

# Analisi della lunghezza degli archi stradali per provincia mediante l'algoritmo "Join by Attributes" tra località censuarie e grafo stradale OSM in QGIS

A cura di Marco Broccoli – broccoli@istat.it

ISTAT – DCME/MEA

12 Dicembre 2023

## Abstract

L'obiettivo di questo studio è implementare un'analisi spaziale per identificare la lunghezza delle carreggiate stradali per provincia, utilizzando dati estratti da OpenStreetMap (OSM) e il censimento ISTAT 2011.

Gli strumenti GIS utilizzati includono l'algoritmo "Join by Attributes" e il software QGIS. Il processo analitico prevede la creazione di una relazione spaziale tra poligoni rappresentanti le località censuarie e il grafo stradale, modellato come shape lineari.

Attraverso la funzione `$length` di QGIS, viene calcolata la lunghezza di ogni arco stradale. Successivamente, questi archi vengono classificati per provincia integrando i dati demografici con le geometrie del grafo stradale.

L'aspetto computazionale di questo metodo viene esaminato per valutare le prestazioni in termini di efficienza e gestione dei dati su larga scala.

L'obiettivo finale dello studio è fornire una metodologia per la costruzione di indicatori utili alla sicurezza stradale, utilizzando come base la lunghezza degli archi stradali e la distribuzione delle carreggiate per provincia.

I risultati di questa ricerca possono costituire un punto di partenza per ulteriori studi sulla sicurezza infrastrutturale e l'ottimizzazione della pianificazione territoriale.

## 1. Introduzione

La crescente disponibilità di dati geografici, tra cui i dati open-source come quelli di OpenStreetMap (OSM), offre nuove opportunità per l'analisi spaziale a supporto della pianificazione urbana e della sicurezza stradale. L'integrazione tra questi dataset e i dati censuari consente di analizzare dettagliatamente la distribuzione delle infrastrutture stradali e valutare la sicurezza in base alla distribuzione delle carreggiate.

In Italia, il censimento ISTAT 2011 fornisce informazioni accurate e georeferenziate sulle località, utili per correlare le caratteristiche demografiche e infrastrutturali. D'altra parte, OSM, una piattaforma di mappatura collaborativa, permette di accedere a dati aggiornati relativi alla rete stradale, utili per monitorare la situazione attuale delle infrastrutture. QGIS, un software open-source per la gestione dei sistemi informativi geografici, offre strumenti avanzati per combinare e analizzare questi dati, come l'algoritmo "Join by Attributes".

Questo studio mira a identificare la lunghezza degli archi stradali in ciascuna provincia italiana utilizzando un approccio GIS. La lunghezza di ogni arco viene calcolata utilizzando la funzione `$length` di QGIS, mentre il join spaziale con i dati del censimento permette di associare gli archi stradali alle rispettive province. Infine, i dati aggregati verranno utilizzati per costruire indicatori per la sicurezza stradale.

## 2. Metodologia

### 2.1. Dati Utilizzati

- **Poligoni delle località:** I poligoni che rappresentano le località italiane sono stati estratti dal censimento ISTAT 2011. Questi dati contengono informazioni relative alla popolazione e alle caratteristiche geografiche di ciascuna località.
- **Grafo stradale:** Il grafo stradale è stato estratto da OpenStreetMap (OSM) al 2024, rappresentando la rete stradale attuale in Italia. Gli archi stradali sono rappresentati come linee geometriche (shapefile lineare) e contengono attributi quali la classificazione funzionale (strade principali, secondarie, ecc.).

### 2.2. Algoritmo "Join by Attributes"

L'algoritmo "Join by Attributes" in QGIS è uno strumento potente che consente di combinare due layer spaziali (o tabelle) basati su attributi comuni, rendendolo essenziale per l'integrazione di dati geografici complessi. In particolare, si tratta di un'operazione di tipo "join", che differisce dall'analisi spaziale pura, poiché la combinazione dei dati avviene in base a campi di attributi piuttosto che a una relazione spaziale tra geometrie.

In questo studio, l'obiettivo è unire i dati delle località censuarie, rappresentate come poligoni, con il grafo stradale di OSM, rappresentato come linee. La chiave per effettuare il "join" è un attributo condiviso dai due dataset: l'identificatore della provincia.

L'algoritmo esegue un'operazione di tipo "one-to-many" o "many-to-one", a seconda del contesto, permettendo di associare record da due tabelle o layer in base a un campo comune. Nel nostro caso, l'attributo comune è il codice o il nome della provincia, presente sia nel layer delle località censuarie che nel grafo stradale. Di seguito viene descritto il funzionamento dell'algoritmo:

- **Identificazione dell'attributo comune:** Prima di eseguire il "join", bisogna identificare un attributo comune ai due dataset. Nel nostro caso, entrambi i dataset contengono un campo che rappresenta la provincia (ad esempio, `provincia_code` o `provincia_name`).
- **Creazione del "join":** L'algoritmo "Join by Attributes" consente di collegare il layer delle località (poligoni) con il grafo stradale (linee). Questo è importante per associare a ogni arco stradale la provincia in cui si trova, utilizzando il campo provincia come chiave per unire i dataset.

Selezione del primo layer: Il primo dataset selezionato contiene le località censuarie. Ogni località è rappresentata da un poligono, e l'informazione della provincia è già presente tra gli attributi di ciascun poligono.

Selezione del secondo layer: Il secondo dataset contiene il grafo stradale, con linee che rappresentano le strade. Ogni segmento di strada è caratterizzato da attributi che descrivono la tipologia della strada (autostrade, provinciali, locali), ma non possiede informazioni sulla provincia.

- **Tipologia di join:** In questo caso si esegue un join "many-to-one", dove molte strade appartenenti a una determinata area geografica (provincia) vengono collegate a un singolo record nel layer delle località. Poiché ogni poligono rappresenta una singola località/provincia, il join permette di "attribuire" la provincia corretta a ciascun arco stradale.
- **Unione degli attributi:** L'algoritmo non modifica le geometrie dei layer coinvolti, ma associa i campi di attributi del layer delle località censuarie al layer stradale. Questo significa che dopo il join, gli archi stradali avranno un nuovo attributo: il codice o il nome della provincia associata. Questo passaggio è cruciale per consentire analisi successive, come il calcolo della lunghezza totale delle strade per provincia o la classificazione funzionale delle carreggiate.

