



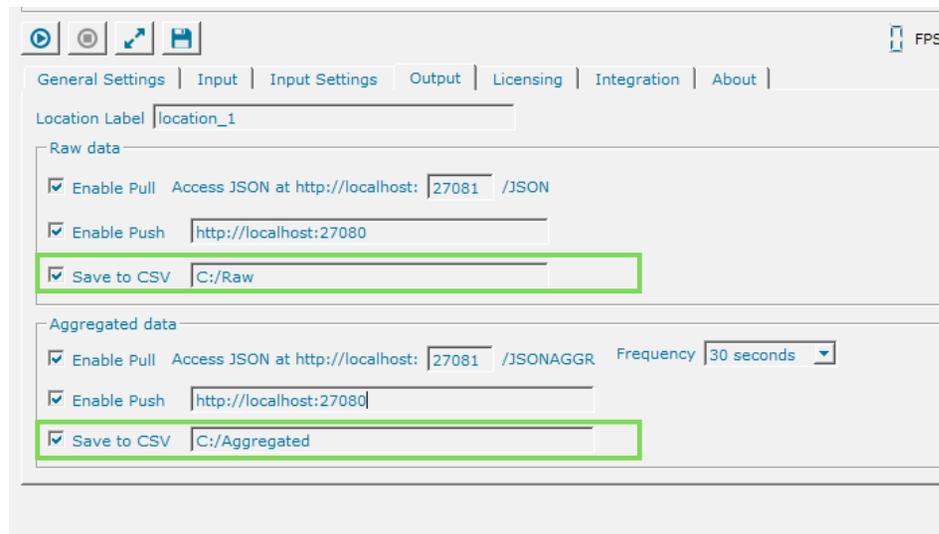
DeepSight Toolkit

Data Aggregation

v6.6.0

Using the CSV DUMPER

A CSV file is a *comma separated values* file which allows data to be saved in a table structured format. The Toolkit generates two CSV files: one containing **raw data** gathered during the real-time analysis, and another with **pre-aggregated data** for each person ID. The raw data can be used for post-processing.



To enable CSV exports go to the *Output* tab and make sure the checkbox **Save data to CSV** is checked. By default, the CSV files will be automatically generated in the Toolkit configuration folder after each analysis. Users can also specify the location where the CSV files will be stored. Every row in the **raw** data CSV file contains the analysis for **1 frame**. In the **aggregated** CSV file each row contains data for one **individual ID/person**.

The table on the left shows an example of what the raw output looks like. As you can see, there are multiple rows for ID 1. This is because the Toolkit analyzes every frame.

The table on the right shows an example of aggregated data for the same analysis. As you can see, there is only one row for ID 1.

Raw CSV Output



| frame | timestamp | id | age | gender |
|-------|------------|----|-----|--------|
| 1 | 1576838867 | 1 | 38 | -99 |
| 3 | 1576838867 | 1 | 38 | -97 |
| 4 | 1576838867 | 1 | 38 | -98 |
| 5 | 1576838867 | 1 | 38 | -98 |
| 6 | 1576838868 | 1 | 38 | -99 |
| 7 | 1576838868 | 1 | 38 | -99 |
| 8 | 1576838868 | 1 | 37 | -99 |
| 9 | 1576838868 | 1 | 37 | -99 |
| 10 | 1576838868 | 1 | 37 | -99 |
| 11 | 1576838868 | 1 | 36 | -99 |
| 12 | 1576838868 | 1 | 36 | -99 |
| 13 | 1576838868 | 1 | 36 | -99 |
| 18 | 1576838869 | 1 | 36 | -99 |
| 19 | 1576838869 | 1 | 36 | -99 |
| 20 | 1576838869 | 1 | 36 | -99 |

Aggregated CSV Output



| id | timestamp | age | gender |
|----|------------------|-----|--------|
| 1 | 19/12/2019 10:47 | 33 | m |
| 2 | 19/12/2019 10:47 | -1 | u |
| 3 | 19/12/2019 10:47 | 43 | f |
| 4 | 19/12/2019 10:47 | 40 | f |
| 5 | 19/12/2019 10:48 | -1 | u |
| 6 | 19/12/2019 10:48 | 24 | f |
| 7 | 19/12/2019 10:48 | -1 | u |
| 8 | 19/12/2019 10:48 | 59 | f |
| 11 | 19/12/2019 10:48 | 40 | m |
| 12 | 19/12/2019 10:48 | -1 | u |
| 13 | 19/12/2019 10:48 | 37 | m |
| 14 | 19/12/2019 10:48 | 40 | m |
| 16 | 19/12/2019 10:48 | 55 | f |
| 17 | 19/12/2019 10:48 | 52 | f |
| 18 | 19/12/2019 10:48 | 42 | f |



Aggregation thresholds

In order to accurately calculate the aggregated data for each ID, the Toolkit applies a few thresholds which help filtering out results with the highest confidence.

