13	14	15	16		
Valor t:	-32.5457849	-141.5660099	82.6370958	-11.2371227	-18.0463374	196.0035081	-0.9240955		
Valor p:	0.0	0.0	0.0	0.0	0.0	0.0	0.3564490		
Resultado:	Se rechaza H0	Se rechaza H0	Se rechaza H0	Se rechaza H0	Se rechaza H0	Se rechaza H0	No se rechaza H0		

Nota. Extracto de resultado de elaboración y ejecución propia

En la Tabla 18.2 se presentan los resultados de las pruebas de hipótesis de las pruebas t de Student y la prueba F de Fisher, las cuales se pueden interpretar de la siguiente manera.

De las pruebas t de Student, para la mayoría de las variables, se rechaza la hipótesis nula (H_0), lo que sugiere que existe una relación significativa entre esas variables y la variable dependiente (*CapitaldeTrabajo*); con excepción de las variables *RelaciónDeudaCapital* y *CoberturaDeLaDeudax*, las cuales no rechazan H_0 y sugieren que no hay suficiente evidencia para afirmar una relación significativa entre ellas y *CapitaldeTrabajo*.

De la prueba F de Fisher, ya que el valor p estimado es mayor que el nivel de significancia de α , con un nivel de confianza del 95% no se puede rechazar la hipótesis nula (H_0). Esto significa que no hay suficiente evidencia para afirmar que existe una relación significativa entre las variables independientes del modelo y *CapitaldeTrabajo*.

7. Conclusiones

La presente investigación abordó de manera exhaustiva la evaluación del capital de trabajo en medianas empresas de Bogotá, identificando las variables que son determinantes para su estimación; junto con los principales desafíos y oportunidades para mejorar la gestión financiera en este contexto.

Partiendo de la hipótesis de que una gestión eficiente del capital de trabajo puede mejorar significativamente la sostenibilidad y crecimiento de estas empresas, se empleó una metodología robusta, basada en la recopilación y análisis de datos provenientes de fuentes oficiales y estimación de ratios financieros.

Se identificó que, siguiendo una metodología poco usada en modelos financieros, a través de machine learning se logró seleccionar de un pool de 106, las variables que aportan tanto positiva como negativamente el comportamiento del *CapitaldeTrabajo*. Esto hace que esta investigación plantee una alternativa a la tradicionalidad en los procesos de finanzas, los cuales, usualmente se basan en la utilización de datos o variables de estudios anteriores, o en el uso de regresiones stepwise para la selección de las variables independientes.

Se implementó un modelo de regresión lineal múltiple para determinar los factores críticos que afectan la eficiencia del capital de trabajo. La investigación se basó en estudios anteriores como los de Jensen, M. C., & Meckling, W. (1976). Jensen, M. C. (2002). y Damodaran, A. (2009); sobre teorías de estructura de capital, maximización de las rentabilidades para los accionistas y las observaciones sobre valoración de medianas empresas.

Se validó empíricamente la relación entre una gestión efectiva del capital de trabajo y la mejora del rendimiento financiero de las medianas empresas con el desarrollo del modelo estadístico evaluado.

El modelo de regresión lineal múltiple integró variables clave relacionadas con la gestión del inventario, los plazos de cobro y pago, y otros aspectos financieros relevantes para la valoración de las empresas.

También se encontró la existencia de una relación inversa entre el CapitaldeTrabajo y el EVA Baker, H. D. (2009). El cual, sugiere que la implementación de prácticas eficientes en la gestión del capital, mejoran el valor calculado para EVA, mientras que aquellas que dejan excedentes de capital sin utilizar, o excesos de CapitaldeTrabajo, se destruye valor en la compañía.

Esto sugiere que al mejorar estos aspectos, las empresas no solo optimizan su liquidez y rentabilidad, sino que también incrementan su capacidad para crear valor para los stakeholders.

La evidencia obtenida sugiere que las medianas empresas en Bogotá pueden beneficiarse significativamente de la implementación de prácticas de gestión financiera más rigurosas y basadas en datos.

Sin embargo, visto al tejido empresarial bogotano como una singularidad, dificulta un modelo único. Esto quiere decir que, las complejidades de las diferentes industrias disminuyen la significancia del modelo estimado, y que deben ser vistas según las especificidades de las necesidades del sector o de la empresa.

El modelo estadístico proporcionado en esta tesis confirma que es capaz, con base a los resultados obtenidos, tener cierta capacidad para explicar la variabilidad en el Capital de Trabajo. Donde, mediante un coeficiente de determinación (R^2) igual a 0.410, presenta que las variables independientes del modelo, explican aproximadamente el 41% del comportamiento de la variable dependiente.

Sin embargo, dados los resultados de la prueba F y el coeficiente de determinación estimados para los valores entrenados, se sugiere que no debe ser considerado como único camino predictivo, pues es posible que se necesiten mejoras en el modelo o la inclusión de otras variables para mejorar su capacidad predictiva.