L'utilizzo dell'algoritmo "Join by Attributes" offre diversi vantaggi rispetto a operazioni di join basate esclusivamente su geometrie o relazioni spaziali:

- **Efficienza computazionale:** Unire due dataset utilizzando attributi alfanumerici è generalmente meno oneroso dal punto di vista computazionale rispetto a un join spaziale, poiché non richiede il calcolo delle intersezioni geometriche o la valutazione delle distanze spaziali tra oggetti. Ciò è particolarmente rilevante quando si lavora con dataset di grandi dimensioni, come nel caso di un intero paese.
- **Precisione dei risultati:** Il join basato su attributi evita errori dovuti a piccole discrepanze nelle geometrie, come la posizione esatta di un arco stradale rispetto ai confini della località. Poiché i dati censuari sono già associati a un attributo (provincia) standardizzato e universalmente riconosciuto, l'accuratezza del join è garantita.
- **Facilità di gestione dei dati:** Una volta completato il join, i dati possono essere analizzati senza bisogno di ulteriori operazioni geometriche complesse. La gestione dei dati per provincia diventa immediata, poiché ogni arco stradale ha un attributo associato che lo collega alla provincia corretta.

Nonostante l'efficacia dell'algoritmo, esistono alcune problematiche che possono sorgere durante il processo di join:

- **Dati mancanti o errati:** Se uno dei dataset non contiene valori coerenti nel campo della provincia, il join potrebbe non riuscire o produrre risultati parziali. In tal caso, è importante eseguire un controllo preliminare per assicurarsi che i valori degli attributi corrispondano correttamente nei due dataset.
- **Dati multipli per una località:** Nel caso in cui un arco stradale attraversi più province o località, il join potrebbe produrre duplicati o assegnare l'arco a una sola provincia. Per risolvere questo problema, sarebbe necessario implementare un'analisi spaziale più avanzata, come un "clip" spaziale per dividere l'arco nelle sezioni pertinenti a ciascuna provincia.

Dopo aver completato il join, il dataset risultante può essere utilizzato per eseguire analisi più complesse. In particolare, l'associazione di ogni strada a una provincia permette di calcolare la lunghezza delle strade aggregate per provincia, nonché la classificazione funzionale degli archi stradali, fondamentali per la costruzione degli indicatori per la sicurezza stradale.

### 2.3. Calcolo della lunghezza degli archi (\$length)

Dopo aver completato il **join spaziale** tra il grafo stradale e le località, è possibile procedere con il calcolo della lunghezza di ogni arco stradale. Per questo scopo, QGIS mette a disposizione la funzione `$length`, uno strumento integrato per misurare automaticamente la distanza tra i vertici di ogni geometria lineare (in questo caso, gli archi stradali) e restituire un valore esprimibile in unità metriche, come i metri.

La funzione `$length` fa parte del set di espressioni di QGIS e consente di calcolare la lunghezza di un oggetto geometrico lineare (come una polilinea o una linea) direttamente all'interno del sistema di coordinate in uso. Questo passaggio è fondamentale per ottenere dati quantitativi utilizzabili nelle analisi successive, come la somma della lunghezza delle strade per provincia o per tipologia di strada.

Ogni segmento di strada nel grafo stradale di OpenStreetMap (OSM) è rappresentato da un arco che, dal punto di vista geometrico, è descritto come una linea che collega una serie di vertici (punti). La funzione `$length` calcola la lunghezza complessiva di ogni arco sommando le distanze tra tutti i vertici che costituiscono il segmento di strada. Il risultato è la lunghezza totale dell'arco.

Un aspetto importante del calcolo della lunghezza è il sistema di coordinate in cui i dati sono proiettati. Se il grafo stradale è proiettato in un sistema di coordinate geografiche (ad esempio, con latitudine e longitudine), il calcolo della lunghezza potrebbe non essere accurato, poiché le distanze tra punti espressi in gradi non corrispondono a distanze reali sulla superficie terrestre. Per ottenere misure precise, è necessario che il layer sia proiettato in un sistema di coordinate planari (ad esempio, UTM), dove le distanze vengono espresse in metri o chilometri. In QGIS, la funzione `$length` può essere configurata per restituire la lunghezza in unità coerenti con il sistema di coordinate del progetto o del layer, garantendo che i valori siano corretti per le analisi successive.

Spesso, i dati geografici vengono proiettati in questo sistema di coordinate denominato UTM (Universal Transverse Mercator), poiché suddivide la superficie terrestre in zone che permettono di ottenere misurazioni accurate in metri. Il grafo stradale estratto da OSM può essere convertito in UTM per garantire che la lunghezza degli archi venga calcolata correttamente.

La precisione del calcolo dipende anche dalla qualità e dalla risoluzione del grafo stradale. In OpenStreetMap, le strade sono rappresentate da linee spezzate (polilinee), formate da più vertici. Un numero maggiore di vertici lungo una strada permette di rappresentarne la forma in modo più dettagliato e accurato, migliorando così la precisione del calcolo della lunghezza totale.

Dopo il calcolo della lunghezza per ogni arco, questi valori diventano immediatamente utilizzabili per una vasta gamma di analisi.

Con la lunghezza calcolata per ogni arco e l'associazione di ciascun arco alla provincia tramite il join spaziale, è possibile sommare la lunghezza di tutte le strade all'interno di una determinata provincia. Questo permette di ottenere la lunghezza totale della rete stradale per ogni area amministrativa, facilitando la comparazione tra province.

Ogni arco stradale nel dataset di OSM contiene informazioni relative alla classe funzionale della strada, come autostrade, strade provinciali o comunali. La funzione `$length` permette di calcolare la lunghezza delle strade per ciascuna di queste categorie. Ad esempio, si può sommare la lunghezza di tutte le autostrade (classe funzionale A1) o di tutte le strade urbane locali (classe funzionale C1) per ottenere una visione complessiva della distribuzione delle infrastrutture stradali in una provincia.

Conoscendo la lunghezza totale delle strade in una provincia e la superficie della provincia stessa, è possibile calcolare la **densità stradale**, espressa in metri di strada per chilometro quadrato ( $\text{m}/\text{km}^2$ ). Questo indicatore è utile per comprendere la densità e l'accessibilità delle infrastrutture stradali in diverse aree del territorio e per correlare tali dati con altri fattori, come la popolazione o gli incidenti stradali.

