

1. Data Preparation

Data was extracted through `read_html()` from the Wikipedia page and the correct table was loaded .

```
+-----+-----+
|Rank|      Bank_name|Market_Cap_USDBn|
+-----+-----+
| 1|JPMorgan Chase| 432.92|
| 2|Bank of America| 231.52|
| 3|Industrial and Co...| 194.56|
| 4|Agricultural Bank...| 160.68|
| 5|HDFC Bank| 157.91|
+-----+
only showing top 5 rows

root
|-- Rank: long (nullable = true)
|-- Bank_name: string (nullable = true)
|-- Market_Cap_USDBn: double (nullable = true)
```

2. Data Cleaning

2.1 Handle Missing Values

No missing values were found in the dataset

2.2 Fixing Columns

No duplicates or wrong column type found

2.2 Handle Outliers

1 Outlier is present in the dataset and was left to stay since the dataset is too small and the value is valid .

```
Outliers in Market_Cap_USDBn column:
+-----+
|Rank|      Bank_name|Market_Cap_USDBn|
+-----+
| 1|JPMorgan Chase| 432.92|
+-----+
```

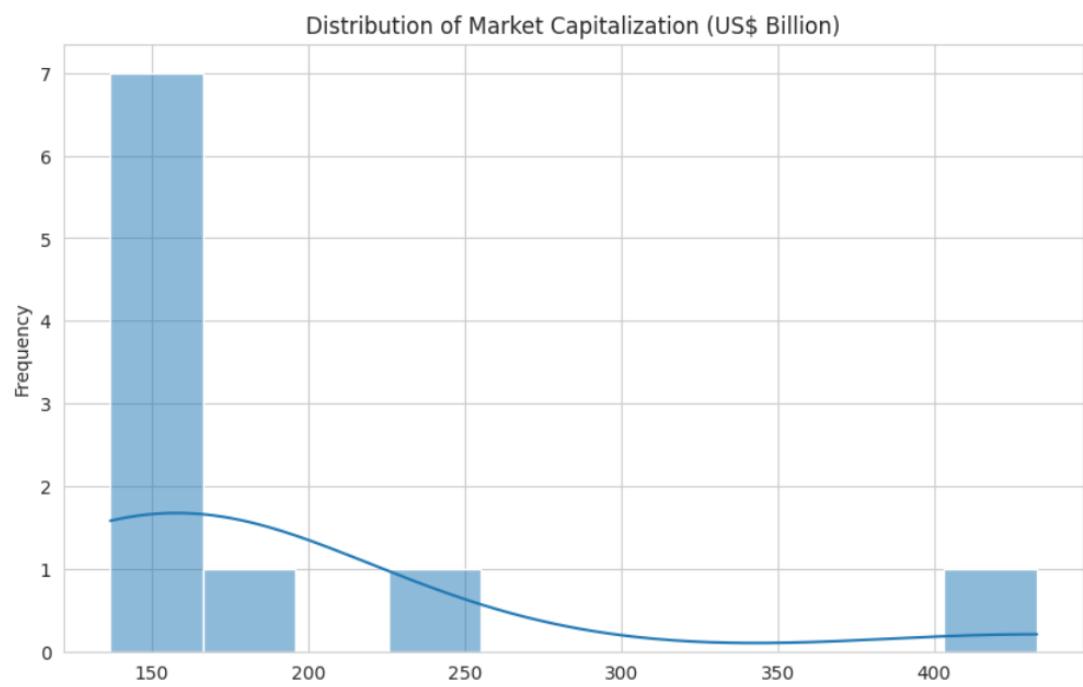
3. Exploratory Data Analysis

3.1. Conversion from PySpark to Pandas DataFrame

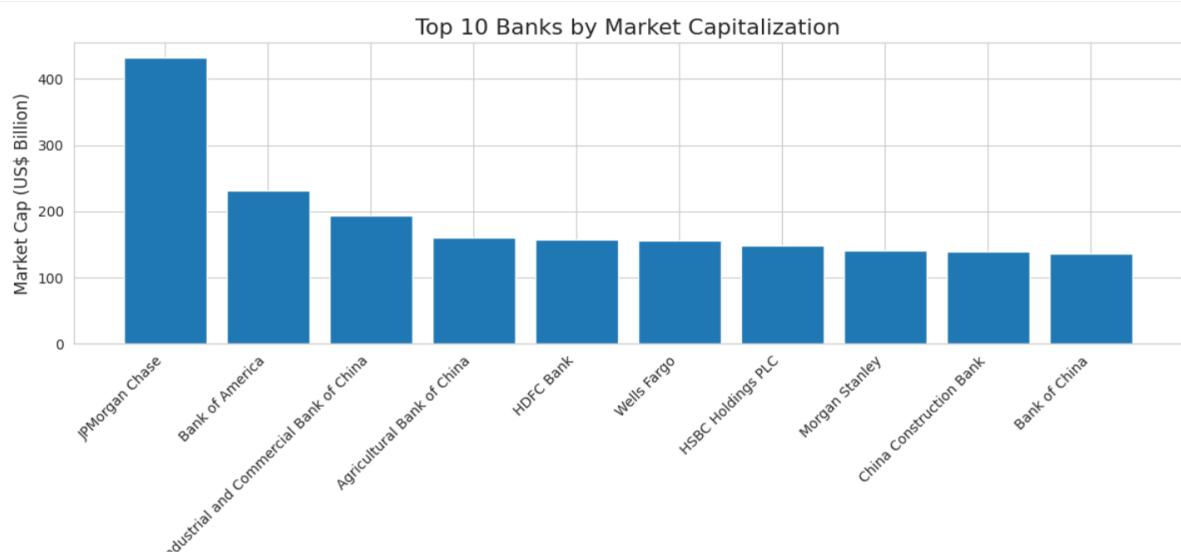
Conversion was done using `toPandas()` method

3.2. Market Capitalization

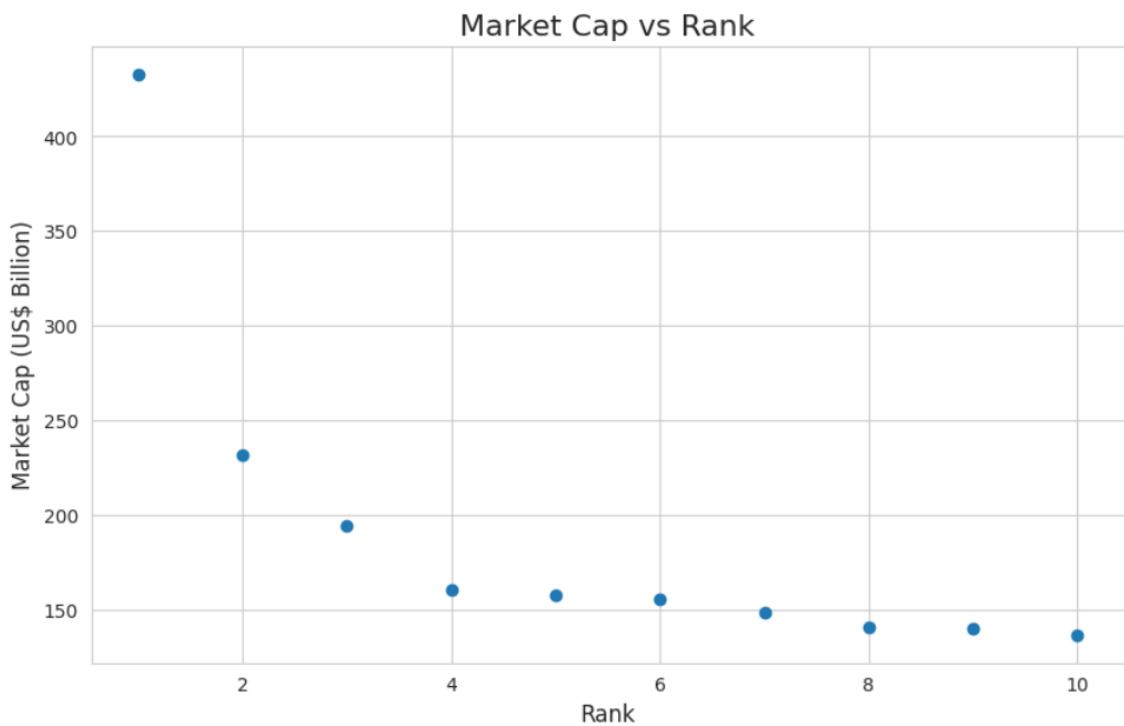
Most of the banks lied between 100 to 200 US Billion dollar market cap range



