

Social Media Performance Analysis

Streaming processing using fake stream
data, Kafka, Spark, Postgresql, and Grafana

Farah Shadra

Introduction

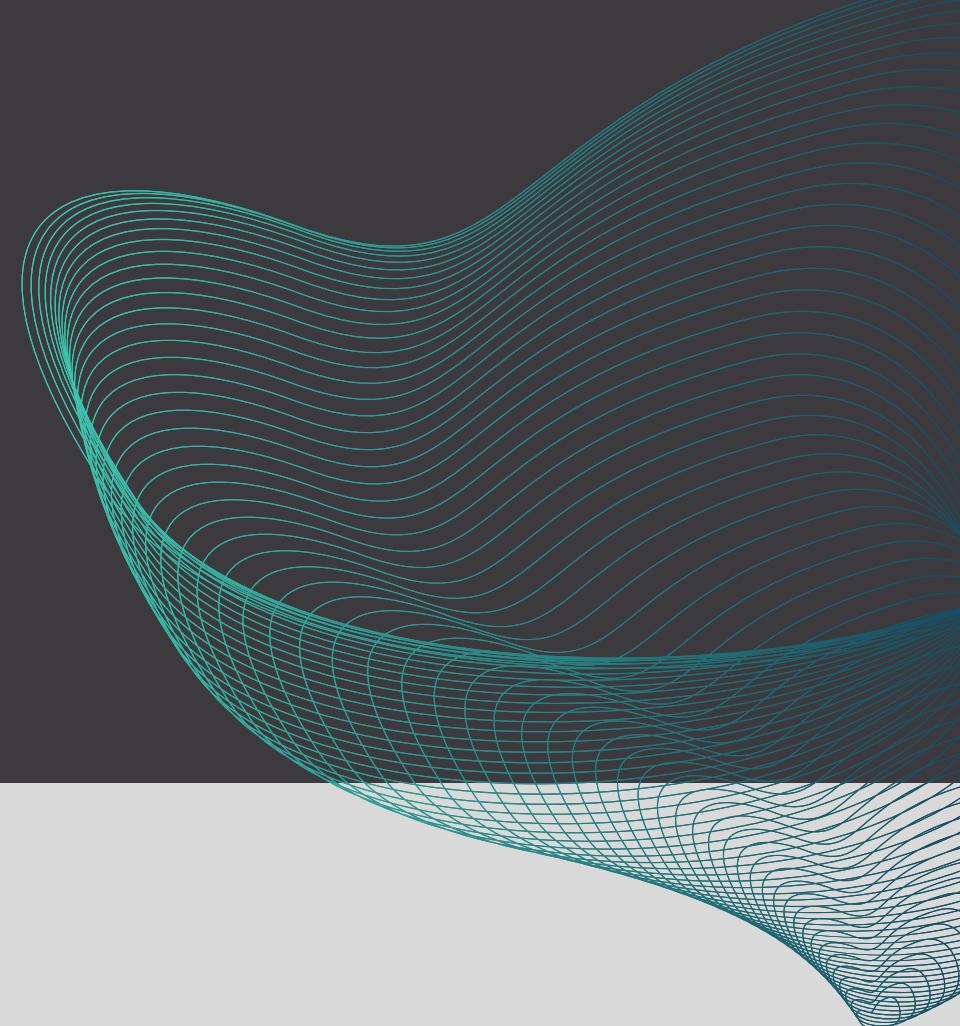
Why we should do this?

The utilization of fake data allows for **comprehensive testing and development of data pipelines** without the need for real, potentially sensitive information. The approach facilitates the measurement and analysis of performance at each stage, aiding in a better **understanding of how each component interacts.**

