



# ANÁLISIS DE CAPACIDADES FINANCIERAS EN ECONOMÍAS SECTORIALES PARTICIPANTES EN ECOSISTEMAS UNIVERSITARIOS DE INNOVACIÓN

*Study of economic agglomerations and sector dynamics in the town of Chapinero based on cluster analysis*

**Gustavo Andrés Romero Duque**<sup>1</sup>

Ingeniero de producción, Fundación Universitaria Los Libertadores

**Ruth Milena Suárez Castro**<sup>2</sup>

Ingeniera industrial, Fundación Universitaria Los Libertadores

**Harold Germán Rodríguez Celis**<sup>3</sup>

Administrador de Empresas, Fundación Universitaria Los Libertadores

## RESUMEN

La presente investigación busca un acercamiento inicial al estudio de contextos en los ecosistemas universitarios de innovación, específicamente analizando un actor importante: el tejido productivo que tiene estrechos lazos con Instituciones de Educación Superior (IES). Para lo anterior, se determinó analizar desde factores financieros, empresas que establecieron actividades de relacionamiento con IES durante el periodo 2013 a 2017.

Se hace necesario empezar a estudiar estas empresas; ver su representatividad financiera y considerar inicialmente su enclave territorial mediante técnicas de clustering particionado, combinando Análisis de Componentes Principales (PCA) y método de K-medias, las cuales son técnicas de análisis multivariado para datos con múltiples variables continuas y que permiten medir el comportamiento de dinámicas empresariales generadas desde las diferentes empresas, en este caso, aquellas que tienen relación o convenios con IES.

Para lo anterior, se desarrolló un estudio piloto desde una IES, la Fundación Universitaria Los Libertadores. Encontrándose alrededor de 216 empresas sobre las que, mediante técnicas de análisis multivariado, se examinaron diversas cuentas de sus Estados financieros para lograr definir la existencia y tipologías de cambios financieros entre 2013-2017 y si existe relación con la ubicación espacial entre actores.

**PALABRAS CLAVE:** Clustering, Dinámicas empresariales, Ecosistemas de innovación, Estrategia, PCA.

## ABSTRACT

This research seeks an initial approach to the study of contexts in university innovation ecosystems, specifically analyzing an important actor: the productive fabric that has close ties with Higher Education Institutions (IES). For the above, it was determined to analyze from financial factors, companies that established relationship activities with IES during the period 2013 to 2017.

<sup>1</sup>Gustavo Andrés Romero Duque, Ingeniero de producción Fundación Universitaria Los Libertadores, FULL Colombia garomerod@libertadores.edu.co

<sup>2</sup>Ruth Milena Suárez Castro, Ingeniera industrial Fundación Universitaria Los Libertadores, FULL Colombia rmsuarezc@libertadores.edu.co

<sup>3</sup>Harold Germán Rodríguez Celis, Administrador de Empresas Fundación Universitaria Los Libertadores, FULL Colombia hgrodriquezc@libertadores.edu.co

It is necessary to start studying these companies; see its financial representativeness and initially consider its territorial enclave using partitioned clustering techniques, combining Principal Component Analysis (PCA) and K-means method, which are multivariate analysis techniques for data with multiple continuous variables and that allow behavior to be measured of business dynamics generated from different companies, in this case, those that have a relationship or agreements with IES.

For the above, a pilot study was developed from an IES, the Los Libertadores University Foundation. Finding around 216 companies on which, using multivariate analysis techniques, various accounts of their financial statements were examined to be able to define the existence and typologies of financial changes between 2013-2017 and if their relationship with the spatial location between actors could be configured.

Getting to know these aspects would allow to contribute in the design of strategies aimed at providing the necessary local conditions to promote the territory and its economic orientations in an advantageous way and that better than the HEIs can be participants in the study of territorial dynamics, the improvement of methods and techniques of territorial recognition and the analysis of its behavior, especially for the generation of purposeful scenarios for the city.

**KEYWORDS:** Clustering, Business dynamics, Innovation ecosystems, Strategy, PCA

**SUMARIO:** 1. Introducción. 1.1. Revisión de literatura 1.2. Propósito de la investigación 2. Método. 2.1.1. Estrategias metodológicas. 2.1.2. Población y muestra. 2.1.3. Instrumentos de recogida de información. 2.1.4. Procedimiento. 3. Resultados. 4. Discusión y conclusiones. 5. Referencia bibliográficas.

**CONTENTS:** 1. Introduction. 1.1. literature review 1.2. Purpose of the investigation 2. Method 2.1.1. Methodological strategies. 2.1.2. Population and sample. 2.1.3. Information collection instruments. 2.1.4. Process. 3. Results. 4. Discussion and conclusions. 5. Bibliographic reference.

Traducción de Gustavo Romero (Profesional en Ingeniería de producción – Fundación Universitaria Los Libertadores

## 1. INTRODUCCIÓN

### 1.1. REVISIÓN DE LITERATURA

Ecosistemas de innovación universitaria

La literatura existente sobre los ecosistemas de innovación universitaria es extensa. Para comprender de mejor manera la producción intelectual en el campo de interés particular de este estudio es fundamental organizar la literatura encontrada a partir de los distintos enfoques de análisis en los que se basan diversidad de autores. Dicha agrupación es un esfuerzo con la intención de explicar la variedad de tendencias que se han desarrollado en los últimos años. En este documento agrupamos estos enfoques de acuerdo con las metodologías tendientes a comprender la influencia de los ecosistemas de innovación universitarios en el territorio de su influencia a nivel empresarial.

El proceso de exploración bibliográfica identificó al menos cinco perspectivas que utilizan las redes de interacción para la innovación en universidades: colaboración de los intermediarios universitarios para el fomento de la innovación, interacción entre universidades y los conglomerados industriales para la transferencia de conocimiento a través de mecanismos de innovación con sectores periféricos, la influencia de las prácticas organizativas en la transición a una triple hélice híbrida, la simbiosis entre la universidad y el distrito de innovación en el contexto de la creación de lugares y por último, la función que desempeña el gobierno en una parque científico universitario.

La perspectiva de colaboración de los intermediarios universitarios para el fomento de la innovación se centra en los roles que juegan las organizaciones intermediarias de innovación, para las universidades que

desean aumentar sus solicitudes de nuevas creaciones de la propiedad industrial. Se han publicado artículos bajo este enfoque, como el realizado por Temel et al. (2021) en los países de Turquía, Croacia y Reino Unido, con el propósito de analizar para distintas formas de industria, los mecanismos de colaboración académica, el papel del intermediario universitario, el papel de las organizaciones, para una economía emergente en el contexto de la actividad de solicitudes de patentes. Para dicha investigación Temel y otros, desarrollaron dos modelos diferentes, a través de los cuales, analizaron las solicitudes de patente y las patentes registradas universitarias. Para su desarrollo, analizaron por departamento el registro de patentes de los últimos cinco años, en los que estudiaron las variables de solicitudes de patente, patentes registradas, el tener oficina de patentes, tener oficina de transferencia tecnológica, tener política pública definida de propiedad intelectual, la interacción con la región geográfica y el tipo de organización intermediaria entre los centros de Colaboración Universidad-Industria y la oficina de licencias tecnológicas que tiene la universidad.

Los resultados de los análisis de esta investigación condujeron a demostrar, que los mecanismos formales de gestión de los centros de intermediación universitaria tuvieron una influencia significativa en el desempeño en innovación académica de las universidades. Sin embargo, la influencia de la implementación de la regulación y el clima de innovación en el rendimiento de la innovación no fue significativo. Por lo tanto, se rechazó la posibilidad de que la implementación de la regulación en este tipo de centros en las universidades afecta el desempeño de la innovación académica de las mismas. Igualmente, no se logró confirmar que el clima de innovación en las universidades afecte

positivamente el desempeño en innovación académica de las respectivas universidades. Los otros resultados demostraron que los mecanismos formales de gestión de los centros de intermediación aduanera tuvieron una fuerte influencia en el desempeño de la innovación académica en las universidades subvencionadas por estos centros. Además, los resultados sugirieron que las universidades subvencionadas por los centros de intermediación tenían más ventajas para desarrollar sus programas y así mejorar su innovación académica.

Por último, se concluyó en este estudio, que las universidades subvencionadas por los centros de intermediación universitaria tienen más ventajas en el desarrollo de su contexto y mejora del rendimiento de la innovación académica. El mecanismo de la gestión formal de los centros de intermediación se convierte entonces en el factor más importante y crucial para mejorar el desempeño en innovación académica. Del mismo modo se logró demostrar, que la contratación de más expertos de la industria para buscar socios potenciales para los centros de intermediación podría facilitar una mayor interacción entre las universidades y la industria (Temel et al., 2021).

Lo anterior concuerda con el hallazgo de Zuniga (2011) quien identificó que, en las economías desarrolladas, el tamaño y el crecimiento de los vínculos entre la industria y la universidad también ha significado una proliferación de esquemas y tipos de actividad colaborativa, especialmente en países con un sistema de gobernanza y gobierno más descentralizado. Otros estudios han demostrado que las organizaciones intermediarias universitarias han facilitado el proceso de relacionamiento con empresas locales y nacionales como lo afirma (Rothaermel et al., 2007).

En relación con la perspectiva de interacción entre universidades y los conglomerados industriales para la transferencia de conocimiento a través de mecanismos de innovación con sectores periféricos, Fernández-Esquinas et al. (2016) exploraron la combinación de mecanismos utilizados por las empresas en Andalucía, región periférica de España, cuando tratan de interactuar con las universidades, con el propósito de identificar de qué manera contribuye a la transferencia de conocimiento, la relación entre universidades empresas y empresas cuando se utiliza un análisis de la combinación de mecanismos de transferencia, mediante un sistema de innovación periférico.

Para el abordaje de su estudio con este propósito, los autores agruparon la información en cuatro categorías principales: las características de la empresa, la capacidad de innovación y actividades, los vínculos con el sector universitario, por último, la ubicación geográfica y relaciones de empresa. Para identificar los determinantes de cada tipo del mecanismo de interacción, se aplicaron procedimientos multivariados a fin de encontrar las dimensiones de las interacciones universidad-industria.