La lunghezza delle strade per classe funzionale è un elemento chiave per costruire indicatori di sicurezza. Ad esempio, la presenza di lunghi tratti di autostrada o di strade ad alta velocità in una provincia potrebbe correlarsi con un aumento del rischio di incidenti. Indicatori come il **rapporto tra la lunghezza delle strade principali (autostrade)** e le strade locali possono fornire informazioni su dove concentrare le politiche di sicurezza stradale, come l'installazione di barriere protettive o miglioramenti alla segnaletica.

QGIS consente di automatizzare il calcolo della lunghezza per ogni arco in modo massivo e rapido. Questo è particolarmente utile quando si lavora con dataset estesi come il grafo stradale di un'intera nazione.

I risultati del calcolo della lunghezza possono essere facilmente combinati con altre analisi spaziali, come buffer, analisi di densità o la modellizzazione di rete (network analysis). La funzione `$length` può quindi essere vista come un passaggio intermedio ma fondamentale per operazioni GIS più complesse.

Oltre al calcolo della lunghezza lineare, la funzione `$length` può essere combinata con altre espressioni o strumenti di calcolo di campi in QGIS per ottenere misure specifiche per vari contesti analitici, come la ponderazione della lunghezza in base alla classe funzionale o alla tipologia di superficie stradale.

L'uso della funzione `$length` in QGIS per il calcolo della lunghezza degli archi stradali è un passaggio cruciale per la quantificazione delle infrastrutture stradali e per lo sviluppo di indicatori che possono supportare l'analisi della sicurezza stradale e la pianificazione del territorio. Questa funzione fornisce una base solida per ulteriori analisi, permettendo di esplorare le correlazioni tra la

lunghezza delle strade, la loro classificazione funzionale e altri fattori critici come il rischio di incidenti e l'accessibilità delle aree urbane e rurali.

## 2.4. Classificazione per provincia

Una volta eseguito il **join spaziale** tra il layer delle località (province) e il grafo stradale, il passo successivo è quello di **aggregare gli archi stradali per provincia**. Questa operazione consente di raggruppare gli archi (o segmenti di strada) in base alla provincia in cui si trovano, rendendo possibile l'analisi delle infrastrutture stradali a livello amministrativo e territoriale.

L'attributo "provincia" associato al layer delle località censuarie diventa essenziale per classificare ogni arco stradale all'interno della sua provincia di appartenenza. L'operazione di **join by attributes**, descritta precedentemente, ha già legato ciascun arco stradale alla relativa provincia utilizzando un campo comune, come il codice o il nome della provincia. Una volta completata questa fase, ogni arco stradale contiene l'informazione sulla provincia in cui è localizzato, permettendo analisi territoriali più approfondite.

L'aggregazione degli archi stradali per provincia rappresenta un passaggio chiave nel workflow analitico. Questo permette di **raggruppare** gli archi stradali secondo il loro attributo "provincia" e calcolare vari indicatori geografici e infrastrutturali su base provinciale. In pratica, si tratta di una procedura che consente di **suddividere** la rete stradale su scala territoriale più ridotta, ossia le singole province, mantenendo al contempo il dettaglio spaziale e funzionale del grafo stradale.

Dopo aver aggregato gli archi stradali per provincia, è possibile utilizzare la funzione `$length`, descritta in precedenza, per calcolare la lunghezza complessiva delle strade per ciascuna provincia. Questo passaggio genera un indicatore quantitativo della **lunghezza totale delle strade** (espressa in metri o chilometri) all'interno dei confini di ogni provincia.

QGIS offre strumenti nativi per eseguire operazioni di aggregazione dei dati. La funzione "Aggregate" permette di raggruppare gli archi stradali in base a un determinato attributo, in questo caso "provincia", e di applicare funzioni di calcolo come la somma delle lunghezze degli archi.

Si crea un raggruppamento basato sull'attributo "provincia", e si sommano le lunghezze di tutti gli archi stradali associati a quella provincia.

Per ogni provincia, viene calcolata la somma delle lunghezze di tutti gli archi, generando un singolo valore che rappresenta la lunghezza totale della rete stradale in quella provincia.

Questo processo fornisce un risultato finale dove, per ogni provincia, è disponibile la **lunghezza totale della rete stradale**. Il risultato è una tabella che contiene le seguenti colonne principali:

- **Provincia:** Nome o codice della provincia.
- **Lunghezza totale delle strade:** La somma della lunghezza di tutti gli archi stradali all'interno della provincia (espressa in metri o chilometri, a seconda del sistema di coordinate utilizzato).

Ad esempio, per la provincia di Milano, il risultato potrebbe indicare che ci sono **5.000 chilometri di strade** in totale, mentre per una provincia più piccola come Lodi, la lunghezza totale potrebbe essere significativamente inferiore.

Una volta calcolata la lunghezza totale delle strade per ciascuna provincia, diventa possibile condurre una serie di analisi geografiche e infrastrutturali su larga scala, come:

1. **Confronto tra Province:** Con le lunghezze totali delle strade per provincia, è possibile fare un **confronto quantitativo** tra le diverse province, mettendo in evidenza le aree con reti stradali più estese. Ad esempio, si può notare che province altamente urbanizzate o più densamente popolate, come Milano o Roma, avranno probabilmente una lunghezza totale di strade molto superiore rispetto a province rurali o montane.

Questo confronto fornisce anche informazioni utili per comprendere la distribuzione delle infrastrutture stradali a livello regionale o nazionale, evidenziando eventuali disparità nella disponibilità di vie di comunicazione tra aree diverse.

2. **Analisi della Densità Stradale:** La lunghezza totale delle strade per provincia può essere ulteriormente utilizzata per calcolare la **densità stradale**, che rappresenta la quantità di strada (in metri o chilometri) per chilometro quadrato di superficie della provincia. La densità stradale è un indicatore importante per valutare il grado di sviluppo delle infrastrutture di trasporto in una determinata area.

La formula per calcolare la densità stradale è la seguente:

$$\text{Densità stradale} = \text{Lunghezza totale delle strade} / \text{Superficie della provincia (in km}^2\text{)}$$

Questo indicatore permette di evidenziare se un'area è ben servita da infrastrutture stradali o se vi è una carenza di strade, contribuendo così alla pianificazione delle politiche di trasporto e alla gestione del territorio.

