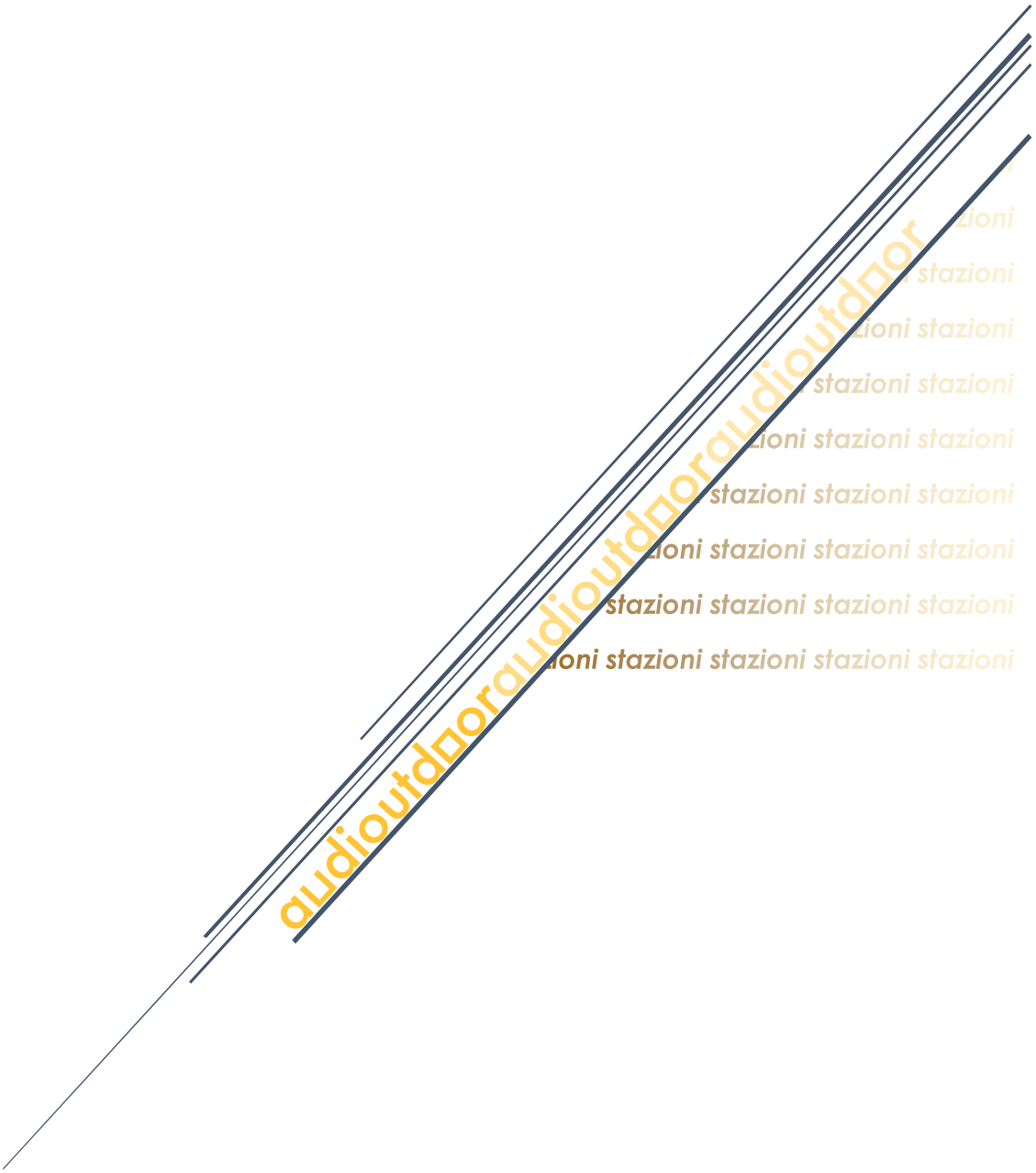
Metodologia della ricerca Audioutdoor

Modulo stazioni



Luglio 2026

Ambiente stazioni

Il modulo stazioni

Viene qui riportata la metodologia di calcolo dell'audience per ambienti chiusi, in questo caso *ad hoc* per le stazioni ferroviarie che, a differenza di aeroporti e metropolitane, non dispongono di dati censuari sui frequentatori; il principale dato a disposizione è rappresentato dalla sensoristica installata dai possessori degli spazi, anche per altri scopi.
Le regole che portano all'installazione della sensoristica sono condivise e approvate da Audioutdoor, così come le regole che calcolano gli eventuali momenti di mancato funzionamento del sensore.
Inoltre, Audioutdoor procede tramite sistematici controlli attraverso società terze al controllo della bontà dei processi di rilevamento da parte dei sensori.

Fonti dato necessarie per la metodologia "ambienti chiusi con sensori"

1. Dati telco

Il dato primario è quello telco, che fornisce gli individui presenti all'interno dei confini dello specifico ambiente (nel caso di Grandi Stazioni, di ogni singola stazione, con profili di età, sesso e provenienza sul territorio per il solo target individui italiani che vengono, attraverso l'elaborazione GfK analoga a tutti gli ambienti Audioutdoor, riportate alle 2016 combinazioni sociodemografiche necessarie per il funzionamento della ricerca).
Il dato telco rilevato considera gli individui con una presenza di almeno 60 secondi continuativi.
Le ulteriori informazioni relative alla condizione professionale delle donne, e alla responsabilità degli acquisti alimentari vengono invece ricavate da altre fonti dato, come meglio illustrato nella metodologia.
Questo dato costituisce la base per la stima degli utenti unici.

2. Sensoristica

Per quanto riguarda la rilevazione con sensoristica, questa è in capo alla concessionaria, ma con garanzie per Audioutdoor sia sulle metodologie di rilevazione dettagliate, sia sul corretto funzionamento dei sensori.
In primis, vengono esaminate da Audioutdoor tutte le problematiche che sorgono da parte della società che gestisce i sensori, emerse durante il periodo dell'installazione (tra cui, il mancato funzionamento per periodi di breve o lunga durata, eventuali limiti massimi di quantificazione delle presenze, e simili…). In secondo luogo, vengono esaminati i fattori correttivi e/o i processi di Data Quality applicati per raggiungere una misurazione corretta.
Questo passaggio è oggetto di un capitolato che la società gestore fornisce ad Audioutdoor, costantemente aggiornata.
La concessionaria si fa garante nei confronti di Audioutdoor che la rilevazione tramite sensore è anonimizzata e rispetta le regole di legge sulla privacy.