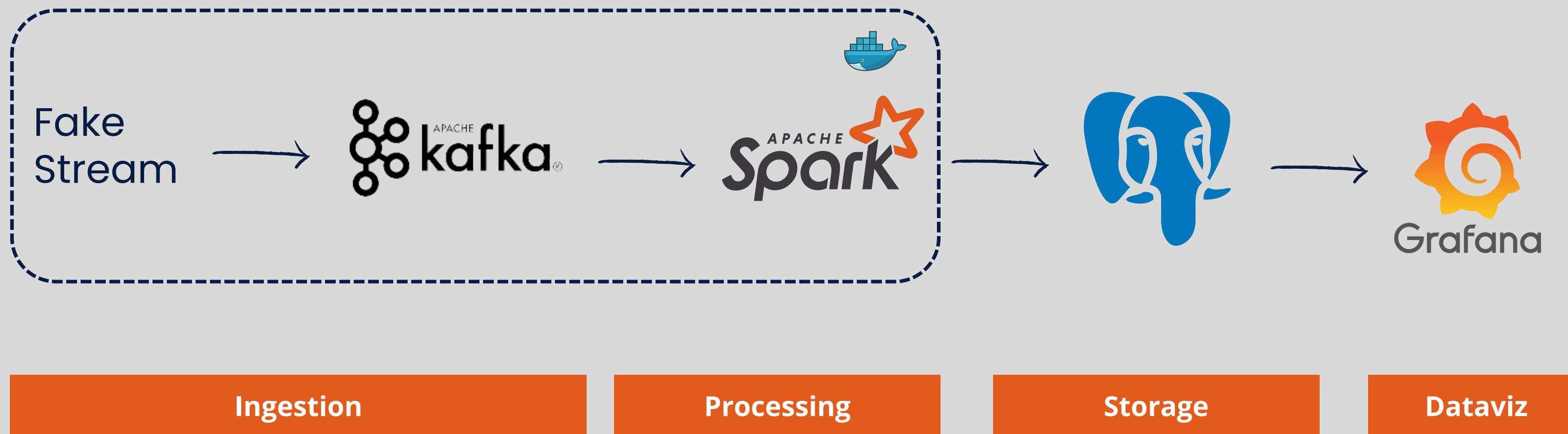
From metrics to insight...

Processing social media data offers several **advantages and insights for businesses:**

- Understanding User Behavior
- Sales Performance
- Conversion Rate and Bounce Rate
- Regional Analysis



Stack



Ingestion

Fake Stream

ID	Kota	Sosial_Media	Usage_count	Sales	Visitors	PageVisitors	Conversion	Ts
1dd12	Jakarta	X	800	45	1000	250	50	1705323357
1dd13	Tangerang	Instagram	200	123	500	300	60	1705323357
1dd13	Bogor	Facebook	300	5	250	120	20	1705323357

Event Producer



```
b'{"ID": "3d131cdc-0611-46c1-9ff7-9861a9c6c75f", "Kota": "Bogor", "Sosial_Media": "Facebook", "Usage_Count": 101,
"Sales": 124, "Visitors": 586, "PageVisitors": 178, "Conversion": 446, "Ts": 1706980463}!'
=====
RecordMetadata(topic='test-topic', partition=0, topic_partition=TopicPartition(topic='test-topic', partition=0),
offset=12192, timestamp=1706980585998, log_start_offset=2604, checksum=None, serialized_key_size=-1,
serialized_value_size=200, serialized_header_size=-1)
=====
b'{"ID": "4bc50707-5676-4dbf-9d5b-e21c3699d860", "Kota": "Depok", "Sosial_Media": "Instagram", "Usage_Count": 986,
"Sales": 141, "Visitors": 570, "PageVisitors": 79, "Conversion": 132, "Ts": 1706980353}!'
=====
RecordMetadata(topic='test-topic', partition=0, topic_partition=TopicPartition(topic='test-topic', partition=0),
offset=12193, timestamp=1706980591004, log_start_offset=2604, checksum=None, serialized_key_size=-1,
serialized_value_size=200, serialized_header_size=-1)
=====
```

Processing

Listen from Kafka Topic, handled late data,
cleaning, transformation, load to DB.

Event consume

Calculate **BounceRate** and **ConversionRate** from **visitors**,
pagevisitors, and **conversion** column

```
# Parse Kafka messages and apply transformations
parsed_df = stream_df.select(from_json(col("value").cast("string"), kafka_schema).alias("parsed_value")) \
    .select("parsed_value.*") \
    .withColumn("Ts", from_unixtime("Ts").cast(TimestampType())) \
    .withColumn("ConversionRate", expr("(Conversion/Visitors)*100")) \
    .withColumn("BounceRate", expr("(PageVisitors/Visitors)*100")) \
    .withWatermark("Ts", "5 minutes") # Set watermark with a 5-minute handled late data
....
def write_to_postgresql(batch_df, batch_id):
    print(f"Processing batch {batch_id}")
    batch_df.show(truncate=False)

batch_df.write \
    .format("jdbc") \
    .option("url", f"jdbc:postgresql:// {PG_HOST}/{PG_DATABASE}") \
    .option("dbtable", "Socmed_user_sales") \
    .option("user", PG_USER) \
    .option("password", PG_PASSWORD) \
    .mode("append") \
    .save()
```

