

ETL Bank Data Analysis

World's Top 10 Banks

– Insight & Analytics Report



Submitted by : Suyash Nagar

September 2025

Contents

1.	Overview	3
2.	Business Value	3
3.	DataSet	4
4.	Technical Stack.....	4
5.	Data Preparation.....	5
5.1	Logging Function	5
5.2	Initialize Spark session	5
5.3	Wikipedia Bank Data Extraction (URL)	5
5.4	Load exchange rates (CSV)	6
6.	Data Cleaning.....	6
6.1	Handle Missing Values.....	6
6.2	Fixing Columns.....	6
6.3	Handle Outliers	7
6.4	Saving Data to AWS S3.....	7
7.	Exploratory Data Analysis.....	8
7.1	Conversion from PySpark to Pandas DataFrame	8
7.2	Market Capitalization.....	8
7.3	Top 10 Banks	8
7.4	Market Cap vs Bank Ranking	9
7.5	Market Cap Analysis	9
7.6	Market Cap Quartile Distribution.....	10
7.7	Cumulative Market Share Analysis	10
7.8	Categorising Banks.....	11
7.9	Visualise Market Share Distribution.....	11
8.	Banking Data ETL Querying.....	12
8.1	Data Processing	12
8.2	Market Capitalization Analysis	13
8.3	Market Concentration Analysis	14
8.4	Market Capitalization Distribution	16
8.5	Comparative Size Analysis	18
8.5	Market Growth Analysis	19
8.6	Market Dominance Analysis.....	20

8.7	Segment-Wise Bank Analysis	21
8.8	Performance Dashboard.....	22
9.	Report Conclusion: Strategic Insights from Global Bank Market Leader	23

1. Overview

- Develop a structured approach to analyzing market capitalization trends, currency conversions, and global banking performance
- Task is to examine historical banking data to derive actionable insights that can drive strategic growth. Your analysis will help identify the top 10 largest banks by market capitalization, convert market cap values into multiple currencies (USD, GBP, EUR, INR), and store the processed data for easy retrieval.
- By leveraging big data analytics and cloud services, financial institutions can streamline operations, enhance decision-making, and maximize revenue opportunities.

2. Business Value

The banking industry operates in a highly competitive and globalized market where financial institutions must continuously monitor their market position, performance, and currency fluctuations. To stay competitive, banks must leverage data-driven insights to optimize their financial strategies, assess market trends, and make informed decisions. In this assignment, you will analyse banking data to uncover patterns in market capitalization, currency conversions, and global rankings. With Apache Spark's ability to handle large datasets efficiently, financial institutions can process vast amounts of data in real-time, helping them make faster and more informed decisions.

As an analyst at a financial institution, your task is to examine historical banking data to derive actionable insights that can drive strategic growth. Your analysis will help identify the top 10 largest banks by market capitalization, convert market cap values into multiple currencies (USD, GBP, EUR, INR), and store the processed data for easy retrieval. To enhance scalability and accessibility, you will utilize AWS S3 for data storage, ensuring seamless integration with cloud-based analytics tools. By leveraging big data analytics and cloud services, financial institutions can streamline operations, enhance decision-making, and maximize revenue opportunities.

3. DataSet

The data for this project can be accessed from the following [wikipedia link](#).

The dataset used in this analysis comprises global banking data extracted from a Wikipedia page (List of Largest Banks) archived as of September 2023.

It includes rankings of the world's largest banks based on market capitalization (in USD billions) and is structured into a single table with three columns. The data was scraped programmatically, cleaned, and transformed using PySpark to support cross-currency analysis. Exchange rate data (for USD, GBP, EUR, and INR) was sourced from a CSV file to enable currency conversions. The processed dataset is stored in CSV and SQLite formats for accessibility and efficient querying.

The dataset consists of a single table (Largest_banks) with the following three key attributes:

- **Rank:** Numerical position of the bank based on market capitalization.
- **Bank Name:** Name of the financial institution.
- **Market Cap (USD Billion):** Market valuation of the bank in USD billions.

This dataset enables analysis of global banking trends, cross-currency valuations, and the relative market dominance of financial institutions.

An additional data on market exchange rates is available to compare the market cap dealing with different currencies.

4. Technical Stack

- a. Cloud : Amazon Web Service (AWS EC 2)
- b. Storage : AWS S3 and Google Content Drive
- c. Code : Pyspark and Python

5. Data Preparation

5.1 Logging Function

```
6. #Logging Function
7. def logging_progress(msg):
8.     timeformat = '%Y-%h-%d-%H:%M:%S'
9.     timestamp = datetime.now().strftime(timeformat)
10.
11.     logpath='/content/sample_data/IIIT_Assignments/Banking/Banking_Analys
12.     is.txt'
13.     os.makedirs(os.path.dirname(logpath), exist_ok=True)
14.     with open(logpath, 'a') as f:
15.         f.write(f"{timestamp}:{msg}\n")
```

5.2 Initialize Spark session

Initiated 'BankingDataAnalysis'

5.3 Wikipedia Bank Data Extraction (URL)

```
wiki_url =
"https://web.archive.org/web/20230908091635/https://en.wikipedia.org/wiki/Lis
t_of_largest_banks"
```

- `pd.read_html(wiki_url)` : Load Data using URL
- `wiki_spark_df = spark.createDataFrame(wiki_df)` : Convert to PySpark
- `wiki_spark_df.show(3)`

```
Top 3 Rows
+---+-----+-----+
|Rank|      Bank_Name|Market_Cap_USD_Billion|
+---+-----+-----+
| 1|  JPMorgan Chase|        432.92|
| 2|  Bank of America|        231.52|
| 3|Industrial and Co...|        194.56|
+---+-----+-----+
```

- `wiki_spark_df.printSchema()`

```
Bank Schema
root
|-- Rank: long (nullable = true)
|-- Bank_Name: string (nullable = true)
|-- Market_Cap_USD_Billion: double (nullable = true)
```

5.4 Load exchange rates (CSV)

- `exchange_rate_df = spark.read.csv("/content/sample_data/IIIT_Assignments/Banking/exchange_rate(in).csv", header=True, inferSchema=True)` from Google Drive

Exchange Rate Data	
Currency	Rate
EUR	0.93
GBP	0.8
INR	82.95

6. Data Cleaning

6.1 Handle Missing Values

```
# Check for missing values
missing_values = wiki_spark_df.select(
    [count(when(isnull(c),c)).alias(c) for c in wiki_spark_df.columns]
)
logging_progress("Missing Values Checked")
missing_values.show()
```

Rank	Bank_Name	Market_Cap_USD_Billion
0	0	0

There are no missing values

6.2 Fixing Columns

- There are no Duplicates
- Column format is correct, and no conversion is required

6.3 Handle Outliers

```
# Write code for outlier analysis
quantiles = wiki_spark_df.approxQuantile("Market_Cap_USD_Billion", [0.25, 0.75], 0.1)
Q1, Q3 = quantiles[0], quantiles[1]
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
outliers = wiki_spark_df.filter((col("Market_Cap_USD_Billion") < lower_bound) | (col("Market_Cap_USD_Billion") > upper_bound))
outliers.show()
```

Rank	Bank Name	Market_Cap_USD_Billion
1	JPMorgan Chase	432.92

There is presence of outlier JP

- JP Morgan Chase, with \$432.92B, stands out to be high outlier
- JP Morgan dominance is statistically exceptional, making it as a global leader

