

**UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO**

**DIPARTIMENTO DI INFORMATICA**

**CORSO DI LAUREA MAGISTRALE**

**IN**

**DATA SCIENCE**

**Caso di studio di**

**Gestione di dati strutturati e non strutturati**

**Docenti: Studente:**

**Corrado Loglisci Marco Casamassima 730346**

**Anno Accademico 2022-2023**

Sommario

[Analisi del problema 3](#_Toc135996185)

[Dataset 5](#_Toc135996186)

[Modellazione concettuale 8](#_Toc135996187)

[Dizionario dei dati 8](#_Toc135996188)

[Modellazione logica 10](#_Toc135996189)

[Pre-processing 11](#_Toc135996190)

[Province 11](#_Toc135996191)

[Decessi 11](#_Toc135996192)

[Comuni 12](#_Toc135996193)

[Popolazioni 12](#_Toc135996194)

[Indicatori 13](#_Toc135996195)

[Stazioni, rilevazioni e rileva 13](#_Toc135996196)

[Data Definition Language 15](#_Toc135996197)

[Data Manipulation Language 18](#_Toc135996198)

[Query Language 19](#_Toc135996199)

[Figure 25](#_Toc135996200)

[Strumenti utilizzati 27](#_Toc135996201)

[Bibliografia 28](#_Toc135996202)

# Analisi del problema

L'inquinamento atmosferico, causato dalle attività umane, rappresenta un fattore di rischio ampiamente riconosciuto per la salute umana e per gli ecosistemi. Negli Stati occidentali, la lotta contro l'inquinamento atmosferico ha una storia di oltre sessant'anni di studi e ricerche volti a comprendere i meccanismi degli effetti dannosi degli inquinanti, nonché a individuare strategie e tecniche di mitigazione. Le reti di monitoraggio costituiscono il principale strumento per la valutazione della qualità dell'aria, intesa come il complesso di attività mirate a verificare il rispetto dei valori limite e l'ottenimento degli obiettivi stabiliti al fine di prevenire, eliminare o ridurre gli effetti negativi dell'inquinamento atmosferico sulla salute umana e sull'ecosistema in un determinato territorio statale [1].

Secondo uno studio dell’Organizzazione Mondiale della Sanità (OMS) ogni anno muoiono circa 7 milioni di persone a causa dell’inquinamento ambientale [2] e circa 9 persone su 10 al mondo respirano aria con elevati livelli di inquinanti [3] che causano malattie respiratorie e polmonari, leucemia, problemi cardiovascolari, malattie cardiache, ictus, cancro ai polmoni, ecc.

Un indicatore che può essere utilizzato per valutare la qualità dell’aria è la quantità del particolato atmosferico, ossia una miscela complessa di particelle solide e liquide di sostanze organiche ed inorganiche sospese in aria come, ad esempio, alcuni metalli tossici (piombo, cadmio e nichel) che hanno un’elevata tossicità [4]. In generale quanto più piccola è la dimensione delle particelle tanto maggiore è la loro capacità di penetrare nei polmoni e dunque di produrre effetti dannosi sulla salute umana. Il particolato è suddiviso in base al diametro aerodinamico:

* Particelle PM10: con diametro aerodinamico inferiore a 10 µm, penetrano nel tratto superiore delle vie aeree (cavità nasali, faringe e laringe);
* Particelle PM2.5: con diametro aerodinamico inferiore a 2.5 µm, possono giungere fino alle parti inferiori dell’apparato respiratorio.

Le sorgenti di queste particelle possono essere di tipo naturale (erosione del suolo, vulcani, incendi boschivi, dispersione di pollini, etc.) o antropogenico (industrie, riscaldamento, traffico veicolare e processi di combustione in generale) [5].

Secondo l’OMS, per il particolato non è possibile definire un valore limite al di sotto del quale non si verificano nella popolazione effetti sulla salute: per questo motivo le concentrazioni di PM10 e PM2.5 nell'aria dovrebbero essere mantenute a livelli più bassi possibile [6]. Tuttavia, le linee guida dell'OMS sulla qualità dell'aria, aggiornate al 2019, stabiliscono delle soglie di concentrazioni di particolato atmosferico consigliate da raggiungere. In particolare:

* Per il PM2.5: le concentrazioni medie annuali non devono superare i 10 µg/m3, mentre l’esposizione media giornaliera non deve superare i 25 µg/m3 per più di 3 - 4 giorni all'anno.
* Per il PM10: le concentrazioni medie annuali non devono superare i 20 µg/m3, mentre l’esposizione media giornaliera non deve superare i 50 µg/m3 per più di 3 - 4 giorni all'anno.