The thresholds are;

- Minimum detection time more than 1 second
- Face size less than 20 pixels = unknown age & gender
- Headpose above +30 and -30 degrees in yaw = unknown age & gender

If you get unknown readings, we suggest to filter out these results (see below) and only use results with high confidence or re-position your camera to capture faces from a more frontal angle.

| | id | timestamp | age | gender | impressions | views | attention_sessions | average_attention_time | attention_time | dwel_time | total_detection_time | smile_viewing | smile_maximum |
|---|----|--------------------|-----|--------|-------------|-------|--------------------|------------------------|----------------|-----------|----------------------|---------------|---------------|
| ✗ | 1 | 2020/2/25 11:07:47 | -1 | u | 1 | 0 | 0 | 0 | 0 | 4.379 | 4.379 | 0 | 4 |
| | 2 | 2020/2/25 11:07:47 | -1 | u | 1 | 0 | 0 | 0 | 0 | 4.379 | 4.379 | 0 | 81 |
| ✓ | 1 | 2020/2/25 11:08:27 | 41 | m | 1 | 1 | 1 | 682 | 682 | 45.870 | 45.870 | 7 | 56 |
| | 2 | 2020/2/25 11:08:27 | 30 | f | 1 | 1 | 12 | 728 | 8.743 | 42.669 | 42.669 | 98 | 100 |

How does the aggregation frequency work?

Once a new person is detected (ID 1) the raw data for each frame is aggregated in memory until the person leaves the frame. The aggregation frequency does not affect this in any way as the aggregated CSV row will only be written once the person leaves the frame for a minimum of 5 seconds.

The only exception occurs when DeepSight Toolkit is stopped or closed. This will write any available aggregated data from memory to the aggregated CSV (will also be included in the HTTP output - push and pull).

Since the aggregated output relies on the person to leave the frame, we do not suggest using it for real-time triggers. Instead we recommend using the raw data output for this. The aggregated output is designed for reporting purposes (e.g. pushing aggregated data to a Cloud dashboard).

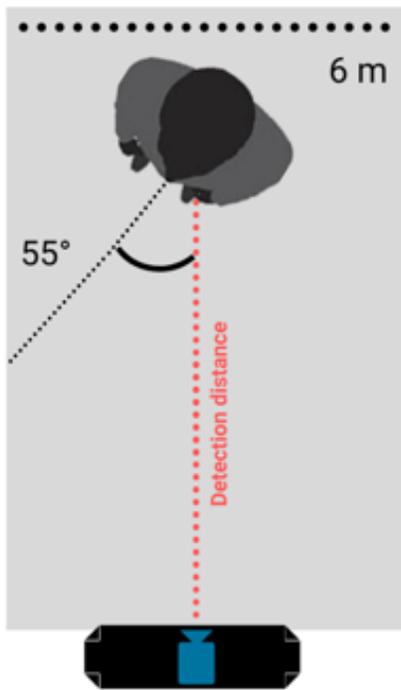
[LEARN MORE
ABOUT RAW DATA](#)

How do we calculate the aggregated metrics?

| METRIC | DESCRIPTION |
|---|---|
| People # of people detected | Count of all faces detected. Head pose +/- 90° from center line (faces within this range will be given an ID number). |
| Impressions # of people likely to see the ad | Count of all impressions detected, headpose +/- 45° from center line. |
| Viewers # of people viewing the ad | Count of all viewers detected, varying head pose depending on distance from screen. |
| Detection Time Sum of time detected | Total detection duration of all faces detected, head pose +/- 90° from center line. |
| Dwell Time Sum of time spent by impressions | Cumulative detection time per impression , headpose +/- 45° from center line. |
| Attention Time Sum of time spent by viewers | Cumulative detection time per viewer , varying head pose depending on distance from screen. |
| Attention session Sum of attention sessions | Count of sessions where viewing = 1. Repeated looks per viewer. |
| Attention time per session Average attention time per session | Average attention time per session per viewer. |

