Indice di qualità dell’aria nell’agglomerato di Milano

Alla luce delle nuove linee guida dell’OMS 2021 a che punto è Milano (short-term)? Le misure adottate per migliorare la qualità dell’aria sono efficaci? (da rivedere)

**Introduzione**

La qualità dell’aria è un fattore molto importante per la protezione dell’ambiente ma soprattutto per la salute umana. Secondo la WHO ogni anno nel mondo 4,2 milioni di morti sono riconducibili ad una esposizione ad aria inquinata [REF].

L’inquinamento atmosferico per essere ridotto a livello globale deve essere trattato a livello locale. Nella Pianura Padana la presenza di Alpi e Appennini determina condizioni meteorologiche che favoriscono l’accumulo di inquinanti al suolo, in particolare in Lombardia la situazione è critica rispetto agli obiettivi della normativa italiana ed europea. Il monitoraggio degli inquinanti nella regione è affidato ad ARPA Lombardia che dispone di diverse stazioni di controllo distribuite nelle diverse zone in cui è suddiviso il territorio. Infatti, come prevede la direttiva europea 08/50/CE recepita dal D.Lgs. 155/10, deve esserci una suddivisione della regione in zone e agglomerati sulle quali valutare il rispetto dei valori obiettivo e dei valori limite dei diversi inquinanti. Delle 8 zone individuate nella regione Lombardia, noi ci concentreremo su quella chiamata Agglomerato di Milano che comprende le città di Milano, Monza, Como e territori limitrofi.

Gli inquinanti monitorati dalle XX stazioni dell’agglomerato di Milano e usati per definire la qualità dell’aria sono: particolato (PM 2.5 e PM 10), biossido di azoto (NO2), biossido di zolfo (SO2) e ozono atmosferico (O3). I dati orari e giornalieri delle concentrazioni vengono mediati all’interno dell’agglomerato e confrontati con una scala (da ‘molto scarsa’ a ‘molto buona’) che ne identifica la pericolosità. La qualità dell’aria complessiva è poi determinata considerando il punteggio peggiore tra i diversi inquinanti.

**Domanda di ricerca**

Alla luce delle nuove linee guida sulla qualità dell’aria rilasciate a fine 2021 dalla WHO e in vista delle nuove restrizioni sulle emissioni, in che condizioni si trova l’aria nella zona di Milano? Come è evoluto l’indice di qualità dell’aria, che riassume le concentrazioni dei principali inquinanti, dal rilascio delle linee guida precedenti ormai 15 anni fa? Inoltre, cercheremo di valutare se le restrizioni dovute alla pandemia di COVID-19 hanno avuto qualche effetto sulla qualità dell’aria.

**Fonti datasets**

* **Dati sensori aria** (piú datasets dal 2005 al 2021) https://www.dati.lombardia.it/Government/ARPA-LOMBARDIA-elenco-dataset-pubblicati/8ask-gxyr/data
* **Stazioni qualità dell’aria** https://www.dati.lombardia.it/Ambiente/Stazioni-qualit-dell-aria/ib47-atvt
* **Dataset zonizzazione** https://www.arpalombardia.it/sites/DocumentCenter/Documents/ARIA-Rete-rilevamento/allegato%20dgr%20zonizzazione%202011.pdf
* **Tabella indice di qualità dell’aria (IQA)** https://www.arpalombardia.it/Pages/Aria/Modellistica/Indice-qualit%C3%A0-aria.aspx?firstlevel=Modellistica

**Descrizione dati**

I dataset riportati sopra contengono una descrizione delle stazioni, dei sensori, delle zone e le singole misurazioni orarie tra gli anni 2016 e 2021.

Il dataset Stazioni qualità dell’aria contiene uno storico dei sensori installati negli anni sul suolo lombardo. Sono presenti degli attributi per descrivere il sensore stesso come IdSensore, NomeTipoSensore e UnitaMisura che definiscono la grandezza misurata. Poi sono presenti degli attributi che descrivono la stazione in cui si trova il singolo sensore come IdStazione, NomeStazione, DataStart e DataStop e degli attributi che la localizzano: Provincia, Comune, Lat, Lng e Quota.