Il sensore rileva, oltre alle quantità di persone che si trovano nell'area di visibilità (definita dalle regole di visibilità Audioutdoor), soprattutto il tempo di permanenza all'interno del cono di visibilità del singolo impianto.
Non tutti gli impianti pubblicitari sono dotati di sensori; in questi casi, si applica un modello probabilistico per fornire queste informazioni anche agli impianti che ne sono sprovvisti.
Vi è un minimo di sensori necessari per offrire una buona copertura di ciascun ambiente, che dipende dalla potenza dei sensori, dal numero di impianti e dalla loro dislocazione, nonché dalla dimensione in metri quadri dell'ambiente stesso.
Audioutdoor valuta, caso per caso, che questa condizione sia soddisfatta, in quanto non è possibile a priori determinare una regola generale.
Audioutdoor attiva controlli di società terze sulle corrette applicazioni.

La concessionaria indica ad Audioutdoor il numero dei sensori, l'ampiezza dell'area, le mappe dettagliate degli ambienti e il posizionamento degli impianti con le caratteristiche anagrafiche.

Inoltre, vengono condivise con Audioutdoor sia le dimensioni di ciascuno spazio calpestabile, sia la dimensione delle singole celle che costituiscono il reticolato delle zone calpestabili.
Nelle mappe sono riportati eventuali ostacoli che sono considerati come limiti di non visibilità dell'impianto. La distanza di visibilità soggetta a ostacoli può variare per impianti identici.

Questi dati rilevano le OTS delle persone che transitano nelle zone di interesse calpestabili. La granularità minima di aggregazione è oraria per garantire la redistribuzione degli individui unici giornalieri nelle OTS generate in ciascuna zona.
Il cono di visibilità è quello stabilito per ogni singolo formato da Audioutdoor per qualsiasi ambiente (140° e raggio varabile in funzione della diagonale e dell'angolo tra la base e la diagonale)

3. Produzione degli incroci tra individui e impianti

Audioutdoor ha affidato il trattamento dei dati telco e dei dati provenienti dai sensori per la produzione degli incroci tra individui e impianti alla società Target Research.
Il file prodotto, contenente un giorno medio, è uguale a quello generato per tutti gli altri ambienti Audioutdoor (id individuo, id impianto, ora di inizio contatto, ora di fine contatto e durata del contatto), e sottoposto attraverso elaborazione di GfK e MCS all'espansione a 365 giorni e calcolo della copertura multiambiente (metodologie e algoritmi già presenti sul sito di Audioutdoor per tutti gli altri ambienti).

Per semplificare la produzione file degli incroci sono stati definiti dei percorsi tipo tramite la previa definizione di sottoaree dello specifico ambiente a cui vengono associate delle probabilità di essere attraversate può produrre N percorsi tipo. Ciascun individuo viene associato ad un singolo percorso, rispettando la distribuzione di probabilità di seguire un determinato percorso ed un numero di impianti che si trovano su quel percorso.
Nel produrre queste associazioni individui-percorsi che servono per arrivare all'incrocio individuo-impianto, viene posta particolare attenzione nel rispettare la distribuzione delle informazioni sociodemografiche degli individui che provengono da fonti telco.

Quindi, il file degli incroci ha tante righe quante sono le OTS generate in un giorno medio nello specifico ambiente.

Dati e file input

1. Dati telco: Wind-Tre

I dati di flusso di Wind-Tre sono rilevati tramite le sim che sono presenti all'interno del poligono che definisce i limiti della stazione. Tutte le sim W3 che entrano all'interno di questo poligono vengono conteggiate grazie alla triangolazione operata dalle 3 antenne telefoniche più vicine alla stazione in oggetto (errore approssimativo di rilevazione di 150 metri).
I dati vengono poi riparametrati tramite la quota di mercato W3 per stimare il totale del flusso di persone in stazione. Quindi, se in una stazione vengono rilevate 15.000 persone in un giorno e W3 ha il 30% di quota di mercato: 15.000/30% = 50.000 persone stimate in stazione. W3 fornisce anche informazioni sociodemografiche dei possessori delle sim: le fasce d'età, il sesso, se popolazione straniera o italiana.

- Per gli italiani si conosce anche la provenienza della sim in termini di domicilio prevalente (comune dove ha passato la maggior parte delle notti nelle ultime quattro settimane), quindi da dove arriva l'individuo quando accede nell'area di rilevazione dell'ambiente.
- Per gli stranieri solo il loro totale.

Questi dati vengono caricati ogni notte su Data Warehouse e si riferiscono al giorno precedente (p.es. il martedì si vedono i dati del lunedì).
W3 fornisce gli individui unici con diverse granularità:

- Unici al minuto
- Unici all'ora
- Unici al giorno
- Unici alla settimana

- Unici bisettimanali
- Unici mensili

Al fine del calcolo del numero degli unici nel giorno medio (file degli individui) viene elaborata la media degli "unici al giorno" di un intero anno (e.g. ultimi 12 mesi rolling) per ciascuno degli ambienti considerati, ovvero le 14 stazioni ferroviarie di pertinenza GSR. La tabella degli "unici orari" viene invece utilizzata per definire la distribuzione degli unici giornalieri medi precedentemente definiti.

I rapporti tra le diverse granularità di unici ci dicono quanto è il tasso di ritorno medio nel periodo considerato. Per esempio, se divido la somma degli unici giornalieri di una settimana per il valore degli unici settimanale fornito da W3 avrò un valore che mi indica qual è il tasso di ritorno medio di quegli individui (visite). Per esempio, somma unici giornalieri di 7 giorni = 140; Unici settimanali 100. Tasso di ritorno = 140/100 = 1,4. Quindi in media gli individui di quella settimana hanno fatto 1,4 visite (la maggior parte è venuta una sola volta, qualcuno due o più volte). In questo modo possiamo calcolare i tassi di ritorno (visite) in qualsiasi periodo che va dall'ora al mese.

NOTA: Le OTS invece rappresentano il numero di volte in cui coloro che passano in stazione sono esposti ad una pubblicità. In un giorno quindi una persona viene registrata come 1 unico, però potrebbe essere entrata in stazione 2 volte (2 visite) e potrebbe essere stata esposta a 12 schermi in lassi di tempo diversi (potenzialmente anche 2 volte agli stessi 6 impianti se fa il tragitto inverso la sera perché è un pendolare), generando quindi 12 OTS.

Per i dati sociodemografici non presenti nei dati telco (responsabili acquisti e donna che lavora e non lavora) questi vengono acquisiti tramite integrazione del dato telco condiviso con elaborazioni affidate a GfK, la quale si occupa di espandere il dato presente disaggregandolo per le ulteriori variabili necessarie, mantenendo invariati i volumi complessivi.