Los resultados permitieron a los autores concluir, que cuando se visualizan diferentes formas en que las empresas de una periferia interactúan con las universidades, se observan que, en los sistemas de innovación general, surge una amplia gama de interacciones. La más común son las que se relacionan

con la formación y el intercambio de recursos humanos, así como trabajos de consultoría. El desarrollo de los proyectos de I + D y el uso de instalaciones universitarias son menos habituales. Las actividades menos frecuentes son la explotación de patentes, la creación de spin-off y participación en joint ventures. Estos resultados apoyan la tesis de que las universidades son una fuente importante de información tácita de conocimiento, a través de la formación de los recursos humanos necesarios, por lo que la contratación de trabajadores calificados para mejorar el uso de los resultados a partir de proyectos colaborativos y trabajos de consultoría es de vital importancia. La principal contribución de este estudio es el desarrollo de una tipología de empresas que proporciona una serie de implicaciones teóricas y políticas para las interacciones universidad-industria (Fernández-Esquinas et al., 2016).

Lo anterior apoya los estudios de Cooke & Piccaluga (2004) quienes habían encontrado que los sistemas periféricos de innovación tienen mayores complejidades para transformar la investigación y el desarrollo en productos objeto de propiedad industrial, por lo que las universidades debían esforzarse para lograr mayores beneficios económicos. Los autores afirmaban entonces, que este problema se presentaba en algunos ecosistemas de innovación de países desarrollados, haciendo que las regiones disminuyeran sus posibilidades en términos de capacidades de absorción y en generación comercial como consecuencia de la innovación. Los argumentos que encontraron para este comportamiento lo explicaron identificando que las industrias de este modelo se caracterizan por concentrar bajas y medias tecnologías, las cuales, tienen un rol preponderante en los procesos de innovación, particularmente en términos de empleo y producción.

El ecosistema entonces se determina como un escenario compuesto principalmente por pequeñas empresas y con pocos grandes usuarios de tecnología en los mercados locales, haciendo que el sector público domine la inversión que se realice en I + D visibles en instalaciones científicas, lo que dificulta en estos sistemas periféricos, la transferencia de los resultados de investigación al sector productivo y el mercado como productos intangibles de propiedad intelectual. Este fenómeno típico de la innovación universitaria en contextos europeos se identifica según Dosi et al (2006) como la conocida paradoja europea, cuando describen que los países de la Unión Europea desempeñan un papel de liderazgo mundial en términos de producción científica de alto nivel, pero se quedan atrás en la capacidad de convertir esta fortaleza en innovaciones generadoras de riqueza. Este comportamiento los autores lo explican como debilidades significativas presentes en el sistema europeo de ciencia, en el que se encuentra una investigación rezagada con respecto a los EE. UU. en varias áreas y, en segundo lugar, una industria europea relativamente débil, caracterizada en promedio por la presencia de sectores basados en las TIC y las biotecnologías, los cuales tienen una menor propensión a innovar, y por una relativa escasa participación en oligopolios internacionales en muchas ocupaciones de los procesos de innovación.

Otra de las perspectivas se basa en la influencia de las prácticas organizativas que tienen los actores de la

innovación en la transición hacia un sistema de triple hélice híbrido. Investigadores como Sarpong et al (2017) han elaborado desarrollos teóricos bajo este enfoque a partir de trabajos exploratorios, con el propósito de identificar cómo las prácticas organizativas de diversas esferas institucionales (universidades, empresas y gobierno) facilitan o impiden la transición al modelo de innovación híbrido de triple hélice en el contexto de Malasia. Para esto, llevaron a cabo una investigación de corte cualitativo en la que participaron 27 actores estratégicos, 12 de universidades de Malasia intensivas en investigación, 9 organizaciones consideradas motores de la industria, y 6 altos funcionarios gubernamentales. Los participantes de la investigación habían pasado un promedio de 10 años trabajando en su institución de afiliación. El trabajo cualitativo se basó en la discusión individual de su trabajo diario y el papel de la organización para cual trabajaban en relación con la manera cómo apropiaban y establecían la política nacional de innovación de Malasia. Los resultados permitieron identificar que Malasia todavía se encuentra en un proceso de transición para lograr una triple hélice completa. De igual manera hallaron que si bien las tres esferas institucionales están comprometidas a lograr un modelo híbrido de triple hélice, se encuentran en prácticas que impiden su capacidad para organizar la distribución de sus capacidades y recursos para apoyar sus iniciativas de innovación. Además, sus prácticas organizativas demuestran esquemas diferente acerca de cómo logra fortalecer sus colaboraciones y alianzas.

El enfoque de triple hélice es un modelo sucesivo de las relaciones de colaboración entre los tres eslabones institucionales tradicionales como son las universidades, la industria y el gobierno persigue la innovación a través de la interacción de todos. El modelo incorpora según Etzkowitz (2003) un proceso donde el gobierno planea y dirige la relación entre la industria y la academia en busca de innovación. La industria por su parte desarrolla el papel principal de la innovación, mientras que la función de la universidad se limita a la docencia y la investigación por lo que tiene pocos incentivos de participación, lo que hace que, el conocimiento generado por éstas sea limitado ya que la academia y la investigación suelen estar desconectadas de las necesidades de la industria.

El modelo se transformó para convertirse en un enfoque en el cual, las universidades, el gobierno y las empresas trabajan independientemente en ámbitos institucionales distintos. En este, el gobierno se limita a abordar las fallas del mercado, mientras que las universidades se dedican a la investigación básica y formación de mano de obra. Entre tanto, el modelo híbrido de triple hélice, que representa el último enfoque de esta perspectiva, posee un interés constante de adopción crece en el mundo, debido a que reúne combinaciones de sus dos antecesores, destacándose en virtud que se configura como una red que fomenta el movimiento hacia relaciones de colaboración y vinculante entre los tres grandes esferas institucionales sumado a otras organizaciones diversas y de distintas disciplinas en las que la política de innovación es el resultado de sus interacciones en lugar de estar organizado y planeado por el gobierno, de tal modo que, de acuerdo con Mueller (2006) se impulsa la capitalización del conocimiento en el sentido que las universidades tienen la posibilidad de asumir un papel generativo en la dirección regional desarrollo económico a través de actividades de emprendimiento

académico que comparten características comunes con los roles tradicionales de la industria y el estado en la regulación económica.

Una cuarta perspectiva es la que hace referencia a la simbiosis entre la universidad y el distrito de innovación en el contexto de la creación de lugares. Bajo este modelo los distritos orientados a la innovación se configuran bajo un nuevo tipo de uso del suelo, donde los actores públicos y privados trabajan para fomentar, atraer, y retener la inversión y el talento que tienen por objeto revitalizar las zonas urbanas e impulsar las actividades de la economía del conocimiento y la innovación. Este modelo se ha convertido en una estrategia urbana popular en todo el mundo. Frecuentemente los distritos se ubican junto con las universidades para actuar como un ancla fundamental en sus matrices socioespaciales y operativas. Esta es una tendencia aún poco estudiada (Pancholi et. al, 2020).

Para su comprensión Pancholi y otros, desarrollaron un estudio de corte cualitativo en el cual, analizaron cómo las partes interesadas perciben la innovación en los distritos de Australia. Los resultados de su investigación les permitió identificar que las universidades desempeñan un papel fundamental como facilitadores de la creación de lugares a través de la integración social en los distritos de innovación; además, encontraron que las universidades actúan como plataformas para la colaboración y el intercambio de conocimiento mediante la adopción de enfoques colaborativos; adicionalmente, hallaron que las universidades ayudan a desarrollar un sentido de la confianza en la comunidad y el establecimiento de distritos de innovación haciéndolas más, cohesionadas y prósperas; Por último llegaron a la conclusión, que las universidades contribuyen a la madurez sociocultural y organizativa de los distritos de innovación.

En relación con la perspectiva del rol que desempeña el gobierno en un parque científico universitario, Yang et al (2021) estudiaron los factores clave que afectan la estabilización del ecosistema de innovación verde en el gobierno de China y los efectos de las políticas de regulación ambiental. Para el desarrollo de su trabajo los autores emplearon la teoría basada en juegos con el fin de revelar la estrategia del juego y los cambios ocurridos durante el proceso de innovación ecológica, además de examinar el mecanismo de juego evolutivo de la regulación ambiental en las alianzas de innovación ecológica entre la universidad y la industria. En el proceso metodológico utilizaron la simulación con el fin de afectar los parámetros que le permitiera simular los factores clave que inciden en la estabilidad del ecosistema de innovación verde a través de un modelo basado en múltiples agentes. De este modo, los investigadores construyeron un modelo de optimización de juego evolutivo para analizar los impactos de la colaboración gobierno-universidad-industria en un ecosistema de innovación verde basados en la teoría de basado en la teoría de juegos evolutivos.

Los resultados permitieron identificar que, en el ecosistema de innovación verde, los grupos industriales y los grupos universitarios eventualmente evolucionarán hacia las estrategias de Innovación colaborativa, aunque la probabilidad es baja para poder evolucionar a un estado estable evolutivo ideal. Otro resultado se

relaciona con el mecanismo de mercado, donde mayores costos por defecto y la relación de distribución de los costos de I + D junto con los beneficios de la innovación verde, son factores clave que afectan la estabilización del ecosistema de innovación verde y pueden evitar que los grupos industriales y universitarios elijan una estrategia de alianza de traición. El estudio realizado los llevó a concluir que si el gobierno chino introduce regulaciones ambientales más estrictas, los grupos industriales y los grupos universitarios eventualmente realizarán colaboraciones en actividades de innovación verde, y viceversa, además concluyen que, la adopción de una combinación de políticas de regulación ambiental será la mejor opción para promover más grupos industriales y grupos universitarios para compartir sus diversas capacidades y recursos en el ecosistema de innovación verde estudiado.

#### Técnicas de análisis de aglomeraciones

Existen diversas técnicas del análisis multivariado que pueden aplicarse para el estudio de agrupamientos industriales (entre pares) conocidas como análisis de conglomerados o análisis clúster, actualmente se consideran técnicas propias de la minería de datos que para identificar y estudiar grupos de objetos con características similares.

