



UNIVERSITA' DEGLI STUDI DI
NAPOLI FEDERICO II

Scuola Politecnica e delle Scienze di Base
Corso di Laurea Magistrale in Ingegneria Informatica

Information Systems and Business Intelligence

ANALISI DATASET "temp_humid_data"

Anno Accademico [2023-2024]

Prof.ssa Flora Amato

Studenti
Riccardo Romano
matr. M63001489

Giovanni Riccardi
matr. M63001480

Francesco Panariello
matr. M63001433

Indice

Indice.....	II
Traccia Elaborato PowerBI.....	3
Scopo Analisi	4
Metodologia di Sviluppo.....	4
Analisi Sheet1	5
Anteprima dei Dati e Prime Analisi.....	5
Analisi Giornate Calde e Giornate Fredde.....	6
Analisi Giornate Umide e Giornate Secche	10
Analisi del Cambiamento di Temperatura in base al Mese	12
Analisi Dati Sheet3	15
Anteprima dei Dati e Prime Analisi.....	15
Analisi Mediante Utilizzo del Widget ChatBot.....	16
Analisi Approfondita su Adult Males	18
Analisi Predittiva di no. Adult Males.....	20
Confronto tra Sheet1 e Sheet3	22
Conclusioni	23

Traccia Elaborato PowerBI

Power BI su Dataset Allegato:

- Utilizza Microsoft Power BI per creare report e dashboard.
- Include diverse visualizzazioni di dati come grafici, grafici a torta, mappe di calore, ecc.
- Esplora le funzionalità di Power BI come la funzione chatbot (tasto apposito in Power BI), analisi predittiva se appropriato.
- Assicurati che il report sia ben organizzato e presenti i dati in modo chiaro e comprensibile.

Scopo Analisi

Questo studio presenta un'analisi pluriennale condotta su un dataset denominato "temp_humid_data". Esso comprende dati relativi alla temperatura, all'umidità e alla presenza di adult males, raccolti durante un arco temporale di due anni. Lo scopo principale di questo studio è esplorare le relazioni, le tendenze stagionali e le possibili implicazioni derivanti da questi attributi.

Metodologia di Sviluppo

L'analisi dei dati è stata condotta attraverso l'impiego della piattaforma PowerBI, un potente strumento di analisi e visualizzazione dei dati.

Il dataset "temp_humid_data" è suddiviso in due componenti principali: "Sheet1" e "Sheet3", i quali si differenziano soprattutto per il fatto di essere selezionati su due annate differenti, ovvero 2022 per il primo e 2023 per il secondo. Questa distinzione ha permesso di concentrare l'analisi sui due componenti, evidenziandone le particolarità specifiche.

Analisi Sheet1

Lo Sheet1 è composto da dati quali: time, temperature_mean, humidity_mean.

A partire da questi dati sono state fatte analisi accurate su diversi aspetti, come: l'andamento mediante grafico dei due attributi, l'osservazione dei valori medi ed uno studio sui giorni più caldi/freddi e sui giorni più umidi/secchi. Tali dati sono riferiti all'anno 2022.

Anteprima dei Dati e Prime Analisi

Come prima cosa, è stata resa visibile un'anteprima dei dati per poter osservare in maniera dinamica e diretta la composizione dello Sheet1 del Dataset.

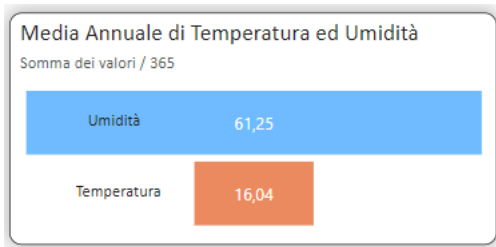
I dati sono stati resi osservabili mediante PowerBI attraverso l'utilizzo di una tabella, ecco come sono resi disponibili i dati:

Anno	Mese	Giorno	Somma di relativehumidity_mean	Somma di temperature_mean
2022	marzo	31	94	10,34
2022	dicembre	24	94	9,34
2022	maggio	6	93	13,89
2022	dicembre	10	93	5,40
2022	dicembre	22	93	8,71
2022	dicembre	31	93	9,85

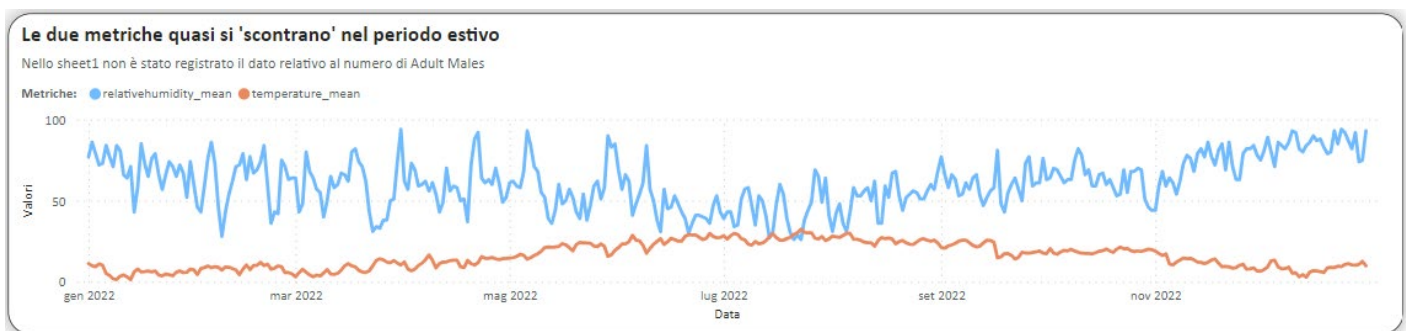
A partire da questi dati, sono state conseguite le prime analisi come:

- l'osservazione della media annuale di Temperatura ed Umidità
- l'osservazione dell'andamento di Temperatura ed Umidità

La rappresentazione della media annuale è stata resa disponibile mediante il tool “Imbuto” presente su PowerBI, in modo da poter visualizzare con facilità e praticità i valori dei due attributi:



Per quanto riguarda l’andamento dei due attributi invece, è stato utilizzato un grafico a linee relative alle metriche `relativehumidity_mean` e `temperature_mean`. Si specifica inoltre che vi è l’assenza della metrica Adult Males nello studio di questo periodo, infatti sarà presente solamente nello Sheet3.



