

Rome Accident Analysis

Anni 2006-2021



<https://github.com/matfe/RomeAccidentAnalysis>



luglio 2024

Autori: M. Ferri e D. Napoli



Il progetto

Questo progetto si propone di analizzare gli incidenti stradali che si sono verificati nell'ambito del territorio di Roma capitale nel periodo temporale 2006-2021. La scelta di questo tema è giustificata dal fatto che un'analisi di questi eventi può dare utili e oggettive informazioni al decisore politico. Questi, pertanto, potrà attuare politiche mirate per migliorare la sicurezza stradale basandosi su evidenze fattuali, ad esempio attraverso l'individuazione delle strade più a rischio e delle aree geografiche a maggiore densità di incidenti e con un numero di vittime più elevato. La disponibilità di dati storici, inoltre, può fornire utili indicazioni statistiche sull'andamento dell'incidentalità negli anni e su eventuali tendenze significative



Gli obiettivi

Il progetto, attraverso operazioni di pulizia, trasformazione e normalizzazione dei dati grezzi e, successivamente, mediante un'analisi di dettaglio dei dati aggregati, ha come obiettivo principale la generazione di indicatori statistici che consentano di descrivere il fenomeno dell'incidentalità stradale nel territorio di Roma capitale nell'arco temporale dal 2006 al 2021. Nel dettaglio, i principali *deliverable* di progetto saranno:

- ✓ Grafici che mostrano l'andamento dell'incidentalità nel tempo, le strade con il maggior numero di incidenti e il maggior numero di vittime, le eventuali correlazioni degli incidenti con le condizioni meteorologiche, il fondo stradale e le condizioni di visibilità ecc...
- ✓ Mappe di incidentalità
- ✓ Grafi

Dataset 1/2

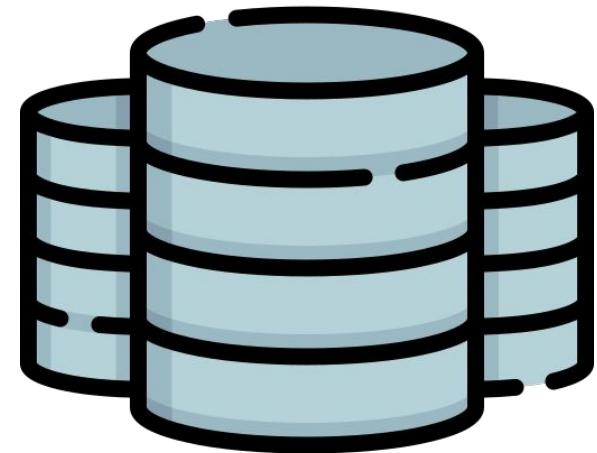
Sono stati utilizzati i dataset resi disponibili sul portale open data del Comune di Roma:

<https://dati.comune.roma.it/catalog/organization/roma-capitale-area-tematica-supporto-all-amministrazione>

Nel dettaglio:

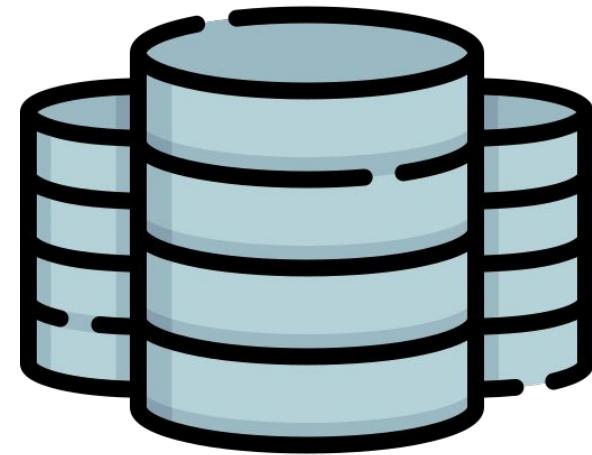
Ciascun dataset contiene l'elenco degli incidenti stradali che sono avvenuti nel territorio di Roma Capitale nell'anno di riferimento.

Ciascun dataset contiene tutti gli incidenti stradali in cui è intervenuta una pattuglia di qualunque Gruppo della Polizia Locale di Roma Capitale. Sono pertanto esclusi solo gli incidenti in cui i soggetti coinvolti abbiano raggiunto una conciliazione. Il dataset non comprende gli incidenti avvenuti sul Grande Raccordo Anulare di Roma Capitale



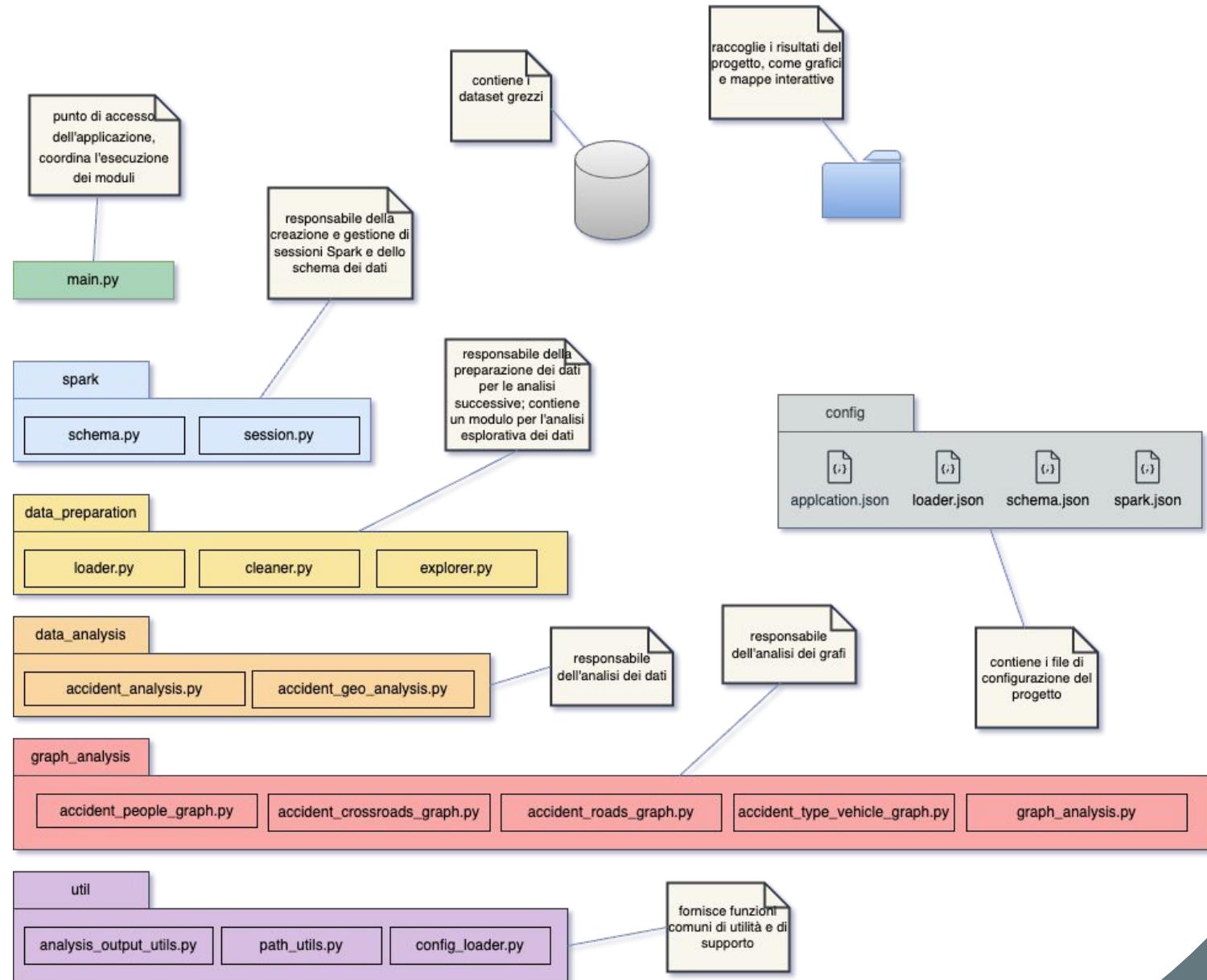
Dataset 2/2

Il dataset sugli incidenti stradali fornisce informazioni dettagliate su ogni incidente registrato. Ogni incidente è identificato da un "Protocollo" unico, che è associato alle persone coinvolte. Le informazioni includono data e ora dell'incidente, localizzazione (strade, coordinate GPS), natura dell'incidente e condizioni stradali e atmosferiche. Gli esiti sono documentati in termini di numero di feriti, morti e illesi. Per ogni persona coinvolta, sono riportati il tipo e lo stato del veicolo, il ruolo (conducente, passeggero, ecc.), il sesso, il tipo di lesione, l'utilizzo di dispositivi di sicurezza e l'eventuale decesso.



