

UNIVERSITÀ DEGLI STUDI DI  
NAPOLI FEDERICO II



CORSO DI LAUREA MAGISTRALE IN INGEGNERIA  
INFORMATICA

DIPARTIMENTO DI INGEGNERIA ELETTRICA E DELLE  
TECNOLOGIE DELL'INFORMAZIONE

**Information Systems and  
Business Intelligence**  
Trend Analysis - Agritech Domain

**Candidati:**

Vincenzo D'Angelo M63001595  
Giorgio Di Costanzo M63001579

**Professoressa:**

Flora Amato

Anno Accademico 2024/2025

# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Trasformazione Dati</b>	<b>1</b>
2.1	Dati meteo . . . . .	1
2.1.1	Aggregazione giornaliera Dati Meteo . . . . .	2
2.2	Dati catture . . . . .	3
<b>3</b>	<b>Relazioni tra tabelle</b>	<b>4</b>
<b>4</b>	<b>Visualizzazione dei dati</b>	<b>6</b>

# 1 Introduzione

L'obiettivo di questo documento è effettuare un'analisi approfondita delle informazioni presenti nei dataset di *Imola* e *Cicalino*, di cui si è ampiamente discusso in precedenza (si rimanda alla documentazione). L'analisi si concentra sull'esplorazione delle relazioni tra le variabili, l'identificazione di pattern significativi e la visualizzazione dei dati in modo sintetico ed efficace, al fine di supportare ulteriori decisioni e modelli predittivi.

l'analisi è stata condotta avvalendosi di Power BI [1], uno strumento di business intelligence sviluppato da Microsoft, progettato per analizzare e visualizzare dati in modo interattivo. Consente agli utenti di connettersi a diverse fonti di dati, trasformarli, analizzarli e presentarli tramite report e dashboard dinamici.

## 2 Trasformazione Dati

I file sono stati forniti in formato HTML, pertanto si è reso necessario un passaggio preliminare di estrazione delle tabelle. I dati sono stati successivamente sottoposti ad opportune modifiche, tra cui operazioni di casting, aggregazioni e ristrutturazioni delle informazioni, per garantire coerenza e adeguatezza alle analisi successive.

### 2.1 Dati meteo

Una volta caricato, il dataset dei dati meteo si presenta nella seguente struttura.

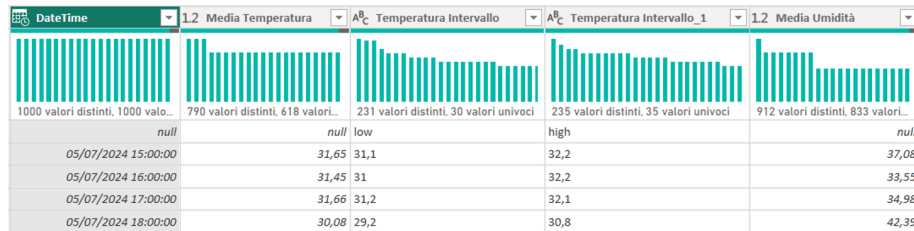


Figura 1: Dati meteo grezzi

Per renderlo adeguato all'analisi sono stati eseguiti i seguenti passaggi:

- Eliminazione della prima riga;
- Separazione della colonna DateTime in due colonne distinte: Data e Ora;
- Conversione dei tipi di dati (casting):
  - Data: Testo → Data
  - Ora: Testo → Ora
  - Temperatura intervallo Low: Testo → Numero decimale

- Temperatura intervallo High: Testo → Numero decimale
- Ridenominazione delle colonne.

