|  |
| --- |
| **Business Requirements Document** |

The objective of this Business Requirements Document (BRD) is to establish clear communication between business process owners and Business Services Program Managers/ Application Services Project Managers while they jointly establish the capabilities needed to be delivered as part of the project. To this end, the requirements must be written so that both perspectives have a clear understanding of what is being proposed for delivery.

It will be used by the Business Analyst at the start of the project to determine effort required for delivering the capabilities, it will be used during design to ensure that all capabilities are included in the delivery, and it will be referenced by the test designers to ensure that all agreed capabilities are being delivered.

|  |
| --- |
| **Project Name: SimilarWeb**  **Business Unit: Strategic Insights**  **Program Manager: Stacy Anapol**  **Project Manager/Business Analyst: Sheetal Abbi** |

**Document Revision History**

This section lists revisions for this Systems Requirements Specification and descriptions of particular changes made.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Version** | **Section/Page** | **Update Description** | **Contact Name** |
| 02/04/2021 | 1.0 | All | Draft | Tyler Post |
| 02/10/2021 | 1.2 | All | Minor changes and added details on licenses available currently via API to be worked on as Phase 2. | Sheetal |
| 05/05/2022 | 2.0 | All | Phase 2 Similar Web | Sheetal |
| 05/24/2022 | 3.0 | All | Scope changed from 1 metric to 23 metrics | Sheetal |
|  |  |  |  |  |

**Document Approval History**

The section provides a means of tracking the review and approval of this Systems Requirements Specification. This is also the list of who has authority to approve changes in the document.

| **Name** | **Project Role** | **Date** | **Responsibility** |
| --- | --- | --- | --- |
| Qi Zhang |  |  |  |
| Austin Wielenga |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# **Business Process flows (optional)**

This section captures existing diagrams in showing the current state and the proposed end-state of the business processes impacted by this project. (Embed diagram files directly here. Avoid adding Links to external locations)

|  |  |
| --- | --- |
| Title of Diagram | Diagram Files |
|  |  |
|  |  |
|  |  |

# **Requirement List**

List out each requirement of this project using the “Description” column - this describes the information, actions or process that the business needs information systems to support.

Use the “Outline Number” column to show relationships between the sub-level-requirements.

Use the “Scope” Column to determine what is IN or OUT of scope

Identify who is responsible for the requirement using the “Owner” column, this is person you go to for an answer or who is ultimately responsible. Should only be one person. Examples: Business Process Owners (BPO), Key Business Users etc.

Assess the priority in the “Priority” column using the following terms:

• M - **Must-Have**: Describes a requirement that must be satisfied in the final solution for the solution to be considered a success.

• N - **Nice-to-Have:** Describes a requirement which is considered desirable but not necessary. This will be included if time and resources permit.

## **General Overview**

SimilarWeb tracks web traffic for vendor websites (e.g. homedepot.com, lowes.com, amazon.com, truevalue.com). Some key metrics that are provided from SimilarWeb are monthly visits, unique visitors, visit duration, bounce rate, key words, conversion rates, and differentiation between mobile and desktop.

The data pertaining to conversion rate and online traffic is currently being used by Strategic Insights as input for the Demand Sensing Model. To fulfill the demand sensing model immediate needs, the data is downloaded from online portal for ingestion into Caspian as Phase 1 of the Project.

As Phase 2 of the project, we plan to bring in all the data available to us via API considering the **limitations** we have in terms of API hits -

* US and Canada platform – limitation of 2000 monthly API hits
* LATAM – LATAM - could be in Portuguese and may require translation- limitation of 500 monthly API hits
* EMEA, NZ and Australia – limitation of 5000 monthly API hits

The API hit is calculated as 1 Metric x 1 Site x 1 country x 1 device type

## **Current Existing Process**

All reports and data required for Phase 1 and Phase 2 are currently manually accessed and downloaded through online user interface /web portal.