Architettura applicativa

L'applicazione ha un'architettura modulare in cui ogni pacchetto include moduli e file correlati tra loro



Preparazione dei dati 1/4

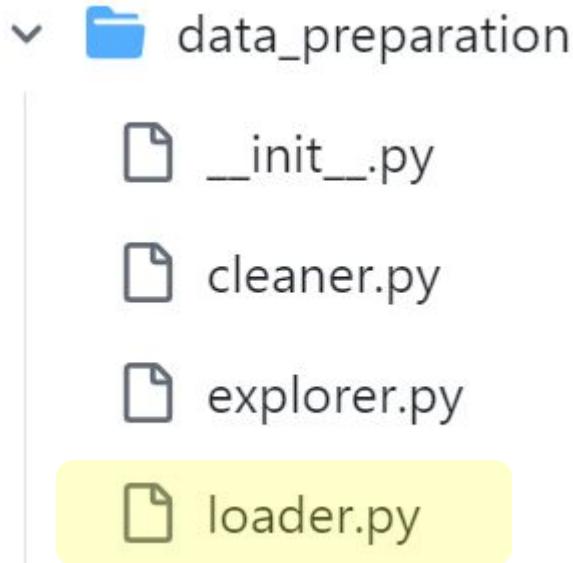
- ▼  data_preparation
 -  _init__.py
 -  cleaner.py
 -  explorer.py
 -  loader.py

Il pacchetto data_preparation contiene i moduli necessari per preparare i dati per la successiva fase di analisi.
Esso si compone dei seguenti moduli:

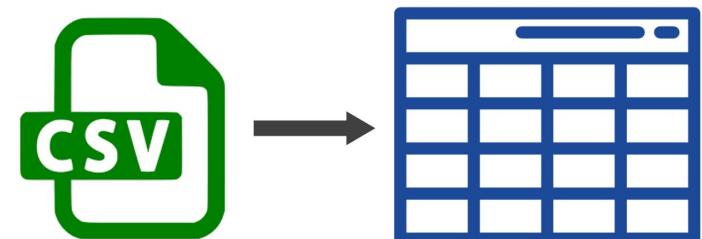
loader
cleaner
explorer



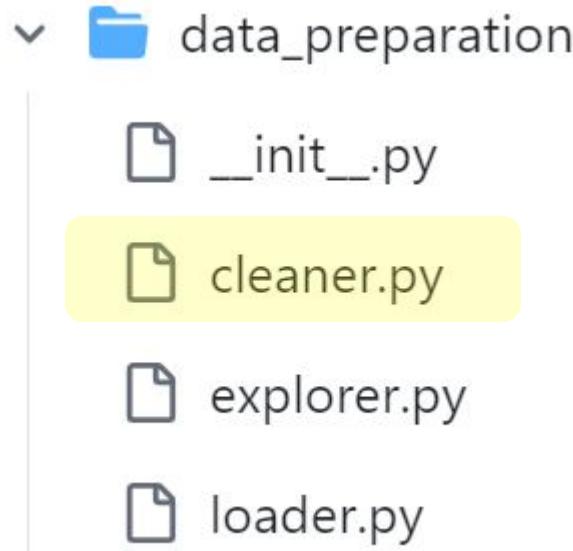
Preparazione dei dati 2/4



Il modulo **loader** si occupa del caricamento e dell'aggregazione dei dati dai file in formato csv relativi a ciascuna annualità del periodo di interesse. I dati sono caricati in un DataFrame.



Preparazione dei dati 3/4



Il modulo ***cleaner*** lavora sui dati grezzi caricati nel DataFrame. Carica configurazioni da un file JSON, rinomina colonne, sostituisce valori nulli, normalizza testo, rimuove colonne specifiche e righe duplicate. Tronca e converte DataOralIncidente in timestamp, sostituisce valori secondo un dizionario.



Data Cleaning

Preparazione dei dati 4/4

- ▼  data_preparation
 -  _init__.py
 -  cleaner.py
 -  explorer.py
 -  loader.py

Il modulo **explorer** esegue un'analisi esplorativa dei dati attraverso un insieme di compiti. Nel dettaglio, si occupa di contare i valori mancanti e unici e di salvare i risultati delle analisi in file in formato *html*.



Analisi dei dati 1/3

- ▼  data_analysis
 -  _init__.py
 -  accident_analysis.py
 -  accident_geo_analysis.py

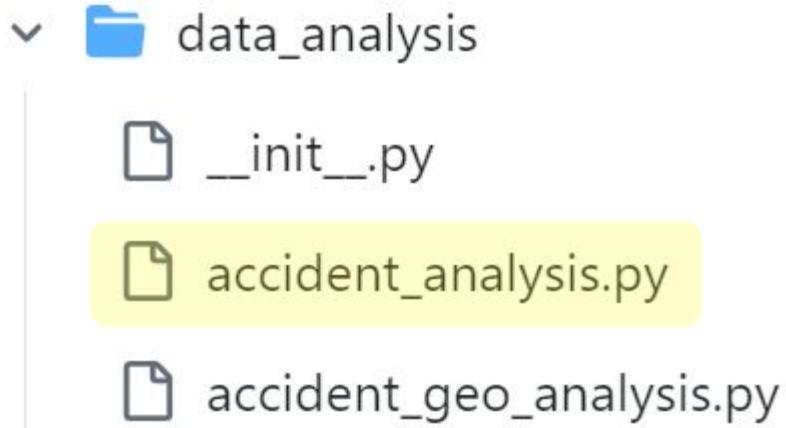
Il pacchetto `data_analysis` esegue l'analisi degli incidenti stradali a Roma, utilizzando PySpark per l'elaborazione dei dati e strumenti di visualizzazione per rappresentare tendenze temporali e distribuzioni geografiche.

Esso si compone dei seguenti moduli:

accident_analysis
accident_geo_analysis



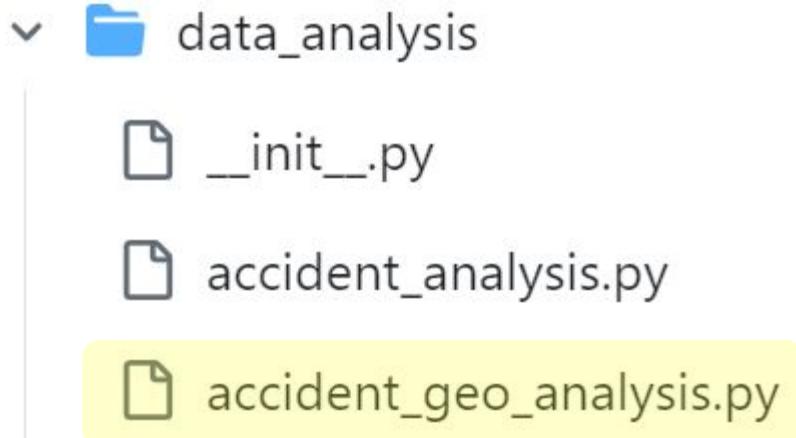
Analisi dei dati 2/3



Il modulo ***accident_analysis*** esegue l'analisi degli incidenti stradali, elaborando e aggregando i dati. Visualizza i risultati evidenziando tendenze temporali, fattori di rischio e distribuzione oraria degli incidenti.



Analisi dei dati 3/3



Il modulo ***accident_geo_analysis*** esegue l'analisi geografica degli incidenti stradali, visualizzando mappe interattive che evidenziano le zone critiche e la natura degli incidenti.



