



FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

**MODELO DE ANÁLISIS CLÚSTER APLICADO A INSTITUCIONES BANCARIAS
DEL SISTEMA FINANCIERO PERUANO: UNA HERRAMIENTA PARA
IDENTIFICAR PERFILES DE RIESGO EN EL BANCO GNB**

**Línea de investigación:
Estadística y bioestadística**

Modalidad de Suficiencia Profesional para optar por el Título Profesional
de Licenciada en Estadística

Autora

Gomez Ccapa, Enma Rayza

Asesor

Aniceto Capristán, Anne Elizabeth

ORCID: 0000-0001-7683-056X

Jurado

Estrada Cantero, Jeanette Nazaria

Rivas Arguelles, José Walter

Ruiz Arias, Raúl Alberto

Lima - Perú

2025

MODELO DE ANÁLISIS CLÚSTER APLICADO A INSTITUCIONES BANCARIAS DEL SISTEMA FINANCIERO PERUANO: UNA HERRAMIENTA PARA IDENTIFICAR PERFILES DE RIESGO EN EL BANCO GNB

INFORME DE ORIGINALIDAD

28%

INDICE DE SIMILITUD

28%

FUENTES DE INTERNET

2%

PUBLICACIONES

16%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1

dspace.ucuenca.edu.ec

Fuente de Internet

2%

2

docplayer.es

Fuente de Internet

2%

3

datatab.es

Fuente de Internet

2%

4

www.bancognb.com.pe

Fuente de Internet

2%

5

www.coursehero.com

Fuente de Internet

2%

6

hdl.handle.net

Fuente de Internet

1%

7

prezi.com

Fuente de Internet

1%

8

revistes.ub.edu

Fuente de Internet

1%



Universidad Nacional
Federico Villarreal

VRIN | VICERRECTORADO
DE INVESTIGACIÓN

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

**MODELO DE ANÁLISIS CLÚSTER APLICADO A INSTITUCIONES
BANCARIAS DEL SISTEMA FINANCIERO PERUANO: UNA
HERRAMIENTA PARA IDENTIFICAR PERFILES DE RIESGO EN EL
BANCO GNB**

Línea de Investigación:

Estadística y bioestadística

Modalidad de Suficiencia Profesional para optar por el Título

Profesional de Licenciada en Estadística

Autor:

Gomez Ccapa, Enma Rayza

Asesor:

Aniceto Capristán, Anne Elizabeth

ORCID: 0000-0001-7683-056X

Jurado:

Estrada Cantero, Jeanette Nazaria

Rivas Arguelles, José Walter

Ruiz Arias, Raúl Alberto

Lima - Perú

2025

Dedicatoria

Este trabajo está dedicado a mi familia conformada por mis padres, Luis y Mercedes quienes han estado siempre al pie del cañón apoyándome en todo sentido; mi hermana Lorena, compañera incondicional en buenos y sobre todo malos momentos en mi desarrollo personal y profesional. Sin el apoyo de estas personas, llegar a este punto de mi carrera no hubiera sido posible.

INDICE

I.	Introducción	1
1.1.	Trayectoria del autor	2
1.2.	Descripción de la empresa	3
1.2.1.	Visión	4
1.2.2.	Misión	4
1.3.	Organigrama de la empresa.....	4
1.4.	Áreas y funciones desempeñadas	6
II.	Descripción de una actividad específica	9
2.1.	Aplicación y dimensiones analizadas.....	9
2.1.1.	Objetivo general	9
2.1.2.	Objetivos específicos.....	10
2.2.	Marco metodológico/teórico.....	10
2.2.1.	Métodos de análisis multivariado	10
2.2.1.1.	Técnicas de análisis de dependencia	13
2.2.1.2.	Técnicas de análisis de interdependencia	15
2.2.2.	Análisis de conglomerados o clúster	18
2.2.2.1.	Pasos para el análisis de conglomerados	19
2.2.2.2.	Medidas de Similaridad.....	20
2.2.2.3.	Estandarización de los datos.....	21
2.2.2.4.	Formación de grupos	22
2.3.	Definición de términos básicos	24
2.4.	Variables y dimensiones analizadas	25
2.5.	Resultados encontrados	27

2.5.1. Análisis descriptivo	27
2.5.2. Desarrollo de Análisis por Conglomerados o <i>Clúster</i>	31
2.5.2.1. Método del centroide	31
2.5.2.2. Método del vecino más cercano	33
2.5.2.3. Método del vecino más lejano	34
2.5.2.4. Método de vinculación promedio	35
2.5.2.5. Método de Ward	37
2.5.3. Elección del número de grupos	39
III. Aportes más destacables a la Empresa/Institución	41
IV. Conclusiones	46
V. Recomendaciones	48
VI. Referencias bibliográficas	49
VII. Anexos	51

ÍNDICE DE FIGURAS

Figura 1 Organigrama de Banco GNB Perú	5
Figura 2 Proceso de realización de un análisis de conglomerados	20
Figura 3 Ejemplo de dendograma para un análisis cluster de 7 observaciones	23
Figura 4 Historial de conglomeración método del centroide	24
Figura 5 Factores y variables para el análisis	27
Figura 6 Boxplot distribuciones de variables	29
Figura 7 Matriz de correlación	30
Figura 8 Dendograma método del centroide	32
Figura 9 Dendograma método del vecino más cercano	34
Figura 10 Dendograma método del vecino más lejano	35
Figura 11 Dendograma método vinculación promedio	37
Figura 12 Dendograma método de Ward	38
Figura 13 Frecuencias para la elección del método de cluster	40
Figura 14 Participación de créditos sobre activos-Grupo 6	42
Figura 15 Participación de pasivos sobre activos-Grupo 6	42
Figura 16 Ratio de costos-Grupo 6	43
Figura 17 Ratio de Margen-Grupo 6	44
Figura 18 Ratio de Rentabilidad-Grupo 6	44
Figura 19 Ratio de Solvencia-Grupo 6	45

ÍNDICE DE TABLAS

Tabla 1 Clasificación de técnicas multivariadas	13
Tabla 2 Estadísticos descriptivos de las variables	28
Tabla 3 Resultados Test de Bartlett	30
Tabla 4 Historial de conglomeración método del centroide	32
Tabla 5 Historial de conglomeración método del vecino más cercano.....	33
Tabla 6 Historial de conglomeración método del vecino más lejano	34
Tabla 7 Historial de conglomeración método de vinculación promedio	36
Tabla 8 Historial de conglomeración método de ward	38
Tabla 9 Formación de grupos	41

Resumen

El presente trabajo que lleva por nombre: “Modelo de análisis clúster aplicado a instituciones bancarias del Sistema Financiero peruano: una herramienta para identificar perfiles de riesgo en el Banco GNB”, surgió por la necesidad de formar grupos de entidades del sistema Financiero Peruano como herramienta para identificar y analizar perfiles de riesgos en el área de Gestión de Riesgo de crédito del Banco GNB. La metodología utilizada es el análisis de conglomerados o *cluster*, técnica estadística que forma parte de los métodos multivariados. El análisis de *cluster* en la industria bancaria se enmarca en la vasta literatura empírica que aplica distintas técnicas estadísticas para analizar el comportamiento estratégico de los bancos (Koller, 2001). Sin embargo, estudios recientes, han aplicado el análisis de conglomerados para identificar grupos de bancos en economías emergentes (Terrones y Vargas, 2013), en este sentido, este trabajo, tuvo como objetivo principal, la formación de grupos de entidades en base a variables de equilibrio financiero, generación de margen y costos, rentabilidad y solvencia. El análisis se realizó sobre una base conformada por las dieciséis entidades que conforman el sistema financiero peruano y las variables ya mencionadas anteriormente. Los resultados fueron la formación de seis grupos de entidades: el grupo de los bancos comerciales compuesto por Citibank y ICBC, el grupo de los bancos Retail y mype con Mibanco, Ripley y Falabella, los de mayor participación en el mercado como BCP, BBVA, Scotiabank e Interbank, los bancos especializados como GNB, Pichincha y Santander y de forma aislada quedaron como grupo unitario Bank of China y Alfin Banco por ser entidades que recién han iniciado operaciones en la banca peruana.

Palabras clave: sistema financiero, análisis por conglomerados, perfil de riesgo.

Abstract

The present work, which is called: “Cluster analysis model applied to banking institutions of the Peruvian Financial System: a tool to identify risk profiles in Banco GNB”, arose from the need to form groups of entities of the Peruvian Financial System as a tool. to identify and analyze risk profiles in the Credit Risk Management area of Banco GNB. The methodology used is cluster analysis, a statistical technique that is part of multivariate methods. Cluster analysis in the banking industry is part of the vast empirical literature that applies different statistical techniques to analyze the strategic behavior of banks (Koller, 2001). However, recent studies have applied cluster analysis to identify groups of banks in emerging economies (Terrones and Vargas, 2013). In this sense, this work had as its main objective the formation of groups of entities based on variables. financial balance, margin generation and costs, profitability and solvency. The analysis was carried out on a basis made up of the sixteen entities that make up the Peruvian financial system and the variables already mentioned above. The results were the formation of six groups of entities: the group of commercial banks composed of Citibank and ICBC, the group of Retail and mype banks with Mibanco, Ripley and Falabella, those with the greatest market participation such as BCP, BBVA, Scotiabank and Interbank, specialized banks such as GNB, Pichincha and Santander and, in isolation, Bank of China and Alfin Banco remained as a unitary group as they are entities that have recently begun operations in Peruvian banking. Key words: Analysis cluster, Risk Management, determining factors of risk profile

Keywords: Financial System, analysis cluster, risk profile.

I. INTRODUCCIÓN

El sistema financiero comprende todas las instituciones del sistema bancario que operan en un país. En el Perú, está integrado por el Banco Central de Reserva del Perú, el Banco de la Nación, las empresas bancarias y la banca de fomento en liquidación (BCRP, 2023). En específico, las empresas bancarias que se dedican a la intermediación financiera persiguen diferentes objetivos, entre los que se encuentran el análisis permanente del posicionamiento que tiene cada entidad dentro del marco bancario.

En la práctica, dicho análisis es realizado por las áreas de Gestión de Riesgo. En su Reglamento de Gestión de Riesgo de Crédito 3780-2011, la Superintendencia de Banca y Seguros (SBS), insta a las entidades financieras a establecer su mercado objetivo y analizar el posicionamiento en el mercado (SBS, 2011). A partir de este reglamento es que surge el análisis de perfiles de riesgo. Dicho análisis es realizado por cada entidad para saber cómo está en relación con las demás empresas financieras. Esto es importante ya que, a través de factores como el equilibrio financiero, generación de margen, rentabilidad y solvencia, la empresa será capaz de saber cuán sólida es para el logro de sus objetivos.

El presente informe se desarrolla dentro del área de Seguimiento de Portafolio de la entidad financiera Banco GNB Perú. Tiene como objetivo, desarrollar un modelo basado en el análisis de *cluster* o conglomerados para formar grupos de entidades financieras que permita clasificar a Banco GNB junto con otras de perfil de riesgo parecido.

El sistema financiero peruano consta de 16 entidades financieras que tienen características que las hacen diferentes entre sí, como la participación de créditos y depósitos, indicadores financieros como la rentabilidad, costos y solvencia, por lo cual realizar un análisis comparado de GNB frente a cada una de las 15 entidades restantes resulta poco práctico.

Por este motivo, surge la necesidad de implementar una metodología que permita agrupar a las entidades financieras y situar a GNB junto con otras que reúnan características similares.