Questo elaborato ha l’obiettivo di creare un database col fine di aiutare gli utenti a visionare e analizzare i dati relativi all’inquinamento in Puglia e alle morti causate da malattie che potrebbero essere correlate all’inquinamento ambientale nelle varie regioni del Paese. In questo studio si considerano solo i dati relativi al 2020. In particolare, i possibili stakeholder del sistema possono essere:

* Gli amministratori comunali e provinciali, i quali possono utilizzare il database come sistema di supporto alle decisioni per garantire una migliore salute dei cittadini del proprio territorio, ad esempio:
  + Promuovendo l'uso dei trasporti pubblici: i mezzi di trasporto pubblico, come i bus e i treni, possono ridurre il numero di veicoli in circolazione, diminuendo così le emissioni di gas nocivi nell'aria;
  + Promuovendo la mobilità a piedi e in bicicletta;
  + Limitando il traffico veicolare: chiudendo alcune strade al traffico o riducendo il limite di velocità a 30 km/h. Soluzioni già presa in considerazione da alcuni comuni come Trento, Bologna e Torino [7] [8];
  + Aumentando le aree verdi in città.
* I cittadini, i quali possono prestare attenzione all’importanza di vivere in un mondo sostenibile.
* I ricercatori, i quali possono utilizzare i dati per condurre studi ed esperimenti con l'obiettivo di individuare eventuali soluzioni al problema.

# Dataset

Il database utilizzato in questo elaborato è stato creato a partire da dataset provenienti da diverse fonti di dati, tra cui:

* il numero di morti nelle diverse province della Puglia nel 2020, estratto dal sito dell'Istituto Nazionale di Statistica (ISTAT);
* la popolazione delle varie città della Puglia nel 2020, estratta dal sito ISTAT;
* i valori di PM10 e PM2.5 registrati dalle diverse stazioni di rilevamento dell'inquinamento in Puglia nel 2020, provenienti dal sito dell'Istituto Superiore per la Protezione Ambientale (ISPRA);
* i confini delle province e città della Puglia, provenienti dal sito ISTAT[[1]](#footnote-1) in formato shapefile.

Di seguito si descrivono approfonditamente i dataset utilizzati.

* **DCIS\_CMORTE1\_EV\_15052023111151999[[2]](#footnote-2)**: contiene i dati relativi al numero di decessi nelle varie province della Puglie causati da diverse malattie, considerando solo quelle che potrebbero essere correlate all'inquinamento ambientale, suddivisi per sesso. Secondo il National Institute of Environmental Health Sciences, alcune delle malattie causate dall'inquinamento ambientale includono il cancro, le malattie cardiovascolari, le malattie respiratorie, il diabete mellito, l'obesità e i disturbi riproduttivi, neurologici e del sistema immunitario [9]. Perciò sono state estratte solo le morti causate da:
  + tumori maligni;
  + malattie endocrine;
  + nutrizionali e metaboliche;
  + malattie del sistema circolatorio;
  + malattie del sistema respiratorio;
  + malattie dell'apparato digerente;
  + malattie del sistema e del tessuto connettivo;
  + malattie dell'apparato genitourinario.

Le colonne sono:

* **ITTER107**: codice statistico dell’unità amministrativa territoriale;
* **Territorio**: provincia della Puglia;
* **TIPO\_DATO15**: codice del tipo di dato richiesto (DEATH);
* **Tipo dato**: tipo di dato richiesto (morti);
* **CAUSEMORTE\_SL**: codice della causa di morte;
* **Causa iniziale di morte** **– European Short List**: nome della causa di morte;
* **SEXISTAT1**: codice del sesso;
* **Sesso**: nome del sesso;
* **TIME**: anno dei dati richiesto;
* **Seleziona periodo**: anno dei dati richiesto;
* **Value**: numero di morti.
* **puglia\_PM10\_2020[[3]](#footnote-3)**: contiene i dati relativi alle registrazioni effettuate dalle stazioni di rilevamento di PM10 nel territorio pugliese nel 2020. La latitudine e la longitudine sono stati calcolati utilizzando le coordinate in formato E-WGS84-UTM33 e N-WGS84-UTM33, ricavate da <https://www.arpa.puglia.it/moduli/output_immagine.php?id=4887>, e il sito <https://coordinates-converter.com/it>.