3. **Identificazione delle Aree con Infrastrutture Stradali Insufficienti:** La lunghezza totale delle strade può essere confrontata con altri indicatori socio-economici, come la popolazione o il traffico veicolare, per identificare province che presentano infrastrutture stradali insufficienti rispetto al loro fabbisogno. Questo è particolarmente rilevante nelle aree in cui la popolazione è in crescita o dove vi è una forte domanda di mobilità.

Le province con una bassa lunghezza totale di strade rispetto alla loro popolazione potrebbero necessitare di investimenti nelle infrastrutture di trasporto, mentre le province con una lunghezza elevata ma una densità di popolazione bassa potrebbero dover ottimizzare la gestione delle infrastrutture esistenti.

4. **Valutazione della Sicurezza Stradale:** L'aggregazione degli archi per provincia permette anche di esaminare la **sicurezza stradale**. Si può ad esempio confrontare la lunghezza delle strade principali (come autostrade e strade statali) con quella delle strade secondarie e locali, per valutare quali tipi di strade prevalgono in ciascuna provincia. Province con una rete stradale composta principalmente da strade locali potrebbero avere un rischio maggiore di incidenti a causa della minore qualità delle infrastrutture.

Inoltre, conoscendo la lunghezza delle strade per provincia, è possibile combinare queste informazioni con dati sugli incidenti stradali per identificare i tratti di strada più pericolosi e pianificare interventi mirati per migliorare la sicurezza.

L'aggregazione degli archi stradali per provincia, seguita dalla somma delle lunghezze, fornisce un'analisi quantitativa robusta e dettagliata della rete stradale distribuita a livello territoriale. Questo



approccio offre una visione completa delle infrastrutture stradali e contribuisce alla pianificazione territoriale, alla gestione delle infrastrutture di trasporto e alla costruzione di indicatori per la sicurezza stradale.

### 3. Aspetti Computazionali

L'uso di **algoritmi GIS** per l'analisi spaziale su dataset di grandi dimensioni, come nel caso del **censimento nazionale delle località** e del **grafo stradale di OpenStreetMap (OSM)**, richiede una serie di accorgimenti per garantire che l'analisi venga eseguita in modo efficiente, ottimizzando sia il tempo di calcolo che la gestione delle risorse hardware. In particolare, l'algoritmo **"Join by Attributes"**, utilizzato in questo studio, si è dimostrato una scelta efficace per l'unione di dati spaziali complessi. Tuttavia, il suo utilizzo su scala nazionale presenta delle sfide computazionali che devono essere affrontate.

L'algoritmo **"Join by Attributes"** in QGIS è uno strumento utilizzato per unire due dataset spaziali (come località censuarie e grafo stradale) in base a un attributo comune, consentendo di integrare informazioni alfanumeriche e spaziali. Nel contesto di un dataset esteso come quello del censimento nazionale e delle strade di OSM, il processo di join comporta un carico computazionale significativo, principalmente a causa delle dimensioni dei dati e della complessità spaziale delle geometrie.

#### 1. Efficienza dell'Algoritmo:

- **Scalabilità:** Il **"Join by Attributes"** è progettato per gestire dataset di varie dimensioni, e la sua efficienza dipende da una serie di fattori, inclusi la dimensione del dataset, il numero di attributi da confrontare, e la complessità geometrica dei dati spaziali. In questo studio, l'efficienza dell'algoritmo è stata valutata attraverso il tempo di esecuzione e l'uso delle risorse di calcolo.
- **Ottimizzazione spaziale:** L'algoritmo in QGIS supporta l'uso di indici spaziali per migliorare l'efficienza nel processo di unione. Gli indici spaziali organizzano i dati in modo da ridurre il numero di confronti necessari tra le geometrie, accelerando notevolmente il processo di join. Questo è particolarmente importante quando si lavora con milioni di geometrie, come nel caso del grafo stradale nazionale.

Uno dei principali fattori che influenza l'esecuzione dell'algoritmo è il **tempo di calcolo**, che varia in funzione delle dimensioni del dataset e delle risorse hardware disponibili.

#### 1. Dimensione del dataset:

- Il censimento nazionale, che comprende milioni di località censuarie distribuite su tutto il territorio, e il grafo stradale di OSM, che contiene dettagli su autostrade, strade secondarie e urbane, formano un dataset di grandi dimensioni. La complessità aumenta esponenzialmente quando si cerca di unire tali dataset basati su attributi come la posizione e la provincia.
- Per dataset di scala nazionale, il tempo di esecuzione dell'algoritmo **"Join by Attributes"** può variare da diversi minuti a diverse ore, a seconda delle risorse computazionali e dell'ottimizzazione dei dati.

#### 2. Impatto delle risorse hardware:

- Le **risorse hardware** a disposizione influenzano fortemente i tempi di esecuzione. L'algoritmo richiede un uso intensivo di CPU, RAM e memoria di archiviazione, poiché deve caricare e confrontare grandi volumi di dati spaziali.

- Sistemi con processori multicore, grande capacità di memoria RAM e dischi SSD veloci possono ridurre significativamente i tempi di esecuzione. Tuttavia, senza un'adeguata ottimizzazione del dataset, anche hardware di alto livello può essere limitato dalla quantità di dati da elaborare.

### 3. **Parallelizzazione:**

- QGIS supporta in parte la **parallelizzazione** dei processi, il che significa che su macchine multi-core, l'algoritmo può eseguire calcoli su più core in parallelo, riducendo ulteriormente il tempo di calcolo. Tuttavia, non tutte le operazioni di join sono completamente parallelizzabili, e quindi l'efficienza finale dipende dalla configurazione specifica del software e dell'hardware.

La **gestione della memoria** è un altro aspetto cruciale quando si lavora con grandi dataset spaziali. L'uso di risorse di memoria efficienti non solo riduce i tempi di esecuzione, ma previene anche il rischio di esaurimento della memoria, che potrebbe portare a errori di sistema o crash durante il processo di calcolo.

#### 1. **Utilizzo di Indici Spaziali:**

- QGIS consente di creare **indici spaziali**, strutture di dati che migliorano la velocità di accesso ai dati spaziali durante l'esecuzione di operazioni come il join spaziale. Questi indici (ad esempio, un R-Tree) permettono di ridurre il numero di confronti necessari tra le geometrie, accelerando notevolmente il processo di join su dataset di grandi dimensioni.
- La costruzione di un indice spaziale richiede un certo tempo e risorse, ma il beneficio in termini di velocità complessiva dell'operazione è notevole. In un contesto come quello di uno studio nazionale, la costruzione di indici spaziali per entrambi i dataset (località censuarie e grafo stradale) è essenziale per ottimizzare il processo.