Con los resultados obtenidos de este modelo, se destaca que Banco GNB comparte grupo con Banco Pichincha, Banbif y Santander. Este grupo se destaca por tener un equilibrio de participación entre activos y depósitos, niveles de rentabilidad holgados en comparación a la competencia y solidez financiera a pesar de no tener una alta cuota de participación sobre el sistema financiero.

1.1. Trayectoria del autor

El autor es bachiller en la carrera profesional de Estadística de la Universidad Nacional Federico Villareal, magister en Métodos Cuantitativos para la gestión de datos en Organizaciones de la Universidad de Buenos Aires.

Comenzó su carrera como practicante de Seguimiento de Riesgos en Banbif (2012), apoyando en actividades de ordenamiento de datos, elaboración de cuadros estadísticos con nivel descriptivo y análisis de indicadores de riesgo a fin de dar soporte a los analistas. Seguido de esto, en esta misma entidad, ascendió al puesto de asistente de seguimiento de riesgos (2014), enfocándose en el seguimiento del portafolio de banca personal. En este contexto, el profesional tuvo la oportunidad de aplicar sus conocimientos como estadística a través del manejo de modelos de score para el seguimiento de tarjetas de crédito, elaborados bajo la metodología logit. Al mismo tiempo, parte de las actividades de este puesto consistía en elaboración de análisis descriptivo de los productos que forman parte la cartera del banco.

En el 2015, se dio el cambio a Banco GNB, entidad donde ha estado laborando estos últimos 8 años. En esta entidad su línea de carrera empezó como analista de Seguimiento de Portafolio de Créditos reportando a la gerencia de Gestión de Riesgos.

Sus principales funciones como analista consistieron en elaborar informes acerca de clientes con alta probabilidad de incumplimiento, elaboración de tablero de indicadores de riesgos y darle seguimiento a los mismos, así como la implementación de metodologías para gestionar el riesgo de crédito. A partir del ', se dio su ascenso al puesto de Especialista de Seguimiento de Portafolio. En este cargo, sus responsabilidades estuvieron enfocadas en el trabajo de los datos con enfoque de Big Data, dado el crecimiento de la cartera. Así mismo estuvo encargada de la elaboración de modelos de segmentación de clientes, estimación de parámetros de riesgos (Probabilidad de *default*), modelos econométricos para el cálculo de la pérdida esperada en base a variables macroeconómicas y perfiles de riesgo de entidades financieras que permitan hacer el seguimiento de la entidad dentro del sistema financiero.

1.2. Descripción de la empresa

Banco GNB Sudameris cuenta con más de 100 años de historia en el sistema financiero de Colombia y está posicionado como una de las mejores empresas del mercado financiero en dicho país. La presencia de la compañía en Perú y América Latina muestra su interés en explorar nuevos mercados y expandir su presencia en toda la región.

Es así como Banco GNB Perú es una unidad financiera del grupo GNB Sudameris, que inició operaciones en nuestro país en octubre de 2013. El banco cuenta con 12 agencias en todo el país (8 en Lima y 4 en provincias) y más de 500 empleados. Ofrece al público una amplia gama de productos y servicios dirigidos a personas naturales y jurídicas.

1.2.1. Visión

Proporcionar a los clientes soluciones financieras eficaces a través de productos y servicios con una atención personalizada que logre satisfacer sus necesidades.

1.2.1. Misión

Ser reconocidos como un banco comprometido, enfocado en la excelencia y dedicados a ofrecer la mejor calidad de servicio al cliente con un personal calificado.

1.3. Organigrama de la empresa

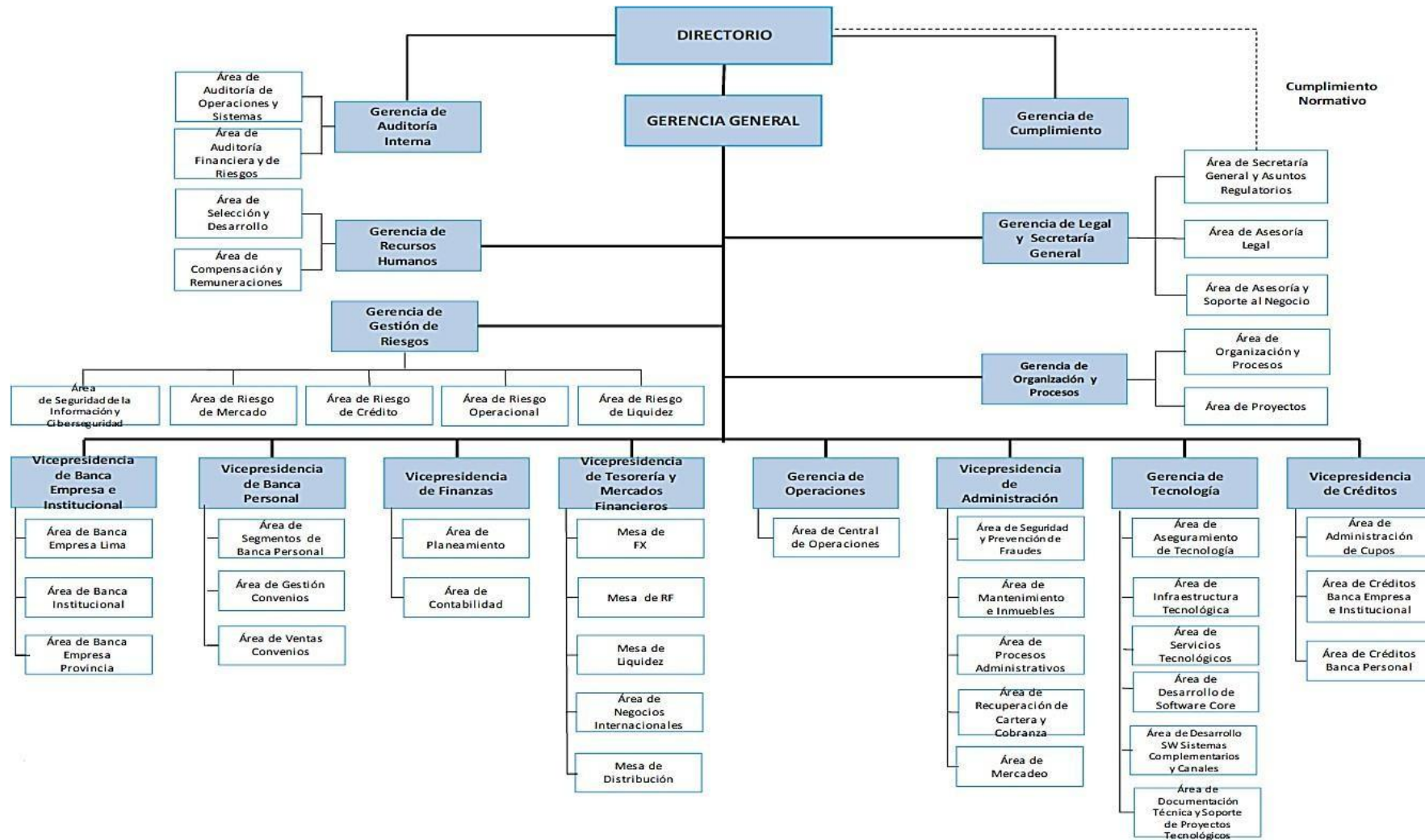
El Banco GNB, consta de 6 gerencias y 8 vicepresidencias, las cuales reportan a la gerencia general. Cada una de estas unidades desempeñan roles estratégicos, de soporte y de contacto directo con el cliente.

El autor de este trabajo viene trabajando hace más de ocho años para la Gerencia de Gestión de Riesgos, en el área de Riesgo de Crédito, cuyo rol es brindar soporte analítico al banco y maneja subdivisiones como:

- Área de seguridad de la información
- Área de riesgo de mercado
- Área de riesgo de crédito
- Área de riesgo operacional
- Área de riesgo de liquidez

Figura 1

Organigrama de Banco GNB Perú



1.4. Áreas y funciones desempeñadas

Banbif: Practicante de Seguimiento de Riesgos (Abr-12 / Dic-13)

- Apoyo en actividades relacionadas al seguimiento de indicadores de riesgo de crédito.
- Seguimiento mensual de concentración de cartera por actividad económica.
- Apoyo en el cálculo del indicador de sobre endeudamiento.
- Apoyo en el análisis de cosechas y cálculo del Indicador de Riesgo como métrica de la evolución de la cartera a nivel de producto.

Banbif: Asistente de Seguimiento de Riesgos (Ene-14 / Ene-15)

- Administración del sistema de señales de alerta para Tarjetas de crédito y revisión individual de clientes con alta probabilidad de incumplimiento.
- Apoyo en las principales actividades relacionadas al seguimiento de la calidad de cartera de créditos del banco.
- Elaboración de reportes para el seguimiento de la cartera de créditos de los clientes de la banca comercial (leasing, vencidos, sistema de vigilancia, señales de alerta entre otros).
- Análisis de cosechas y cálculo del Indicador de Riesgo como métrica de la evolución de la cartera a nivel de producto.
- Construcción del tablero de mora por oficinas de la banca comercial con la finalidad de ver el comportamiento de la cartera atrasada a nivel de Banca, Zona, Oficina y Ejecutivo de Negocios.

Banco GNB: Analista de Seguimiento de Portafolio (Feb-15 / Ene-19)

- Identificación de clientes con alta probabilidad de pérdida a través de la medición de variables de riesgo.
- Seguimiento a los indicadores y límites de riesgo crediticio del portafolio de Banca Comercial y Empresas.
- Informe analítico del gasto de provisiones (tendencias y resultados) para las diferentes líneas de negocio incluyendo perspectivas por sector, productos y equipos.
- Seguimiento de los principales indicadores de riesgo a fin de que estos se encuentren dentro de los límites establecidos.
- Responsable de la elaboración y envío de reportes regulatorios a la SBS, así como el envío de estos a otras entidades como ASBANC y COFIDE.

Banco GNB: Especialista de Seguimiento de Portafolio (Ene-19)

- Análisis de indicadores de riesgo crediticio a nivel portafolio (Big data aplicada a la gestión de riesgos).
- Informes de requerimiento de capital por riesgo de crédito, tendencia y evolución por producto y línea de negocio.
- Elaboración de modelos para la segmentación de clientes, estimación de pérdida esperada y concentración crediticia.
- Informes de competencia en el sistema financiero y modelo de perfil de riesgo de crédito.
- Responsable del sistema de administración de riesgo país. Actualización y seguimiento de calificaciones e indicadores macroeconómicos por país.
- Elaboración de tableros semáforo de alertas crediticias para el portafolio.

- Elaboración de informes de análisis de cartera y clientes con mayor impacto de cara a directorio y comité ejecutivos.
- Responsable del sistema de administración de riesgo cambiario crediticio. Evolución de ratio de dolarización, por la línea de negocio, escenarios estrés de cartera con *shock* cambiario.

II. Descripción de una actividad específica

2.1. Aplicación y dimensiones analizadas

Con el fin de probar las competencias en el campo de la Estadística el profesional desarrollo el análisis de conglomerados para formar grupos de entidades del Sistema Financiero Peruano, en base a variables de perfil de riesgos.

Las variables de perfil de riesgo se usaron para este análisis se dividen en cuatro dimensiones: equilibrio financiero, generación de margen y costos, rentabilidad ajustada al riesgo y solvencia. Cada una de estas posee variables específicas. Por un lado, equilibrio financiero está compuesta por participación de créditos y depósitos que tiene cada banco como parte de sus activos. El factor generación de margen y costes comprende margen neto y costes estructurales, que son indicadores financieros que miden las ganancias obtenidas por la empresa, así como los costes que vienen a ser los gastos realizados para sostener el negocio. La rentabilidad ajustada al riesgo está representada por dos variables que son los Ratios de APR (Activos Ponderados por Riesgo) y la rentabilidad. Finalmente, la Solvencia se mide con la variable Ratio de Solvencia o apalancamiento, el cual es un indicador que mide la solidez de sus activos (cartera) que tienen las entidades financieras.