Si può notare dunque l’andamento delle due metriche e di come, nel periodo estivo, si avvicinino notevolmente, proprio a risaltare il legame della Temperatura e dell’Umidità in quell periodo.

Analisi Giornate Calde e Giornate Fredde

Dopo una prima analisi generale, è stata presa in considerazione un’analisi più dettagliata per comprendere quanti giorni sono stati caldi e quanti giorni invece, sono stati freddi. Ciò permette anche di rispondere ad una domanda: “Ci sono stati più giorni caldi o freddi?”

Lo studio è stato compiuto andando ad aggiungere una nuova colonna all'interno dello Sheet1, ovvero la colonna “**FreddoCaldo**”, tale colonna contiene due valori:

- Giornata Calda, quando la temperatura media è al di sopra dei 20°
- Giornata Fredda, quando la temperatura media è uguale o al di sotto dei 20°

La colonna è stata aggiunta con la seguente regola:

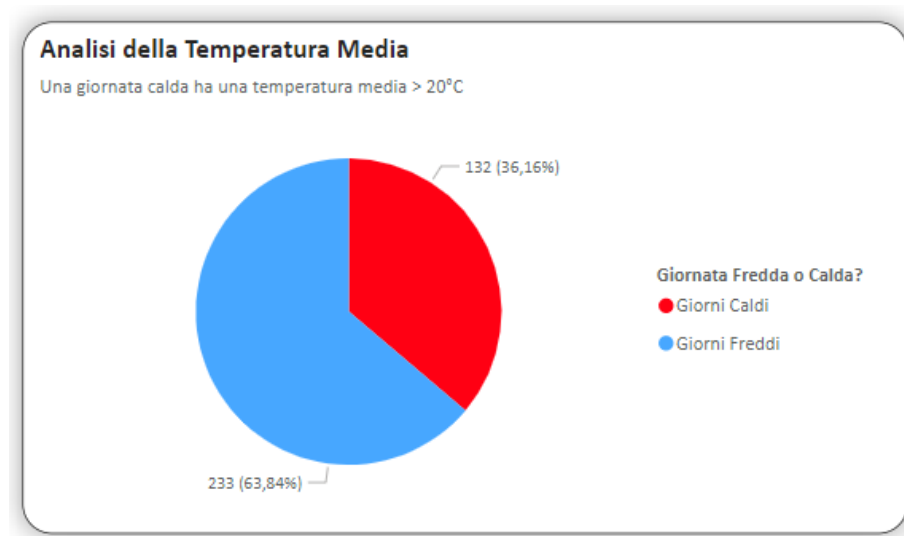
```
= Table.AddColumn("#Inserito giorno dell'anno", "FreddoCaldo", each if [temperature_mean] > 20 then "Giorni Caldi" else if [temperature_mean] <= 20 then "Giorni Freddi" else null)
```

La colonna si presenta nel seguente modo:

ABC 123	FreddoCaldo
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi
	Giorni Freddi
	Giorni Freddi
	Giorni Freddi
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi
	Giorni Caldi

Con l'utilizzo di questa colonna, è stata instanziato un grafico a torta nel quale sono visibili in maniera chiara e pratica i giorni caldi ed i giorni freddi, con la loro percentuale.

Si noti come 132 giorni sono stati caldi, dunque il 36,16%, mentre 233 giorni siano stati freddi/miti e dunque il 63,84%.



Un'altra importante ed interessante analisi che è stata possibile fare, è quella di studiare il numero di giorni veramente caldi e quelli veramente freddi.

Anche per questo studio è stata instanziata una nuova colonna denominata EstremiTemperatura nella quale vengono riportati i valori “Molto Caldo”, “Molto Freddo”, “Temperatura Normale” a seconda del valore della temperatura.

La colonna è stata creata seguendo la seguente regola:

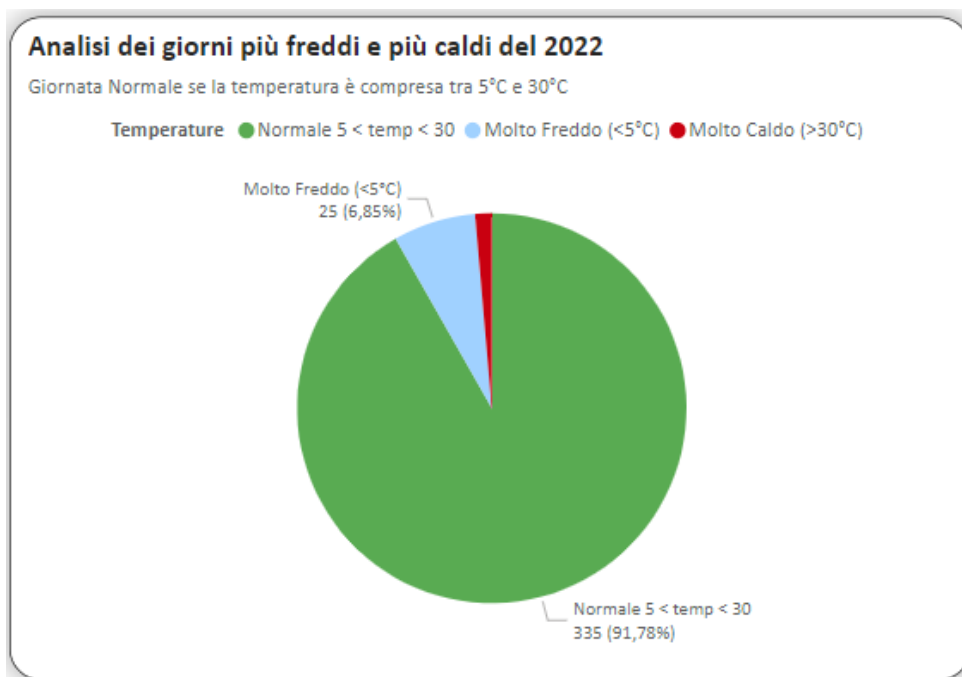
```
= Table.AddColumn("#Colonna condizionale aggiunta", "EstremiTemperatura", each if [temperature_mean] <= 5 then "Molto Freddo (<5°C)" else if [temperature_mean] >= 30 then "Molto Caldo (>30°C)" else "Normale 5 < temp < 30")
```