Pueden utilizarse dos estrategias para realizar este tipo de análisis: Clúster jerárquico, utilizado para identificar grupos de observaciones similares en un conjunto de datos y el particionamiento en clúster conocido como algoritmo k-means, diseñado para dividir un conjunto de datos en varios grupos previamente seleccionados. Las dos estrategias utilizan previamente el Análisis de Componentes Principales (PCA) cuando se tiene un conjunto de datos con múltiples variables no correlacionadas, eliminando el ruido que puede conducir a una agrupación más estable y se facilita la interpretación en el análisis de la similitud de variables (Kassambara, 2017).

#### Métodos de agrupación Jerárquico y K-Medias

Son métodos exploratorios cuya finalidad es identificar agrupaciones dentro de un conjunto de elementos, donde los elementos de cada grupo presentan similitudes (Gil, 2018), grupos homogéneos al interior del clúster (Avendaño, Avendaño, William & Cárdenas, 2014). A estos métodos también se les conoce como algoritmos de clasificación y se dividen en métodos de clasificación jerárquica y métodos no jerárquicos (De la Fuente Fernández, 2011), están incluidos entre las técnicas de machine learning y su forma de cálculo no se desarrolla sobre el cumplimiento de supuestos de linealidad, ni simetría de las variables (Marín, 2014), es posible desarrollar los métodos aun cuando se tengan variables de tipo categórico y permite aplicar diferentes técnicas para la estimación de la matriz de distancias.

Los métodos de clasificación jerárquica se clasifican a su vez en aglomerativos o ascendentes y divisivos o descendentes. La característica principal de los métodos aglomerativos o ascendentes consiste en que utilizan un criterio de agrupación que va agrupando elementos en cada paso hasta llegar a una agrupación que engloba a la totalidad de elementos. En el caso de los métodos divisivos o descendentes se toma a la totalidad de

elementos en una sola agrupación y de acuerdo con el criterio de agrupación definido se divide dicha agrupación en grupos de menor tamaño, siendo los más utilizados: Método de la distancia mínima (vecino más próximo), Método de la distancia máxima (vecino más lejano), Método de la distancia media (vinculación intergrupos), Métodos basados en el centroide, Métodos basados en la mediana (Muñuzuri & Lara, 2014) y Método de Ward (Gallardo, 2013).

En cuanto a los métodos de clasificación no jerárquicos, el más utilizado es el método de agregación alrededor de centros móviles o método K-medias. Es un método de agrupación de individuos no de variables, de acuerdo con las distancias que existen entre ellos. Para su desarrollo es necesario definir un valor de k grupos (centroides), el cual se puede determinar a priori o se puede calcular a través del método del codo. Conociendo el número de K clústeres se procede a clasificar a los individuos, según con la distancia que existe entre cada individuo y su centroide más próximo (Marín, 2014), siendo que los elementos o individuos no pueden pertenecer a más de un grupo (Gil, 2018).

El método general bajo el cual funcionan consiste en los siguientes pasos: Seleccionar K medias iniciales, siendo K el número de clústeres deseados, luego, asignar cada observación al clúster que le sea más cercano, proceso que se realiza de manera automática. Posteriormente, se reasigna o reubica cada observación a uno de los K clúster de acuerdo con alguna regla de parada, en seguida se detiene el proceso de reubicación si no hay reasignación de los puntos o si la reasignación satisface la regla de parada (Carrasco, 2020).

Este tipo de métodos se han desarrollado a partir de técnicas de modelado resultantes de integrar la estadística con aprendizaje automático, con la intención de aplicarlos en sistemas reales, por ejemplo, en sectores económicos, obteniendo como resultado el convertir datos disponibles de un sistema en nuevo conocimiento, facilitando la comprensión profunda del comportamiento del sistema y también la toma de decisiones. Lo anterior, encaminado a la generación de estrategias de relacionamiento y generación de alianzas con las partes interesadas que permitan beneficios mutuos como pueden ser el incremento de los ingresos, la reducción de costos, la reducción de riesgos, entre otras posibilidades (Bucheli & Thompson, 2014).

#### Importancia del análisis de componentes principales PCA en el análisis clúster

El análisis de componentes principales PCA es una técnica estadística y de aprendizaje automático para datos multivariados que juega un papel de importancia en la exploración de la estructura subyacente de un conjunto de datos disponibles dados, y ha venido usándose de manera amplia para analizar una gran variedad de áreas de ingeniería y ciencia. Fue creada por Pearson y Spearman en 1901 y desde entonces ha sido utilizada como una herramienta de solución de problemas en diferentes campos de aplicación, como, por ejemplo: análisis de señales, meteorología e investigaciones atmosféricas y análisis semánticos latentes (Chávez-Chong, Sánchez-García & DelaCerde, 2015).

Se considera un método de análisis factorial, que analiza las correlaciones existentes entre variables e identifica estructuras de datos básicas. Es un método de extracción de factores de un conjunto de datos iniciales a fin de reducir las dimensiones del conjunto de variables en estudio y que se utiliza cuando las variables presentan correlación entre sí. Como resultado del PCA, se obtienen componentes no correlacionadas, donde cada componente es una combinación lineal de los datos de las variables de origen y donde las variables del estudio ya reducidas se representan a través de vectores (Solano, Alandete & Viloria, 2019).

Aplicar PCA junto con análisis clúster facilita la visualización e interpretación de los datos. Esta técnica genera análisis netamente descriptivos, que obtienen una representación con reducción de las dimensiones de un conjunto de datos y por eso se considera un procedimiento inicial que ayuda a eliminar el ruido de los datos al equilibrar el efecto de varios grupos de variables.

Visualizar los datos en el mapa del PCA, es apropiado para interpretar fácilmente los Clústeres generados por partición, como es el caso del algoritmo K-medias. La ventaja de la representación en el mapa es que se puede realizar con un número alto de grupos, de tal forma que los individuos son coloreados según el clúster al que pertenecen, teniendo en cuenta los centroides con los que se realiza la agrupación, así como también son representadas las variables según la tendencia de los componentes principales, todo en el mismo marco de referencia, el gráfico Biplot.

El biplot es un gráfico en el cual se representan las observaciones o individuos en posiciones dadas por las primeras dos componentes principales, a su vez sobre el mismo plano se representan superpuestos los puntos que representan a las variables descritas en las tablas en cada columna de la matriz de datos X y se ubican en posiciones que permiten interpretar relaciones existentes entre las variables y los individuos. Este tipo de gráfico se puede describir como un plano factorial que refleja en una dimensión reducida las variables de mayor importancia que provienen de un conjunto de datos organizado en forma matricial. Su principal característica radica en que logra una representación conjunta; donde aparecen superpuestos en el mismo gráfico los puntos fila o individuos y los vectores columna o variables, situación que lo hace útil para los análisis clúster y para la representación de componentes principales.

#### Software R y Paquete FactoMineR

R es un lenguaje de alto nivel que fue desarrollado en los laboratorios Bell (González Tagle, 2016), para realizar cálculos estadísticos integrando variadas técnicas estadísticas como son pruebas, modelos, análisis y gráficas; es un software libre que permite la gestión de datos y en el cual los usuarios pueden colaborar para ampliar su funcionamiento, a través del entorno de análisis y programación con que cuenta el software.

Entre las principales características que posee R se encuentran: existencia de un lenguaje de programación simple, capacidad de almacenamiento y manejo de datos, existencia de operadores para cálculo sobre variables y matrices, colección de herramientas para realizar el

análisis de los datos, gráficas para favorecer el análisis de datos, la posibilidad de realizar la programación de bucles para análisis de conjuntos sucesivos de datos y la posibilidad de descargar variados paquetes para el análisis de datos (González Tagle, 2016).

FactoMineR es un paquete de R usado para realizar análisis de datos exploratorios multivariados, ha sido desarrollado por François Husson, Julie Josse, Sébastien Lê, d'Agrocampus Rennes y J. Mazet, en el Applied mathematics department del Agrocampus Rennes University France IRMAR. El paquete realiza una amplia variedad de análisis de datos, como Análisis de Componentes Principales (PCA), análisis de correspondencia (CA) y análisis de correspondencia múltiple (MCA), Análisis factorial de datos mixtos (FAMD), Análisis factorial múltiple (MFA), Clustering jerárquico sobre componentes principales (HCPC) y Clustering particionado.

Las características más sobresalientes del paquete están en la posibilidad de tomar en cuenta diferentes tipos de variables (cualitativas y cuantitativas), así como la posibilidad de tomar diferentes tipos de estructura de datos (partición de variables, jerarquización de variables y partición de individuos) (Husson, Josse, Lê, Mazet, 2008).

Cuando los individuos están representados con variables cuantitativas el análisis multivariado requerido es PCA, en el caso en que los individuos del estudio estén representados con dos variables categóricas organizadas en tablas de contingencia, se requiere el uso de Análisis de correspondencias (CA) y, por último, cuando las variables se representan con variables categóricas se requiere el Análisis de correspondencia múltiple (MCA).

El paquete Factoextra, es el paquete complemento de FactoMineR que permite la extracción y visualización de los resultados cuando se aplican análisis exploratorios en datos multivariados. Sus principales usos se dan cuando se requiere visualizar los resultados de análisis de reducción de dimensiones, cuando es necesario extraer datos de resultados de análisis de reducción de dimensiones, cuando es importante la visualización para el análisis de agrupamiento.

Las ventajas de la visualización de los datos en las técnicas de análisis multivariado radican en la facilidad para el reconocimiento de características y las relaciones existentes entre los datos y la identificación de los patrones o tendencias emergentes del estudio. La visualización a partir del paquete Factoextra se da a partir de la representación de los datos, junto con sus atributos y relaciones a elementos gráficos como son puntos, líneas, formas y colores.

## 1.2. PROPÓSITO DE LA INVESTIGACIÓN

La presente propuesta busca aportar sobre el análisis de contextos en ecosistemas de innovación universitarios al revisar como son los comportamientos presentados en empresas que pertenecen o participan en estos ecosistemas y que están asentadas territorialmente en cercanías a estas Instituciones.

A partir del uso de técnicas de clustering particionado, se realiza un análisis de las dinámicas económicas que presentan estas empresas en un territorio durante un periodo de tiempo mediante el análisis de sus capacidades financieras y, aunque existen diferentes trabajos que estudian las dinámicas territoriales, estos se centran más en aspectos de los factores productivos, del mercado o de políticas económicas (Fujita & Thisse, 1996) (Krugman & Venables, 1996) (Manrique, 2006) (Bernal, 2013).