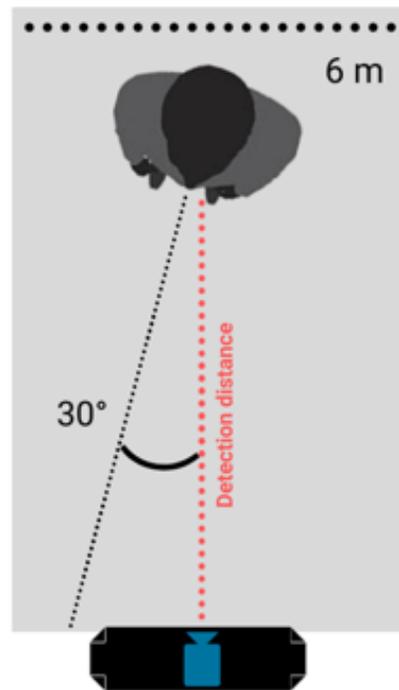
People count/OTS, Impressions and Viewers explained

OTS



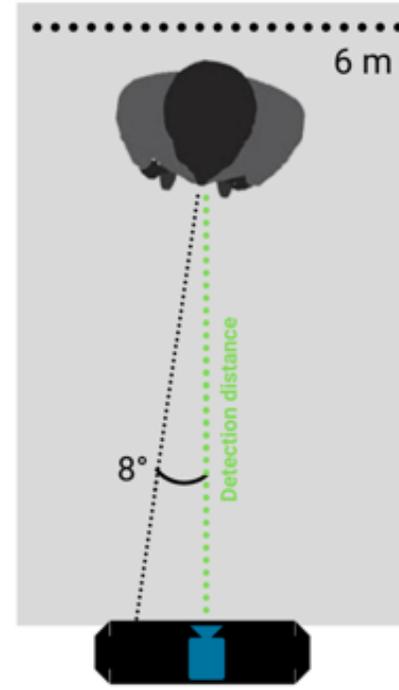
Detected, not viewing

Impressions

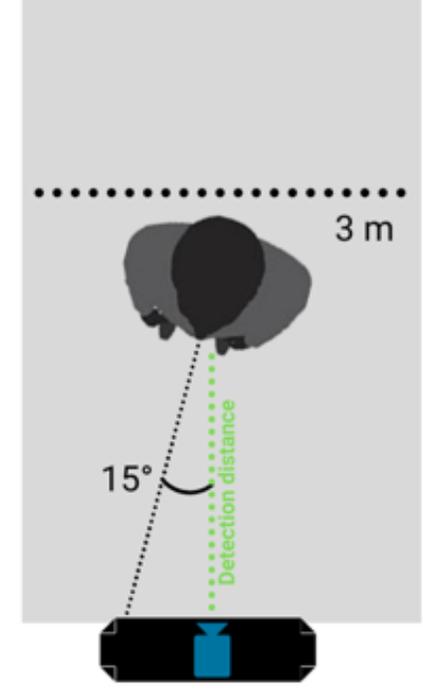


Detected, not viewing

Views



Detected, viewing



Detected, viewing

How do we calculate the aggregated metrics?

| METRIC | DESCRIPTION |
|--|---|
| Gender Female, male, unknown | Faces above 20 pixels in size and headpose within +-30 degrees in yaw will be classified as either male (m) or female (m). Faces outside of this range will be classified as unknown (u). |
| Age Exact age in years | Faces above 20 pixels in size and headpose within +-30 degrees in yaw will be given an exact age value. Faces outside of this range will be given an unknown age (-1). |
| Smile Viewing Maximum smile per viewer | The highest smile value per viewer. |
| Smile Maximum Maximum smile per person | The highest smile value per ID number. |

Aggregated data CSV Item Description

ID: this is the unique identification number assigned by the Toolkit to a **detected person**. This ID number is only temporary therefore if someone exits the frame and then returns after some time, they will be assigned a new ID. This way the analysis remains anonymous and there is no re-identification of people.