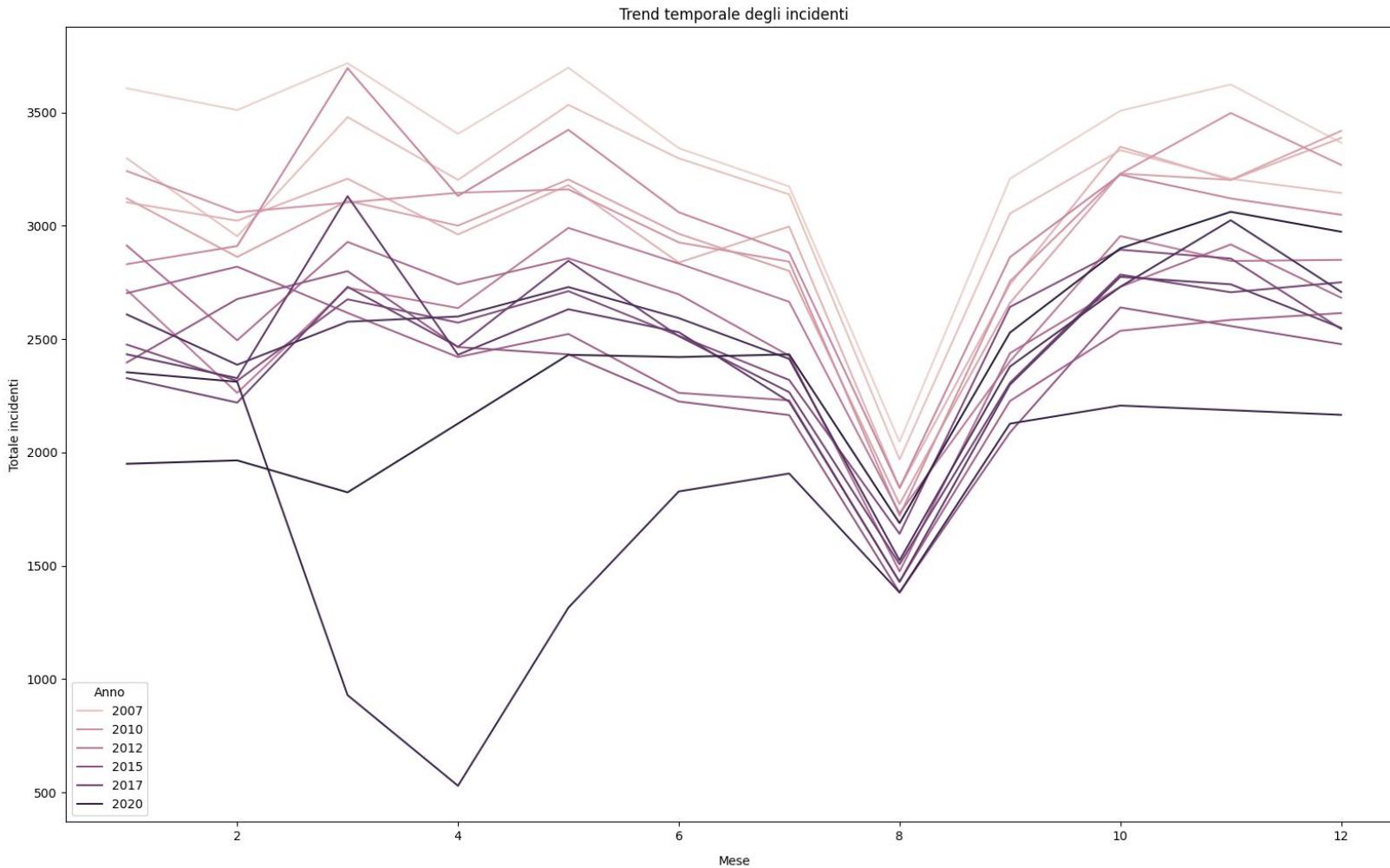
Risultati

analisi esplorativa

Sono stati analizzati i dati relativi a 1.201.165 persone coinvolte in 503.627 incidenti stradali. Numerosi valori mancanti sono stati rilevati in campi chiave come localizzazione e condizioni atmosferiche. Questi dati sono stati fondamentali per la pulizia del dataset e per identificare i campi con troppi valori nulli da rimuovere. L'analisi dei valori unici delle variabili qualitative ha aiutato a rimuovere dati sporchi. Il campo "Protocollo" non è univoco ed è associato alle persone coinvolte negli incidenti, consentendo di capire meglio le dinamiche e le conseguenze degli incidenti.

Risultati

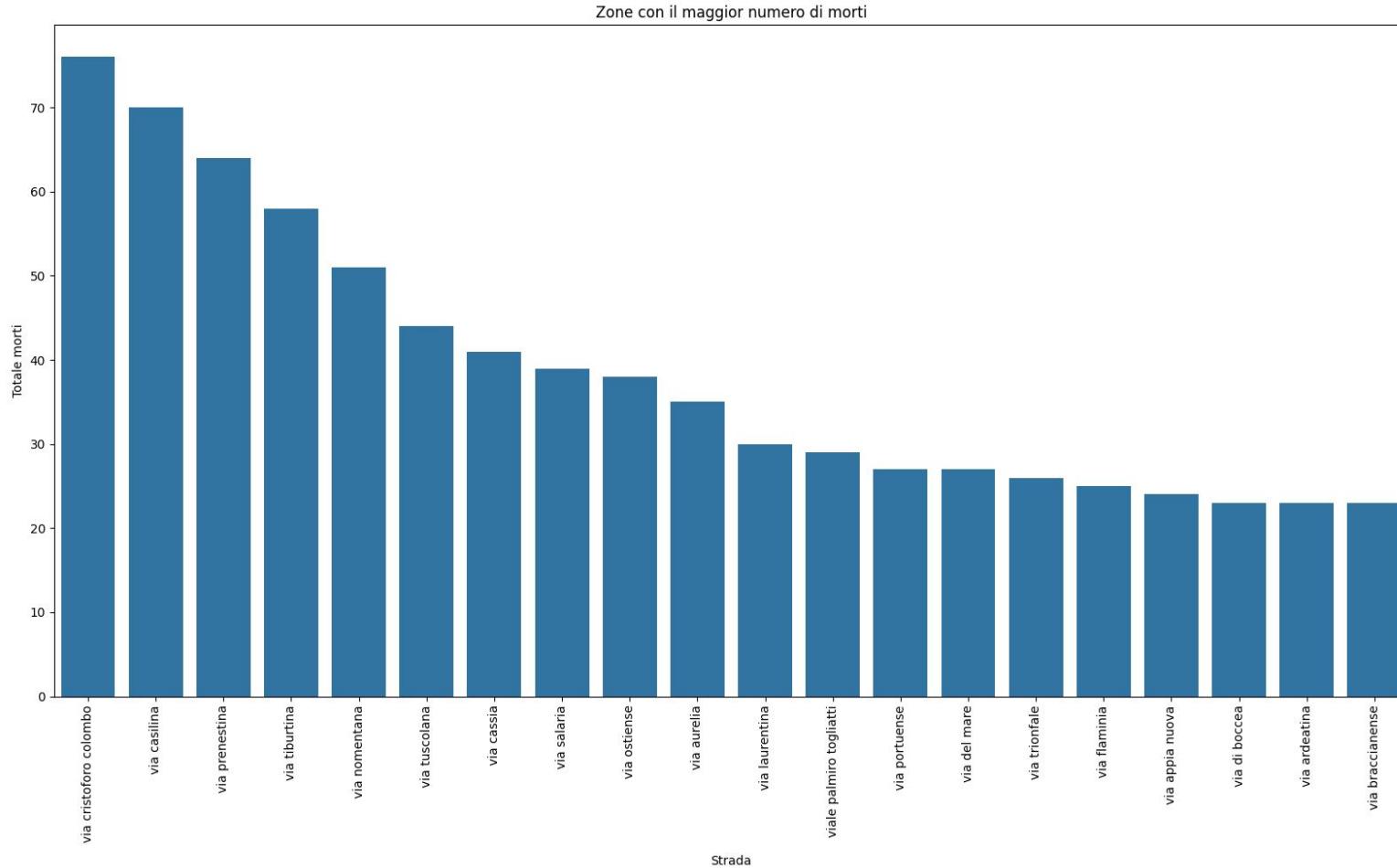
andamento temporale dei sinistri stradali



Osservando l'andamento degli incidenti, si rileva immediatamente, per tutte le annualità, un calo repentino dei sinistri stradali nel mese di agosto il che si spiega facilmente con la riduzione del traffico veicolare dovuta all'esodo vacanziero. Si rileva, inoltre, una riduzione drastica degli incidenti nell'anno 2020 a partire da marzo/aprile per l'imposizione del lockdown dovuto all'emergenza pandemica da Covid-19.

