# BING NEWS DATA ANALYTICS

## SENTIMENT ANALYSIS

## BING SEARCH API V7

Leveraging the Bing Search API v7 solely for sentiment analysis within Bing News enhances understanding of public sentiment towards topics, events, or entities. It provides valuable insights for sentiment tracking and trend analysis, aiding in decision-making and understanding public perception.

## DATA FACTORY

Data Factory seamlessly integrates with Bing Search API v7, enabling efficient extraction, transformation, and loading (ETL) of data from Bing search results. With its scalable infrastructure, it simplifies the ingestion process, empowering organizations to leverage Bing search data for various analytics and insights.

## ONE LAKE

To store data efficiently, we utilize a data lake architecture. This centralized repository allows for the storage of structured, semi-structured, and unstructured data in its native format, facilitating flexibility and scalability for storing data collected from the Bing Search API v7.

## SYNAPSE DATA ENGINEERING

Leveraging Synapse Data Engineering, we streamline data transformation processes in our project. Its scalable infrastructure enables efficient processing of data sourced from the Bing Search API v7, ensuring seamless integration and optimization for analytics and insights generation.

## DATA SCIENCE

Utilizing Synapse ML for sentiment analysis on data enables automated extraction of sentiment from text data sourced from the Bing Search API v7. This facilitates efficient understanding of public sentiment towards various topics, aiding decision-making and trend analysis.

## POWER BI

For reporting, Power BI offers a dynamic and intuitive platform. Integrating Synapse ML sentiment analysis results with Power BI enables visual representation of sentiment trends sourced from Bing Search API v7 data, empowering stakeholders with actionable insights through interactive dashboards and reports.

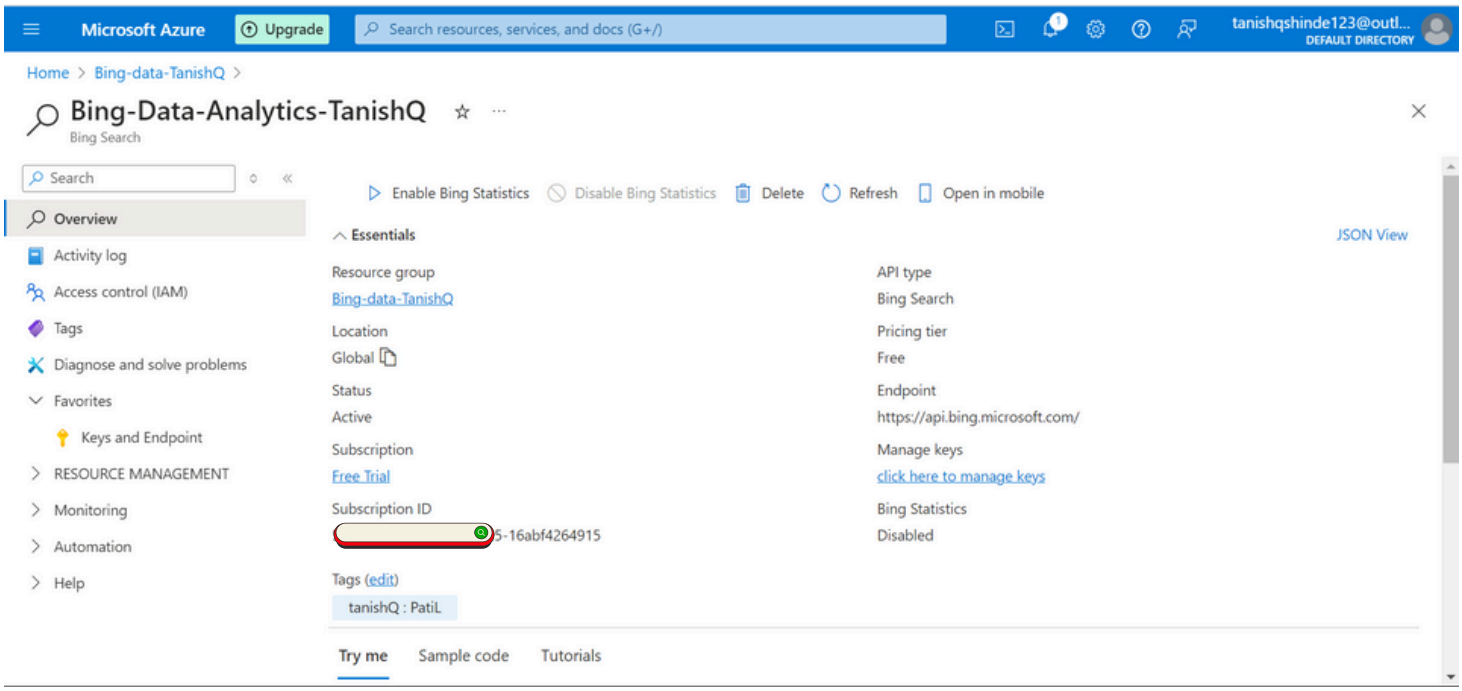
## DATA ACTIVATOR

Utilizing Data Activator, we configure alerts for real-time monitoring of data sourced from the Bing Search API v7. This ensures timely notifications based on predefined criteria, enabling proactive response to significant changes or events in the data, enhancing decision-making and operational efficiency.

## MICROSOFT TEAMS

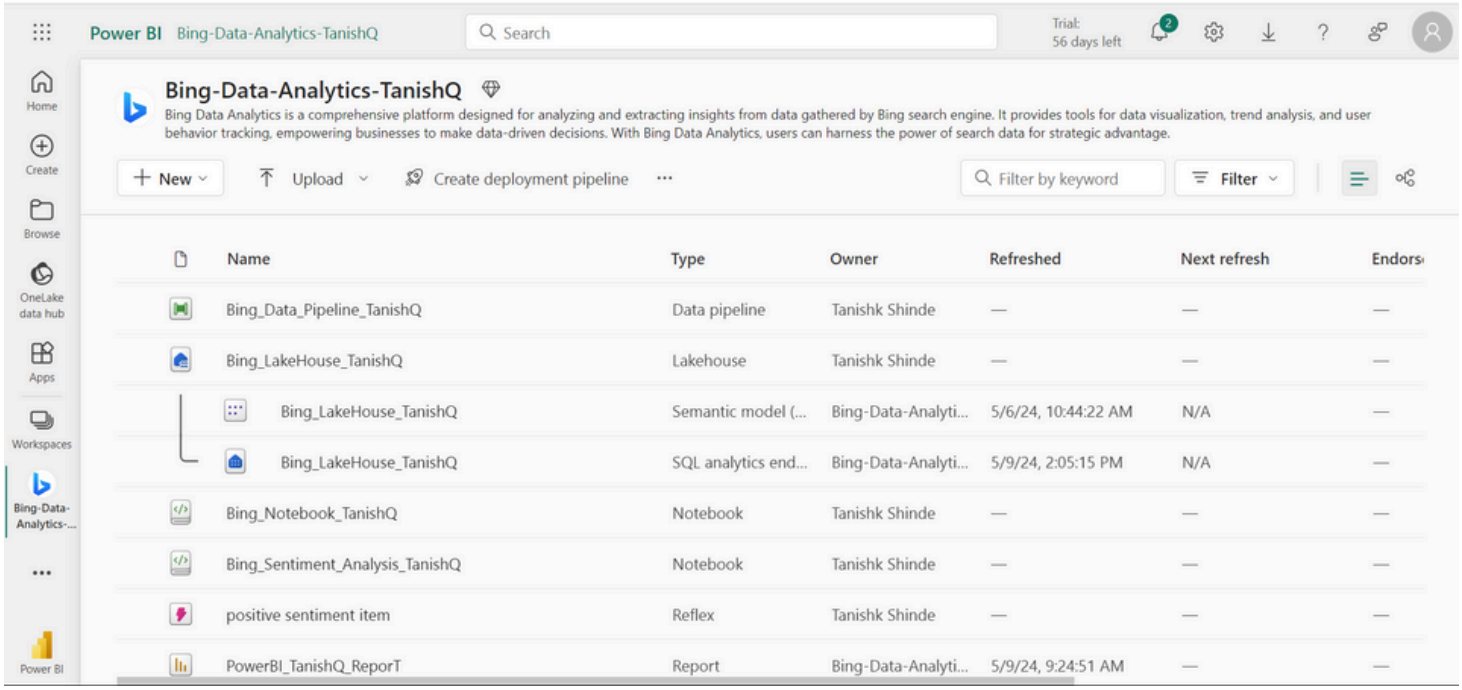
To receive alerts on Microsoft Teams, we integrate Data Activator with Teams' notification system. This setup ensures that alerts triggered by significant events or changes in data sourced from the Bing Search API v7 are promptly delivered to relevant Teams channels or individuals, facilitating collaboration and timely decision-making within the organization.