2.1.1. *Objetivo general*

- Formar grupos de Entidades Financieras en función a variables de equilibrio financiero, generación de margen, rentabilidad ajustada al riesgo y solvencia; que sean lo más parecidas entre sí y diferentes entre los mismos, para así saber que entidades tienen un perfil de riesgo parecido al Banco GNB.

2.1.2. *Objetivos específicos*

- Analizar con técnicas descriptivas exploratorias, las variables de equilibrio financiero, generación de margen, rentabilidad ajustada al riesgo y solvencia, las mismas aportarán a la formación de grupos de Entidades Financieras del Perú.
- Proponer un modelo de perfil de riesgos para Banco GNB, en base al análisis de conglomerados o clúster para las Entidades Financieras del Perú, que tenga como valor agregado, la formación de grupos teniendo en cuenta variables determinantes como equilibrio financiero, generación de margen, rentabilidad ajustada al riesgo y solvencia.

2.2. Marco metodológico/teórico

Este apartado tiene como finalidad desarrollar las metodologías que se utilizó en este trabajo. Partiendo de conceptos generales hasta la metodología usada para cumplir con los objetivos propuestos.

2.2.1. *Métodos de análisis multivariado*

El análisis multivariante tiene como objetivo el estudio estadístico de varias variables medidas en elementos de una sola población (Peña, 2002). Según las necesidades analíticas, esta técnica persigue los siguientes objetivos:

- Resumir el conjunto de variables en unas pocas nuevas variables, construidas como transformaciones de las originales, con la mínima pérdida de información.
- Encontrar grupos en los datos si existen.
- Clasificar nuevas observaciones en grupos definidos.
- Relacionar dos conjuntos de variables.

Es habitual en la práctica, el uso de análisis bivariado, pero cuando entra a tallar una tercera variable o más, el análisis adquiere mayor dificultad, haciendo los modelos más

complejos. Sin embargo, el desarrollo de la informática en los últimos años ha favorecido el uso de técnicas multivariadas con paquetes estadísticos para el análisis de datos (Lozares y López, 1991). Las ventajas que ofrece el análisis multivariado sobre el bivariado son puestas en manifiesto por Cattell:

- Ahorro en el almacenamiento de datos
- Las conclusiones estadísticas son más consistentes
- Desarrollar conceptos teóricos con mayor sustento
- Mejor percepción conceptual (Cattell,1966).

El análisis multivariante deriva del análisis bivariado y este a su vez del análisis univariante. En este sentido, al ser una extensión, el análisis multivariante es una técnica que propone analizar e interpretar las relaciones entre diferentes variables a través de la construcción de modelos complejos que expliquen relaciones de dependencia o interdependencia (Meneses, 2019).

Clasificación de técnicas multivariantes:

Es importante clasificar la técnica multivariada que va a utilizarse, para dar una solución adecuada al problema que se desea resolver.

Según Dillon y Goldstein (1984), antes de elegir entre una técnica u otra, es necesario responder las siguientes preguntas:

- ¿La investigación responde a un problema de dependencia entre variables o interdependencia de estas?
- ¿Cómo están medidas las variables implicadas, en escalas no métricas o métricas?
- Si se está ante un problema de dependencia, ¿Cuántas relaciones se plantean entre las variables dependientes e independientes? ¿Cuántas variables dependientes existen?

Existen diferentes criterios para clasificar las técnicas estadísticas multivariadas. Por un lado, se tienen aquellas con enfoque descriptivo que atienden a objetivos científicos que buscan extraer información y aquellas que generan conocimiento con enfoque inferencial. Ambos métodos pueden asociarse con objetivos específicos de reducción de información y extracción de indicadores, clasificación y construcción de grupos y predicción de relaciones entre variables (Peña, 2002).

Un segundo enfoque son las técnicas de análisis de dependencia e interdependencia. Estas consideran la naturaleza o tipo de variables involucradas (Quero e Inciarte, 2012). Lo descrito en base a los tipos de técnicas multivariadas se resume en el siguiente cuadro:

Tabla 1*Clasificación de técnicas multivariadas*

<i>Técnicas de Dependencia</i>	<i>Técnicas de Independencia</i>
Análisis de regresión lineal múltiple	Análisis de Componentes principales
Análisis Discriminante	Análisis factorial
Análisis de correlación canónica	Análisis cluster o conglomerados
	Análisis de correspondencias

Fuente: Elaboración propia

Para fines prácticos, en el siguiente apartado, se va a abordar de forma sintetizada cada método, dado que no está dentro de los objetivos de este trabajo profundizar en todas las técnicas de análisis multivariado.

2.2.1.1. Técnicas de análisis de dependencia. Las técnicas de análisis de dependencia buscan la existencia o ausencia de relaciones entre dos grupos de variables. Entre las más importantes se tienen:

Análisis de regresión lineal múltiple: busca analizar la relación existente entre una variable dependiente métrica y un conjunto de variables independientes que pueden ser métricas o no métricas. El modelo de regresión lineal debe cumplir con algunos supuestos:

- Supuesto de Linealidad: debe haber una relación del tipo lineal entre las variables dependientes e independientes.
- Supuesto de Independencia: los errores que arroja el modelo deben ser independientes.

- Supuesto de Homocedasticidad: los errores que arroja el modelo deben tener varianza constante.
- Supuesto de Normalidad: Cada variable del modelo debe seguir una distribución normal.
- Supuesto de No colinealidad: las variables explicativas no deben estar correlacionadas.

Los pasos para la aplicación de un modelo de regresión lineal según Vila, Torrado

y Reguant son:

- i. Selección de la variable dependiente que se va a predecir.
- ii. Selección de la o las variables explicativas o independientes.
- iii. Verificación de los supuestos del modelo.
- iv. Interpretación de resultados.
- v. Verificación de la Bondad de ajuste del modelo (Vila, Torrado y Reguant, 2019).

Análisis discriminante: el objetivo de análisis es el mismo que la técnica de regresión lineal con la diferencia que en este caso, la variable dependiente es no numérica. Esta técnica fue propuesta por R. Fisher para saber si existen diferencias significativas entre grupos de sujetos para un conjunto de variables medidas sobre ellos; si existen, explicar en qué dirección se producen y proporcionar procedimientos para clasificar sistemáticamente nuevas observaciones.

Los pasos por seguir para realizar un análisis discriminante (AD) según Ruiz comprenden:

- i. Planteamiento del problema.
- ii. Verificar la existencia de diferencias entre grupos.
- iii. Fijar el número y composición de las dimensiones de discriminación entre los grupos analizados.

- iv. Decidir cuáles son las variables clasificadoras que explican la mayor parte de las diferencias observadas.
- v. Desarrollar un proceso sistemático para clasificar objetos de origen desconocido en grupos de análisis.
- vi. Determinar la significancia estadística obtenida de los resultados en el proceso de clasificación (Ruiz, 2013).

Análisis de correlación canónica: el fin de esta técnica es determinar la existencia de asociación lineal entre un conjunto de variables independientes métricas y otro conjunto de variables dependientes que pueden ser métricas y no métricas. Con respecto a los objetivos de este método, Badii y Castillo, establecen:

- i. Determinar si dos conjuntos de variables son independientes entre sí o viceversa y el grado de relación que puede existir entre los dos conjuntos. Obtener un conjunto de ponderaciones para cada conjunto de variables criterio y variables predictoras, para que las combinaciones lineales de cada conjunto estén correlacionadas de forma máxima.
- ii. Interpretar la naturaleza de cualquier relación existente entre los conjuntos de variables dependientes y variables predictoras. (Badii y Castillo, 2007)

2.2.1.2. Técnicas de análisis de interdependencia

Análisis de componentes principales: Esta es una técnica de reducción de dimensionalidad cuyo objetivo básico es construir una combinación lineal de las variables originales que explique la mayor cantidad posible de información recopilada de las variables originales. El análisis de componentes principales comprende las siguientes etapas según Gurrea:

- i. Análisis de la matriz de correlaciones. Es necesario crear una matriz de correlación porque cuando las variables están altamente correlacionadas, existe información redundante y una pequeña cantidad de factores pueden explicar una gran parte de la variación total.
- ii. Selección de factores. Se selecciona el primer factor como el factor con mayor cantidad de variación, y así sucesivamente hasta seleccionar el porcentaje de variación que el investigador considera adecuado.
- iii. Análisis de la matriz factorial. Consiste en representar los componentes en forma de matrices. Dicha matriz tendrá tantas columnas como componentes principales y tantas filas como variables.
- iv. Interpretación de los factores. Los coeficientes de los factores deben estar cerca de 1, y las variables deben tener coeficientes altos en un solo factor y ningún factor debe tener coeficientes similares.
- v. Cálculo de las puntuaciones factoriales. Se trata de puntuaciones que incluyen los componentes principales de cada caso, permitiendo representarlos gráficamente (Gurrea, 2000).

Análisis factorial exploratorio (AFE): al igual que en el análisis de componentes principales, se busca reducir la dimensionalidad, pero el interés se centra en establecer que causas latentes (factores) pueden estar causando la correlación entre las variables manifiestas, es decir, entre las variables observadas. Según Khan (2006), el término AFE se refiere tanto a un conjunto de técnicas estadísticas como a una única técnica interdependiente utilizada para reducir una gran cantidad de indicadores operativos a una pequeña cantidad de variables conceptuales. (Blalock, 1966).

Se recomienda hacer un análisis exploratorio previo a fin de detectar valores atípicos o extremos. Los pasos por seguir para el análisis factorial son los siguientes:

- i. Cálculo de la matriz de correlaciones. Los métodos de cálculo más conocidos son el criterio de Kaiser y Scree
- ii. Cálculo de Comunalidades. Es la variabilidad explicada por los factores
- iii. Cálculo de Matriz de componentes. Da a conocer las cargas de los factores sobre las variables.
- iv. Cálculo Matriz de rotación. Se calcula para facilitar la interpretación de los factores.
- v. Rotación varimax. Se calcula con fines analíticos para asegurar que las cargas ya sean altas o bajas sean las correctas.

Análisis de correspondencias: permite visualizar gráficamente mediante tablas de contingencias relaciones entre dos variables no métricas. Esta técnica está diseñada para estudiar las relaciones entre cualquier número de modalidades de distintas variables categóricas (Parra, 1996). Para poder llevar a cabo el análisis de correspondencias se tienen los siguientes pasos:

- i. Elaborar una tabla de contingencias. Contiene las frecuencias absolutas simples que son los elementos de la intersección entre filas y columnas, frecuencias absolutas relativas para filas y columnas que determinan los perfiles, así como las frecuencias marginales relativas que determinan las masas.
- ii. Cálculos de distancias Chi cuadrado. Se toma de la tabla de contingencias calculando así distancias fila y distancias columna.
- iii. Cálculo de inercias e interpretación. El cálculo de inercias facilita la interpretación del análisis de correspondencia al ser una medida de dispersión. Está asociada a la masa y la distancia entre cada perfil y cada perfil promedio.

Es una medida de dispersión, es decir una inercia baja indica que las variables no estarían correlacionadas, así como una inercia alta determina alta correlación de estas.