#### 2. **Segmentazione dei Dati in Batch:**

- Per gestire meglio la memoria, QGIS permette di suddividere dataset di grandi dimensioni in **batch** più piccoli. Invece di processare l'intero dataset nazionale in un'unica operazione, i dati possono essere suddivisi per regioni o province e processati separatamente. Questo approccio riduce il carico di memoria per ogni singola operazione, rendendo più gestibile l'intero processo.
- Una volta completati i calcoli per ciascun batch, i risultati possono essere uniti nuovamente per ottenere un output finale a livello nazionale. Questa tecnica è utile soprattutto su sistemi con capacità di memoria limitate, evitando il sovraccarico di RAM e aumentando la stabilità del processo.

#### 3. **Uso di File Temporanei e Cache:**

- Durante le operazioni di calcolo, QGIS fa uso di **file temporanei** e di una **cache** di memoria per mantenere i dati temporanei necessari per le operazioni in corso. Ottimizzare l'uso della cache e gestire i file temporanei su dischi veloci (come SSD) può ridurre i tempi di accesso ai dati e migliorare la velocità complessiva del processo.
- È importante configurare correttamente la gestione dei file temporanei in QGIS per evitare che la memoria venga sovraccaricata o che lo spazio su disco si esaurisca, soprattutto quando si lavora con dataset spaziali di grandi dimensioni.

Possibili strumenti avanzati per l'ottimizzazione computazionale in QGIS possono essere così sintetizzati:

#### 1. **Processing Toolbox e GDAL:**

- La **Processing Toolbox** di QGIS offre accesso a strumenti avanzati per l'ottimizzazione computazionale, inclusi quelli basati su **GDAL** (Geospatial Data

Abstraction Library). Questi strumenti permettono di gestire efficientemente grandi dataset, eseguendo operazioni come la riproiezione, il clipping e il merge di geometrie spaziali in modo più efficiente dal punto di vista computazionale.

- Utilizzare le librerie GDAL direttamente o attraverso la Processing Toolbox può ridurre significativamente i tempi di calcolo rispetto agli strumenti standard di QGIS, soprattutto quando si gestiscono dataset estesi.

## 2. Python e Script Automatizzati:

- Per ottimizzare ulteriormente l'analisi computazionale su grandi dataset, QGIS permette di creare **script Python** personalizzati, che possono gestire in modo più flessibile e controllato il flusso di lavoro, specialmente in operazioni ripetitive o su dataset segmentati.
- Python consente di introdurre controlli di memoria più avanzati e di integrare la parallelizzazione in modo più efficace, migliorando l'efficienza complessiva del processo.

L'esecuzione di algoritmi GIS come "**Join by Attributes**" su dataset di grandi dimensioni, come il censimento nazionale delle località e il grafo stradale di OSM, presenta sfide computazionali legate alla gestione del tempo di calcolo e della memoria. L'ottimizzazione tramite indici spaziali, segmentazione dei dati in batch e l'uso di risorse hardware avanzate permette.

## 4. Risultati

### 4.1 Lunghezza degli Archi per Provincia

L'analisi ha prodotto una serie di tabelle contenenti la lunghezza totale degli archi stradali per provincia, classificati per tipologia funzionale. Gli archi sono stati raggruppati secondo la classe funzionale presente nei dati OSM, includendo categorie quali:

- **Strade principali (A1, A2):** autostrade e strade di grande percorrenza.
- **Strade secondarie (B1, B2):** strade minori, provinciali e regionali.
- **Strade locali (C1, C2):** strade urbane, di quartiere o comunali.

I risultati mostrano che le province con la maggiore estensione di carreggiate, in termini di lunghezza totale degli archi stradali, sono quelle con ampie reti di autostrade e strade di percorrenza veloce, come Roma, Milano e Torino. In queste aree, le infrastrutture stradali non solo supportano traffico locale, ma fungono anche da snodi cruciali per il traffico interprovinciale e nazionale.

*Tabella 1: Esempio di lunghezza degli archi per provincia (dati sintetici)*

Provincia	Lunghezza totale archi (m)	Classe Funzionale A1 (m)	Classe Funzionale B1 (m)	Classe Funzionale C1 (m)
Roma	320,540	120,000	80,000	120,540
Milano	280,000	110,000	70,000	100,000
Torino	250,000	100,000	60,000	90,000

## 4.2 Distribuzione delle Carreggiate per Classe Funzionale

La distribuzione delle carreggiate per classe funzionale ha evidenziato alcune differenze chiave tra le province. Le province con maggior numero di strade locali (classe funzionale C1) mostrano un'alta densità di infrastrutture urbane, con brevi tratti di strade che servono principalmente le popolazioni locali. Al contrario, le province con più chilometri di autostrade (classe funzionale A1) presentano una maggiore predisposizione per il traffico commerciale e interregionale.

## 4.3 Indicatori per la Sicurezza Stradale

Uno degli obiettivi principali di questa ricerca è la costruzione di indicatori per la sicurezza stradale basati sulla lunghezza e sulla distribuzione delle carreggiate. Gli indicatori sono stati costruiti tenendo conto di vari parametri, tra cui la lunghezza degli archi per classe funzionale e la densità delle strade in rapporto alla popolazione provinciale. Gli indicatori elaborati includono:

- **Densità di carreggiata ( $\text{m}/\text{km}^2$ ):** lunghezza totale delle strade in relazione alla superficie della provincia.
- **Lunghezza media degli archi per classe funzionale:** indica l'estensione media delle strade classificate come A1, B1, C1 in ciascuna provincia.
- **Indice di rischio per carreggiata:** basato su fattori quali la lunghezza delle strade di grande percorrenza (A1 e A2) e la frequenza di incidenti stradali storici (dati incidentali provenienti dalla rilevazione statistica Istat/ACI).

*Tabella 2: Esempio di indicatori per la sicurezza stradale*

Provincia	Densità carreggiate ( $\text{m}/\text{km}^2$ )	Lunghezza media A1 (m)	Indice di rischio (su 100)
Roma	1200	4000	85
Milano	1500	3800	80
Torino	1100	4200	75

I dati indicano che le province con una maggiore densità di infrastrutture stradali mostrano un maggiore rischio per incidenti, specialmente nelle zone urbane ad alta densità.