In questo modo, a seconda se la temperatura media è più bassa di 5° o superiore ai 30°, si ottengono informazioni sul tipo di giornata che si sta analizzando.

La colonna viene così rappresentata:

ABC 123	EstremiTemperatura
	Normale 5 < temp < 30
	Molto Freddo (<5°C)
	Molto Freddo (<5°C)
	Molto Freddo (<5°C)
	Molto Freddo (<5°C)
	Normale 5 < temp < 30
	Normale 5 < temp < 30
	Molto Freddo (<5°C)
	Molto Freddo (<5°C)
	Normale 5 < temp < 30
	Normale 5 < temp < 30
	Normale 5 < temp < 30
	Normale 5 < temp < 30
	Normale 5 < temp < 30

Con l'utilizzo di questa nuova colonna, è stato generato un grafico a torta per poter visualizzare con facilità la distinzione tra il numero di giornate molto calde e quelle molto fredde.



Si può notare agilmente come 25 giorni siano stati molto freddi, 5 giorni molto caldi, e come per i restanti 335 giorni la temperatura sia stata compresa tra i 5° e i 30°.

Analisi Giornate Umide e Giornate Secche

Allo stesso modo del precedente sottocapitolo, sono state analizzate le giornate anche relativamente al valore dell'Umidità Media, in modo da studiare quando una giornata è stata umida e quando invece secca. Anche qui è possibile poi rispondere alla domanda “*Ci sono state più giornate secche o umide?*”.

Qui sono state istanziate due colonne, una chiamata **UmidoSecco**, ed un'altra chiamata **EstremiUmidita**. In questo modo è stato possibile studiare quali giornate sono state umide/secche e quali invece siano state molto umide o molto secche, proprio come nel caso della temperatura.

ABC 123 EstremiUmidita	ABC 123 UmidoSecco
Molto Umido (>65%)	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi
Normale 50 < humidity < 65	Giorni Secchi
Molto Umido (>65%)	Giorni Umidi
Normale 50 < humidity < 65	Giorni Umidi
Molto Secco (<50%)	Giorni Secchi
Molto Secco (<50%)	Giorni Secchi
Normale 50 < humidity < 65	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi
Molto Umido (>65%)	Giorni Umidi

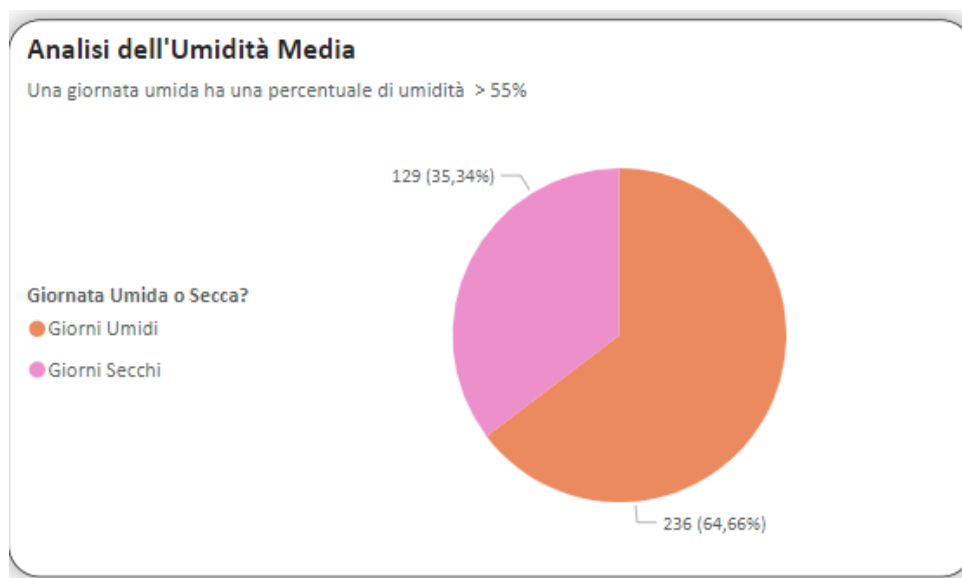
Anche in questo caso, è stata necessaria una regola per la creazione della colonna:

```
= Table.AddColumn("#Aggiunta colonna personalizzata", "UmidoSecco", each if [relativehumidity_mean]
    55 then "Giorni Umidi" else if [relativehumidity_mean] <= 55 then "Giorni Secchi" else null)
```