En resumen, aunque el modelo tiene cierta capacidad para explicar la variabilidad en el Capital de Trabajo, y pese a los resultados de las pruebas de hipótesis, es una herramienta valiosa para los gerentes para la toma de decisiones basadas en las industrias de Bogotá. La investigación ofrece una comprensión más profunda sobre una gestión efectiva en el capital de trabajo. De igual manera, los resultados empíricos soportan la hipótesis de que una gestión eficiente del capital de trabajo contribuye directamente a la rentabilidad y viabilidad de las medianas empresas en entornos competitivos y dinámicos como el de Bogotá.

Por último, se recomienda que las empresas adopten un enfoque proactivo en la gestión del capital de trabajo para mejorar la productividad de la capital, y que adopten más y mejores prácticas que hagan más competente y estandarizadas a las industrias del sector.

También, se ofrece el modelo desarrollado en esta investigación a la ciudadanía como base para el análisis continuo y la toma de decisiones; sin olvidar que este no es la prueba o camino único debido a los hallazgos de su ajuste. Siendo estas últimas capaces de cambiar

en el futuro si se adoptan políticas financieras robustas, enfocadas en optimizar los ciclos de efectivo y mejorar las políticas de generación para los accionistas, siempre con el soporte de análisis de datos y herramientas analíticas avanzadas.

Referencias

- Adams, R. B., & Almeida, H. (2005). Powerful CEOs and their impact on corporate performance. *The Review of Financial Studies*, 18(4), 1403-1432.
- Agresti, A. (2012). *Categorical data analysis* (Vol. 792). John Wiley & Sons.
- ANIF. (2017). Acciones de mejoramiento de las Pymes en 2016-2017. *Coyuntura Pyme*, (40-41).
- ANIF. (2019). *Gran Encuesta Pyme Nacional I-2019, lectura nacional: 1er semestre*. Bogotá, Colombia: ANIF.
- ANIF. (2021). *Gran Encuesta Pyme Nacional II-2020, lectura nacional: 2do semestre*. Bogotá, Colombia: ANIF.
- ANIF. (2019). *Los servicios financieros digitales en América Latina*. Bogotá, Colombia: ANIF para FELABAN.
- Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, 5(9), 754-767.
- Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23(1), 49-60.
- Baker, H. D. (2009). EVA revisited. *Journal of Financial Education*, 35, 1-22. Obtenido de www.jstor.org/stable/41948621

Banco de la República de Colombia. (2024). Tasas de colocación por modalidad de crédito.

1.2.6. Histórico para un tipo de cuenta periodicidad mensual. Gerencia Técnica.

Información extraída de la bodega de datos Serankua el 15/04/2024 18:20:58.

Bolsa de Valores de Colombia. (2023). Guía de preguntas frecuentes sobre a2censo

(Código: GI-PQ-EI-0007, Versión 4.0, Fecha de Vigencia: 05/07/2023). Recuperado

de <https://assets.a2censo.com/questions.pdf>

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal*

Statistical Society Series B: Statistical Methodology, 26(2), 211-243.

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: Design,*

innovation, and discovery (2nd ed.). Wiley.

Chen, S., & Dodd, J. (1997). Economic value added (EVA™): An empirical examination of

a new corporate performance measure. *Journal of Managerial Issues*, 9(3), 318-333.

Recuperado de <http://www.jstor.org/stable/40604150>

Chesher, A., & Peters, S. (1994). Symmetry, regression design, and sampling distributions.

Econometric Theory, 10(1), 116-129.

Coase, R. H. (1937). The nature of the firm. *Economica*, 4, 386-405.

Congreso de la República de Colombia. (2011). Ley 1450 de 2011, artículo 43. Recuperado

de <https://www.suin-juriscol.gov.co/viewDocument.asp?id=1680917>.

Corporación Financiera Internacional. (2010). *Guía práctica de gobierno corporativo:*

Experiencias del Círculo de Empresas de la Mesa Redonda Latinoamericana.

Washington, DC: Corporación Financiera Internacional.

- Correa R., J. S., & Murillo O., J. H. (2015). *Escritura e investigación académica: Una guía para la elaboración del trabajo de grado* (2a ed.). Bogotá, Colombia: Editorial CESA.
- Damodaran, A. (2007). *Return on capital (ROC), return on invested capital (ROIC) and return on equity (ROE): Measurement and implications*. New York, NY, United States: Stern School of Business, New York University. Obtenido de <https://ssrn.com/abstract=1105499> o <http://dx.doi.org/10.2139/ssrn.1105499>
- Damodaran, A. (2009). *Valuing young, start-up and growth companies: Estimation issues*. New York, NY, United States: Stern School of Business, New York University.
- Deloof, M. (2003). Does working capital management affect profitability of Belgian firms? *Journal of Business Finance & Accounting*, 30(3-4), 573-588.
- Dyreng, S., Mayew, W., & Schipper, K. (2017). Evidence of manager intervention to avoid working capital deficits. *Contemporary Accounting Research*, 34(2), 697–725.
- Eljelly, A. M. (2004). Liquidity-profitability tradeoff: An empirical investigation in an emerging market. *International Journal of Commerce and Management*, 14(2), 48-61.
- Fajardo, J. (2017). El mercado de capitales al servicio de las Pymes a través de un innovador modelo de negocio. *Coyuntura Pyme*, 12-17.