# BING SEARCH API V7

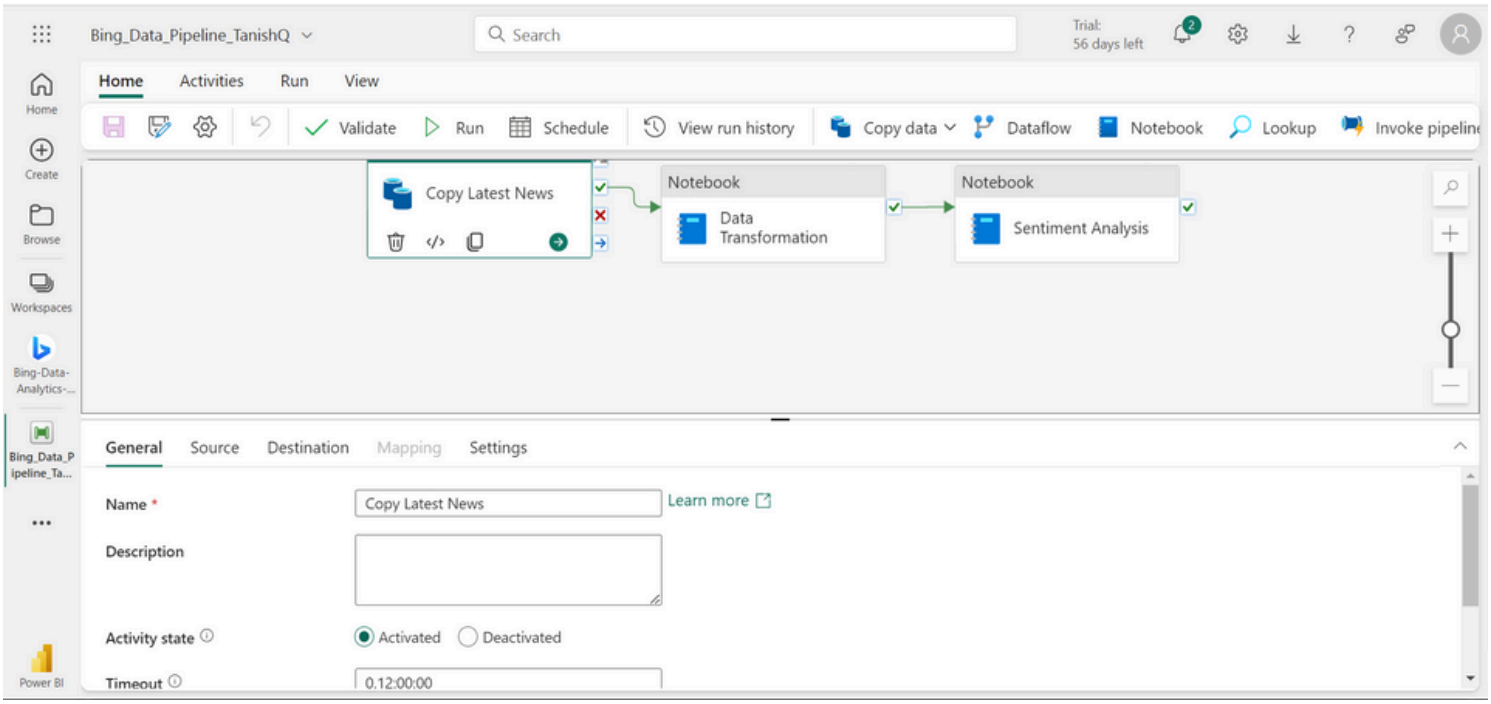



MICROSOFT FABRIC

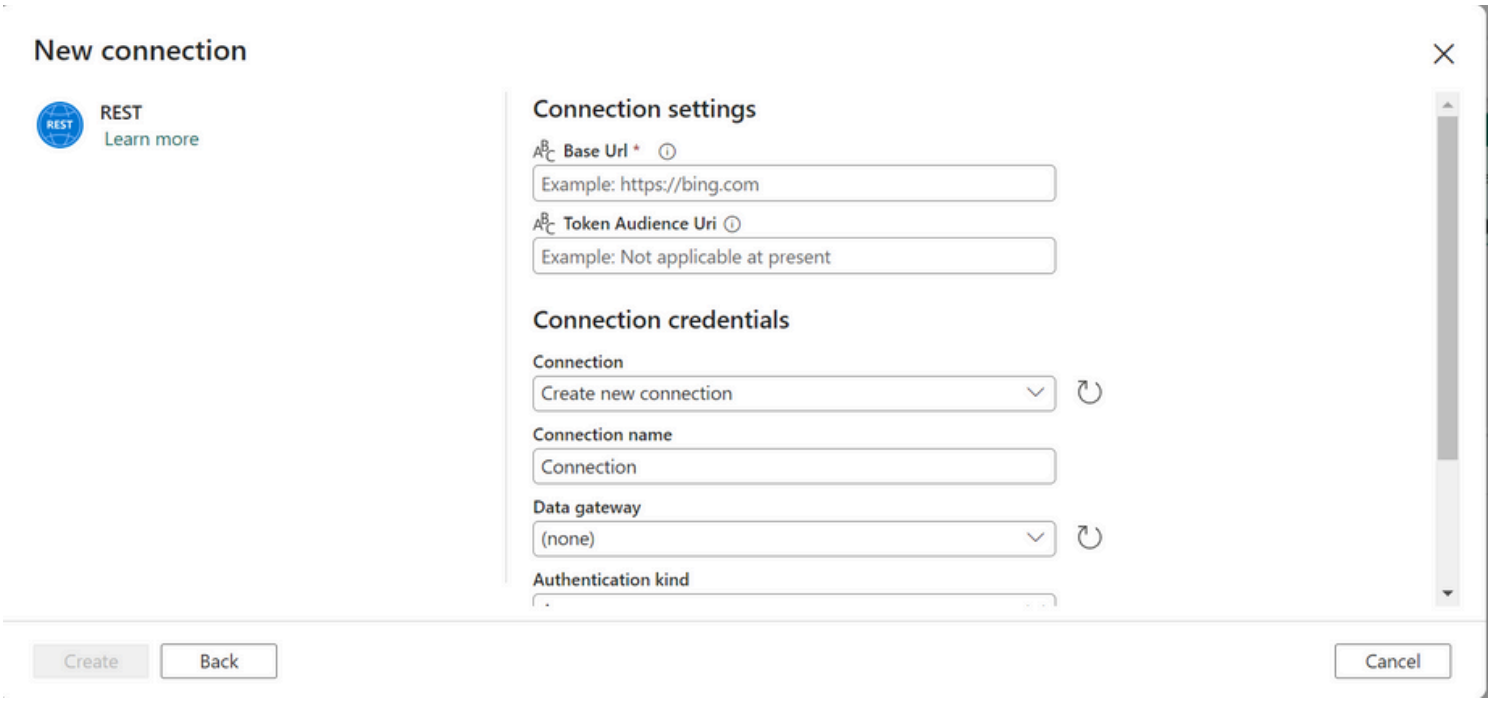
# WORK_SPACES
# BING-DATA-ANALYTICS-TANISHQ



# DATA FACTORY




Azure Data Factory to make REST API calls to the Bing Search API v7 for retrieving information. This involves setting up a linked service in Azure Data Factory to connect to the Bing Search API v7 endpoint and defining appropriate activities within pipelines to invoke the REST API and retrieve the desired data. With this setup, you can efficiently integrate data from Bing Search API v7 into your data workflows within Azure Data Factory.