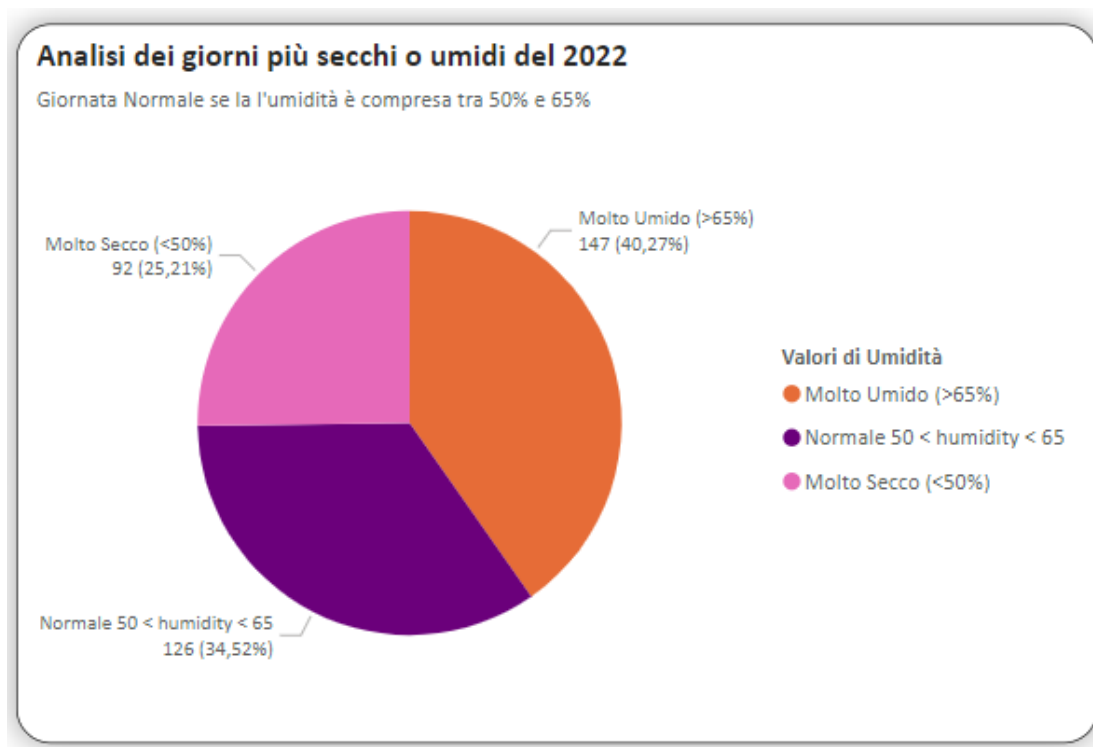
Una per la distinzione dei Giorni Umidi da quelli Secchi in base al valore della Temperatura Media, ed uno per i Giorni molto umidi o molto secchi:

```
= Table.AddColumn("#Colonna condizionale aggiunta1", "EstremiUmidita", each if [relativehumidity_mean]
    <= 50 then "Molto Secco (<50%)" else if [relativehumidity_mean] >= 65 then "Molto Umido (>65%)"
    else "Normale 50 < humidity < 65")
```

È stato presentato poi un grafico a torta a partire dalla colonna *UmidoSecco*, nel quale si può notare come 129 giorni siano stati umidi, dunque il 35,34% e come 236 giorni siano stati secchi/normali 64,66%.

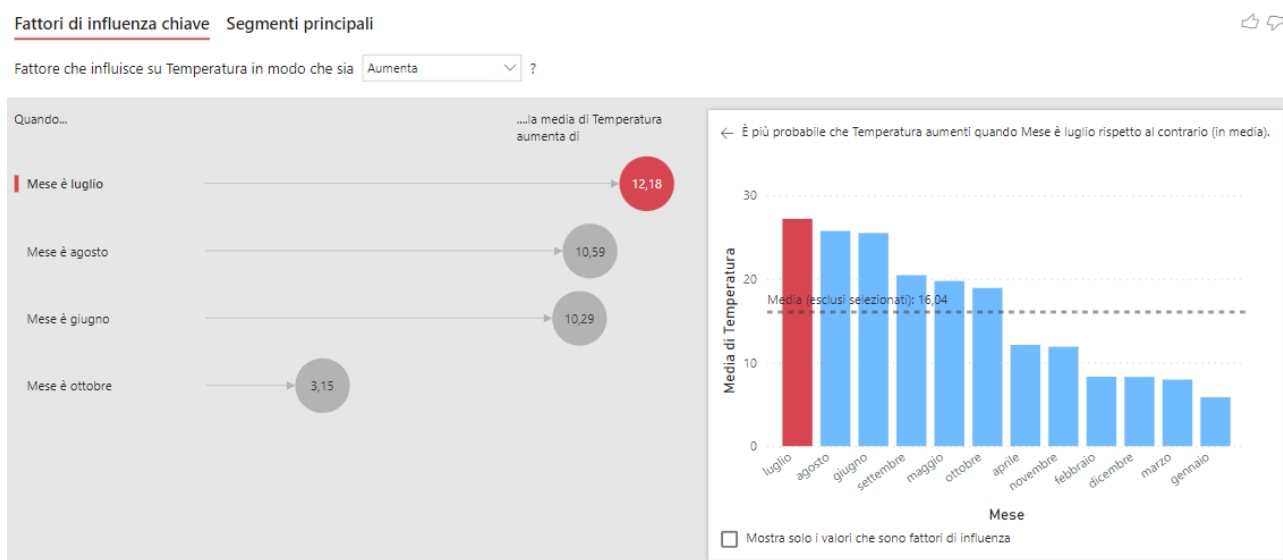


Un altro grafico è stato rappresentato invece con l'utilizzo di *EstremiUmidita* per studiare quando un giorno è stato davvero umido e quando davvero secco.