2. Dati da sensori (telecamere): Quividi

Nel caso di Grandi Stazioni, i dati provengono da sensori installati da Quividi [1], un software di Facial Recognition. Si tratta di una tecnologia basata su algoritmi, che, tramite il rilevamento del passaggio di persone davanti agli schermi, permette di ottenere dati sui Watchers e sulle Opportunity To See (OTS), rispettivamente coloro che guardano lo schermo per un intervallo di tempo di almeno 0.1 secondi e coloro che si trovano nella zona di visibilità dello schermo (e che hanno appunto avuto anche solo "l'opportunità di vedere la pubblicità trasmessa dallo schermo").

Le OTS sono una stima delle possibili visualizzazioni della pubblicità, date dalla moltiplicazione dei Watchers per un fattore di conversione. Questo aggiustamento è necessario per comprendere tutti coloro che sono passati nel cono di visibilità dello schermo. Equivale al traffico nella zona di rilevazione, ma è un valore più elevato rispetto ai soli Watchers effettivamente riconosciuti dalla telecamera e dal sistema Quividi.

Quividi VidiReports misura l'attenzione considerando la durata cumulata di diverse rilevazioni dello stesso volto. Se la stessa persona esce dal cono di visibilità per poi rientrarvi, viene considerata come nuovo Watcher. La tecnologia è in grado di rilevare anche il genere e l'età dei passanti, la durata per cui i Watchers hanno prestato attenzione allo schermo (durata in decimi di secondo del viso rivolto verso lo schermo: "Attention Time") e Tempo di permanenza all'interno del cono di visibilità della tecnologia (Dwell Time, anch'essa in decimi di secondo). I Watchers hanno granularità minima, con dettaglio del timestamp di inizio e fine di permanenza all'interno della zona di visibilità dello schermo e relativa durata. Le OTS rilevate vengono aggregate ad intervalli di 15 minuti. Ai fini del calcolo dell'audience, le OTS vengono aggregate per ora. Watchers e OTS sono resi disponibili su DWH il giorno successivo rispetto alle rilevazioni.

Le telecamere consentono di riconoscere persone fino ad 8 metri di profondità ed hanno un'ampiezza di visione tra i 60 e i 70 gradi. Nel calcolo dell'audience, vengono considerate solo le persone con il volto rivolto verso lo schermo, non quelle di spalle e non i volti che potrebbero essere presenti in alcuni poster sullo sfondo, che vengono filtrati dalla tecnologia come falsi positivi. I dati forniti da Quividi sono sempre anonimizzati e resi in forma aggregata, nel rispetto della normativa vigente sulla privacy.".
Le OTS del giorno precedente vengono caricate ogni notte su DWH.

[1] Un altro software affermato nel panorama della sensoristica per la *data detection and facial recognition* è Blimp

OTS_Clean: Le OTS_Clean vengono generate tramite una procedura di sostituzione dei dati mancanti (Progetto Data Quality) implementato direttamente nel Data Warehouse (store procedure scritta in Mircosoft SQL). È stato necessario implementare questa procedura in quanto le telecamere dotate di tecnologia Quividi sono ovviamente soggette a malfunzionamenti di vario tipo (problemi della telecamera, problemi dell'impianto digitale su cui sono installate o problemi del software Quividi). Queste problematiche comportano l'interruzione delle rilevazioni dei dati necessari alla valorizzazione delle dashboard (in particolare le OTS) fino a che il problema non viene risolto, da remoto, se è un problema di software, o con un intervento da parte di una squadra di manutenzione in loco, se è un problema fisico della telecamera o dell'impianto su cui quest'ultima è installata. Di conseguenza, senza una procedura di sostituzione dei valori mancanti vi sarebbero dei veri e propri "buchi" di valorizzazione per tutto il periodo di malfunzionamento delle telecamere che alimentano la valorizzazione di una determinata zona.

La procedura di sostituzione dati implementata funziona nel seguente modo: i dati delle OTS vengono aggregati per granularità oraria e si controlla che le 4 rilevazioni contenenti dati non siano mancanti (le OTS per ogni impianto dotato di telecamera vengono restituiti ogni 15 minuti). Se le 4 rilevazioni che compongono l'ora sono mancanti (manca la riga per quell'impianto per quel quarto d'ora nella tabella originaria fact_quividi_ots) allora viene innescato il meccanismo di sostituzione delle OTS. In particolare, gli elementi utilizzati per la produzione delle OTS_Clean, che hanno granularità oraria, sono due:

1. Viene calcolato il peso % condizionato all'ora e al giorno della settimana che la specifica telecamera ha avuto nel tempo sul totale delle OTS rilevate nella stazione di riferimento da quando è stata installata. Quindi, se ho un dato mancante nell'ora che va dalle 8 alle 9 del mattino per la telecamera 89472, l'algoritmo calcola quanto è il peso percentuale di quella telecamera sul totale OTS di stazione rilevate in quella specifica ora di quello specifico giorno (calcola la media dei pesi % nello storico disponibile per quella telecamera, p.es. 1,4%).
2. Il totale delle OTS rilevate da tutte le telecamere funzionanti dalle 8 alle 9 del mattino del giorno in cui si è verificato il malfunzionamento sulla specifica telecamera nella stazione di riferimento (del giorno in cui vi è quindi il valore mancante). P.es. dalle 8 alle 9 di oggi nella stazione di Bari Centrale sono state rilevate 2.500 OTS.

Dalla moltiplicazione delle due si ottiene il valore di sostituzione per quella specifica telecamera in quello specifico intervallo temporale: la telecamera, nel suo storico, il lunedì dalle 8 alle 9 ha un peso dell'1,4% sul totale OTS della stazione, quindi 1,4% x 2.500 = 35 OTS.
È stata scelta questa procedura perché è molto solida anche rispetto a sostituzioni operate tramite Time Series analysis (attraverso il metodo robust STL decomposition e imputazione tramite interpolazione), in quanto, se il dato manca per molto tempo (p.es. 3 mesi), queste ultime non garantiscono elevata affidabilità con una tale granularità (per giorno e per ora). Il metodo di sostituzione appena spiegato permette invece di non sovrastimare il valore mancante (non si generano quindi valori anomali) perché il peso % della specifica telecamera viene sempre moltiplicato per le OTS rilevate realmente il giorno del malfunzionamento della telecamera in stazione. In questo modo se quel giorno la stazione è molto poco affollata per sciopero dei treni, per esempio, alla telecamera con valori mancanti verrà attribuita una % del totale delle OTS rilevate quel giorno, così da non sovrastimare il valore mancante. Le OTS_Clean vengono generate ogni notte non appena finisce lo scarico dati della tabella OTS.