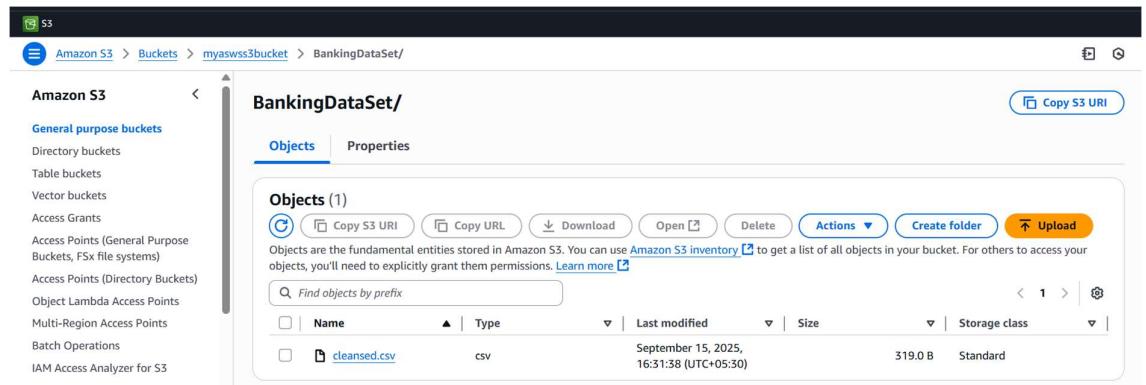
6.4 Saving Data to AWS S3

Save cleansed Data to S3 Storage in AWS EC2

- Install boto3 for AWS
- Define a function, declare
AWS_ACCESS_KEY_ID
AWS_SECRET_ACCESS_KEY

```
df= wiki_spark_df.toPandas()
# Upload Cleansed Data to AWS S3
upload_df(df, "cleansed") # upload to S3
```

Saved to Google Drive at: /content/sample_data/IIIT_Assignments/cleansed.csv
Uploaded to S3: s3://myaws3bucket/BankingDataSet/cleansed.csv



Amazon S3 > Buckets > myaws3bucket > BankingDataSet/

BankingDataSet/

Objects (1)

Name	Type	Last modified	Size	Storage class
cleansed.csv	csv	September 15, 2025, 16:31:38 (UTC+05:30)	319.0 B	Standard

File Successfully uploaded to AWS S3

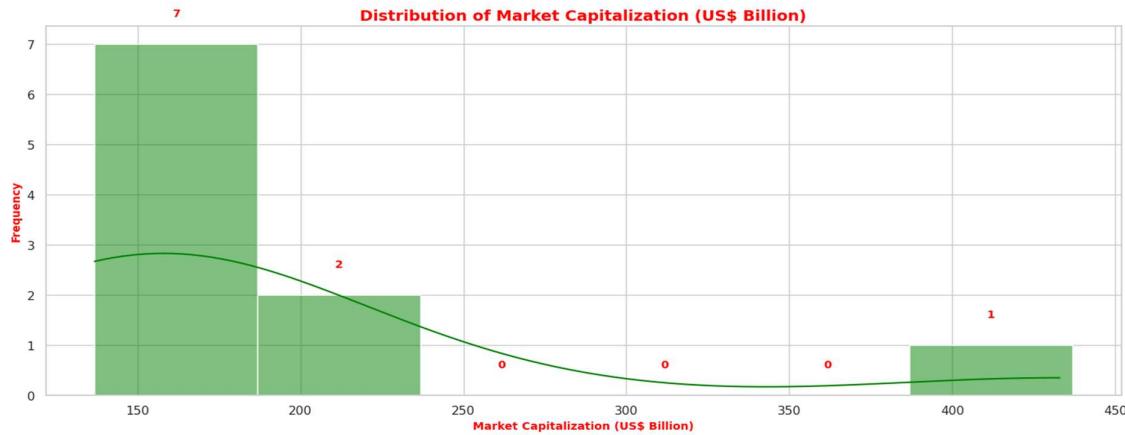
7. Exploratory Data Analysis

7.1 Conversion from PySpark to Pandas DataFrame

```
bank_pd_df = wiki_spark_df.toPandas()
```

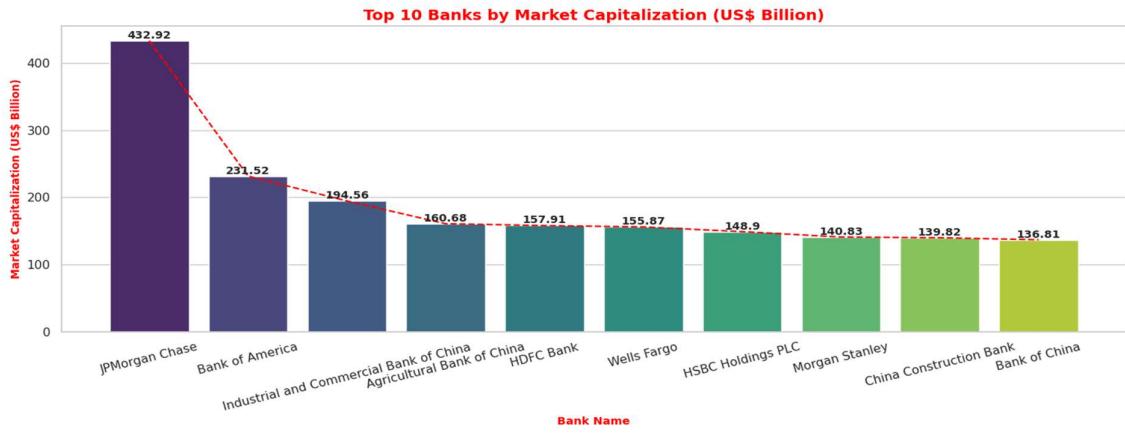
Convert PySpark to pandas data frame for visualization

7.2 Market Capitalization



- 70% of Banks lies in lower range of 125-175 billion, followed by 20% in 175 – 225 billion and only 10% in 400+ Billion
- In 225- 375 billion range there are no Banks

7.3 Top 10 Banks



- JP Morgan being leader by a wide margin, clearly dominating global banks
- With Redline clear indication of sharp decline can be observed
- Overall market is dominated by US and China Bank and India being there with HDFC