Risultati

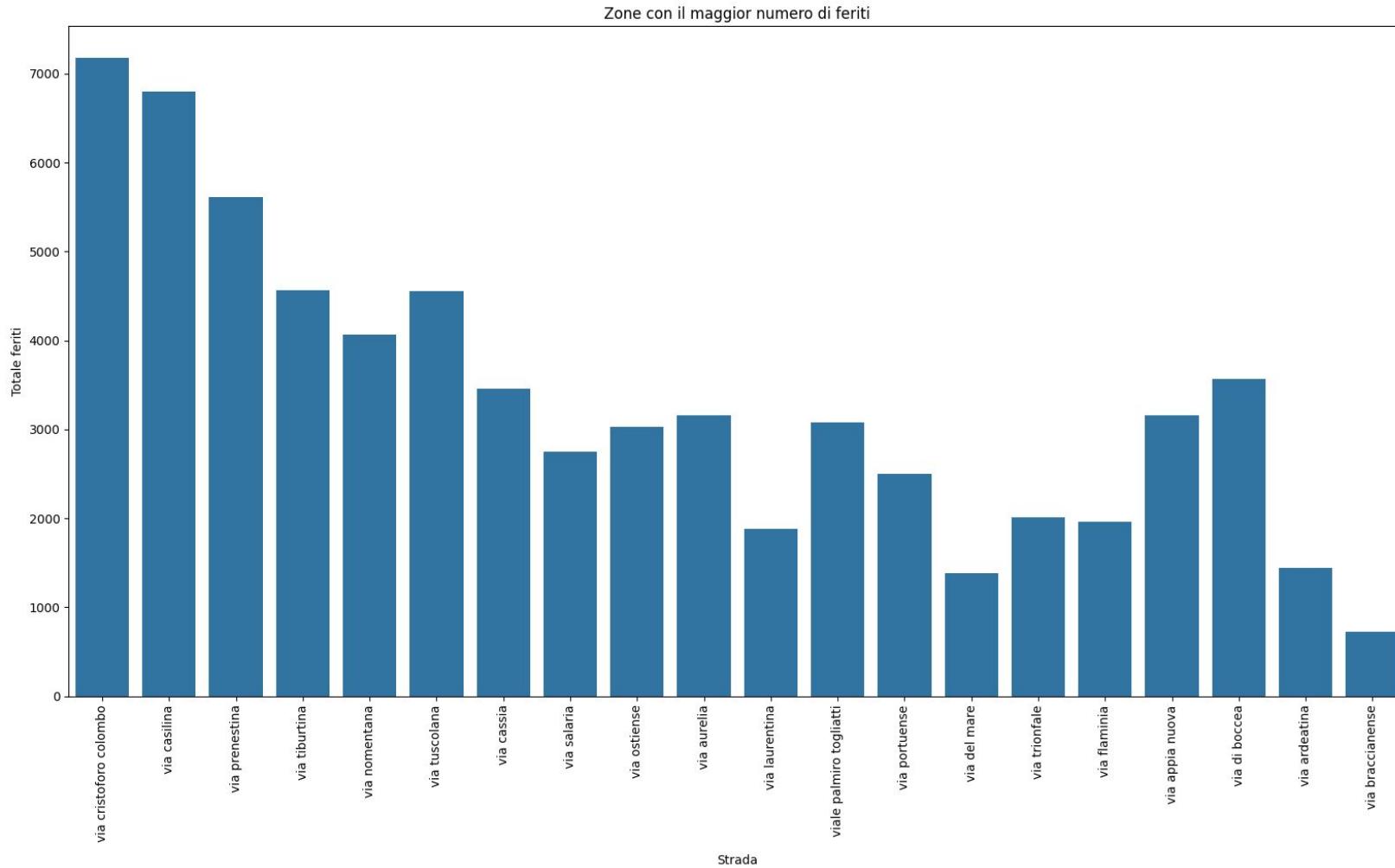
strade con il maggior numero di morti



Nel periodo di interesse considerato (2006-2021) si rileva che le strade su cui si sono registrati gli incidenti più gravi, ovvero con il maggior numero di morti, sono quelle ad elevato scorrimento nelle quali si raggiungono velocità più elevate. Il triste primato appartiene alla via Cristoforo Colombo

Risultati

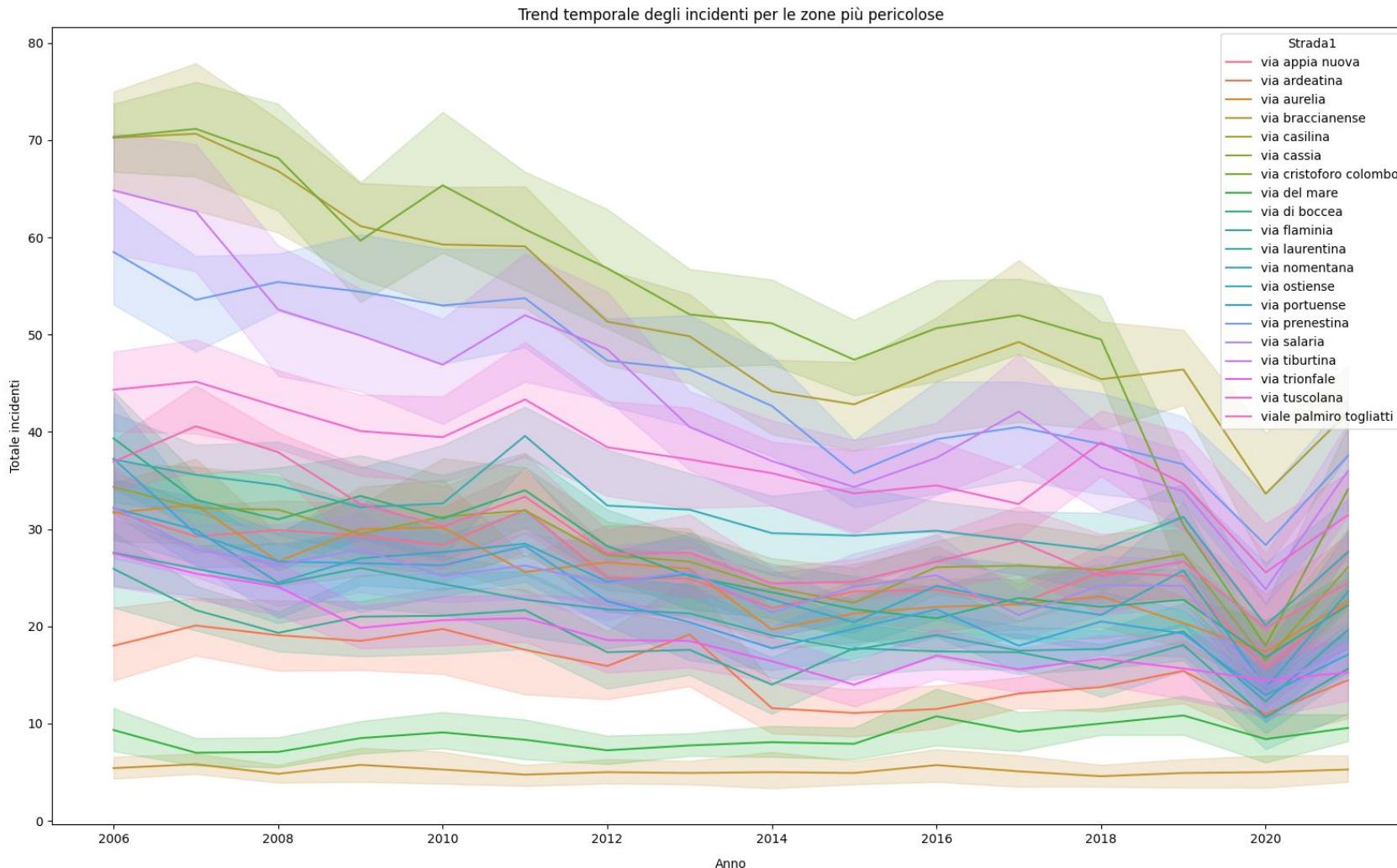
strade con il maggior numero di feriti



Anche per il numero di feriti, si rileva un andamento simile. Prima, per numero di feriti, la via Cristoforo Colombo, seguita dalla via Casilina e dalla via Prenestina