Timestamp: the date and time at which an ID was first detected. This is expressed in a dd/mm/yyyy hh:mm:ss format.

Age: the age value for a specific ID.

Gender: the gender value for a specific ID (m = male, f= female, u = unknown).

Impressions: the number of people detected as a qualified impression - head pose range within the +45° / -45°.

Views: a binary value that changes to 1 whenever a person's head position is detected as looking towards the focal point (camera). The viewing angle changes with the distance from the camera. Detections further away from the camera will have a narrower 'Viewing' angle which provides far more accurate estimations.

Attention sessions: the number of times a detected person has been 'Viewing'. Dividing the number of attention sessions by the number of viewers gives you an average number of views per ID.

Average attention time: the average duration of an attention sessions per ID.

Attention time: the total amount of time that a detected person has spent looking towards the object of interest (camera). This is expressed in seconds with 3 decimals.

Total Detection time: cumulative detection duration per ID expressed in seconds with 3 decimals.

Dwell sessions: the number of times a detected person has been 'Dwelling'. Dividing the total dwell time by the number of dwell sessions gives you an average amount of dwell per ID.

Average dwell time: the average amount of time that a detected person has spent facing the object of interest (camera) per dwell session. A new session starts when a detected person leaves the 'dwell' state and returns within 45° of the camera during the same detection.

Dwell time: the total amount of time that a detected person has spent facing the object of interest (camera). A person is considered as facing the camera whenever their head pose (yaw) is between +45° and -45°.



Smile viewing: the maximum smile value (in %) while a person was detected as 'Viewing'.

Smile maximum: the maximum smile value (in %) while a person was detected. This can be compared to the maximum 'Smile when viewing' to assess the effect of the displayed content on the overall mood of your audience.

Face_mask: binary value (1 or 0) per ID based on the average mask detection confidence throughout the detection. If the average mask confidence value is above 0.75 that specific ID will be aggregated as mask 1 (**yes mask**). For IDs with an average below 0.75 the aggregated CSV will show 0 (**no mask**).

3. Using the PULL MECHANISM

The HTTP interface can be used to connect any Business Intelligence platform or CMS to the Toolkit. Go to '**Output**' Tab and click on the fields: '**Enable Pull**'. You have the option to either get **raw** or **aggregated** data.

The JSON response can be accessed through a GET request on localhost for your external application. You can use this function for instance for Digital Signage or narrowcasting applications which trigger custom and relevant ad content to the right people at the right time.

For the raw data JSON output you can access: **<http://localhost:27080/JSON>**

For the aggregated JSON output you can access: **<http://localhost:27080/JSONAGGR>**

On the next page you can find a screenshot highlighting the Output tab along with the boxes that need to be checked for the HTTP interface to function.

General Settings | Input | Input Settings | **Output** | Licensing | About

Location Label

Raw data

Enable Pull Access JSON at http://localhost: /JSON

Enable Push

Save to CSV

Aggregated data

Enable Pull Access JSON at http://localhost: /JSONAGGR Frequency ▾

Enable Push

Save to CSV

Toolkit Output tab

JSON aggregated example output

```

{
  "submission_time": <date_time of submission in YYYY/MM/DD hh:mm:ss>,
  "detection_count": <number of detected people>,
  "source": <location_label>,
  "aggregator_freq": <frequency at which data is reported in ms>,
  "state": <reported even if no detections, 1 if Toolkit is on>,
  "aggregated_data": [{
    "id": <person ID>,
    "timestamp": <date_time of detection in YYYY/MM/DD hh:mm:ss>,
    "age": <age_estimation in years>,
    "gender": <gender estimation (m = male, f = female, u = unknown)>,
    "impressions": <count of impressions>,
    "views": <count of views>,
    "attention_sessions": <count of attention sessions>,
    "average_attention_time": <average attention time per session in ms>,
    "attention_time": <total attention time in ms>,
    "dwell_sessions": <count of dwell sessions>,
    "average_dwell_time": <average dwelling time per session in ms>,
    "dwell_time": <total dwelling time in ms>,
    "total_detection_time": <total detection time in ms>,
    "smile_viewing": <maximum smile while viewing>,
    "smile_maximum": <maximum smile overall>,
    "face_mask": <mask detector estimation (1 = wearing mask, 0 = not wearing mask)>
  ]
}

