

Entrega resultados

En este documento se hará presentación de los resultados preliminares del proyecto.

En esta sección se presentan los principales hallazgos del análisis exploratorio de los datos financieros del sistema bancario colombiano. El objetivo del EDA es entender la calidad de los datos, identificar patrones, detectar valores atípicos y construir una primera interpretación económica antes de aplicar modelos predictivos.

Descripción general de los datos usados.

Para la realización de nuestro proyecto se recolectaron los siguientes datos: Activo, pasivo, patrimonio, activos líquidos mensuales promedio (ALM PromedioMes), requerimientos de liquidez promedio (RLN PromedioMes), tasa de política monetaria (TPM) y el IRL. En este proyecto usamos 13 diferentes entidades bancarias y se usó las diferentes variables mencionadas anteriormente en un periodo determinado que va de marzo del 2020 a diciembre del 2022.

El dataframe que trabajamos encontramos una columna con el nombre de “Entidad” el cual nos da el nombre de la entidad bancaria como por ejemplo “DAVIVIENDA”. Esta columna sirvió como columna llave para relacionar las bases de datos que desde un inicio venían separadas. Para poder unir las bases de datos mediante esta columna nos aseguramos de que los nombres coincidieran, que no haya errores tipo que en una base el nombre del banco sea “Davivienda” y en la otra base sea “DAVIVIENDA”, para evitar esos errores se decidió manejar esta columna en mayúsculas todos los nombres y además se eliminaron prefijos innecesarios como lo es “Banco” al inicio de cada entidad para mejor visualización del nombre de la entidad.

En este dataframe se decidió usar 2 columnas que relacionan cada dato a un mes y año a cada banco, con el fin de poder ver el comportamiento de nuestras variables cuantitativas a lo largo del tiempo para cada entidad. La razón por la cual se decidió hacer en 2 columnas separadas fue principalmente para poder analizar con mayor facilidad los datos en determinados ciclos económicos que sufre la economía colombiana en determinados periodos del año donde las dinámicas económicas son diferentes.

Ahora vamos a pasar con nuestras variables cuantitativas, en primer lugar, encontramos los “Activos”, es el total de activos de cada banco, que son recursos y bienes que posee la entidad incluyendo cartera de créditos, inversiones, efectivo, etc. Su uso en esta investigación sirve para medir el tamaño de cada entidad y evaluar su solidez patrimonial, recordando que el activo es igual al pasivo más el patrimonio y además la mayor parte de los activos de un banco están compuestos por deuda.

Como segunda variable cuantitativa encontramos los “Pasivos”. Los pasivos en una entidad bancaria son el total de las obligaciones que tiene la entidad donde principalmente son depósitos de clientes y deudas con otras entidades, se entiende que el valor de los pasivos en una entidad bancaria sea alta debido que su principal actividad económica es emitir y captar deuda como motor de la economía con el fin de dar dinamismo al mercado.

Como tercera variable cuantitativa tenemos el “Patrimonio”, el patrimonio es el capital propio del banco. Esta variable se puede usar para medir la capacidad que tiene un banco para absorber perdidas y ayuda a constituir un indicador clave de solvencia. Cabe aclarar que el banco no usa sus propios recursos para generar valor, sino que usa los recursos que capta mediante deuda para generar valor, es por eso por lo que el patrimonio no se usa en la actividad del análisis de eficiencia en su actividad económica.

Nuestra cuarta variable cuantitativa es “ALM PromedioMes”, esta variable nos indica los activos líquidos mensuales promedios que cuenta el banco en un momento específico del tiempo y representa los recursos fácilmente convertibles en efectivo ya sea el efectivo en caja, depósitos en el banco de la república y títulos liquidez. Esta variable se usa principalmente para medir la disponibilidad de liquidez inmediata de cada banco.

La quinta variable cuantitativa es el “RLN PromedioMes”, es la suma de las salidas esperadas de efectivo de una entidad ya sea en obligaciones, retiros y vencimientos de acuerdos, y a eso se le resta la suma de todas las entradas proyectadas. Esta variable se usa principalmente para medir la presión de liquidez que enfrenta la entidad en un periodo dado.