I dataset Dati sensori aria riportano per i diversi anni i valori misurati dai sensori. In particolare, gli attributi presenti sono IdSensore, Valore e Data che ha frequenza giornaliera per PM2.5 e PM10 e frequenza oraria per gli altri inquinanti. Per distinguere questa cosa c’è l’attributo IdOperatore che indica se il dato rappresentato è un valore puntuale o una media. Inoltre, per ogni misurazione è indicato VA se il dato è valido e NA se il dato è invalido; c’è infatti un processo di validazione dei dati da parte degli operatori di ARPA Lombardia prima della pubblicazione per identificare possibili comportamenti anomali.

Il dataset zonizzazione l’abbiamo ricavato a partire dal PDF “ZONIZZAZIONE DEL TERRITORIO REGIONALE IN ZONE E AGGLOMERATI PER LA VALUTAZIONE DELLA QUALITÀ DELL’ARIA AMBIENTE” che riporta l’elenco dei comuni della regione con una colonna per descrivere la zona in cui è stato catalogato.

Infine, la tabella indice di qualità dell’aria (IQA) permette di fornire una stima immediata e sintetica sullo stato dell’aria a partire dalle concentrazioni nel breve periodo. Gli intervalli presenti catalogano ogni inquinante in una scala di 5 valori da ‘molto scarsa’ a ‘molto buona’. Questi valori sono stati scelti da ARPA Lombardia secondo le specifiche dall’Agenzia Europea per l’Ambiente e secondo i valori limiti indicati dalle linee guida della WHO. I primi tre livelli di qualità (Molto buona, buona e accettabile) si trovano al di sotto del valore limite giornaliero (o nelle 8 ore nel caso dell’ozono) mentre gli altri due livelli sono oltre il limite.

**Processo di sviluppo**

Partendo dal dataset con le misurazioni per prima cosa abbiamo rimosso le righe con valori non validi, queste corrispondevano a circa il 10% del totale. In seguito, abbiamo raggruppato le misurazioni per IdSensore e giorno cosí da poter calcolare media e massimo per ogni giorno. Queste due misure ci serviranno per calcolare l’indice giornaliero: per l’ozono viene utilizzato il massimo valore giornaliero mentre per gli altri inquinanti si usa la media giornaliera.

Avendo cosí ridotto le dimensioni del dataset abbiamo proceduto all’unione con il dataset dei sensori, sfruttando il campo IdSensore, in modo da poter collegare il valore al tipo di sensore e alla stazione. In questo modo siamo in grado di identificare e tenere solo i valori che corrispondono a sensori di nostro interesse (PM2.5, PM10, NO2, SO2, O3). A questo ha seguito una fase di controllo per i sensori con frequenza oraria (NO2, SO2) per rimuovere le medie composte da meno di 8 osservazioni (8 ore su 24); abbiamo ritenuto questo un valore sufficiente poiché nel corso della giornata le fluttuazioni non sono eccessive.

Il dataset cosí pulito l’abbiamo quindi unito al dataset della zonizzazione rispetto all’attributo Comune. Questo passaggio ci ha quindi permesso di ricavare solo i valori relativi alla zona di interesse (Agglomerato di Milano). L’ultimo raggruppamento l’abbiamo quindi compiuto rispetto agli attributi Data e NomeTipoSensore così da mediare i valori di massimo e media dei 5 sensori all’interno della zona per ogni giorno.

L’ultimo passaggio è stato poi quello di attribuire gli indici di qualità presenti nella tabella ad ogni inquinante per ogni giorno cosí da avere valori indipendenti dei diversi inquinanti. La misura complessiva della qualità dell’aria l’abbiamo ricavata poi considerando l’inquinante con l’indice peggiore.

**Tools utilizzati**

Per i passaggi descritti nel processo di sviluppo è stato utilizzato Python con le librerie Pandas e Numpy. Per la creazione delle visualizzazioni e la loro composizione nella dashboard è stato utilizzato Tableau.

**Findings**

..

**Processo e risultati della valutazione di qualità**

..