Debido a los retos que han tenido que afrontar durante los últimos años, los generadores de estas unidades productivas [la mayoría son Pequeñas unidades dedicadas desde desarrollo, fabricación y comercialización de productos y servicios], han tenido no solo que atreverse a generar dinámicas de desarrollo de propuestas de valor para los mercados, sino entender y desarrollar procesos de acercamiento e integración con otras unidades territoriales, caso específico, con las IES. Conllevando a afrontar no sólo desafíos operativos, tácticos o estratégicos sino también, desafíos en el tiempo; siendo con esto posible, que se estén generando condiciones para ser eficientes en términos de las dinámicas que impone el mercado para la realización de los bienes y servicios.

## 2. MÉTODO

### 2.1 ESTRATEGIAS METODOLÓGICAS

La presente investigación, tiene definido su inicio desde los enfoques exploratorio y correlacional (Hernández, 2014) (Hurtado, 2000).

La metodología para evaluar los comportamientos de las empresas que participan del ecosistema de innovación de una IES, incluye una etapa de búsqueda de información financiera en bases de datos especializadas, para su organización y análisis a nivel de micro datos de cuentas de varios estados financieros para los años 2013 a 2017 sobre las que, mediante técnicas de análisis multivariado, se examinaron diversas cuentas de sus Estados financieros para lograr definir la existencia y tipologías de cambios financieros entre esos años.

Se hace necesario empezar a estudiar estas empresas; ver su representatividad financiera y considerar inicialmente su enclave territorial mediante técnicas de clustering particionado, combinando Análisis de Componentes Principales (PCA) y método de K-medias, las cuales son técnicas de análisis multivariado para datos con múltiples variables continuas y que permiten medir el comportamiento de dinámicas empresariales generadas desde las diferentes empresas, en este caso, aquellas que tienen relación o convenios con IES.

En los métodos relacionados con el agrupamiento jerárquico, los grupos se construyen dividiendo incrementalmente los patrones utilizando un enfoque de arriba hacia abajo o de abajo hacia arriba. Existen 2 maneras de llevar a cabo el método jerárquico, el primero es, el aglomerativo, y el segundo, es el agrupamiento jerárquico divisivo. El aglomerativo se utiliza con un enfoque de abajo hacia arriba, en el que se construye grupos iniciando con un solo objeto y luego fusionando

estos grupos atómicos en grupos más grandes, hasta que todos los objetos finalmente se encuentran en un solo grupo, o de otra manera, hasta que ciertos grupos cumplan las condiciones de terminación. De manera diferente, el agrupamiento jerárquico divisivo utiliza el enfoque de arriba hacia abajo, a partir del cual, se divide el grupo que contiene todos los objetos en grupos más pequeños, hasta que cada objeto forma una agrupación por sí solo o hasta que satisfaga determinadas condiciones de terminación (Roux, 2018).

El análisis clúster es una técnica de análisis multivariado con capacidad para reducir una gran cantidad de información y organizarla en clústeres que comparten patrones de comportamiento similares. Se considera una técnica exploratoria y esta categorizada como aprendizaje no supervisado, en razón a que, para el proceso de clasificación no utiliza ningún modelo estadístico (Castro Heredia, Carvajal Escobar, & Ávila Díaz, 2012). El análisis clúster se clasifica en técnicas particionales y técnicas jerárquicas.

Por su parte, el clustering basado en particiones o particional como se le conoce comúnmente, es totalmente opuesta a la agrupación jerárquica según Saxena et al. (2017), de manera específica en variantes de K-medias, donde la selección inicial del centro del grupo es un punto importante y crucial. La dependencia del grupo final se basa totalmente en los centros del grupo inicial; por lo tanto, este proceso está delineado para ser más significativo en toda la operación de agrupamiento (Patibandla & Veeranjanyulu, 2020). Como lo argumentan estos autores, los métodos de agrupación en clúster basados en conjuntos resuelven los desafíos de los métodos basados en particiones.

La técnica de clustering particional utiliza dos métodos, el primero de ellos es el de agrupación difusa de C-medias, un método de agrupamiento que permite que un punto pertenezca a dos o más clústeres a diferencia de K- significación en los que solo se asigna un clúster a cada punto. El segundo de ellos es el de agrupación de K-medias o algoritmo K-means, uno de los algoritmos de agrupación en clúster más sencillos, con evaluación comparativa que se aplica principalmente para resolver los problemas de agrupamiento. En este procedimiento, el conjunto de datos dado se clasifica por medio de un número de clústeres definido por el usuario, K (Herreros, 2020).

El análisis clúster particional requiere iniciar con la determinación del número de clústeres a conformar, una forma de lograrlo es mediante la aplicación de la técnica del codo, siendo el algoritmo de K medias el tipo de análisis clúster más conocido. El análisis jerárquico utiliza funciones de enlace para organizar los clústeres, iniciando por considerar n grupos unitarios y posteriormente fusionando dichos grupos hasta lograr un solo grupo con todos los individuos.

El resultado esperado consiste en identificar la conformación de aglomeraciones derivadas de las similitudes presentadas en su comportamiento financiero, evaluando cuatro variables que corresponden a cuentas de sus estados financieros: Utilidad neta, activo, pasivo y patrimonio, todo esto entre las empresas de los sectores estudiados en la localidad de Chapinero en cinco años diferentes. Este análisis, permite encontrar y examinar

la dinámica sectorial de la localidad de Chapinero entre los periodos de estudio.

### **2.1.2. Población y muestra**

Para el desarrollo de la investigación, se tomó como objeto de estudio, las empresas que tienen algún tipo de relacionamiento con una IES en particular: la Fundación universitaria Los Libertadores. Esta muestra no estadística está compuesta por 216 empresas con participación económica en alrededor de 16 sectores económicos y que se encuentran ubicadas en cercanías a Los Libertadores, sobre todo, en la localidad de Chapinero.

Los sectores de las empresas encontradas y analizadas en la localidad de Chapinero que hacen parte del estudio, son: Silvicultura, Administración, Agricultura, Almacenamiento y transporte, Comercio, Construcción y minería, Muebles, Textil, Inmobiliario, Maquinaria y equipo, Metal, Plástico y caucho, Prendas, Publicidad, Químico y Salud.

### **2.1.3. Instrumentos de recogida de información**

Luego de indagar el número de empresas que comprendía el total de convenios que se tenían con información completa desde la Fundación Universitaria Los Libertadores, se determinó buscar la información bajo factores financieros, de estas empresas, en bases de datos especializadas (Cámara de Comercio de Bogotá y Superintendencia de Industria y Comercio), logrando un total de 216 empresas con información relevante para el estudio.

Se tomó información financiera de estas de 216 empresas, siendo la primera observación lograda, encontrar que estaban desarrollando actividades económicas en 16 sectores ubicados en la localidad de Chapinero los cuales están clasificados en: Silvicultura, Administración, Agricultura y ganadería, Almacenamiento y transporte, Comercio, Construcción y minería, Fabricación de muebles, Fabricación textil, Inmobiliario, Maquinaria y equipo, Metales, Papel y cartón, Plástico y caucho, Prendas de vestir, Publicidad, Químico, Salud.

Seguidamente, el análisis documental se fue elaborando a partir de una metodología que agrupa datos en clases empleando clustering particionado. El propósito de utilizar este tipo de técnicas de acuerdo con la literatura existente consiste en conseguir alcanzar una partición de los objetos en grupos de tal manera, que todos los objetos correspondan a alguno de los  $k$  clústeres posibles y que además los clústeres no tengan ningún elemento en común (Larranaga et al., 2012).

Mediante estas técnicas se pueden desarrollar estudios en diversos campos. Uno de ellos es el empresarial en el cual, a través de análisis de sectores económicos o en mercadotecnia para encontrar grupos objetivo, es posible con el clustering ayudar a agrupar (personas o entes) para segmentar grupos etarios de acuerdo con diferentes criterios en términos de características geográficas, sociodemográficas, conductuales, psicográficas, etc., y en función de sus cualidades homogéneas y similares de cada grupo, encontrar distintas formas de relacionarlas con los criterios de los productos, los servicios o los sectores mismos (Huang et al., 2012).

La selección de esta técnica obedeció a que el clustering ofrece un conjunto de técnicas multivariadas que son funcionales en la clasificación de grupos de objetos o individuos en grupos que son homogéneos. En la actualidad, el problema de la agrupación de particiones es uno de los problemas más difíciles en la investigación (Xiao & Xiao, 2010).

Para este estudio se analizó el sector de tejido productivo a partir de técnicas de clustering particionado, en la que se combina el Análisis de Componentes Principales (PCA) y el método de K-medias, con ayuda de la técnica de análisis multivariante para datos con múltiples variables continuas, con el propósito de dimensionar el comportamiento de la dinámica empresarial generada por este tipo de sector y que tiene relación con la institución Universitaria seleccionada para este estudio.

Con el ánimo de evaluar si el análisis multivariado es significativo se aplicó la prueba KMO (Kaiser, Meyer y Olkin), que relaciona los valores de correlación simple observados entre las variables contra indicadores de correlación

parcial. Se considera que los valores de KMO deben estar entre  $0.5 \leq KMO \leq 1$ , siendo aceptable un KMO de 0.5 y siendo apropiado un KMO de 1. Cuando el valor de KMO se encuentran en ese rango indica que es apropiado aplicar el análisis multivariado a los datos del estudio.

### **2.1.4. Procedimiento**

La agrupación en clústeres se conceptualiza en la literatura como un aprendizaje no supervisado, a través del cual, los objetos se agrupan sobre la base de la similitud inherente entre ellos. La teoría de la agrupación de clústeres o clustering como se le conoce normalmente, describe diferentes métodos para agrupar los objetos, como son: jerárquicos, partición, cuadrícula, basada en densidad y basada en el modelo. Las aplicaciones de la agrupación en clústeres se resaltan en diversos campos y la minería de datos (Saxena et al., 2017).

El análisis se desarrolló a través el software R Studio, con el paquete Factminer, utilizando métodos del análisis multivariado-denominados: Análisis de componentes principales (PCA) y análisis de clúster, mediante clustering por particionamiento. Estas técnicas específicamente hacen relación a métodos de clasificación automática que tiene como finalidad la agrupación de elementos de una muestra de acuerdo con la similitud existente entre ellos (Kassambara, 2017). En este caso se aplicó el algoritmo K-means o también llamado K-means clustering, el cual agrupa los elementos de una muestra en K clusters distintos, con la característica de que el valor de k se determina con antelación, en el caso de estudio k=4.