Le colonne sono:

* + - **Provincia**: nome della provincia pugliese dove è situata la stazione di rilevamento;
    - **Comune**: nome del comune pugliese dove è situata la stazione di rilevamento;
    - **Nome della stazione**: nome della stazione di rilevamento;
    - **Tipo di zona**: codice del tipo di zona (U: urbana, S: suburbana, R: rurale);
    - **Tipo di stazione**: codice del tipo di stazione di rilevamento (T: traffico, F: fondo, R: rurale);
    - **Giorni di superamento del valore limite giornaliero [50]**: giorni in cui è stata superata la soglia di 50 µg/m3 di PM10;
    - **Valore medio annuo**: valore medio di PM10 registrato nell’anno;
    - **Massimo**: valore massimo rilevato;
    - **Latitudine**: latitudine della stazione di rilevamento;
    - **Longitudine**: longitudine della stazione di rilevamento.
* **puglia\_PM2\_5\_2020[[4]](#footnote-4)**: contiene i dati relativi alle registrazioni effettuate dalle stazioni di rilevamento di PM2.5 nel territorio pugliese nel 2020. La latitudine e la longitudine sono stati calcolati utilizzando le coordinate in formato E-WGS84-UTM33 e N-WGS84-UTM33, ricavate da <https://www.arpa.puglia.it/moduli/output_immagine.php?id=4887>, e il sito <https://coordinates-converter.com/it>.

Le colonne sono:

* + - **Provincia**: nome della provincia pugliese dove è situata la stazione di rilevamento;
    - **Comune**: nome del comune pugliese dove è situata la stazione di rilevamento;
    - **Nome della stazione**: nome della stazione di rilevamento;
    - **Tipo di zona**: codice del tipo di zona (U: urbana, S: suburbana, R: rurale);
    - **Tipo di stazione**: codice del tipo di stazione di rilevamento (T: traffico, F: fondo, R: rurale);
    - **Giorni di superamento del valore limite giornaliero [25]**: giorni in cui è stata superata la soglia di 25 µg/m3 di PM10;
    - **Valore medio annuo**: valore medio di PM10 registrato nell’anno;
    - **Massimo**: valore massimo rilevato;
    - **Latitudine**: latitudine della stazione di rilevamento;
    - **Longitudine**: longitudine della stazione di rilevamento.
* **DCIS\_POPRES1\_15052023120359710[[5]](#footnote-5)**: contiene i dati relativi alla popolazione maschile nei vari comuni pugliesi nel 2020.

Le colonne sono:

* **ITTER107**: codice statistico dell’unità amministrativa territoriale;
* **Territorio**: comune pugliese;
* **TIPO\_DATO15**: mese in cui si riferisce la rilevazione della popolazione (JAN);
* **Tipo di indicatore demografico**: tipo di indicatore demografico (popolazione al 1° gennaio);
* **SEXISTAT1**: codice ISTAT del sesso (1);
* **Sesso**: nome del sesso (maschi);
* **ETA1**: codice dell’età (TOTAL);
* **Età**: età (totale);
* **STATCIV2**: indicatore dello stato civile (99);
* **Stato** **civile**: stato civile (totale);
* **TIME**: anno a cui si riferiscono i dati;
* **Seleziona** **periodo**: anno a cui si riferiscono i dati;
* **Value**: valore della popolazione.
* **DCIS\_POPRES1\_15052023120546025[[6]](#footnote-6)**: contiene i dati relativi alla popolazione femminile nei vari comuni pugliesi nel 2020.

Le colonne sono:

* **ITTER107**: codice statistico dell’unità amministrativa territoriale;
* **Territorio**: comune pugliese;
* **TIPO\_DATO15**: mese in cui si riferisce la rilevazione della popolazione (JAN);
* **Tipo di indicatore demografico**: tipo di indicatore demografico (popolazione al 1° gennaio);
* **SEXISTAT1**: codice ISTAT del sesso (2);
* **Sesso**: nome del sesso (femmine);
* **ETA1**: codice dell’età (TOTAL);
* **Età**: età (totale);
* **STATCIV2**: indicatore dello stato civile (99);
* **Stato** **civile**: stato civile (totale);
* **TIME**: anno a cui si riferiscono i dati;
* **Seleziona** **periodo**: anno a cui si riferiscono i dati;
* **Value**: valore della popolazione.