### Phase 1

**Conversion Rate**

Currently, a SimilarWeb license holder uses the SimilarWeb online user interface to input the desired parameters for the “Conversion Rate” report and download the data into an Excel report. This file comes with many tabs but only the “Direct” Tab in the excel file is required by the business. This data is then delta loaded to the existing table using the primary key of domains, time period and segment.

Data from the SimilarWeb platform is limited to the past 25 months only, and the monthly data within is often updated. Thus, with each delta load all data entries from before this time period will remain and only the previous 25 months from each data pull will need to be dropped and reloaded.

**Online Traffic**

The license holder uses the user interface to inputs the desired parameters for the “Online Traffic” report and downloads the data into an Excel report. The file comes with many tabs but only the tabs labeled “Monthly” are needed – and the fields are consistent between all these tabs.

* Monthly\_data\_amazon.com
* Monthly\_data\_lowes.com
* Monthly\_data\_homedepot.com
* Monthly\_data\_truevalue.com

The tabs are combined into one and loaded to the existing table using the primary key of domains, channel, and time period.

### Phase 2

All reports and data required for Phase 2 are currently manually accessed and downloaded through online user interface /web portal. These reports are then used in various Dashboards by business. The Phase 2 of this document will explore the opportunity to use API Key and Similar web’s API platform to perform data extraction based on the permissible limits set as per contracts for each region.

## **BUSINESS REQUIREMENT & Proposed Solution**

### Phase 1

**One-time Historical Load**

The business has collected SimilarWeb data beyond 25 months prior in offline Excel files. Please refer to the below two embedded files for the one-time historical load data and to start the tables in the database, with the same logic as the regular loads listed below. These files can also be made available in the SFTP site when set up.

*Historical data:*



**Regular Monthly Loads**

On a monthly basis (around the 10th business day of each month) the excel files need to be pulled from an SFTP site (Dev team to create the SFTP site). Files will be dropped into the SFTP site from a business user who holds a license. The data needs to be loaded into two tables and under one schema. The tables names should be Conversion\_Rate and Online\_Traffic. The Schema name should be SimilarWeb. The two embedded files in the proposed solution section are the desired structures of the tables.

Embedded below are the formats that the business will receive from the SimilarWeb platform. Please use the following excel tabs for each table listed when ingesting. **The data in other tabs are not required by the business.**

* conversion\_rate
  + Direct
* online\_traffic
  + Monthly\_data\_amazon.com
  + Monthly\_data\_lowes.com
  + Monthly\_data\_homedepot.com
  + Monthly\_data\_truevalue.com

*SimilarWeb UI report structure:*



**Delta Logic**

Data from the SimilarWeb platform is limited to the past 25 months only, and the monthly data within is often updated. Thus, with each delta load all data entries from before this time period will remain and only the previous 25 months from each data pull will need to be updated.

For example, data pulled on Nov 2021 contains data from Oct 2019 – Sep 2021. Thus, existing data from prior to Oct 2019 would have to be kept, while existing data after Oct 2019 would need to be updated.

Sample output tables for denoting format are embedded below – this is the expected table format for the business.





Please note that the proposed solution is in lieu of the time constraint for Demand Sensing project and is being proposed as Phase 1 of the solution. As part of Phase 2 for this project, we will interview the business stakeholders and bring in relevant business data using the API links available to us for all the licenses we have data available for (US, Australia, Canada, LATAM and EMEA).

### Phase 2

The API platform provides data related to web analytics services and offers its users information on their clients' and competitors' web traffic and performance. We currently have separate contracts to get data from the below countries–

* US/Canada platform allowed 2000 API hits
* EMEA, NZ and Australia and allowed 5000 API hits
* LATAM allowed 500 API hits

The formula to calculate an API hit is 1 Metric x 1 Site x 1 country x 1 device type

Since business has finalized 23 metrics (listed below) to collect data from, the number of websites(sites) that can be considered for each platform as per the permissible limit are as follows –

* US/Canada Platform can include **86 sites** in total (2000/23)
* EMEA, NZ and Australia Platform can include **217** sites in total (5000/23)
* LATAM platform can include **21** sites in total (500/23)

**Enclosing the list of all Websites (URL’s) provided by the Business in the excel file below - Please note that while the list of sites are within the permissible limit for US/Canada platform, the business is working to condense the sites list for EMEA and LATAM platforms.**

****

**Metrics**

Business users would like to get the data related to the below 23 metrics from each platform/country.