Para la selección del valor k se aplicó método del codo o elbow method para encontrar el número de clústeres con que se debía realizar el estudio. El método K-means aplica un algoritmo que calcula la media (centroide) de cada una de las variables que se estudian en los elementos de la muestra y luego asigna los elementos al clúster cuyo centroide está más cercano, asegurando que se reduzca la varianza interna del clúster (Amat, 2017).

Se aplicó la prueba KMO para el supuesto de homogeneidad de varianzas, a fin de corroborar que los datos fueran adecuados para realizar el análisis multivariado por análisis de componentes principales (PCA), obteniendo un valor KMO de 0.5 que se considera aceptable.

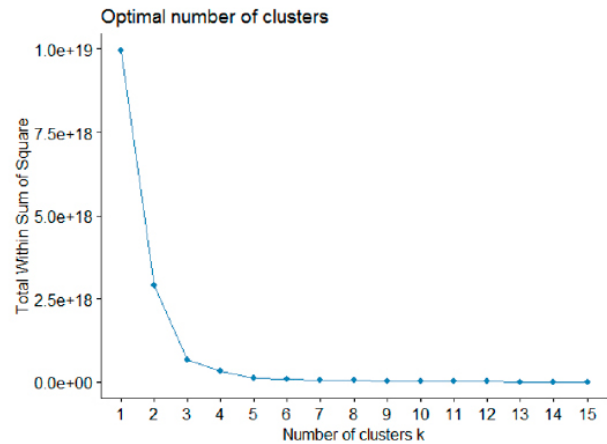
Luego, se procedió a establecer la existencia de relaciones entre los sectores de acuerdo con cada variable de estudio definidas así: Utilidad neta, activos, pasivos y patrimonio obtenidos por cada sector en el periodo 2013-2017. Igualmente se realizó un análisis Clúster CA, el cual buscó identificar el grado de cercanía o similitud entre sectores, teniendo en cuenta distancias obtenidas por la técnica k-means por tener variables continuas.

La variable de Utilidad neta hace referencia en este estudio a la rentabilidad de una empresa después de contabilizar todos los costos e impuestos. Los activos indican el valor de las propiedades que pertenecen a empresas del sector, por su parte los pasivos se refieren al total de obligaciones y deudas con las que las empresas del sector financian su actividad, y por último el patrimonio es el conjunto de bienes y derechos que posee una empresa en un periodo determinado, para el caso del estudio la periodicidad es anual.

### 3. RESULTADOS

Para el análisis de componentes principales (PCA-Biplot) y el análisis de clúster (Clúster Plot) se utilizó el software R Studio, aplicando la función K-means. Para ello, previamente se aplicó método del codo o elbow method que indicó el número de clusters con que se debía realizar el estudio, se encontró que con 4 clusters se logra que el within (suma de cuadrados interna de cada clúster) se estabilice, de acuerdo a la Fig. 1.

Fig. 1 Número óptimo de Clúster.



Fuente: Completamente generada por los autores, 2021

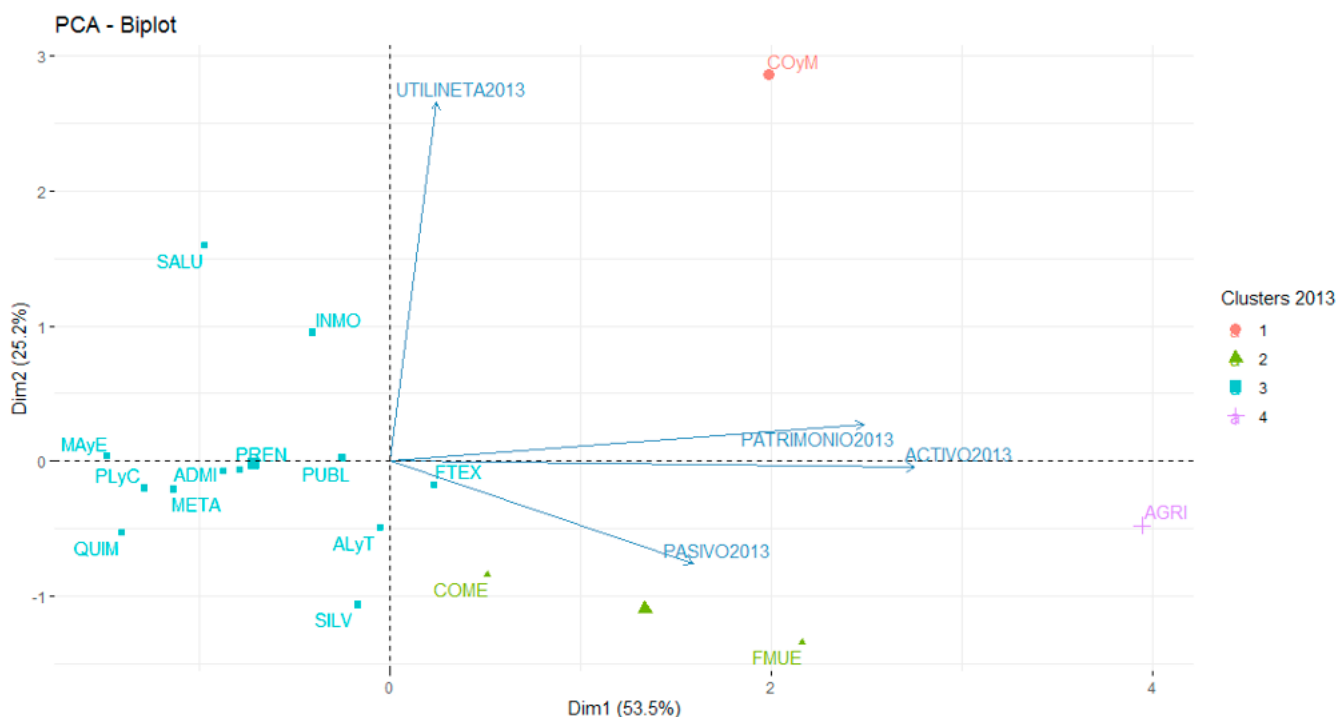
Como resultado de cada análisis, se contrastaron los componentes principales (PCA- Biplot) que contienen el mayor valor de varianza de los datos; un biplot se puede interpretar de la siguiente manera: un individuo (sector) que está en el mismo lado de una variable representada por medio de un vector (Utilidad, activo, pasivo o patrimonio) dada tiene un valor alto para esta variable; un individuo que está en el lado opuesto de una variable dada tiene un valor bajo para esta variable. Por otra parte, el tamaño del ángulo formado entre el conjunto de vectores, indica si existe alta correlación entre las variables de estudio, que para este caso son: activos, pasivos, patrimonio y utilidad neta.

Dinámica sectorial con agrupación jerárquica en componentes principales para el periodo 2013 a 2017

Se observa en la Fig. 2 PCA- Biplot del año 2013, las dos primeras componentes principales obtenidas de PCA que explican el 78.7% de la variabilidad de los datos. Como se puede observar en la figura, se presentan dos componentes y cuatro variables.

La componente Dim1 con el 53,5% explica de manera significativa las variables patrimonio, activo y pasivo, representados en vectores que crecen hacia el lado positivo de la Dim1 (eje X), evidenciando correlaciones positivas con las variables menos con la utilidad. Valores altos de esta componente implicaría que ese año el sector destaca en crecimiento del activo, pasivo y patrimonio. En ese mismo sentido, la componente Dim2 (eje Y) con el 25,2% estaría asociada de manera directa (correlación positiva) con la utilidad de las empresas que componen cada sector y en consecuencia valores altos de esta componente indican incremento de la utilidad de cada sector.

**Fig. 2 Análisis de componentes principales y Clúster K-Means en localidad de Chapinero 2013**



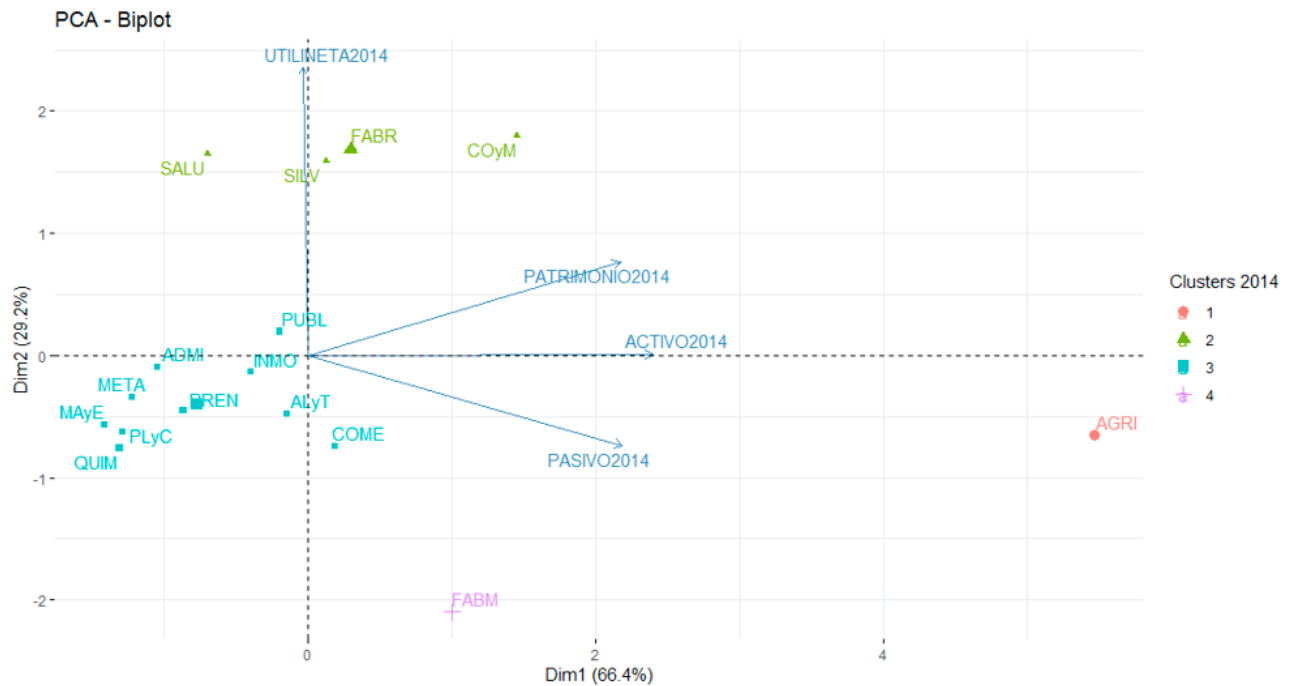
Fuente: Completamente generada por los autores, 2021

La figura 2 presenta el comportamiento de los sectores para el año 2013, en este caso se encuentra que el clúster 1 en color rojo está compuesto únicamente por el sector de construcción y minería que evidencia un muy alto valor de las cuentas de utilidad neta, y altos valores en las cuentas de patrimonio, activo y pasivo, lo que indica el buen desempeño financiero de este sector para el año 2013. El clúster 2 en color verde representa 2 sectores, comercio y fabricación de muebles, los cuales tienen un valor bajo de utilidad neta, sin embargo, presentan valores mayores a la media general en cuentas del pasivo, activo y patrimonio, lo que puede indicar una estrategia de gestión basada en el crecimiento del patrimonio y de los activos por medio de pasivos y de no repartición de utilidades a los socios.

