E-Commerce: The trend of Digital Goods in the AOL Dataset 2006

Ebna Sina
Krishnakumar Mudaliar
Chimagbanwe Umeh
Ravi Pandit
Sunbla Khan

Business Intelligence

M.Sc. Data Science

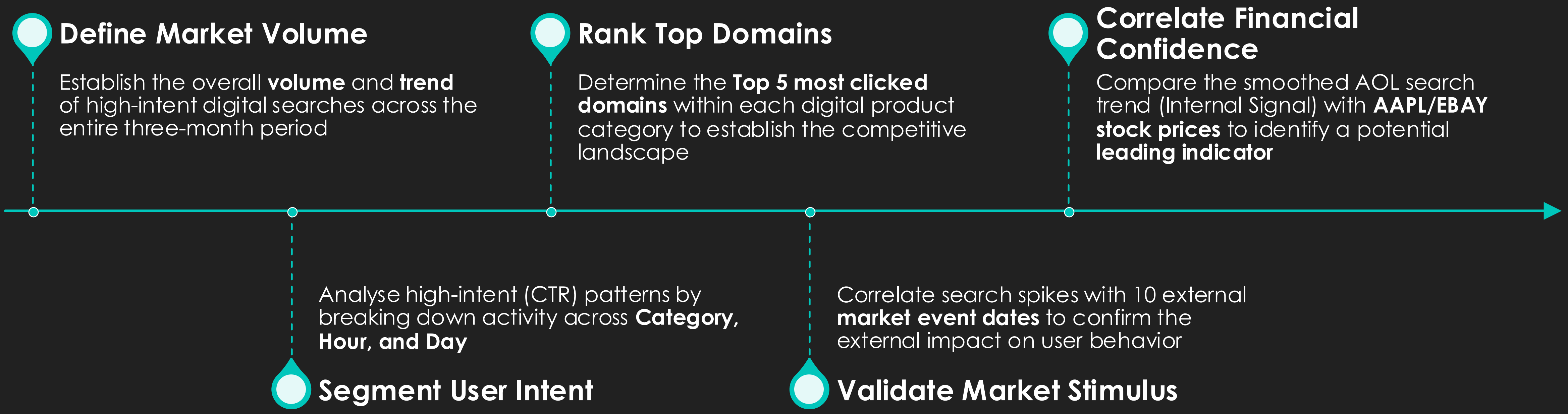
Supervised By: Alexander Löser

December 08, 2025

E-commerce on Digital goods

- **Intangible Products:** Digital goods are non-physical items (e.g., software, music, e-books, subscriptions) sold and delivered purely over the internet.
- **Unique Economics:** Characterized by a **zero marginal cost** (nearly free to produce additional units) and instant global distribution, leading to high-profit margin potential.
- **Historical Pivot:** The Q2 2006 period represents the crucial phase where the market pivoted from physical media (CDs, DVDs) to digital downloads and early subscription models, establishing today's digital economy foundations.

Analysis Progression



Define Market Volume

Establish the overall **volume** and **trend** of high-intent digital searches across the entire three-month period

Segment User Intent

Analyse high-intent (CTR) patterns by breaking down activity across **Category, Hour, and Day**

Rank Top Domains

Determine the **Top 5 most clicked domains** within each digital product category to establish the competitive landscape

Validate Market Stimulus

Correlate search spikes with 10 external **market event dates** to confirm the external impact on user behavior

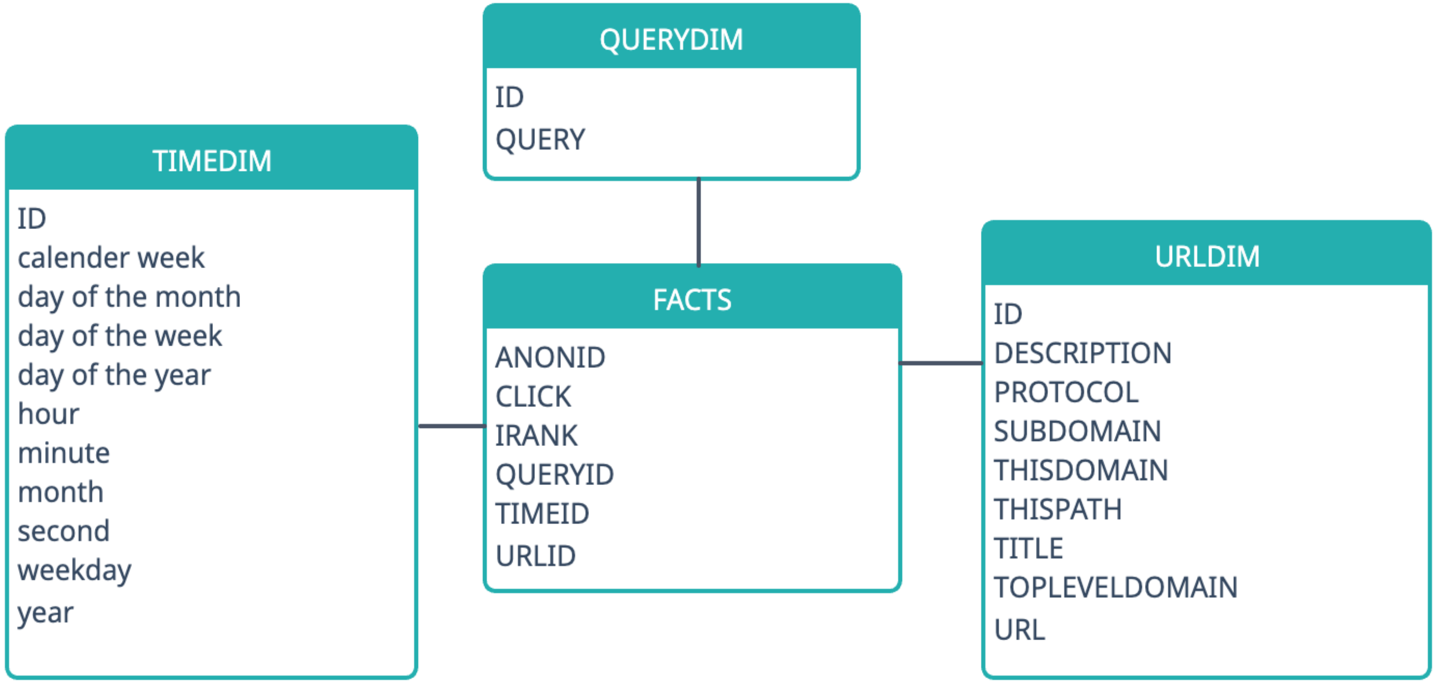
Correlate Financial Confidence

Compare the smoothed AOL search trend (Internal Signal) with **AAPL/EBAY stock prices** to identify a potential **leading indicator**

QUESTION 1

What was the overall volume and trend of digital content-related searches across all weeks, and what was the **total contribution by month?**

Initial Star Schema



QUERYDIM
ID
QUERY

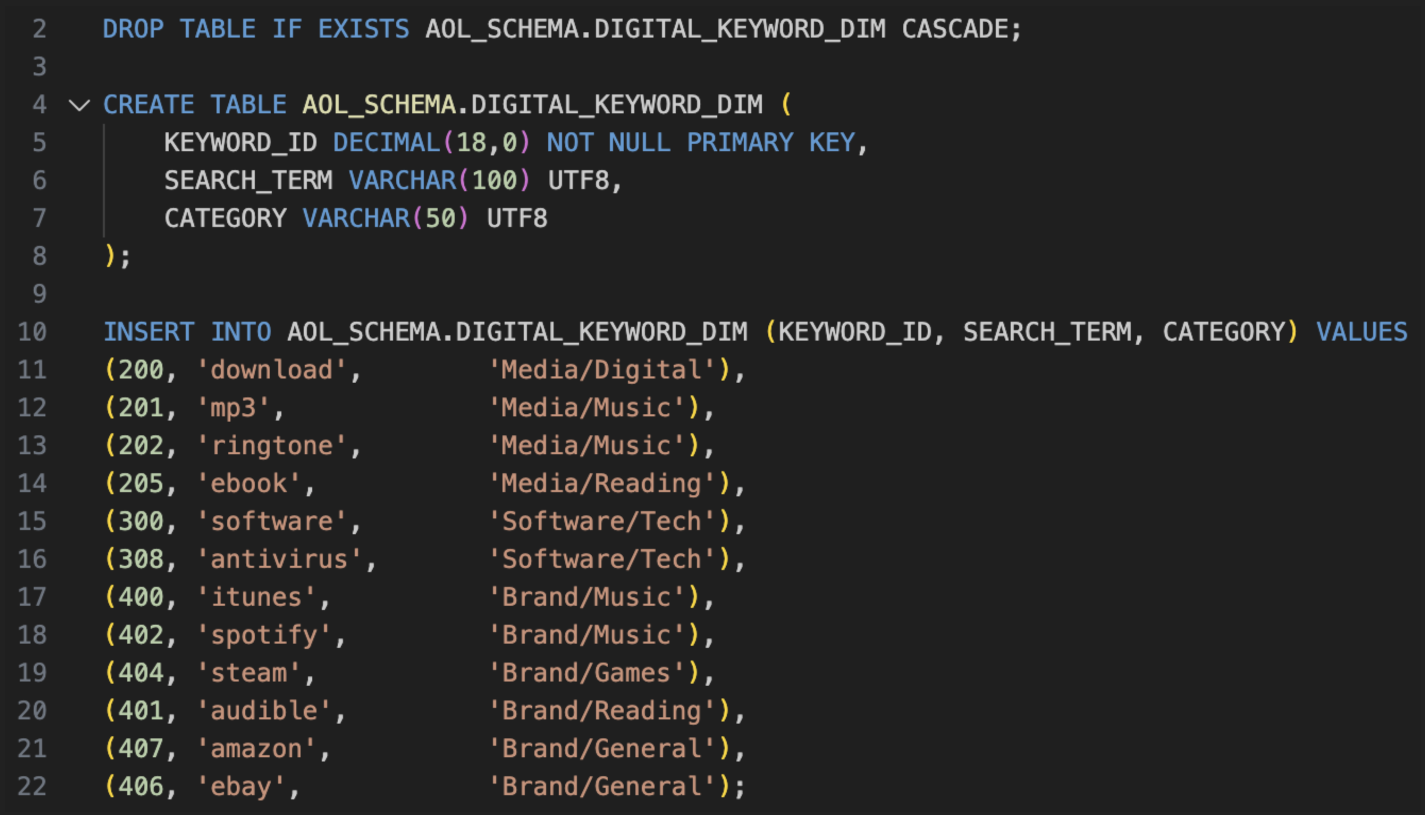
TIMEDIM
ID
calender week
day of the month
day of the week
day of the year
hour
minute
month
second
weekday
year

FACTS
ANONID
CLICK
IRANK
QUERYID
TIMEID
URLID

URLDIM
ID
DESCRIPTION
PROTOCOL
SUBDOMAIN
THISDOMAIN
THISPATH
TITLE
TOPLEVELDOMAIN
URL

Q1: Volume analysis of digital content-related searches

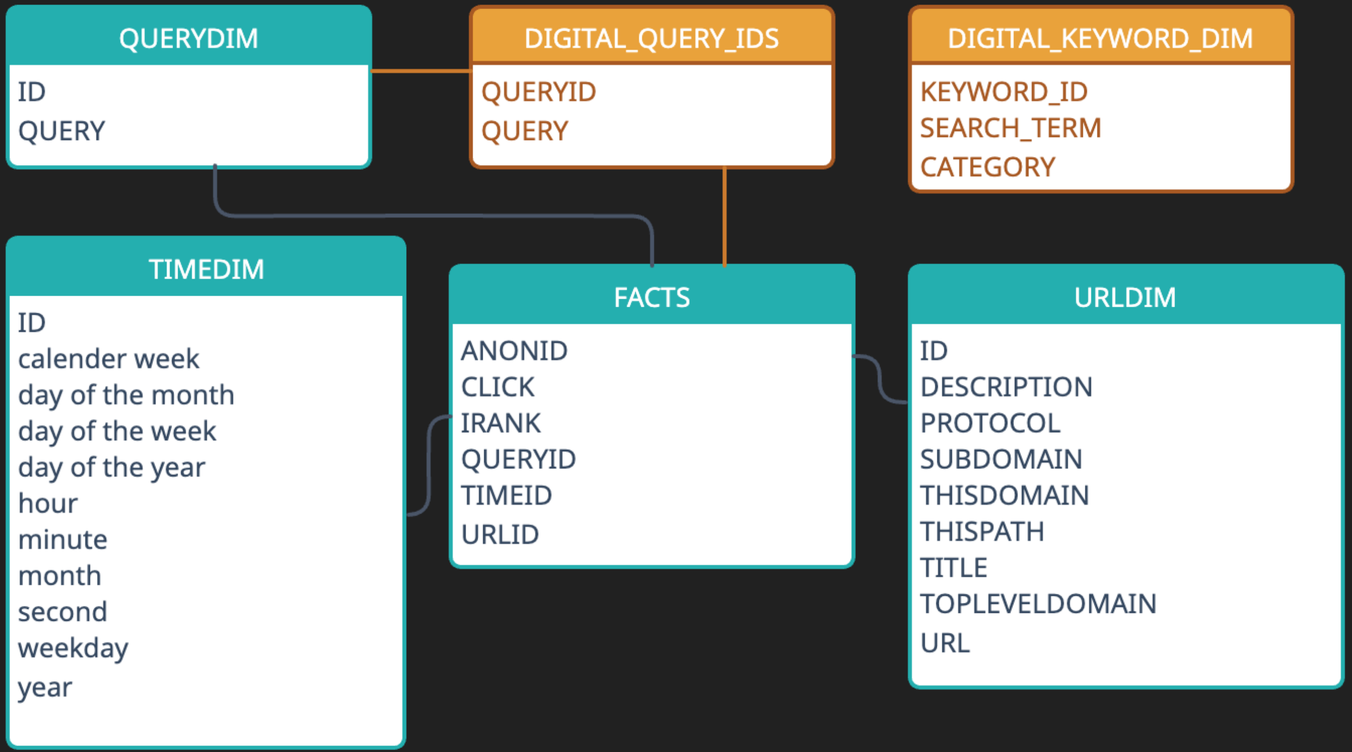
Define the keywords used for querying the term **"digital content"** and their Category in a table DIGITAL_KEYWORD_DIM



```sql
DROP TABLE IF EXISTS AOL_SCHEMA.DIGITAL_KEYWORD_DIM CASCADE;

CREATE TABLE AOL_SCHEMA.DIGITAL_KEYWORD_DIM (
    KEYWORD_ID DECIMAL(18,0) NOT NULL PRIMARY KEY,
    SEARCH_TERM VARCHAR(100) UTF8,
    CATEGORY VARCHAR(50) UTF8
);

INSERT INTO AOL_SCHEMA.DIGITAL_KEYWORD_DIM (KEYWORD_ID, SEARCH_TERM, CATEGORY) VALUES
(200, 'download',        'Media/Digital'),
(201, 'mp3',             'Media/Music'),
(202, 'ringtone',        'Media/Music'),
(205, 'ebook',           'Media/Reading'),
(300, 'software',        'Software/Tech'),
(308, 'antivirus',       'Software/Tech'),
(400, 'itunes',          'Brand/Music'),
(402, 'spotify',         'Brand/Music'),
(404, 'steam',           'Brand/Games'),
(401, 'audible',         'Brand/Reading'),
(407, 'amazon',          'Brand/General'),
(406, 'ebay',            'Brand/General');
```