Fisher, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Publisher*.

Fisher, R. A. (1966). *The design of experiments* (Vol. 21). Edinburgh: Oliver and Boyd.

Greene, W. H. (2018). *Econometric analysis*. Pearson.

Haitovsky, Y. (1969). Multicollinearity in regression analysis: Comment. *The Review of Economics and Statistics*, 51(4), 486-489.

Harris, M., & Raviv, A. (2008). A theory of board control and size. *The Review of Financial Studies*, 21(4), 1797-1832.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer Series in Statistics.

Hodges Jr, J. L. (1958). The significance probability of the Smirnov two-sample test. *Arkiv för Matematik*, 3(5), 469-486.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.

Jacobson, R. (1987). The validity of ROI as a measure of business performance. *The American Economic Review*, 77(3), 470-478.

Jensen, M. C. (2002). Value maximization, stakeholder theory, and the corporate objective function. *Business Ethics Quarterly*, 12(2), 235-256.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305-360.

- Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1/2), 149-176.
- Kafeel, A. J., Ud Din, M., Waris, A., Tahir, M., & Khan, S. (2020). Working capital management and firms' profitability: Dynamic panel data analysis of manufactured firms. *Journal of Financial Risk Management*, 9(4), 1-15.
- Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione (On the empirical determination of a distribution law). *Giornale dell'Istituto Italiano degli Attuari*, 4(1-2), 83-91.
- Ministerio de Comercio, Industria y Turismo de la República de Colombia - MinTIC. (s.f.). Guía de contratación pública para micro y pequeñas empresas - MIPYME -. Recuperado el 13 de abril de 2024, de <https://www.aplicaciones-mcit.gov.co/guiapymes/c1i3.html>
- McConville, D. (1994, abril 18). All about EVA. *Industry Week*, 55-58.
- McKinney, W. (2022). Python for data analysis: *Data wrangling with Pandas, NumPy, and Jupyter* (3rd ed.). O'Reilly Media.
- Melkumova, L. E., & Shatskikh, S. Y. (2017). Comparing ridge and LASSO estimators for data analysis. *Procedia Engineering*, 201, 746-755.
- Nourayi, M. M., & Daro, F. (1996). Performance evaluation and measurement issues. *Journal of Managerial Issues*, 8(2), 206-217.

- Núñez, E., Steyerberg, E. W., & Núñez, J. (2011). Estrategias para la elaboración de modelos estadísticos de regresión. *Revista Española de Cardiología*, 64(6), 501-507.
- Oke, J., Akinkunmi, W. B., & Etebefia, S. O. (2019). Use of correlation, tolerance and variance inflation factor for multicollinearity test. *GSJ*, 7(5), 652-659.
- Pearson, K. (1892). *The grammar of science*. Adam and Charles Black.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pope, P. T., & Webster, J. T. (1972). The use of an F-statistic in stepwise regression procedures. *Technometrics*, 14(2), 327-340.
- Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, 23(4), 3-13.
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.
- Sampieri, R. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. Ciudad de México, México: McGraw Hill e Interamericana Editores, S.A. de C.V.

- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *In 9th Python in Science Conference*.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611.
- Sharma, A. K., & Kumar, S. (2011). Effect of working capital management on firm profitability: Empirical evidence from India. *Global Business Review*, 12(1), 159-173.
- SciPy. (2022). `scipy.stats.ks_2samp`. Recuperado de https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ks_2samp.html
- Smith, K. (1980). Profitability versus liquidity tradeoffs in working capital management. *Readings on the management of working capital*. West Publishing Company, New York, St. Paul.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251.
- Trefethen, L. N., & Bau, D. (2022). Numerical linear algebra. Society for Industrial and Applied Mathematics.
- Trujillo, A., Guzmán, A., & Prada, F. (2015). *Juntas directivas en el desarrollo del gobierno corporativo*. Bogotá, Colombia: Editorial CESA.
- Vance, D. E. (2002). *Financial analysis and decision making*. Reino Unido: McGraw Hill LLC.

- Van Horne, J. C., & Wachowicz, J. M. (2004). *Fundamentals of financial management* (13th ed.). Pearson Education.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., & SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261-272.
<https://doi.org/10.1038/s41592-019-0686-2>
- Walbert, L. (1994). The Stern Stewart Performance 1000: Using EVA to build market value. *Journal of Applied Corporate Finance*, 7(1), 109-120.
- Yeo, I.-K., & Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954-959.
<https://doi.org/10.1093/biomet/87.4.954>