3.3. Top 10 Banks

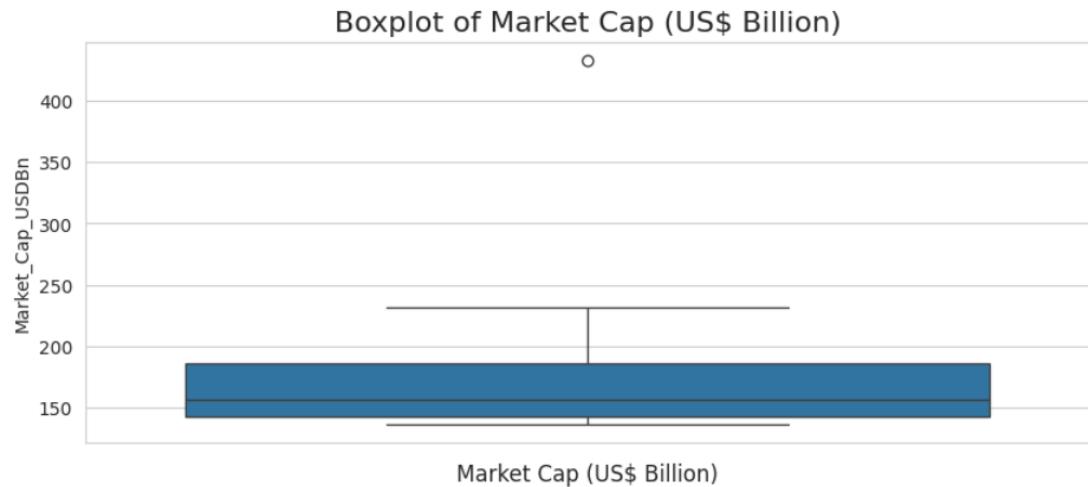


3.4. Market Cap vs Bank Ranking



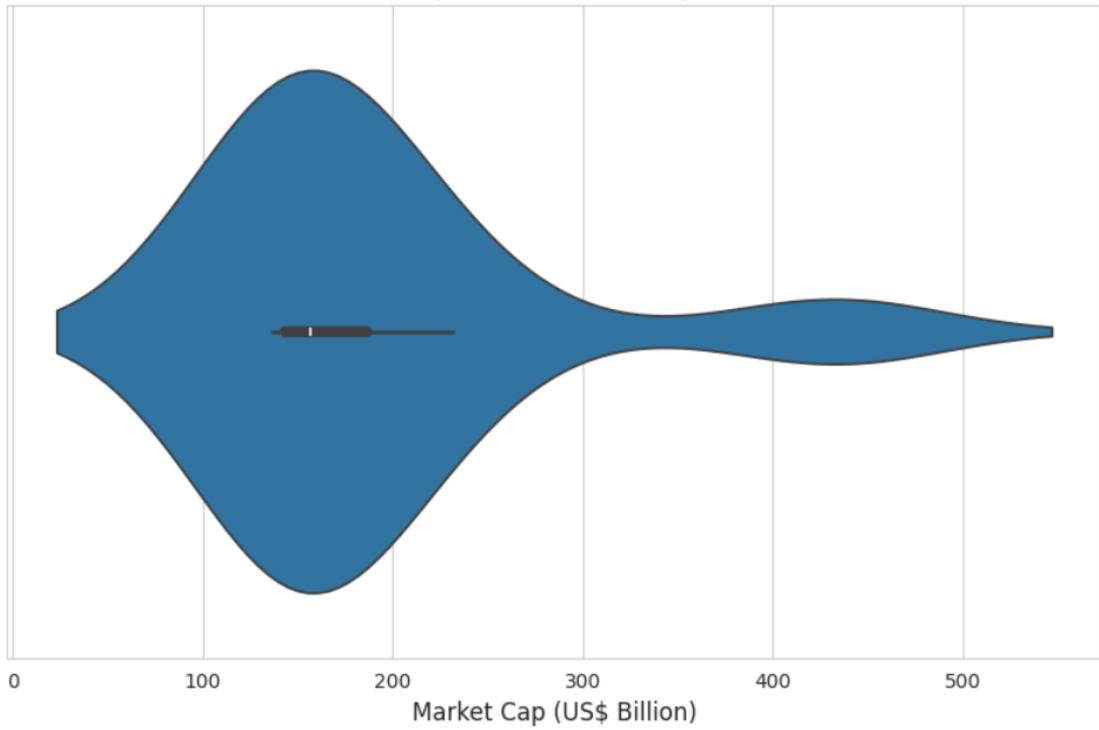
Market cap and Rank have strong negative correlation .