An image of the DDL for the AOL_SCHEMA.DIGITAL_KEYWORD_DIM

Extends Schema to create a Materialize view of the related query IDs



QUERYDIM
ID
QUERY

DIGITAL_QUERY_IDS
QUERYID
QUERY

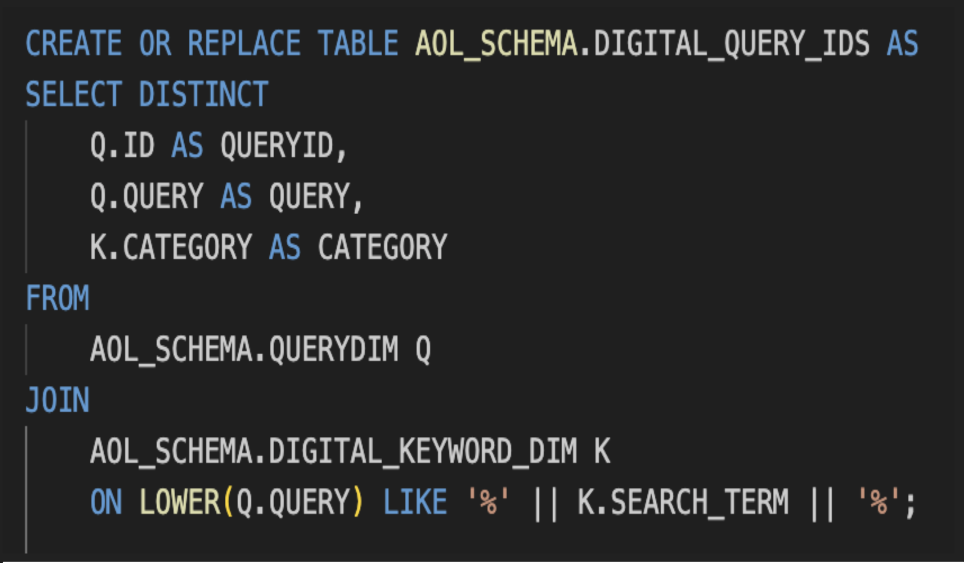
DIGITAL_KEYWORD_DIM
KEYWORD_ID
SEARCH_TERM
CATEGORY

TIMEDIM
ID
calender week
day of the month
day of the week
day of the year
hour
minute
month
second
weekday
year

FACTS
ANONID
CLICK
IRANK
QUERYID
TIMEID
URLID

URLDIM
ID
DESCRIPTION
PROTOCOL
SUBDOMAIN
THISDOMAIN
THISPATH
TITLE
TOPLEVELDOMAIN
URL

DIGITAL_QUERY_IDS: hold queries on digital content for faster query with FACTs, about 68,380 records, instead of 10 million total records from the **QUERYDIM** table.



```
CREATE OR REPLACE TABLE AOL_SCHEMA.DIGITAL_QUERY_IDS AS
SELECT DISTINCT
    Q.ID AS QUERYID,
    Q.QUERY AS QUERY,
    K.CATEGORY AS CATEGORY
FROM
    AOL_SCHEMA.QUERYDIM Q
JOIN
    AOL_SCHEMA.DIGITAL_KEYWORD_DIM K
    ON LOWER(Q.QUERY) LIKE '%' || K.SEARCH_TERM || '%';
```

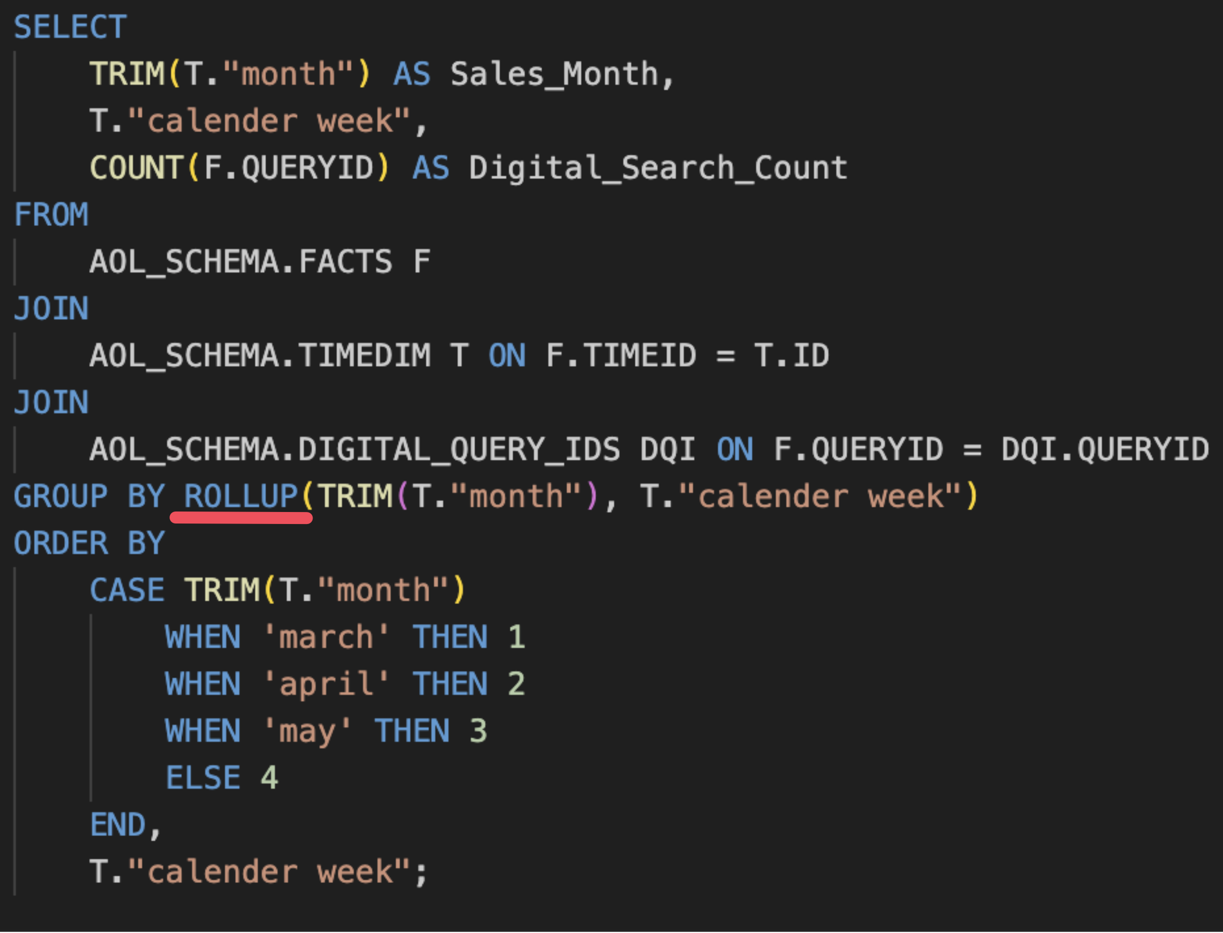
An image of the DDL for the AOL_SCHEMA.DIGITAL_QUERY_IDS

Q1: Volume analysis of digital content-related searches

- Join TIMEDIM and DIGITAL_QUERY_IDS

- Group by Month and week using ROLLUP to get the break-downs

- Order by Month also making sure the month is sorted well

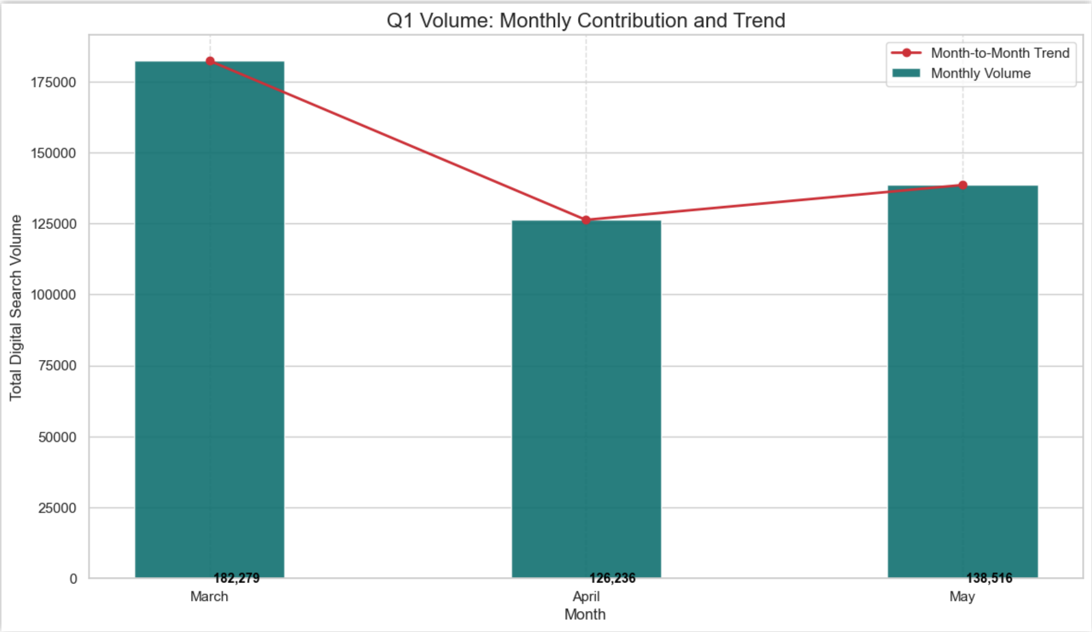


```
SELECT
    TRIM(T."month") AS Sales_Month,
    T."calender week",
    COUNT(F.QUERYID) AS Digital_Search_Count
FROM
    AOL_SCHEMA.FACTS F
JOIN
    AOL_SCHEMA.TIMEDIM T ON F.TIMEID = T.ID
JOIN
    AOL_SCHEMA.DIGITAL_QUERY_IDS DQI ON F.QUERYID = DQI.QUERYID
GROUP BY ROLLUP(TRIM(T."month"), T."calender week")
ORDER BY
    CASE TRIM(T."month")
        WHEN 'march' THEN 1
        WHEN 'april' THEN 2
        WHEN 'may' THEN 3
        ELSE 4
    END,
    T."calender week";
```

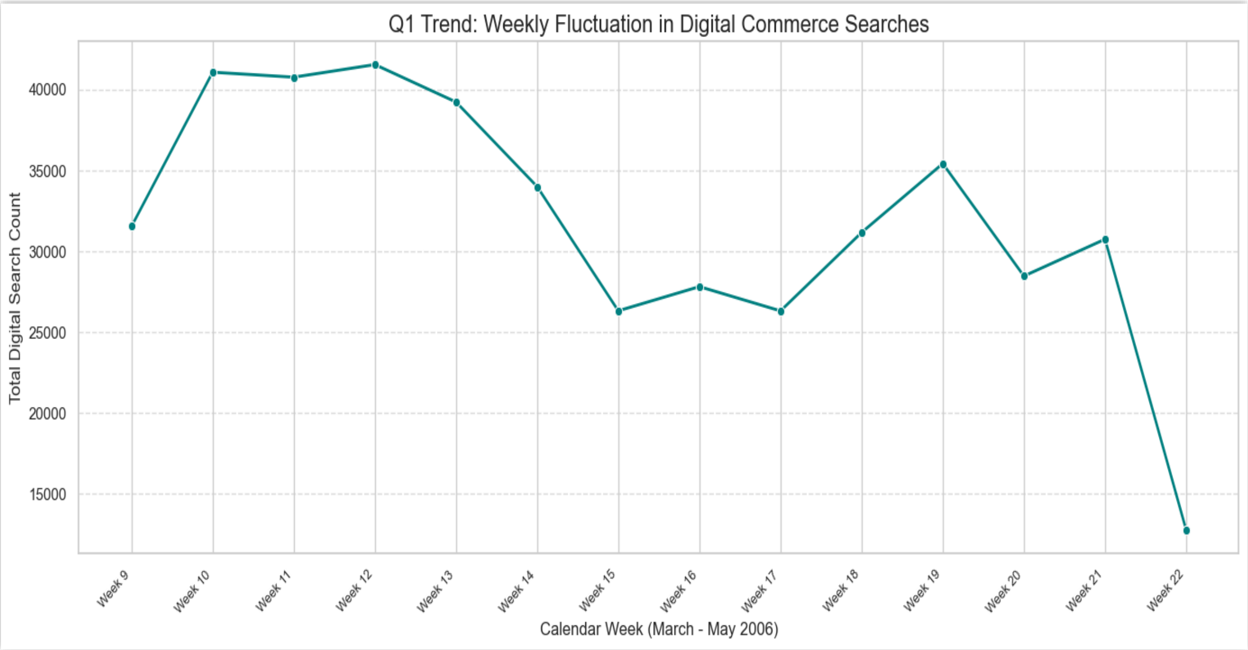
An image of the DQL for the monthly and weekly analysis of searches on digital content

Q1: Visualization of the search volumes

Monthly trend



Weekly trend



March has the highest search for digital-related content
A decline in April and an incline in May