```

```

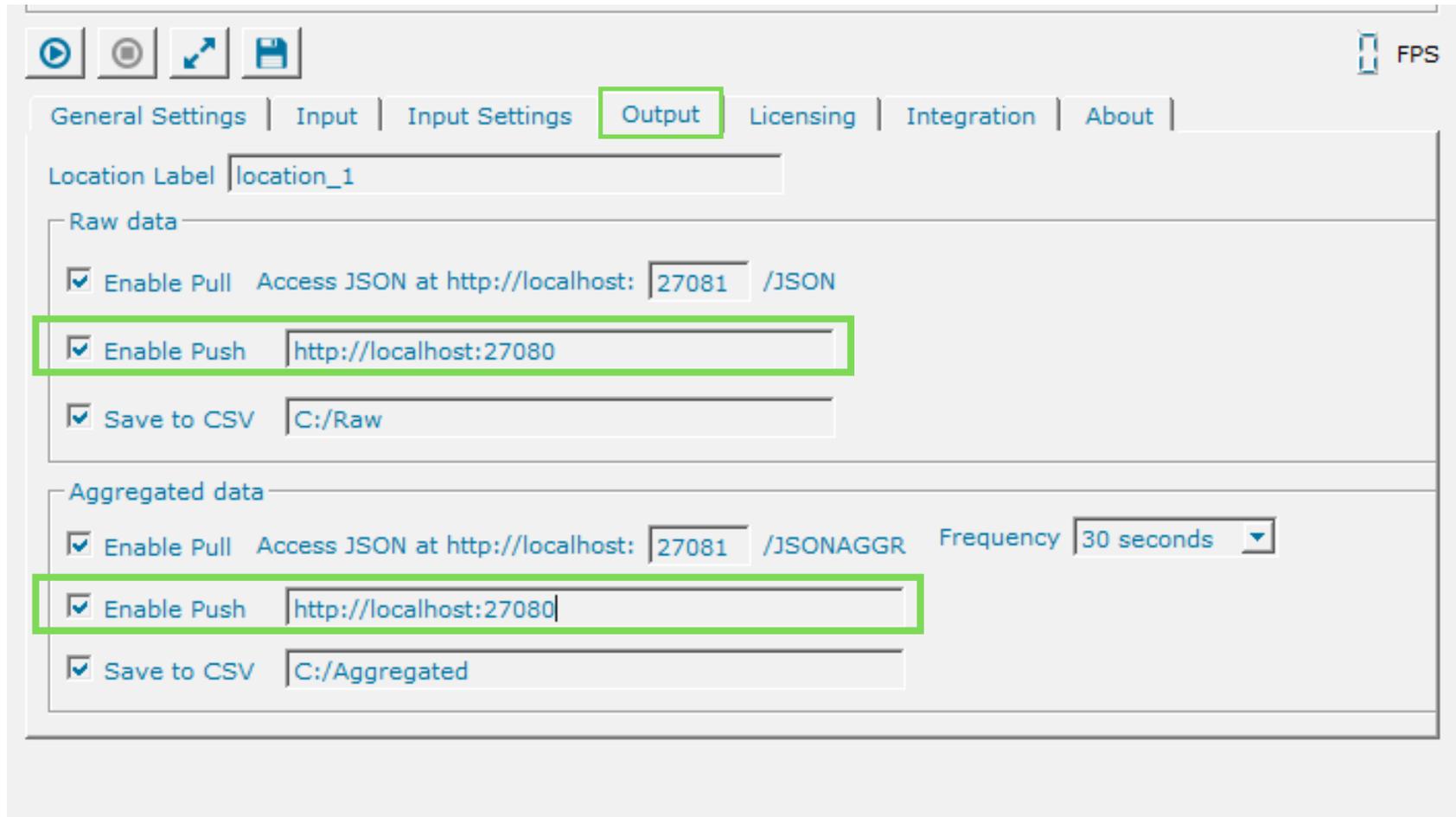
{
  "submission_time": "2021-03-31 12:43:16+02:00",
  "detection_count": 1,
  "source": "location_1",
  "aggregator_freq": 30000,
  "state": 1,
  "aggregated_data": [{
    "id": 7,
    "timestamp": "2021-03-31 12:42:53+02:00",
    "age": 28,
    "gender": "m",
    "impressions": 1,
    "views": 1,
    "attention_sessions": 1,
    "average_attention_time": 7.618,
    "attention_time": 7.618,
    "dwell_sessions": 1,
    "average_dwell_time": 7.618,
    "dwell_time": 7.618,
    "total_detection_time": 7.618,
    "smile_viewing": 38,
    "smile_maximum": 38,
    "face_mask": 0
  ]
}

```

Using the PUSH Mechanism

The push mechanism allows for messages to be sent from the Toolkit to external applications (eg. **CMS systems**) instead of requiring the external application to actively ask the Toolkit for changing status (pulling). Therefore the Toolkit can push data into any external application or API.