3.5. Market Cap Analysis



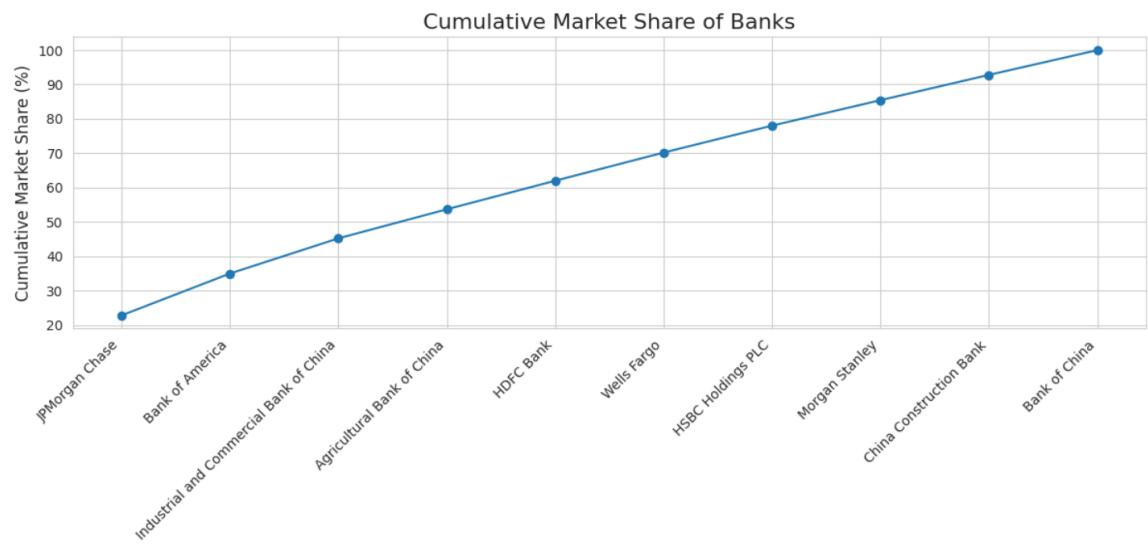
3.6. Market Cap Quartile Distribution

Market Cap Distribution by Quartile



Violin plot used for Market cap distribution by quartile

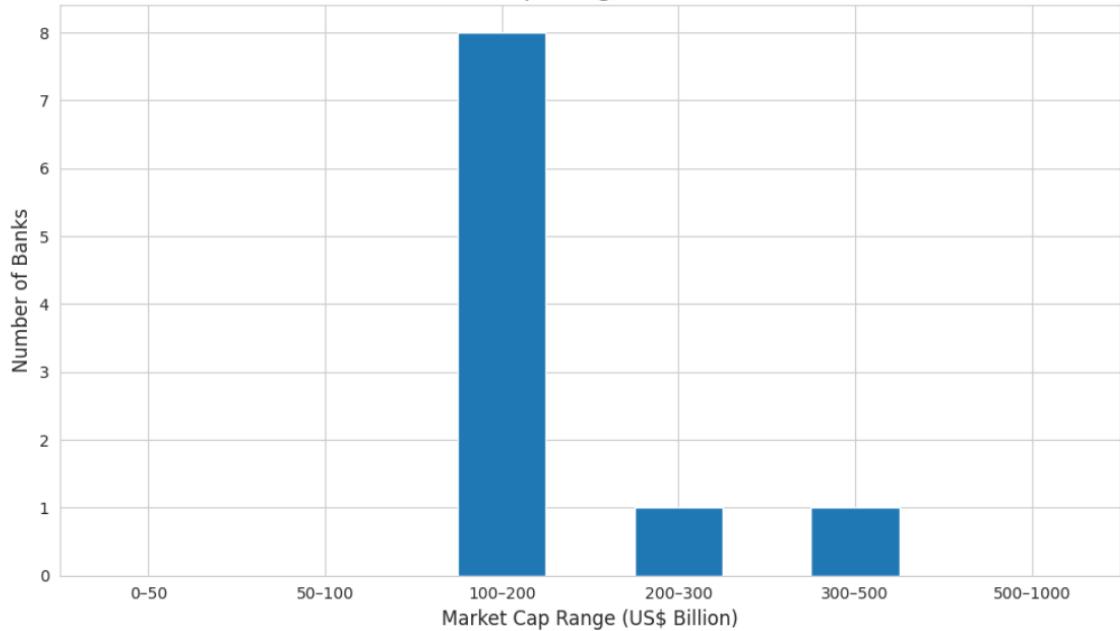
3.7. Cumulative Market Share Analysis



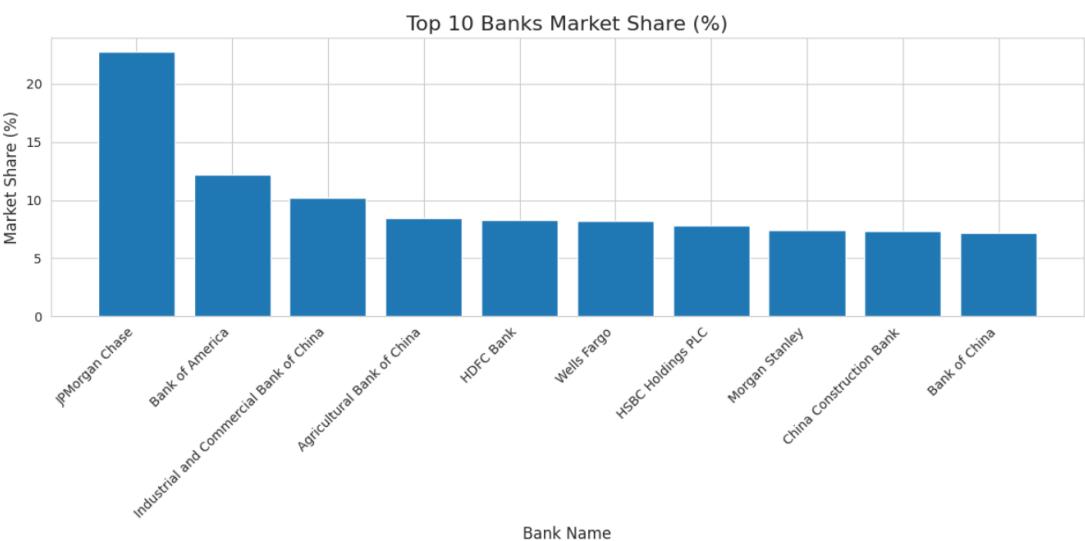
3.8. Categorising Banks

...

Market Cap Range Distribution



3.9. Visualise Market Share Distribution



4. ETL and Querying

4.1. Market Capitalization Analysis

Rank	Bank_Name	MarketCap	Growth_vs_Median	Growth_vs_TopBank	Growth_Rank	Gap_From_Previous
1	JPMorgan Chase	432.92	276.03000000000003	1.0	1	NULL
2	Bank of America	231.52	74.63000000000002	0.5347870276263513	2	201.4
3	Industrial and Co...	194.56	37.670000000000016	0.4494132865194493	3	36.96000000000001
4	Agricultural Bank...	160.68	3.7900000000000205	0.3711540238381225	4	33.87999999999995
5	HDFC Bank	157.91	1.0200000000000102	0.3647556130462903	5	2.7700000000000102
6	Wells Fargo	155.87	-1.019999999999981	0.3600434260371431	6	2.039999999999992
7	HSBC Holdings PLC	148.9	-7.989999999999981	0.34394345375589025	7	6.96999999999999
8	Morgan Stanley	140.83	-16.059999999999974	0.32530259632264624	8	8.069999999999993
9	China Construction...	139.82	-17.069999999999993	0.3229696017739998	9	1.0100000000000193
10	Bank of China	136.81	-20.079999999999984	0.316016816039915	10	3.009999999999991