3. Altri file necessari allo sviluppo del sistema

GSR_Database_impianti: il DB impianti è un file Excel in cui vengono costantemente aggiornati gli impianti presenti nelle stazioni, sia tradizionali (sprovvisti di telecamera) sia digitali (la maggior parte dotati di telecamere).
Attualmente, le telecamere di GSR sono all'incirca 900 Cam attive dislocate nelle 14 stazioni ferroviarie. Ad ogni asset sono associate le caratteristiche anagrafiche principali, tra cui la tipologia di prodotto (GOTV, Digitotem, Banner, MaxiLed etc.), la grandezza dello schermo, la categoria (digitale o tradizionale), la posizione all'interno delle aree piano della stazione, la data di installazione e lo stato attuale di attività dell'impianto (attivo o non attivo). Inoltre, per ciascun impianto digitale, sono presenti le location Quividi associate alle relative telecamere. È il file che incrocia quindi le informazioni di Broadsign, software utilizzato per l'erogazione delle pubblicità sugli impianti e le informazioni Quividi, software utilizzato per la rilevazione dell'audience.

Assunzioni del modello

Visibilità impianti: una prima assunzione di questo modello è stata fatta in merito alla visibilità degli impianti, dal momento che la distanza da cui un impianto può essere visto varia in base alla superficie dello schermo e dipende direttamente dalla visibilità dell'occhio umano. La formula utilizzata (che viene da ricerche americane sulla visibilità degli schermi digitali outdoor e validata nel paper Quividi) è: $\sqrt{superficie\ mq} \times 23$.

Dalle ricerche e dalle fonti esaminate è lecito assumere almeno un raddoppio della visibilità reale umana rispetto alla telecamera Quividi, passando così da 8 metri a 16 metri di distanza per uno schermo GOTV 42" (*Tabella 1*). Essendo l'area di visibilità riconducibile geometricamente ad una porzione di cerchio, all'aumentare del raggio, l'area aumenta più che proporzionalmente. Nello specifico, raddoppiando la misura del raggio, l'area del cerchio quadruplica, secondo la formula: $Area = raggio^2 \times \pi$. In questo modo, si assume che il rapporto tra la visibilità della tecnologia Quividi e la visibilità dell'occhio umano sia 1:4.

Inoltre, la visuale umana nota le presenze di impianti/persone/oggetti con un'ampiezza di quasi 180°, a differenza della telecamera Quividi che è in grado di cogliere gli stessi elementi con un angolo massimo di 70°. L'incremento dovuto alla maggiore ampiezza dell'angolo di visibilità umana viene applicato in maniera ridotta (un cono con un angolo al vertice di 140°) perché "la massima visibilità" per l'occhio umano è una situazione ideale che non rispecchia la realtà per diverse limitazioni (posizione osservatore molto laterale rispetto allo schermo e quindi impossibilità di leggere il contenuto se non ridirezionando lo sguardo, altre persone interposte, etc.). Il cono di visibilità con angolo di 140° è considerato lo standard Audioutdoor.

Per una maggiore coerenza con la realtà, sono state applicate anche delle correzioni ai coni di visibilità, tagliati sulla base di elementi che potrebbero ostacolare la visione degli schermi. Laddove vi siano ostacoli dati dalla struttura architettonica della stazione, come muri, negozi, colonne, etc., osservabili utilizzando le mappe, sono state apportate correzioni per singolo impianto per evitare che gli venissero attribuite delle OTS in aree dove l'impianto non è effettivamente visibile.

Tabella 1: Equivalenza fra pollici della diagonale schermo e superficie dello schermo in metri quadri

Formato 16:09	Mq	Distanza di Visibilità (m)
40 pollici	0.44	15,3
42 pollici	0.49	16,0
55 pollici	0.83	20,9
65 pollici	1.16	24,8
72 pollici	1.43	26,7
75 pollici	1.55	28,7
86 pollici	2.04	32,8

NB: la tabella non è esaustiva di tutti gli impianti presenti in GSR, è solo esemplificativa dell'applicazione della formula precedentemente descritta

Software utilizzati

1. SQL

Il linguaggio SQL è utilizzato per gestire il progetto di data quality, descritto in precedenza, tramite il quale si generano le OTS_Clean. Inoltre, le fonti dato primarie, quali OTS, OTS_Clean, Dati Telco, sono tutti presenti su DWH e vengono aggiornati con cadenza giornaliera. Tutte le procedure che generano l'audience sono salvate in molteplici step di vari processi SQL che aggiornano le strutture dell'audience con cadenza giornaliera (Individui e Audience). Mentre per la produzione del file degli incroci tra individui e impianti pubblicitari è stato utilizzato Python.

2. Python

Python è stato utilizzato perché più efficiente da un punto di vista computazionale per la produzione del file degli incroci che prevedeva molteplici loop per stazione, per circuito di asset e per ora della giornata. È stato utilizzato Python su Visual Studio in una Virtual Machine. Questa procedura, che scrive comunque la tabella di output (file degli incroci) su DWH, viene rilanciata all'occorrenza, come per esempio l'aggiunta, la dismissione o la modifica di alcune caratteristiche degli impianti pubblicitari o la modifica sostanziale di alcuni percorsi nelle stazioni per lavori o ristrutturazioni. Altrimenti il file è aggiornato secondo necessità, in particolare quando si raggiunge un certo numero di impianti da aggiungere/modificare e ricaricato su piattaforma Audioutdoor.

Output prodotti (con metodologia)

1. Tabelle delle dimensioni e Tracciato Record Impianti

Le tabelle delle dimensioni, utili alla creazione del file "Tracciato Record Impianti" sono state prodotte a partire dai file di riferimento di Audioutdoor, trovando le corrispondenze con il Database Impianti di GSR, contenente tutte le informazioni di anagrafica degli impianti. L'unico dato basato su un'assunzione riguarda il raggio nel file "Formati impianto", che è pari a $rad(base_m \times altezza_m) \times 23$, mentre l'ampiezza del cono di visibilità è impostata a 140 secondo lo standard di Audioutdoor.
NB: il calcolo $rad(base_m \times altezza_m) \times 23$ è assimilabile agli scaglioni di visibilità di Audioutdoor per gli impianti di piccole e medie dimensioni. Sebbene per gli impianti di grandi dimensioni quelli di Audioutdoor siano sensibilmente più elevati ciò non costituisce un problema perché in ambienti chiusi, seppur ampi, non è possibile raggiungere distanze di visibilità molto elevate per la presenza di ostacoli strutturali.