If the push feature is enabled, the Toolkit will send an **HTTP POST** request to the endpoint defined by the user every second. The endpoint needs to be a **valid HTTP server** able to receive HTTP POST requests. The body of the HTTP POST request is a JSON object containing a summary of the last second of analysis. You can add the target URL directly in the Toolkit interface.

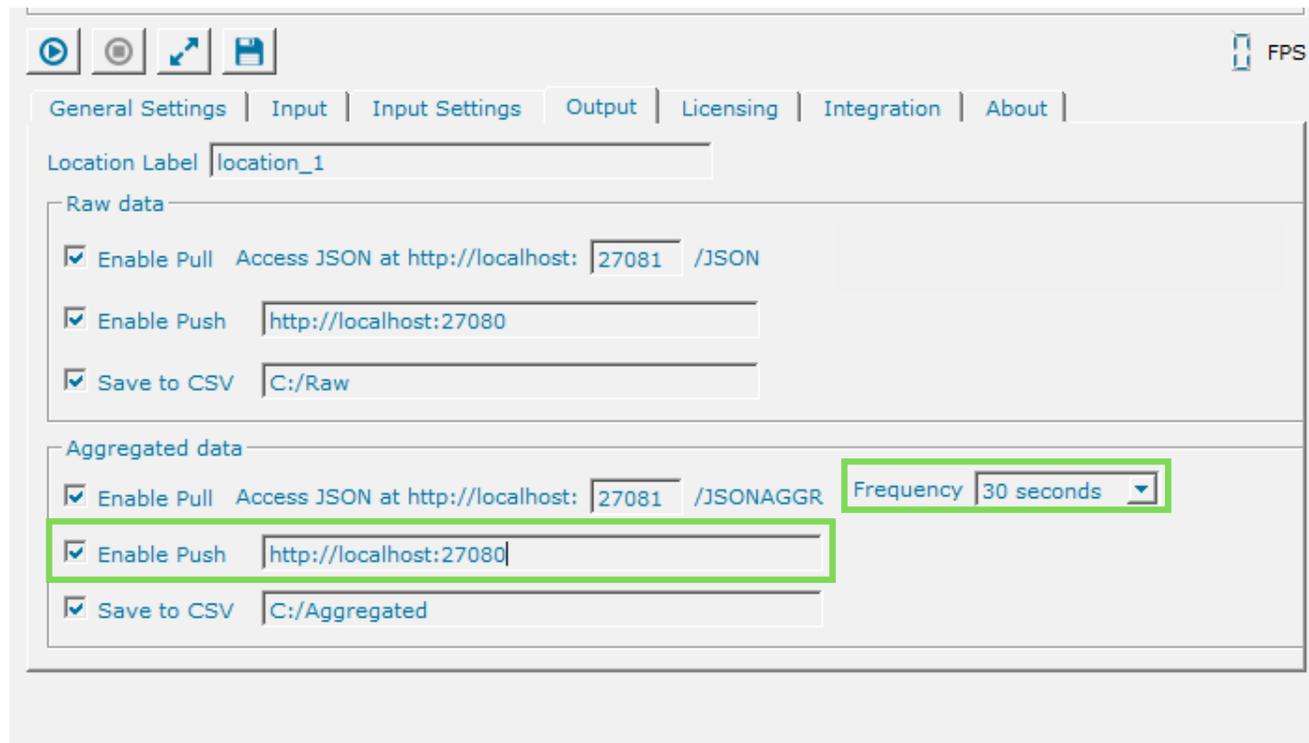


The screenshot shows the 'Output' tab of a configuration tool. At the top, there are navigation icons (play, stop, refresh, save) and a 'FPS' icon. Below these are tabs for 'General Settings', 'Input', 'Input Settings', 'Output' (highlighted), 'Licensing', 'Integration', and 'About'. The 'Location Label' is set to 'location_1'. The 'Raw data' section has three options: 'Enable Pull' (checked) with 'Access JSON at http://localhost: 27081 /JSON', 'Enable Push' (checked and highlighted) with 'http://localhost:27080', and 'Save to CSV' (checked) with 'C:/Raw'. The 'Aggregated data' section has three options: 'Enable Pull' (checked) with 'Access JSON at http://localhost: 27081 /JSONAGGR' and 'Frequency' set to '30 seconds', 'Enable Push' (checked and highlighted) with 'http://localhost:27080', and 'Save to CSV' (checked) with 'C:/Aggregated'.

Toolkit Output tab

The aggregated JSON output will be sent to your server's location every 30 seconds, 1 minute, 5 minutes, or 10 minutes, depending on the **Frequency** that you select from the drop down box on the right hand side of the *Output* tab. This will contain data aggregated per ID for any detections that occurred during the aggregation period.

The Toolkit will aggregate data for each ID 5 seconds after the last detection occurs in case the same person returns into the video frame.



The screenshot shows the 'Output' tab of the Toolkit configuration interface. The 'Location Label' is set to 'location_1'. The 'Raw data' section has three options checked: 'Enable Pull' (Access JSON at http://localhost:27081 /JSON), 'Enable Push' (http://localhost:27080), and 'Save to CSV' (C:/Raw). The 'Aggregated data' section also has three options checked: 'Enable Pull' (Access JSON at http://localhost:27081 /JSONAGGR), 'Enable Push' (http://localhost:27080), and 'Save to CSV' (C:/Aggregated). The 'Frequency' dropdown menu is set to '30 seconds' and is highlighted with a green box. The 'FPS' icon is visible in the top right corner.