El clúster 3 en color turquesa presenta dos tipos de comportamiento de los sectores que tienen valores por debajo de la media general en términos de los activos, patrimonio y pasivo, sin embargo es evidente que los sectores inmobiliario, maquinaria y equipo, publicidad presentaron utilidades positivas aunque bajas o cercanas a cero, mientras que los sectores plástico y caucho, administración, prendas, fabricación de textiles, metal, químico, almacenamiento y transporte, y silvicultura, presentaron muy bajo nivel de utilidad neta para el año 2013. El clúster 4 en color morado representa al sector agricultura que presentó como estrategia altos niveles de activos, patrimonio y pasivos esto evidencia que este sector aumento su nivel de endeudamiento a costa de las utilidades netas.

Para el año 2014, la Fig. 3 presenta un PCA- Biplot con dos componentes principales que contienen el 95.6% (Dim1 66.49% y Dim2 29.2%) de variabilidad de la información que permite analizar el comportamiento de los 16 sectores de estudio agrupados en 4 clústeres. El clúster 1 (rojo) muestra que al sector de agricultura que presenta bajo nivel de utilidad, y alto nivel de pasivos, activos y del patrimonio, situación que pudo suceder para el año 2014 debido a que el sector trabajó con la estrategia de endeudamiento y de no repartición de la utilidad con los socios a fin de incrementar sus activos y su patrimonio.

**Fig. 3 Análisis de componentes principales y Clúster K-Means en la localidad de Chapinero 2014**

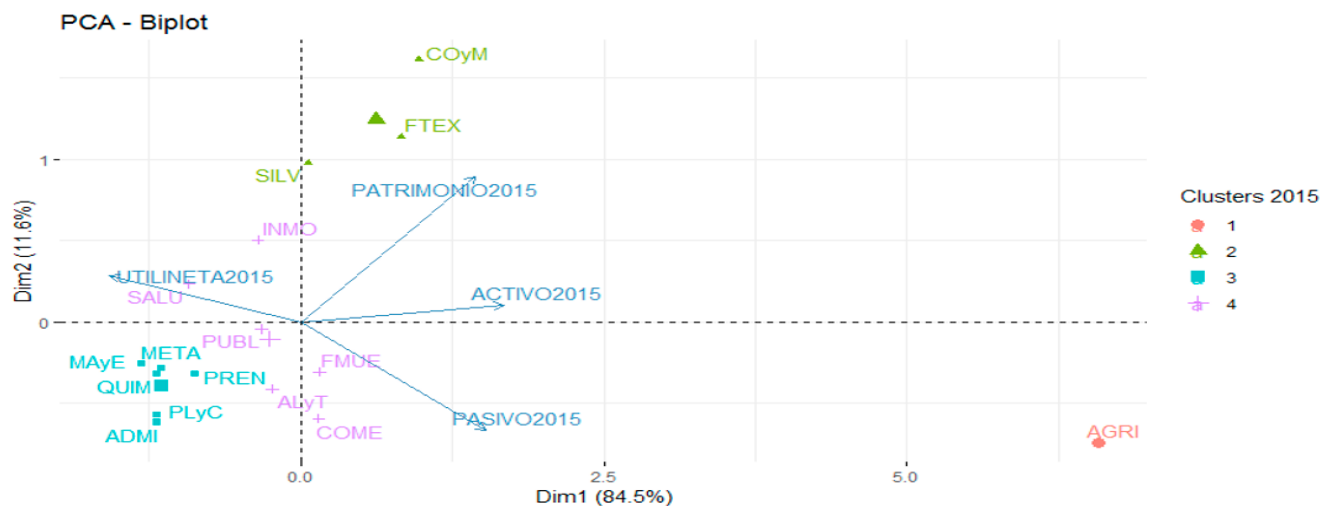


Fuente: Completamente generada por los autores, 2021

El clúster 2 en esta figura (verde, representa 4 sectores que son: silvicultura, fabricación textil, construcción y minería y, salud, encontrando que sectores presentan mejores resultados de utilidad neta, en comparación con el año 2013.

En el clúster 3 (turquesa) de esta figura, se encuentran 10 sectores de la localidad de Chapinero, mostrando 2 casos de acuerdo con la utilidad neta resultante, con una característica común que es bajo valor en las cuentas de pasivo, activo y patrimonio para el 2014. Los sectores del primer caso con utilidad neta positiva, aunque baja está el sector de publicidad, en el segundo caso con utilidad con muy bajo valor y en algunos casos con pérdidas están: Administración, Almacenamiento y transporte, Comercio, Inmobiliario, Maquinaria y equipo, Metal, Plástico y caucho, prendas, químico. El sector de fabricación de muebles para el año 2014 se encuentra clasificado en el clúster 4 (morado), debido a que presenta muy bajo valor de utilidad, al estar en una ubicación inversa al crecimiento de la utilidad neta, presenta valores mayores a la media en los pasivos activos y patrimonio, siendo este un comportamiento similar al presentado en el 2013, sin embargo debido a que no presentó utilidades nuevamente, se puede considerar que es un sector que está sobreviviendo a través del endeudamiento.

**Fig. 4 Análisis de componentes principales y Clúster K-Means en localidad de Chapinero 2015**



Fuente: Completamente generada por los autores, 2021

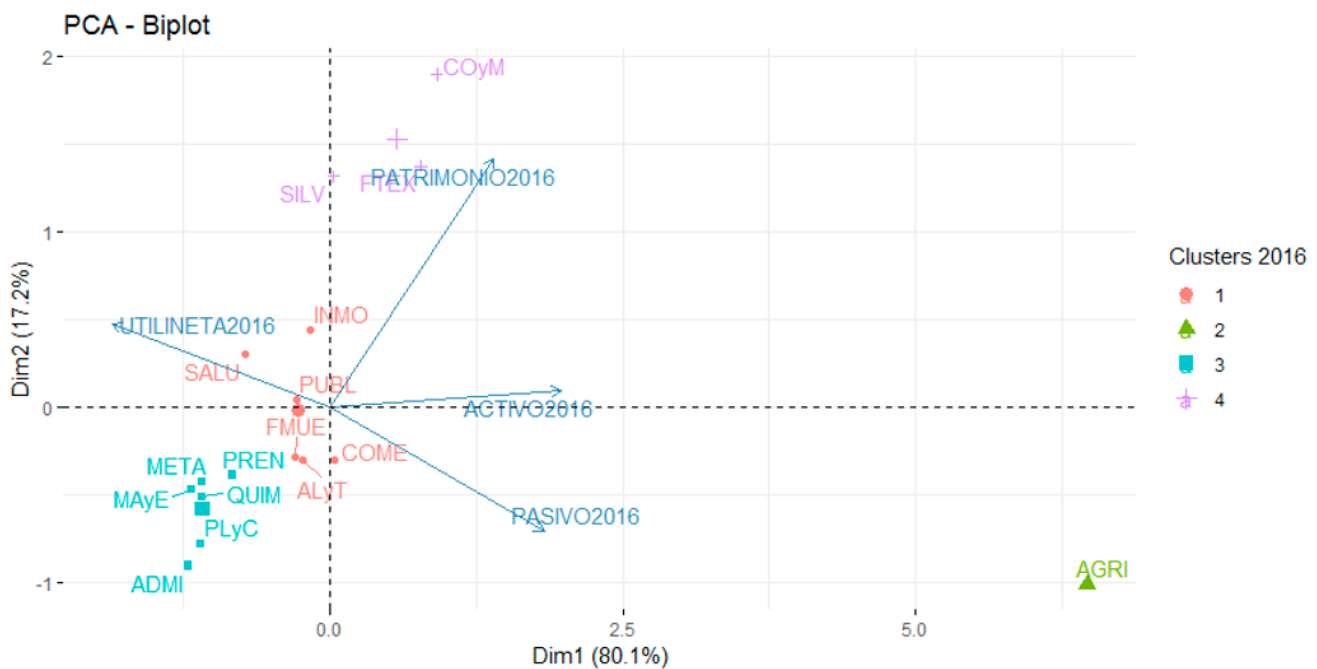
La Fig. 4 presenta un PCA- Biplot con dos componentes principales que contienen el 96.1% (Dim1 84.5% y Dim2 11.6%) de variabilidad de los 16 sectores de estudio para el año 2015 agrupados en 4 clústeres. Para el año 2015 se observa un cambio en la distribución y orientación de los vectores de cada variable en estudio. El vector de utilidad neta presenta una ubicación en el lado negativo de la componente principal Dim1, lo que pone de manifiesto que una parte importante de los sectores tuvo un desempeño bajo de la utilidad neta para ese año.

El clúster 1 (rojo) corresponde al sector de agricultura, el cual para el año 2015, presenta el mayor nivel de pérdidas en comparación con los demás sectores, con un alto nivel de pasivos, activos y patrimonio.

Para el año 2015 el clúster 2 (verde) está conformado en este caso por los sectores (silvicultura, construcción y minería, y Fabricación textil) los cuales presentan un buen desempeño de las utilidades netas y del patrimonio, mientras que el vector de pasivos disminuye para dichos sectores.