Analisi del Cambiamento di Temperatura in base al Mese

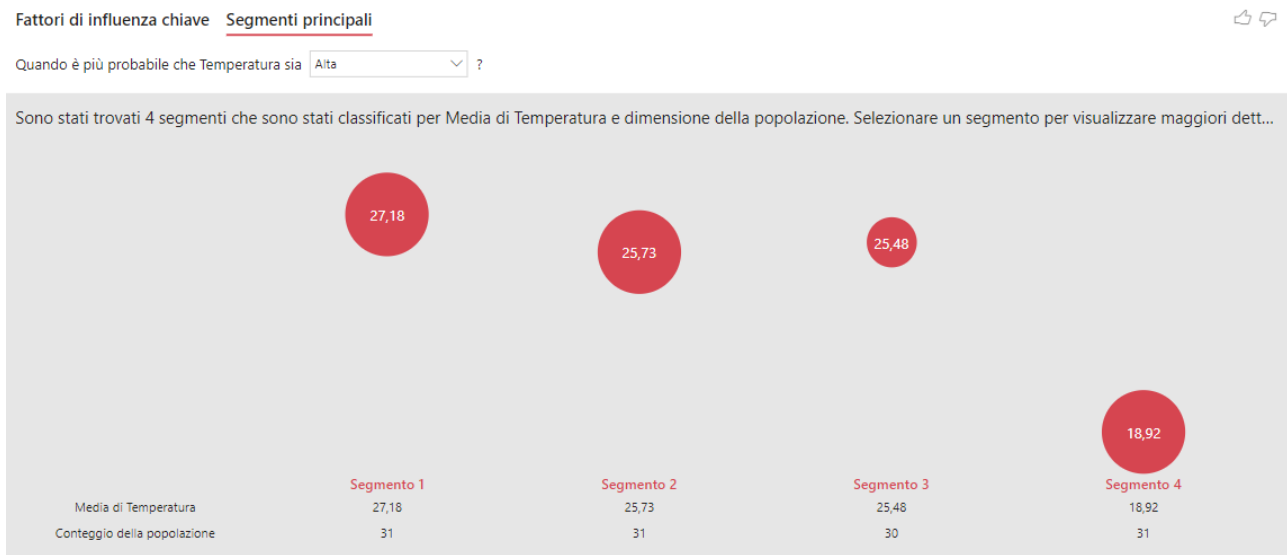
Dato che lo Sheet1 presente nel DataSet presenta i valori di Temperatura Media nel corso di ogni mese, può risultare utile lo studio della variazione della temperatura mese per mese.



Con l'utilizzo di questo grafico, sono evidenziati i *Fattori di Influenza Chiave*, qui viene evidenziato come in alcuni mesi, il valore della temperatura cambi.

Si nota inoltre che è stato adattato un menù a tendina nel quale è possibile scegliere se vedere quanto **Aumenta** la temperatura o quanto **Diminuisce**.

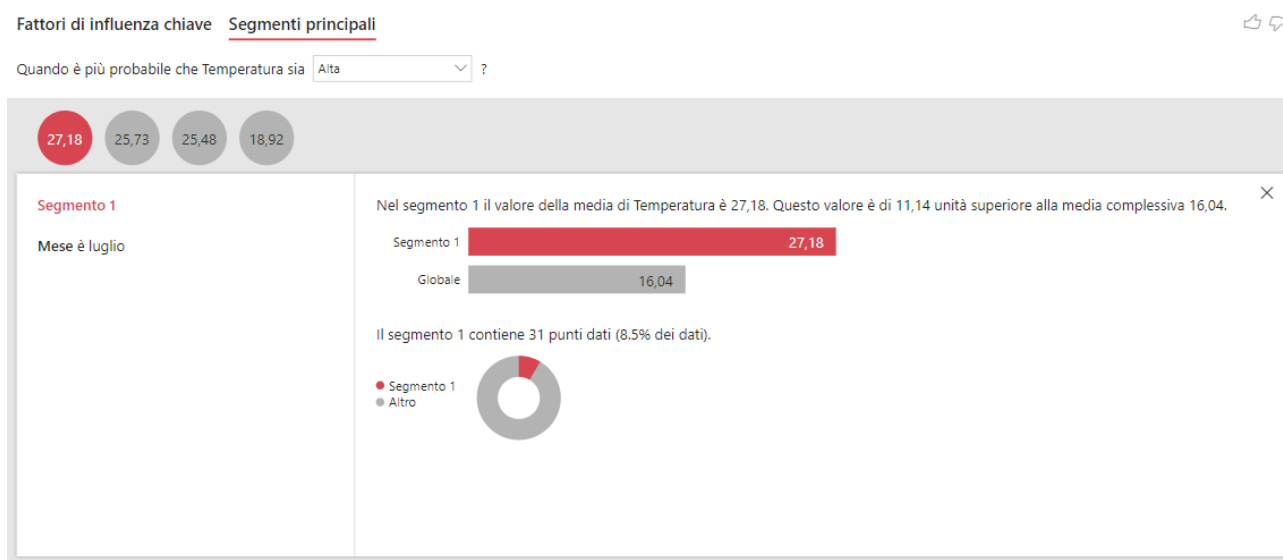
La sezione *Segmenti Principali* invece, presenta la possibilità di poter visualizzare quando c'è più probabilità di trovare mesi in cui la temperatura è alta:



A partire da questo grafico è possibile notare 4 distinti segmenti che sono stati differenziati tra di loro per:

- Media di Temperatura
- Conteggio dei Punti Dati

È possibile anche ottenere più informazioni facendo click su uno di questi segmenti.



In questa nuova schermata, è possibile osservare come, nella parte sinistra, viene “suggerito” il tipo di mese in base alla temperatura.

Nella parte destra invece, viene affrontato uno studio sul segmento scelto in cui vengono generate informazioni in merito a:

- Temperatura Media
- Di quante unità supera la media globale
- Quanti punti dati contiene

Mediante questo grafico, è pratico e rapido visualizzare le differenze relative alla Temperatura Media di ogni mese.

Analisi Dati Sheet3

All'interno del DataSet, come è stato anticipato prima, è presente un altro Sheet, nel quale, a differenza di quello precedente, oltre i valori di `temperature_mean`, `relativehumidity_mean`, `data`, è disponibile anche un ulteriore attributo, ovvero `no. Adult Males`.

Tale valore, offre informazioni di tipo “intero” in base al numero di adulti maschi di insetti per periodo. Si noti inoltre come questi dati, siano riferiti all'anno 2023, mentre lo Sheet1 ai dati dell'anno 2022.