2. File degli individui

Per prima cosa è stata prodotta una tabella che replica i codici relativi alle provenienze, da associare a ciascun individuo rilevato tramite dati Telco. La creazione del file individui ha come punto di partenza una tabella con le informazioni fornite da telco Wind-Tre circa gli individui rilevati in ciascuna stazione a livello giornaliero, cui si associano la provenienza (con le specifiche illustrate precedentemente) e le informazioni relative al responsabile d'acquisto provenienti da GfK (tramite le distribuzioni sociodemografiche presenti in entrambe le fonti dato).
È stato creato un processo in SQL che consente di unire tutte le specifiche degli individui provenienti dalle tre fonti sopra citate e che raggruppa un numero totale di individui per tutte le combinazioni possibili di tali caratteristiche, in modo da produrre il file finale con una riga per ciascun individuo rilevato in ciascuna stazione in ciascun giorno del periodo considerato. Le informazioni attribuite a ciascun soggetto sono: genere, età, provenienza (ID_AREA e CAP) e responsabile d'acquisto donna che lavora e donna che non lavora per 2.016 celle GfK. A ciascun individuo è associato anche la fascia oraria di accesso in stazione (p.es. le 8 indica che l'individuo è entrato in stazione dalle 8:00 alle 8.59), seguendo la distribuzione oraria degli unici di ciascuna stazione fornita dal dato telco (vedi sezione DATI E FILE INPUT).
Ogni individuo, inoltre, viene associato solo ad un determinato percorso possibile nella stazione, o spazio chiuso, che ha visitato (vedi paragrafo successivo). Il percorso che compie il singolo individuo costituisce un ulteriore vincolo nella produzione del file degli incroci, perché facendo un determinato percorso quello specifico individuo potrà generare OTS solo sugli impianti presenti in quel percorso.

3. Creazione di percorsi per il file degli incroci

Questo è uno step propedeutico alla creazione del file degli incroci. È infatti fondamentale capire come si muovono gli individui all'interno degli spazi per definire delle associazioni tra individui e impianti che rispettino il più possibile la realtà. Operazione necessaria per la creazione di percorsi basati su un approccio matriciale delle aree in cui sono divisi gli spazi (piani o macroaree di ciascuna stazione):

1. Definizione di aree in tutti gli spazi di interesse aperti al pubblico. La suddivisione non deve essere guidata necessariamente da un criterio numerico, come per esempio la suddivisione in aree che abbiano all'incirca gli stessi metri quadri, ma piuttosto secondo logiche di accessibilità e funzionali (per esempio un corridoio chiuso

con due entrate viene diviso dagli atri che collega, oppure la zona dove arrivano un determinato tipo di treni, come i regionali, viene divisa da quella dei treni ad alta velocità). Per fornire un esempio concreto, il piano binari della stazione di Milano centrale è stato suddiviso in 6 aree, 3 aree fronte binari e 3 aree lungo i binari. A ciascuna delle suddette aree è stato dato un nome e sono state associate le celle di 25 metri quadri che compongono il reticolato calpestabile delle stazioni. Ciascuna di queste aree avrà quindi un numero di OTS generate in quell'area dal modello probabilistico utilizzato per l'attribuzione spaziale dell'audience a partire dalla sensoristica a disposizione della concessionaria. Con il sistema Total Audience[2], Il numero di OTS (la densità di OTS) presente in quell'area è funzione delle rilevazioni delle cam e del numero di impianti pubblicitari (anche tradizionali sprovvisti di cam) presenti in quella determinata area.

2. Censimento di tutti gli accessi possibili allo spazio di interesse. In questo modo si comprende quale delle aree definite al punto precedente costituiscono un punto di accesso/uscita oppure sono solo aree di passaggio all'interno di un possibile percorso effettuato da un individuo.
3. Determinazione del numero di persone che accedono ed escono da ciascun punto di accesso della stazione. Per esempio, fatto 100 il numero di individui presenti in stazione nel giorno medio, quanti accedono ed escono da ciascun punto di accesso?
4. Definizione della probabilità di accedere in ciascuna area limitrofa a quella di accesso. L'approccio è sequenziale, quindi una volta definite le aree immediatamente adiacenti a quella di accesso, è necessario associare delle probabilità di attraversamento di quelle aree (le aree adiacenti devono fare somma 100). Questo processo deve essere ripetuto per ogni area, definendo la probabilità di attraversare le aree successive immediatamente adiacenti (o eventualmente tornare indietro all'area di origine). Per esempio, se da una specifica area di accesso si può passare in altre 3 aree, è necessario stimare qual è la probabilità di accedere a ciascuna di quelle aree. Si veda per esempio l'*immagine 2*: in questa matrice sono riportati per semplicità solo due piani della stazione di Milano centrale (piano ammezzato, suddiviso in 3 aree, e piano binari suddiviso in 6 aree); le aree appartenenti a ciascun piano vengono riportate in una matrice, posizionandole sia sulle righe, sia sulle colonne. A queste viene aggiunta una riga evidenziata in giallo che rappresenta le aree in cui vi sono punti di accesso in stazione e una colonna, evidenziata in arancione, che rappresenta i punti di uscita della stazione (ne viene riportato solo uno, con la percentuale di individui che utilizza quella zona come uscita, pari all'8%, perché si vuole spiegare la dinamica di uno specifico percorso).

Immagine 2: esempio dell'approccio matriciale per la creazione di percorsi

Approccio Matriciale - Definizione dei percorsi tramite attraversamento probabilistico delle aree		**Piano ammezzato**			**Piano binari unito**						**ESCE**
		Ammezzato Centrale	**Ammezzato IV Novembre**	**Ammezzato Savoia**	**Binari Centrale**	**Binari IV Novembre**	**Binari Savoia**	**Mosaici Centrale**	**Mosaici IV Novembre**	**Mosaici Savoia**	**Punto di uscita**
ENTRA	**Punto di entrata**				**35%**	**20%**	**15%**		**15%**	**15%**	
Piano ammezzato	**Ammezzato Centrale**										
	Ammezzato IV Novembre										
	Ammezzato Savoia										
Piano binari unito	**Binari Centrale**					20%	20%	**50%**			***10%***
	Binari IV Novembre										
	Binari Savoia										
	Mosaici Centrale	10%		35%					**30%**	25%	
	Mosaici IV Novembre	20%	20%			25%		15%			***20%***
	Mosaici Savoia										

NOTA: nell'immagine 2 è definito un solo possibile percorso effettivo per facilitare la comprensione della metodologia applicata.