Risultati

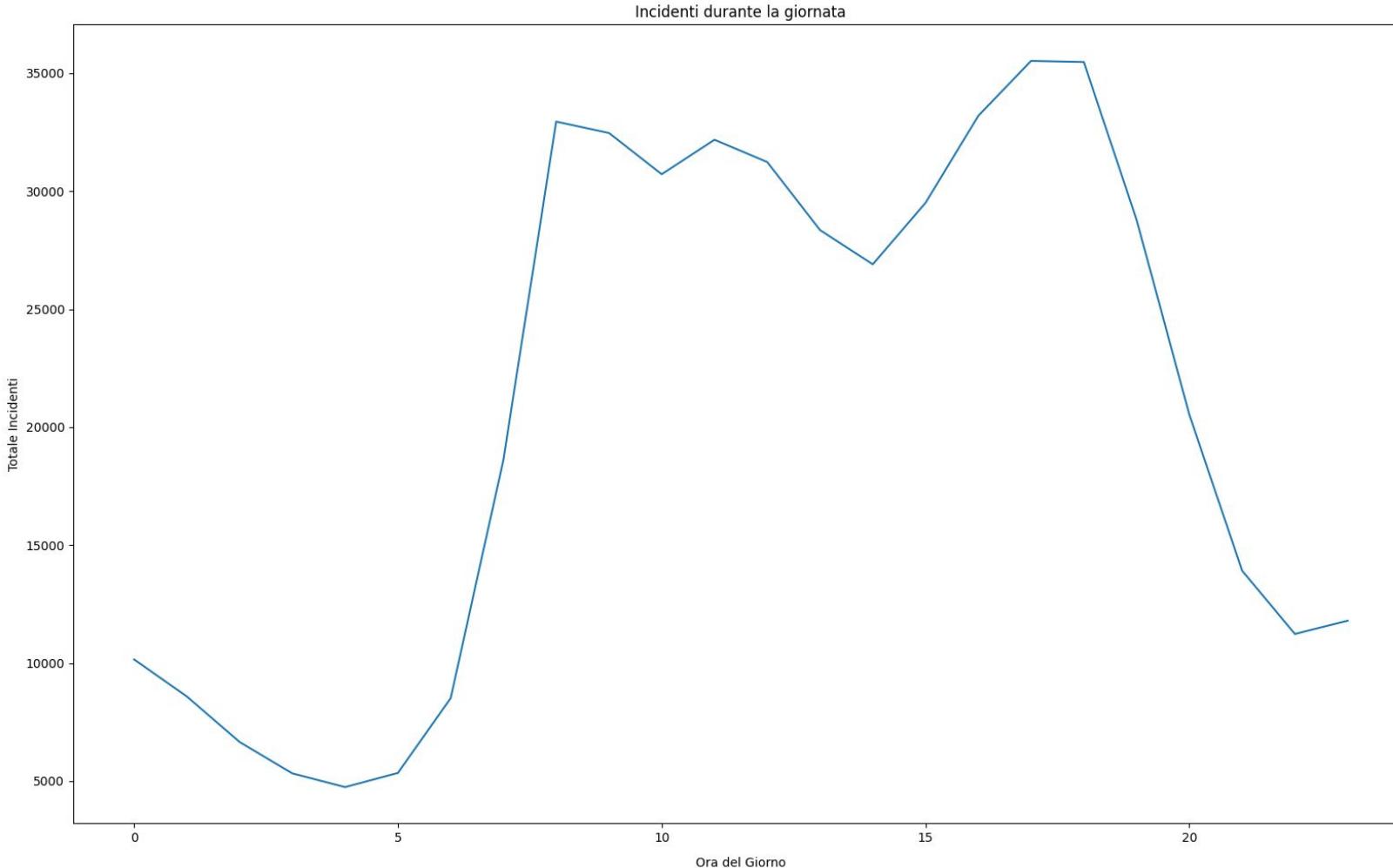
andamento temporale degli incidenti
sulle strade più pericolose



Osservando l'andamento degli incidenti nel periodo di interesse, con riferimento alle strade più pericolose, si rileva una generale decrescita dei sinistri nel tempo. Questa tendenza potrebbe essere spiegata dai concomitanti effetti di interventi pubblici per aumentare la sicurezza stradale e di ammodernamento del parco veicoli circolante con più efficaci sistemi di sicurezza (vedasi sistemi ADAS)