El clúster 3 (turquesa) en este caso está compuesto por 6 sectores Maquinaria y equipo, metal, plástico y caucho, prendas, administración y, químico, los cuales presentan bajo nivel en la cuenta de patrimonio y de los activos, mientras mantiene crecimiento en la utilidad neta y el pasivo. Por último, el clúster 4 (morado) está compuesto los sectores almacenamiento y transporte, comercio, y textil, publicidad y salud, los cuales presentan en común la reducción de los vectores de valores muy cercanos a la media general para todas las cuentas de utilidad neta, patrimonio, activo y pasivo. Un aspecto de importancia en el año 2015 es la mejora en las utilidades de los sectores que venían apareciendo rezagados en los dos años anteriores, sin embargo, al ser este un análisis exploratorio, no permite identificar las situaciones que permitieron ese cambio en el comportamiento de la variable utilidad.

**Fig. 5 Análisis de componentes principales y Clúster K-Means en localidad de Chapinero 2016**



*Fuente: Completamente generada por los autores, 2021*

Para el año 2016, la Fig. 5 presenta un PCA- Biplot con dos componentes principales que contienen el 97.3% (Dim1 80.1% y Dim2 17.2%) de variabilidad de los 16 sectores de estudio para el año 2016 agrupados en 4 clústeres. El clúster 1 (rojo) representa a 6 sectores entre los que están almacenamiento y transporte, comercio, fabricación de muebles, inmobiliario, publicidad y salud, estos sectores presentan valores muy cercanos a la media general para cada una de las cuentas.

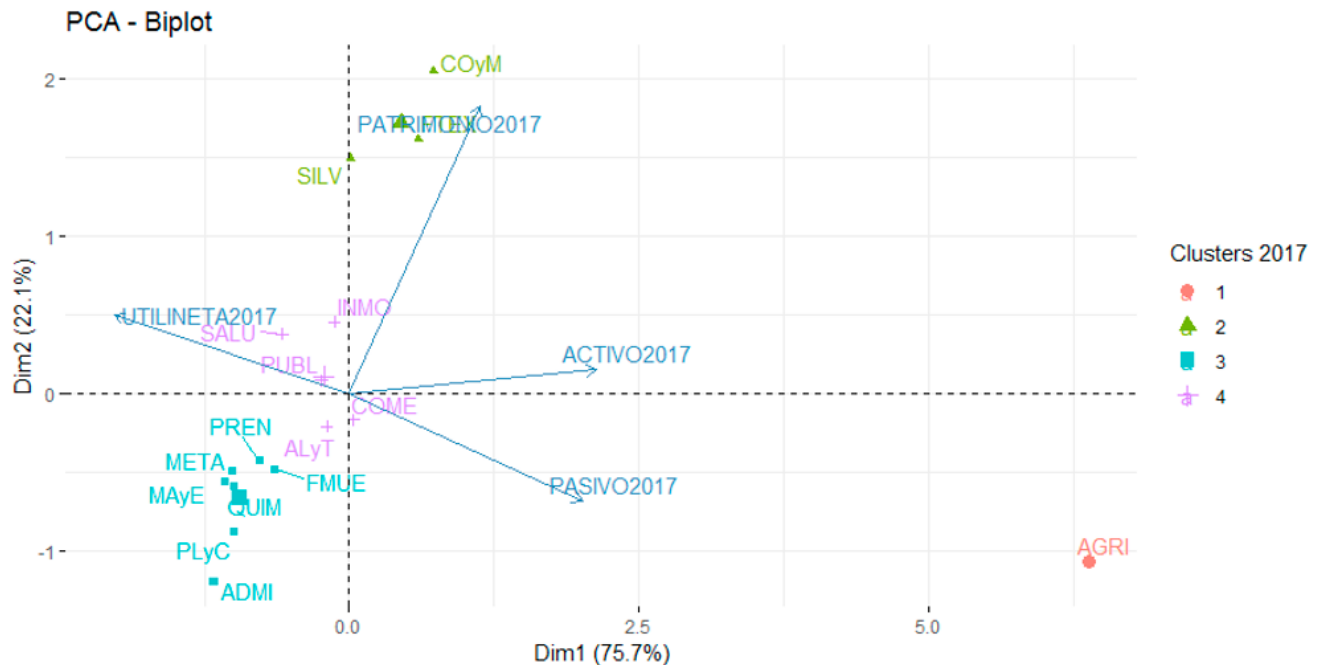
El clúster 2 (verde) corresponde al sector de agricultura con el más alto nivel de pasivos, con el mayor nivel de pérdidas, así como alto nivel de activos, pero con una disminución en el nivel de patrimonio, situación que muestra que el sector presenta serias dificultades que además de llevarlo a no generar utilidad, le está representado para el 2016 alto endeudamiento.

El clúster 3 (turquesa) representa 6 sectores que son: administración, maquinaria y equipo, metal, plástico y caucho, prendas y químico, donde es evidente que los sectores presentan un muy bajo nivel en las cuentas de patrimonio y activos, mientras mantiene utilidades netas positivas y moderadas y bajos niveles de pasivo. Al parecer la estrategia de estos sectores está mantener el nivel de endeudamiento bajo a costa de los activos y del patrimonio de cada sector.

En cuanto al análisis del clúster 4 (morado), este está conformado por 3 sectores que son silvicultura, construcción y minería y fabricación textil, los cuales presentan valores elevados del patrimonio, con utilidad neta positiva y reducción del pasivo. La estrategia de este sector está enmarcada en aumentar el patrimonio reduciendo a la vez el nivel de endeudamiento.

La Fig. 6 presenta un PCA- Biplot con dos componentes principales que contienen el 97.8% (Dim1 75.7% y Dim2 22.1%) de variabilidad de los 16 sectores de estudio para el año 2017 agrupados en 4 clústeres. En el clúster 1 (rojo) se tiene el sector agricultura con un nivel de pérdida notorio en comparación con los demás sectores y un alto nivel de pasivos, situación que se ha mantenido constante desde el año 2015.

**Fig. 6. Análisis de componentes principales y Clúster K-Means en localidad de Chapinero 2017**



Fuente: Completamente generada por los autores, 2021

En el clúster 2 (verde) representa a 3 sectores que presentan alto nivel de patrimonio, así como utilidades netas bajas pero positivas. Los sectores son silvicultura, construcción y minería, y fabricación textil. En el clúster 3 (turquesa) se encuentran ubicados los sectores administración, maquinaria y equipo, metal, plástico y caucho, químico, metal, y prendas, los cuales han mantenido su nivel alto de patrimonio para el año 2017, manteniendo utilidad neta cercana al promedio de utilidad de todos los sectores.

Para finalizar en el clúster 4 (morado) se encuentran ubicados los sectores almacenamiento y transporte, comercio, publicidad, inmobiliario y salud, el comportamiento de los sectores mencionados está dado por tener valores cercanos a la media en todas las cuentas de patrimonio y utilidad neta baja pero positiva y reducciones en el pasivo, por lo cual se puede mencionar que su nivel de crecimiento, como el de endeudamiento es moderado.

Al revisar la dinámica sectorial en términos de la utilidad neta generada por cada sector de la localidad de Chapinero, se encuentra que los sectores que mantuvieron niveles positivos altos de utilidad neta en los 5 años estudiados fueron: construcción y minería, publicidad y sector salud. Los sectores que mantuvieron niveles altos de utilidad neta por 4 años fueron: silvicultura, fabricación textil, e inmobiliario.

Los sectores que mantuvieron niveles altos de utilidad durante tres años continuos desde 2015 a 2017 fueron: administración, maquinaria y equipo, metal, plástico y caucho, prendas y químico, evidenciando una dinámica creciente y beneficiosa en las empresas que desarrollan actividades económicas semejantes. Por el contrario, los siguientes sectores presentaron utilidad neta con valores bajos tan solo 1 año de los 5 estudiados: almacenamiento y transporte, comercio y fabricación de muebles. Sin embargo, el sector que en los años estudiados presentó mayores y constantes pérdidas fue el sector de agricultura, lo que indica que este sector no está generando utilidades.

A su vez, la dinámica de los sectores en relación con el pasivo muestra que en el periodo de estudio los sectores de agricultura, almacenamiento y transporte, y comercio, presentaron los pasivos elevados de manera constante, poniendo en evidencia que son los sectores de la localidad de Chapinero con el mayor nivel de endeudamiento; en tanto que, el resto de los sectores lograron mantener bajos niveles en sus cuentas de pasivos especialmente en los últimos 3 años.

Los sectores de agricultura, construcción y minería, fabricación textil, y silvicultura presentaron un nivel de patrimonio, creciente a lo largo del periodo de estudio, en comparación con los demás sectores que mantuvieron bajo o constante el patrimonio. Este mismo comportamiento se dio en relación con el activo para los mismos sectores, incluyendo al sector comercio.

#### 4. DISCUSIÓN Y CONCLUSIONES

Lograr conocer las tipologías de las aglomeraciones presentes entre las empresas estudiadas, tiene implicaciones fundamentales en términos de estrategias sectoriales. Esto es aún más importante si se considera el contexto creciente de integración regional y mundial al que se está abocado como país emergente y sus resultados permitirían contribuir al diseño de estrategias orientadas a proveer condiciones locales necesarias para impulsar el territorio y sus orientaciones económicas de forma ventajosa y que mejor que la Fundación Universitaria Los Libertadores pueda ser participe del estudio de dinámicas territoriales, el perfeccionamiento de métodos y técnicas de reconocimiento territorial y el análisis de su comportamiento, en especial para la generación de escenarios propositivos para la ciudad.

El análisis multivariado al ser una herramienta adecuada para el análisis de las dinámicas empresariales y sectoriales, favoreció la identificación de patrones de comportamiento sectorial, los cuales, al ser realizados en un periodo de cinco años dan una noción de la tendencia de cada sector a los cambios. Al identificar el comportamiento y las relaciones subyacentes, se logró determinar las estrategias de gestión financiera compartidas por cada sector, sin embargo, el resultado obtenido en cada análisis, evidencia desigualdad entre sectores económicos en términos de las variables de estudio, manteniendo lo identificado por la mayoría de los autores referenciados.