The API documentation/access link is listed below -

[**https://documenter.getpostman.com/view/5388671/RzfcNs8W?version=latest#85dbdaa3-4c00-4cf3-a4b8-37cf7c133d41**](https://documenter.getpostman.com/view/5388671/RzfcNs8W?version=latest#85dbdaa3-4c00-4cf3-a4b8-37cf7c133d41)

**API Key = 350f31c8aa40494cb3435d4d92cc3076**

### **Data Pull Example**

**Metric = Desktop Traffic - Visit, Country = US; Site = dewalt.com**

**API Key =** 350f31c8aa40494cb3435d4d92cc3076

**API link for getting last 24 months History**

[https://api.similarweb.com/v1/website/**dewalt.com**/traffic-and-engagement/visits?api\_key=350f31c8aa40494cb3435d4d92cc3076&start\_date=2020-04&end\_date=2022-04&country=**US**&granularity=monthly&state=&main\_domain\_only=false&format=json&show\_verified=false&mtd=false](https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/visits?api_key=350f31c8aa40494cb3435d4d92cc3076&start_date=2020-04&end_date=2022-04&country=US&granularity=monthly&state=&main_domain_only=false&format=json&show_verified=false&mtd=false)

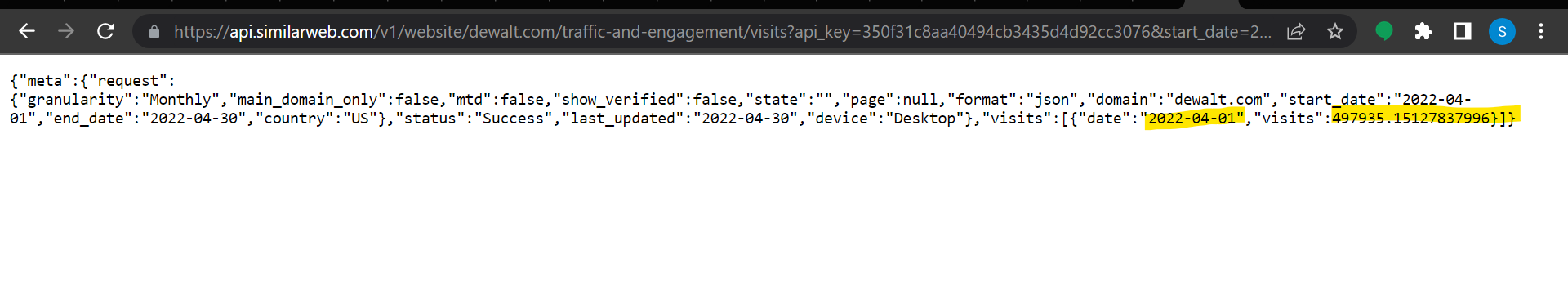
**Data Example (JSON) for the above API data pull (History data for last 24 months)**

****

**API link for getting the data for Delta (1 month data)**

<https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/visits?api_key=350f31c8aa40494cb3435d4d92cc3076&start_date=2022-04&end_date=2022-04&country=US&granularity=monthly&state=&main_domain_only=false&format=json&show_verified=false&mtd=false>