# Modellazione concettuale

Immagine che contiene diagramma, testo, linea, Piano

Descrizione generata automaticamente

## Dizionario dei dati

Di seguito viene descritto il diagramma E-R

* Entità **Stazioni**: stazione di rilevamento degli inquinanti. Ha come attributi:
  + ID\_Stazione: primary key. Codice identificativo della stazione di rilevamento;
  + Nome: nome della stazione di rilevamento;
  + Tipo\_zona: tipo di zona in cui si trova la stazione di rilevamento;
  + Tipo\_stazione: tipo della stazione di rilevamento;
  + Posizione: posizione geografica della stazione.
* Entità **Comuni**: comune della Puglia. Ha come attributi:
  + ID\_Comune: primary key. Codice identificativo del comune;
  + Nome: nome del comune;
  + Posizione: posizione geografica del comune.
* Relazione **Appartiene\_A**: relazione che lega le entità Stazione e Comune con una relazione N:1 con partecipazione totale di Stazione. Indica l’appartenenza della stazione al comune.
* Entità **Popolazioni**: popolazione registrata dal comune. Ha come attributi:
  + ID\_Popolazione: primary key. Identificativo della popolazione;
  + Numero\_maschi: numero di maschi nella popolazione;
  + Numero\_femmine: numero di femmine nella popolazione.
* Relazione **Ha**: relazione che lega le entità Comune e Popolazione con una relazione 1:1 con partecipazione totale di entrambe le entità. Ha come attributo:
  + Anno: anno a cui si riferisce la popolazione del comune.
* Entità **Province**: provincia della puglia. Ha come attributi:
  + ID\_Provincia: primary key. Codice identificativo della provincia;
  + Nome: nome della provincia;
  + Posizione: posizione geografica della provincia.
* Relazione **Situato\_in**: relazione che lega le entità Comune e Provincia con una relazione N:1 con partecipazione totale di entrambe le entità.
* Entità **Decessi**: decessi registrati. Ha come attributi:
  + ID\_Decessi: primary key. Codice identificativo dei decessi;
  + Nome\_malattia: nome della malattia a cui si riferiscono i decessi;
  + Numero\_decessi\_maschi: numero di maschi morti;
  + Numero\_decessi\_femmine: numero di femmine morte.
* Relazione **Registra**: relazione che lega le entità Provincia e Decessi con una relazione 1:N con partecipazione totale di entrambe le entità. Ha come attributo:
  + Anno: anno a cui si riferiscono i decessi registrati.
* Entità **Rilevazioni**: rilevazione dell’indicatore di inquinamento effettata dalla stazione. Ha come attributi:
  + ID\_Rilevazione: primary key. Codice identificativo della rilevazione;
  + Anno: anno in cui è stata effettuata la rilevazione;
  + Giorni\_superamento: giorni in cui è stata superata la soglia massima dell’indicatore d’inquinamento consigliata dall’OMS;
  + Media\_annuale: media dei valori registrati dell’indicatore dell’inquinamento;
  + Massimo: valore massimo rilevato nell’anno.
* Relazione **Effettua**: relazione che lega le entità Stazione e Rilevazione con una relazione 1:N con partecipazione totale di Rilevazione.
* Entità **Indicatori**: indicatore d’inquinamento. Ha come attributi:
  + Nome: primary key. Nome dell’indicatore d’inquinamento;
  + Sigla: sigla dell’indicatore d’inquinamento;
  + Soglia\_giorni\_superamento: soglia di giorni che l’OMS consiglia di non superare;
  + Soglia\_media\_annuale: soglia della media di valori annuali che l’OMS consiglia di non superare.
* Relazione **Relativo\_a**: relazione che lega le entità Rilevazione e Indicatore con una relazione N:1 con partecipazione totale di Rilevazione.
* Relazione **Rileva**: relazione che lega le entità Stazione e Indicatore con una relazione N:N.

# Modellazione logica

Immagine che contiene testo, schermata, Carattere, diagramma

Descrizione generata automaticamente

Dalla traduzione del modello concettuale al modello logico sono state individuate 7 tabelle:

Poiché è presente una relazione 1:N tra “Province” e “Decessi” nel modello concettuale, sono state create le tabelle “province” e “decessi” con quest’ultima contenente una chiave esterna per “province”. Inoltre, è stato inserito l’attributo “anno” della relazione “registra” in “decessi”. Stesso tipo di conversione è stato fatto per le relazioni 1:N tra “comuni” e “province”, “stazioni” e “comuni”, “rilevazioni” e “stazioni”, e “rilevazioni” e “indicatori”. La relazione 1:1 tra “comuni” e “popolazioni” è stata gestita inserendo in “comuni” gli attributi di “popolazioni” e l’attributo “anno” di “ha”. Per quanto riguarda la relazione N:N tra “stazioni” e “indicatori”, è stata creata la tabella “rileva”, la quale possiede come chiave primaria le chiavi primarie delle due tabelle.