Risultati

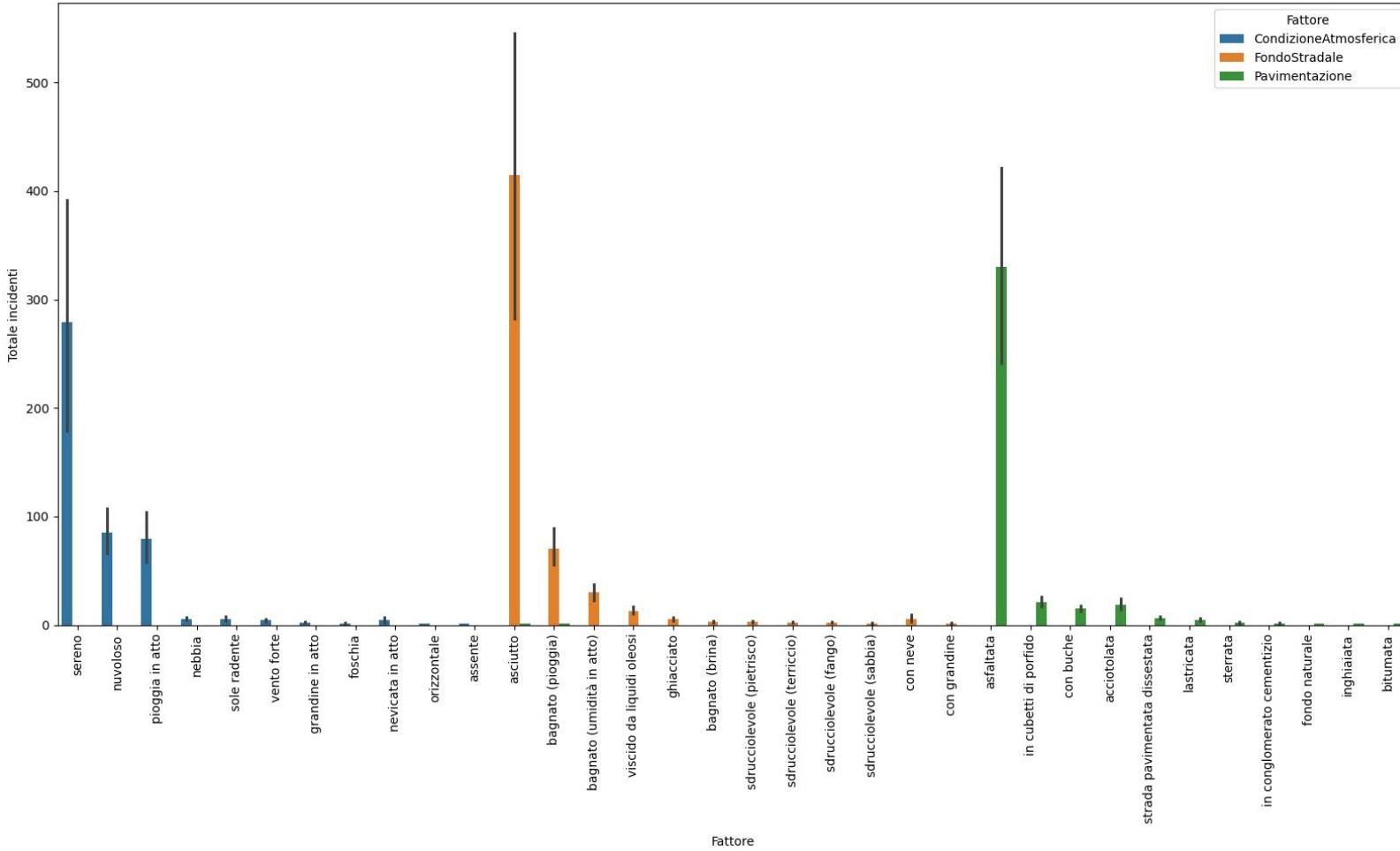
andamento orario degli incidenti



Osservando l'andamento degli incidenti nelle varie fasce orarie si rileva un comportamento caratteristico che segue la tendenza del traffico veicolare. Nelle ore notturne c'è meno traffico e si registrano meno incidenti. A partire dalle prime luci dell'alba e per tutta la mattinata c'è una crescita, poi c'è una regione di «quasi plateau» nelle ore centrali e una progressiva riduzione in serata

Risultati

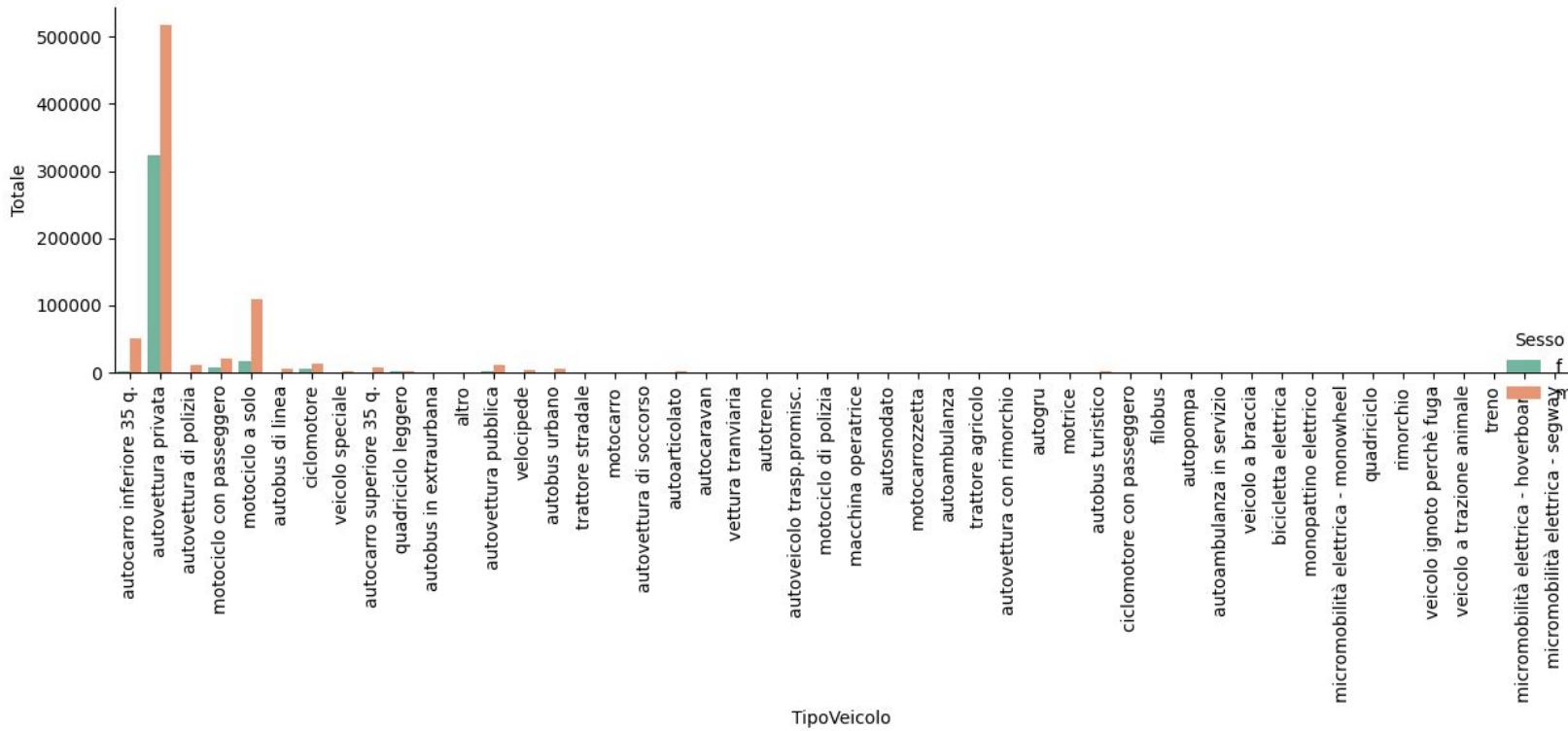
fattori che contribuiscono agli incidenti



Si nota una maggiore incidenza di incidenti in condizioni atmosferiche serene e su strade asciutte, seguiti da strade bagnate e pavimentazione asfaltata. Anche se meno frequenti, non vanno ignorati gli incidenti in condizioni di pioggia, nebbia o su pavimentazioni meno comuni come ciottolato e buche, che rappresentano rischi specifici da approfondire per migliorare la sicurezza stradale.

Risultati

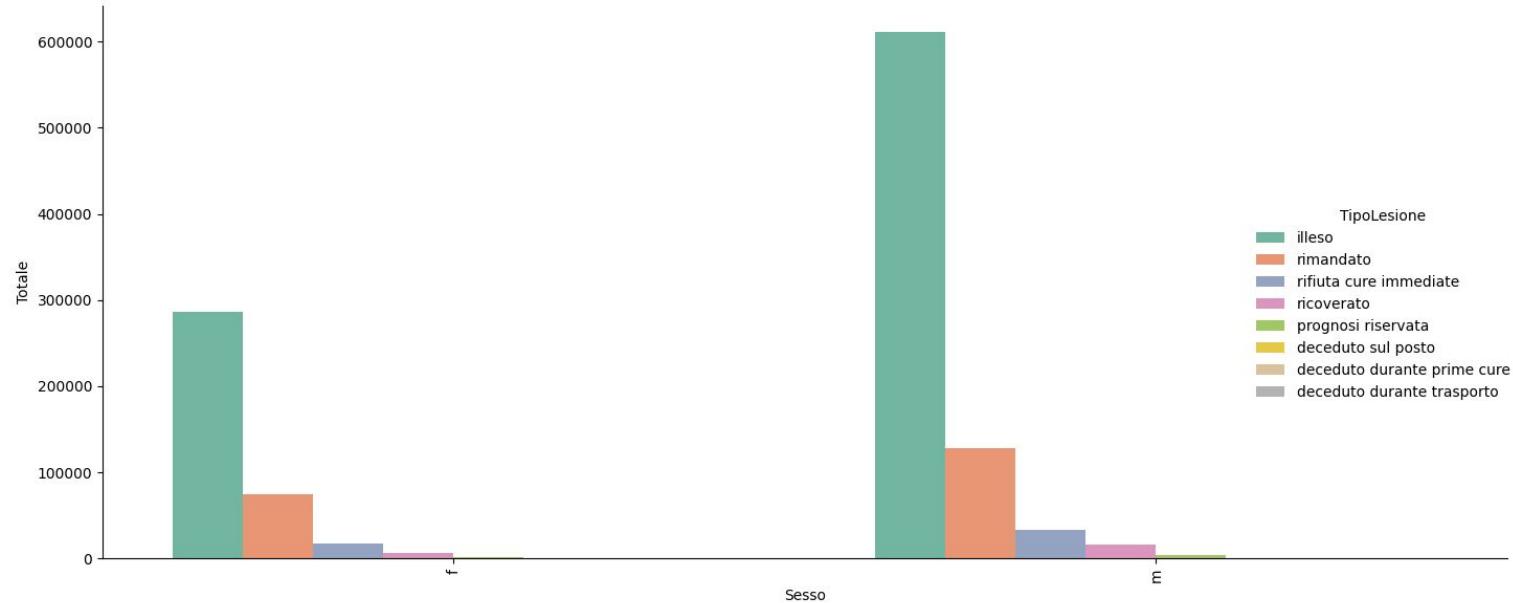
incidenti per tipo veicolo e sesso



Si osserva una significativa differenza di genere nella distribuzione degli incidenti per tipo di veicolo. Le autovetture private sono le più coinvolte, con una prevalenza di conducenti maschi. Anche i motocicli e le biciclette mostrano una maggioranza maschile. Le autovetture a noleggio e i veicoli di micromobilità elettrica sono utilizzati in misura pressoché uguale tra i due sessi