	Data	Ora	1.2 Media Temperatura	1.2 Temperatura Intervallo Low	1.2 Temperatura Intervallo High	1.2 Media Umidità
1	05/07/2024	15:00:00	31,65	31,1	32,2	37,08
2	05/07/2024	16:00:00	31,45	31	32,2	33,55
3	05/07/2024	17:00:00	31,66	31,2	32,1	34,98
4	05/07/2024	18:00:00	30,08	29,2	30,8	42,39
5	05/07/2024	19:00:00	28,78	26,5	29,6	47,54
6	05/07/2024	20:00:00	24,26	22,4	26	59,08
7	05/07/2024	21:00:00	20,1	18,6	21,9	76,51
8	05/07/2024	22:00:00	17,6	17	18,4	87,61
9	05/07/2024	23:00:00	16,42	16	16,9	91,33
10	06/07/2024	00:00:00	15,44	14,9	15,9	92
11	06/07/2024	01:00:00	14,72	14,6	14,9	93,9
12	06/07/2024	02:00:00	14,31	14,1	14,5	95,22

Figura 2: Dati meteo processati

### 2.1.1 Aggregazione giornaliera Dati Meteo

Successivamente, per ogni tabella, è stata creata una nuova versione che aggrega le informazioni su base giornaliera.

Variabile	Definizione
Temperatura Media	Media dei valori di temperatura registrati nelle ore del giorno.
Temperatura Massima	Valore massimo delle temperature registrate durante le ore del giorno.
Temperatura Minima	Valore minimo delle temperature registrate durante le ore del giorno.
Umidità Media	Media dei valori di umidità registrati nelle ore del giorno.
Escursione Termica	Differenza tra i valori di temperatura massima e minima registrati durante le ore del giorno.
Escursione Umidità	Differenza tra i valori di umidità massima e minima registrati durante le ore del giorno.

Tabella 1: Descrizione delle variabili meteorologiche giornaliere

	Data	Ora	1.2 Media Temperatura	1.2 Temperatura Intervallo Low	1.2 Temperatura Intervallo High	1.2 Media Umidità
1	05/07/2024	15:00:00	31,65	31,1	32,2	37,08
2	05/07/2024	16:00:00	31,45	31	32,2	33,55
3	05/07/2024	17:00:00	31,66	31,2	32,1	34,98
4	05/07/2024	18:00:00	30,08	29,2	30,8	42,39
5	05/07/2024	19:00:00	28,78	26,5	29,6	47,54
6	05/07/2024	20:00:00	24,26	22,4	26	59,08
7	05/07/2024	21:00:00	20,1	18,6	21,9	76,51
8	05/07/2024	22:00:00	17,6	17	18,4	87,61
9	05/07/2024	23:00:00	16,42	16	16,9	91,33
10	06/07/2024	00:00:00	15,44	14,9	15,9	92
11	06/07/2024	01:00:00	14,72	14,6	14,9	93,9
12	06/07/2024	02:00:00	14,31	14,1	14,5	95,22

Figura 3: Dati meteo aggregati

## 2.2 Dati catture

Una volta caricato, il dataset dei dati sulle catture si presenta nella seguente struttura:

A <sup>B</sup> <sub>C</sub> DateTime	A <sup>B</sup> <sub>C</sub> Numero di insetti	A <sup>B</sup> <sub>C</sub> Nuove catture (per evento)	A <sup>B</sup> <sub>C</sub> Recensito	A <sup>B</sup> <sub>C</sub> Evento
58 valori distinti, 58 valori univ...	5 valori distinti, 1 valori univ...	4 valori distinti, 1 valori univoci	1 valori distinti, 0 valori univ...	2 valori distinti, 0 valori univ...
06.07.2024 06:01:00	0	0	Si	
07.07.2024 06:04:00	0	0	Si	
08.07.2024 06:03:00	0	0	Si	
09.07.2024 06:05:00	0	0	Si	
09.07.2024 06:13:59			Si	Cleaning

Figura 4: Dati meteo grezzi

Per renderlo adeguato all'analisi sono stati eseguiti i seguenti passaggi:

- Divisione della colonna DateTime in Data e Ora;
- Rimozione colonna Ora;
- Sostituzione del valore nella colonna Recensito: 'Si' è stato sostituito con 1;
- Sostituzione del valore nella colonna Evento: 'Cleaning' è stato sostituito con 1 e le celle vuote sono state impostate a 0;
- Sostituzione del valore nelle colonne Nuove catture e Numero di insetti: i valori 'null' sono stati sostituiti con 0;
- Conversione dei tipi di dati (casting):
  - Data: Testo → Data
  - Numero Insetti: Testo → Numero intero
  - Nuove catture (per evento): Testo → Numero intero
  - Recensito: Testo → Numero intero
  - Evento: Testo → Numero intero
- Combinazione delle righe in cui l'evento è pari a '1' (Cleaning) con la riga precedente, che rappresenta lo stesso giorno.

	1 <sup>2</sup> <sub>3</sub> Data	1 <sup>2</sup> <sub>3</sub> Numero di insetti	1 <sup>2</sup> <sub>3</sub> Nuove catture (per evento)	1 <sup>2</sup> <sub>3</sub> Recensito	1 <sup>2</sup> <sub>3</sub> Evento
1	06/07/2024	0	0	1	0
2	07/07/2024	0	0	1	0
3	08/07/2024	0	0	1	0
4	09/07/2024	0	0	1	1
5	10/07/2024	0	0	1	0
6	11/07/2024	0	0	1	0
7	12/07/2024	0	0	1	0
8	13/07/2024	0	0	1	0
9	14/07/2024	1	1	1	0

Figura 5: Dati meteo processati

### 3 Relazioni tra tabelle

Un grafico delle relazioni è una visualizzazione che mostra come le tabelle sono collegate tra loro all'interno di un modello di dati.

Ogni tabella nel grafico è rappresentata da un blocco rettangolare che mostra i campi (colonne) contenuti all'interno della tabella. Le linee che collegano le tabelle indicano le relazioni tra esse. Queste linee rappresentano i collegamenti tra colonne di diverse tabelle, basate su chiavi comuni. Le linee che collegano le tabelle mostrano anche la cardinalità delle relazioni. Le linee delle relazioni spesso sono orientate per indicare la direzione in cui i dati possono essere filtrati o aggregati durante le analisi.



Figura 6: Relazioni tra tabelle

Le tabelle *Dati meteo storici* e *Dati meteo storici (2)* sono collegate rispettivamente alle tabelle *Grafico delle catture* e *Grafico delle catture (2)* tramite la colonna *Data*. Poiché i dati meteo sono organizzati su base oraria mentre i dati delle catture sono su base giornaliera, la cardinalità della relazione è molti a uno. Tale relazione consente di effettuare analisi incrociate tra i dati meteo e i dati sulle catture, permettendo di esplorare come le condizioni meteo orarie influenzano le catture giornaliere.

Le tabelle *Dati meteo storici* e *Dati meteo storici (2)* sono collegate rispettivamente alle tabelle *Dati meteo storici aggregati* e *Dati meteo storici aggregati 2* tramite la colonna *Data*. Poiché i dati meteo nelle prime due tabelle sono organizzati su base oraria, mentre nelle tabelle aggregate sono organizzati su

base giornaliera, la cardinalità della relazione è molti a uno. Questa relazione consente di eseguire analisi che combinano i dati meteo orari con quelli giornalieri, permettendo una visione più dettagliata e completa delle informazioni meteo nel tempo.

Le Tabelle *Dati meteo storici aggregati* e *Dati meteo storici aggregati 2* sono collegate tra di loro tramite la colonna `Data`. Poiché entrambe le tabelle contengono dati aggregati su base giornaliera, la cardinalità della relazione è uno a uno. Questa relazione consente di eseguire un'analisi comparativa tra i dati meteo presenti nel primo e nel secondo dataset, entrambi organizzati e aggregati a livello giornaliero.