Toolkit Output tab

JSON aggregated example output

```

{
  "submission_time": <date_time of submission in YYYY/MM/DD hh:mm:ss>,
  "detection_count": <number of detected people>,
  "source": <location_label>,
  "aggregator_freq": <frequency at which data is reported in ms>,
  "state": <reported even if no detections, 1 if Toolkit is on>,
  "aggregated_data": [{
    "id": <person ID>,
    "timestamp": <date_time of detection in YYYY/MM/DD hh:mm:ss>,
    "age": <age_estimation in years>,
    "gender": <gender estimation (m = male, f = female, u = unknown)>,
    "impressions": <count of impressions>,
    "views": <count of views>,
    "attention_sessions": <count of attention sessions>,
    "average_attention_time": <average attention time per session in ms>,
    "attention_time": <total attention time in ms>,
    "dwell_sessions": <count of dwell sessions>,
    "average_dwell_time": <average dwelling time per session in ms>,
    "dwell_time": <total dwelling time in ms>,
    "total_detection_time": <total detection time in ms>,
    "smile_viewing": <maximum smile while viewing>,
    "smile_maximum": <maximum smile overall>,
    "face_mask": <mask detector estimation (1 = wearing mask, 0 = not wearing mask)>
  ]
}

```

```

{
  "submission_time": "2021-03-31 12:43:16+02:00",
  "detection_count": 1,
  "source": "location_1",
  "aggregator_freq": 30000,
  "state": 1,
  "aggregated_data": [{
    "id": 7,
    "timestamp": "2021-03-31 12:42:53+02:00",
    "age": 28,
    "gender": "m",
    "impressions": 1,
    "views": 1,
    "attention_sessions": 1,
    "average_attention_time": 7.618,
    "attention_time": 7.618,
    "dwell_sessions": 1,
    "average_dwell_time": 7.618,
    "dwell_time": 7.618,
    "total_detection_time": 7.618,
    "smile_viewing": 38,
    "smile_maximum": 38,
    "face_mask": 0
  ]
}

```

To know more about the Toolkit CSV output you can download a sample file below:

DOWNLOAD
RAW CSV

DOWNLOAD
AGGREGATED
CSV

Every row in the raw data CSV file contains the analysis for **each frame**.
In the aggregated CSV file each row contains data for **each individual ID/person**.

You can also watch a short walk-through video on how to use the aggregated CSV file.

WATCH A VIDEO
about CSV
output