7.4 Market Cap vs Bank Ranking



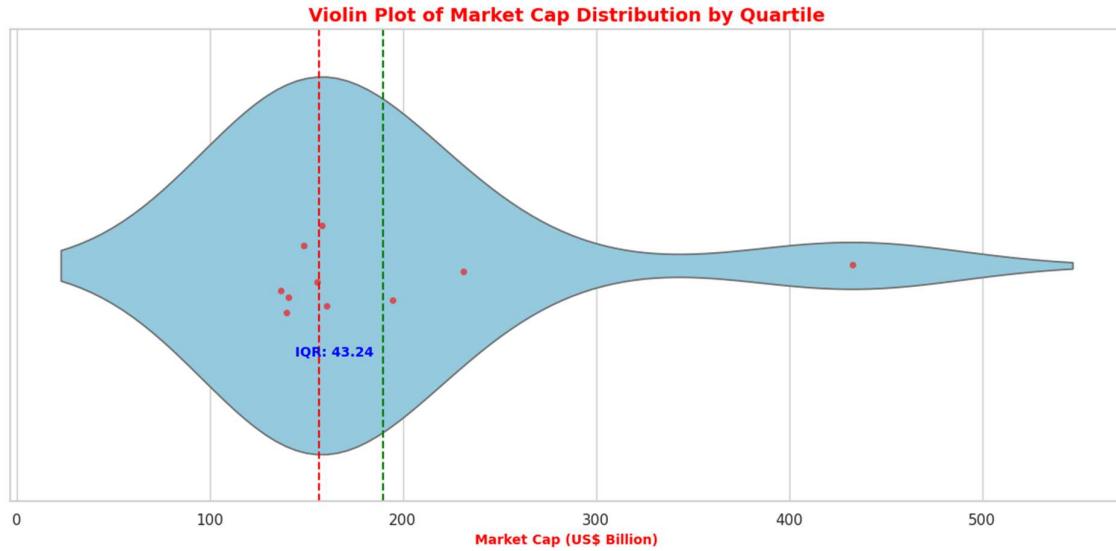
- Bubble Chart between Bank global ranking and Market Capitalization (USD), represents the magnitude in capitalization
- Clear steep drop from rank 1 and 2 is visible by sheer size and drop
- Lower order banks are in near cluster

7.5 Market Cap Analysis



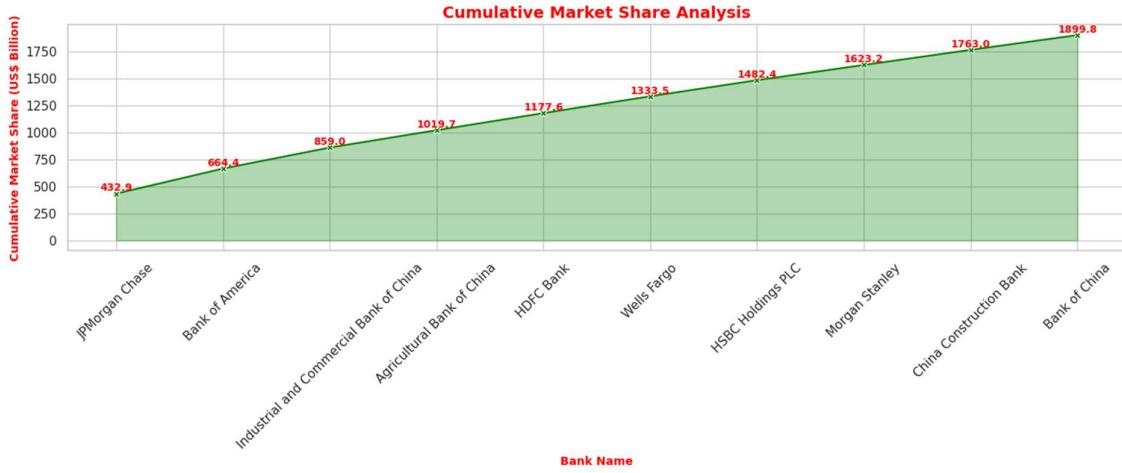
- With \$432.92B, JP Morgan chase is statistically isolated and outlier
- Median: ~\$156.89B
- Mean: ~\$189.98B (pulled upward by JPMorgan)
- Outlier: JPMorgan Chase, far beyond the upper quartile

7.6 Market Cap Quartile Distribution



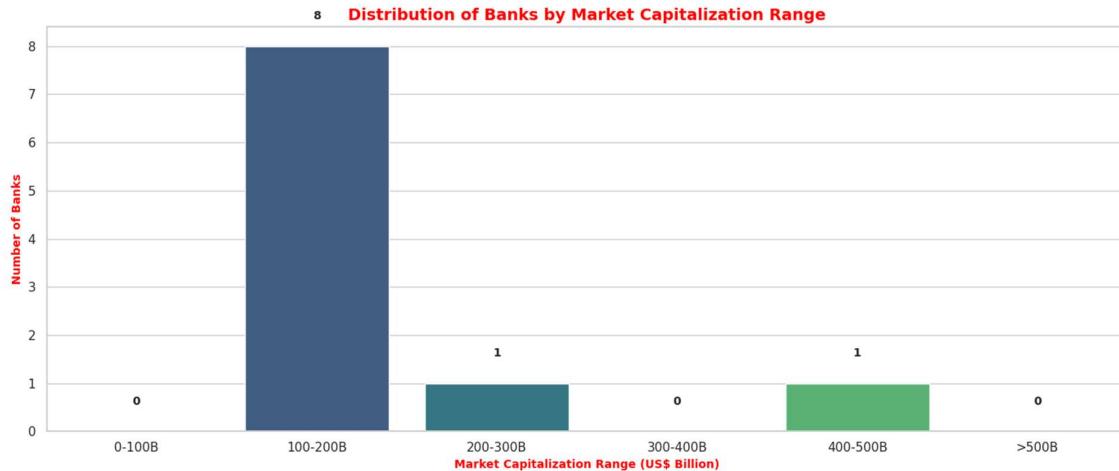
- Violin plot is bottom heavy, clearly showing right-skewed distribution
- Most banks are concentrated in lower quartile, which could be seen with density bulge(150-200) Billion towards bottom
- Long tail to right due to heavy outlier of JP Morgan Chase

7.7 Cumulative Market Share Analysis



- J.P Morgan Chase is setting initial jump of (432) Billion setting dominant tone
- The slope flattens slightly after JP Morgan and Bank of America, showing lower contribution from rank 3 onwards
- The final cumulative total reached \$1889.8 Billion

7.8 Categorising Banks



- Bar chart yet again confirms strong concentration of banks in lower ranks with 100-200 Billion range
- Only Bank of America is in 200+ Club and JP Morgan leading club by 400+ making it a global leader

7.9 Visualise Market Share Distribution



- Donut Chart highlights market share hierarchy, with JP Morgan taking biggest slice of cake with 22.8%
- The next major contributor is Bank of America with 12.2
- Remaining banks share between 10% - 7%

8. Banking Data ETL Querying

8.1 Data Processing

8.1.1 Initialize Spark session and Read Data from AWS S3

Read Cleansed data from S3 and combine with Exchange rates

8.1.2 Process Exchange Rate

Rank	Bank_Name	Market_Cap_USD_Billion	Market_Cap_EUR_Billion	Market_Cap_GBP_Billion	Market_Cap_INR_Billion
1	JPMorgan Chase	432.92	402.62	346.34	35910.71
2	Bank of America	231.52	215.31	185.22	19204.59
3	Industrial and Co...	194.56	180.94	155.65	16138.75
4	Agricultural Bank...	160.68	149.43	128.54	13328.41
5	HDFC Bank	157.91	146.86	126.33	13098.63
6	Wells Fargo	155.87	144.96	124.7	12929.42
7	HSBC Holdings PLC	148.9	138.48	119.12	12351.26
8	Morgan Stanley	140.83	130.97	112.66	11681.85
9	China Construction...	139.82	130.03	111.86	11598.07
10	Bank of China	136.81	127.23	109.45	11348.39

- Process Currency for EUR, GBP and INR

8.1.3 Write Data to sqlite3

```
▶ import sqlite3
# Write data to SQL Database
def write_SQLLite(df, dbname, table_name):
    df_SQL_pd = df.toPandas()
    conn = sqlite3.connect(dbname)
    df_SQL_pd.to_sql(table_name, conn, if_exists='replace', index=False)
    conn.close()
    print(f"Data written to SQLite table '{table_name}' in database '{dbname}'")

#Read Data from SQL
def read_SQLLite(queries, dbname):
    conn = sqlite3.connect(dbname)
    for query in queries:
        print(f"\nExecuting: {query}")
        df = pd.read_sql_query(query, conn)
        print(df)
    conn.close()
```