Le tabelle *Dati meteo storici aggregati* e *Dati meteo storici aggregati 2* sono collegate rispettivamente alle tabelle *Grafico delle catture* e *Grafico delle catture (2)* tramite la colonna `Data`. Poiché entrambe le tabelle contengono dati aggregati su base giornaliera, la cardinalità della relazione è uno a uno. Tale relazione consente di effettuare analisi incrociate tra i dati meteo aggregati e i dati sulle catture, permettendo di esplorare come le condizioni meteo giornaliere influenzano le catture giornaliere.

Le Tabelle *Grafico delle catture* e *Grafico delle catture 2* sono collegate tra di loro tramite la colonna `Data`. Poiché entrambe le tabelle contengono dati aggregati su base giornaliera, la cardinalità della relazione è uno a uno. Questa relazione consente di analizzare in modo congiunto i dati delle catture provenienti dalle due tabelle, facilitando confronti e approfondimenti tra le diverse informazioni sulle catture raccolte.

## 4 Visualizzazione dei dati

Le visualizzazioni in Power BI trasformano i dati complessi in informazioni facilmente comprensibili, utilizzando grafici, diagrammi e mappe. Questi strumenti offrono una panoramica chiara e immediata di pattern, tendenze e anomalie, rendendo più facile l'interpretazione di grandi volumi di dati. Inoltre, Power BI consente di interagire con i dati in tempo reale, offrendo all'utente la possibilità di selezionare, filtrare ed esplorare i dati in modo approfondito. Questo approccio dinamico permette di adattare l'analisi alle necessità specifiche, garantendo un'analisi più dettagliata e flessibile.

Il grafico che segue mostra l'andamento della Temperatura Minima (in blu chiaro) e della Temperatura Massima (in blu scuro) per ogni giorno del dataset *Dati meteo storici aggregati*.

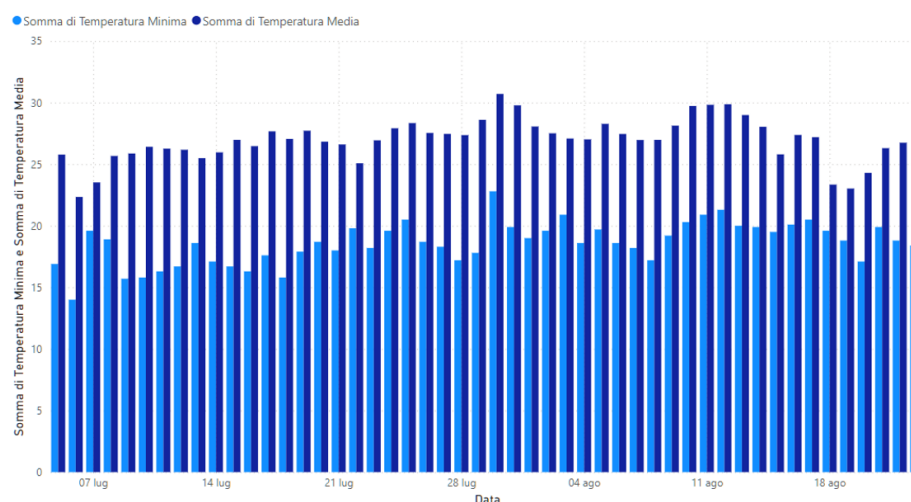


Figura 7: Temperatura minima e temperatura massima per data

Si osserva un'oscillazione costante sia nella temperatura minima che nella temperatura massima nel corso delle settimane.

Tra il 28 luglio e l'11 agosto, si registrano valori più elevati nella temperatura massima, con picchi particolarmente visibili. Durante questo periodo, anche la temperatura minima mostra un incremento. Dopo il 14 agosto, si osserva un calo più evidente nei valori massimi, mentre i valori minimi rimangono stabili. La differenza tra le temperature massime e minime tende a variare, indicando giorni con maggiore o minore escursione termica.