# Pre-processing

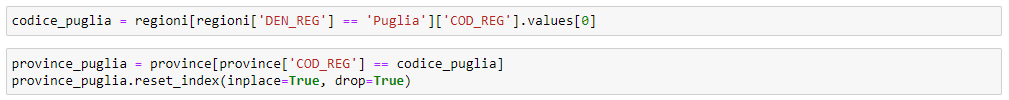
La fase di pre-processing è necessaria per preparare i dati csv alle tabelle che saranno create tenendo in considerazione il modello logico.

## Province

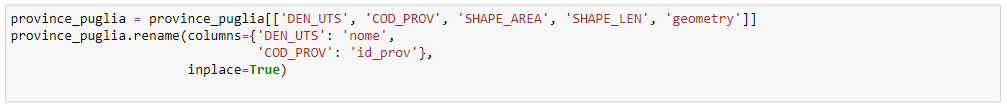
* Lettura dei dataset relativi alle regioni e alle province:



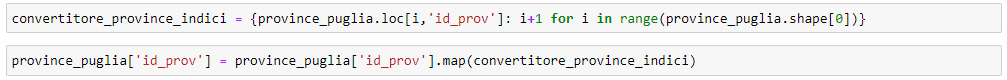
* Estrazione delle province pugliesi:



* Ridenominazione delle colonne di province:



* Creazione degli indici delle province:



* Esportazione dello shapefile sulle province:



## Decessi

* Lettura del dataset relativo ai decessi in puglia:



* Selezione delle colonne rilevanti:



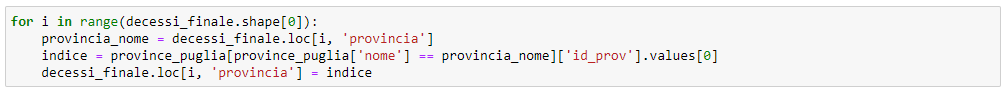
* Modifica del dataset in modo da avere come colonne “nome\_malattia”, “numero\_decessi\_maschi”, “numero\_decessi\_femmine”, “anno”, “provincia”:

Immagine che contiene testo, Carattere, linea, schermata

Descrizione generata automaticamenteImmagine che contiene testo, schermata, Carattere, bianco

Descrizione generata automaticamente

* Sostituzione del nome della provincia nei decessi con il corrispondente indice della tabella province:



* Esportazione del file csv sui decessi:



## Comuni

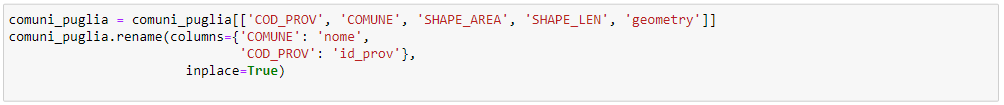
* Lettura del dataset relativo ai comuni:



* Estrazione dei comuni pugliesi:



* Ridenominazione delle colonne di comuni:



* Creazione degli indici delle province:



* Creazione degli indici dei comuni:



* Esportazione dello shapefile sui comuni:



## Popolazioni

* Lettura dei dataset relativi alle popolazioni nelle regioni pugliesi:



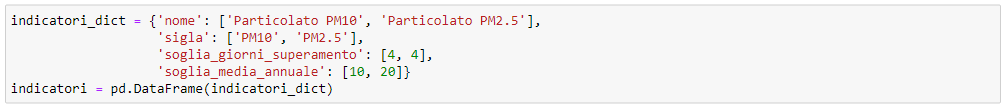
* Merge dei dataset relativi alla popolazione femminile e alla popolazione maschile:

Immagine che contiene testo, Carattere, schermata

Descrizione generata automaticamente

## Indicatori

* Creazione del dataframe relativo agli indicatori:

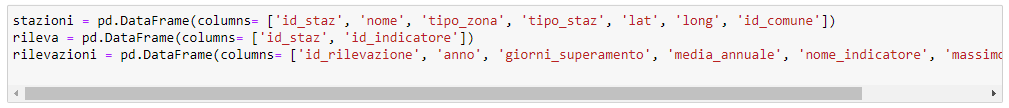


* Esportazione del file csv degli indicatori:



## Stazioni, rilevazioni e rileva

* Creazione dei dataframe relativi a stazioni, rileva e rilevazioni:



* Lettura dei dataset relativi alle rilevazioni:



* Aggiunta del . nella latitudine e longitudine delle stazioni (###### => ##.####):