- Define Read/Write functions for SQLite3
- Run the ETL Pipeline and execute queries

```

Data written to SQLite table 'Banks_Analysis' in database 'Banks_Analysis.db'

Executing: SELECT * FROM Banks_Analysis
   Rank          Bank_Name  Market_Cap_USD_Billion \
0      1  JPMorgan Chase            432.92
1      2  Bank of America            231.52
2      3  Industrial and Commercial Bank of China  194.56
3      4  Agricultural Bank of China            160.68
4      5  HDFC Bank                  157.91
5      6  Wells Fargo                155.87
6      7  HSBC Holdings PLC            148.90
7      8  Morgan Stanley              140.83
8      9  China Construction Bank            139.82
9     10  Bank of China                136.81

Executing: SELECT AVG(Market_Cap_USD_Billion) FROM Banks_Analysis
AVG(Market_Cap_USD_Billion)
0 189.982

Executing: SELECT Bank_Name FROM Banks_Analysis LIMIT 5
   Bank_Name
0  JPMorgan Chase
1  Bank of America
2  Industrial and Commercial Bank of China
3  Agricultural Bank of China
4  HDFC Bank

```

- Load data from S3
- Perform Currency Calculations
- Data was written to SQL and successfully read back in queries
- Display SQL execution results

8.2 Market Capitalization Analysis

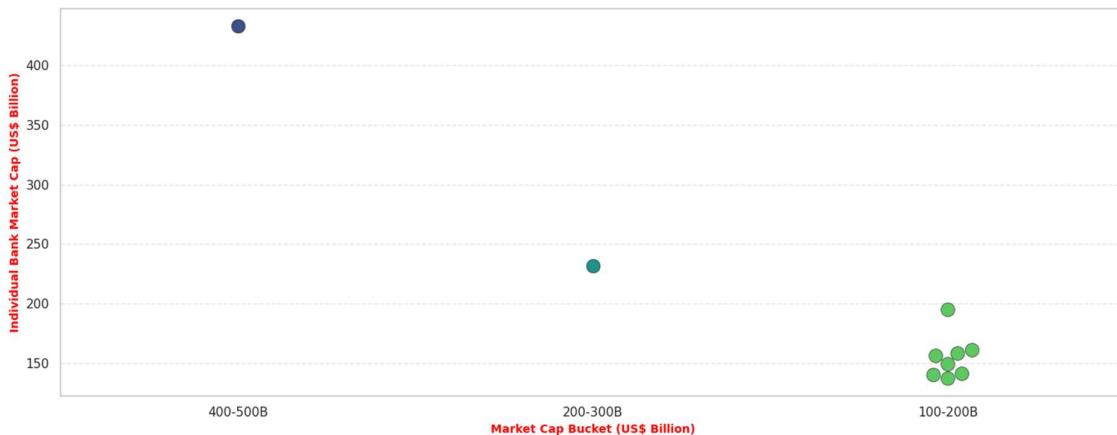
Performs a ranked analysis of banks by market capitalization, calculating each bank's position (Global_rank) and its relative gap from the market leader, preceding bank, and successor.

Rank	Bank_Name	Market_Cap_USD_Billion	Cumulative_Market_Share_USD_Billion	Market_Cap_Range	Market_Share_%	Market_Cap_EUR_Billion	Market_Cap_GBP_Billion	Market_Cap_INR_Billion
1	JPMorgan Chase	432.92	432.92	400-500B	22.79	402.62	346.34	35910.71
2	Bank of America	231.52	664.44	200-300B	12.19	215.31	185.22	19204.46
3	Industrial and Commercial Bank of China	194.56	859.0	100-200B	10.24	180.94	155.65	16138.75
4	Agricultural Bank of China	160.68	1019.68	100-200B	8.46	149.43	128.54	13328.41
5	HDFC Bank	157.91	1177.59	100-200B	8.31	146.86	126.33	13098.63
6	Wells Fargo	155.49	1333.46	100-200B	8.2	144.96	124.7	12929.42
7	HSBC Holdings PLC	148.9	1482.36	100-200B	7.84	138.48	119.12	12211.26
8	Morgan Stanley	140.83	1623.19	100-200B	7.11	130.97	111.66	11681.87
9	China Construction Bank	139.82	1763.01	100-200B	7.36	130.03	111.86	11598.07
10	Bank of China	136.81	1899.82	100-200B	7.2	127.23	109.45	11348.35

Query: Advanced Market Cap Analysis with Growth Metrics

Bank_Name	Market_Cap_USD_Billion	Global_rank	Gap_from_Leader	Gap_from_preceeding	Gap_from_Successor
JPMorgan Chase	432.92	1	0.0	NULL	-201.4
Bank of America	231.52	2	-201.4	-201.4	-36.96
Industrial and Co...	194.56	3	-238.36	-36.96	-33.88
Agricultural Bank...	160.68	4	-272.24	-33.88	-2.77
HDFC Bank	157.91	5	-275.01	-2.77	-2.04
Wells Fargo	155.87	6	-277.05	-2.04	-6.97
HSBC Holdings PLC	148.9	7	-284.02	-6.97	-8.07
Morgan Stanley	140.83	8	-292.09	-8.07	-1.01
China Construction...	139.82	9	-293.1	-1.01	-3.01
Bank of China	136.81	10	-296.11	-3.01	NULL