Questa visualizzazione permette di individuare facilmente periodi con temperature più elevate, monitorare l'escursione termica giornaliera e comprendere l'andamento delle temperature minime e massime nel corso del tempo.



I grafici sottostanti rappresentano la Media di Temperatura e la Media di Umidità per ora del dataset *Dati meteo storici*, evidenziando l'andamento di questi due indicatori durante le 24 ore della giornata.

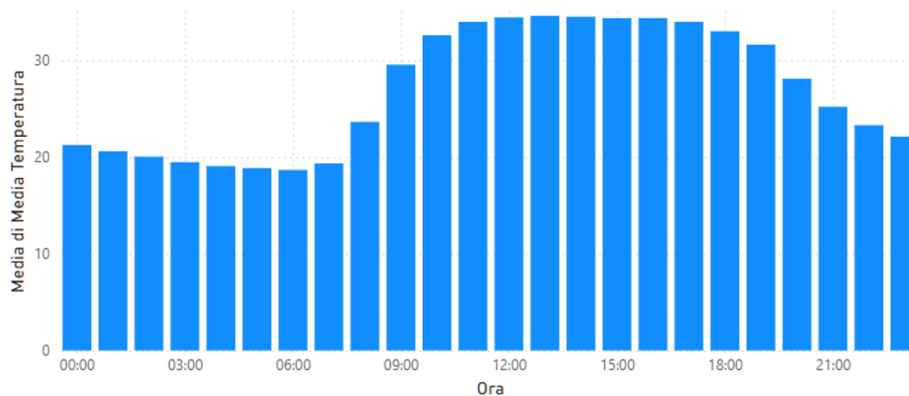


Figura 8: Temperatura media per data

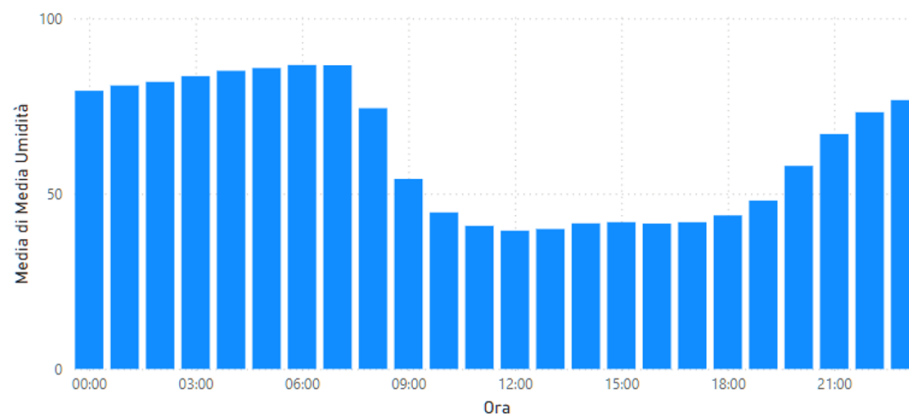


Figura 9: Umidità media per data

Il grafico in alto mostra l'andamento della temperatura media nelle diverse ore del giorno. Come evidenziato, la temperatura tende ad aumentare progressivamente durante la mattinata, raggiungendo un picco nel pomeriggio, per poi diminuire la sera e la notte.

Il grafico in basso illustra l'andamento della media dell'umidità nelle diverse ore. Si osserva che l'umidità è generalmente più alta nelle prime ore del mattino e durante la notte, per poi diminuire durante il giorno, con i valori più bassi nel pomeriggio.

Entrambi i grafici sono utili per osservare i pattern temporali dei dati meteo, in particolare per capire come la temperatura e l'umidità variano durante il giorno.

Gli scatterplot riportati mettono in relazione alcune informazioni del dataset *Dati meteo storici aggregati* al numero di nuove catture. Ogni punto dei grafici rappresenta un evento di cattura e la dimensione del cerchio è proporzionale al numero di nuove catture.