Análisis *cluster* o conglomerados: El propósito de esta técnica no es agrupar variables, sino agrupar observaciones de manera que las observaciones en cada grupo sean similares con respecto a las variables utilizadas para agrupar, y las observaciones entre grupos sean lo más diferentes posible con respecto a esas variables. El análisis de conglomerados se presenta en dos tipologías: métodos de agrupamiento jerárquico y métodos de agrupamiento no jerárquico. La agrupación jerárquica incluye métodos jerárquicos aglomerados y métodos jerárquicos divididos. Por otro lado, los métodos no jerárquicos se utilizan para agrupar objetos por un número de grupo específico.

2.2.2. Análisis de conglomerados o *clúster*

Análisis de conglomerados o *cluster*, tiene como objetivo agrupar objetos o individuos en grupos homogéneos en base a similitudes entre ellos. Este método también forma parte de los métodos de clasificación automática no supervisada. Esta técnica se encarga de resolver tres tipos de problemas (Peña, 2002):

- Partición de datos:

En primer lugar, se tiene un conjunto de datos heterogéneos y se desea formar un número específico de grupos de modo que cada elemento pertenezca exactamente a un grupo y esté clasificado, y cada grupo sea internamente homogéneo.

- Construcción de jerarquías:

Dada su similitud, es deseable estructurar jerárquicamente los elementos de un conjunto de datos. La clasificación jerárquica significa que los datos se organizan en niveles de modo que los niveles más altos contengan a los niveles más bajos.

- Clasificación de variables

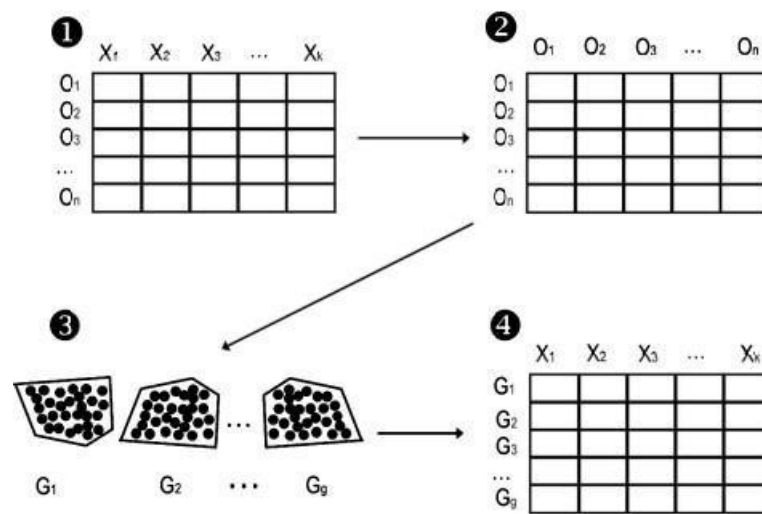
Si se dispone de un gran número de variables, sería interesante dividir las en grupos. Las variables pueden formarse en grupos o estructurarse jerárquicamente. Este trabajo no aborda este tipo de problemas ya que el objetivo es crear grupos de individuos.

2.2.2.1. Pasos para el análisis de conglomerados. El análisis de conglomerados se diferencia de otras técnicas multivariadas por no conocer a priori el número de grupos, esto se deriva de las observaciones. Aldas y Uriel (2017), establecen los siguientes pasos a seguir para el análisis cluster:

- El analista dispone de una cantidad de observaciones con una cantidad de variables.
- A continuación, se elige una medida que diga en qué grado cada par de observaciones se parece entre sí. A esta medida se le llama medida de similitud o similaridad.
- El siguiente paso es formar grupos que contengan observaciones que sean más similares entre sí según la medida de similitud calculada previamente. Esto requiere elegir entre dos tipos de análisis de conglomerados (jerárquico y no jerárquico) y seleccionar el método de agrupamiento apropiado para el tipo de análisis seleccionado.
- Finalmente, los grupos identificados deben ser descritos y comparados entre sí. Para ello basta con ver qué valor medio toman las variables utilizadas en el análisis en cada grupo resultante.

Figura 2

Proceso de realización de un análisis de conglomerados



Fuente: Aldas y Uriel (2017)

A continuación, se van a describir las decisiones que será necesario adoptar en cada uno de los pasos descritos. Cabe mencionar que solo se va a poner foco en los métodos utilizados para realizar el presente trabajo.

2.2.2.2. Medidas de Similitud. Si las variables utilizadas para caracterizar las observaciones son métricas o numéricas, se encuentran disponibles las siguientes medidas de similitud:

- **Distancia euclídea:** Si se consideran dos observaciones i y j de las n posibles y si llamamos x_{jp} el valor de la variable x_p de las k existentes en dichas observaciones, la distancia euclídea D_{ij} entre ambas se calcularía del siguiente modo:

$$D_{ij} = \sqrt{\sum_{p=1}^k (x_{ip} - x_{jp})^2}$$

- Distancia euclídea al cuadrado: elevar al cuadrado la distancia euclídea, hace más sencillo el cálculo de las medidas de similaridad, sobre todo cuando se trata de métodos de agrupación como Ward o del centroide.

$$D_{ij} = \sum_{p=1}^k (x_{ip} - x_{jp})^2$$

Otras medidas utilizadas para el análisis de conglomerados son la distancia de Minkowski, distancia de Manhattan y para datos binarios se tiene: índice de Jaccard, coeficiente de Sokal y Michener, coeficiente de Hamann entre otros.

2.2.2.3. Estandarización de los datos. Las medidas de similaridad son muy sensibles a las unidades de medida propias de cada variable. Es necesario corregir el efecto de los datos, para evitar influencias no deseables. Se tienen:

- Puntuación Z: Los datos se estandarizan restando de cada observación de una variable particular la media de esa variable para el grupo de observaciones, y dividiendo el resultado por la desviación estándar. Por tanto, una variable estandarizada tiene una media de 0 y una desviación estándar de 1.
- Rango 1: El valor de una variable determinada para cada observación se divide por el rango de esa variable para el grupo de observación. De esta forma, el rango de variación de una variable así estandarizada se reduce a un intervalo de valores de 1.
- Rango de 0 a 1: El valor de una variable particular para cada observación se estandariza restando el valor más bajo tomado por esta variable dentro del conjunto de observaciones y dividiéndolo por el rango. Por tanto, el valor mínimo de la variable es 0 y el valor máximo es 1.

2.2.2.4. **Formación de grupos.** Los algoritmos de agrupación

existentes responden a dos grandes enfoques:

Métodos no Jerárquicos

Inicialmente se establece a priori un número de grupos y los individuos se van clasificando en cada uno de esos ellos. La elección de estos grupos se hace con conocimiento del problema de investigación.

Métodos Jerárquicos

Se opera paso a paso para formar los grupos definitivos. Se manejan 2 enfoques:

- **Aglomerativo:** Se parte con grupos de menor tamaño y así se van formando los grupos. Inicialmente, cada individuo es un grupo.
- **Desagregativo:** Se parte de un grupo de mayor tamaño y se van desagregando elementos. Se empieza con todos los individuos en un solo grupo. Métodos de agrupamiento:
- **Método del centroide:** Enfoque Aglomerativo. Utiliza la distancia euclídea al cuadrado. Se van uniendo iterativamente las observaciones más cercanas, representadas por su respectivo promedio, este proceso arroja el historial de conglomeración y una representación gráfica llamada dendograma, donde el punto de corte que va a indicar el número de grupos es aquella parte del gráfico donde se observa un gran salto.
- **Método del vecino más cercano (Single Linkage):** Enfoque Aglomerativo. Utiliza la distancia euclídea al cuadrado. Para medir las distancias entre dos grupos, formando grupos a raíz de menor distancia, considerando la menor distancia entre los individuos de cada grupo.

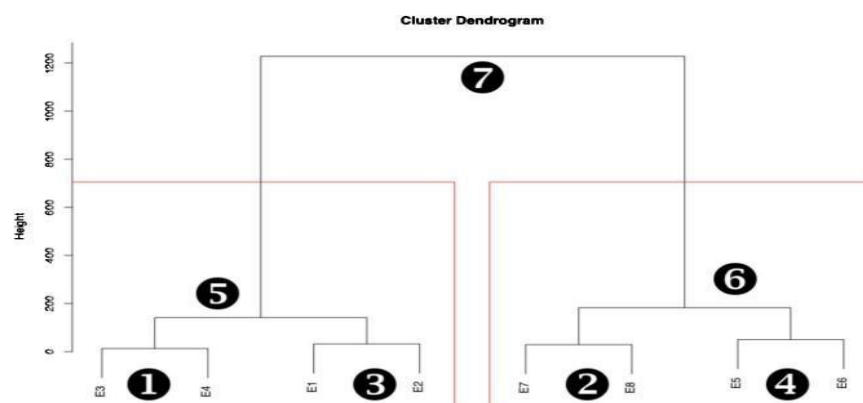
- Método del vecino más lejano (Complete Linkage): Enfoque Aglomerativo. Utiliza la distancia euclídea al cuadrado. Es similar al anterior, con la diferencia que la distancia entre dos grupos se mide por la distancia entre sus miembros más alejados.
- Método de la vinculación promedio (Average Linkage): Considera el promedio de las distancias entre los elementos de los grupos que se van formando.
- Método de Ward: Enfoque Desagregativo. Busca la máxima homogeneidad dentro de cada grupo. Evalúa todas las posibles combinaciones, para todas ellas se calcula una distancia asociada a la distancia euclídea al cuadrado. Se elige la solución con la menor SCT.

Historial de conglomeración y dendograma

El historial de conglomeración muestra las distancias sucesivas que se han calculado como parte del proceso de agrupamiento. El dendograma viene a ser una representación gráfica de este que va a permitir determinar de forma razonable el número de grupos o cluster a elegir.

Figura 3

Ejemplo de dendograma para un análisis cluster de 7 observaciones



Fuente: Aldas y Uriel (2017)

Índices para la elección del número de conglomerados o *cluster*

Existen diferentes índices que se han planteado en las últimas fechas. En específico Charrad et al. (2014) ha incorporado en su paquete de R un total de 30 índices. En este trabajo se limita a ampliar en cada uno de estos, sin embargo, en el apartado elección de número de grupos (2.6.3) se van a presentar tablas de frecuencias de los 30 índices por cada método clúster que se va a desarrollar. Esto va a facilitar la elección del número de grupos.

2.3. Definición de términos básicos

Este trabajo se desarrolla en el contexto del Sistema Financiero Peruano, en específico dentro del área de seguimiento de portafolio de Banco GNB. En ese sentido, es necesario definir las terminologías, de cara a entender los resultados.

- Sistema financiero peruano: está compuesto por el Banco Central, Banco de la Nación y empresas bancarias. El modelo *clúster* que se va a desarrollar se centra en las empresas bancarias del sistema financiero, actualmente son las siguientes:

Figura 4

Bancos del Sistema Financiero Peruano

ENTIDAD
BCP
INTERBANK
SCOTIABANK
BBVA
COMERCIO
CITIBANK
BANBIF
RIPLEY
PICHINCHA
MIBANCO
GNB
SANTANDER
ICBC
BANK OF CHINA
ALFIN
FALABELLA

Fuente: Superintendencia de Banca y Seguros (SBS)

- Perfil de Riesgo de la entidad financiera

Es la relación entre los beneficios que se pretende obtener y los riesgos que se está dispuesto a asumir. Por ejemplo, el Banco BBVA, tiene un modelo de perfil de riesgos de acuerdo con su modelo de negocio, teniendo en cuenta factores de mercado, geografía donde opera (BBVA,2015). Esto es particularmente ventajoso para el Banco GNB. Debido a que esta entidad cuenta con un modelo de perfil de riesgo que le permite desarrollar sus actividades en el marco de las estrategias y políticas de gestión y gestión de riesgos establecidas por las personas jurídicas del banco, adaptarse a los cambios de la economía y adaptarse a las regulaciones.