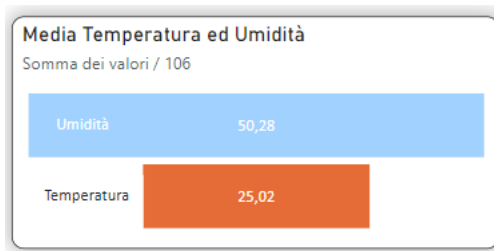
Anteprima dei Dati e Prime Analisi

Un primo passo è stato quello di fornire un'anteprima dei dati dello Sheet3 per rendere più semplice la loro visualizzazione (anche dinamica), ed uno studio iniziale sulla media e sull'andamento delle metriche nel corso dell'anno 2023.

I dati sono stati resi visualizzabili mediante una tabella in questo modo:

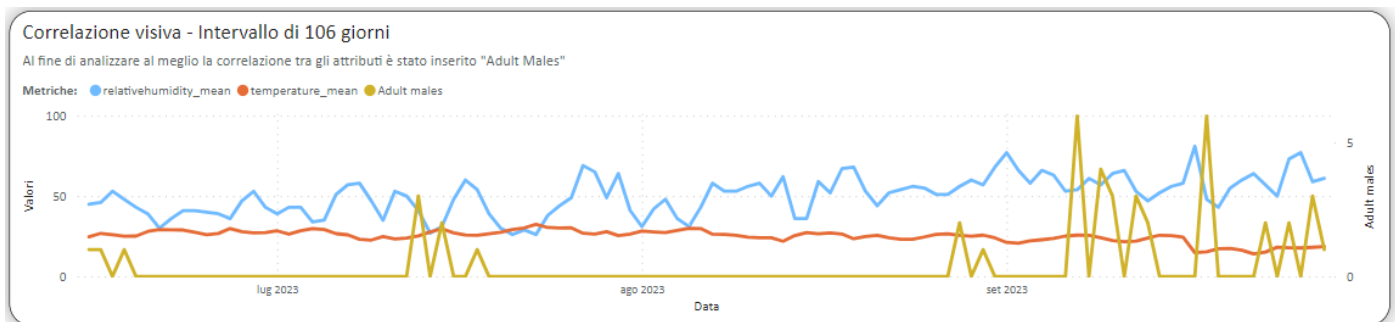
Anno	Mese	Giorno	Somma di relativehumidity_mean	Somma di temperature_mean
2022	marzo	31	94	10,34
2022	dicembre	24	94	9,34
2022	maggio	6	93	13,89
2022	dicembre	10	93	5,40
2022	dicembre	22	93	8,71
2022	dicembre	31	93	9,85

A partire da questi dati, così come con lo Sheet1, è stata presentata la differenza delle medie durante l'anno sia della Temperatura che dell'Umidità:



In questo modo risulta semplice ed intuitiva l'osservazione dei valori medi dei due attributi presi in considerazione.

È stato fatto poi un ulteriore studio sugli attributi relativo all'andamento delle tre metriche presenti nello Sheet3, in modo da graficare il loro andamento durante il 2023 per vedere come siano correlati tra di loro i dati.

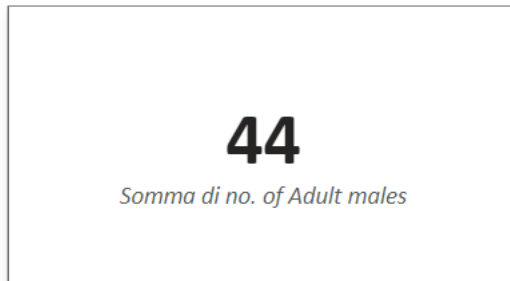


È possibile notare già da un primo sguardo rapido che non vi è alcuna correlazione "importante" tra le tre differenti metriche, soprattutto per quanto riguarda Adult Males.

Analisi Mediante Utilizzo del Widget ChatBot

Un'ulteriore analisi è stata quella affrontata mediante l'utilizzo del widget ChatBot reso disponibile da PowerBI. Mediante questo widget è possibile porre delle "domande" preimpostate al tool per poter visualizzare i dati richiesti o le informazioni richieste.

Un primo esempio è stato quello relativo al numero totale degli Adult Males, ponendo la domanda “How many no. Adult Males are there”:



Si può notare in maniera rapida che il numero totale degli Adult Males è di 44.

Un altro esempio è stato quello di chiedere al ChatBot "Days in Sheet 3 with the highest no. of Adult Males", in modo da visualizzare dunque i giorni con il numero massimo di Adult Males :

Date	no. of Adult males	temperature_mean	relativehumidity_mean	Sheet1.temperature_mean
giovedì 7 settembre 2023	6	25,67	54	25,6
lunedì 18 settembre 2023	6	15,41	48	15,4

Mediante questa tabella è possibile notare che i due giorni con il maggior numero di Adult Males, sono stati giovedì 7 Settembre e lunedì 18 Settembre del 2023.

Un ultimo esempio è stato quello di chiedere "Descending sort Sheet 3 by no. of Adult Males" per visualizzare in maniera decrescente lo Sheet3 in base al numero degli Adult Males:

Date	temperature_mean	relativehumidity_mean	Sheet1.temperature_mean	Sheet1.relat
giovedì 7 settembre 2023	25,67	54	25,67	
lunedì 18 settembre 2023	15,41	48	15,41	
sabato 9 settembre 2023	24,08	57	24,08	
giovedì 13 luglio 2023	25,20	41	25,20	
domenica 10 settembre 2023	22,41	64	22,41	
martedì 12 settembre 2023	22,04	53	22,04	
mercoledì 27 settembre 2023	18,09	59	18,09	
sabato 15 luglio 2023	29,64	31	29,64	
lunedì 28 agosto 2023	25,59	56	25,59	