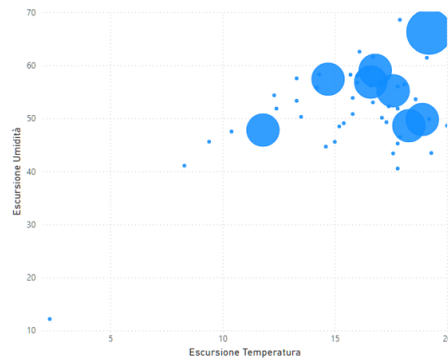


Figura 10: Nuove catture per escursione termica ed escursione umidità

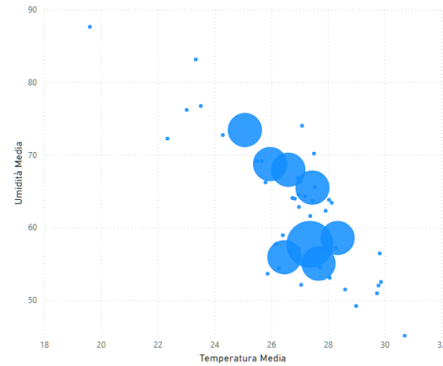


Figura 11: Nuove catture per temperatura media ed umidità media

Questi grafici sono utili per visualizzare eventuali correlazioni tra i dati meteo (come temperatura, umidità, ed escursione) e le catture di insetti. Analizzando la disposizione e la dimensione dei cerchi, è possibile identificare tendenze, come ad esempio quali condizioni meteorologiche potrebbero essere associate a un numero maggiore di catture.

Il grafico a sinistra spiega come l'escursione della temperatura e l'escursione dell'umidità possano o meno influenzare le nuove catture. Si può osservare che la dimensione dei cerchi cresce nella sezione in alto a destra del grafico. Questo suggerisce che eventi con una maggiore escursione sia della temperatura che dell'umidità possano essere associati a un numero maggiore di catture.

Il grafico a destra illustra l'influenza della temperatura media e dell'umidità media sulle nuove catture. Si nota chiaramente che le catture tendono a concentrarsi in corrispondenza di temperature medie comprese tra i 25 e i 29 gradi Celsius e livelli di umidità media tra il 55% e il 75%. Questo evidenzia come queste condizioni climatiche specifiche possano avere un impatto diretto sull'aumento delle catture, suggerendo una possibile correlazione tra queste variabili ambientali e il numero di insetti catturati.

I grafici ad aree in esame riportano i seguenti dati combinati:

- In Viola: Deviazione Standard dell'Umidità giornaliera.
- In Blu: Deviazione Standard della Temperatura giornaliera.
- In Arancione: Nuove catture avvenute nel giorno.

Il grafico in alto illustra le informazioni contenute nei dataset *Dati Meteo Storici* e *Grafico delle catture*, mentre il grafico in basso fa riferimento ai *Dati Meteo Storici (2)* e *Grafico delle catture (2)*.

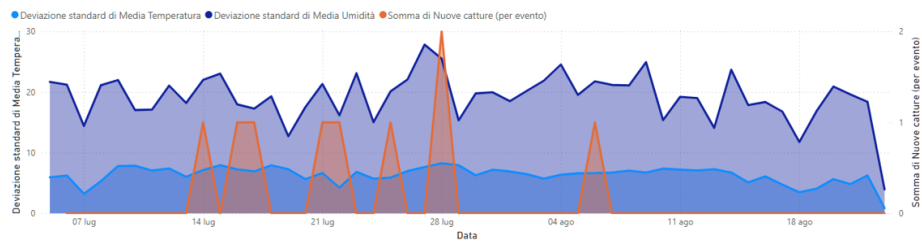


Figura 12: deviazione standard dell'umidità giornaliera, deviazione standard della temperatura giornaliera e nuove catture avvenute nel giorno per data - 1