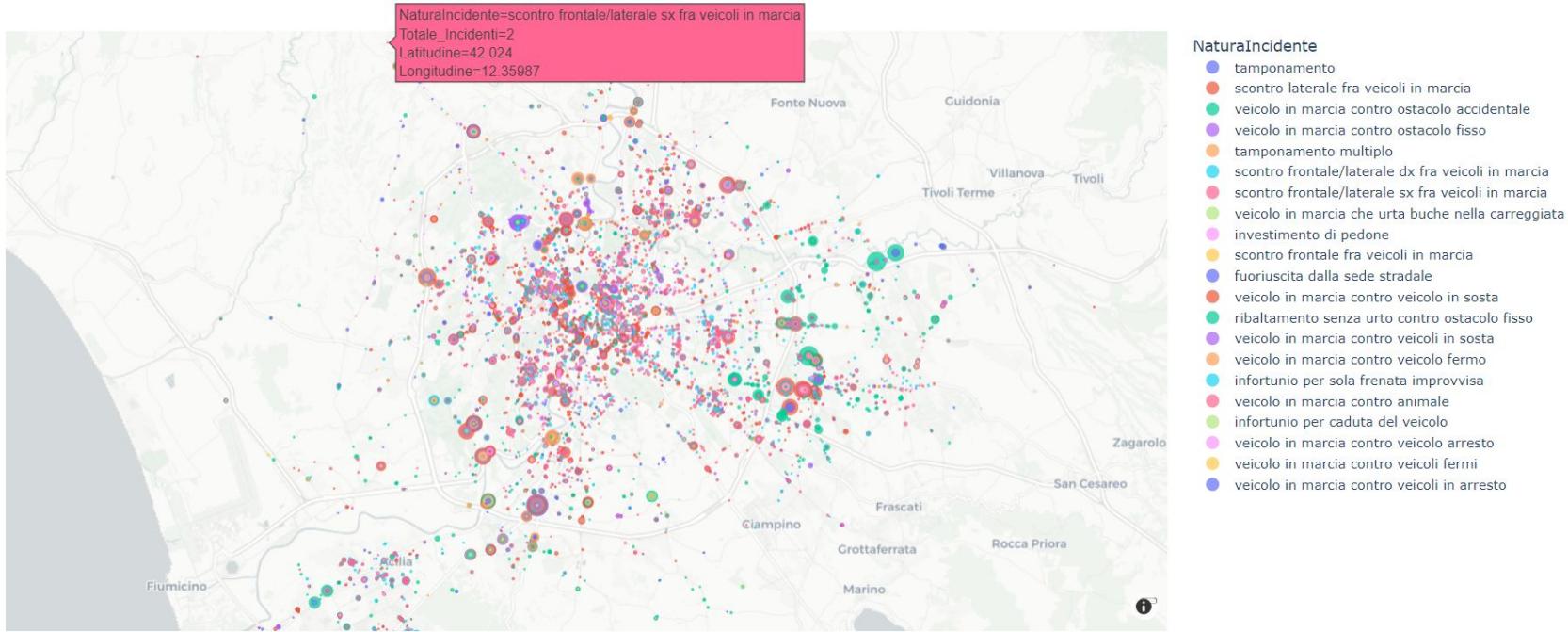
Risultati

esiti degli incidenti per sesso



Si osserva che più uomini risultano illesi rispetto alle donne, forse perché guidano di più. Entrambi i sessi hanno comunque una quota significativa di ricoveri e prognosi riservate. Anche se la mortalità è bassa, è presente. È importante analizzare questi dati per migliorare la sicurezza stradale per tutti.

Distribuzione degli incidenti stradali in base alla natura degli incidenti



Risultati

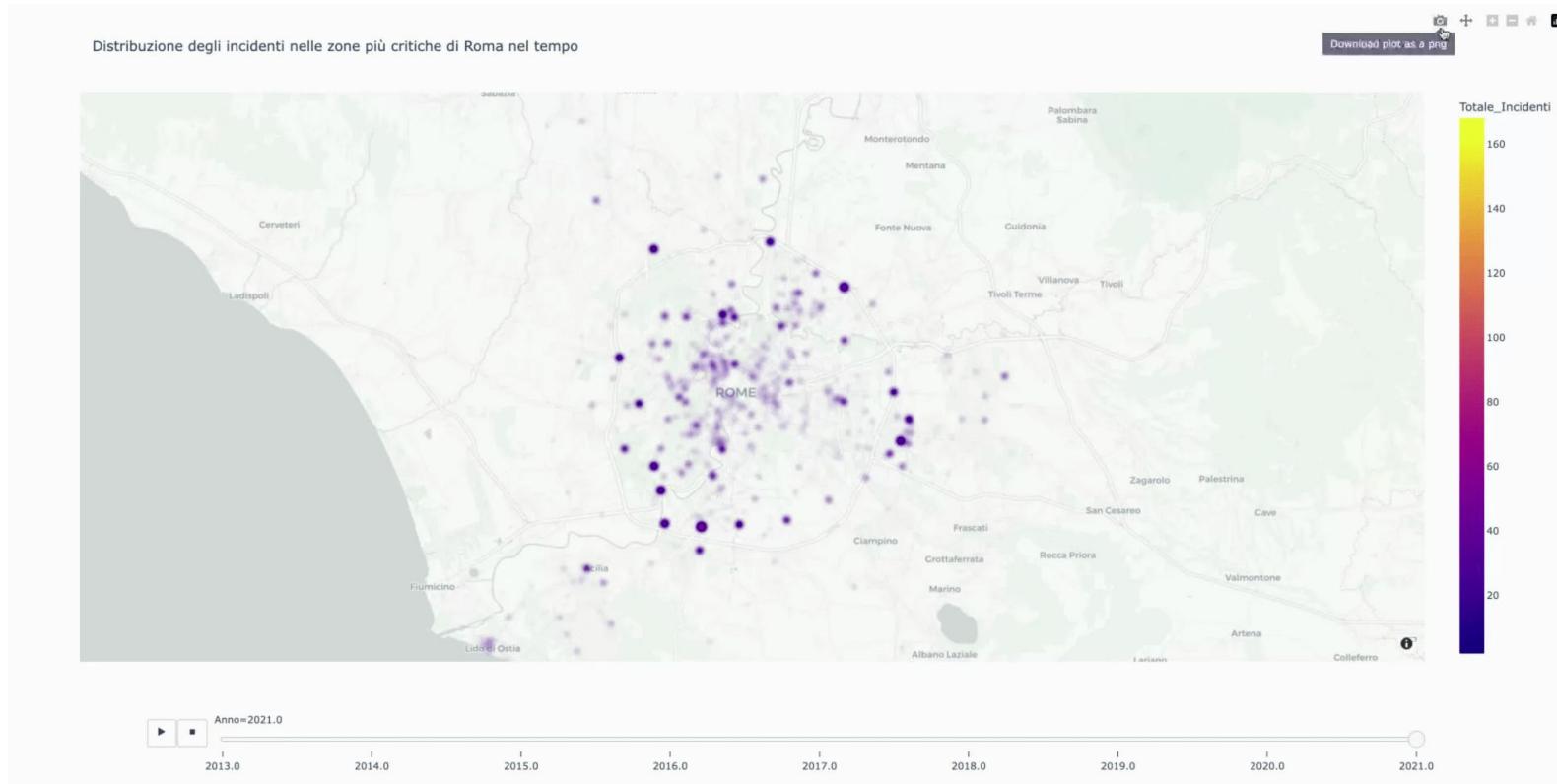
mappa interattiva

incidenti vs tipo incidente

Mostra la distribuzione degli incidenti stradali, utilizzando solo dati con coordinate valide. Evidenzia come alcuni tipi di incidenti siano concentrati in specifiche aree. Ad esempio, nel quadrante est della mappa, le bolle verdi indicano incidenti con veicoli che si sono scontrati contro ostacoli. Questi dati permettono di pianificare interventi mirati come la rimozione degli ostacoli, miglioramenti nella segnaletica stradale e nell'illuminazione nelle aree più critiche.