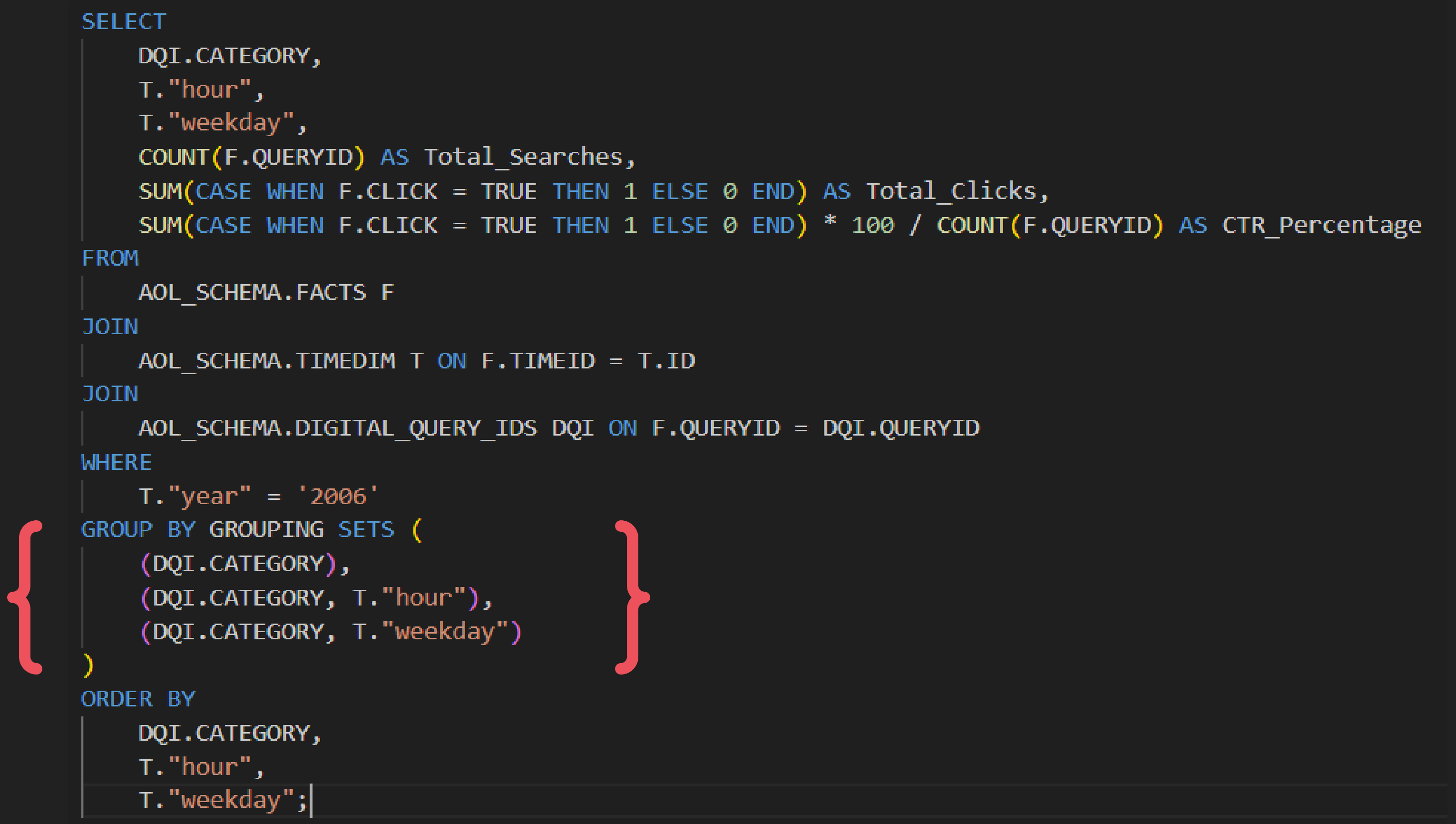
The searches on went from week 9 to 22, with week 12 being the highest and week 22 as the lowest, which could because of not enough data

QUESTION 2

Which specific digital commerce categories (Music/Media, Software, Brands) **drove the highest click-through rate (CTR)**, and how did the user search volume and intent **compare based on the hour and day of the week**?

Q2: Rank the digital commerce categories by user intent (CTR)

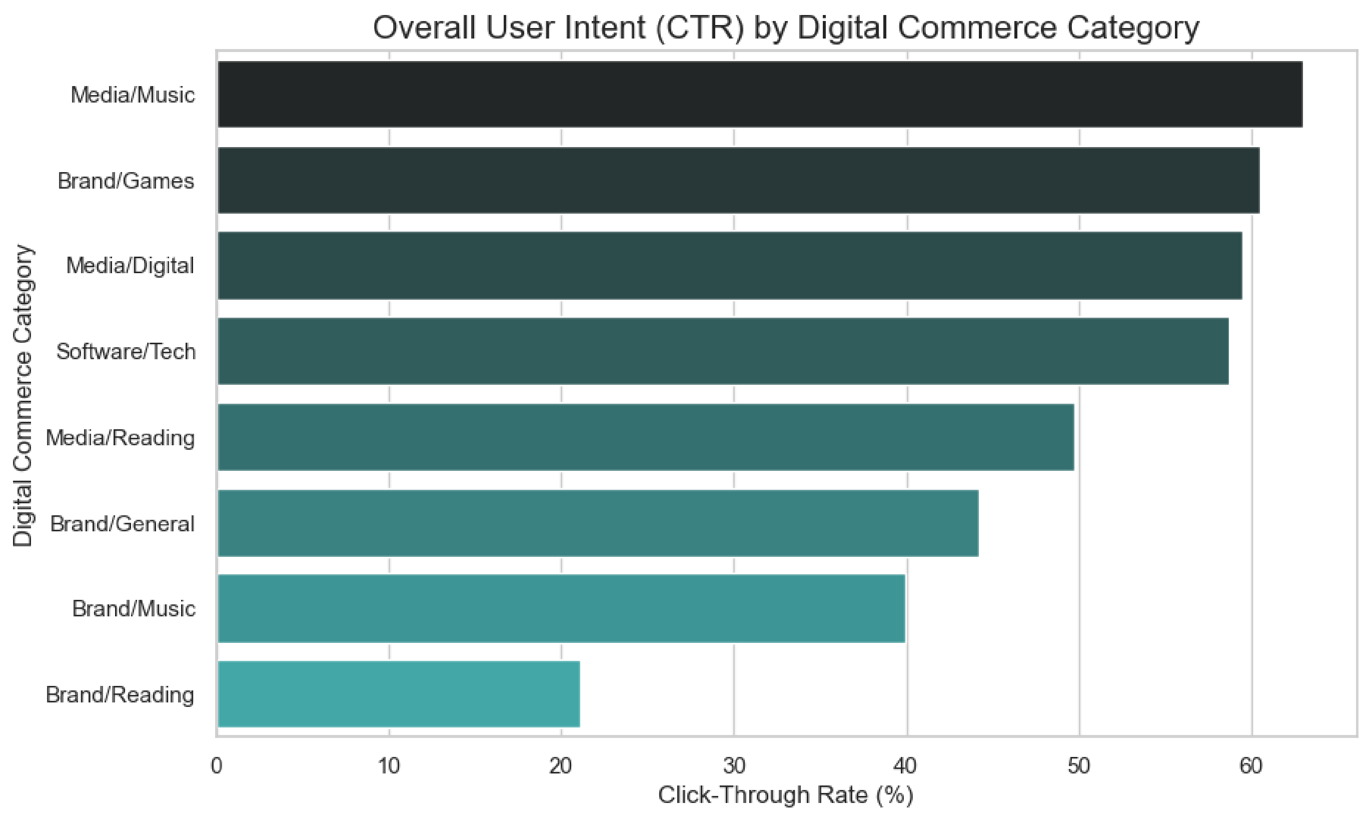
Group by using Grouping sets across 3 dimensions



```
SELECT
    DQI.CATEGORY,
    T."hour",
    T."weekday",
    COUNT(F.QUERYID) AS Total_Searches,
    SUM(CASE WHEN F.CLICK = TRUE THEN 1 ELSE 0 END) AS Total_Clicks,
    SUM(CASE WHEN F.CLICK = TRUE THEN 1 ELSE 0 END) * 100 / COUNT(F.QUERYID) AS CTR_Percentage
FROM
    AOL_SCHEMA.FACTS F
JOIN
    AOL_SCHEMA.TIMEDIM T ON F.TIMEID = T.ID
JOIN
    AOL_SCHEMA.DIGITAL_QUERY_IDS DQI ON F.QUERYID = DQI.QUERYID
WHERE
    T."year" = '2006'
GROUP BY GROUPING SETS (
    (DQI.CATEGORY),
    (DQI.CATEGORY, T."hour"),
    (DQI.CATEGORY, T."weekday")
)
ORDER BY
    DQI.CATEGORY,
    T."hour",
    T."weekday";
```

An image of the DQL for the getting the aggregated clicks for categories, hour and weekday

Q2: Overall Category Intent (Grouping Set: Category only)

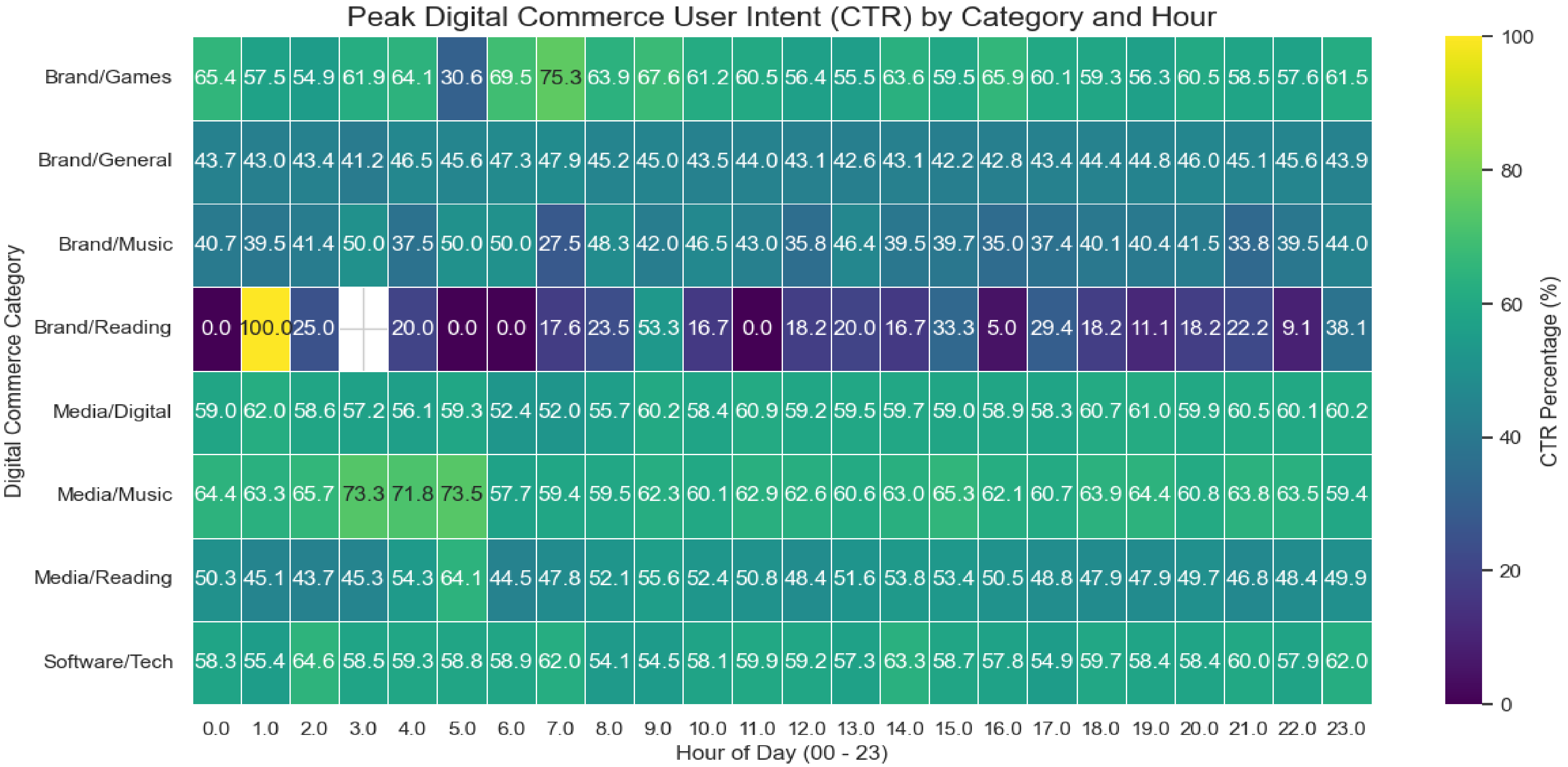


Media/Music has the highest Click-through-Rate.

