Product Viewing Analysis Tool

1. Introduction

This directory contains scripts and files for two distinct types of product viewing analysis:

1. Continuous View Analysis (`long` script)

2. Most Viewed Analysis (`frequent` script)

2. Continuous View Analysis

2.1. The `long.js` Script

This long.js script analyzes:

- Total session duration (up to 5 minutes)

- Individual product view durations

- Page ping counts per product view
- Multiple user sessions tracking

The script produces a JSON output with two main sections, like in this sample:

```json
{
  "timing_by_user": {
    "1": {
      "first_event": "2025-04-12T07:00:00.000Z",
      "last_event": "2025-04-12T07:01:20.000Z",
      "duration_seconds": 80
    },
    "2": {
      "first_event": "2025-04-12T07:00:00.000Z",
      "last_event": "2025-04-12T07:02:20.000Z",
      "duration_seconds": 140
    },
    "3": {
      "first_event": "2025-04-12T07:00:00.000Z",
      "last_event": "2025-04-12T07:04:50.000Z",
      "duration_seconds": 290
    }
  },
  "user_views": {
    "1": {
      "5": {
        "duration_seconds": 80,
        "page_ping": 8
      }
    },
    "2": {
      "10": {
        "duration_seconds": 80,
        "page_ping": 8
      },
      "11": {
        "duration_seconds": 50,
        "page_ping": 5
      }
    },
    "3": {
      "20": {
        "duration_seconds": 290,
        "page_ping": 23
      }
    }
  }
}
```

2.2. Test Data Samples

Some JSON Lines files are provided for testing the `ContinuousViewProcessor`:

2.2.1. long-1.jsonl

- File: long-1.jsonl ← Single user, Single product
- Result: long-1.json

2.2.2. long-2.jsonl

- File: long-2.jsonl ← Single user, Multiple products
- Result: long-2.json

2.2.3. long-3.jsonl

- Input: long-3.jsonl ← Multiple users, Multiple products
- Result: long-3.json

2.3. Testing Scenarios

The sample files help test various aspects of the `ContinuousViewProcessor`:

Time Window Management
- All events within 5-minute window
- Edge case testing near window boundaries

User Behavior Patterns
- Single user viewing multiple products
- Multiple users viewing simultaneously
- Rapid product switching

Page Ping Analysis
- Regular ping intervals (10 seconds)
- Irregular ping patterns
- Missing pings handling

Product View Duration
- Long views (90+ seconds)
- Medium views (30-60 seconds)
- Short views (< 30 seconds)

3. Most Viewed Analysis

The `frequent` script ([see technical details](#)) analyzes customer product viewing patterns to identify the most frequently viewed products within a 5-minute timeframe. This analysis is used by the `MostViewedProcessor` to generate discounts for products that attract high customer interest.

3.1. Purpose

This tool helps identify which products receive the most views from customers by analyzing their behavior. It's particularly useful for:

- Understanding which products attract the most customer interest
- Identifying potential candidates for promotional campaigns
- Analyzing customer engagement patterns through view frequency

3.2. How it Works

3.2.1. Input

The script processes a JSON Lines file containing product viewing events, where each line represents a customer interaction with a product page.

3.2.2. Output

The script generates a JSON file that ranks products based on two key metrics:

- `occurrences`: How many times customers viewed the product
- `total_seconds`: Total time spent viewing the product (in seconds)

Products are sorted by:

1. Number of views (highest to lowest)
2. Total viewing time (highest to lowest) when number of views is equal

3.2.3. Time Window

The analysis focuses on the first 5 minutes of activity, starting from the timestamp of the first event. This helps identify products that generate immediate interest from customers.

3.3. Latest Most Viewed Analysis Results

```
{
  "SP Dunk Low Retro": {
    "occurrences": 12,
    "total_seconds": 1850
```

```
    },
    "SP Air Force 1 Shadow": {
      "occurrences": 10,
      "total_seconds": 1516
    },
    "Total Orange": {
      "occurrences": 7,
      "total_seconds": 958
    },
    "SP Air Max Plus 3": {
      "occurrences": 4,
      "total_seconds": 501
    },
    "SP Flex Runner 2": {
      "occurrences": 3,
      "total_seconds": 547
    }
  }
}
```

This output shows the most frequently viewed products in the analyzed timeframe, with "SP Dunk Low Retro" being the most viewed product with 12 views and 1,850 seconds of total viewing time.

4. Technical Details

The frequent (← Bash wrapper script) is implemented in three different programming languages to demonstrate language-agnostic processing capabilities. All implementations produce identical results:

4.1. Available Implementations

- **JavaScript**: frequent.js - Node.js implementation
- **Python**: frequent.py - Python 3 implementation
- **jq**: frequent.jq - jq implementation

4.2. Usage

The script can be executed using any of these implementations:

```
./frequent js  # Run JavaScript version
./frequent py  # Run Python version
./frequent jq  # Run jq version
```

If no implementation is specified, the script defaults to JavaScript:

```
./frequent
```

All implementations:

- Read JSON Lines from standard input
- Process the first 5 minutes of events
- Sort products by views and viewing time
- Output formatted JSON to standard output