2.4. Variables y dimensiones analizadas

En este trabajo se van a considerar cuatro factores los cuales son:

- Equilibrio Financiero

La manera más práctica de identificar modelos de negocio entre una u otra entidad, es comparando su estructura de activos y pasivos, así los modelos que tienen mayor proporción de créditos estarán orientados a la intermediación financiera frente a entidades cuyos modelos estén orientados a la banca corporativa y de inversión.

Variables de Equilibrio Financiero

Participación de créditos: se mide sobre los activos totales de la entidad financiera. Mide que porcentaje de los activos de la empresa corresponde a los créditos.

Participación de depósitos: se mide sobre los activos totales de la entidad financiera. Mide que porcentaje de los activos de la empresa corresponde a los depósitos.

A este grupo de indicadores también se le conoce como indicadores de estructura financiera.

- Generación de margen y eficiencia en costes

Las entidades financieras generan ganancias medidas por los indicadores de márgenes ya sea bruto o neto. Esto sirve para comparar cuanto genera cada entidad en función a los segmentos orientados, y los costes incurridos para soportar el negocio.

Variables de generación de Margen y eficiencia de Costes

Margen neto: medido sobre los activos, indica la rentabilidad de los activos totales.

Costes estructurales: medido sobre los activos, es el gasto en que la empresa incurre para contratación de personal, servicios recibidos de terceros, impuestos y contribuciones.

- Rentabilidad ajustada al riesgo y perfil

La competencia entre entidades financieras también se mide por la diversidad de riesgo al que la entidad está dispuesta a incurrir. El indicador más representativo para medir esta diversidad de riesgo es el APR (Activos Ponderados por Riesgos) el cual se divide por el activo total de la empresa. Entonces a mayor valor de este indicador, mayor es el riesgo en el sentido que cuenta con mayores activos de ponderación elevada afectos de APR. Variables de Rentabilidad y Perfil

APR: información correspondiente a la suma de los activos de la empresa ponderados según riesgo en miles de soles por cada empresa del sistema financiero.

Rentabilidad: se mide dividiendo el margen bruto sobre el activo. Es el beneficio obtenido con los recursos propios de la entidad financiera.

- Solvencia y Apalancamiento

El indicador más importante de la fortaleza financiera de las entidades de crédito es una adecuada posición de solvencia.

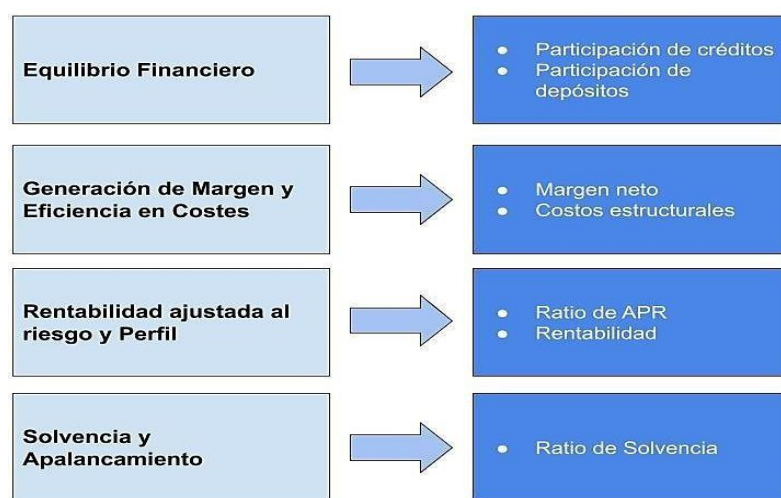
Variables de Solvencia y Apalancamiento

Ratio de Solvencia: nivel de apalancamiento de la empresa, en número de veces. Indica cuan sólida es la entidad, actualmente el regulador (SBS) exige a las entidades financieras tener como mínimo 8% de ratio de solvencia.

En la siguiente figura se describe que indicadores de cada factor se van a considerar en este trabajo:

Figura 5

Factores y variables para el análisis



Elaboración propia

2.5. Resultados encontrados

2.5.1. Análisis descriptivo

Para formar los grupos de entidades financieras, se dispone de una base de datos de la banca múltiple compuesta por 16 entidades (observaciones) y 8 variables (columnas) cuya información corresponde a diciembre del 2022. El análisis estadístico se va a realizar con la herramienta computacional R-Studio.

Tabla 2*Estadísticos descriptivos de las variables*

Distribución de las variables							
Media, desvío y principales cuantiles							
Variable	Promedio	Desvío estándar	Min	p25	Mediana	p75	Max
APR	27.48125	42.873834	0.3	3.200	7.10	26.100	152.4
PART_CREDITOS	58.33750	23.751544	3.8	56.150	66.40	73.300	85.0
PART_DEP	67.80625	10.709402	51.5	63.250	65.35	71.075	90.2
RENTABILIDAD	7.11875	4.556932	0.7	4.900	5.70	10.025	15.3
COSTOS_ESTRUCTURALES	1.67500	1.160172	0.0	0.775	1.55	2.425	4.3
MARGEN_NETO	3.65625	2.954312	-0.4	1.950	3.10	4.300	9.5
PERFIL	94.45000	40.665333	24.0	81.975	85.30	108.875	184.3
RATIO_CAPITAL	20.30625	20.124197	4.1	13.675	14.80	16.500	92.8

Elaboración propia

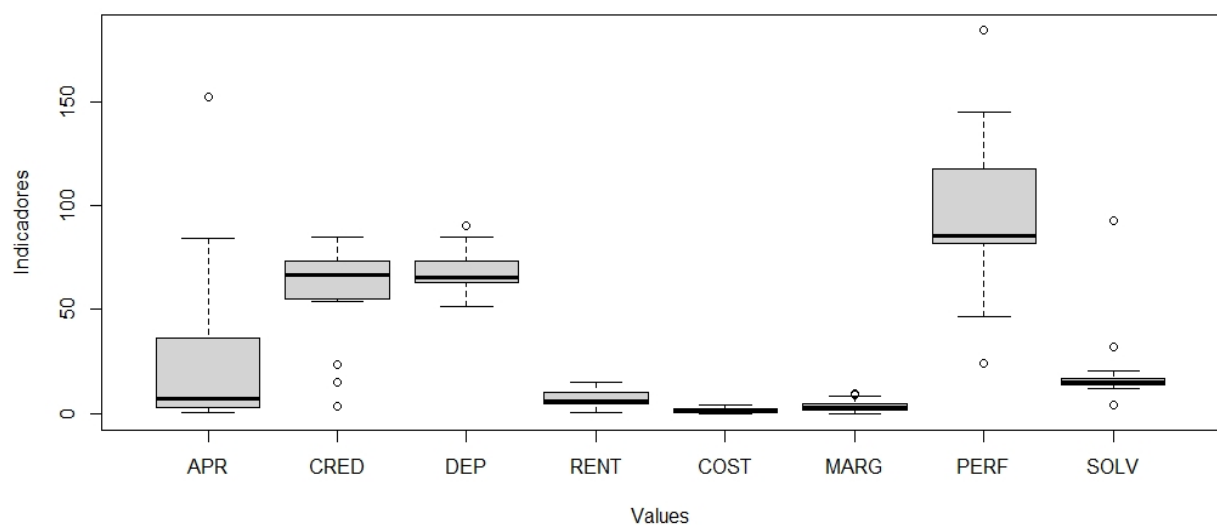
Puede verse que, en promedio, las entidades destinan el 58% de sus activos a la colocación de créditos y el 68% a la captación de depósitos. El recorrido de ambas variables indica que más del 50% de las observaciones (8 bancos) cuenta con una participación en créditos y depósitos superior al 65%, así mismo existen casos extremos, como ICBC (85%) y Mibanco (90%) las cuales están enfocadas en la colocación de depósitos y la intermediación Financiera respectivamente.

Con relación a los indicadores financieros, se observa que los costos estructurales y el margen neto promedian en un 1.7% y 3.7% respectivamente. Por un lado, el nivel de costos o gastos en que incurren las entidades para gestionar su negocio es bajo, lo cual es positivo; sin embargo, el margen promedio se ve afectado por algunas entidades que han registrado valores negativos, es decir no han generado utilidad al cierre del ejercicio 2022. Cabe mencionar que ambos indicadores presentan niveles de volatilidad bajos para las entidades analizadas, en comparación con el resto de los indicadores.

El siguiente gráfico muestra el recorrido de cada variable, donde se observa que variables como Perfil, APR, Créditos y Depósitos toman valores considerablemente superiores a los de variables como rentabilidad, costo, margen y solvencia. Esto evidencia la necesidad de estandarizar las variables, sumado a que su escala de medición es distinta.

Figura 6

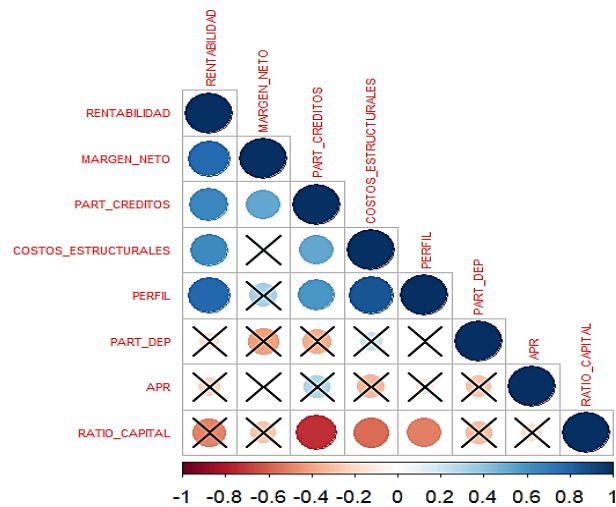
Boxplot distribuciones de variables



Elaboración propia

Análisis de correlaciones

Antes de realizar el análisis de conglomerados, se presenta un análisis de correlación. A pesar de que existen ciertas correlaciones, la mayoría de ellas no son significativas.

Figura 7*Matriz de correlación*

Elaboración propia

El Test de Bartlett verifica si la matriz de correlación definida anteriormente se diferencia significativamente de la matriz identidad y así probar la existencia de correlación entre variables. Se plantea la hipótesis nula y alternativa

h_0 : la matriz de correlacion es igual a la identidad, no existe

correlación h_1 : las matriz de correlacion es diferente a la identidad, existe correlación

Tabla 3*Resultados Test de Bartlett*

<i>Chi cuadrado</i>	<i>p-value</i>	<i>g.l.</i>
929.6428	1.06E-177	28

Elaboración propia

En este caso, se rechaza la hipótesis nula en favor de la alternativa de existencia de correlación entre variables.

2.5.2. *Desarrollo de Análisis por Conglomerados o Clúster*

Antes de iniciar un análisis clúster deberían tomarse tres decisiones:

1. Selección de las variables relevantes para identificar a los grupos.
2. Elección de la medida de proximidad entre los individuos (medida de distancia).
3. Seleccionar el criterio para agrupar individuos en conglomerados (método).

En primer lugar, se van a estandarizar las variables restando su media y dividiendo por desvío estándar. Al estar considerando sólo variables métricas se procede a calcular la matriz de distancias euclídea y euclídea al cuadrado. Se utilizará la distancia euclídea al cuadrado y métodos jerárquicos para determinar el número de *clúster* o conglomerados.

2.5.2.1. Método del centroide. En primer lugar, se analiza el método del centroide. Para ello, se utiliza la matriz de distancias euclídea al cuadrado (Ver anexo 7.2).