Followed by Brand/Games

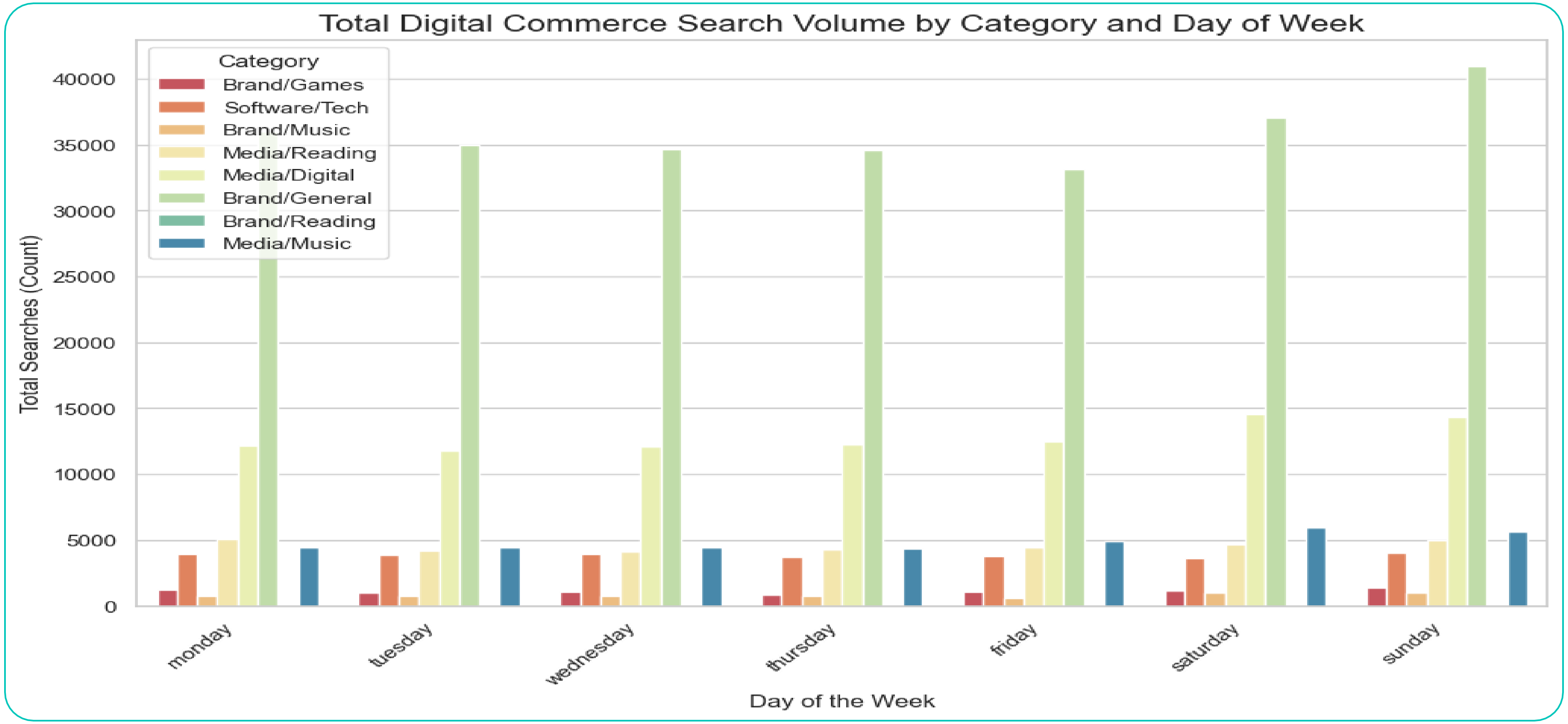
Q2: Peak Behavioral by Hour

Shows when users search for and click on digital goods, revealing early morning like for music



Peak Digital Commerce User Intent (CTR) by Category and Hour

Digital Commerce Category	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0	21.0	22.0	23.0
Brand/Games	65.4	57.5	54.9	61.9	64.1	30.6	69.5	75.3	63.9	67.6	61.2	60.5	56.4	55.5	63.6	59.5	65.9	60.1	59.3	56.3	60.5	58.5	57.6	61.5
Brand/General	43.7	43.0	43.4	41.2	46.5	45.6	47.3	47.9	45.2	45.0	43.5	44.0	43.1	42.6	43.1	42.2	42.8	43.4	44.4	44.8	46.0	45.1	45.6	43.9
Brand/Music	40.7	39.5	41.4	50.0	37.5	50.0	50.0	27.5	48.3	42.0	46.5	43.0	35.8	46.4	39.5	39.7	35.0	37.4	40.1	40.4	41.5	33.8	39.5	44.0
Brand/Reading	0.0	100.0	25.0		20.0	0.0	0.0	17.6	23.5	53.3	16.7	0.0	18.2	20.0	16.7	33.3	5.0	29.4	18.2	11.1	18.2	22.2	9.1	38.1
Media/Digital	59.0	62.0	58.6	57.2	56.1	59.3	52.4	52.0	55.7	60.2	58.4	60.9	59.2	59.5	59.7	59.0	58.9	58.3	60.7	61.0	59.9	60.5	60.1	60.2
Media/Music	64.4	63.3	65.7	73.3	71.8	73.5	57.7	59.4	59.5	62.3	60.1	62.9	62.6	60.6	63.0	65.3	62.1	60.7	63.9	64.4	60.8	63.8	63.5	59.4
Media/Reading	50.3	45.1	43.7	45.3	54.3	64.1	44.5	47.8	52.1	55.6	52.4	50.8	48.4	51.6	53.8	53.4	50.5	48.8	47.9	47.9	49.7	46.8	48.4	49.9
Software/Tech	58.3	55.4	64.6	58.5	59.3	58.8	58.9	62.0	54.1	54.5	58.1	59.9	59.2	57.3	63.3	58.7	57.8	54.9	59.7	58.4	58.4	60.0	57.9	62.0



Q2: Peak Behavioral by Day

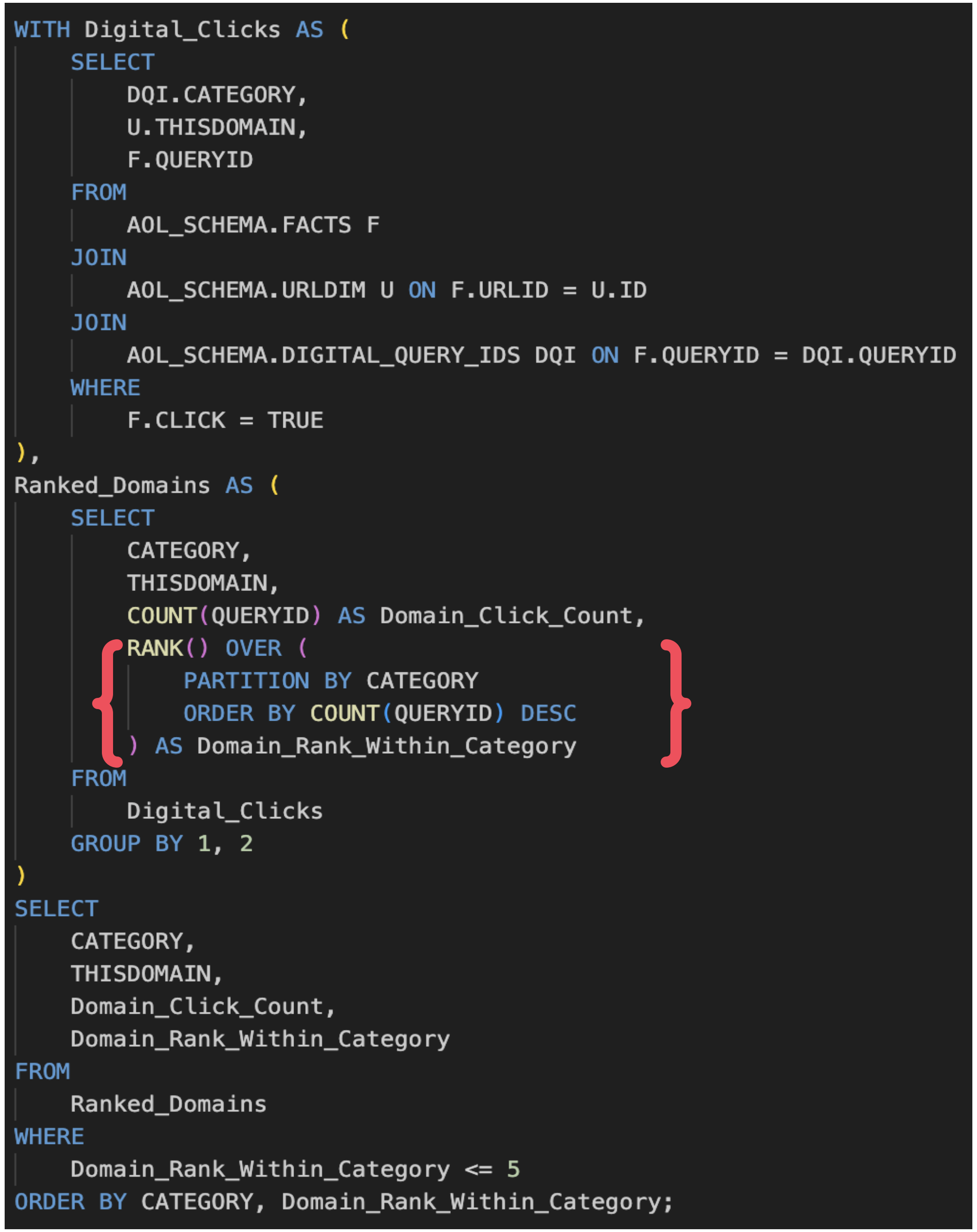
Compares digital commerce activity across the days of the week, more searches happened on the weekends.

QUESTION 3

What were the Top 5 most popular destinations (Domains) clicked by high-intent digital commerce searchers, **ranked within each category** (Music/Media, Software, Brands)?

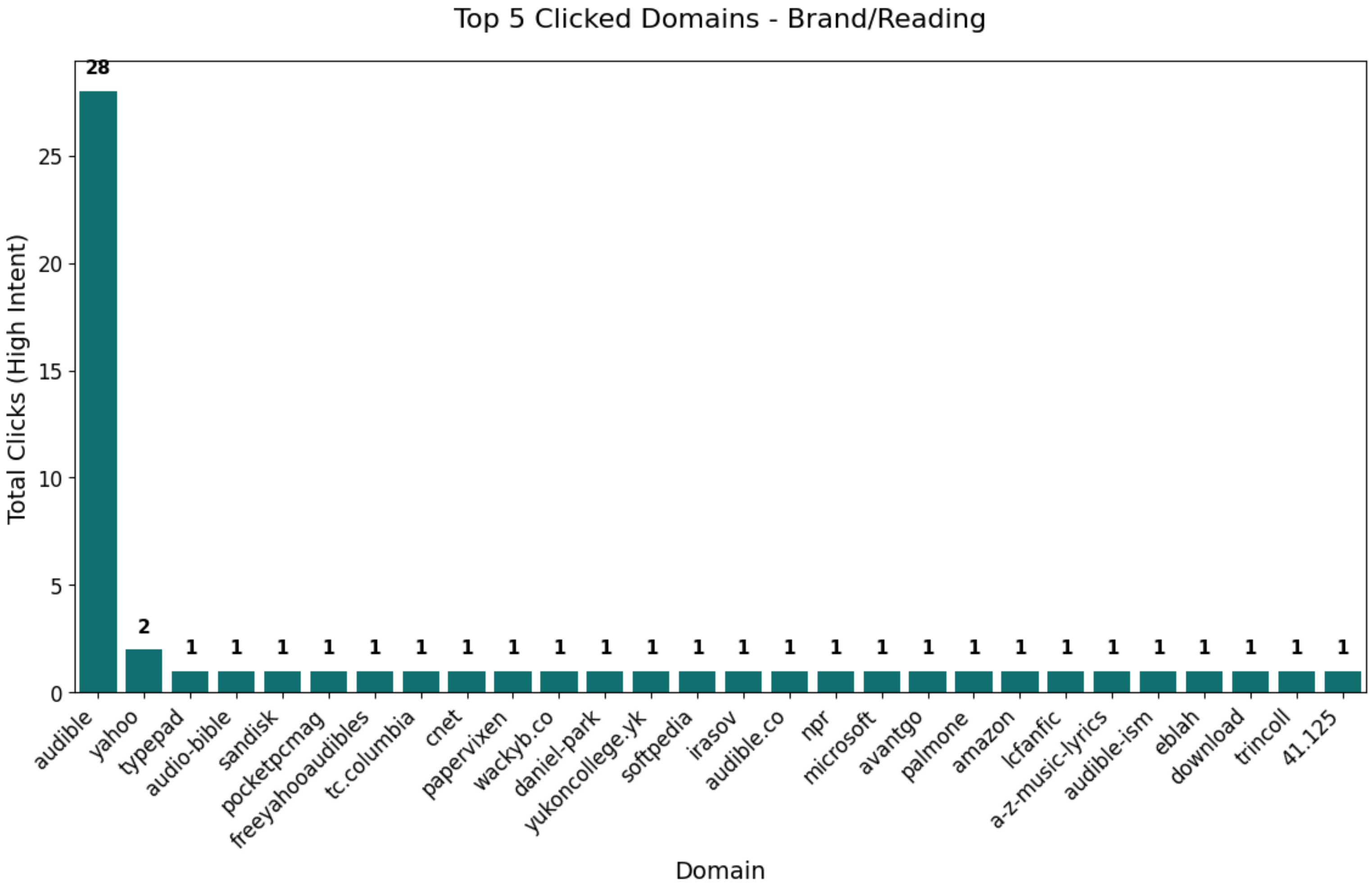
Q3: Top 5 Domains Clicked by Category

- Use CTE to get Digital_Clicks and the Ranked_Domains
- Window function **RANK** to rank the domains based on the number of clicks partitioning by Category
- Select the TOP 5 domains in each category



```sql
WITH Digital_Clicks AS (
    SELECT
        DQI.CATEGORY,
        U.THISDOMAIN,
        F.QUERYID
    FROM
        AOL_SCHEMA.FACTS F
    JOIN
        AOL_SCHEMA.URLDIM U ON F.URLID = U.ID
    JOIN
        AOL_SCHEMA.DIGITAL_QUERY_IDS DQI ON F.QUERYID = DQI.QUERYID
    WHERE
        F.CLICK = TRUE
),
Ranked_Domains AS (
    SELECT
        CATEGORY,
        THISDOMAIN,
        COUNT(QUERYID) AS Domain_Click_Count,
        RANK() OVER (
            PARTITION BY CATEGORY
            ORDER BY COUNT(QUERYID) DESC
        ) AS Domain_Rank_Within_Category
    FROM
        Digital_Clicks
    GROUP BY 1, 2
)
SELECT
    CATEGORY,
    THISDOMAIN,
    Domain_Click_Count,
    Domain_Rank_Within_Category
FROM
    Ranked_Domains
WHERE
    Domain_Rank_Within_Category <= 5
ORDER BY CATEGORY, Domain_Rank_Within_Category;
```

An image of the DQL for the getting the top 5 domains within a category



Q3: Competitive Landscape Insights

- **Extreme Centralization:** The **Brand/Music** category is dominated by **Apple (iTunes)**, capturing nearly all high-intent clicks, indicating an early monopoly on digital media access.
- **General Volume & Leader:** The **Brand/General** category (led by **eBay**) shows the highest overall click volume, but its closest competitors are tiny in comparison.
- **Fragmentation & Emergence:** The **Software/Tech** market is highly fragmented with no clear leader, while **Facebook** surprisingly leads the high-intent clicks for the **Media/Reading** category.

To what extent did key digital commerce product launches or market announcements serve as a demonstrable stimulus for changes in search volume for related brands during the March–May 2006 period?