The Web Search API lets you send a search query to Bing and get back search results that include links to webpages, images, and more. This section provides technical details about the query parameters and headers that you use to request web search results and the JSON response objects that contain them. For examples that show how to make requests, see Bing Web Search overview.

## Endpoints

To request web search results, send a GET request to:

Copy

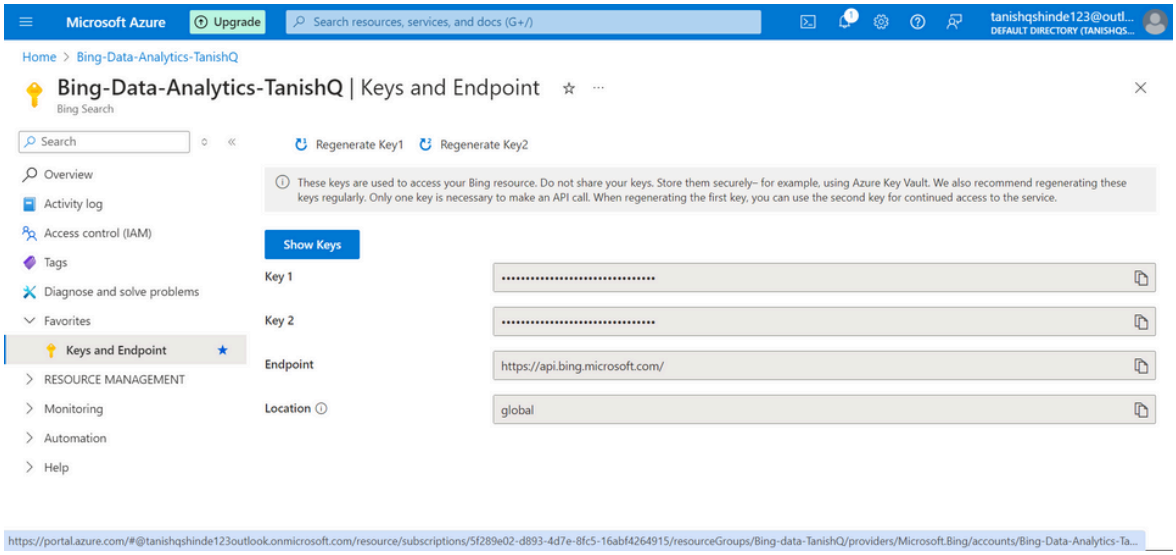
```
https://api.bing.microsoft.com/v7.0/search
```

The request must use the HTTPS protocol.

# REQUEST HEADERS

| | | |
|---|---|---|
| Ocp-Apim-Subscription-Key | Yes | The subscription key that you received when you signed up for this service in Azure Portal while creating a Bing resource. |