Bank-Level Market Cap Distribution by Bucket

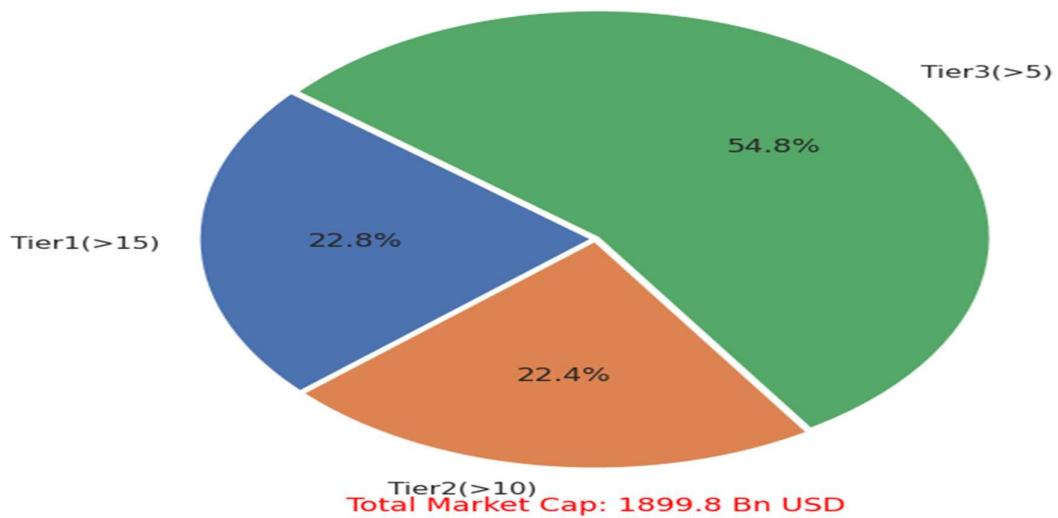
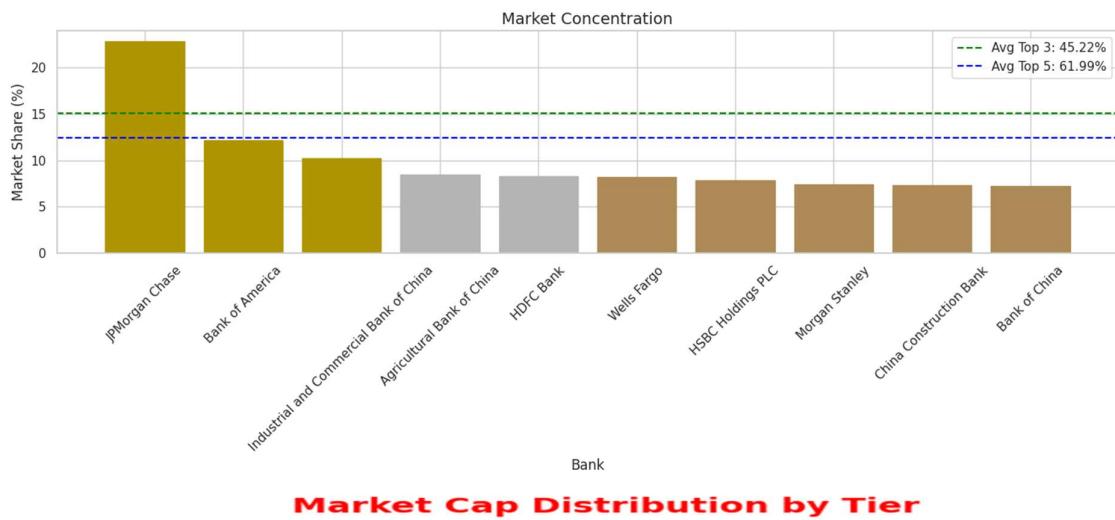


- Swarm plot helped in visually representing individual datapoints
- 3 Key growth metrics, helped in layered comparison of financial positioning for each bank entity
 - Gap from Leader
 - Gap from Preceding
 - Gap from Successor
- JP Morgan Chase leads by a wide margin, with large gaps from immediate peers
- Most bank show tight clustering with neighbours

8.3 Market Concentration Analysis

Market Concentration into tiers (Tier 1, Tier 2, Tier 3, and Others), It provides a clear view of market concentration across performance bands.

Market_Share_Tier	Bank_Count	Total_Market_Share_%	Average_Market_Cap_USD_Billion
Tier1(>15)	1	22.79	432.92
Tier2(>10)	2	22.43	213.04
Tier3(>5)	7	54.78	148.69



- Market Concentration
 - Tier 1 greater than 15%, Market share 22.79 with only 1 Bank
 - Tier 2 greater than 10%, Market share 22.43 with only 2 Bank
 - Tier 3 greater than 5%, Market share 54.78 with only 7 Bank
- The chart reveals a **sharp concentration at the top**, with **JPMorgan Chase** and **Bank of America** exceeding the 10% market share threshold
- Bar Chart Shows Tier into Gold , Silver and Bronze

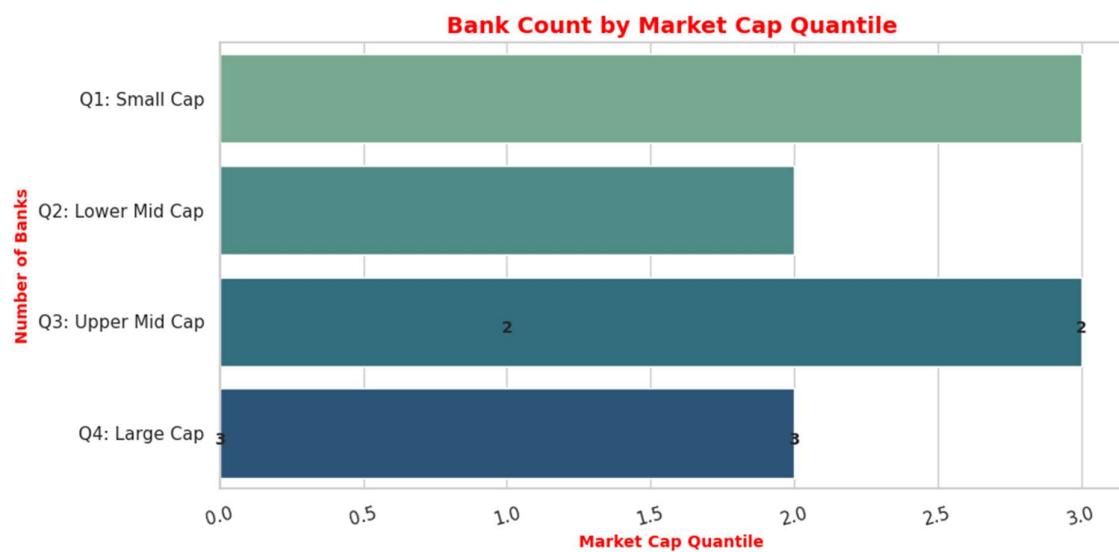
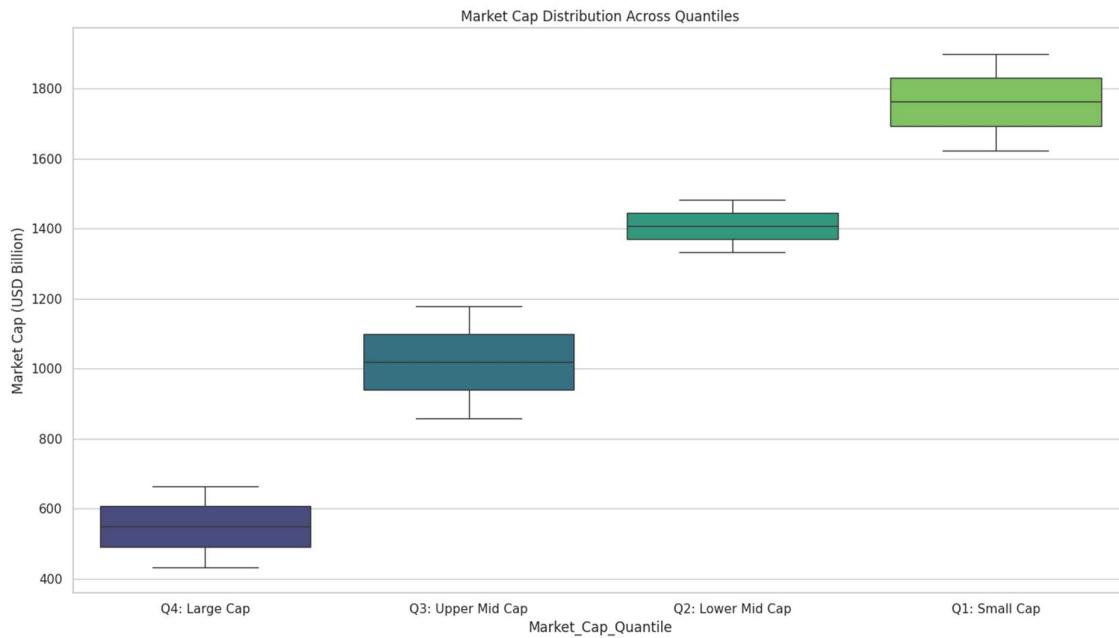
8.4 Market Capitalization Distribution