## 5. Discussione

L'analisi dei dati stradali, in particolare quelli relativi alla lunghezza, alla tipologia delle strade e alla loro distribuzione territoriale, è un passaggio fondamentale per la costruzione di **indicatori di sicurezza stradale**. Questi indicatori sono utilizzati non solo per migliorare la pianificazione urbana, ma anche per lo sviluppo di politiche volte a ridurre il rischio di incidenti stradali. Province come Roma e Milano, caratterizzate da un'alta densità infrastrutturale e da una grande estensione di strade ad alto traffico, rappresentano esempi chiave per comprendere come la rete stradale e l'intensità del traffico influenzino il livello di rischio stradale.

I principali indicatori di sicurezza stradale e di pianificazione urbana sono qui di seguito descritti:

### 1. Distribuzione delle Infrastrutture e Densità Stradale:

- La **densità stradale**, ovvero la quantità di chilometri di strade per chilometro quadrato di territorio, è uno degli indicatori principali utilizzati per valutare la qualità delle infrastrutture di trasporto in un'area. A una maggiore densità stradale può corrispondere una maggiore **intensità del traffico**, con un conseguente aumento del rischio di incidenti.
- Nelle province con alta urbanizzazione come Roma e Milano, l'estensione della rete stradale è accompagnata da una concentrazione di infrastrutture di trasporto che include strade principali, autostrade, tangenziali e vie urbane a scorrimento veloce. La complessità e l'interconnessione di queste infrastrutture, se da un lato favoriscono la mobilità, dall'altro aumentano le opportunità di **conflitti veicolari**, aumentando il rischio di incidenti.

### 2. Strade di Percorrenza Veloce e Rischio di Incidenti:

- Le **strade di percorrenza veloce**, come autostrade e superstrade, sono progettate per garantire elevati flussi di traffico e velocità medie più alte. Tuttavia, queste caratteristiche implicano anche una maggiore probabilità di incidenti gravi, soprattutto in contesti ad alta densità di traffico.
- Le province di **Roma e Milano** sono dotate di estese reti di strade a percorrenza veloce, dove il traffico intenso, specialmente durante le ore di punta, aumenta la probabilità di collisioni multiple e incidenti con esiti gravi. La **statistica degli incidenti stradali** in queste province mostra chiaramente una correlazione tra l'elevata intensità del traffico e la frequenza degli incidenti.

### 3. Indicatori di Sicurezza Stradale:

- La costruzione di indicatori specifici per la **sicurezza stradale** inizia con la raccolta di dati relativi alla rete stradale, come la lunghezza delle carreggiate, la loro classificazione funzionale (autostrade, strade statali, vie urbane, ecc.) e il volume di traffico. Questi dati vengono combinati con le informazioni sugli **incidenti stradali** per calcolare indicatori come:
  - **Indice di incidentalità**: rapporto tra il numero di incidenti e la lunghezza della rete stradale in una determinata area (espressa in incidenti per chilometro).
  - **Indice di gravità degli incidenti**: calcolato in base al numero di feriti e decessi rispetto al numero di incidenti totali.

- **Indice di rischio per tipologia di strada:** che confronta il rischio di incidenti su autostrade, strade urbane e strade rurali.
  - Questi indicatori permettono di identificare le **aree critiche** e di pianificare interventi mirati per ridurre la sinistrosità stradale.
4. **Densità del Traffico e Incidentalità:**
- La **densità del traffico** (numero di veicoli per chilometro di strada) è un altro fattore chiave che contribuisce all'aumento del rischio stradale. In aree metropolitane come Roma e Milano, dove il flusso di veicoli è estremamente elevato, la combinazione di traffico denso e alta velocità incrementa la probabilità di incidenti.
  - In tali contesti, gli indicatori di sicurezza stradale mostrano un rischio maggiore non solo per gli automobilisti, ma anche per pedoni e ciclisti, che spesso si trovano a condividere lo spazio con un traffico veicolare intenso.

Le principali politiche per la sicurezza stradale possono essere raggruppati nei seguenti tre punti:

1. **Sviluppo di Politiche di Mitigazione del Rischio:**

- La costruzione di indicatori dettagliati sulla sicurezza stradale fornisce ai decisori politici informazioni cruciali per sviluppare strategie di mitigazione del rischio. Tali strategie possono includere:
  - **Riduzione della velocità massima consentita** in determinati tratti stradali ad alto rischio.
  - **Adozione di tecnologie intelligenti** per il monitoraggio del traffico, come sistemi di rilevamento automatico della velocità e dei semafori intelligenti.
  - **Realizzazione di infrastrutture di sicurezza**, come barriere di protezione sui tratti autostradali più pericolosi o la creazione di zone a traffico limitato nelle aree urbane.
- Le province con un'elevata estensione di strade ad alto scorrimento, come **Milano e Roma**, possono beneficiare dell'implementazione di **interventi infrastrutturali** mirati, come il miglioramento della segnaletica stradale, l'incremento delle corsie di emergenza e la costruzione di **rotatorie** in aree di intersezione critiche.

2. **Ottimizzazione della Pianificazione Urbana:**

- L'analisi delle infrastrutture stradali, supportata dalla costruzione di indicatori per la sicurezza, fornisce le basi per una **pianificazione urbana più consapevole**. Le città densamente popolate e con reti stradali complesse, come Milano e Roma, devono affrontare la sfida di **gestire la crescita della domanda di mobilità** in modo sicuro ed efficiente.
- Una delle soluzioni proposte è la **creazione di reti di trasporto multimodale** che distribuiscano meglio il traffico veicolare, promuovendo al contempo l'uso del trasporto pubblico e la mobilità sostenibile (come piste ciclabili e percorsi pedonali). Questi approcci riducono la dipendenza dai veicoli privati, contribuendo a diminuire il rischio di incidenti.