# KEYS AND ENDPOINT

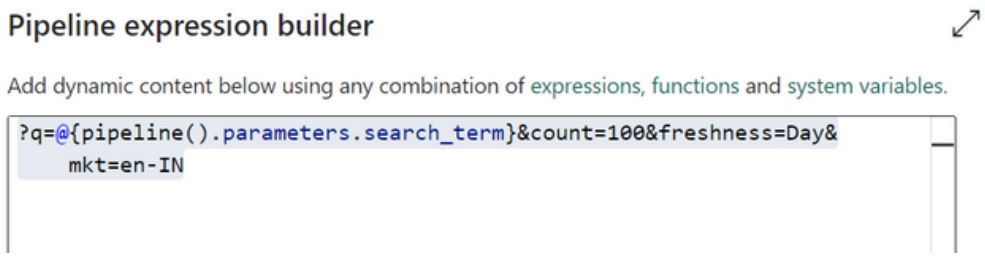


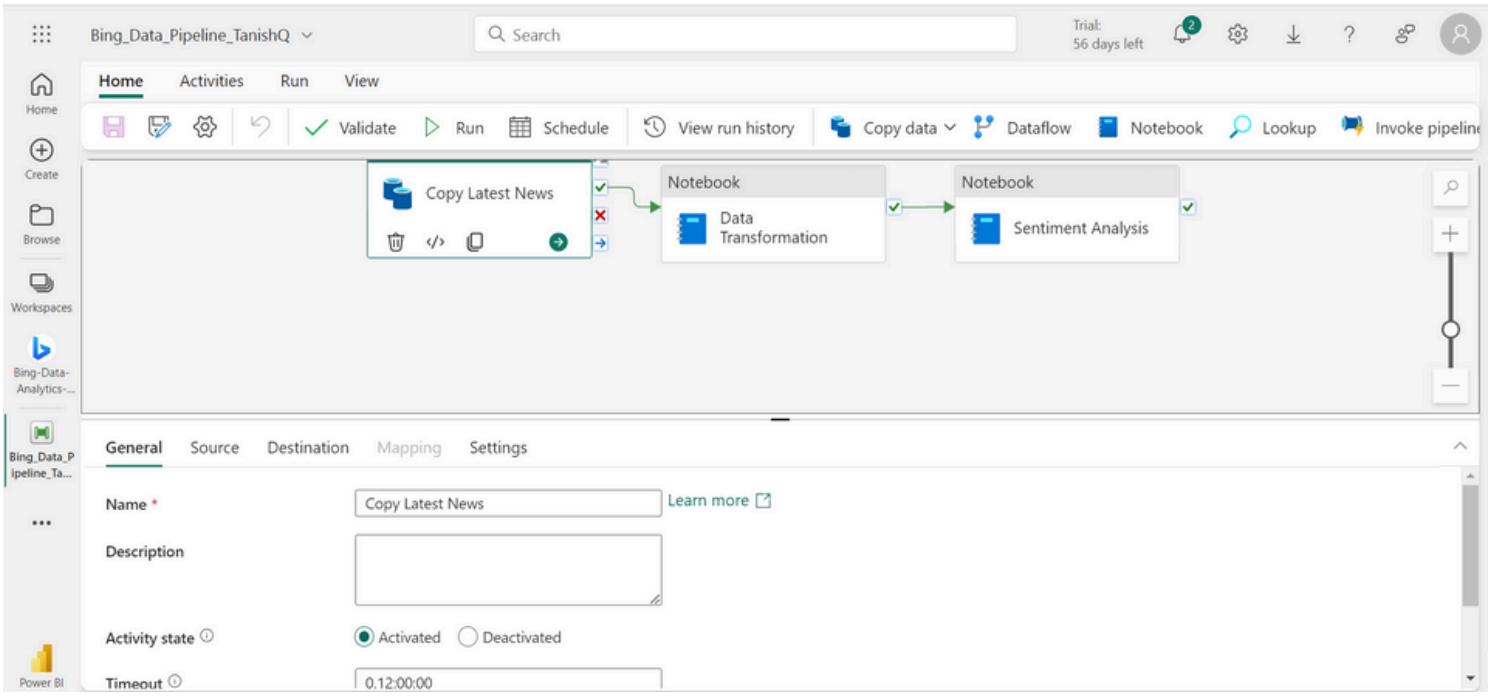
# PREVIEW DATA

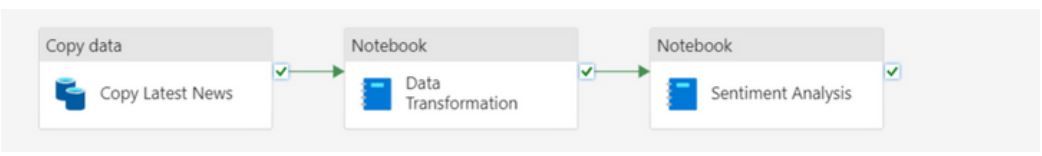
**Preview data**


```
[
  {
    "_type": "News",
    "readLink": "https://api.bing.microsoft.com/api/v7/news/search?q=latest+news",
    "queryContext": {
      "originalQuery": "latest news",
      "adultIntent": false
    },
    "totalEstimatedMatches": 32,
    "sort": [
      {
        "name": "Best match",
        "id": "relevance",
        "isSelected": true,
        "url": "https://api.bing.microsoft.com/api/v7/news/search?q=latest+news"
      },
      {
        "name": "Most recent",
        "id": "date",
        "isSelected": false,
        "url": "https://api.bing.microsoft.com/api/v7/news/search?q=latest+news&sortby=date"
```

# PIPELINE EXPRESSION BUILDER





# SYNAPSE DATA ENGINEERING

```
1  df = spark.read.option("multiline", "true").json("Files/Bing-Latest-News-TanishQ.json")
2  # df now is a Spark DataFrame containing JSON data from "Files/Bing-Latest-News-TanishQ.json".
3  display(df)
```

- Command executed in 819 ms by Tanishk Shinde on 1:33:55 PM, 5/09/24 — PySpark (Python)

Table | Chart | Showing rows 1 - 1 | Inspect | Search

| | _type | queryContext | readLink | sort | totalEstimatedMatches | value |
|---|---|---|---|---|---|---|
| 1 | News | {"adultIntent":false,"... | https://api.bin... | [{"id":"relev... | 39 | [{"datePubli... |

```
1  df= df.select("value")
```

- Command executed in 246 ms by Tanishk Shinde on 1:33:56 PM, 5/09/24 — PySpark (Python)

```
1  display(df)
```

- Command executed in 261 ms by Tanishk Shinde on 1:33:57 PM, 5/09/24 — PySpark (Python)

Table | Chart | Showing rows 1 - 1 | Inspect | Search

| | value |
|---|---|
| 1 | [{"datePubli... |

```
1  from pyspark.sql.functions import explode
2
3  df_explode = df.select(explode(df["value"]).alias("json_Object"))
```

- Command executed in 237 ms by Tanishk Shinde on 1:33:58 PM, 5/09/24 — PySpark (Python)

```
1  display(df_explode)
```

- Command executed in 814 ms by Tanishk Shinde on 1:33:59 PM, 5/09/24 — PySpark (Python)

Table | Chart | Showing rows 1 - 39 | Inspect | Search

| | json_Object |
|---|---|
| 1 | {"datePublished":"... |
| 2 | {"datePublished":"... |
| 3 | {"datePublished":"... |
| 4 | {"datePublished":"... |
| 5 | {"datePublished":"... |
| 6 | {"datePublished":"... |
| 7 | {"datePublished":"... |
| 8 | {"datePublished":"... |

```
1  json_list = df_explode.toJSON().collect()
```
✓ - Command executed in 812 ms by Tanishk Shinde on 1:34:00 PM, 5/09/24 — PySpark (Python)

+ Code + Markdown

```
1  print(json_list)
```
✓ - Command executed in 364 ms by Tanishk Shinde on 1:34:01 PM, 5/09/24 — PySpark (Python)


```
['{"json_Object":{"about":[{"name":"Pickleball","readLink":"https://api.bing.microsoft.com/api/v7/entities/d6a2574a-9b44-74ee-6e04-ba1976d07181"},{"name":"Table tennis","readLink":"https://api.bing.microsoft.com/api/v7/entities/574f0ecc-043a-6f47-a383-39bd82059d6c"},{"name":"Badminton","readLink":"https://api.bing.microsoft.com/api/v7/entities/33a4342a-8aa1-5d93-45d5-5d8c27494e59"},{"name":"Tennis","readLink":"https://api.bing.microsoft.com/api/v7/entities/7a186152-36fa-501a-77a6-71b4f02f4c99"}],"category":"Sports","datePublished":"2024-05-09T03:50:00.0000000Z","description":"Natekar Sports launches World Pickleball League with Sony Entertainment as strategic investor, aiming to promote sport globally.","image":{"thumbnail":{"contentUrl":"https://www.bing.com/th?id=OVFT.ppeiu0svQ-D9lqPwUf4f9i&pid=News","height":393,"width":700}},"name":"Natekar Sports and Gaming set to launch World Pickleball League","provider":[{"_type":"Organization","image":{"thumbnail":
```


```
1  import json
2  news_json = json.loads(json_list[25])
```
✓ - Command executed in 246 ms by Tanishk Shinde on 1:34:02 PM, 5/09/24 — PySpark (Python)

```
1  print(news_json)
```
✓ - Command executed in 240 ms by Tanishk Shinde on 1:34:02 PM, 5/09/24 — PySpark (Python)


```
{'json_Object': {'category': 'Sports', 'datePublished': '2024-05-08T23:41:00.0000000Z', 'description': 'From the cricket world, Mongolia will kickstart their seven-match tour of Japan with the first T20I at the Sano Cricket Ground. Meanwhile, Bangladesh will look to clinch the five-match T20I series against Zimbabwe with a win in the third game at Chattogram.', 'name': 'Sports Highlights: Dortmund Pip PSG 1-0 In Champions League, Enter Final', 'provider': [{'_type': 'Organization', 'image': {'thumbnail': {'contentUrl': 'https://www.bing.com/th?id=ODF.nP2rzyfwERvAkqhL1eUBtg&pid=news'}}, 'name': 'Outlookindia'}], 'url': 'https://www.outlookindia.com/sports/live-scores-today-may-7-2024-ipl-dc-vs-rr-ucl-psg-vs-dortmund-world-schedule-results'}}
```


```
1  print(news_json['json_Object']['description'])
```
✓ - Command executed in 275 ms by Tanishk Shinde on 1:34:03 PM, 5/09/24 — PySpark (Python)

```
From the cricket world, Mongolia will kickstart their seven-match tour of Japan with the first T20I at the Sano Cricket Ground.
Meanwhile, Bangladesh will look to clinch the five-match T20I series against Zimbabwe with a win in the third game at Chattogram.
```

```
1  print(news_json["json_Object"]["name"])
2  print(news_json["json_Object"]["description"])
3  print(news_json["json_Object"]["category"])
4  print(news_json["json_Object"]["url"])
5
6  # Check if "image" key exists before accessing it
7  if "image" in news_json["json_Object"]["provider"][0]:
8      print(news_json["json_Object"]["provider"][0]["image"]["thumbnail"]["contentUrl"])
9  else:
10     print("Image not available")
11
12 print(news_json["json_Object"]["provider"][0]["name"])
13 print(news_json["json_Object"]["datePublished"])
14
```
✓ - Command executed in 262 ms by Tanishk Shinde on 1:34:03 PM, 5/09/24 — PySpark (Python)

```
Sports Highlights: Dortmund Pip PSG 1-0 In Champions League, Enter Final
From the cricket world, Mongolia will kickstart their seven-match tour of Japan with the first T20I at the Sano Cricket Ground.
Meanwhile, Bangladesh will look to clinch the five-match T20I series against Zimbabwe with a win in the third game at Chattogram.
Sports
https://www.outlookindia.com/sports/live-scores-today-may-7-2024-ipl-dc-vs-rr-ucl-psg-vs-dortmund-world-schedule-results
https://www.bing.com/th?id=ODF.nP2rzyfwERvAkqhL1eUBtg&pid=news
Outlookindia
2024-05-08T23:41:00.0000000Z
```

```python
title = []
description = []
category = []
url = []
image = []
provider = []
datePublished = []

for json_str in json_list:
    try:
        article = json.loads(json_str)

        if "json_Object" in article:
            if "name" in article["json_Object"]:
                title.append(article["json_Object"]["name"])
            else:
                title.append(None)

            if "description" in article["json_Object"]:
                description.append(article["json_Object"]["description"])
            else:
                description.append(None)

            if "category" in article["json_Object"]:
                category.append(article["json_Object"]["category"])
            else:
                category.append(None)

            if "url" in article["json_Object"]:
                url.append(article["json_Object"]["url"])
            else:
                url.append(None)

            if "image" in article["json_Object"] and "thumbnail" in article["json_Object"]["image"]:
                image.append(article["json_Object"]["image"]["thumbnail"].get("contentUrl"))
            else:
                image.append(None)

            if "provider" in article["json_Object"] and len(article["json_Object"]["provider"]) > 0:
                provider.append(article["json_Object"]["provider"][0].get("name"))
            else:
                provider.append(None)

            if "datePublished" in article["json_Object"]:
                datePublished.append(article["json_Object"]["datePublished"])
            else:
                datePublished.append(None)

    except Exception as e:
        print(f"Error processing json object: {e}")
```