4.2. Market Concentration Analysis

Rank	Bank_name	MarketCap	MarketSharePerc	MarketShareTier
1	JPMorgan Chase	432.92	22.79	Tier 1: Dominant
2	Bank of America	231.52	12.19	Tier 1: Dominant
3	Industrial and Co...	194.56	10.24	Tier 1: Dominant
4	Agricultural Bank...	160.68	8.46	Tier 2: Strong
5	HDFC Bank	157.91	8.31	Tier 2: Strong
6	Wells Fargo	155.87	8.2	Tier 2: Strong
7	HSBC Holdings PLC	148.9	7.84	Tier 2: Strong
8	Morgan Stanley	140.83	7.41	Tier 2: Strong
9	China Construction...	139.82	7.36	Tier 2: Strong
10	Bank of China	136.81	7.2	Tier 2: Strong

4.3. Market Capitalization Distribution

Rank	Bank_name	MarketCap	Quartile	Quartile_Label
1	JPMorgan Chase	432.92	1	Top 25% MarketCap Banks
2	Bank of America	231.52	1	Top 25% MarketCap Banks
3	Industrial and Commercial Bank of China	194.56	1	Top 25% MarketCap Banks
4	Agricultural Bank of China	160.68	2	Upper-Mid 25% Banks
5	HDFC Bank	157.91	2	Upper-Mid 25% Banks
6	Wells Fargo	155.87	2	Upper-Mid 25% Banks
7	HSBC Holdings PLC	148.9	3	Lower-Mid 25% Banks
8	Morgan Stanley	140.83	3	Lower-Mid 25% Banks
9	China Construction Bank	139.82	4	Bottom 25% Banks
10	Bank of China	136.81	4	Bottom 25% Banks

4.4. Comparative Size Analysis

Rank	Bank_name	MarketCap	MarketSharePct	Size_Category
1	JPMorgan Chase	432.92	22.79	Mega Bank
2	Bank of America	231.52	12.19	Mega Bank
3	Industrial and Commercial Bank of China	194.56	10.24	Mega Bank
4	Agricultural Bank of China	160.68	8.46	Large Bank
5	HDFC Bank	157.91	8.31	Large Bank
6	Wells Fargo	155.87	8.2	Large Bank
7	HSBC Holdings PLC	148.9	7.84	Large Bank
8	Morgan Stanley	140.83	7.41	Large Bank
9	China Construction Bank	139.82	7.36	Large Bank
10	Bank of China	136.81	7.2	Large Bank

4.5. Market Growth Analysis

Rank	Bank_name	MarketCap	Next_Bank	Next_Bank_MarketCap	MarketCap_Gap
1	JPMorgan Chase	432.92	Bank of America	231.52	201.4
2	Bank of America	231.52	Industrial and Commercial Bank of China	194.56	36.96
3	Industrial and Commercial Bank of China	194.56	Agricultural Bank of China	160.68	33.88
4	Agricultural Bank of China	160.68	HDFC Bank	157.91	2.77
5	HDFC Bank	157.91	Wells Fargo	155.87	2.04
6	Wells Fargo	155.87	HSBC Holdings PLC	148.9	6.97
7	HSBC Holdings PLC	148.9	Morgan Stanley	140.83	8.07
8	Morgan Stanley	140.83	China Construction Bank	139.82	1.01
9	China Construction Bank	139.82	Bank of China	136.81	3.01
10	Bank of China	136.81	NULL	NULL	NULL

4.6. Market Dominance Analysis

Rank	Bank_name	MarketCap	MarketSharePct	CumulativeSharePct	DominanceScore
1	JPMorgan Chase	432.92	22.79	22.79	Dominant Leader
2	Bank of America	231.52	12.19	34.98	Strong Player
3	Industrial and Commercial Bank of China	194.56	10.24	45.22	Strong Player
4	Agricultural Bank of China	160.68	8.46	53.68	Moderate Player
5	HDFC Bank	157.91	8.31	61.99	Moderate Player
6	Wells Fargo	155.87	8.2	70.19	Moderate Player
7	HSBC Holdings PLC	148.9	7.84	78.03	Moderate Player
8	Morgan Stanley	140.83	7.41	85.44	Moderate Player
9	China Construction Bank	139.82	7.36	92.8	Moderate Player
10	Bank of China	136.81	7.2	100.0	Moderate Player