Processing batch 3013

ID	Kota	Sosial_Media	Usage_Count	Sales	Visitors	PageVisitors	Conversion	Ts	ConversionRate	BounceRate
352ed520-03fb-4933-9d2f-38019883650d	Bekasi	Instagram	849	68	835	161	364	2024-02-03 17:26:45	43.59281437125748	19.281437125748504
58ff840c-91e2-4c86-ae46-33ba4a796262	Depok	X	351	61	525	205	254	2024-02-03 17:27:44	48.38095238095238	39.04761904761905

Processing batch 3014

ID	Kota	Sosial_Media	Usage_Count	Sales	Visitors	PageVisitors	Conversion	Ts	ConversionRate	BounceRate
0dab3815-7e78-4411-aab5-28c8fe449cfb	Bogor	X	493	142	744	248	153	2024-02-03 17:26:38	20.56451612903226	33.33333333333333

Processing batch 3015

ID	Kota	Sosial_Media	Usage_Count	Sales	Visitors	PageVisitors	Conversion	Ts	ConversionRate	BounceRate
53980b33-45d5-4186-8580-9e8c641e4386	Bogor	Facebook	224	95	821	171	435	2024-02-03 17:23:32	52.98416565164433	20.828258221680876
267176da-dbe3-45f2-a0b4-1db9b3363739	Bogor	X	409	52	956	250	192	2024-02-03 17:24:04	20.0836820083682	26.15062761506276

Processing batch 3016

ID	Kota	Sosial_Media	Usage_Count	Sales	Visitors	PageVisitors	Conversion	Ts	ConversionRate	BounceRate
2b38cd11-2296-4678-97cb-231ad1e1f73f	Jakarta	TikTok	522	64	666	134	489	2024-02-03 17:25:12	73.42342342342343	20.12012012012012

Storage

Neon Serverless Postgresql instance



Limitation: limited storage (3gb per branch) and computing (100 hours/month) if you use free version

socmed_user_sales										
#	ID	Kota	Sosial_Media	Usage_Count	Sales	Visitors	PageVisitors	Conversion	Ts	
1	1f295e09-a523-4055-84c4-06317adcdda9	Bekasi	TikTok	345	66	607	125	434	2024-02-03 11:33:5	
2	6c077156-7bd4-4d46-a456-3ac2ebe7e309	Bogor	X	428	86	704	242	139	2024-02-03 11:32:1	
3	0b16d5e5-4864-4ed0-81e5-1577f9b16e32	Depok	Telegram	62	140	870	202	212	2024-02-03 11:36:0	
4	fa9ecc5a-452c-498f-a9ab-a023ba7d9228	Bogor	TikTok	653	87	686	206	274	2024-02-03 11:32:0	
5	c905544b-06e2-4f79-a208-a4c4cc2053db	Tangerang	Instagram	822	124	810	182	387	2024-02-03 11:33:4	
6	0f3ebd71-ce30-4137-96b4-c54ad15b444b	Jakarta	Instagram	981	109	912	82	209	2024-02-03 11:33:4	
7	6e6f1a89-6192-4325-9fbe-c0fbcd6c63b	Depok	Instagram	964	73	726	63	479	2024-02-03 11:34:5	

Dataviz

Lightweight, real-time, open source...



Grafana



Refresh every 5 seconds to display dynamic data.