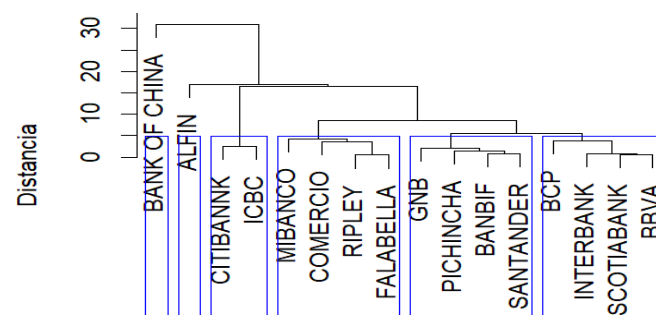
El historial de conglomeración (Tabla 1) muestra que en el primer paso se agrupan Banco Ripley (8) y Falabella (16), con una distancia menor de 0.4677543. Lo cual tiene sentido pues ambos son bancos enfocados en colocar tarjetas de crédito para financiar compras de consumo Retail.

Tabla 4*Historial de conglomeración método del centroide*

Método del centroide		
Primeros 5 pasos		
heigh	merge.1	merge.2
0.4677543	-8	-16
0.6388184	-3	-4
0.7902619	-7	-12
0.8332989	-2	2
1.3741452	-9	3

Elaboración propia

El dendograma en este caso, separa a Bank of China y Alfin banco como grupos de una sola observación. Esto se debe a que ambas entidades han iniciado hace poco tiempo operaciones en Perú (2020). Además, Bank of China está orientado al rubro de operaciones corporativas y comercio exterior, mientras Alfin Banco es un banco orientado a la banca personal.

Figura 8*Dendograma método del centroide*

Matriz de distancias al cuadrado

2.5.2.2. Método del vecino más cercano. Este método también genera como primera agrupación de pares a Banco Ripley (8) y Falabella (16):

Tabla 5

Historial de conglomeración método del vecino más cercano

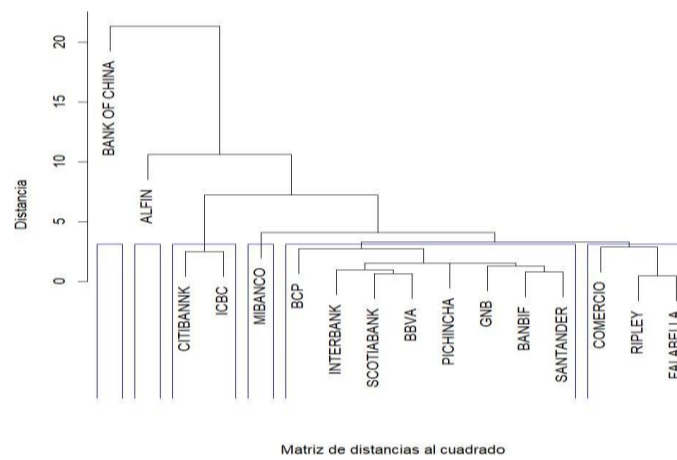
Método del vecino mas cercano		
Primeros 5 pasos		
heigh	merge.1	merge.2
0.4677543	-8	-16
0.6388184	-3	-4
0.7902619	-7	-12
0.9804070	-2	2
1.3144115	-11	3

Elaboración propia

El dendograma en este método, también indica la formación de 6 grupos. Se mantienen Bank of China, Alfin y Mibanco como grupos de uno, siendo este último agrupado de esta manera por tener un elevado nivel de colocaciones dirigido a segmentos micro y pequeños empresarios. Se colocan en un grupo de 6 observaciones a las entidades con mayor representación en el sistema como BBVA y BCP. Por otro lado, existe un grupo de 3 bancos conformado por Banco del Comercio, Ripley y Falabella.

Figura 9

Dendograma método del vecino más cercano



2.5.2.3. Método del vecino más lejano. Se ajusta el modelo del vecino más lejano.

Este método también genera como primera agrupación de pares a Banco Ripley (8) y

Falabella (16):

Tabla 6

Historial de conglomeración método del vecino más lejano

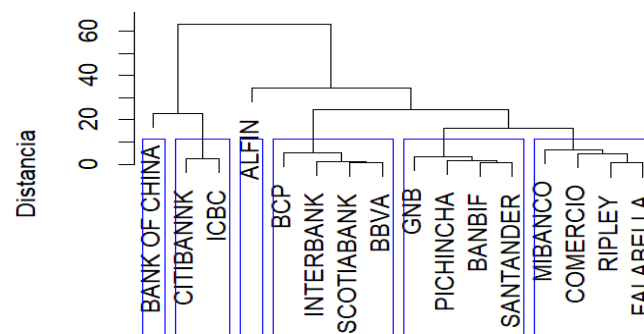
Método del vecino mas lejano		
Primeros 5 pasos		
heigh	merge.1	merge.2
0.4677543	-8	-16
0.6388184	-3	-4
0.7902619	-7	-12
1.0056001	-2	2
1.6188556	-9	3

Elaboración propia

Este método al igual que el vecino más cercano, tiende a dejar por separado a Bank of China y Alfin. Al igual que el método del centroide, agrupa las entidades con mayores niveles de colocación en el sistema correctamente: BCP, Interbank, Scotiabank y BBVA. Por otro lado, junta a Mibanco, Comercio, Ripley y Falabella debido a su orientación a la colocación de créditos masivos.

Figura 10

Dendograma método del vecino más lejano



Matriz de distancias al cuadrado

2.5.2.4. Método de vinculación promedio. Se mantienen las mismas agrupaciones de pares que los otros métodos.

Tabla 7*Historial de conglomeración método de vinculación promedio*

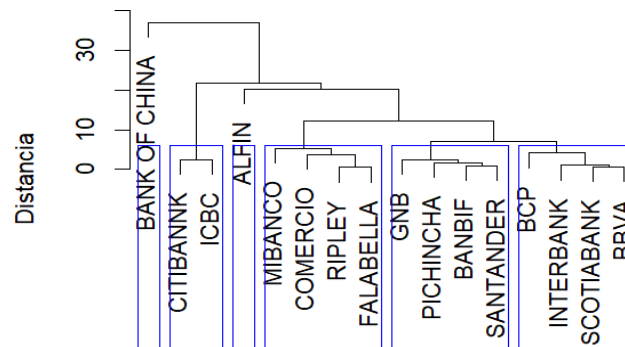
Método de vinculación promedio		
Primeros 5 pasos		
heigh	merge.1	merge.2
0.4677543	-8	-16
0.6388184	-3	-4
0.7902619	-7	-12
0.9930035	-2	2
1.5717106	-9	3

Elaboración propia

Este método mantiene las mismas agrupaciones que el método del vecino más lejano. Separa Bank of China y Alfin como grupos de a uno, los cuatro bancos con mayor concentración crediticia, seguido de bancos especializados en función a los productos financieros en específico que ofrece y finalmente bancos como Falabella y Ripley dirigidos al sector Retail.

Figura 11

Dendograma método vinculación promedio



Matriz de distancias al cuadrado

En todos los casos anteriormente presentados, se mantienen las mismas agrupaciones excepto por el método del vecino más cercano, que propone agrupar a bancos con mayor magnitud de colocaciones en el sistema financiero junto con otros bancos que ofrecen una gama de productos especializados como GNB que está orientado a créditos por convenios.

2.5.2.5. Método de Ward. Con este método la cantidad optima de conglomerados es 6, al igual que los anteriores tiende a agrupar los bancos más grandes, los de Retail como Falabella y Ripley, mientras que GNB comparte grupo con Pichincha, Banbif y Santander.

Tabla 8

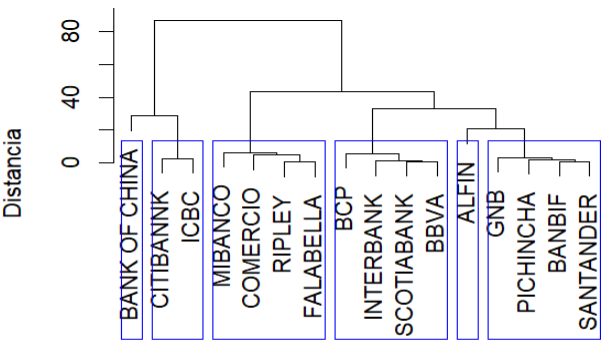
Historial de conglomeración método de ward

Método de Ward		
Primeros 5 pasos		
heigh	merge.1	merge.2
0.4677543	-8	-16
0.6388184	-3	-4
0.7902619	-7	-12
1.1110652	-2	2
1.8321935	-9	3

Elaboración propia

Figura 12

Dendograma método de Ward

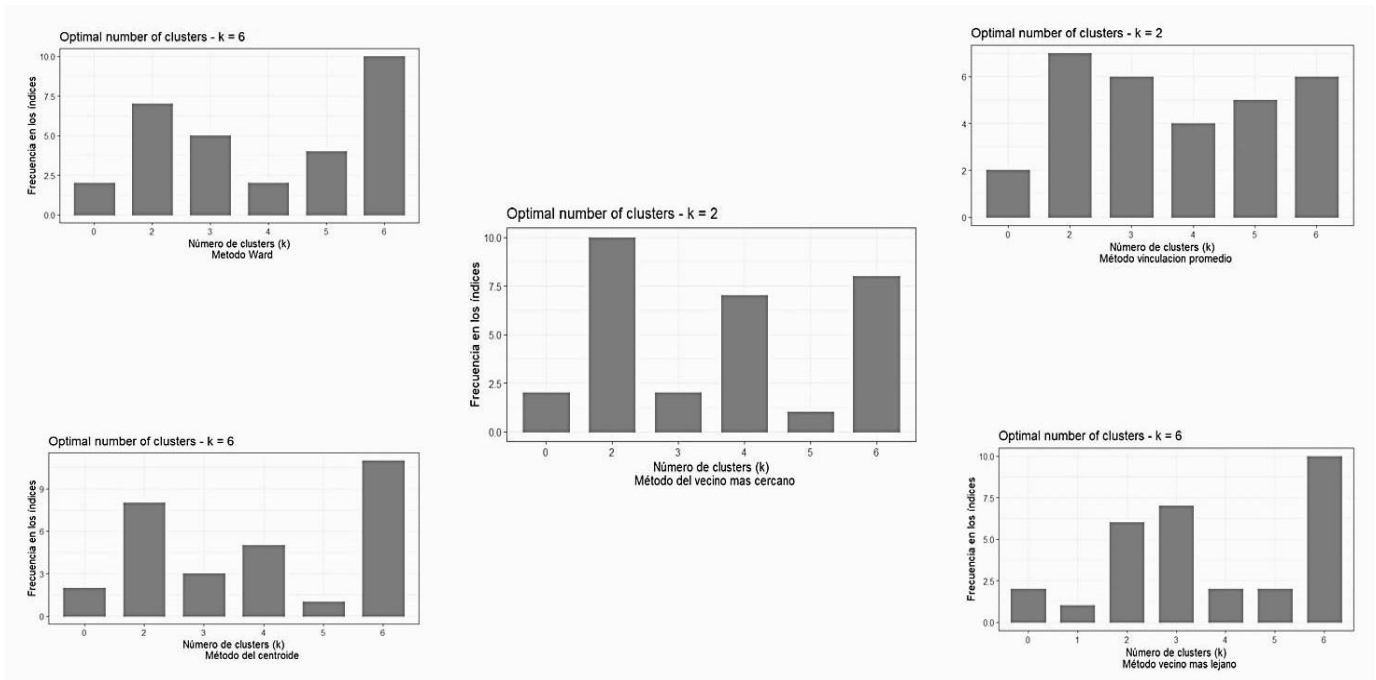


Matriz de distancias al cuadrado

Si se tiene en consideración el conocimiento del giro de negocio que tiene banco GNB, todos los métodos probados arrojan grupos coherentes, excepto el método del vecino más cercano. Así como GNB está enfocado en colocar créditos por convenios, Santander opta por colocar créditos vehiculares, Pichincha y Banbif están más especializados en la mediana empresa y créditos de consumo para personas naturales. Entonces es lógico que dichas entidades estén en el mismo grupo dado su apetito de colocar solo en determinados segmentos.