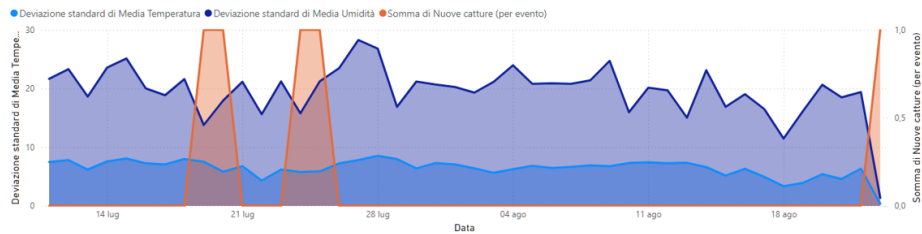


Figura 13: deviazione standard dell'umidità giornaliera, deviazione standard della temperatura giornaliera e nuove catture avvenute nel giorno per data - 2

L'andamento delle Deviazioni Standard delle due variabili è generalmente regolare; tuttavia, si osserva che nella maggior parte dei casi le catture avvengono quando le deviazioni standard superano il loro valore medio.

È inoltre evidente che gli ultimi punti potrebbero rappresentare degli outlier, poiché le deviazioni standard di umidità e temperatura mostrano valori molto bassi rispetto al resto del grafico. Questo comportamento anomalo potrebbe essere dovuto al fatto che non sono stati raccolti campioni per tutte le fasce orarie.

I grafici riportati di seguito illustrano i seguenti dati combinati:

- In Viola: Il numero di insetti
- In Arancione: Le nuove catture avvenute nel giorno
- I picchi Blu: La presenza dell'evento di Cleaning

Il grafico in alto illustra le informazioni contenute nel dataset *Grafico delle catture*, mentre il grafico in basso fa riferimento a *Grafico delle catture (2)*.

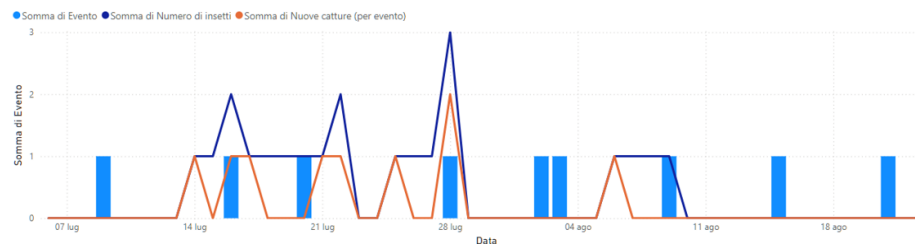


Figura 14: Numero di insetti, nuove catture avvenute nel giorno e presenza di evento di Cleaning - 1

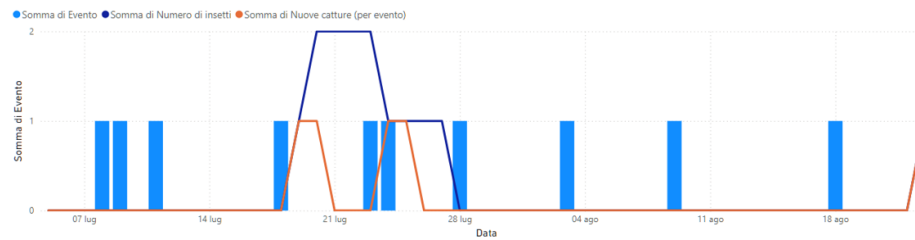


Figura 15: Numero di insetti, nuove catture avvenute nel giorno e presenza di evento di Cleaning - 2

Si può notare come il numero di insetti sia fortemente influenzato dall'evento di Cleaning, mentre non ha alcuna influenza sulle catture future degli insetti.

## Riferimenti

- [1] Microsoft Corporation. *Microsoft Power BI*. Business Intelligence and Data Visualization Tool. 2025. URL: <https://powerbi.microsoft.com/>.