✓ - Command executed in 260 ms by Tanishk Shinde on 1:34:04 PM, 5/09/24

```python
description
```

✓ - Command executed in 250 ms by Tanishk Shinde on 1:34:05 PM, 5/09/24 PySpark (Python)

```
['Natekar Sports launches World Pickleball League with Sony Entertainment as strategic investor, aiming to promote sport globally.',
 'ROASTING THE GOAT — Former NFL star Shawne Merriman told Fox News Digital he thought Netflix's "Greatest Roast of All Time: Tom Brady" was one of the best roasts he\'d ever seen. Continue reading…',
 "Welcome to ESPN India's daily blog, where we aim to bring you all the news from the dynamic world of Indian sports: from the wide variety of Olympic sports to franchise leagues. May has plenty of sporting action taking place - which you can find listed out in ESPN India's sporting calendar.",
```

```python
from pyspark.sql.types import StructType,StructField,StringType
# combine list
data = list(zip(title,description,category,url,image,provider,datePublished))
# define schema

schema = StructType([
    StructField("title",StringType(),True),
    StructField("description",StringType(),True),
    StructField("category",StringType(),True),
    StructField("url",StringType(),True),
    StructField("image",StringType(),True),
    StructField("provider",StringType(),True),
    StructField("datePublished",StringType(),True)

])
df_cleaned = spark.createDataFrame(data,schema=schema)
```

```
display(df_cleaned)
```

- Command executed in 1 sec 536 ms by Tanishk Shinde on 1:34:09 PM, 5/09/24

Table | Chart | Showing rows 1 - 39

| | title | description | category | url | image | provider | datePublished |
|---|---|---|---|---|---|---|---|
| 1 | Natekar Sp... | Natekar Sports la... | Sports | https://ww... | https://www... | Business Line | 2024-05-09T03:50:0... |
| 2 | Fox News S... | ROASTING THE ... | Sports | https://ww... | https://www... | Fox News on ... | 2024-05-08T19:48:0... |
| 3 | Indian Spor... | Welcome to ESP... | Sports | https://ww... | https://www... | ESPN | 2024-05-09T03:30:0... |
| 4 | Alderson Br... | Visit ESPN for Al... | NULL | https://ww... | https://www... | ESPN | 2024-05-09T07:12:0... |
| 5 | Prep sports... | With a 27-2 reco... | Sports | https://ww... | https://www... | Los Angeles Ti... | 2024-05-09T00:31:5... |
| 6 | Disney's Rit... | With ESPN, Disn... | NULL | https://ww... | https://www... | NextTV on MS... | 2024-05-08T17:33:0... |

```
from pyspark.sql.functions import to_date , date_format

df_cleaned_final = df_cleaned.withColumn("datePublished",date_format(to_date("datePublished"),"dd-MMM-yyyy"))
```

- Command executed in 267 ms by Tanishk Shinde on 1:34:09 PM, 5/09/24

```
display(df_cleaned_final)
```

- Command executed in 828 ms by Tanishk Shinde on 1:34:11 PM, 5/09/24

Table | Chart | Showing rows 1 - 39

| | title | description | category | url | image | provider | datePublished |
|---|---|---|---|---|---|---|---|
| 1 | Natekar Sp... | Natekar Sports la... | Sports | https://ww... | https://www... | Business Line | 09-May-2024 |
| 2 | Fox News S... | ROASTING THE ... | Sports | https://ww... | https://www... | Fox News on ... | 08-May-2024 |
| 3 | Indian Spor... | Welcome to ESP... | Sports | https://ww... | https://www... | ESPN | 09-May-2024 |
| 4 | Alderson Br... | Visit ESPN for Al... | NULL | https://ww... | https://www... | ESPN | 09-May-2024 |
| 5 | Prep sports... | With a 27-2 reco... | Sports | https://ww... | https://www... | Los Angeles Ti... | 09-May-2024 |
| 6 | Disney's Rit... | With ESPN, Disn... | NULL | https://ww... | https://www... | NextTV on MS... | 08-May-2024 |
| 7 | Michigan S... | Michigan becam... | Sports | https://ww... | https://www... | Forbes | 08-May-2024 |

```
from pyspark.sql.utils import AnalysisException

try:

    table_name = 'Bing_LakeHouse_TanishQ.tbl_latest_news'

    df_cleaned_final.write.format("delta").saveAsTable(table_name)

except AnalysisException:

    print("Table Already Exists")

    df_cleaned_final.createOrReplaceTempView("vw_df_cleaned_final")

    spark.sql(f""" MERGE INTO {table_name} target_table
                   USING vw_df_cleaned_final source_view

                   ON source_view.url = target_table.url

                   WHEN MATCHED AND
                   source_view.title <> target_table.title OR
                   source_view.description <> target_table.description OR
                   source_view.category <> target_table.category OR
                   source_view.image <> target_table.image OR
                   source_view.provider <> target_table.provider OR
                   source_view.datePublished <> target_table.datePublished

                   THEN UPDATE SET *

                   WHEN NOT MATCHED THEN INSERT *
    """)
```

```
df = spark.sql("SELECT * FROM Bing_LakeHouse_TanishQ.tbl_latest_news LIMIT 1000")
display(df)
```