4.7. Segment-Wise Bank Analysis

MarketCap_Segment	BankCount	TotalSegmentMarketCap	AvgMarketCap	SegmentSharePct
Large Cap	5	900.54	180.11	47.4
Mid Cap	4	566.36	141.59	29.81
Mega Cap	1	432.92	432.92	22.79

4.8. Performance Dashboard

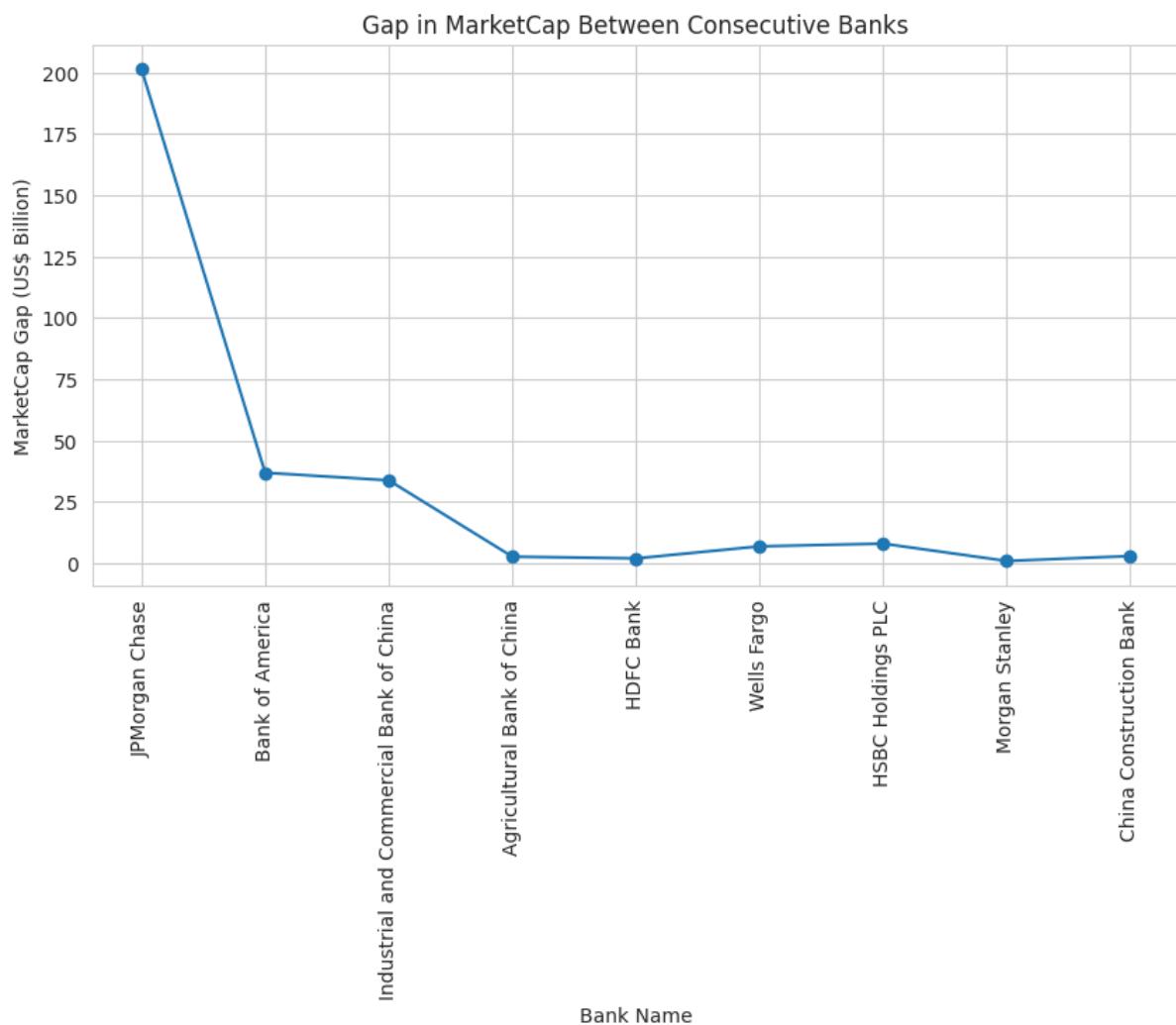
== Market Capitalization Summary ==

Rank	Bank_name	MarketCap	Next_Bank	Next_Bank_MarketCap	MarketCap_Gap
1	JPMorgan Chase	432.92	Bank of America	231.52	201.4
2	Bank of America	231.52	Industrial and Commercial Bank of China	194.56	36.96
3	Industrial and Commercial Bank of China	194.56	Agricultural Bank of China	160.68	33.88