Nell'*immagine 2* è rappresentato uno specifico percorso. Come si vede nella prima riga evidenziata in giallo, in questa versione della stazione semplificata, vi sono 5 aree che contengono punti di accesso (caselle con presenza di una percentuale e che fanno 100% sulla riga) e 4 aree di transito (caselle senza indicazione di una percentuale). Nel percorso in esame uno specifico individuo entra dall'area Binari Centrale (scende da un treno in arrivo a Milano quindi). I colori

2 La metodologia Total Audience è stata validata da un ente esterno (Nutcracker) e presentata in un convegno al Politecnico di Milano. La metodologia sviluppata da GSR per l'attribuzione spaziale dell'audience e la relativa valorizzazione di qualsiasi elemento negli spazi calpestabili e aperti al pubblico, restituisce anche gli unici giornalieri per ogni «pacchetto commerciale» e ciascun impianto che compone il singolo pacchetto commerciale.

indicati nell'immagine servono a seguire l'individuo, quindi ora l'individuo va spostato sulla riga colorata in azzurro, che è appunto Binari Centrale. Da questa area può spostarsi in 4 aree:

- Uscire nuovamente dalla stazione dalla stessa area (riprende un treno nell'area Binari Centrale, 10% di probabilità);
- Andare in una delle aree limitrofe (prende un treno da Binari IV novembre o Binari Savoia, 20% di probabilità ciascuna) che costituiscono dei punti di accesso e quindi uscite dalla stazione;
- Andare in direzione del corpo della stazione, quindi accedere all'area Mosaici Centrale (50% di probabilità).

Anche in questo caso la riga fa somma 100% delle probabilità. Il percorso che stiamo prendendo in considerazione fa passare gli individui che lo percorrono verso l'area Mosaici Centrale, colorata in verde.
A questo punto dobbiamo spostarci nell'area Mosaici Centrale sulle righe per continuare il percorso. Da questa area gli individui possono accedere a 4 aree adiacenti:

- Cambiare piano utilizzando le scale o le scale mobili e accedere alle aree Ammezzato Centrale o Ammezzato Savoia (rispettivamente 10% e 35% di probabilità);
- Spostarsi nel piano nelle due aree limitrofe, Mosaici IV novembre o Mosaici Savoia (rispettivamente 30% e 25% di probabilità)

Ovviamente un possibile percorso avrebbe compreso la possibilità di tornare indietro verso i binari dopo aver visitato l'area centrale di galleria dei mosaici; quindi, anche altre celle dovrebbero essere valorizzate per completare la matrice e definire anche percorsi che tornano indietro. Nell'immagine sono state omesse le celle con percentuali visibili riproporzionate con somma 100 per facilitare la spiegazione della metodologia proposta.
Gli individui che seguono il percorso in oggetto si spostano nell'area di Mosaici IV novembre evidenziata in grigio. L'individuo ora potrebbe visitare le altre aree a cui può accedere dalla suddetta area, oppure uscire dalla stazione in quanto l'area in cui si trova costituisce un punto di accesso alla stazione (cella evidenziata in arancione nell'ultima colonna, con probabilità del 20%). Il percorso in oggetto prevede l'uscita dalla stazione e si conclude, come visibile nell'ultima colonna evidenziata in arancione.
In questo modo si definiscono tutti i possibili percorsi e si riempie la matrice. Ogni individuo del file individui verrà associato ad uno specifico percorso.

4. Note metodologiche integrative per la modellizzazione dei percorsi

Nella metodologia descritta è necessario minimizzare la probabilità che vi siano dei percorsi che ricadano in un "loop infinito" (sono invece ammessi "loop brevi"), quindi è stato inserito nel modello dei percorsi un abbattimento sulla probabilità di tornare in un'area già visitata. Questo abbattimento specifico per ogni spazio (singola stazione) deve essere studiato dalla concessionaria insieme con il comitato tecnico di Audioutdoor, in quanto dipende dalle dinamiche dei flussi all'interno dell'ambiente chiuso di interesse, dalla presenza o meno di percorsi relativamente obbligati e dalla tipologia di ambiente. Per esempio, nelle stazioni, chi ha 30 minuti di attesa tra un treno e l'altro visiterà probabilmente due volte le aree antistanti ai binari perché andrà a fare degli acquisti.

Per modellizzare alcuni di questi fenomeni ci si avvale di tutti i dati e conoscenze in possesso con l'obiettivo di avvicinarsi il più possibile alla descrizione reale del fenomeno dei percorsi. Per esempio, per gli ambienti delle stazioni ferroviarie vengono utilizzate le percentuali dei *train-train users*, ovvero persone che escono da un treno e lo riprendono nel giro di poco tempo per modellizzare quante persone dalle 3 aree in cui sono divisi i binari escono nuovamente da quelle zone. La percentuale di *train users* è però abbattuta, perché alcuni individui che hanno una coincidenza con più di 30 minuti di attesa seguono dei percorsi in stazione e non rimangono fermi.

Per validare il numero di accessi in alcuni punti degli ambienti chiusi di maggior interesse sono stati condotti anche studi con altre tecnologie (contapersone, rilevamento wi-fi e GPS).

Per la definizione della probabilità di passare ad una determinata area limitrofa data l'area in cui l'individuo è posizionato, sono state utilizzate le informazioni relative alle densità di OTS di ciascuna area fornita dal modello di attribuzione spaziale dell'audience. Quest'ultimo è stato utilizzato come una proxy della probabilità di attraversamento di una specifica area data la presenza originaria dell'individuo in un'area adiacente.

Questa illustrata è una possibile metodologia per la creazione di percorsi realistici in assenza della possibilità di rilevarli direttamente. La creazione o il rilevamento di percorsi è uno step fondamentale per l'associazione di OTS generate su determinati impianti da ogni individuo, altrimenti le associazioni delle OTS rilevate tramite sensoristica sarebbero casuali e non realistiche.
Qualora si avesse modo di rilevare direttamente i percorsi che compiono gli individui e i relativi flussi che li attraversano, non sarebbe necessario sviluppare un modello probabilistico come quello matriciale descritto nel presente documento.