✓ - Command executed in 1 sec 532 ms by Tanishk Shinde on 1:34:36 PM, 5/09/24

Table | Chart | Showing rows 1 - 68

| | ABC title | ABC description | ABC category | ABC url | ABC image | ABC provider | ABC datePublished |
|---|---|---|---|---|---|---|---|
| 1 | Fox News S... | ROASTING THE ... | Sports | https://ww... | https://www... | Fox News on ... | 08-May-2024 |
| 2 | Indian Spor... | Welcome to ESP... | Sports | https://ww... | https://www... | ESPN | 09-May-2024 |
| 3 | Alderson Br... | Visit ESPN for Al... | NULL | https://ww... | https://www... | ESPN | 09-May-2024 |
| 4 | Natekar Sp... | Natekar Sports la... | Sports | https://ww... | https://www... | Business Line | 09-May-2024 |

```
%%sql
SELECT count(*) FROM tbl_latest_news
```

✓ - Command executed in 1 sec 467 ms by Tanishk Shinde on 1:34:38 PM, 5/09/24

Table | Chart

| | 12L count(1) |
|---|---|
| 1 | 68 |

# DATA SCIENCE

## SENTIMENT ANALYSIS USING SYNAPSE ML(

```
1  df = spark.sql("SELECT * FROM Bing_LakeHouse_TanishQ.tbl_latest_news LIMIT 1000")
2  display(df)
```

✓ - Command executed in 14 sec 634 ms by Tanishk Shinde on 1:49:42 PM, 5/09/24

Table | Chart | Showing rows 1 - 76

| | ABC title | ABC description | ABC category | ABC url | ABC image | ABC provider | ABC datePubli... |
|---|---|---|---|---|---|---|---|
| 1 | Ex-interpret... | LOS ANGEL... | NULL | https://abc... | https://ww... | ABC | 08-May-2024 |
| 2 | Today in Sp... | LeBron Jam... | Sports | https://apn... | NULL | Associated ... | 08-May-2024 |
| 3 | Alderson Br... | Visit ESPN f... | NULL | https://ww... | https://ww... | ESPN | 09-May-2024 |
| 4 | Indian Spor... | Welcome t... | Sports | https://ww... | https://ww... | ESPN | 09-May-2024 |
| 5 | Michigan S... | Michigan b... | Sports | https://ww... | https://ww... | Forbes | 08-May-2024 |
| 6 | Virginia Sp... | The legal s... | Sports | https://ww... | https://ww... | Forbes | 08-May-2024 |

```
1  # APPLY THE MODEL TO OUR DATAFRAME
2  result = model.transform(df)
```

✓ - Command executed in 786 ms by Tanishk Shinde on 1:49:45 PM, 5/09/24 — PySpark (Python)

```
1  import synapse.ml.core
2  from synapse.ml.services import AnalyzeText
```

✓ - Command executed in 243 ms by Tanishk Shinde on 1:49:43 PM, 5/09/24 — PySpark (Python)

```
1  # IMPORT THE MODEL AND CONFIGURE THE INPUT AND OUTPUT COLUMNS
2
3  model = (AnalyzeText()
4          .setTextCol("description")
5          .setKind("SentimentAnalysis")
6          .setOutputCol("response")
7          .setErrorCol("error"))
```

✓ - Command executed in 248 ms by Tanishk Shinde on 1:49:43 PM, 5/09/24 — PySpark (Python)

```
1  display(result)
```

✓ - Command executed in 6 sec 324 ms by Tanishk Shinde on 1:49:51 PM, 5/09/24

Table | Chart | Showing rows 1 - 76

| | ABC title | ABC description | ABC category | ABC url | ABC image | ABC provider | ABC datePubli... | ANY error | ANY response |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Ex-interpret... | LOS ANGEL... | NULL | https://abc... | https://ww... | ABC | 08-May-2024 | NULL | {"document... |
| 2 | Today in Sp... | LeBron Jam... | Sports | https://apn... | NULL | Associated ... | 08-May-2024 | NULL | {"document... |
| 3 | Alderson Br... | Visit ESPN f... | NULL | https://ww... | https://ww... | ESPN | 09-May-2024 | NULL | {"document... |
| 4 | Indian Spor... | Welcome t... | Sports | https://ww... | https://ww... | ESPN | 09-May-2024 | NULL | {"document... |
| 5 | Michigan S... | Michigan b... | Sports | https://ww... | https://ww... | Forbes | 08-May-2024 | NULL | {"document... |
| 6 | Virginia Sp... | The legal s... | Sports | https://ww... | https://ww... | Forbes | 08-May-2024 | NULL | {"document... |
| 7 | The Sports ... | Yoshinobu ... | Sports | https://ww... | NULL | Los Angele... | 08-May-2024 | NULL | {"document... |

```python
# CREATE SENTIMENT COLUMN
from pyspark.sql.functions import col

sentiment_df = result.withColumn("sentiment",col("response.documents.sentiment"))
```

✓ - Command executed in 235 ms by Tanishk Shinde on 1:49:52 PM, 5/09/24

```python
display(sentiment_df)
```

✓ - Command executed in 3 sec 534 ms by Tanishk Shinde on 1:49:56 PM, 5/09/24

Table | Chart | Showing rows 1 - 76

| | ABC title | ABC description | ABC category | ABC url | ABC image | ABC provider | ABC datePubli... | ANY error | ANY response | ABC sentiment |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Ex-interpret... | LOS ANGEL... | NULL | https://abc.... | https://ww... | ABC | 08-May-2024 | NULL | {"document... | negative |
| 2 | Today in Sp... | LeBron Jam... | Sports | https://apn.... | NULL | Associated ... | 08-May-2024 | NULL | {"document... | neutral |
| 3 | Alderson Br... | Visit ESPN f... | NULL | https://ww.... | https://ww... | ESPN | 09-May-2024 | NULL | {"document... | neutral |
| 4 | Indian Spor... | Welcome t... | Sports | https://ww.... | https://ww... | ESPN | 09-May-2024 | NULL | {"document... | neutral |
| 5 | Michigan S... | Michigan b... | Sports | https://ww.... | https://ww... | Forbes | 08-May-2024 | NULL | {"document... | neutral |
| 6 | Virginia Sp... | The legal s... | Sports | https://ww.... | https://ww... | Forbes | 08-May-2024 | NULL | {"document... | neutral |
| 7 | The Sports ... | Yoshinobu ... | Sports | https://ww.... | NULL | Los Angele... | 08-May-2024 | NULL | {"document... | neutral |

```python
sentiment_df_final = sentiment_df.drop("error","response")
```

✓ - Command executed in 251 ms by Tanishk Shinde on 1:49:57 PM, 5/09/24

```python
display(sentiment_df_final)
```

✓ - Command executed in 3 sec 521 ms by Tanishk Shinde on 1:50:00 PM, 5/09/24

Table | Chart | Showing rows 1 - 76

| | ABC title | ABC description | ABC category | ABC url | ABC image | ABC provider | ABC datePublished | ABC sentiment |
|---|---|---|---|---|---|---|---|---|
| 1 | Ex-interpret... | LOS ANGELES -- ... | NULL | https://abc.... | https://www.... | ABC | 08-May-2024 | negative |
| 2 | Today in Sp... | LeBron James sc... | Sports | https://apn.... | NULL | Associated Pr... | 08-May-2024 | neutral |
| 3 | Alderson Br... | Visit ESPN for Al... | NULL | https://ww.... | https://www.... | ESPN | 09-May-2024 | neutral |
| 4 | Indian Spor... | Welcome to ESP... | Sports | https://ww.... | https://www.... | ESPN | 09-May-2024 | neutral |
| 5 | Michigan S... | Michigan becam... | Sports | https://ww.... | https://www.... | Forbes | 08-May-2024 | neutral |
| 6 | Virginia Sp... | The legal sports ... | Sports | https://ww.... | https://www.... | Forbes | 08-May-2024 | neutral |

```python
from pyspark.sql.functions import col, to_date

sentiment_df_final = sentiment_df_final.withColumn("datePublished",to_date(col("datePublished"),"dd-MMM-yyyy"))
```