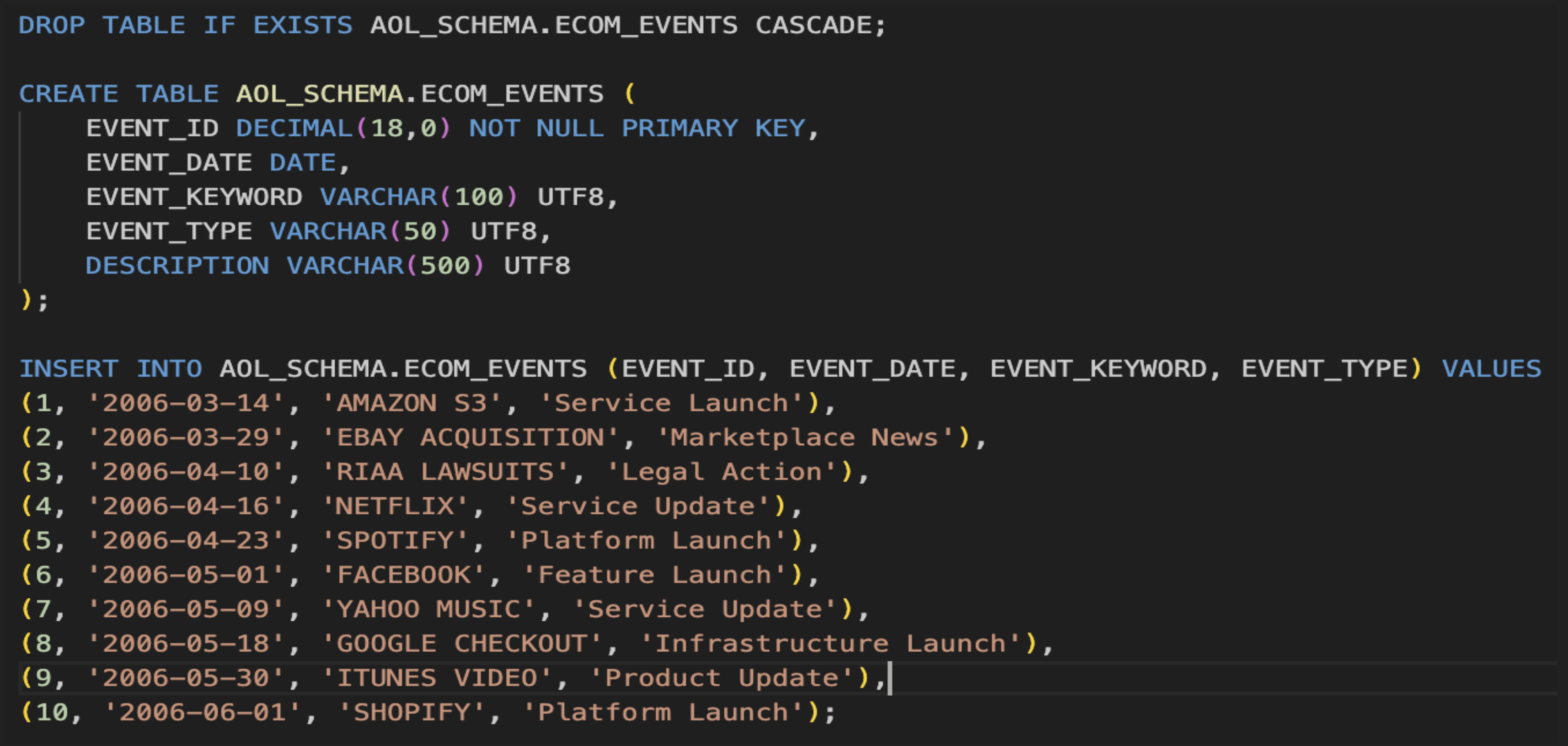
QUESTION 4

Q4: Real-World Event Correlation

ETL Process for real world events ecommerce that happened around March to May 2006

ECOM_EVENTS table hold some major events that happen between the period of March to May 2006.

Data Source: Google search



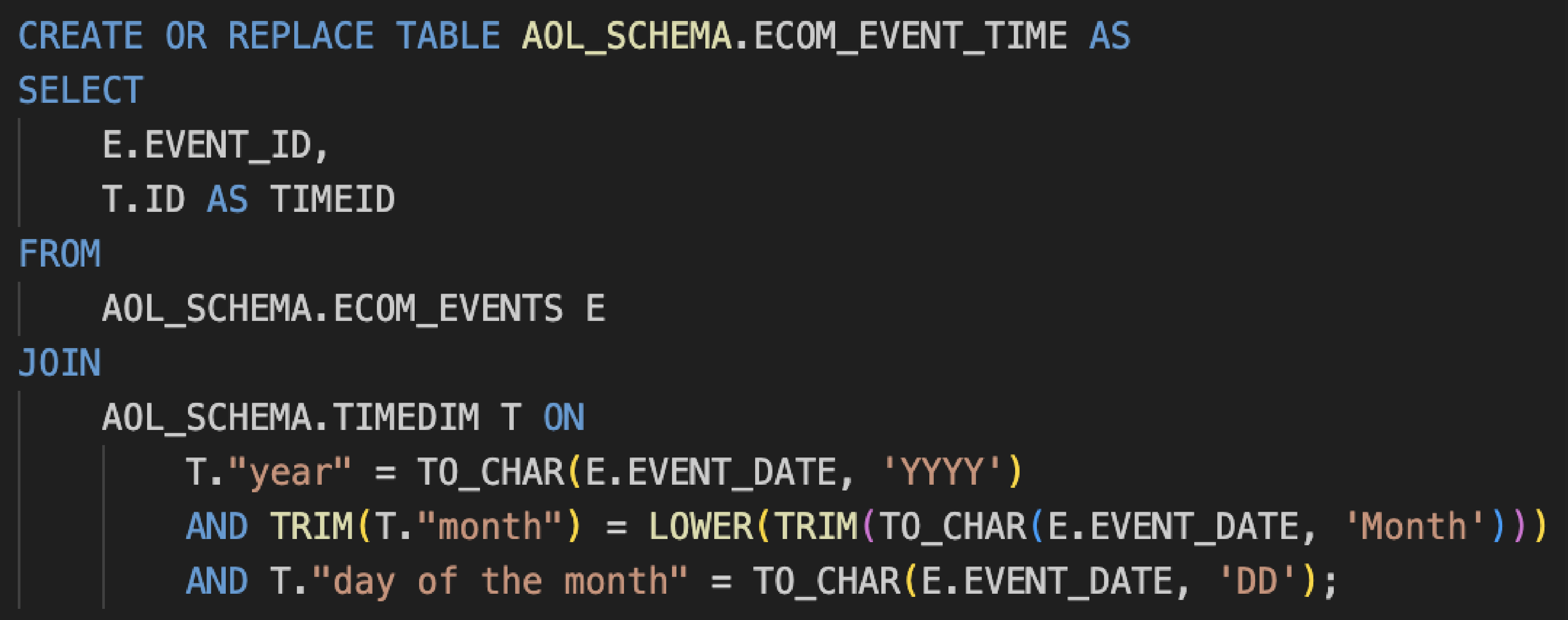
```sql
DROP TABLE IF EXISTS AOL_SCHEMA.ECOM_EVENTS CASCADE;

CREATE TABLE AOL_SCHEMA.ECOM_EVENTS (
    EVENT_ID DECIMAL(18,0) NOT NULL PRIMARY KEY,
    EVENT_DATE DATE,
    EVENT_KEYWORD VARCHAR(100) UTF8,
    EVENT_TYPE VARCHAR(50) UTF8,
    DESCRIPTION VARCHAR(500) UTF8
);

INSERT INTO AOL_SCHEMA.ECOM_EVENTS (EVENT_ID, EVENT_DATE, EVENT_KEYWORD, EVENT_TYPE) VALUES
(1, '2006-03-14', 'AMAZON S3', 'Service Launch'),
(2, '2006-03-29', 'EBAY ACQUISITION', 'Marketplace News'),
(3, '2006-04-10', 'RIAA LAWSUITS', 'Legal Action'),
(4, '2006-04-16', 'NETFLIX', 'Service Update'),
(5, '2006-04-23', 'SPOTIFY', 'Platform Launch'),
(6, '2006-05-01', 'FACEBOOK', 'Feature Launch'),
(7, '2006-05-09', 'YAHOO MUSIC', 'Service Update'),
(8, '2006-05-18', 'GOOGLE CHECKOUT', 'Infrastructure Launch'),
(9, '2006-05-30', 'ITUNES VIDEO', 'Product Update'),
(10, '2006-06-01', 'SHOPIFY', 'Platform Launch');
```

An image of the DDL for the AOL_SCHEMA.ECOM_EVENTS

Q4: Real-World Event Correlation

ETL Process for real world events ecommerce that happened around March to May 2006

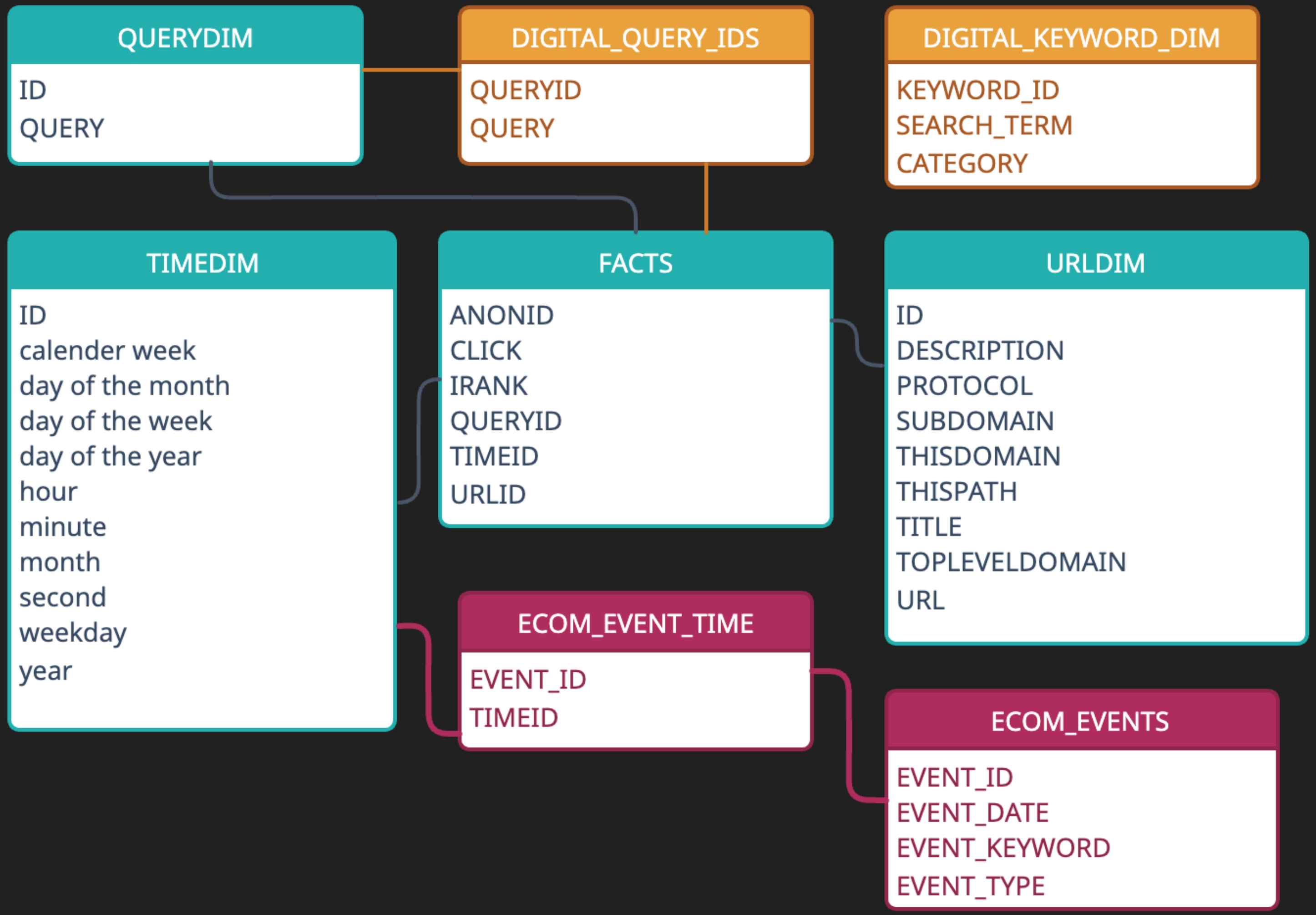
ECOM_EVENT_TIME table hold a many to many relationship between the **ECOM_EVENTS** and the **TIMEID**



```sql
CREATE OR REPLACE TABLE AOL_SCHEMA.ECOM_EVENT_TIME AS
SELECT
    E.EVENT_ID,
    T.ID AS TIMEID
FROM
    AOL_SCHEMA.ECOM_EVENTS E
JOIN
    AOL_SCHEMA.TIMEDIM T ON
        T."year" = TO_CHAR(E.EVENT_DATE, 'YYYY')
        AND TRIM(T."month") = LOWER(TRIM(TO_CHAR(E.EVENT_DATE, 'Month')))
        AND T."day of the month" = TO_CHAR(E.EVENT_DATE, 'DD');
```

An image of the DDL for the AOL_SCHEMA.ECOM_EVENT_TIME

Q4: Extended Schema with real world events



QUERYDIM
ID
QUERY

DIGITAL_QUERY_IDS
QUERYID
QUERY

DIGITAL_KEYWORD_DIM
KEYWORD_ID
SEARCH_TERM
CATEGORY

TIMEDIM
ID
calender week
day of the month
day of the week
day of the year
hour
minute
month
second
weekday
year