**Delta (Json) data Response for last 1 month**

****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outline Number | Scope | Description | Owner | Priority |
| BR 1.0 | IN | **High Level Data Requirements – Functional:** At high level, we need to acquire and ingest SimilarWeb data for two tables and on a recurring basis perform a delta load for the data– **Create a schema named “similarweb” and create the following tables**   * conversion\_rate * online\_traffic   Data Dictionary Format for the tables are included in Appendix A. | SI | M |
| BR 1.1 | IN | **History Requirements & Frequency:**   * conversion\_rate - One-time historical load, Delta Loads monthly * online\_traffic – One-time historical load, Delta Loads monthly   **Frequency:** Data needs to get updated monthly – per the methodology outlined in “Proposed Solution” | SI | M |
| BR 1.2 | IN | Acquisition - The data needs to be extracted from sftp folders monthly and the solution details are listed in Proposed solution section. | SI | M |
| BR 2.0 | IN | **High Level Data Requirements – Functional:** At high level, we need to acquire and ingest SimilarWeb data and on a recurring basis and perform a delta load for the data after the historical data ingest – **Use the schema “similarweb”, created as part of Phase 1 and create the following 14 tables**   1. Traffic\_visit 2. **Pages\_visit** 3. **Average\_visit\_duration** 4. **Bounce\_Rate** 5. **Page\_Views** 6. **Unique\_visitors** 7. **Traffic\_sources\_overview** 8. **Mobile\_web\_referrals** 9. **Mobile\_web\_outgoing\_referrals** 10. **Mobile\_keywords** 11. **Searchvisit\_distribution** 12. **Keyword\_competitors** 13. **Keyword\_analysis\_competitors** 14. **Desktop\_traffic\_sources**   Data Dictionary Format (API links/structures) for the tables are included in Appendix A. | SI | M |
| BR 2.1 | IN | **History Requirements & Frequency:**   * All tables will need to get History data pulled (last 2 years – as part of the API call link – start date, end-date) followed by monthly delta load(by editing the API links with start date – end-date)   **Frequency:** Data needs to get updated monthly – per the methodology outlined in “Proposed Solution” | SI | M |
| BR 2.2 | IN | Acquisition - The data needs to be extracted from API links shared above and the solution details are listed in Proposed solution section. | SI | M |

# **Assumptions / Dependencies / Constraints (optional)**

We are working with a time constraint to cater to Demand Sensing model and hence utilizing the online web portal platform. The vendor has API links available with some limitations(API hits); For this vendor we have multiple licenses that cater to following platforms–

1. US – US, Canada platform – 2000 monthly API hits
2. LATAM – LATAM - could be in Portuguese and may require translation- 500 monthly API hits
3. EMEA, NZ and Australia – 5000 monthly API hits

API hit formula is 1 Metric x 1 Site x 1 country x 1 device type

As part of Phase 2, we will explore the business needs further and utilize the API links and bring in data into Snowflake.

API documentation

<https://documenter.getpostman.com/view/5388671/RzfcNs8W?version=latest#85dbdaa3-4c00-4cf3-a4b8-37cf7c133d41>

# **User Groups (optional)**

In this section, describe the Users (people, organizations, or other entities) that participate in the execution of business processes that interact with the business and/or interact with the new product of application.

|  |  |
| --- | --- |
| User Group Name | Description |
| NA | Not Applicable |

The facet map describes the background info of the User groups that interact with the new product or application.

|  |  |  |  |
| --- | --- | --- | --- |
| Facets | User Group 1 | User Group 2 | User Group 3 |
| Location: |  |  |  |
| Primary Language: |  |  |  |
| Functions:  (what the user will do with the system) |  |  |  |
| Population:  (how many Users at the location) |  |  |  |

**\*\*Detail Functional requirement must be captured in Functional Requirements/SRS document.**

**\*\*Infrastructure requirements need to be documented in Infrastructure Requirements document template. A/S PM should work with I/S PM to document the requirements prior to ISR submission and envision phase gate review.**

## APPENDIX A

**Phase 1**

**Listing the Data Dictionary for all Datasets**

**Table name - Online\_Traffic**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Position** | **Field Name** | **PK?** | **Format** | **Description** |
| 1 | Time Period | X | Datetime | The time for when the values represent. |
| 2 | Domain | X | String | Website URL that the information was gathered from |
| 3 | Channel | X | String | Type of media |
| 4 | Traffic Share |  | Float | Percentage of traffic based upon domain |
| 5 | Channel Traffic |  | Float | How many views per line item |
| 6 | Avg Visit Duration |  | Time | How long a user was on the domain |
| 7 | Pages / Visit |  | Float | Average number of pages visited by the user |
| 8 | Bounce Rate |  | Float | Percentage of users who enter a site and leave without doing anything else |