La sexta variable cuantitativa es la “TPM”, esta variable es la tasa de referencia del banco de la república que sirve como principal instrumento para la política monetaria y da claras señales al mercado de que situación se atraviesa. Esta tasa esta directamente relacionada con las demás tasas del mercado mediante los mecanismos de transmisión y fija las dinámicas del mercado dependiendo la situación y el objetivo del Banco Central.

Y finalmente nuestra variable dependiente la “IRL”, este es el indicador que relaciona la capacidad de cubrir los requerimientos netos de liquidez con los activos líquidos disponibles. Este indicador se encuentra en la superintendencia financiera o se puede construir mediante una fórmula que es $IRL = \frac{ALM}{RLN}$

Si este indicador es mayor a 1 el banco tiene suficientes activos líquidos para cubrir sus obligaciones, si es aproximadamente igual a 1 tiene liquidez ajustada y requiere

vigilancia sobre su situación y si es menor a 1 tiene un riesgo de liquidez alto y el banco no podrá cubrir completamente sus salidas.

Análisis exploratorio.

✦ Dimensiones del dataset: (317, 10)	✦ Valores nulos por columna:
✦ Tipos de datos:	Entidad 0
Entidad object	Activo 0
Activo float64	Pasivo 0
Pasivo float64	año 0
año int64	mes 0
mes object	ALM PromedioMes 0
ALM PromedioMes float64	RLN PromedioMes 0
RLN PromedioMes float64	IRL 0
IRL float64	Patrimonio 0
Patrimonio float64	TPM 0
TPM float64	dtype: int64
dtype: object	✦ Registros duplicados: 0

En la tabla anterior encontramos como definimos los objetos en nuestro dataframe, donde Activo, Pasivo, Patrimonio, ALM, RLN e IRL están definidos como “float” para realizar análisis cuantitativos mas adelante. Por su parte las demás variables se definieron como “object” ya que son únicamente variables que relacionan cada dato a un mes o banco. Año por su parte se definió como entero para poder hacer análisis transversales a lo largo del tiempo.

Además, identificamos una base con 317 observaciones en nuestras 10 columnas definidas, asegurándonos también que no se encontraran valores nulos, con el fin de realizar un buen análisis y no tener algún sesgo en la distribución de los datos que puedan afectar la muestra.

Comportamiento de los datos.

Una vez definida nuestra base de datos, definiendo correctamente el tipo de los datos vamos al análisis exploratorio de los datos.