✓ - Command executed in 263 ms by Tanishk Shinde on 1:58:59 PM, 5/09/24 PySpark (Python)

```python
from pyspark.sql.utils import AnalysisException

try:

    table_name = 'Bing_LakeHouse_TanishQ.tbl_Sentiment_Analysis'

    sentiment_df_final.write.format("delta").saveAsTable(table_name)

except AnalysisException:

    print("Table Already Exists")

    sentiment_df_final.createOrReplaceTempView("vw_sentiment_df_final")

    spark.sql(f""" MERGE INTO {table_name} target_table
                   USING vw_sentiment_df_final source_view

                   ON source_view.url = target_table.url

                   WHEN MATCHED AND
                   source_view.title <> target_table.title OR
                   source_view.description <> target_table.description OR
                   source_view.category <> target_table.category OR
                   source_view.image <> target_table.image OR
                   source_view.provider <> target_table.provider OR
                   source_view.datePublished <> target_table.datePublished

                   THEN UPDATE SET *

                   WHEN NOT MATCHED THEN INSERT *
    """)

```

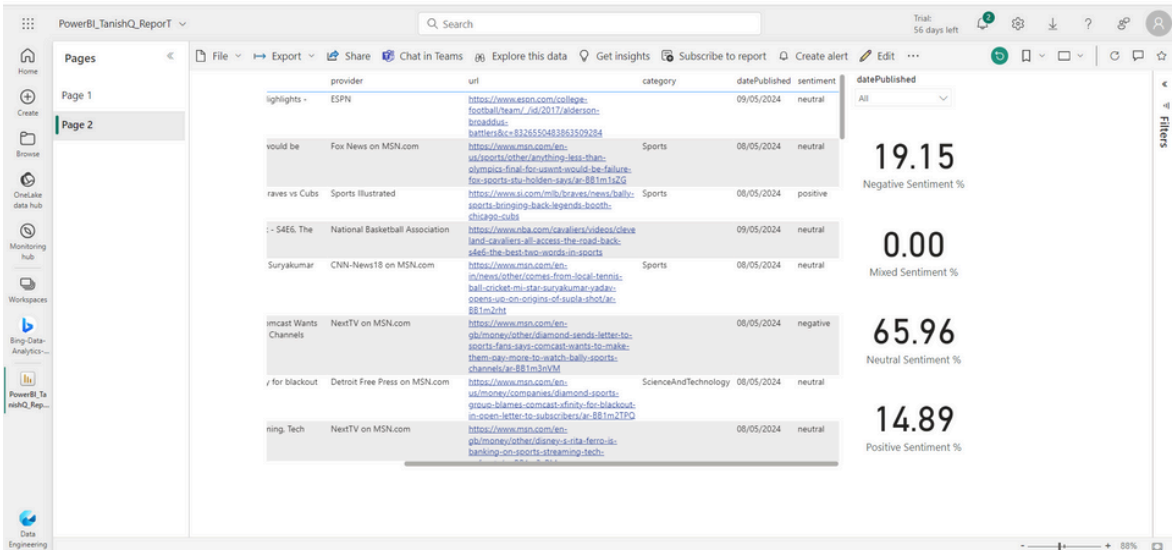
```python
df = spark.sql("SELECT * FROM Bing_LakeHouse_TanishQ.tbl_sentiment_analysis LIMIT 1000")
display(df)
```

✓ - Command executed in 1 sec 584 ms by Tanishk Shinde on 1:57:02 PM, 5/09/24

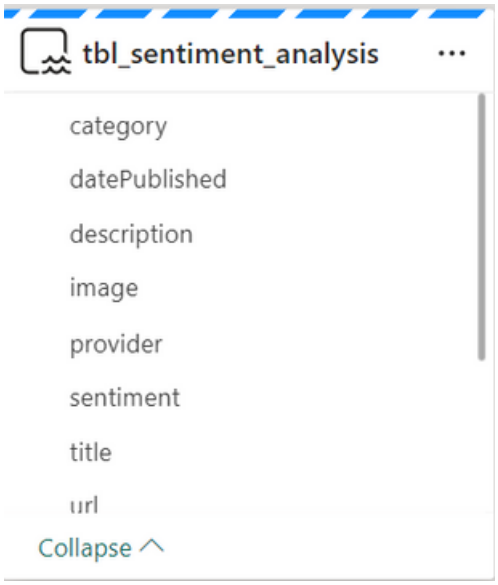
Table | Chart | Showing rows 1 - 76

| | title | description | category | url | image | provider | datePublished | sentiment |
|---|---|---|---|---|---|---|---|---|
| 1 | Ex-interpret... | LOS ANGELES -- ... | NULL | https://abc... | https://www... | ABC | 2024-05-08 | negative |
| 2 | Today in Sp... | LeBron James sc... | Sports | https://apn... | NULL | Associated Pr... | 2024-05-08 | neutral |
| 3 | PSU bank s... | PSU bank stocks,... | Business | https://eco... | https://www... | Indiatimes | 2024-05-06 | negative |
| 4 | iSportz Ann... | ORLANDO, FL / ... | ScienceAndTec... | https://fina... | https://www... | YAHOO!Finance | 2024-05-09 | neutral |
| 5 | PRESS DIGE... | The following ar... | LifeStyle | https://fina... | NULL | YAHOO!Finance | 2024-05-06 | negative |
| 6 | Israel Move... | The order—whic... | World | https://tim... | https://www... | Time | 2024-05-06 | negative |

# POWER BI




# OPEN DATA MODEL



# DATA ANALYSIS EXPRESSIONS (DAX)

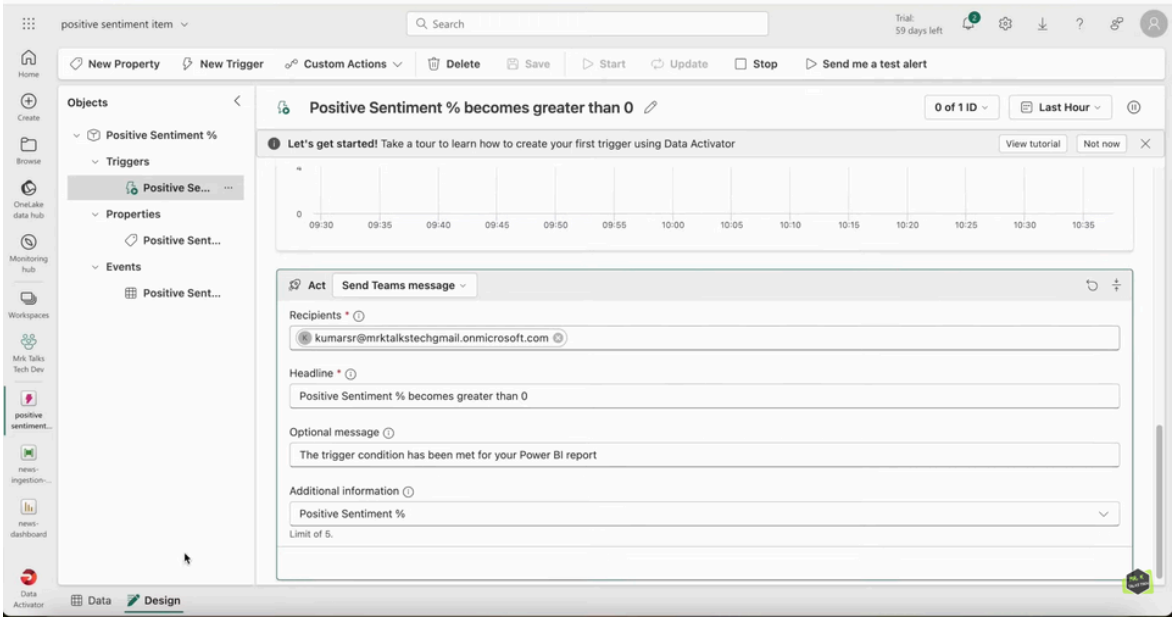
```
Mixed Sentiment % =
IF (
    COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "mixed")) > 0,
    DIVIDE (
        CALCULATE (
            COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "mixed"))
        ),
        COUNTROWS ('tbl_sentiment_analysis')
    )*100,
    0
)
```