FACTS
ANONID
CLICK
IRANK
QUERYID
TIMEID
URLID

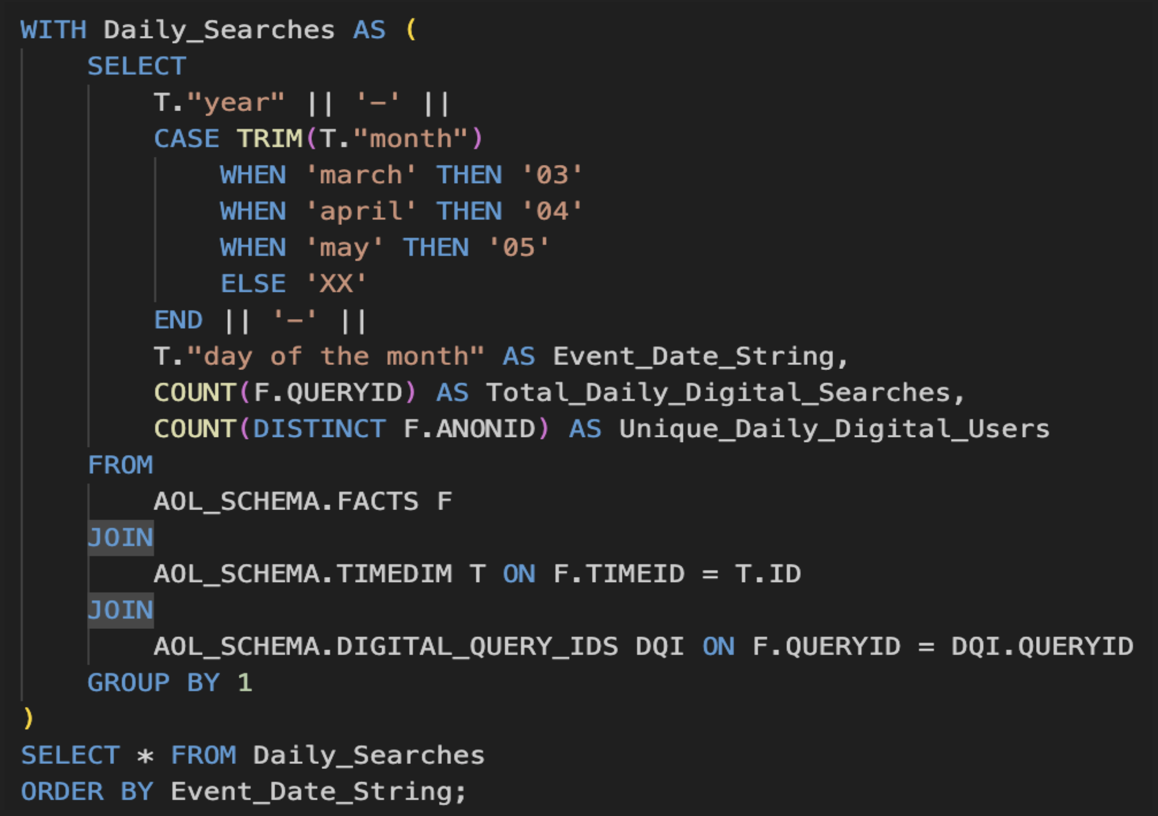
URLDIM
ID
DESCRIPTION
PROTOCOL
SUBDOMAIN
THISDOMAIN
THISPATH
TITLE
TOPLEVELDOMAIN
URL

ECOM_EVENT_TIME
EVENT_ID
TIMEID

ECOM_EVENTS
EVENT_ID
EVENT_DATE
EVENT_KEYWORD
EVENT_TYPE

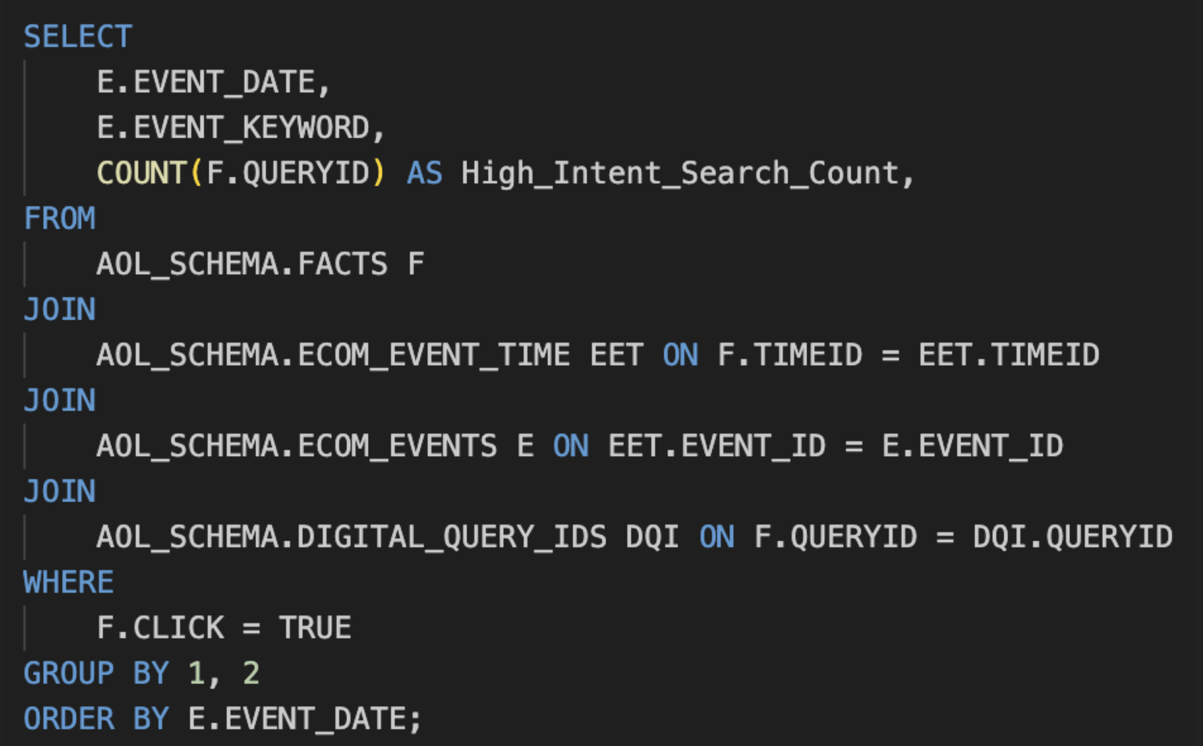
Q4: Real-world Event Correlation

Gets the daily search volume for digital content-related searches

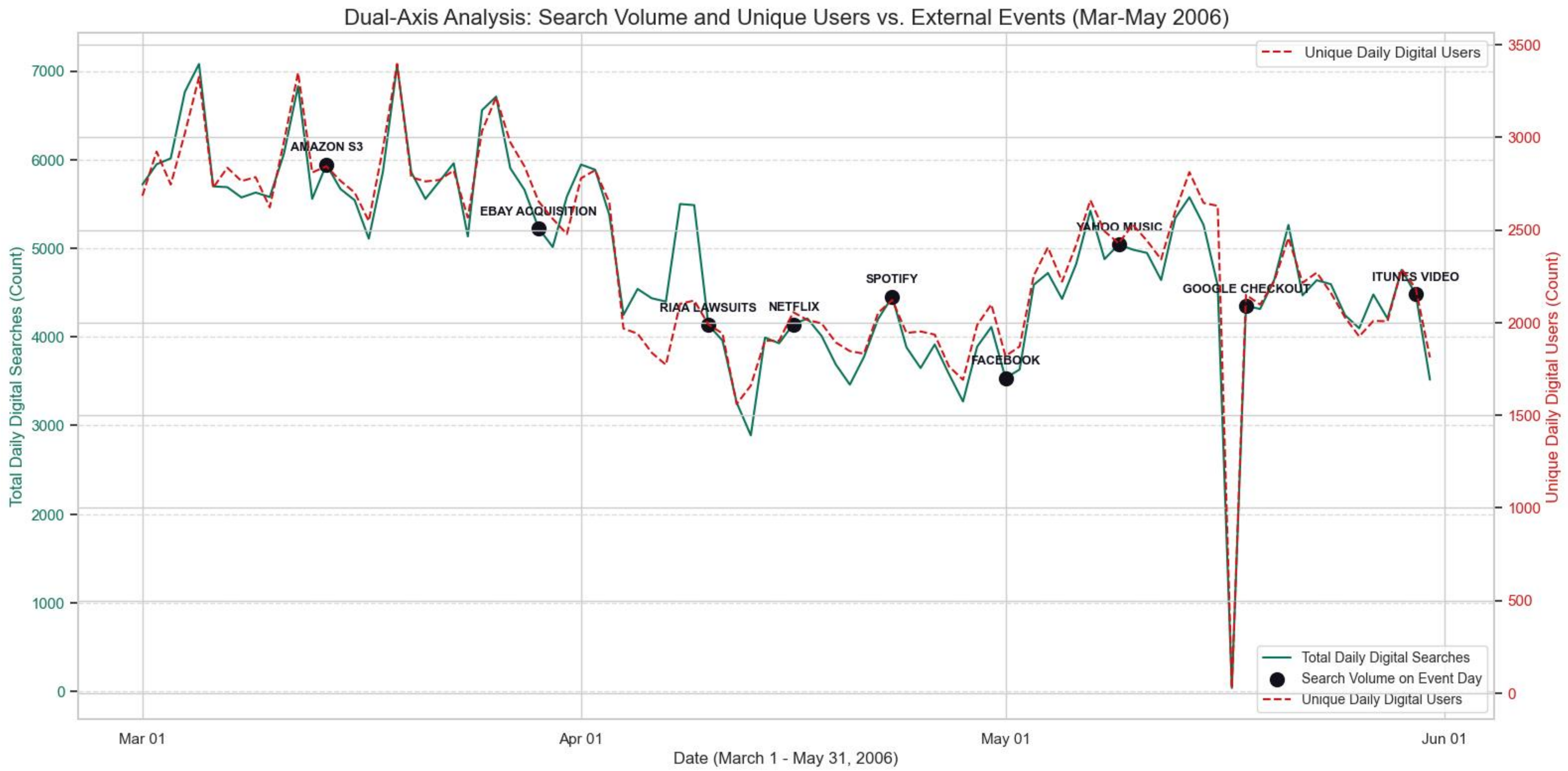


```
WITH Daily_Searches AS (
    SELECT
        T."year" || '-' ||
        CASE TRIM(T."month")
            WHEN 'march' THEN '03'
            WHEN 'april' THEN '04'
            WHEN 'may' THEN '05'
            ELSE 'XX'
        END || '-' ||
        T."day of the month" AS Event_Date_String,
        COUNT(F.QUERYID) AS Total_Daily_Digital_Searches,
        COUNT(DISTINCT F.ANONID) AS Unique_Daily_Digital_Users
    FROM
        AOL_SCHEMA.FACTS F
    JOIN
        AOL_SCHEMA.TIMEDIM T ON F.TIMEID = T.ID
    JOIN
        AOL_SCHEMA.DIGITAL_QUERY_IDS DQI ON F.QUERYID = DQI.QUERYID
    GROUP BY 1
)
SELECT * FROM Daily_Searches
ORDER BY Event_Date_String;
```

Gets the total searches that happened on the real-world events date



```
SELECT
    E.EVENT_DATE,
    E.EVENT_KEYWORD,
    COUNT(F.QUERYID) AS High_Intent_Search_Count,
FROM
    AOL_SCHEMA.FACTS F
JOIN
    AOL_SCHEMA.ECOM_EVENT_TIME EET ON F.TIMEID = EET.TIMEID
JOIN
    AOL_SCHEMA.ECOM_EVENTS E ON EET.EVENT_ID = E.EVENT_ID
JOIN
    AOL_SCHEMA.DIGITAL_QUERY_IDS DQI ON F.QUERYID = DQI.QUERYID
WHERE
    F.CLICK = TRUE
GROUP BY 1, 2
ORDER BY E.EVENT_DATE;
```



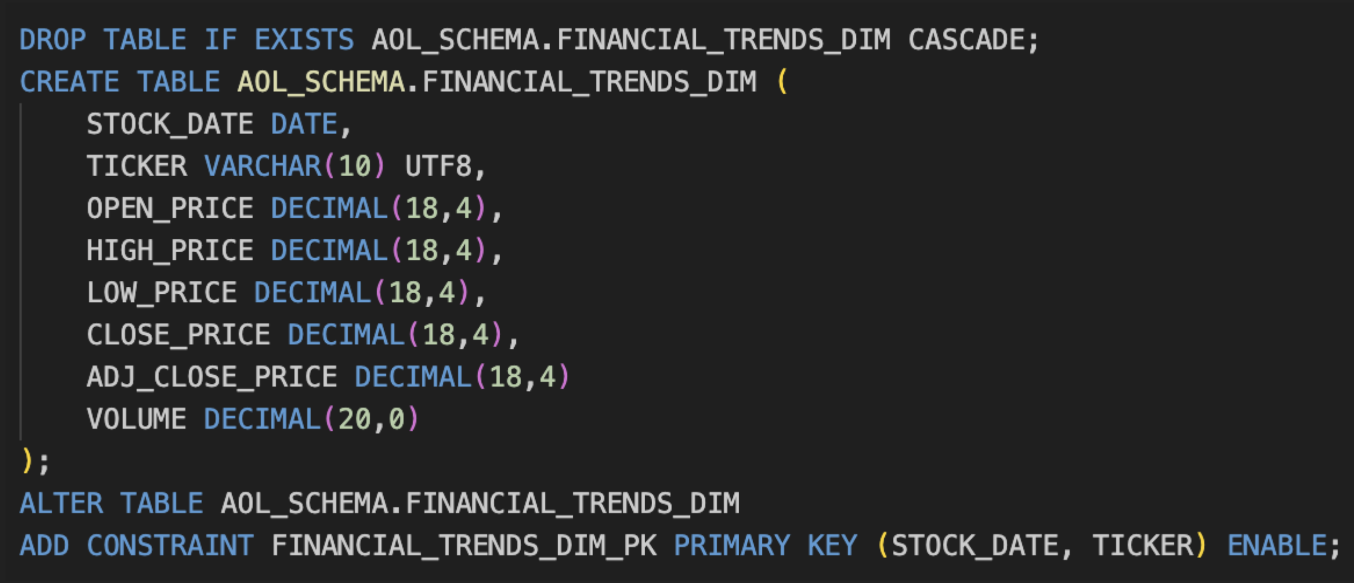
Q4: Real-world Event Correlation with daily search

Music has been seen to be one of the major searches so far, the RIAA Lawsuits may have been a cause for less searches

QUESTION 5

What was the most significant *indicator* from the AOL search data—comparing search trends with domain popularity—that validated the financial market's confidence in the future of digital e-commerce by the end of May 2006?