Vamos a realizar un primer análisis de la media, mediana, desviación estándar, identificando mínimo, máximo y sus respectivos percentiles para finalmente verlos en una grafica distribuidos en una caja de bigotes y posteriormente realizaremos un análisis de la correlación de los datos con nuestra variable endógena.

```

Estadísticas de Activo:
Media: 42,245,822.97
Mediana: 28,794,019.82
Desviación estándar: 39,962,036.95
Mínimo: 0.00
Máximo: 140,619,856.20
Q1 (25%): 15,872,476.70
Q3 (75%): 53,830,769.76
IQR: 37,958,293.06

Estadísticas de Pasivo:
Media: 37,103,603.06
Mediana: 26,105,169.13
Desviación estándar: 34,331,161.55
Mínimo: 0.00
Máximo: 126,230,902.64
Q1 (25%): 14,030,288.25
Q3 (75%): 49,250,336.17
IQR: 35,220,047.92

Estadísticas de Patrimonio:
Media: 5,142,219.91
Mediana: 2,311,159.21
Desviación estándar: 6,282,363.50
...
Máximo: 12.00
Q1 (25%): 1.75
Q3 (75%): 5.00
IQR: 3.25

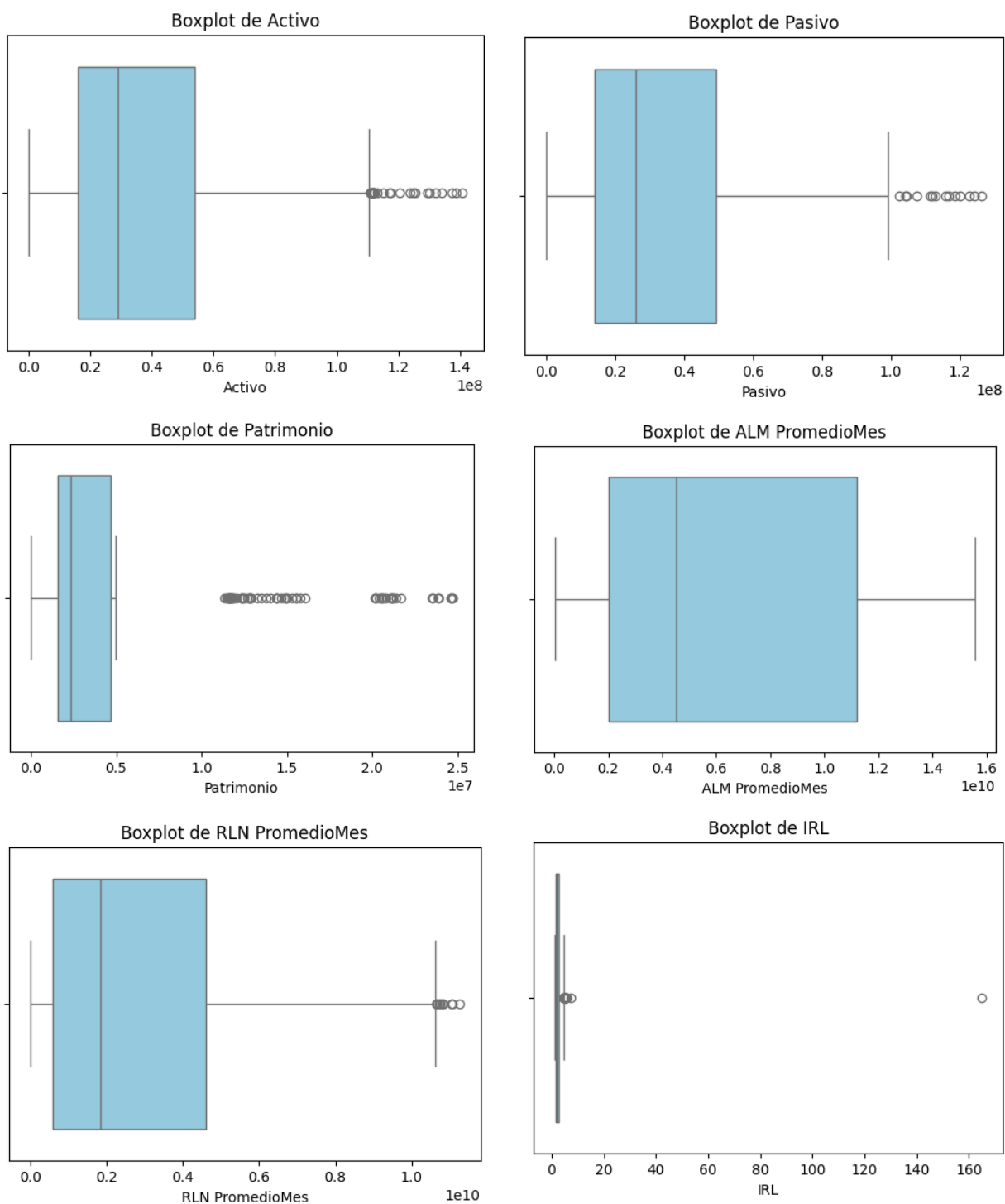
```

Como se evidencia en la tabla anterior encontramos que tanto en los activos, pasivos y patrimonio su media y su mediana están muy alejadas entre sí, lo cual nos da indicios que existen sesgos en la muestra. En los 3 casos la media es mayor a la mediana por lo que intuimos que existen pocos bancos grandes que elevan el promedio de la muestra generando una asimetría positiva. Esto confirma que el sistema esta dominado por unos pocos bancos grandes mientras muchos más son pequeños. Lo anterior también lo vemos en como se comportan los pasivos que también es muy asimétrica y tiene mucha distribución y como los pasivos en los bancos representa en su mayoría los depósitos del publico este resultado refleja que el ahorro de la economía esta muy concentrado en los bancos más grandes. Por su parte el patrimonio también presenta el mismo sesgo reafirmando que existe diferencias significativas entre los bancos y nos muestra que la capacidad de absorción de perdidas frente a choques es muy diferente entre los bancos, donde hay entidades que cuentan con un patrimonio robusto y otros que por el contrario cuentan con un patrimonio más frágil.