<https://farahduta7.grafana.net/goto/ikFdXtIR?orgId=1>

Social Media Performance Analysis

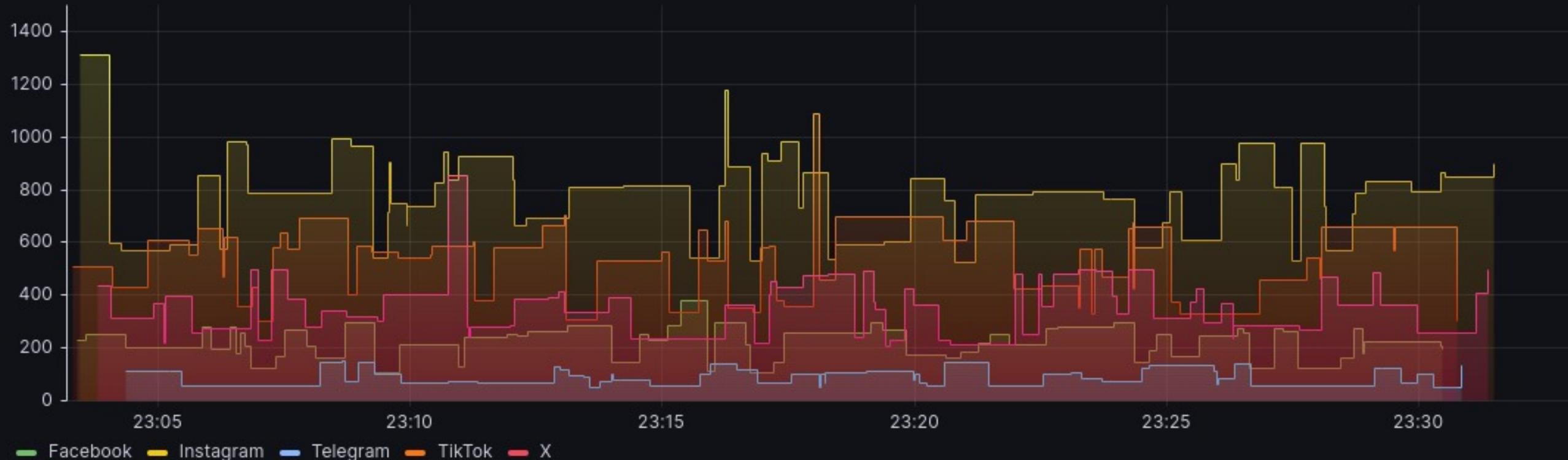
Average Conversion Rate

Platform	Avg. Conversion Rate
Telegram	42.3
Instagram	41.7
X	41.7
Facebook	41.4
TikTok	40.7

Average Bounce Rate

Platform	Avg. Bounce Rate
Telegram	20.9
Instagram	20.5
X	20.5
Facebook	20.3
TikTok	20.7

Real-time Total Usage



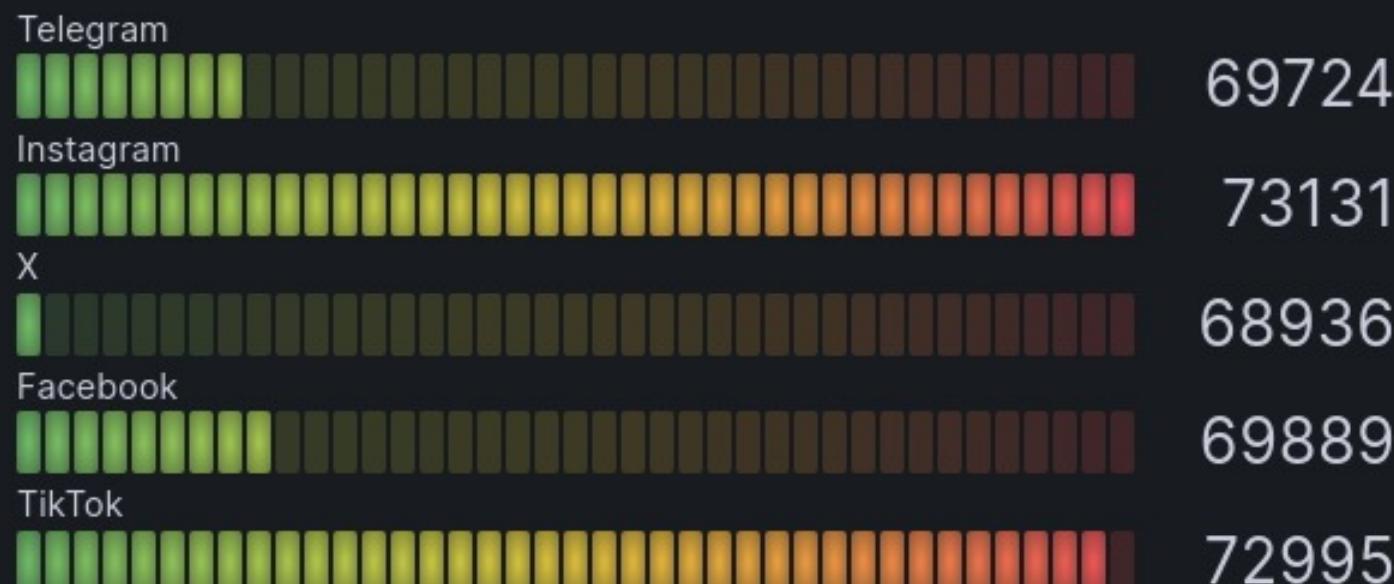
Total Usage

1.36 Mil

Which social media platform has the highest number of visitors in 5 minutes?