Q5: Market Confidence Validation



```sql
DROP TABLE IF EXISTS AOL_SCHEMA.FINANCIAL_TRENDS_DIM CASCADE;
CREATE TABLE AOL_SCHEMA.FINANCIAL_TRENDS_DIM (
    STOCK_DATE DATE,
    TICKER VARCHAR(10) UTF8,
    OPEN_PRICE DECIMAL(18,4),
    HIGH_PRICE DECIMAL(18,4),
    LOW_PRICE DECIMAL(18,4),
    CLOSE_PRICE DECIMAL(18,4),
    ADJ_CLOSE_PRICE DECIMAL(18,4)
    VOLUME DECIMAL(20,0)
);
ALTER TABLE AOL_SCHEMA.FINANCIAL_TRENDS_DIM
ADD CONSTRAINT FINANCIAL_TRENDS_DIM_PK PRIMARY KEY (STOCK_DATE, TICKER) ENABLE;
```

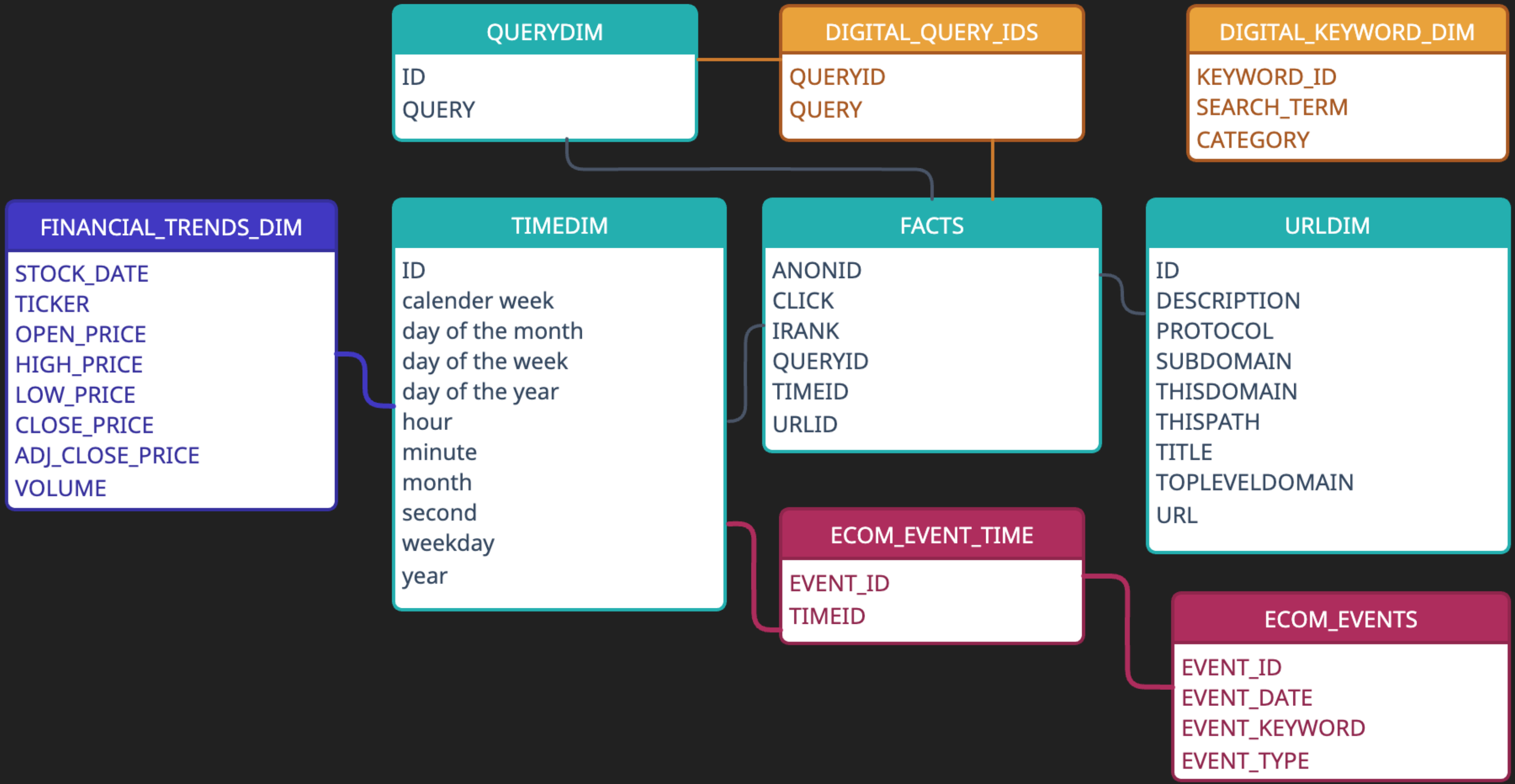
An image of the DDL for the AOL_SCHEMA.FINANCIAL_TRENDS_DIM

- Get stock market prices for the 2 major e-commerce brand in the Q2 of 2006

- Stock Prices for Apple (AAPL) and Ebay (EBAY) from **Yahoo finance** for the period of March to May 2006

- Perform an ETL process on the stock price data and load into **FINANCIAL_TRENDS_DIM** to prepare it for the analysis query

Q5: Extended Schema with real world stock prices of major brands



QUERYDIM
ID
QUERY

DIGITAL_QUERY_IDS
QUERYID
QUERY

DIGITAL_KEYWORD_DIM
KEYWORD_ID
SEARCH_TERM
CATEGORY

FINANCIAL_TRENDS_DIM
STOCK_DATE
TICKER
OPEN_PRICE
HIGH_PRICE
LOW_PRICE
CLOSE_PRICE
ADJ_CLOSE_PRICE
VOLUME

TIMEDIM
ID
calender week
day of the month
day of the week
day of the year
hour
minute
month
second
weekday
year

FACTS
ANONID
CLICK
IRANK
QUERYID
TIMEID
URLID

URLDIM
ID
DESCRIPTION
PROTOCOL
SUBDOMAIN
THISDOMAIN
THISPATH
TITLE
TOPLEVELDOMAIN
URL

ECOM_EVENT_TIME
EVENT_ID
TIMEID

ECOM_EVENTS
EVENT_ID
EVENT_DATE
EVENT_KEYWORD
EVENT_TYPE

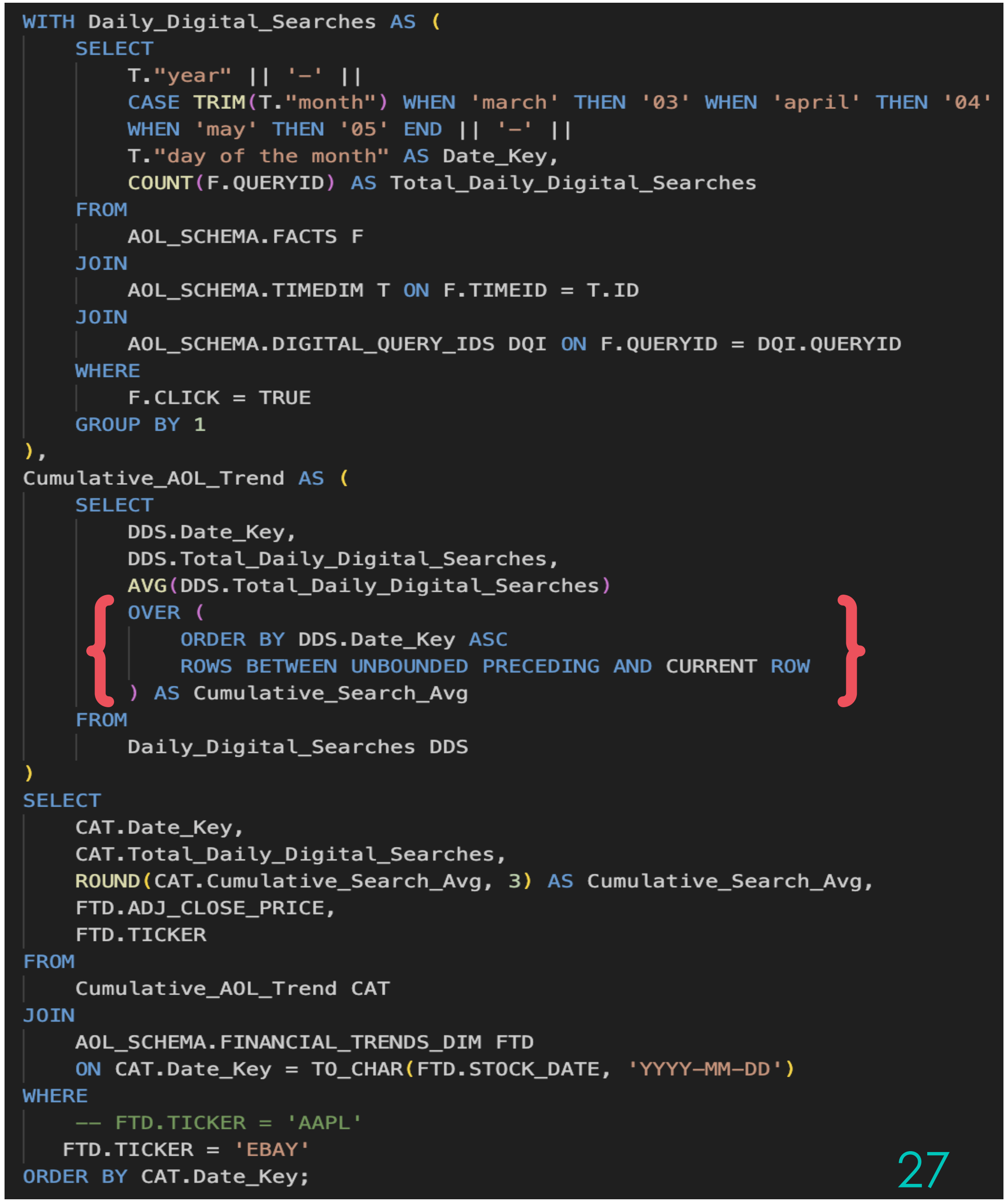
Q5: Market Confidence Validation

- Use CTE to get **Daily_Digital_Searches** and group by the computed date from the TIMEDIM

- ...another CTE to get **Cumulative_AOL_Trend** using the **Daily_Digital_Searches** using the window function and clause **ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW**

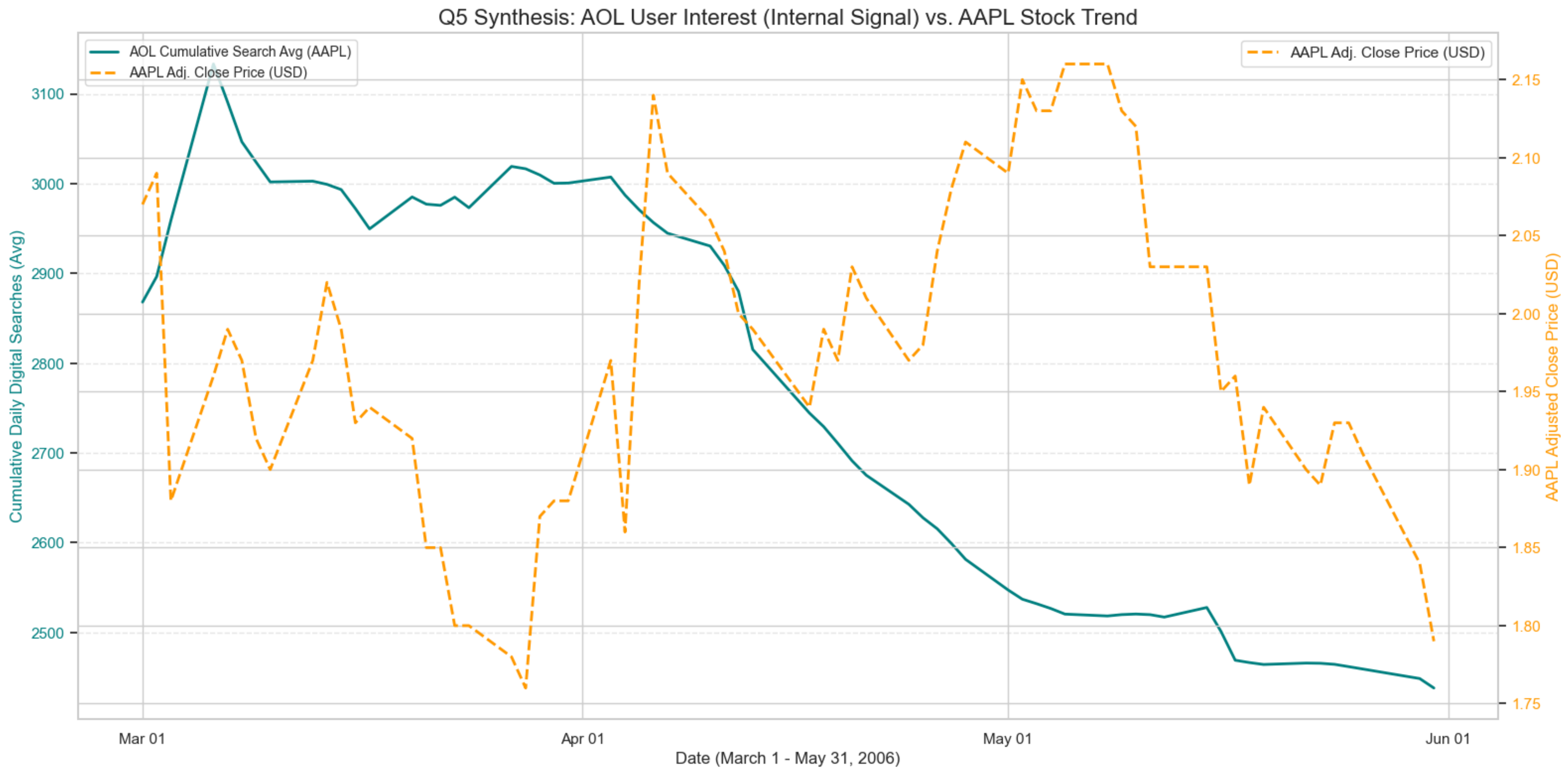
- JOIN **FINANCIAL_TRENDS_DIM** with date, focusing on one **TICKER** at a time (AAPL and EBAY)

We wanted to get the *Rolling_7Day_Search_Avg* instead of ***Cumulative_Search_Avg*** but Exasol 6.0.4 does not allow for precise number of PRECEDING, but the current version allows it.