Immagine che contiene testo, Carattere, schermata, linea

Descrizione generata automaticamente

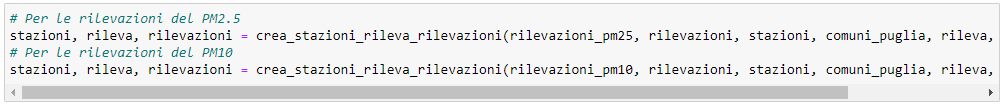
* Ridenominazione delle colonne:

Immagine che contiene testo, Carattere, schermata, linea

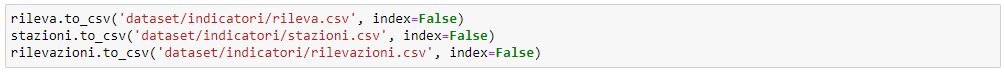
Descrizione generata automaticamente

* Creazione dei dataset:  
  Immagine che contiene testo, schermata, Carattere, numero

  Descrizione generata automaticamente



* Esportazione dei file csv delle stazioni, rileva e rilevazioni:



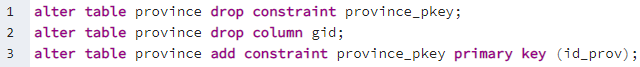
# Data Definition Language

Prima di creare le tabelle, è necessario aver creato il Database InquinamentoAriaPuglia, aggiungere l’estensione postgis e importare i file shapefile in PgAdmin4 utilizzando PostGIS 2.0 Shapefile and DBF Loader Exporter:

Immagine che contiene testo, schermata, schermo, numero

Descrizione generata automaticamente

* Modifica della primary key in province:

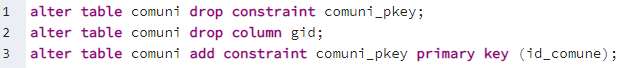


* Creazione tabella decessi:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

* Modifica della primary key in comuni:



* Modifica dei tipi di dati degli attributi:



* Creazione della foreign key id\_prov in comuni:



* Creazione tabella stazioni:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

* Aggiunta dell’attributo geom in stazioni:



* Creazione tabella indicatori:

Immagine che contiene testo, Carattere, schermata, linea

Descrizione generata automaticamente

* Creazione tabella rilevazioni:

Immagine che contiene testo, schermata, Carattere

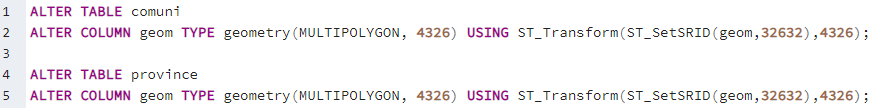
Descrizione generata automaticamente

* Creazione tabella rileva:

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

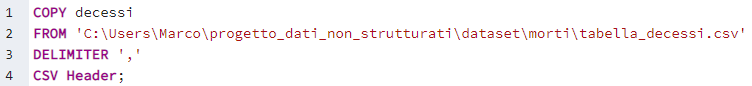
* Conversione del sistema di riferimento dell’attributo geom in province e comuni:



Sono stati creati i file DDL1.sql per la creazione del database e DDL2.sql per le restanti query DDL.

# Data Manipulation Language

* Importazione dei dati in decessi:

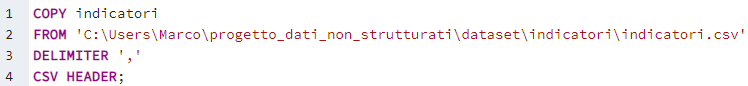


* Importazione dei dati in stazioni:

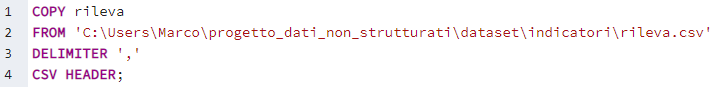
Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

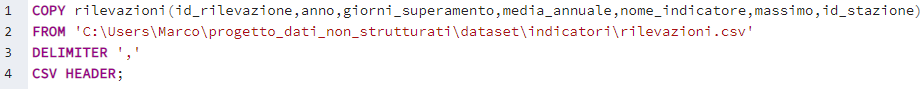
* Importazione dei dati in indicatori:



* Importazione dei dati in rileva:



* Importazione dei dati in rilevazioni:



È stato creato il file DML.sql contenente le query relativi al DML.