2.5.3. Elección del número de grupos

En la sección anterior, se han probado diferentes métodos de agrupación. Cada método sugiere un número de grupo, y se tienen diferentes indicadores para elegir. Los métodos de Hubert y Dindex son métodos gráficos para definir el número de grupos. Entonces se tienen los siguientes gráficos de frecuencias que sugieren un número por cada método probado:

Figura 13*Frecuencias para la elección del número de cluster*

Elaboración propia

Todos los métodos sugieren la formación de 6 conglomerados, excepto el método de vecino más cercano y el de vinculación promedio (2 grupos). Por mayoría se eligió formar 6 grupos, ya que tiene sentido mantener en un solo grupo entidades nuevas como ICBC Y Alfin Banco, otro de las entidades con mayor cuota de mercado en el sistema financiero (BBVA, BCP, Scotiabank, Interbank), bancos especializados en unos cuantos productos o segmentos como Santander (Vehiculares), GNB (Convenios), Banbif (préstamos) y por último el grupo de entidades que se enfocan en financiar compras de personas, personas naturales con negocio y microempresarios conformado por Falabella, Ripley y Mibanco.

III. Aportes más destacables a la Empresa/Institución

Se formaron 6 grupos de entidades financieras, cuya descripción se resume en la siguiente tabla:

Tabla 9

Formación de grupos

Cluster	Entidades	Diferenciación
Cluster 1	Bank of China	Entidad con menos de 5 años de instalación en el sistema financiero peruano. Se especializa en operaciones de comercio exterior y captación de depósitos, a la fecha cuenta con clientes del rubro corporativo y gran empresa
Cluster 2	Alfin Banco	Banco nuevo, empezó a operar entre 2021 y 2022. Está enfocado a colocar créditos a personas naturales emprendedoras y como parte de estrategia de posicionamiento, depósitos a plazo fijo con tasas promocionales.
Cluster 3	Citibank, ICBC	Ambos bancos prestan servicios bancarios comerciales, dirigidos a empresas del rubro mediana, grande y corporativo. Tienen mayor participación de depósitos como parte de sus activos (Entre 80% y 90% sobre el total de sus activos), así mismo, manejan niveles de rentabilidad y costes muy parecidos
Cluster 4	Mibanco, Comercio, Ripley, Falabella	Estas entidades tienen un manejo de costes estructurales alrededor del 2% y ganancias netas que van entre 8% y 10%, Esto se traduce a que por estar dirigido a un público masivo, los recursos destinados a operar el negocio hacen que sean parecidos entre sí. Sus mercado objetivo lo constituyen las personas naturales y mypes
Cluster 5	BCP, Interbank, Scotiabank, BBVA	Bancos con mayor cuota de mercado en el sistema financiero (Concentran más del 75% de las colocaciones). Cuentan con un apalancamiento sólido, lo cual hace que tengan niveles de APR más elevado que el resto de bancos. Los productos financieros que ofrecen están dirigidos tanto a personas naturales como empresas.
Cluster 6	GNB, Pichincha, Banbif, Santander	Bancos con representación media dentro del sistema, están después de los cuatro primeros. Manejan sus costos de manera muy similar, ya que cada uno de estos bancos se especializan en la colocación de productos en específico. Por ejemplo Banco GNB está enfocado en colocar créditos por convenios, mientras que Santander se especializa en créditos vehiculares. Así mismo, sus niveles de retorno oscilan entre 2% y 4%

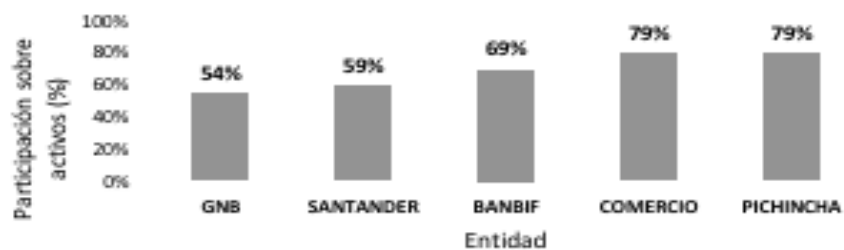
Elaboración propia

En específico, con relación a cada una de las dimensiones y variables analizadas se consiguió identificar lo siguiente para la entidad:

- El equilibrio financiero, representado por las variables participación de créditos y depósitos sobre activos, muestran que banco GNB, tiene más captación de depósitos que créditos. Del grupo donde quedó ubicado (Grupo 6), es el banco que tiene la menor representación de créditos sobre sus activos totales. Caso contrario sucede con la participación de depósitos, dado que GNB es el banco con mayor captación de depósitos en su grupo (78%). Esto permite concluir que es una entidad, que en relación con bancos como Santander o Banbif, centra sus esfuerzos en captar más depósitos en lugar de créditos.

Figura 14

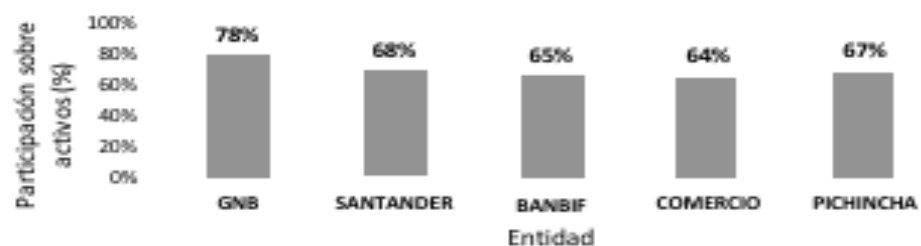
Participación de créditos sobre activos-Grupo 6



Elaboración propia

Figura 15

Participación de pasivos sobre activos-Grupo 6

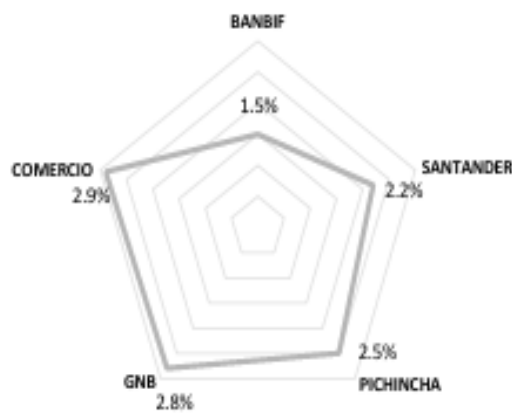


Elaboración propia

- La capacidad para generar margen y eficiencia en costos dice cuanta ganancia es capaz de generar la entidad financiera con el uso eficiente de recursos (gasto en personal, marketing, etc.). El grupo 6 tiene un ratio de costos promedio de 2.4%, y banco GNB se sitúa por encima de esta cifra con una eficiencia en costos del 2.8%, siendo superado por Banco del comercio.

Figura 16

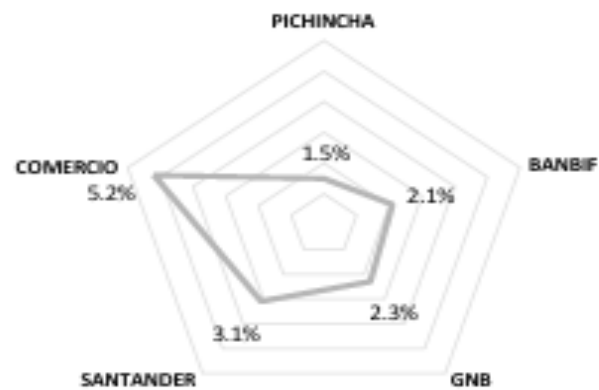
Ratio de costos-Grupo 6



Elaboración propia

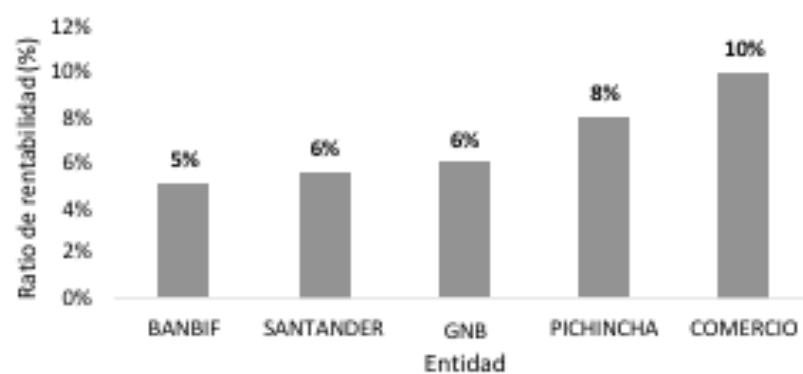
Elaboración propia

El margen es un indicador con un recorrido que va desde 1.5% hasta 5.2%. En el grupo 6, Banco del Comercio es la entidad que genera la mayor ganancia con un retorno de 5.2%. Le siguen Santander y GNB con 3.1% y 2.3% respectivamente.

Figura 17*Ratio de Margen-Grupo 6*

Elaboración propia

- La rentabilidad en el grupo 6, posiciona a Banco GNB como el tercer banco más rentable, superado por Pichincha y Comercio. Esto se debe a que dichas entidades tienen carteras crediticias más grandes en comparación a GNB.

Figura 18*Ratio de Rentabilidad-Grupo 6*

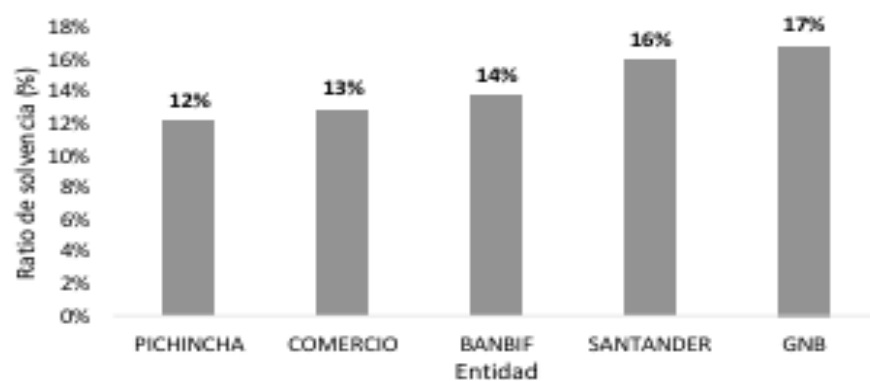
Elaboración propia

- Por último, el ratio de solvencia o ratio de capital es uno de los indicadores más importantes, ya que indica cuan solida es una entidad financiera. En el grupo 6, Banco GNB es la entidad más solvente (17%), lo cual indica que es una empresa financiera con recursos sólidos para afrontar situaciones difíciles como crisis económicas o sectoriales. Le sigue Banco Santander (16%), cuyo respaldo está en pertenecer a un grupo bancario sólido con amplia experiencia en el sector. El resto de entidades cumple con lo exigido por el regulador (mínimo del 8%).