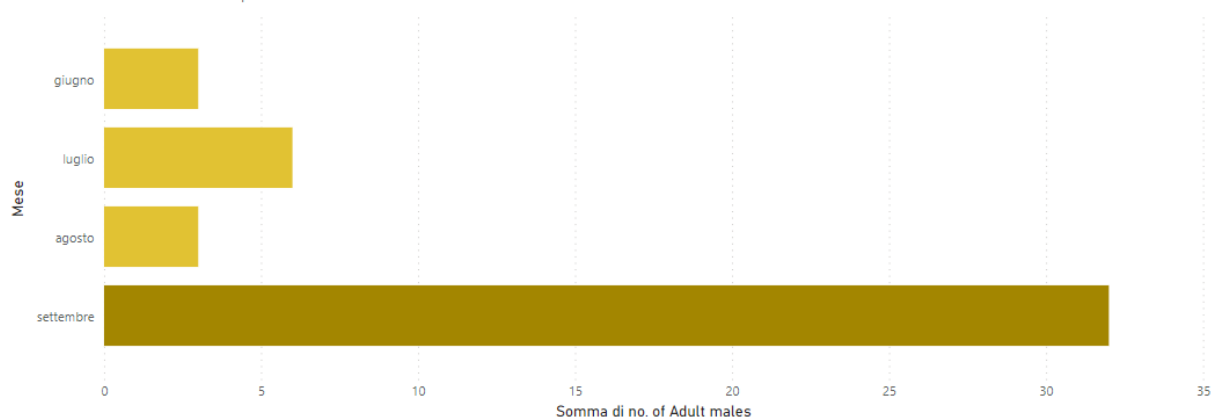
Analisi Approfondita su Adult Males

Avendo approfondito già lo studio sulla Temperatura e sull'Umidità, per lo Sheet3 ci siamo soffermati sull'analisi di Adult Males.

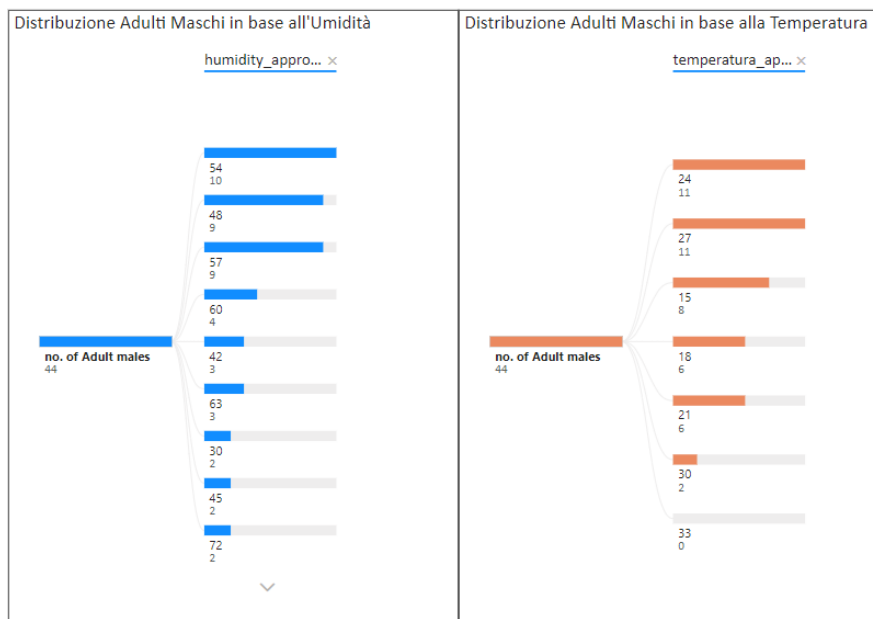
Una prima analisi è stata quella di visualizzare, mediante un grafico a barre in pila, il numero di Adult Males distribuiti nei vari mesi del 2023.

In questo modo, è possibile ottenere importanti informazioni sulla concentrazione degli insetti nel corso dei mesi e dunque, quando è più consigliabile utilizzare ad esempio degli insetticidi.

Somma di no. of Adult males per Mese
Valutiamo il numero di insetti distribuiti per mesi



Si può notare subito che il mese con la più alta concentrazione di Adult Males è Settembre, ciò può indurre magari un contadino o un proprietario di terre, ad utilizzare insetticidi o prestare maggior attenzione, durante tale mese.



Un'ulteriore analisi è stata affrontata per analizzare la distribuzione degli Adult Males in base all'Umidità ed in base alla Temperatura.

Il tutto è stato affrontato andando ad analizzare due nuove colonne, che prendono il nome di *temperatura_approssimata* e *humidity_approssimata*.

Questi due attributi sono stati inseriti considerando che i valori di Temperatura e di Umidità, sono stati approssimati ai 3 valori più vicini, in modo da rendere più leggibile e pratico il widget utilizzato.

Ottenuto mediante tale regola durante la creazione:

```
= Table.AddColumn("#Aggiunta colonna personalizzata", "humidity_approssimata", each Number.Round([relativehumidity_mean]/3)*3)
```

(Procedimento identico per la temperatura)

Grazie a questi grafici ad Alberi di Scomposizione, è stato possibile visualizzare in modo rapido la distribuzione del numero di Adult Males rispetto ad Umidità e Temperatura. Infatti si può notare con facilità che ad un' umidità del 54% corrisponde il numero massimo di Adult Males, mentre ad una temperatura di 24° corrisponde il numero massimo di Adult Males.

Allo stesso modo dello Sheet3, ci siamo soffermati sullo studio dei Fattori di Influenza Chiave , per vedere quali attributi influissero fortemente su Adult Males.

Come era stato accennato in precedenza, prendendo visione del grafico iniziale, relativo all'andamento delle tre metriche, era possibile notare che i dati non erano correlati strettamente tra di loro. Proprio per questo, lo studio dei Fattori di Influenza Chiave, non ha sortito l'effetto desiderato, in quanto il DataSet preso in esame è risultato insufficiente.