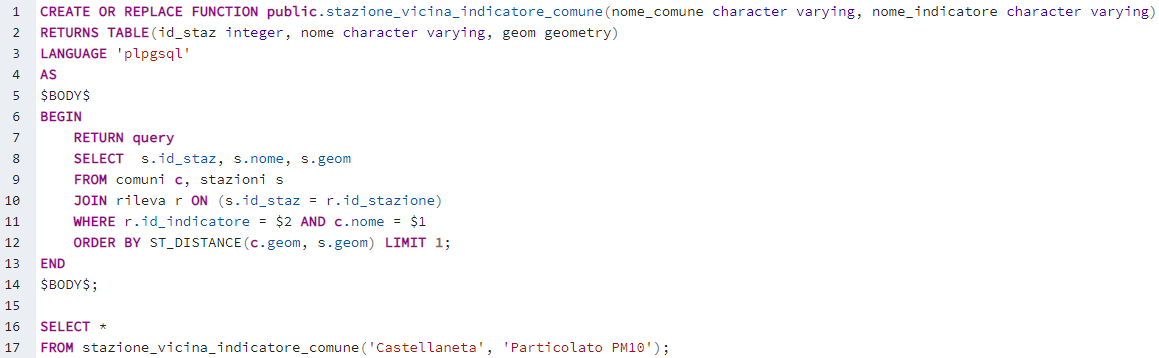
# Query Language

* **stazione\_vicina\_indicatore\_posizione**: mostrare la stazione di rilevamento più vicina ad una data posizione e che rileva un dato indicatore:

Immagine che contiene testo, Carattere, schermata

Descrizione generata automaticamente

* **stazione\_vicina\_indicatore\_comune**: mostrare la stazione di rilevamento più vicina ad un dato comune e che rileva un dato indicatore:



Da questa query risulta che la stazione più vicina a Castellaneta che rileva il Particolato PM10 è Massafra.

* **stazioni\_provincia**: mostrare le stazioni di rilevamento, con i relativi indicatori che rilevano, all’interno di una data provincia:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Da questa query risulta che ci sono 15 stazioni nella provincia di Taranto.

* **stazioni\_vicine\_comune\_distanza**: mostrare le stazioni vicine ad un dato comune di una data distanza:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Da questa query risulta che nel raggio di 100km da Castellaneta ci sono 64 stazioni di rilevamento.

* **distanza\_minima\_stazioni**: mostrare la distanza minima che c’è tra due stazioni e le informazioni di queste ultime:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Da questa query risulta che le due stazioni più vicine in Puglia sono Torchiarolo ENEL e Torchiarolo con una distanza di 595.6m.

* **stazioni\_vicine\_raggio\_centro**: mostrare le stazioni più vicine in un’area dato un punto e un raggio:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

* **num\_stazioni\_superamento\_soglia\_giornaliera\_provincia:** mostrare il numero di stazioni di un dato indicatore e di una data provincia che supera la soglia minima giornaliera:

Immagine che contiene testo, Carattere, linea, schermata

Descrizione generata automaticamente

Da questa query risulta che nella provincia di Bari ci sono 12 stazioni che hanno rilevato un valore di Particolato PM10 maggiore di 50 µg/m3 per più di 4 giorni nel 2020.

* **num\_stazioni\_province:** mostrare il numero di stazioni di un dato indicatore di ogni provincia in ordine decrescente:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Da questa query risulta che il numero di stazioni che rilevano il Particolato PM10 nelle varie province sono: 12 a Bari, 11 a Taranto, 11 a Brindisi, 8 a Lecce, 4 a Foggia e 2 a Barletta-Andria-Trani. Mentre, il numero di stazioni che rilevano il Particolato PM2.5 sono: 7 a Bari, 5 a Lecce, 4 a Taranto, 4 a Brindisi, 2 a Barletta-Andria-Trani e 2 a Foggia.

* **malattia\_province\_media\_superata:** mostrare le malattie con più morti delle province in cui c’è stato il superamento della soglia della media annuale di un dato indicatore e un dato anno, e le relative stazioni:

Immagine che contiene testo, schermata, Carattere, documento

Descrizione generata automaticamente

Da questa query risulta che, nel 2020, ci sono 48 stazioni in cui si supera la soglia media annuale consigliata dall’OMS e che nelle province in cui operano, la malattia con più morti è sono quelle relative al sistema circolatorio.

Per ogni query è stato creato un file .sql.

# Figure

Immagine che contiene testo, mappa, atlante, diagramma

Descrizione generata automaticamente

Immagine che contiene mappa, atlante, testo

Descrizione generata automaticamente

Figura 1 Raffigurazione delle stazioni e dei comuni, e delle stazioni e delle province, rispettivamente nella figura superiore e inferiore.