**Table name - Conversion\_Rate**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Position** | **Field Name** | **PK?** | **Format** | **Description** |
| 1 | Domains | X | String | Website URL that the information was gathered from |
| 2 | Time Period | X | Datetime | The time for when the values represent |
| 3 | Visits |  | Float | Number of total visits to domain |
| 4 | Converted Visits |  | Float | Number of converted visits at domain |
| 5 | Conversion Rate |  | Float | Visits/Converted Visits |
| 6 | Segment | X | String | Amazon Specific to identify ‘Tools and Home Improvement’ Category |

**Phase 2**

**Since all the API links to all tables will be ingested as JSON files and not flat structures enclosing the API links which has structures listed. Also including the Data samples below -**

## **JSON Data Structure in Snowflake**

### **Desktop Visits**

{  
 "meta": {  
 "request": {  
 "granularity": "Monthly",  
 "main\_domain\_only": false,  
 "show\_verified": true,  
 "format": "json",  
 "domain": "dewalt.com",  
 "start\_date": "2020-04-01",  
 "end\_date": "2022-04-01",  
 "limit": null,  
 "country": "US"  
 },  
 "status": "Success",  
 "last\_updated": "2022-04-30"  
 },  
 "visits": [  
 {  
 "**date": "2022-03-01",  
 "visits": 27834071.99703641**  
 },  
 **{  
 "date": "2022-04-01",  
 "visits": 24885059.04897431  
 },** ]  
}