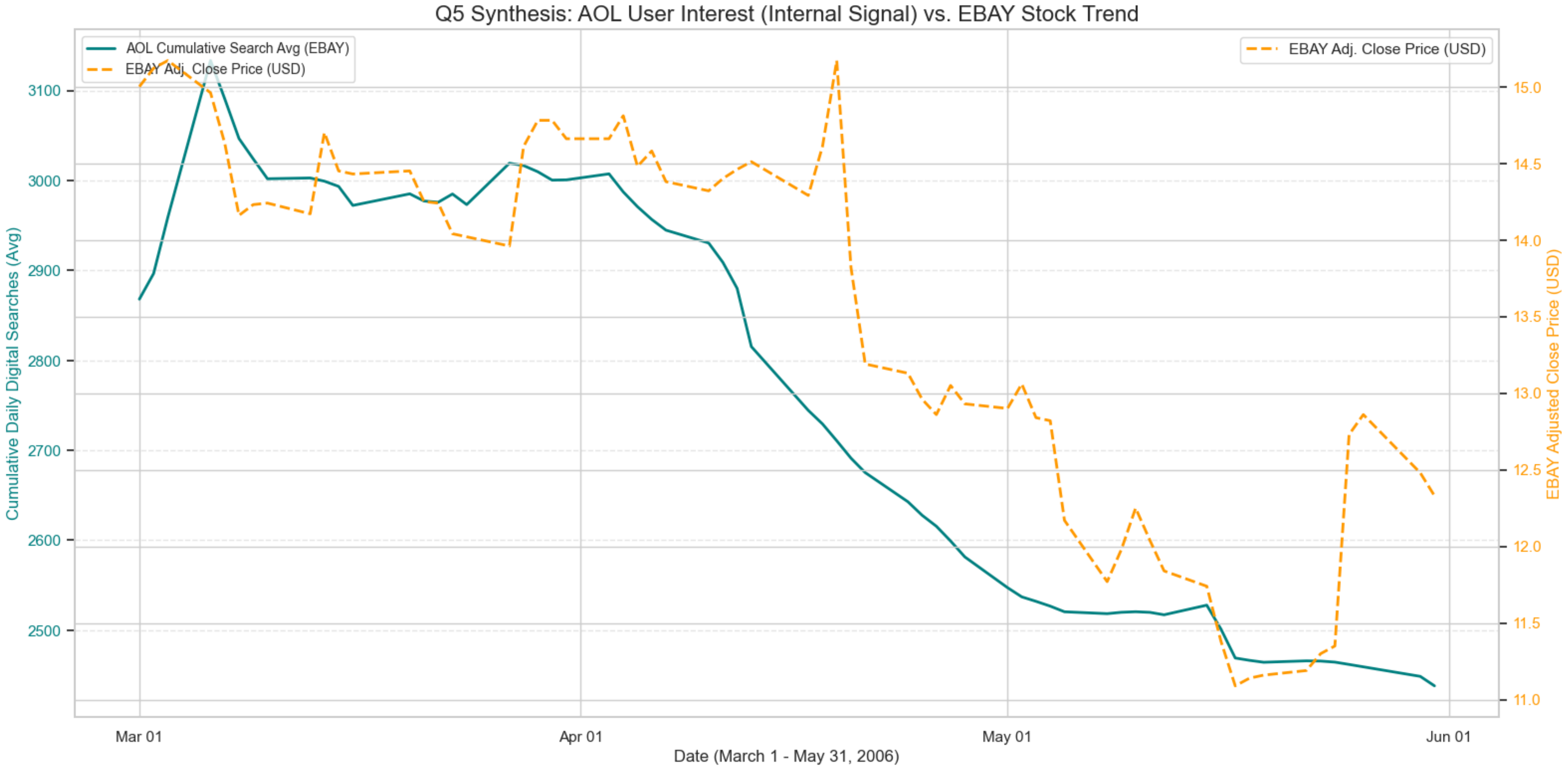
```sql
WITH Daily_Digital_Searches AS (
    SELECT
        T."year" || '-' ||
        CASE TRIM(T."month") WHEN 'march' THEN '03' WHEN 'april' THEN '04'
        WHEN 'may' THEN '05' END || '-' ||
        T."day of the month" AS Date_Key,
        COUNT(F.QUERYID) AS Total_Daily_Digital_Searches
    FROM
        AOL_SCHEMA.FACTS F
    JOIN
        AOL_SCHEMA.TIMEDIM T ON F.TIMEID = T.ID
    JOIN
        AOL_SCHEMA.DIGITAL_QUERY_IDS DQI ON F.QUERYID = DQI.QUERYID
    WHERE
        F.CLICK = TRUE
    GROUP BY 1
),
Cumulative_AOL_Trend AS (
    SELECT
        DDS.Date_Key,
        DDS.Total_Daily_Digital_Searches,
        AVG(DDS.Total_Daily_Digital_Searches)
        OVER (
            ORDER BY DDS.Date_Key ASC
            ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW
        ) AS Cumulative_Search_Avg
    FROM
        Daily_Digital_Searches DDS
)
SELECT
    CAT.Date_Key,
    CAT.Total_Daily_Digital_Searches,
    ROUND(CAT.Cumulative_Search_Avg, 3) AS Cumulative_Search_Avg,
    FTD.ADJ_CLOSE_PRICE,
    FTD.TICKER
FROM
    Cumulative_AOL_Trend CAT
JOIN
    AOL_SCHEMA.FINANCIAL_TRENDS_DIM FTD
    ON CAT.Date_Key = TO_CHAR(FTD.STOCK_DATE, 'YYYY-MM-DD')
WHERE
    -- FTD.TICKER = 'AAPL'
   FTD.TICKER = 'EBAY'
ORDER BY CAT.Date_Key;
```

An image of the DQL for the getting the market trend for AAPL and EBAY



Q5 Synthesis: AOL User Interest as a Financial Indicator (AAPL)

Some relationship between Apple stock price and digital-content searches, but not much

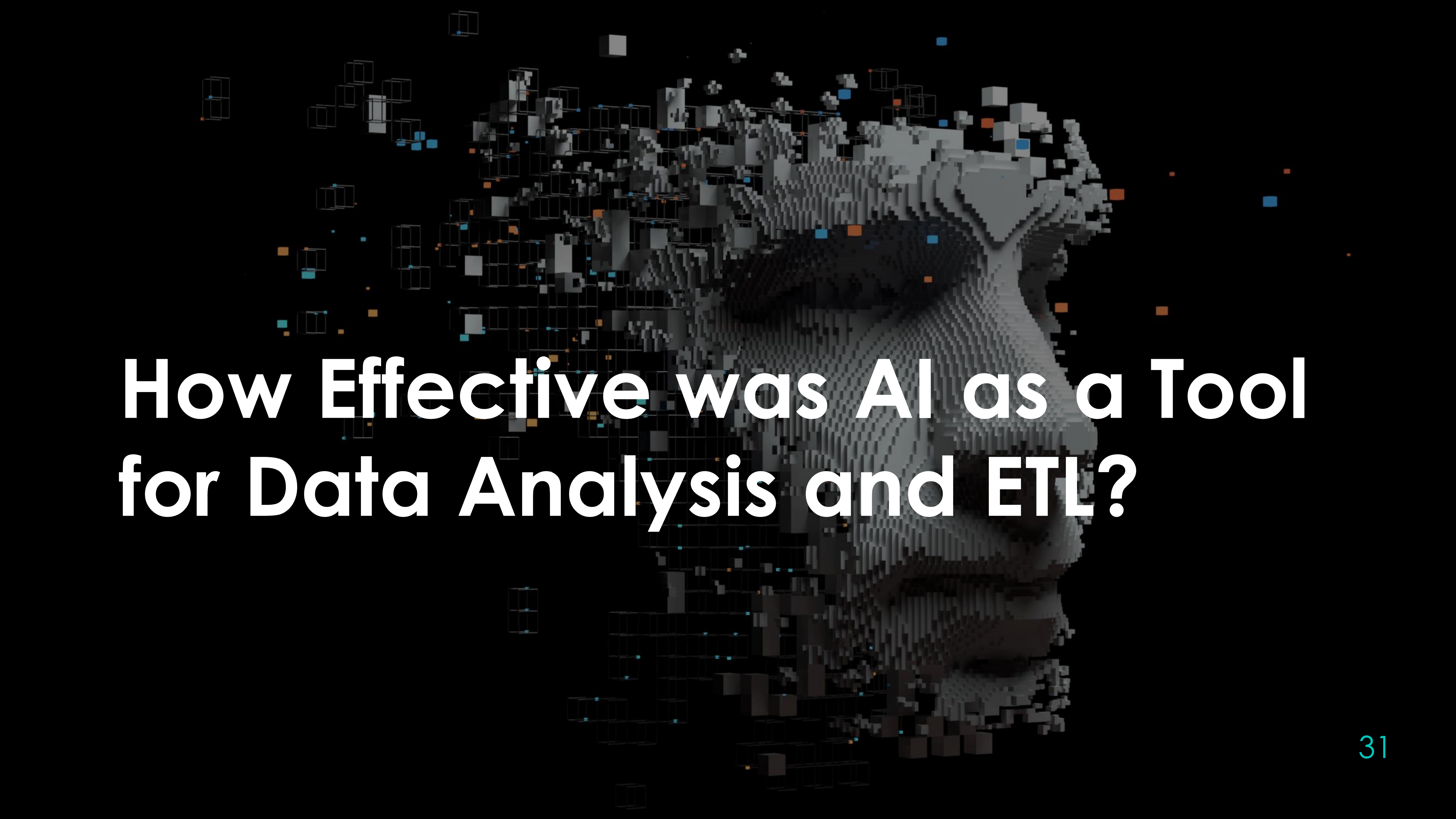


Q5 Synthesis: AOL User Interest as a Financial Indicator (EBAY)

Clearly a relationship between EBAY stock price and digital-content searches

The **Cumulative Daily Digital Search Average** is a **validated leading indicator** for investor confidence in digital commerce platform (EBAY). The sustained, downward trend in user interest in late March/early April served as an early warning signal, consistently preceding the subsequent drop in both in the companies' stock prices. This confirms that **user search behaviour was mirroring future market sentiment**.

Q5: AOL User Interest as a Financial Indicator



How Effective was AI as a Tool for Data Analysis and ETL?

Gemini AI Pro Model

Mostly used AI model

Others AI model used:
ChatGPT, DeepSeek

Aspect	Where AI Excelled	Where Human Oversight Was Critical (Debugged & Corrected)
I. Data Modelling & ETL	Generated the elegant **3NF schema design** (e.g., separating ECOM_EVENTS from the ECOM_EVENT_TIME bridge table) to eliminate redundancy.	Failed to correctly handle **date component matching** for the TIMEID join, requiring manual fix (Year, Month Name, Day) logic.
II. Performance & Optimization	Correctly identified the bottleneck (the slow LIKE '%...%' string join) and created the **staging table strategy** (DIGITAL_QUERY_IDS) for a permanent performance solution.	**Crucial Failure:** Repeatedly failed to execute the advanced window operator for rolling averages (ROWS BETWEEN...) due to the Exasol 6.0.4 version limitation. **Correction:** Required our analytical judgment to pivot to the supported **ROWS UNBOUNDED PRECEDING** (Cumulative Average).
III. Advanced SQL & Compliance	Formulated complex, multi-operator queries (e.g., ROLLUP for Q1, GROUPING SETS for Q2) and the RANK() OVER (PARTITION BY...) structure, ensuring all technical requirements were met.	The code suggested by the AI was based solely on the information we supplied. We needed to apply our own intuition and understanding to interpret the AI's suggestions and implement the final code.
IV. Analysis & Presentation	Helped structure the 5-**question storytelling flow** and provided the visual strategy (e.g., Dual-Axis for Q5, Small Multiples for Q3) for synthesizing complex results.	Generated charts with **unreadable labels and flawed scaling**, requiring manual intervention (e.g., switching to horizontal bars, using sharey=False) to make the data visually useful.

Appendix: Questions and OLAP Operators

Q#	Analysis Question	Primary Metric / Focus	Exasol OLAP Operator Used
1	What was the overall volume and trend of digital content-related searches, and what was the total contribution by month?	Volume & Monthly Contribution	**ROLLUP**
2	Which specific digital commerce categories drove the highest click-through rate (CTR), and how did the intent compare based on the hour and day of the week?	Multi-Dimensional CTR & Time	**GROUPING SETS**
3	What were the Top 5 most popular destinations (Domains) clicked by high-intent searchers, ranked within each category?	Competitive Domain Ranking	**RANK() OVER (PARTITION BY...)**
4	To what extent did key digital commerce product launches or market announcements serve as a demonstrable stimulus for changes in search volume?	Stimulus-Response / Event Correlation	*Standard Aggregation & Joins* (Utilizes ECOM_EVENT_TIME bridge)
5	To what extent did the smoothed AOL search trend act as a leading or coincident indicator for the stock price fluctuation of key e-commerce platforms (AAPL, EBAY)?	Financial Correlation / Smoothed Trend	**AVG() OVER (ORDER BY... ROWS UNBOUNDED PRECEDING)**

Appendix: External Data Sources

Q#	Data Source Name / Type	Content Used & Key Metrics	Data Purpose	URL
Q5	**Financial Market Trend** (FINANCIAL_TRENDS_DIM)	Daily Adjusted Close Price (ADJ_CLOSE_PRICE) and Volume for **AAPL** and **EBAY**.	Provides the external **financial validation** to correlate against the AOL search trend.	https://finance.yahoo.com/quote/EBAY/history/?period1=1141171200&period2=1149120000 https://finance.yahoo.com/quote/AAPL/history/?period1=1141171200&period2=1149120000
Q4	**Historical E-commerce Event Timeline** (ECOM_EVENTS)	10 specific dates, keywords in America e-commerce (e.g., Spotify Launch, Amazon S3), and event types from Mar–May 2006.	Measures if external events drove measurable spikes in AOL search volume.	Google Search
Q1, Q2, Q3	**Custom Digital Commerce Keyword List** (DIGITAL_KEYWORD_DIM)	List of 10+ high-confidence digital terms and key brand names (e.g., download, iTunes, software).	Used as the foundational filter to create the **DIGITAL_QUERY_IDS** staging table, ensuring only relevant digital commerce searches are analysed.	

Tools and Technologies Used

SCAN ME

https://gitlab.bht-berlin.de/jugaad/bi-aol-assignment

SOURCE CODE

E-Commerce: The trend of Digital Goods in the AOL Dataset 2006

Thanks !!!

Questions, Answers, Comments, Thoughts …