3. **Miglioramento dell'Accessibilità e della Sicurezza per gli Utenti Vulnerabili:**

- L'analisi della rete stradale può essere utilizzata per valutare la **sicurezza degli utenti vulnerabili**, come pedoni e ciclisti. In molte province, i tratti stradali ad alta velocità attraversano zone residenziali o commerciali, aumentando il rischio per chi non utilizza veicoli a motore.
- In queste aree, politiche come la **creazione di marciapiedi e piste ciclabili protette**, l'installazione di **semafori pedonali**, e la riduzione della **larghezza delle carreggiate** possono migliorare la sicurezza. A livello urbano, queste misure sono particolarmente rilevanti in città con un traffico intenso come Roma e Milano.

L'analisi dei dati stradali e la costruzione di indicatori per la sicurezza stradale forniscono una base solida per lo sviluppo di politiche di trasporto sicure ed efficienti. Nelle province con una maggiore estensione di strade ad alto scorrimento, come Roma e Milano, gli indicatori di incidentalità mettono in evidenza un rischio elevato, dovuto all'intensità del traffico e alla complessità della rete stradale. L'utilizzo di questi indicatori non solo contribuisce a migliorare la **pianificazione urbana**, ma permette anche di identificare le aree a maggior rischio e di **sviluppare politiche di sicurezza mirate**, con l'obiettivo di ridurre gli incidenti e migliorare la qualità della vita dei cittadini.

## 5.1 Implicazioni per la Sicurezza Stradale

L'analisi della **lunghezza delle strade per classe funzionale** è un elemento cruciale per comprendere la struttura della rete stradale e per valutare il livello di rischio associato alla circolazione veicolare in determinate aree geografiche. Ogni classe funzionale di strada (autostrade, strade statali, provinciali, comunali, ecc.) presenta caratteristiche diverse in termini di capacità, velocità di percorrenza e volume di traffico, e questo incide direttamente sul rischio di incidenti.

La distribuzione per provincia e l'identificazione delle aree critiche avviene secondo i seguenti criteri:

### 1. Lunghezza delle Strade per Classe Funzionale:

- Le strade vengono classificate in base alla loro **funzione** all'interno della rete viaria, e le diverse classi funzionali presentano un livello di rischio variabile. Per esempio, le **autostrade (classe A)**, che sono progettate per veicoli che viaggiano a velocità elevate e con poche intersezioni, tendono ad avere un rischio maggiore di **incidenti gravi**, anche se la probabilità di collisione può essere inferiore rispetto alle strade urbane più trafficate.
- D'altro canto, le **strade locali (classi B1 e B2)**, pur avendo un traffico generalmente meno intenso, possono presentare rischi maggiori a causa della presenza di numerosi punti di accesso secondari, incroci e attività pedonale.

### 2. Distribuzione Geografica delle Infrastrutture Stradali:

- La distribuzione delle strade per classe funzionale varia notevolmente da provincia a provincia, con un'infrastruttura stradale di **alta capacità concentrata in aree metropolitane** o di grande transito, come Roma, Milano, Napoli e Torino. Queste aree sono caratterizzate da una grande quantità di autostrade, tangenziali e strade di scorrimento veloce, che servono un volume significativo di traffico giornaliero.
- In provincia, invece, le **strade provinciali e comunali** costituiscono la maggior parte della rete stradale, rappresentando un fattore di rischio più rilevante a causa del **maggior numero di intersezioni** e del traffico locale meno regolamentato.

**La valutazione del rischio e le politiche di sicurezza mirate può essere applicata mediante** l'analisi della lunghezza delle strade per classe funzionale e la loro distribuzione geografica non solo fornisce una visione dettagliata dell'infrastruttura stradale, ma permette di individuare le **aree ad alto rischio** e di sviluppare politiche di sicurezza stradale efficaci. Alcune delle misure che emergono da questa analisi includono:

### 1. Miglioramento della Segnaletica nelle Aree ad Alta Densità di Traffico:

- La segnaletica stradale è uno degli strumenti più efficaci per ridurre il rischio di incidenti, soprattutto nelle aree con un'elevata densità di traffico. Le **autostrade** e le **strade principali** (classi A1 e A2), che attraversano le province più urbanizzate,

possono beneficiare di un **miglioramento della segnaletica**, con l'installazione di segnali più visibili e aggiornati per avvisare i conducenti di curve pericolose, uscite autostradali, e riduzioni di velocità.

- Inoltre, in aree con traffico intenso, la **segnaletica elettronica dinamica** può fornire informazioni in tempo reale su condizioni del traffico, incidenti o rallentamenti, contribuendo a prevenire situazioni di pericolo.

## 2. **Installazione di Barriere di Sicurezza lungo le Autostrade Principali:**

- Le **barriere di sicurezza** sono fondamentali per ridurre la gravità degli incidenti, soprattutto su **autostrade ad alta capacità** come l'**A1** (Autostrada del Sole), che collega le principali città italiane. Le barriere possono prevenire che i veicoli coinvolti in incidenti fuoriescano dalla carreggiata, riducendo il rischio di collisioni frontali o di ribaltamenti.
- Nelle aree più critiche, dove il traffico è particolarmente intenso o la conformazione del terreno è più pericolosa (curve strette, discese ripide, ecc.), l'installazione di **barriere di sicurezza avanzate**, come quelle in acciaio o con tecnologie di assorbimento dell'energia, può significativamente ridurre la mortalità legata agli incidenti.

## 3. **Riduzione degli Accessi Secondari su Strade Principali (B1 e B2):**

- Le **strade principali** delle classi **B1 e B2** spesso presentano un numero elevato di **accessi secondari**, come incroci con strade minori, entrate a parcheggi o collegamenti a edifici residenziali e commerciali. Questi punti di accesso sono noti per essere **punti critici** dove si verifica un'elevata percentuale di incidenti, in particolare in presenza di traffico intenso e velocità elevate.
- Una delle politiche più efficaci per ridurre il rischio in queste aree è la **limitazione del numero di accessi secondari**, sia attraverso una regolamentazione più severa nella progettazione delle strade, sia con la **creazione di rotatorie** che riducono la velocità e regolano meglio il flusso del traffico. Questo approccio è particolarmente rilevante nelle zone rurali o semiurbane, dove la **commistione tra traffico locale e traffico ad alta velocità** è più comune.