**14 tables proposed as below for the 23 metrics;**



**Other Data Structures can be extracted from the**

****

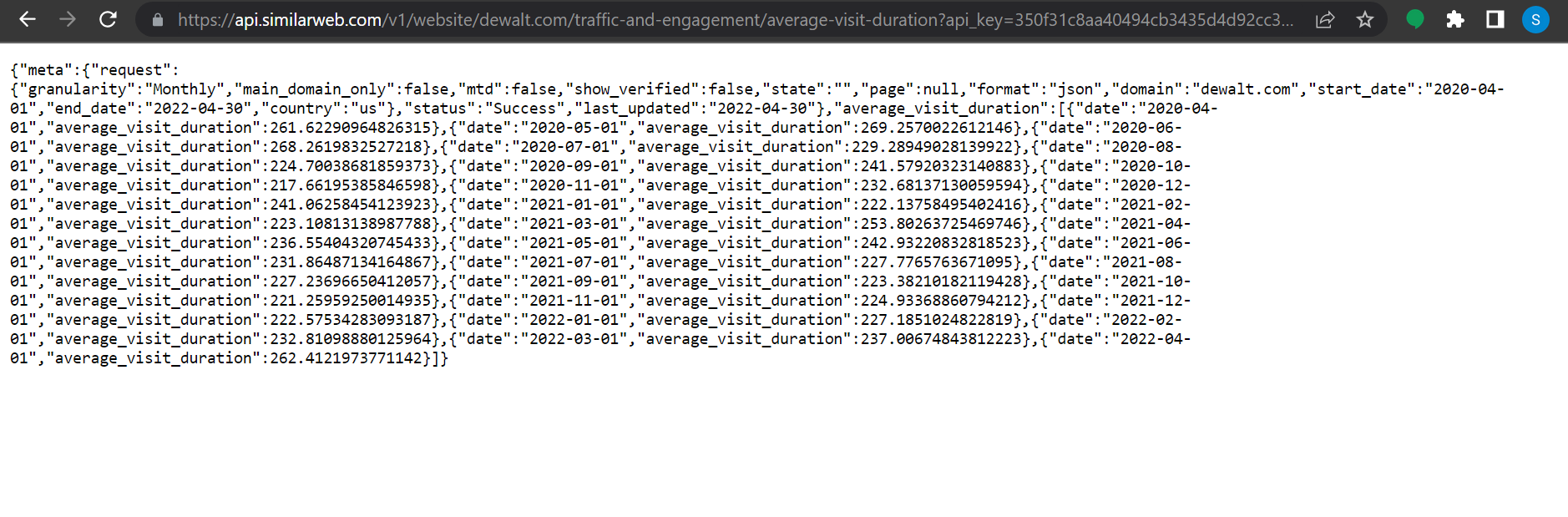
**Data Samples**

### Metric = Desktop Traffic Average visit duration, **Country = US; site = dewalt.com**

**API link for getting last 24 months History**

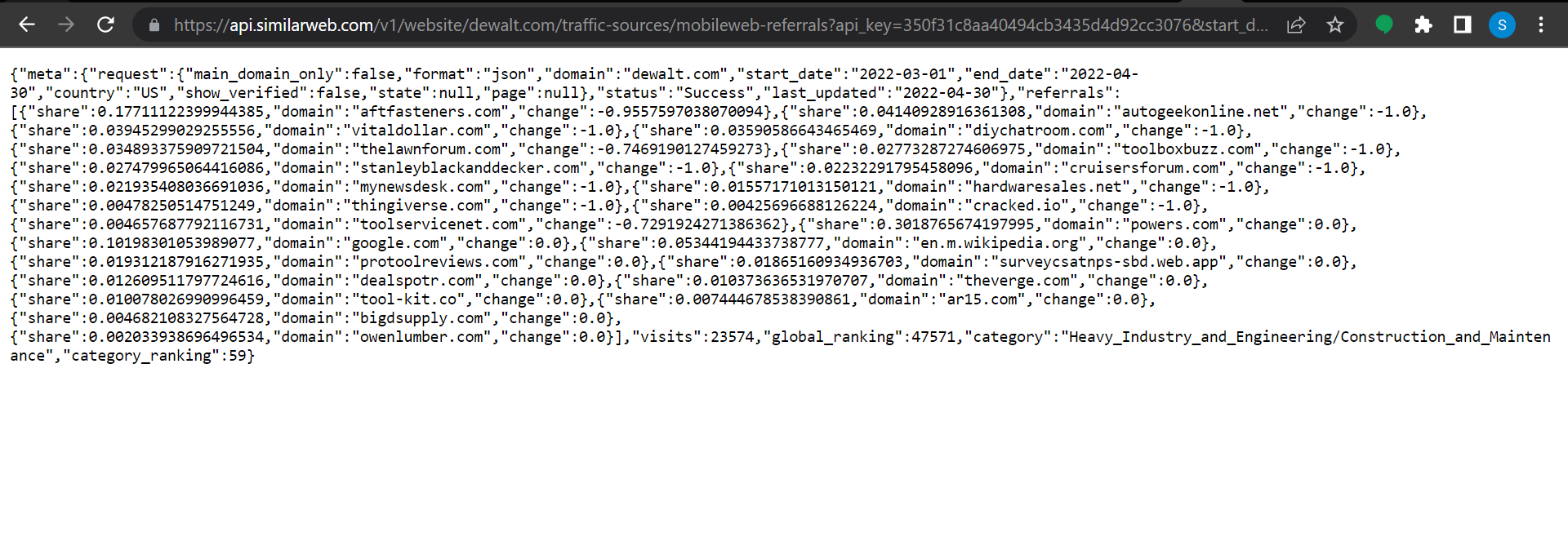
<https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/average-visit-duration?api_key=350f31c8aa40494cb3435d4d92cc3076&start_date=2020-04&end_date=2022-04&country=us&granularity=monthly&state=&main_domain_only=false&format=json&show_verified=false&mtd=false>

**Data Example (JSON) for the above API data pull (History data for last 24 months)**

****

### Metric = Mobile Web referrals**, Country = US; site = dewalt.com**

**Json data file -**

****

### **Mobile Web Referrals**

**Data Header and Body Structure different for this data extract**

{

"meta”: {

"request”: {

"main\_domain\_only”: false,

"format":"json",

"domain":"dewalt.com",

"start\_date":"2022-03-01",

"end\_date":"2022-04-30",

"country":"US",

"show\_verified":false,

"state":null,

"page":null},

"status":"Success",

"last\_updated":"2022-04-30"},

"referrals":