File degli incroci

Il file degli incroci crea il raccordo tra gli individui identificati tramite la creazione del file al punto precedente e le OTS generate su ciascun impianto in ciascuna ora media del giorno medio. Dunque, gli input in questa fase sono il file degli individui e una tabella che contiene le OTS orarie di ogni impianto pubblicitario del giorno medio del periodo considerato.
Di seguito le informazioni necessarie alla produzione del file degli incroci:

1. File OTS:
   - Posizione dell'impianto all'interno della stazione (coordinate)
   - Area di stazione in cui è presente l'impianto (vedi paragrafo precedente sui percorsi)
   - Numero di OTS generate ogni ora da ciascun impianto (con applicazione di una curva logaritmica descritta nel paragrafo "*Adjustment logaritmico delle OTS propedeutico alla produzione del file degli incroci*")

Le suddette informazioni provengono dalle telecamere presenti nell'ambiente di interesse che sono dotate di tecnologia Quividi e dalla mappatura degli impianti pubblicitari sulle planimetrie degli spazi calpestabili aperti al pubblico.

Adjustment logaritmico delle OTS propedeutico alla produzione del file degli incroci

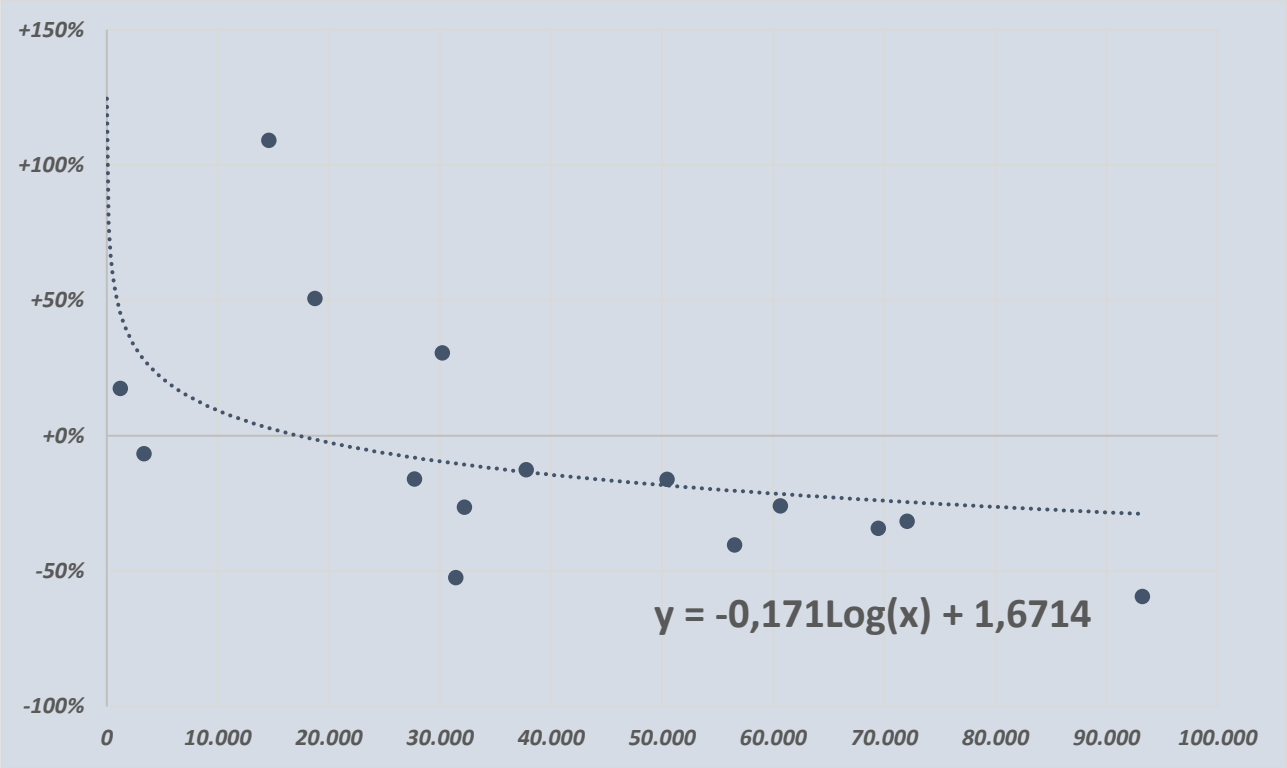
Prima dell'esecuzione dello script Python per la produzione del file degli incroci, è stato introdotto uno step di adjustment delle OTS applicato alla tabella delle OTS orarie per impianto. Tale passaggio è propedeutico alla generazione degli incroci, in quanto il file delle OTS rappresenta uno degli input principali dell'algoritmo che associa ciascuna OTS a uno specifico individuo, nel rispetto dei vincoli dati dalla fascia oraria e dal percorso assegnato.
L'esigenza di introdurre questo adjustment nasce dalla necessità di anticipare quello che produrrà il cambio di algoritmo Quividi (un nuovo algoritmo più preciso di quello attuale che verrà introdotto dalla concessionaria nei prossimi mesi) e al contempo rendere più graduale la correzione delle OTS rispetto a una logica per fasce, che avrebbe prodotto effetti discontinui e potenzialmente troppo impattanti al passaggio da una classe all'altra. Per questo motivo è stata adottata una funzione logaritmica interpolante, costruita a partire dalle stime fornite da Quividi in fase di test che simula ciò che accadrà con l'adozione del nuovo algoritmo (con esclusione degli outlier). La funzione consente di applicare una correzione continua: gli impianti con valori di OTS più contenuti ricevono un adjustment positivo, mentre gli impianti con valori di OTS più elevati ricevono un adjustment progressivamente negativo. In questo modo si ottiene un effetto più "morbido", senza gradini, e si contribuisce a mitigare il fenomeno delle frequenze elevate.
La funzione applicata è la seguente:

$$Y = -0{,}171 \times \log(OTS\ impianto) + 1{,}6714 - 1 + K$$

Dove:
- Y rappresenta la percentuale di adjustment da applicare alle OTS dell'impianto;
- La componente "-1" consente di trasformare la funzione interpolante in una variazione percentuale;
- K nella versione attuale della metodologia è posto pari a 0, così da rispettare la funzione interpolante originaria senza traslazioni della curva.

Operativamente, la curva viene applicata alla tabella delle OTS prima della produzione del file degli incroci. Di conseguenza, lo script Python non utilizza più direttamente le OTS originarie, ma le OTS già corrette tramite adjustment logaritmico. Solo successivamente ciascuna riga del file OTS viene duplicata in funzione del numero di OTS corrette e associata agli individui filtrati per stazione, fascia oraria e percorso (come dettagliato in seguito). Questo consente di mantenere invariata la logica metodologica di associazione tra individui e impianti, introducendo però una normalizzazione preliminare dei volumi OTS coerente con gli effetti attesi dell'evoluzione dell'algoritmo Quividi.