El PCA al realizarse por el algoritmo de k-means, buscó y logró agrupar los datos en torno a un número preestablecido de Clústeres (Dominguez, Herrera, & Mendoza, 2018), según el nivel de proximidad definido por distancias Euclídeas (Fernández, 2019) y al reducir la dimensionalidad de los datos, se logró evidenciar interrelaciones existentes entre las observaciones, así como comportamientos típicos que develan patrones de similitud y que derivan la identificación de agrupamientos y relaciones más fácil que lo obtenido con el análisis univariado (Camarero, Cerbán, Turias, González, & Camarero, 2016).

Ahora bien, se debe tener en cuenta que en la presente investigación se omitieron aspectos tributarios e incentivos fiscales que pudieran existir por falta de información al respecto. Estudios previos han identificado la dificultad para incluir variables de este tipo, pues gran parte de los acuerdos fiscales y tributarios se logran a través de acuerdos privados entre gobierno y empresa.

Para finalizar, este capítulo es derivado del proyecto de investigación “Modelo de capacidades de innovación de un ecosistema universitario bajo dinámicas de stakeholders basado en simulación de agentes”, financiado por “Fundación Universitaria los Libertadores”. Año de inicio: 2021, año de finalización: 2021



## REFERENCIAS

- **Almansa, M. B. (2010).** *Inteligencia territorial para una redefinición eficiente de las políticas públicas.* *Trab. Rev. andaluza Relac. laborales* (23), 75–94
- **Amajid, G., Souaf, M., & Elwazani, Y. (2016).** *Territorial marketing and its effects on Development, approach from the literature.* *Revue marocaine de Recherche en Management et marketing*, 13, 111–129.
- **Amat, R. J. (2017).** *Clustering y heatmaps: aprendizaje no supervisado.*
- **Avendaño Prieto, B., Avendaño Prieto, G., William, C., & Cárdenas-Avendaño, A. (2014).** *Guía de referencia para investigadores no expertos en el uso de estadística multivariada\**. *Diversitas*.
- **Barrionuevo, C., & Brichetti, P. (2012).** *Aportes conceptuales para pensar los Observatorios de Inteligencia y Desarrollo Territorial. XI INTI International Conference (págs. 1–11).* *La Plata (Argentina): Universidad Nacional de La Plata.*
- **Bedoya, E., Vargas, L. E., Gonzalez, H., & Ceveriche, C. (2017).** *Análisis de componentes principales para caracterización estructural de pymes manufactureras de Cartagena de Indias, Colombia.* *Lámpsakos*(17), 52-59.
- **Benites, J. (2009).** *Innovación y competitividad: convirtiendo las ideas en valor.* *San Salvador: FAO*
- **Bernal, C. S. (2013).** *Atributos territoriales y la localización de las aglomeraciones industriales en Bogotá.*
- **Bogotá D.C.: Universidad de Los Andes.**
- **Bonabeau, E. (2009).** *Decisions 2.0 : The Power of Collective Intelligence.* *MIT Sloan Manag. Rev.*, 50(2), 45–52.
- **Bucheli, Herbert & Thompson Wayne (2009).** *Statistics and Machine Learning at Scale New Technologies Apply Machine Learning to Big Data.* *SAS*
- **Camarero, O. A., Cerbán, J. M., Turías, I., González, N., & Camarero, A. (2016).** *Clasificación de los puertos españoles mediante indicadores de explotación utilizando análisis de conglomerados.* *INGECUC*, 41-49.
- **Carrasco Oberto, Gonzalo. (2020).** *Clúster No jerárquico versus CART y BIPLLOT.* *Universidad de Salamanca. Departamento de Estadística.*
- **Chávez-Chong, C.O., Sánchez-García, J.E. y DelaCerde-Gastélum, J. (2015).** *Análisis de componentes principales funcionales en series de tiempo económicas.* *Revista Internacional de Gestión del Conocimiento y la Tecnología. ISSN 2255-5684 Vol.3 (2).* 1-13.
- **De la Fuente Fernández, S. (2011).** *Análisis de conglomerados. Obtenido de Facultad de ciencias económicas y empresariales UAM: [http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOM\\_ERADOS/conglomerados.pdf](http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOM_ERADOS/conglomerados.pdf)*
- **De La Hoz, E., & López, P. L. (2017).** *Aplicación de Técnicas de Análisis de Conglomerados y Redes Neuronales Artificiales en la Evaluación del Potencial Exportador de una Empresa.* *Información tecnológica*, 28(4), 67-74.
- **Domínguez, E. D., Herrera, T. F., & Mendoza, A. (2018).** *Métodos multivariantes aplicados a la efectividad de los disolventes en la industria farmacéutica.* *NGECUC*, 14(1), 133-140.



## REFERENCIAS

- **Fernández, J. M. (2019).** *Manual abreviado de estadística multivariante.* Badajoz, España: Universidad de Extremadura, Departamento de matemáticas.
- **Ferrán Aranz, M. (2011).** *Una metodología de minería de datos para la agrupación de series temporales: aplicación al sector de la construcción residencial.*
- **Fujita, M., & Thisse, J. (1996).** *Economics of Agglomeration.* *Journal of the Japanese and International Economies*, 10, 339-378.
- **Gallardo, M. (2013).** *Ampliación de Análisis de datos multivariantes.* Obtenido de Universidad de Granada: <http://www.ugr.es/~gallardo/>
- **Gil Martínez, C. (Junio de 2018).** *Métodos de clustering.* Obtenido de R Pubs: [https://github.com/CristinaGil/Estadística\\_machine\\_learning\\_R](https://github.com/CristinaGil/Estadística_machine_learning_R)
- **Girardot, J.-J. (2010).** *Inteligencia Territorial y Transición Socio-Ecológica.* *Trabajo*, 23(23), 15-39.
- **González-cabo, V., Murgueitio, M., & Cruz, C. L. (2014).** *Una mirada a lo regional ¿Competitividad versus Marketing Territorial ?* *Int. Rev. Bussiness Res. Pap.*, 5(1837-5685), 1-26.
- **González Tagle, M. (Agosto de 2016).** *Introducción a R y RStudio.* . Obtenido de [https://rstudio-pubs-static.s3.amazonaws.com/195980\\_3f4cd84bc3ca434daeec55c6c211d13e.html](https://rstudio-pubs-static.s3.amazonaws.com/195980_3f4cd84bc3ca434daeec55c6c211d13e.html)
- **Gutiérrez, T. A. (2011).** *Estudios del territorio: potencialidad socio-espacial para los procesos de desarrollo.* *Anekumene*(2), 119-135.
- **Hernández, S. R. (2014).** *Metodología de la investigación.* México D.F.: Mc Graw Hill.
- **Herreros Orcajo, H. (2020).** *Sistema de recomendación de League of Legends mediante agrupación y redes de neuronas (Bachelor's thesis).*
- **Huang, JJ, Tzeng, GH y Ong, CS (2007).** *Segmentación de marketing mediante agrupación de vectores de soporte. Sistemas expertos con aplicaciones*, 32 (2), 313-317.
- **Hurtado, d. B. (2000).** *Metodología de la investigación holística.* Caracas: Sypal
- **Husson F, Josse J, Lê S, Mazet J. (2008).** *FactoMineR: An R Package for Multivariate Analysis.* Obtenido de <http://www.jstatsoft.org/>
- **Kagan, Y. Y., & Jackson, D. D. (1991).** *Long-term earthquake clustering.* *Geophysical Journal International*, 104(1), 117-133.
- **Kassambara, A. (2017).** *Practical guide to principal component methods in R*
- **Krugman, P., & Venables, A. J. (1996).** *Globalization and the inequality of nations.* *The Quarterly Journal of Economics*, CX.5(5), 857-880.



## REFERENCIAS

- Larranaga, P., Inza, I., & Moujahid, A. (2012). Tema 14. Clustering.
- MacCallum, D., Moulart, F., Hillier, J., & Vicari, H. S. (2012). *Social Innovation and Territorial Development*. Ashgate.
- Manrique, O. (2006). Fuentes de las economías de aglomeración: Una revisión bibliográfica. *Cuadernos de economía*, 53-73.
- Gallardo, M. (2013). Ampliación de Análisis de datos multivariantes. Obtenido de Universidad de Granada: <http://www.ugr.es/~gallardo/>
- Marín, J. M. (2014). Análisis de conglomerados (i) El procedimiento conglomerados de K medias. Obtenido de Universidad Carlos III de Madrid: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/21conglk.pdf>
- Patibandla, R. S. M., & Veeranjanyulu, N. (2020). Asimrank based ensemble method for resolving challenges of partition clustering methods.
- Muñuzuri Sanz, J., & Lara Albín, J. d. (Septiembre de 2014). Obtenido de <http://bibing.us.es/proyectos/abreproy/5453/fichero/PFC+tecnicas+clusterizacion.pdf>
- Rousseaux, F., Saurel, P., & Petit, J. (2014). Knowledge Engineering or Digital Humanities? Territorial Intelligence, a Case in Point (Vol. 514). *Stud. Comput. Intell.*,.
- Roux, M. (2018). A comparative study of divisive and agglomerative hierarchical clustering algorithms. *Journal of Classification*, 35(2), 345-366.
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664-681.
- Schneider, S., & Peyré, T. (2006). *Territorio y enfoque territorial: de las referencias cognitivas a los aportes aplicados al análisis de los procesos*. Buenos Aires: Ciccus.
- Spielman, S. E. (2014). Spatial Collective Intelligence? credibility, accuracy, and Volunteered Geographic Information. *Cartogr. Geogr. Inf. Sci.*, 41(2), 115-124.
- Solano, E., Alandete, N. Vilorio, A. (2019). Análisis de componentes principales en la competitividad en Colombia. *Revista ibérica de sistemas y tecnologías de la información RISTI*. 24, 260-271.
- Tomadoni, C. (2007). A propósito de las nociones de espacio y territorio. *Gestión y Ambiente*, 10(4), 53-65.
- Tritiño, V. M. (1998). Patrimonio arquitectónico, cultura y territorio. *Dialnet*, 95-104.
- Xiao, J. y L. Xiao, L. (2010) "A Research of the Partition Clustering Algorithm", *Seminario internacional de 2010 sobre procesamiento de información de inteligencia y computación confiable*, págs. 551-553, doi: 10.1109 / IPTC.2010.148.