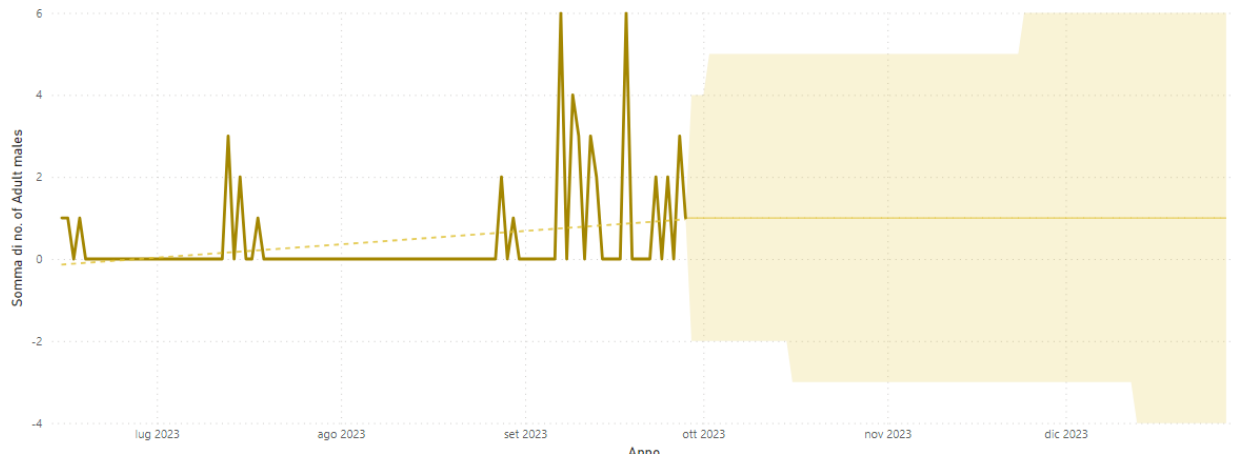
Analisi Predittiva di no. Adult Males

Sebbene i dati non siano strettamente correlati tra di loro, ed il DataSet risulti insufficiente per lo studio di un'analisi predittiva, il team ha comunque preso in considerazione l'idea di visualizzare graficamente la "probabile" predizione del numero di maschi adulti nel corso dei prossimi tre mesi successivi, ovvero Ottobre, Novembre e Dicembre.

Utilizzando un grafico a linee ed il tool di "Prediction", è stato possibile generare tale grafico a partire dai dati iniziali:

Analisi sulla Previsione Futura dei prossimi tre mesi del numero di Adult Males

Si può notare come la previsione non sia affatto precisa dato che il Dataset contiene un numero limitato di campioni

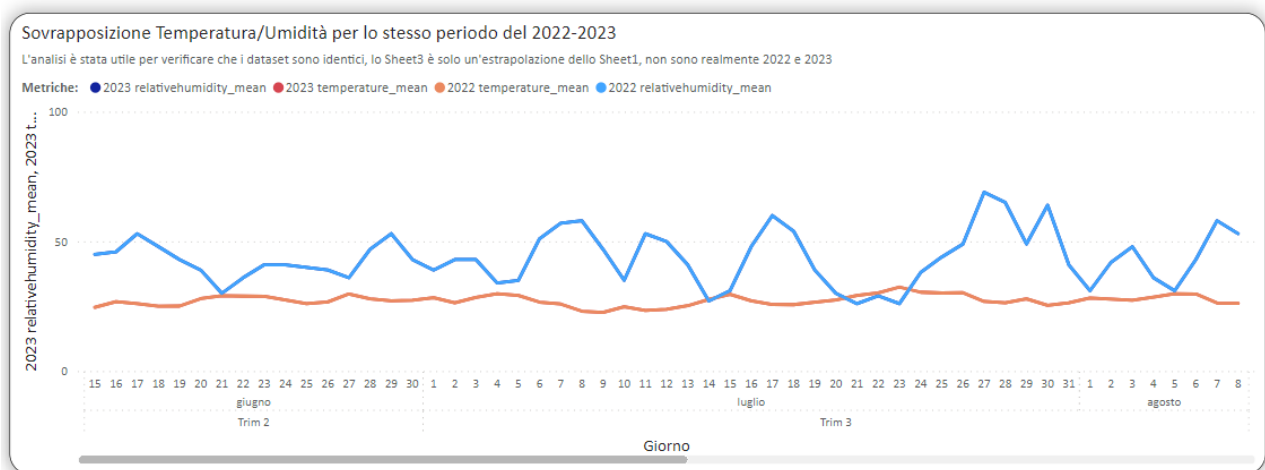


L'analisi infatti, è risultata infruttuosa, generando una previsione “esagerata” e poco precisa.

Confronto tra Sheet1 e Sheet3

Come ultimo passo, è stato processato un confronto tra i due Sheet.

Per poter applicare il confronto, sono state implementate altre due colonne all'interno dello Sheet3, le quali sono semplicemente l'importazione dei dati di Temperatura Media ed Umidità Media.



Effettuando questo confronto abbiamo notato che i due Sheet non sono distribuiti su due anni differenti, ma anzi, lo Sheet3 è semplicemente un'estrapolazione dello Sheet1, infatti i valori di temperatura media e di umidità media sia per quanto riguarda il 2022 che il 2023, sono identici per tutto il grafico, infatti si noti che le linee sono sovrapposte.

Conclusioni

In conclusione, l'analisi condotta ha dimostrato chiaramente i vantaggi distintivi offerti dall'impiego di PowerBI nell'analisi dei dati considerati nel presente studio. Le funzionalità intuitive e avanzate di questa piattaforma hanno rappresentato un valore significativo nel processo di esplorazione, visualizzazione e comprensione dei dati relativi al DataSet in esame.

L'utilizzo di PowerBI ha consentito una rapida e accurata esplorazione delle variazioni stagionali di temperatura e umidità, offrendo strumenti grafici potenti per identificare pattern e trend rilevanti nel corso del periodo analizzato. La facilità nell'elaborazione di modelli predittivi basati sui dati storici ha fornito una prospettiva rapida sulle condizioni climatiche future, sebbene a causa della non sufficienza del DataSet, siano stati ottenuti risultati non validi.