Perform quartile-based market segmentation of banks into Small Cap , Medium and Large caps per quartile buckets, enabling tiered insights into market structure and distribution.

percentile	banks_in_quartile	min_cap	max_cap	avg_cap	total_cap
1	3	136.81	140.83	139.15	417.46
2	3	148.9	157.91	154.23	462.68
3	2	160.68	194.56	177.62	355.24
4	2	231.52	432.92	332.22	664.44

Quantiles: [140.83, 155.87, 194.56]

Bank_Name	Market_Cap_USD_Billion	Market_Cap_Quantile
JPMorgan Chase	432.92	Q4: Large Cap
Bank of America	231.52	Q4: Large Cap
Industrial and Co...	194.56	Q3: Upper Mid Cap
Agricultural Bank...	160.68	Q3: Upper Mid Cap
HDFC Bank	157.91	Q3: Upper Mid Cap
Wells Fargo	155.87	Q2: Lower Mid Cap
HSBC Holdings PLC	148.9	Q2: Lower Mid Cap
Morgan Stanley	140.83	Q1: Small Cap
China Construction...	139.82	Q1: Small Cap
Bank of China	136.81	Q1: Small Cap

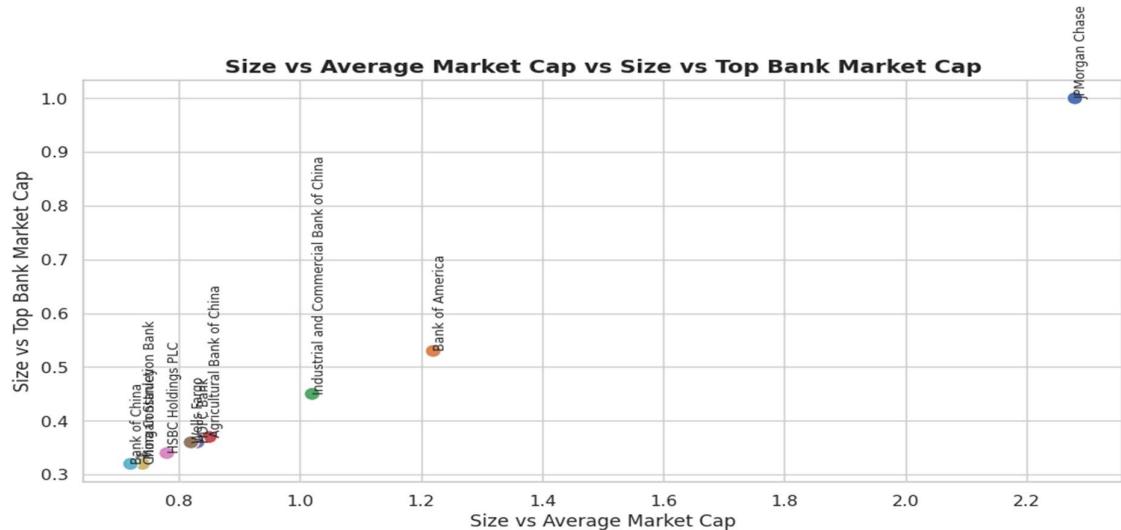


- Clear signs of bottom-heavy distribution
- Q4 (Large Cap) includes only two banks—JPMorgan Chase and Bank of America—with significantly higher market caps.
- Q1 (Small Cap) contains four banks clustered tightly below the \$140.83B threshold, indicating a dense lower tier.
- The mid-cap tiers (Q2 and Q3) show gradual transitions, but the overall structure confirms a top-heavy distribution with sharp financial asymmetry.
- The horizontal bar chart shows a balanced distribution of banks across quantiles, with Q1 (Small Cap) and Q4 (Large Cap) each containing 3 banks, while Q2 and Q3 hold 2 banks each.

8.5 Comparative Size Analysis

market capitalization against the **average** and **maximum** values, calculating relative size ratios (Size_vs_Avg, Size_vs_Top) and classifying banks into **small, medium, or large** categories.

Bank Name	Market Cap USD Billion	Size_vs_Avg	Size_vs_Top	Market Size Category
JPMorgan Chase	432.92	2.28	1.0	Large (>= 1.5x Average)
Bank of America	231.52	1.22	0.53	Medium (0.8x - 1.5x Average)
Industrial and Commercial Bank of China	194.56	1.02	0.45	Medium (0.8x - 1.5x Average)
Agricultural Bank of China	160.68	0.85	0.37	Medium (0.8x - 1.5x Average)
HDFC Bank	157.91	0.83	0.36	Medium (0.8x - 1.5x Average)
Wells Fargo	155.87	0.82	0.36	Medium (0.8x - 1.5x Average)
HSBC Holdings PLC	148.9	0.78	0.34	Small (< 0.8x Average)
Morgan Stanley	140.83	0.74	0.33	Small (< 0.8x Average)
China Construction Bank	139.82	0.74	0.32	Small (< 0.8x Average)
Bank of China	136.81	0.72	0.32	Small (< 0.8x Average)

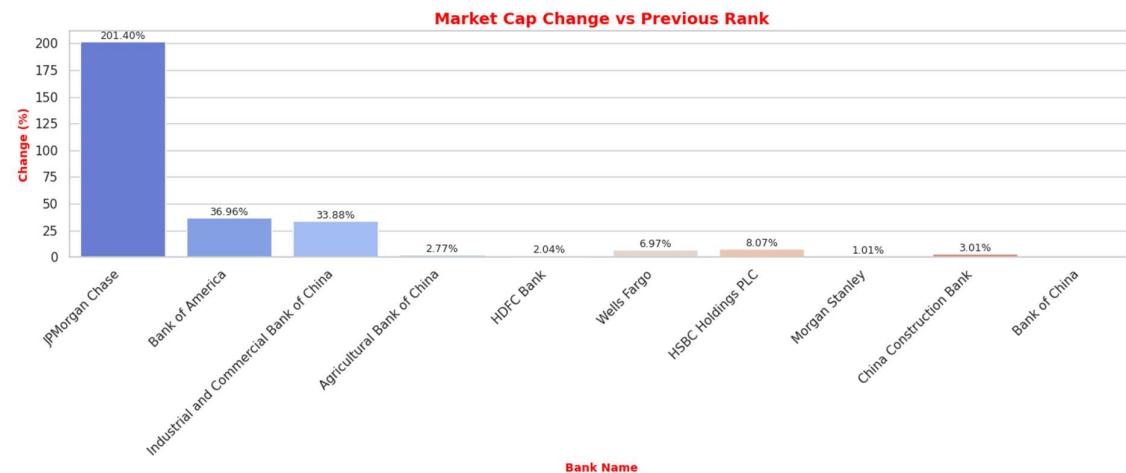


- JPMorgan Chase is the only bank classified as “Large”, with a market cap more than $1.5 \times$ the average, while most others fall into the “Medium” category
- Presence of sharp size disparity, where one dominant player significantly skews the average
- The scatter plot clearly illustrates JPMorgan Chase’s dominance, positioned farthest from the origin with the highest ratios in both dimensions—Size vs Average and Size vs Top Bank.
- All other banks cluster in the lower-left quadrant, indicating limited deviation from the average and significant distance from the top, reinforcing a highly skewed market structure with one outsized leader and a compressed competitive field.