Del análisis anterior encontramos diferencias significativas entre los bancos y esto no será de utilidad para poder comprender que el sistema financiero colombiano no

funciona bajo competencia perfecta y no todos tienen las mismas capacidades de afrontar y responder frente a las crisis.

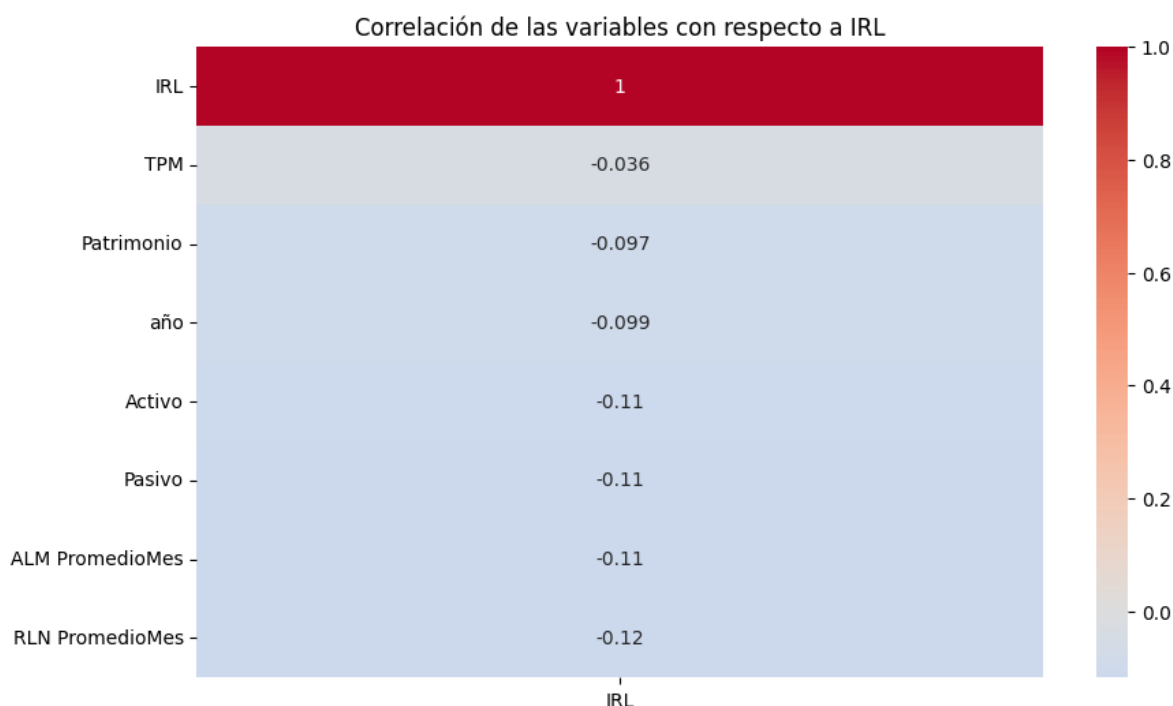
A continuación, se presentaran los “boxplot” para una más fácil interpretación:



Siguiendo la logica del analisis anterior encontramos que efectivamente existe valores atipicos a la derecha de la distribucion que afectan el mercado y absobern gran parte del mercado

debido que poseen un poder de mercado superior a las de otras entidades, esto tambien puede ser señal de deficiencias en el sistema que no permite la libre competencia dentro del mercado. Todo lo anterior bajo la premisa que el sector financiero colombiano esta altamente regulado y no funciona como un mercado convencional.

Por otro lado identificamos que los indicadores no estan tan dispersos lo cual es buena señal que el mercado es sano y de momento la mayoría de los individuos tienden a seguir el mismo riesgo de liquidez pese a sus diferencias. A excepcion de un dato atipico.



Por último, analizamos las correlaciones de nuestras variables independientes con nuestra variable dependiente y encontramos poca correlación. Cosa que va a hacer que nuestra modelo no sea fiable y será de vital importancia ajustar este problema de correlación usando otras variables más correlacionadas con el IRL.