## 4. **Interventi di Manutenzione Stradale:**

- Il **mantenimento delle condizioni delle strade** è un altro aspetto essenziale per garantire la sicurezza, soprattutto su strade di scorrimento veloce e ad alto volume di traffico. Strade deteriorate o con pavimentazioni irregolari aumentano il rischio di incidenti, soprattutto per i veicoli a due ruote.
- Interventi come il **rifacimento periodico del manto stradale**, l'aggiunta di **strisce riflettenti** e l'installazione di **dossi** nelle aree urbane possono ridurre notevolmente il rischio di sinistri. Nelle province con strade più vecchie e meno curate, questi interventi diventano cruciali per migliorare la sicurezza complessiva.

## 5. **Monitoraggio e Analisi del Traffico:**

- L'utilizzo di **sistemi di monitoraggio del traffico** basati su **tecnologie GIS**, come telecamere, sensori e software di analisi del flusso veicolare, permette alle amministrazioni locali di raccogliere dati dettagliati sulle dinamiche del traffico e di individuare i punti più critici.
- Questi strumenti possono essere impiegati per ottimizzare la distribuzione del traffico, ridurre i tempi di congestione e aumentare la sicurezza su tratti stradali critici. Le informazioni raccolte possono essere utilizzate per pianificare interventi di **manutenzione mirata**, riorganizzare la segnaletica o implementare **sistemi di gestione del traffico in tempo reale**.

L'analisi della lunghezza delle strade per classe funzionale e la loro distribuzione per provincia fornisce una base solida per la **valutazione del rischio stradale**. L'identificazione delle aree con una



concentrazione maggiore di infrastrutture di alta capacità, come le autostrade o le principali arterie di traffico, permette di focalizzare le politiche di sicurezza su interventi mirati. Questi includono il miglioramento della segnaletica, l'installazione di barriere di sicurezza e la riduzione degli accessi secondari su strade principali. Tali interventi sono essenziali per ridurre la sinistrosità, specialmente nelle aree metropolitane e nelle province con un'alta densità stradale. Ciò può essere più agevolmente rappresentata con indicatori sintetici di rischio all'esposizione ad incidente stradale secondo stratificazioni temporali del fenomeno (giorno della settimana, mese, giorni lavorativi o festivi, ora del giorno, ecc...) o per tipologia di utente della strada (automobilista, motociclista, ciclista, pedone, ecc...). Le politiche di prevenzione da applicare da parte degli organi preposti può variare sensibilmente.

## 6. Conclusioni

Questo studio ha dimostrato l'efficacia dell'uso di strumenti GIS, come l'algoritmo "Join by Attributes" in QGIS, per unire dati spaziali complessi provenienti da diverse fonti e calcolare la lunghezza delle infrastrutture stradali. Il calcolo della lunghezza degli archi stradali e la loro aggregazione per provincia fornisce una base utile per la costruzione di indicatori per la sicurezza stradale.

L'analisi mostra che le province con una maggiore concentrazione di strade ad alta capacità presentano un più alto rischio di incidenti, suggerendo la necessità di politiche di sicurezza mirate. Gli indicatori elaborati potrebbero essere utilizzati come strumenti di supporto per decisioni politiche e per interventi infrastrutturali volti a migliorare la sicurezza.

Le statistiche sperimentali prodotte con i dati della rilevazione sull'incidentalità stradale, rappresentano un primo fattivo passo avanti nell'ambito della sinergia tra fonti dati di tipo amministrativo e fonti Big Data. Futuri sviluppi di questo studio potrebbero includere l'integrazione dei dati di traffico, oltre a un'analisi temporale delle modifiche alla rete stradale valutando così l'impatto in termini di significatività nella riduzione dell'incidentalità stradale.

## Bibliografia

1. QGIS Documentation - "QGIS User Guide." (2023). Disponibile online: [https://docs.qgis.org/latest/en/docs/user\\_manual/index.html](https://docs.qgis.org/latest/en/docs/user_manual/index.html)
2. OpenStreetMap Contributors - "OpenStreetMap: The Free Wiki World Map." (2023). Disponibile online: <https://www.openstreetmap.org>
3. ISTAT - "Censimento della Popolazione e delle Abitazioni, 2011." Istituto Nazionale di Statistica (2011).
4. Smith, B. et al. - "Spatial Join Techniques in Geographic Information Systems: A Comparative Review." *International Journal of Geographical Information Science* 22(3): 359-382 (2010).
5. Schilling, A. - "OpenStreetMap and QGIS Integration: A Comprehensive Approach to Spatial Analysis." *GIScience & Remote Sensing* 51(5): 567-580 (2014).
6. ESRI - "Spatial Join in GIS: Understanding the Fundamentals." (2020). Disponibile online: <https://www.esri.com/en-us/arcgis>
7. Safe Roads Foundation - "Statistical Indicators for Road Safety: Methodologies and Application." *Journal of Road Traffic Research* 29(1): 12-19 (2022).
8. Ali M.H., M.M Jaber, S.K. Abd, et al., 2022. Big data analysis and cloud computing for smart transportation system integration. *Multimedia Tools and Application*, Springer, 1-18, Volume 22 <https://doi.org/10.1007/s11042-022-13700-7>
9. Broccoli M., S. Bruzzone, 2019 e 2021. *Utilizzo di open street map per il calcolo di indicatori per l'incidentalità stradale sulla rete viaria italiana*. In Statistiche Sperimentali. Roma: Istat <https://www.istat.it/it/archivio/231732> ; <https://www.istat.it/it/archivio/257382>
10. Broccoli M., R. Giannini, 2021. *Stima dei chilometri percorsi sul territorio nazionale con tool di webscraping (iMacros) da fonte Big Data*. In 14a Conferenza Nazionale di Statistica, Sessione Poster. Roma: Istat [023.jpg \(2160×3840\) \(istat.it\)](https://www.istat.it/it/archivio/231732)
11. Narváez-Villa P., B. Arenas-Ramírez, J. Mira, F. Aparicio-Izquierdo, 2021. *Analysis and Prediction of Vehicle Kilometers Traveled: A Case Study in Spain*, Int. J. Environ. Res. Public Health, 18(16), 8327; <https://doi.org/10.3390/ijerph18168327>
12. National Highway Traffic Safety Administration. 2018. Road Safety and Vehicle Mileage: A Comprehensive Stud, in *NHTSA Technical Report*, No. 2018-TR-123.
13. Ramin A., 2020. Application of big data in transportation safety analysis using statistical and deep learning methods. In *PhD dissertation, University of Tennessee*. [https://trace.tennessee.edu/utk\\_graddiss/5858](https://trace.tennessee.edu/utk_graddiss/5858)