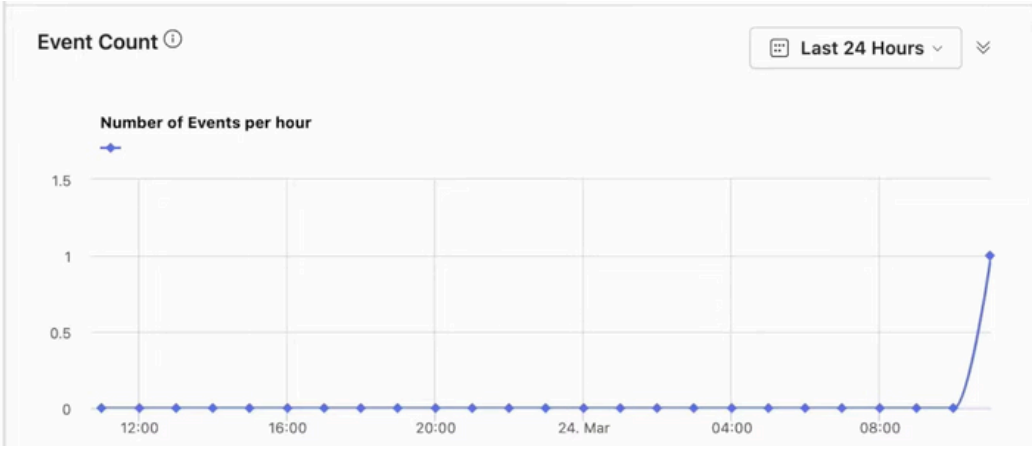
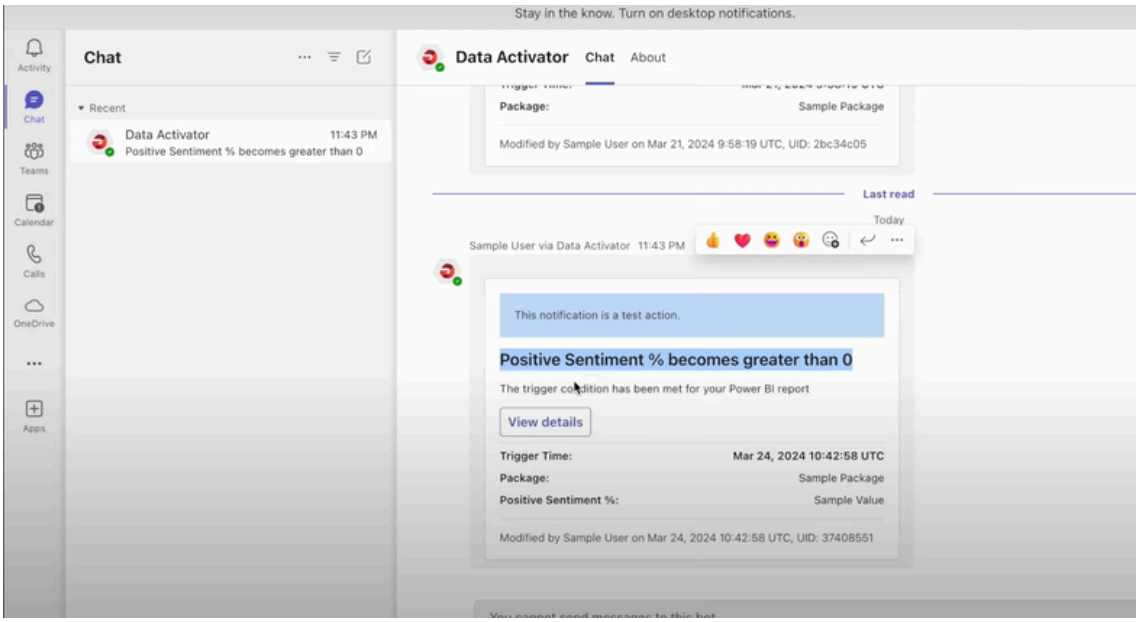
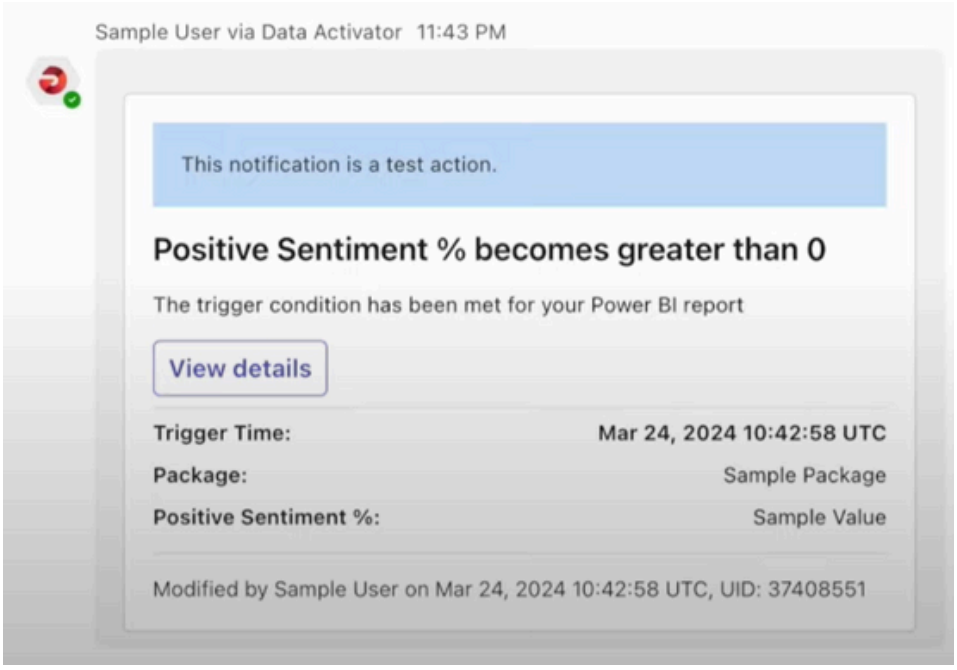
```
Negative Sentiment % =
IF (
    COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "negative")) > 0,
    DIVIDE (
        CALCULATE (
            COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "negative"))
        ),
        COUNTROWS ('tbl_sentiment_analysis')
    )*100,
    0
)
```

```
Neutral Sentiment % =
IF (
    COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "neutral")) > 0,
    DIVIDE (
        CALCULATE (
            COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "neutral"))
        ),
        COUNTROWS ('tbl_sentiment_analysis')
    )*100,
    0
)
```

```
Positive Sentiment % =
IF (
    COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "positive")) > 0,
    DIVIDE (
        CALCULATE (
            COUNTROWS ( FILTER ('tbl_sentiment_analysis', 'tbl_sentiment_analysis'[Sentiment] = "positive"))
        ),
        COUNTROWS ('tbl_sentiment_analysis')
    )*100,
    0
)
```

# DATA ACTIVATOR






# MICROSOFT TEAMS