8.5 Market Growth Analysis

Calculate Absolute and percentage gap in market capitalization between each bank and its immediate successor in rank. Using Spark window functions

Market Growth and Gap Analysis Results:				
	Bank_Name	Market_Cap_USD_Billion	Next_Market_Cap	Gap_vs_Next
				gap_percentage
	JPMorgan Chase	432.92	231.52	201.4
	Bank of America	231.52	194.56	36.96
	Industrial and Co...	194.56	160.68	33.88
	Agricultural Bank...	160.68	157.91	2.77
	HDFC Bank	157.91	155.87	2.04
	Wells Fargo	155.87	148.9	6.97
	HSBC Holdings PLC	148.9	140.83	8.07
	Morgan Stanley	140.83	139.82	1.01
	China Constructio...	139.82	136.81	3.01
	Bank of China	136.81	NULL	NULL



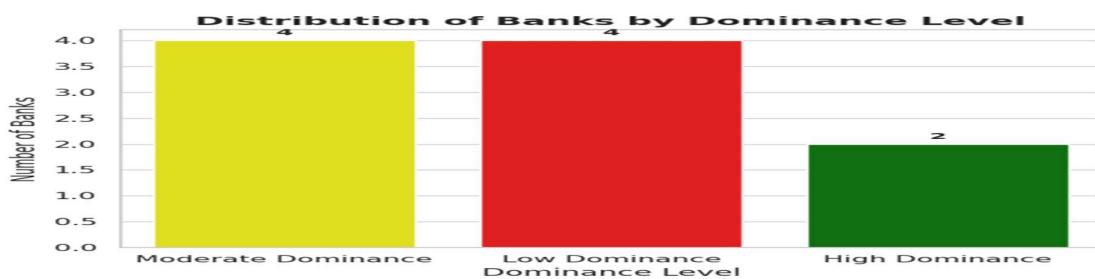


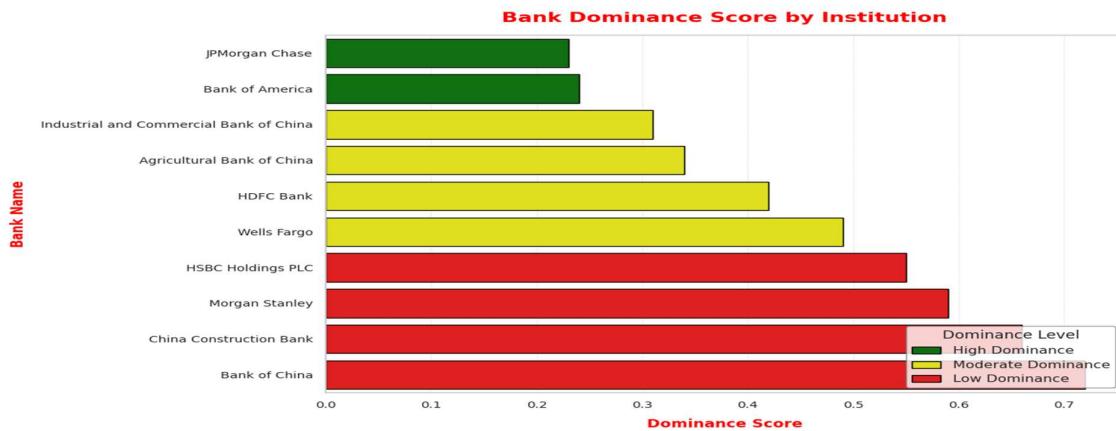
- Massive drop-off of \$201.4B (46.52%) between JPMorgan Chase and Bank of America
- Subsequent gaps are notably smaller, with most banks showing single-digit percentage declines, indicating a steep initial fall followed by a more gradual tapering across the leaderboard

8.6 Market Dominance Analysis

Dominance Score for each bank using its market cap and rank relative to the total market, then classifies banks into High, Moderate, or Low Dominance tiers

rank	Bank_Name	Market_Cap_USD_Billion	Dominance_Score	Dominance_Level
1	JPMorgan Chase	432.92	0.23	High Dominance
2	Bank of America	231.52	0.24	High Dominance
3	Industrial and Co...	194.56	0.31	Moderate Dominance
4	Agricultural Bank...	160.68	0.34	Moderate Dominance
5	HDFC Bank	157.91	0.42	Moderate Dominance
6	Wells Fargo	155.87	0.49	Moderate Dominance
7	HSBC Holdings PLC	148.9	0.55	Low Dominance
8	Morgan Stanley	140.83	0.59	Low Dominance
9	China Construction...	139.82	0.66	Low Dominance
10	Bank of China	136.81	0.72	Low Dominance





- **High Dominance:** Only **2 banks**, including JPMorgan Chase, exert outsized influence.
- **Moderate Dominance:** **4 banks** occupy the strategic middle ground—likely balancing scale with agility.
- **Low Dominance:** Another **4 banks** fall into this tier, suggesting limited market leverage despite respectable capitalization.

8.7 Segment-Wise Bank Analysis

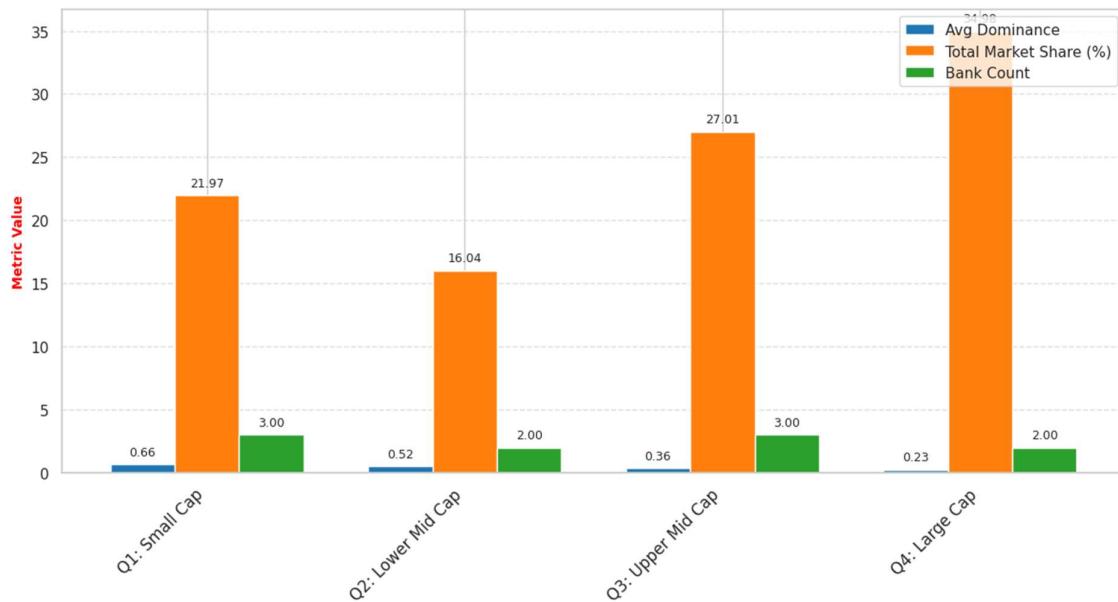
Market_Size_Category	Number_of_Banks	Average_Market_Cap_USD_Billion
Large (>= 1.5x Average)	1	432.92
Medium (0.8x - 1.5x)	5	180.11
Small (< 0.8x Average)	4	141.59

Market_Cap_Quantile	Avg_Dominance	Total_Market_Share	Bank_Count
Q1: Small Cap	0.65	21.97	3
Q2: Lower Mid Cap	0.52	16.04	2
Q3: Upper Mid Cap	0.35	27.01	3
Q4: Large Cap	0.235	34.98	2

- **Q4 banks dominate the market share**, but their **low dominance scores** reflect concentrated power in fewer hands.
- **Q1 banks**, despite smaller caps, show **higher strategic dominance**, possibly due to niche positioning or regional agility.
- **Q2 and Q3** offer a balanced mix—ideal for **partnerships, scaling strategies, or regulatory focus**.