Figura 19

Ratio de Solvencia-Grupo 6

Ratio de Solvencia-Grupo 6



Elaboración propia

IV. Conclusiones

- Con el modelo de análisis de conglomerados, se lograron identificar 6 grupos de entidades financieras que conforman ahora el modelo de perfil de riesgos del Banco GNB. Se tienen los siguientes grupos: dos de un solo componente (*Clúster 1 y 2*) que vienen a ser Bank of China y Alfin Banco, el grupo de los comerciales (*Clúster 3*) compuesto por Citibank y ICBC, el grupo de los bancos Retail y mype (*Cluster 4*) con Mibanco, Ripley y Falabella, los de mayor participación en el mercado como BCP, BBVA, Scotiabank e Interbank (*Cluster 5*), y los bancos especializados como GNB, Pichincha y Santander (*Cluster 6*).
- El análisis descriptivo realizado, sugiere que las entidades del sistema financiero peruano dedican en promedio más esfuerzos a la captación de depósitos que a la colocación de créditos. También se caracteriza por tener niveles de costos bajos que oscilan entre 1.7% y 3.7%, esto se traduce en que las entidades se caracterizan por hacer que su funcionamiento sea eficiente con el menor costo posible. Así mismo, en ratio de solvencia promedio que poseen las entidades es del 20%. Esto es un buen indicio de solidez por parte de las empresas bancarias, ya que lo exigido por el regulador a cada entidad es el 8%.
- El modelo desarrollado en base al análisis de conglomerados, dio lugar al modelo de perfiles de riesgos, como herramienta para la gestión de riesgos que sirve para identificar al banco GNB como parte de un grupo junto con otras entidades, con las cuales comparte características similares de equilibrio financiero, generación de margen, rentabilidad, y solvencia.

- Según el factor de equilibrio financiero o composición, se tiene que, de las 16 entidades, Citibank y ICBC tienen más del 80% en depósitos como parte de sus activos totales. Caso contrario sucede con Mibanco, que tiene el 85% de sus activos destinado a la colocación de créditos para emprendedores y Mype.
- Con relación al factor de generación de margen y eficiencia en costes. El cluster de los bancos Retail y mype reflejan claramente como Falabella, Ripley y Mibanco emplean sus recursos de forma similar para llevar adelante sus modelos de negocio.
- En términos de rentabilidad y perfil de riesgos medido por Activos ponderados por riesgos. Los bancos del *cluster* 5 son aquellos que tienen mayor diversidad de riesgos, y al mismo tiempo tienen una rentabilidad promedio favorable frente a la competencia.
- El factor solvencia indica cuan sólidas son las entidades para afrontar situaciones difíciles como crisis o recesiones. En ese sentido, las 16 entidades financieras, incluyendo banco GNB, tienen en promedio una solvencia del 20%, indicador que se maneja con holgura ante las recomendaciones de los estándares internacionales y la regulación peruana (8.5%).

V. Recomendaciones

- El modelo de conglomerados aplicado a las entidades del sistema financiero peruano se puede compartir con otras áreas del banco GNB como finanzas o negocios a fin de que puedan elaborar informes analíticos con base a agrupaciones formadas con sustento estadístico.
- Los grupos identificados se deben tener en cuenta como punto de partida, para elaborar otros análisis de indicadores importantes como la morosidad, cartera en riesgo, colocaciones totales entre otras. Esto a fin de que dichos análisis sean más enfocados en grupos de entidades con características similares.
- Debido a que el modelo dio como parte de sus resultados la formación de dos grupos con un solo integrante (ICBC y Alfin Banco), se recomienda actualizar el mismo con periodicidad anual y poner foco en estas entidades a fin de ver que grupo podrían integrar.

VI. Referencias bibliográficas

- Badii, & Castillo. (2007). Análisis de correlación canónica (ACC) e investigación científica. *Innovaciones de negocios*.
- Banco Central de Reserva del Perú. (s.f.). *Glosario*. Obtenido de Banco Central de Reserva del Perú: <https://www.bcrp.gob.pe/publicaciones/glosario/e.html>
- BBVA. (s.f.). *Informe con Relevancia Prudencial 2015*. Obtenido de <https://shareholdersandinvestors.bbva.com/microsites/pilarIII2015/es/2/pre.html>
- Blalock, H. M. (1996). *Estadística Social*. México.
- Brissett, P. R. (2018). *La importancia para la gestión de riesgos entidades financieras*. Bogotá, Colombia.
- Cattell, R. (1966). *Handbook of Multivariate Experimental Psychology*. Chicago: Rand McNally.
- Galicia. (2003). Los enfoques del riesgo de crédito. México: Instituto del Riesgo Financiero.
- Goldstein, D. (1984). Multivariate Analysis, methods and applications.
- Gurrea, M. (2000). Análisis de componentes principales. *Proyecto e-Math financiado por la Secretaría de Estado de Educación y Universidades (MECD)*.
- Jácome Ortega, X. O. (2015). Método y Análisis a la Supervisión Bancaria: Monitor de la Banca en Ecuador.
- Jainaga, García, & Aguirre. (s.f.). *Cash Management: Análisis empírico de una cultura financiera*.
- Khan, J. (2006). *Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. The counseling psychologist*.
- Lopez, & Dávila. (2020). *Propuesta para mejorar la calidad de la cartera crediticia en los procesos de aprobación y seguimiento en la banca comercial Interbank*. Lima, Perú.
- Lozares, C., & Pedro, L. (1991). El Análisis Multivariado: Definición, Criterios y Clasificación. *Universidad Autónoma de Barcelona*, 9-29.
- Meneses, J. (2019). *Introducción al Análisis Multivariante*. Barcelona: UOC.
- Parra, J. (1996). Modelo de análisis de correspondencias múltiples. *Revista de Ciencias Sociales*, 183-196.
- Peña, D. (2002). *Análisis de Datos Multivariantes*. Madrid.
- Perú, B. C. (s.f.). *Glosario de Términos económicos*. Obtenido de [https://www.bcrp.gob.pe/publicaciones/glosario/s.html#:~:text=Sistema%20financiero%20bancario%20\(Banking%20financial,banca%20de%20fomento%20en%20liquidaci%C3%B3n](https://www.bcrp.gob.pe/publicaciones/glosario/s.html#:~:text=Sistema%20financiero%20bancario%20(Banking%20financial,banca%20de%20fomento%20en%20liquidaci%C3%B3n)
- Quero, M., & Inciarte, K. (2012). Clasificación de técnicas estadísticas multivariantes. *Telos*, 275-286.
- Ruiz Aranibar, G. (2013). Análisis discriminante. *Revista Varianza*, 1-10.

SBS, S. d. (2011). Reglamento para la Gestión de Riesgo de Crédito 3780-2011. Lima, Perú. Aldás, J., & Uriel, E. (2017). *Análisis Multivariante aplicado con R*. Madrid, España: Paraninfo.

SBS. (2015). Glosario de términos e indicadores financieros. Obtenido de <https://intranet2.sbs.gob.pe/estadistica/financiera/2015/Setiembre/SF-0002-se2015.PDF#:~:text=13.,iniciar%20acci%C3%B3n%20judicial%20o%20arbitral>.

Sola, I., & Ruiz, D. (2015). *Marco competitivo del sector bancario español en la union bancaria*. Madrid, España.

Vila, Torrado, & Reguant. (2019). Análisis de regresión lineal múltiple con SPSS: un ejemplo práctico. *REIRE*.

VII. Anexos

7.1. Base de datos de entidades financieras utilizados para el análisis

FECHA	ENTIDAD	APR	PART. CREDITOS	PART. DEPÓSITOS	RENTABILIDAD	COSTOS ESTRUCTURALES	MARGEN NETO	PERFIL	RATIO CAPITAL	GRUPO
31/12/2021	BCP	152.4	65.8	65.5	4.9	0.8	3.4	82	14.9	Grupo 5
31/12/2021	INTERBANK	57.6	62.9	64.9	6	1.3	4	84.2	15.9	Grupo 5
31/12/2021	SCOTIABANK	72.5	73.6	57.5	5.3	1.3	3.1	94.1	14.7	Grupo 5
31/12/2021	BBVA	84.4	73.2	62.2	4.9	0.7	3.5	83	14.1	Grupo 5
31/12/2021	COMERCIO	3.3	78.8	63.6	9.9	2.9	5.2	145.3	12.8	Grupo 6
31/12/2021	CITIBANK	3.6	23.4	79.9	2.9	0.4	2.7	46.7	20.7	Grupo 3
31/12/2021	BANBIF	15.6	68.6	65.2	5	1.5	2.1	81.9	13.7	Grupo 6
31/12/2021	RIPLEY	2.9	67	65	15.2	2.4	8.2	136.1	13.6	Grupo 4
31/12/2021	PICHINCHA	11.3	78.9	67.1	7.9	2.5	1.5	99.8	12.1	Grupo 6
31/12/2021	MIBANCO	14.1	85	52.7	13.3	1.6	9	88.9	16.4	Grupo 4
31/12/2021	GNB	4.9	53.9	78.2	5.9	2.8	2.3	82	16.8	Grupo 6
31/12/2021	SANTANDER	8.2	58.6	68	5.5	2.2	3.1	86.4	15.9	Grupo 6
31/12/2021	ICBC	1.4	14.9	90.2	0.7	0.1	0.6	24	32.2	Grupo 3
31/12/2021	BANK OF CHINA	0.3	3.8	51.5	0.8	0	0.7	49.7	92.8	Grupo 1
31/12/2021	ALFIN	1.2	56.9	84.7	10.4	4.3	-0.4	184.3	4.1	Grupo 2
31/12/2021	FALABELLA	6	68.1	68.7	15.3	2	9.5	142.8	14.2	Grupo 4

Fuentes: Estadísticas SBS

7.2. Matriz de distancias euclídeas al cuadrado

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00															
2	5.20	0.00														
3	4.43	0.98	0.00													
4	2.72	1.01	0.64	0.00												
5	19.71	7.15	8.00	11.13	0.00											
6	18.24	8.48	13.77	11.92	21.51	0.00										
7	10.76	1.52	2.56	3.39	6.44	7.61	0.00									
8	23.59	10.30	12.87	15.24	2.89	23.99	11.74	0.00								
9	14.36	3.80	4.62	6.65	3.27	13.45	1.62	8.85	0.00							
10	19.99	8.70	9.45	11.22	6.59	25.19	10.67	4.05	10.45	0.00						
11	16.66	5.20	8.77	9.84	7.16	7.20	3.24	11.89	2.78	16.31	0.00					
12	12.96	2.15	4.25	5.55	4.83	7.20	0.79	9.27	1.52	10.49	1.31	0.00				
13	27.20	17.98	24.71	21.60	35.56	2.50	16.15	39.48	23.31	41.62	13.63	15.82	0.00			
14	38.84	28.67	30.93	31.42	45.14	21.35	28.03	49.39	36.93	44.47	32.94	28.47	22.75	0.00		
15	34.62	21.45	24.25	27.93	10.90	29.43	18.18	17.46	11.21	32.24	10.60	14.44	38.63	63.31	0.00	

7.3 Promedio de variables por grupo

Variable	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Total Entidades
APR	0.30	1.20	2.50	7.67	91.73	8.66	27.48
Participacion de créditos	3.80	56.90	19.15	73.37	68.88	67.76	58.34
Participacion de depósitos	51.50	84.70	85.05	62.13	62.53	68.42	67.81
Rentabilidad	0.80	10.40	1.80	14.60	5.28	6.84	7.12
Costos	0.00	4.30	0.25	2.00	1.03	2.38	1.68
Margen neto	0.70	-0.40	1.65	8.90	3.50	2.84	3.66
Perfil	49.70	184.30	35.35	122.60	85.83	99.08	94.45
Ratio de Solvencia	92.80	4.10	26.45	14.73	14.90	14.26	20.31