Facebook
5004

Total Transaction

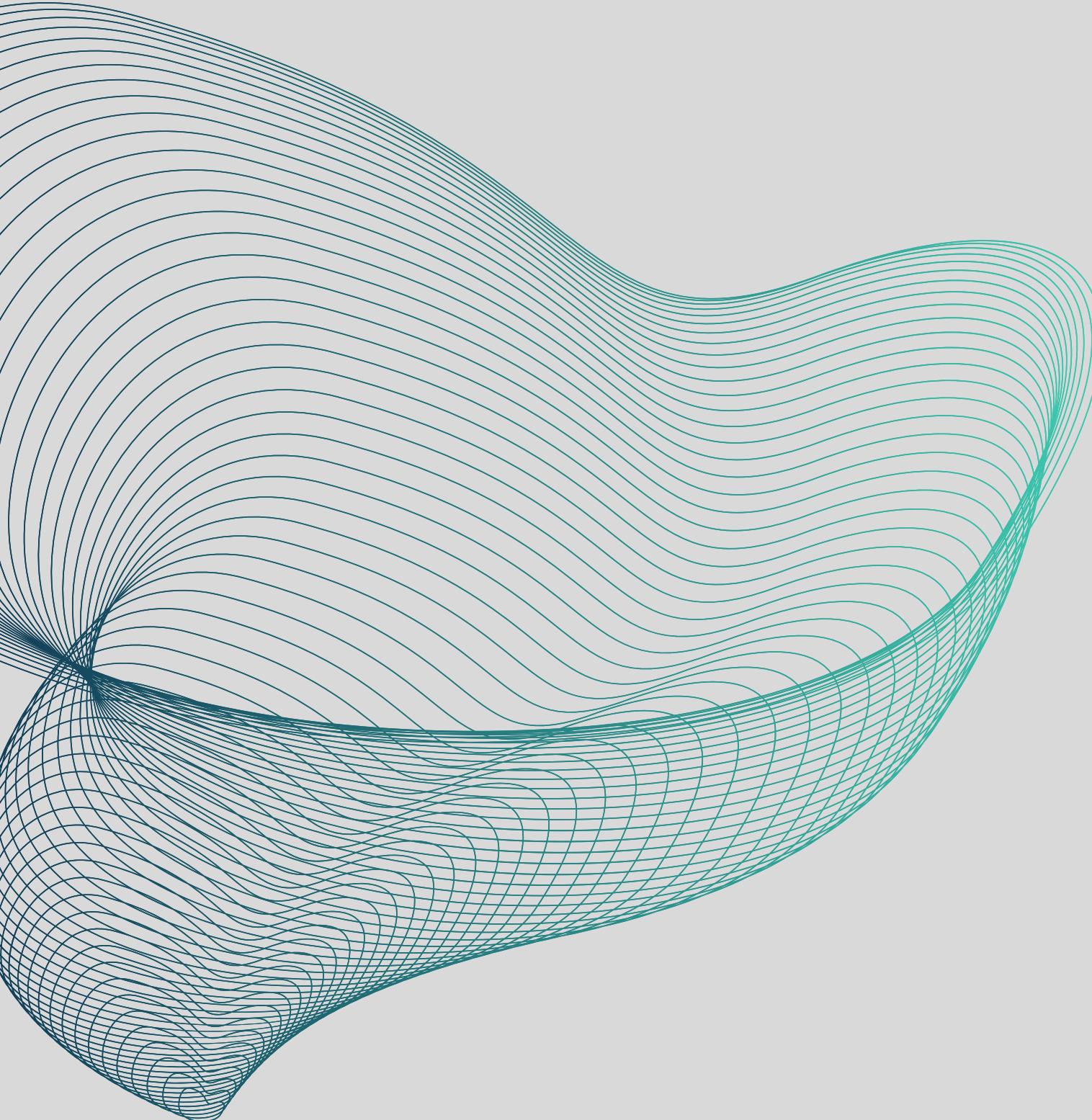


Which city has the highest number of users?

City	Social_Media
Bogor	Facebook
Depok	Instagram
Depok	Telegram
Tangerang	TikTok
Tangerang	X

Which city has the highest number of sales?

City	Social_Media
Bogor	Facebook
Tangerang	Instagram
Depok	Telegram
Tangerang	TikTok
Tangerang	X



Thanks!

Any feedback?

Farah Shadra

farahduta7@gmail.com