[

{

"share":0.17711122399944385,"domain":"aftfasteners.com","change":-0.9557597038070094},

{

"share":0.04140928916361308,"domain":"autogeekonline.net","change":-1.0},

{

"share":0.03945299029255556,"domain":"vitaldollar.com","change":-1.0},

{

"share":0.03590586643465469,"domain":"diychatroom.com","change":-1.0},

{

"share":0.034893375909721504,"domain":"thelawnforum.com","change":-0.7469190127459273},{"share":0.02773287274606975,"domain":"toolboxbuzz.com","change":-1.0},{"share":0.027479965064416086,"domain":"stanleyblackanddecker.com","change":-1.0},{"share":0.02232291795458096,"domain":"cruisersforum.com","change":-1.0},{"share":0.021935408036691036,"domain":"mynewsdesk.com","change":-1.0},{"share":0.01557171013150121,"domain":"hardwaresales.net","change":-1.0},{"share":0.00478250514751249,"domain":"thingiverse.com","change":-1.0},{"share":0.00425696688126224,"domain":"cracked.io","change":-1.0},{"share":0.004657687792116731,"domain":"toolservicenet.com","change":-0.7291924271386362},{"share":0.3018765674197995,"domain":"powers.com","change":0.0},{"share":0.10198301053989077,"domain":"google.com","change":0.0},{"share":0.05344194433738777,"domain":"en.m.wikipedia.org","change":0.0},{"share":0.019312187916271935,"domain":"protoolreviews.com","change":0.0},{"share":0.01865160934936703,"domain":"surveycsatnps-sbd.web.app","change":0.0},{"share":0.012609511797724616,"domain":"dealspotr.com","change":0.0},{"share":0.010373636531970707,"domain":"theverge.com","change":0.0},{"share":0.010078026990996459,"domain":"tool-kit.co","change":0.0},{"share":0.007444678538390861,"domain":"ar15.com","change":0.0},{"share":0.004682108327564728,"domain":"bigdsupply.com","change":0.0},{"share":0.002033938696496534,"domain":"owenlumber.com","change":0.0}],"visits":23574,"global\_ranking":47571,"category":"Heavy\_Industry\_and\_Engineering/Construction\_and\_Maintenance","category\_ranking":59}

**More Data Extraction Examples**

**Metric = Desktop Traffic - Visit**

**Country = US; site = dewalt.com**

**API link for getting last 24 months History**

**https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/visits?api\_key=350f31c8aa40494cb3435d4d92cc3076&start\_date=2020-04&end\_date=2022-04&country=US&granularity=monthly&state=&main\_domain\_only=false&format=json&show\_verified=false&mtd=false**

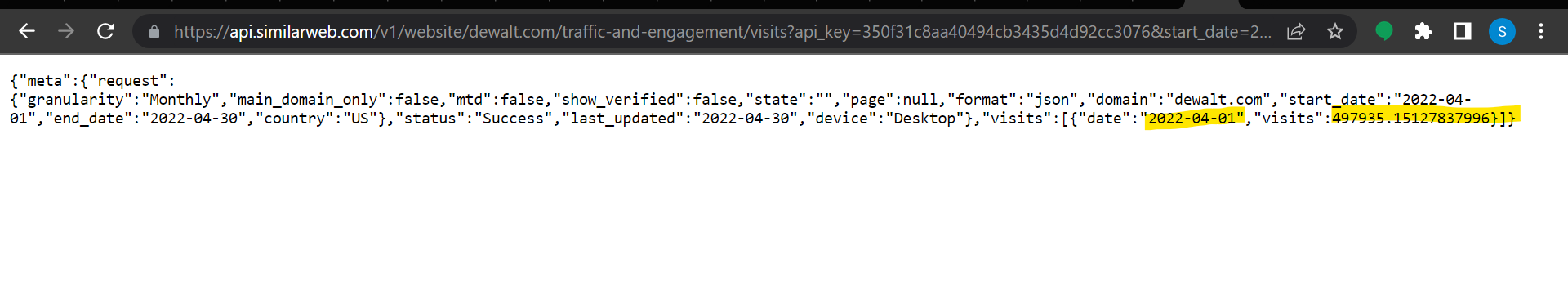
**API link Result for the above (History data for last 24 months)**