Immagine 3: curva logaritmica utilizzata per l'adjustment Quividi (Stima cambio dell'algoritmo)



La curva rappresenta l'andamento dell'adjustment applicato alle OTS degli impianti. All'aumentare delle OTS rilevate, la percentuale di correzione tende progressivamente a ridursi e a diventare negativa, consentendo di abbassare i valori degli impianti con OTS più elevate e di incrementare quelli con OTS più contenute.

2. File Individui:

- Informazioni sociodemografiche e comportamentali degli individui (ogni riga di questo file è un individuo);
- Fascia oraria di accesso in stazione;
- Percorso assegnato all'individuo.

3. Tabella ponte Percorsi-Aree-Impianti:

Ogni impianto è in una determinata area di stazione, ma ovviamente più percorsi attraversano la stessa area, e quindi si rende necessaria una struttura che associ a ciascun percorso le aree attraversate e gli asset presenti in quelle aree. Questa struttura costituisce un vincolo fondamentale nella produzione degli incroci perché consente ad un individuo che fa un determinato percorso di incontrare solo gli impianti effettivamente presenti in quel percorso.
La difficoltà principale consiste nell'attribuire le OTS a delle interazioni specifiche nel tempo con gli individui «unici» rilevati tramite tecnologia telco e attribuiti tramite adattamento della metodologia di calcolo degli unici. Per rispondere alla complessità di questa operazione si è deciso di costruire un algoritmo su Python.
Quest'ultimo, tramite dei loop, prende in considerazione una specifica riga del file delle OTS alla volta, quindi un determinato impianto di una stazione, una determinata ora e le OTS generate da quell'asset in quell'ora (p.es. il Digitotem 03 di Milano ha generato 100 OTS tra le 8:00 e le 8:59 di mattina). Questa specifica riga viene duplicata N volte, dove N è il numero di OTS (in questo caso pari a 100). La procedura a questo punto seleziona solo gli unici presenti a Milano in quella determinata fascia oraria dal file degli individui. Tramite la tabella ponte vengono ulteriormente filtrati solo gli individui associati ad un percorso che incrocia l'area dove è situato il Digitotem 03. A questo punto l'algoritmo procede con una estrazione casuale con reinserimento dal set di individui filtrato che viene ripetuta N volte, dove N è il numero di OTS di quell'impianto in quell'ora. Come da esempio, in questo caso le righe (OTS) che necessitano dell'associazione a

degli individui che la abbiano generate sono 100. Gli individui a cui sono associati percorsi più lunghi sono presenti in più loop e quindi generano coerentemente più OTS. Quando ciascuna riga (OTS) ha un individuo associato la procedura continua il loop con un nuovo asset o con lo stesso in una diversa ora della giornata. L'esecuzione dello script python si arresta quando ciascuna OTS ha un individuo associato.

4. Finalizzazione file degli incroci con ultime informazioni temporali

Ogni OTS ora ha un individuo specifico che l'ha generata, ma manca ancora il tempo di inizio e fine della specifica OTS e la relativa durata. Per generare i suddetti timestamp con una metodologia probabilistica che risultasse il più coerente possibile, si è proceduto in tre step:

- Associazione del tempo di permanenza totale nello spazio di interesse (stazione) a ciascun individuo con relativo timestamp di entrata e di uscita. Per generare questi campi è stato utilizzato come parametro noto il tempo medio di permanenza (dato telco) degli individui per ciascuna stazione. È stata costruita quindi una distribuzione normale per ogni stazione con media uguale al tempo di permanenza medio del dato telco. Il tempo di permanenza minimo è settato a 60 secondi, mentre il tempo massimo varia in base alla dimensione dell'ambiente chiuso di interesse (dimensione dei percorsi e delle stazioni). È presente un vincolo nella generazione del tempo di permanenza di ciascun individuo: persone che effettuano percorsi più lunghi e hanno generato più OTS vengono in media associate ad un tempo di permanenza superiore alla media della distribuzione, viceversa per chi effettua percorsi più brevi. L'associazione del tempo di permanenza a ciascun individuo è propedeutica alla creazione dei timestamp delle singole OTS.
- Ora che il file degli incroci è stato arricchito con i timestamp di entrata e di uscita di ciascun individuo è possibile generare dei timestamp di inizio e fine per ciascuna OTS che siano compresi nell'intervallo di permanenza dello specifico individuo. Questi timestamp non sono mutualmente esclusivi (le durate di due OTS possono sovrapporsi per lo stesso individuo) e seguono la logica del percorso seguito dall'individuo. Se il percorso inizia nell'area 3 e finisce in area 5 l'individuo incontrare prima gli asset presenti in area 3 e poi quelli presenti in area 5 e i timestamp delle OTS rispetteranno questo vincolo.
- Applicazione del CAP sulle frequenze residue. A valle dell'adjustment logaritmico delle OTS viene inoltre applicato un controllo finale sulle frequenze residue, tramite l'introduzione di un CAP pari a 10 sui singoli impianti. Il CAP non viene applicato a livello di pacchetto commerciale, poiché un pacchetto composto da numerosi impianti può legittimamente generare una frequenza superiore a 10 qualora un percorso intercetti più volte impianti appartenenti allo stesso circuito. Il controllo viene invece effettuato a livello di singolo impianto, con l'obiettivo di limitare eventuali casi residui di frequenza anomala. La frequenza di ciascun impianto viene calcolata come rapporto tra le OTS generate dall'impianto e gli individui unici associati:

  $$Frequenza_{impianto} = {}^{OTS_{impianto}}\Big/_{Unici_{impianto}}$$

  Se tale valore risulta superiore a 10, le OTS dell'impianto vengono ridotte fino al valore massimo compatibile con una frequenza pari a 10. Si eliminano le interazioni (righe di incroci) di individui che vedono lo stesso asset più di una volta, conservando quindi almeno un'interazione tra individuo e impianto così da mantenere invariato il numero di unici che intercetta quello specifico impianto.
  Il CAP agisce quindi esclusivamente in riduzione e non incrementa mai le OTS. Questo passaggio ha la funzione di contenere i casi estremi eventualmente non risolti dall'adjustment logaritmico, garantendo maggiore coerenza del dato finale prima del caricamento delle strutture in piattaforma Audioutdoor.

Il file degli incroci è a questo punto completo, si filtrano le informazioni superflue ed è pronto, previo controlli di coerenza automatizzati, ad essere caricato in piattaforma Audioutdoor.
Queste procedure di creazione timestamp temporali e controllo delle frequenze sono implementate anch'esse tramite script Python.