Immagine che contiene testo, mappa

Descrizione generata automaticamente

Immagine che contiene mappa, testo, atlante

Descrizione generata automaticamente

Figura Raffigurazione delle stazioni e dei comuni, e delle stazioni e delle province, rispettivamente nella figura superiore e inferiore. I colori delle stazioni dipendono dal valore massimo di PM10 rilevato nel 2020

Dalle figure risultano che le stazioni sono situate soprattutto intorno a città come Bari, Taranto e Lecce. Inoltre, a Bari le stazioni hanno registrato un valore massimo di PM10, durante il 2020, molto maggiore rispetto alle stazioni situate in altri punti.

# Strumenti utilizzati

Gli strumenti utilizzati per la realizzazione del progetto sono:

* Python 3.9.13: per la fase di pre-processing;
* PGAdmin4, PostGIS 2.0 Shapefile and DBF Loader Exporter, PostgreSQL 9.6 e 11, e postgis 2.5.3 e 3.3.2: per le fasi di DDL, DML e QL;
* QGIS Desktop 3.28.3: per la visualizzazione grafica delle stazioni, province e comuni.

# Bibliografia

|  |  |
| --- | --- |
| [1] | S. n. p. l. p. dell'ambiente, «PARTE-INIZIALE-STATO-E-TREND.pdf,» 02 12 2020. [Online]. Available: https://www.snpambiente.it/wp-content/uploads/2020/12/PARTE-INIZIALE-STATO-E-TREND.pdf. |
| [2] | «500 actions to take, including 82 measures to reduce both outdoor and indoor air pollution that can help prevent 7 million premature deaths worldwide,» 2 09 2021. [Online]. Available: https://www.who.int/multi-media/details/500-actions-to-take-including-82-measures-to-reduce-both-outdoor-and-indoor-air-pollution-that-can-help-prevent-7-million-prematuer-deaths-worldwide. |
| [3] | «9 out of 10 people worldwide breathe polluted air, but more countries are taking action,» 2 05 2018. [Online]. Available: https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action#:~:text=Air%20pollution%20levels%20remain%20dangerously,outdoor)%20and%20household%20air%20pollution.. |
| [4] | A. R. P. A. V. d'Aosta, «ll particolato atmosferico - PM10 e PM2.5,» [Online]. Available: https://www.arpa.vda.it/it/aria/l-inquinamento-atmosferico/2536-il-particolato-atmosferico. |
| [5] | A. Lombardia, «PM10 E PM2,5,» [Online]. Available: https://www.arpalombardia.it/Pages/Aria/Inquinanti/PM10-PM2,5.aspx. |
| [6] | «PM10 - Particolato atmosferico o polveri sottili,» [Online]. Available: https://www.issalute.it/index.php/la-salute-dalla-a-alla-z-menu/p/pm10-particolato-atmosferico-o-polveri-sottili#effetti-sulla-salute. [Consultato il giorno 28 02 2018]. |
| [7] | R. News, «A Trento a 30 all'ora, ecco il piano del Comune per la sicurezza,» [Online]. Available: https://www.rainews.it/tgr/trento/articoli/2023/04/a-trento-a-30-allora-ecco-il-piano-del-comune-per-la-sicurezza-1a2c52be-39f6-4c89-97b7-7d4866fc3b07.html. |
| [8] | I. Macchi, «In Italia crescono le città con limite a 30 km/h,» [Online]. Available: https://www.alvolante.it/news/italia-crescono-citta-limite-30-kmh-381060. |
| [9] | N. I. o. E. H. Sciences, «Air Pollution and Your Health,» [Online]. Available: https://www.niehs.nih.gov/health/topics/agents/air-pollution/index.cfm. [Consultato il giorno 06 05 2023]. |

1. https://www.istat.it/it/archivio/222527 [↑](#footnote-ref-1)
2. Fonte: http://dati.istat.it/Index.aspx?QueryId=26453# [↑](#footnote-ref-2)
3. Fonte: https://annuario.isprambiente.it/sys\_ind/777 [↑](#footnote-ref-3)
4. Fonte: https://annuario.isprambiente.it/sys\_ind/781 [↑](#footnote-ref-4)
5. Fonte: http://dati.istat.it/Index.aspx?DataSetCode=DCIS\_POPRES1# [↑](#footnote-ref-5)
6. Fonte: http://dati.istat.it/Index.aspx?DataSetCode=DCIS\_POPRES1# [↑](#footnote-ref-6)