****

**API link for getting the data for Delta (1 month data)**

[**https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/visits?api\_key=350f31c8aa40494cb3435d4d92cc3076&start\_date=2022-04&end\_date=2022-04&country=US&granularity=monthly&state=&main\_domain\_only=false&format=json&show\_verified=false&mtd=false**](https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/visits?api_key=350f31c8aa40494cb3435d4d92cc3076&start_date=2022-04&end_date=2022-04&country=US&granularity=monthly&state=&main_domain_only=false&format=json&show_verified=false&mtd=false)

**Delta (Json) data for last 1 month**

****

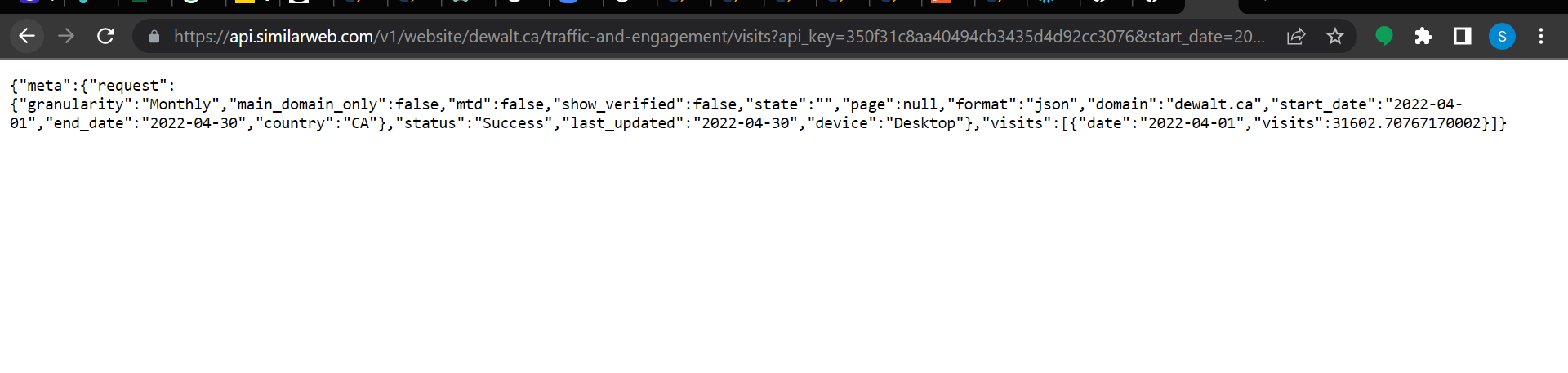
****

**Metric = Desktop Traffic - Visit**

**Country = Canada; site = dewalt.ca**

**API Key = 350f31c8aa40494cb3435d4d92cc3076**

<https://api.similarweb.com/v1/website/dewalt.ca/traffic-and-engagement/visits?api_key=350f31c8aa40494cb3435d4d92cc3076&start_date=2022-04&end_date=2022-04&country=CA&granularity=monthly&state=&main_domain_only=false&format=json&show_verified=false&mtd=false>





## **Metric = Desktop Traffic Average visit duration**

**Country = US; site = dewalt.com**

## [**https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/average-visit-duration?api\_key=350f31c8aa40494cb3435d4d92cc3076&start\_date=2020-04&end\_date=2022-04&country=us&granularity=monthly&state=&main\_domain\_only=false&format=json&show\_verified=false&mtd=false**](https://api.similarweb.com/v1/website/dewalt.com/traffic-and-engagement/average-visit-duration?api_key=350f31c8aa40494cb3435d4d92cc3076&start_date=2020-04&end_date=2022-04&country=us&granularity=monthly&state=&main_domain_only=false&format=json&show_verified=false&mtd=false)

## 

## **Mobile Web referrals Data Pull**

## 