4	Agricultural Bank of China	160.68	HDFC Bank	157.91	2.77

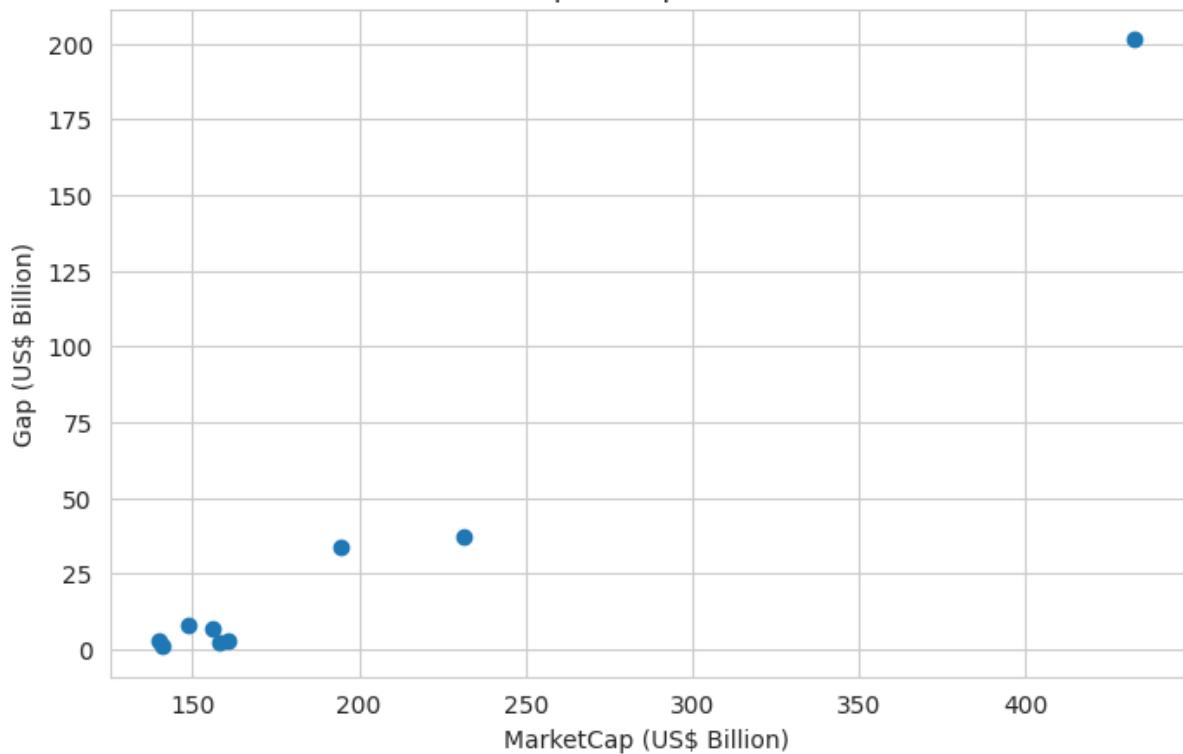
5	HDFC Bank	157.91	Wells Fargo	155.87	2.04	
6	Wells Fargo	155.87	HSBC Holdings PLC	148.9	6.97	
7	HSBC Holdings PLC	148.9	Morgan Stanley	140.83	8.07	
8	Morgan Stanley	140.83	China Construction Bank	139.82	1.01	
9	China Construction Bank	139.82	Bank of China	136.81	3.01	
10	Bank of China	136.81	NULL	NULL	NULL	

+-----+-----+-----+-----+

+



MarketCap vs Gap to Next Bank



MarketCap Gap Between Consecutive Banks