8.8 Performance Dashboard

Market Cap Quantile Performance Analysis



- Q1 : Small Cap
 - Average Dominance of 0.66 (Highest)
 - Market Share of 21.97%
 - Count of Banks 2
- Q2 : Lower Mid Cap
 - Average Dominance of 0.60
 - Market Share of 16.04%
 - Count of Banks 2
- Q3 : Upper Mid Cap
 - Average Dominance of 0.66 (Highest)
 - Market Share of 21.97%
 - Count of Banks 3
- Q4 : Large Cap
 - Average Dominance of 0.63
 - Market Share of 34.98% (Highest)
 - Count of Banks 3

Q1 Large Cap – Dominates market share but doesn't have dominance score – suggesting concentration by top ranked banks like JP Morgan and Bank of America

9. Logging

```
code_log.txt x ...
1 2025-Sep-15-07:13:02 : ETL process pipeline started
2 2025-Sep-15-07:13:02 : ETL process pipeline completed
3 2025-Sep-15-07:13:04 : Starting : Advanced Market Cap Analysis with Growth Metrics
4 2025-Sep-15-07:13:07 : Completed : Advanced Market Cap Analysis with Growth Metrics
5 2025-Sep-15-07:13:07 : Starting : Market Concentration Analysis
6 2025-Sep-15-07:13:08 : Completed : Market Concentration Analysis
7 2025-Sep-15-07:13:09 : Starting : Statistical Distribution Analysis
8 2025-Sep-15-07:13:20 : Completed : Statistical Distribution Analysis
9 2025-Sep-15-07:13:21 : Starting Comparative Size Analysis
10 2025-Sep-15-07:13:23 : Starting Market Growth Analysis and Gap Identification
11 2025-Sep-15-07:34:48 : ETL process pipeline started
12 2025-Sep-15-07:34:48 : ETL process pipeline completed
13 2025-Sep-15-07:34:59 : Starting : Advanced Market Cap Analysis with Growth Metrics
14 2025-Sep-15-07:35:00 : Completed : Advanced Market Cap Analysis with Growth Metrics
15 2025-Sep-15-07:35:15 : Starting : Market Concentration Analysis
16 2025-Sep-15-07:35:16 : Completed : Market Concentration Analysis
17 2025-Sep-15-07:36:02 : Starting : Statistical Distribution Analysis
18 2025-Sep-15-07:36:06 : Completed : Statistical Distribution Analysis
19 2025-Sep-15-07:37:21 : Starting Comparative Size Analysis
20 2025-Sep-15-07:38:15 : Starting Market Growth Analysis and Gap Identification
21 2025-Sep-15-07:42:06 : Market Dominance Analysis
22 2025-Sep-15-07:43:17 : Market Dominance Analysis
23 2025-Sep-15-07:43:43 : Market Dominance Analysis
24 2025-Sep-15-07:45:51 : Market Dominance Analysis
25 2025-Sep-15-07:46:16 : Market Dominance Analysis
26 2025-Sep-15-07:48:21 : Starting Segment Performance Analysis
27 2025-Sep-15-07:51:48 : ETL process pipeline started
28 2025-Sep-15-07:51:48 : ETL process pipeline completed
29 2025-Sep-15-07:51:49 : Starting : Advanced Market Cap Analysis with Growth Metrics
30 2025-Sep-15-07:51:50 : Completed : Advanced Market Cap Analysis with Growth Metrics
31 2025-Sep-15-07:51:50 : Starting : Market Concentration Analysis
32 2025-Sep-15-07:51:51 : Completed : Market Concentration Analysis
33 2025-Sep-15-07:51:52 : Starting : Statistical Distribution Analysis
34 2025-Sep-15-07:51:59 : Completed : Statistical Distribution Analysis
35 2025-Sep-15-07:52:00 : Starting Comparative Size Analysis
36 2025-Sep-15-07:52:03 : Starting Market Growth Analysis and Gap Identification
37 2025-Sep-15-07:52:05 : Market Dominance Analysis
38 2025-Sep-15-07:54:34 : Market Dominance Analysis
39 2025-Sep-15-07:55:18 : Market Dominance Analysis
40 2025-Sep-15-07:55:38 : Starting Segment Performance Analysis
41 2025-Sep-15-07:55:51 : Starting Segment Performance Analysis
42 2025-Sep-15-08:11:20 : ETL process pipeline started
43 2025-Sep-15-08:11:20 : ETL process pipeline completed
44 2025-Sep-15-08:11:36 : Starting : Advanced Market Cap Analysis with Growth Metrics
45 2025-Sep-15-08:11:39 : Completed : Advanced Market Cap Analysis with Growth Metrics
46 2025-Sep-15-08:12:14 : Starting : Advanced Market Cap Analysis with Growth Metrics
47 2025-Sep-15-08:12:15 : Completed : Advanced Market Cap Analysis with Growth Metrics
48 2025-Sep-15-08:12:35 : Starting : Market Concentration Analysis
49 2025-Sep-15-08:12:37 : Completed : Market Concentration Analysis
50 2025-Sep-15-08:13:15 : Starting : Market Concentration Analysis
51 2025-Sep-15-08:13:19 : Completed : Market Concentration Analysis
52 2025-Sep-15-08:15:12 : Starting : Market Concentration Analysis
53 2025-Sep-15-08:17:57 : Starting : Market Concentration Analysis
54 2025-Sep-15-08:18:16 : Starting : Market Concentration Analysis
55 2025-Sep-15-08:19:45 : ETL process pipeline started
```

10. Report Conclusion: Strategic Insights from Global Bank Market Leader

1. This comprehensive analysis of the world's top 10 banks by market capitalization reveals a highly skewed and concentrated market structure, dominated by a few financial giants.
2. JPMorgan Chase emerges as the unequivocal leader, with a market cap exceeding \$432B, almost double of its nearest neighbour
3. JP Morgan also distorted financial statistic by being a heavy outlier
4. Market Concentration: Tier 1 – JP Morgan bank alone holds 22.8% of total market share, Tier 2 (Bank of America & ICBC) together contributes to 22.4%, remaining 7 banks combined over half of market share.
5. Dominance Disparity: Despite their size, Tier 1 banks show lower dominance scores, suggesting that market share does not always translate to strategic influence.
6. Size Categorization: Only JPMorgan qualifies as “Large” ($>1.5 \times$ average market cap), while most banks fall into the “Medium” category. This reinforces the presence of a single outsized player and a tightly clustered mid-tier.
7. Growth Gaps: The absolute and percentage drop between JPMorgan and Bank of America is staggering—\$201.4B or 46.5%. Subsequent gaps taper off, forming a steep-to-flat curve across the leaderboard.