Risultati

mappa interattiva incidenti nelle zone più critiche nel tempo



Dal 2013 al 2021, si osserva una variazione nella distribuzione degli incidenti nelle zone più critiche di Roma. Utilizzando dati con coordinate valide, l'analisi evidenzia aumenti in alcune aree e riduzioni in altre. È essenziale approfondire queste variazioni per migliorare la sicurezza stradale e pianificare interventi mirati.

Risultati

analisi su grafi - metriche

È stata effettuata un'analisi delle relazioni degli incidenti stradali tramite grafi e calcolo delle seguenti metriche con GraphFrame:

- **PageRank**: determina l'importanza relativa di ciascun nodo nel grafo
- **Degree (Totale, In-Degree, Out-Degree)**: Misura il numero di collegamenti entranti e uscenti di ciascun nodo
- **Conteggio dei Triangoli**: Identifica sottostrutture di tre nodi completamente connesse
- **Componenti Fortemente Connessi**: Identifica gruppi di nodi interconnessi dove ogni nodo è raggiungibile da ogni altro nodo del gruppo.
- **Propagazione delle Etichette**: Algoritmo di clustering che raggruppa nodi in comunità basate sulle connessioni
- **Motif Finding**: Identifica schemi ricorrenti nei grafi

Nota: le analisi sui grafi, tranne il grafo degli incroci stradali, sono state effettuate su un campione del 30% di nodi e archi, per garantire la gestibilità dei dati, dato che non si hanno a disposizione calcolatori con risorse adeguate per elaborare grafi molto densi.

Risultati

analisi su grafi - grafo degli incroci stradali

- Le strade con il più alto PageRank, come Via Cristoforo Colombo e Via Casilina, sono i nodi più importanti nel grafo degli incidenti agli incroci.
- Le strade con il più alto Degree Totale sono le più frequentemente coinvolte negli incidenti.
- I nodi con alto conteggio dei triangoli indicano cluster di strade che sono frequentemente interconnesse tramite incidenti.

id	pagerank	id degree	id triang	a	e	b
via casilina 20.702990291987277	20.0333902209953	via casilina 252	via cristoforo co... 92	{via salvino sern... {via ostiense, st...		
via cristoforo co.... 19.89156632828565		via prenestina 227	via prenestina 83	{via isole del ca... {via isole del ca... {via delle baleni...		
via prenestina 19.89156632828565		via tiburtina 197	via casilina 75	{largo giulio one... {largo giulio one... {via dei campi sp...		
via trionfale 19.057552404578797		via nomentana 185	via ostiense 69	{via gallarate, s... {via gallarate, v... {via casilina, st...		
via salaria 18.174123286478057		via tuscolana 185	via tuscolana 67	{via bernardino r... {via bernardino r... {via federico di ...		
via tuscolana 16.56914124714795		id inDegree	via nomentana 55	{via di villa ada... {via di villa ada... {via lariana, str...		
via tiburtina 15.427275703898587		via casilina 91	via appia nuova 52			
via di torrevecchia 14.172933586467426		via prenestina 78	via laurentina 45			
via nomentana 13.336013661731446		via cristoforo co... 77	via flaminia 44			
via flaminia 13.270044240715238		via tuscolana 66				
via appia nuova 12.32767584303554		via tiburtina 66				
via portuense 11.532271092161562		via nomentana 59				
via cassia 10.786210187146613		via trionfale 53				
viale palmiro tog... 10.58409651800013		via cassia 52				
via ostiense 10.48442409596368		id outDegree				
via laurentina 10.359526939223606		via casilina 161				
via di boccea 10.04596568314704		via prenestina 149				
via aurelia 9.137987559457585		via tiburtina 131				
viale di trastevere 9.026057983825805		via nomentana 126				
viale dei romagnoli 8.825650322789064		via tuscolana 119				
		via cassia 118				
		via portuense 93				

Risultati

analisi su grafi – grafo delle persone coinvolte negli incidenti

- I conducenti e i passeggeri sono i nodi più importanti nel grafo delle persone coinvolte, come indicato dal loro alto PageRank.
- La maggior parte degli incidenti coinvolge conducenti, seguiti dai passeggeri e dai pedoni, come indicato dal loro alto Degree Totale.
- La propagazione delle etichette mostra come le persone coinvolte si raggruppino in diverse comunità.

id	pagerank	id degree	id outDegree	a	e	b
conducente 179796.82881104108		conducente 501653	1683250 3	{1680886, accident} {1680886, passeg... {passeggero, person}		
passeggero 43973.31245483054		passeggero 192993	1724948 3	{1681025, accident} {1681025, passeg... {passeggero, person}		
pedone 7465.196339997912		pedone 32919	1682835 3	{1682619, accident} {1682619, passeg... {passeggero, person}		
passeggero istrut... 45.950744006423484		passeggero istrut... 224	1711318 3	{1684407, accident} {1684407, passeg... {passeggero, person}		
pedone sconosciuto 28.338018875052484		pedone sconosciuto 131	1683783 3	{1690827, accident} {1690827, passeg... {passeggero, person}		
pedone non identi... 1.3828917174759352		1684000 3	1725338 3	{1713748, accident} {1713748, passeg... {passeggero, person}		
passeggero non id... 0.8468522569559475		1711682 3		{1673783, accident} {1673783, passeg... {passeggero, person}		
1665119 0.5405439938016686		1820691 3		{1674214, accident} {1674214, passeg... {passeggero, person}		
1665122 0.5405439938016686		1675452 3		{1674549, accident} {1674549, passeg... {passeggero, person}		
1665133 0.5405439938016686		1672430 3				
1665109 0.5405439938016686		1683250 3				
1665136 0.5405439938016686		1685766 3				
1665256 0.5405439938016686		1681297 3				
1665166 0.5405439938016686		1707305 3				
1665110 0.5405439938016686		1724948 3				
1665245 0.5405439938016686		1668835 3				
1665139 0.5405439938016686		+-----+-----+				
1665257 0.5405439938016686		conducente 501653				
1665137 0.5405439938016686		passeggero 192993				
1665295 0.5405439938016686		pedone 32919				
		passeggero istrut... 224				
		pedone sconosciuto 131				
		pedone non identi... 3				
		passeggero non id... 2				
		+-----+-----+				
		1680842 1012				
		1680854 25769804797				
		1680910 25769804800				
		1680934 8589941528				
		1680939 25769804806				
		1680976 1027				

Risultati

analisi su grafi - grafo dei tipi di veicoli coinvolti

- Le autovetture private e i motocicli sono i nodi più importanti nel grafo dei veicoli coinvolti, come indicato dal loro alto PageRank.
 - Le autovetture private sono di gran lunga il tipo di veicolo più frequentemente coinvolto, come indicato dal loro alto Degree Totale.
 - La propagazione delle etichette mostra come i diversi tipi di veicoli si raggruppano in varie comunità di

autovettura privata	160762.6103368363	a	{1681224, incidente} {1681224, autovet... {autovettura di p...
motociclo a solo	33848.7679511622	e	{1681688, incidente} {1681688, autobus... {autobus urbano, ...
autocarro inferio...	11572.30332216674	b	{1683620, incidente} {1683620, autocar... {autocarro inferi...
ciclomotore	4819.960313129114		{1684987, incidente} {1684987, autocar... {autocarro inferi...
motociclo con pas...	4735.362469754453		{1687850, incidente} {1687850, autocar... {autocarro inferi...
autovettura pubblica	3034.4074610490748		{1675809, incidente} {1675809, autocar... {autocarro superi...
autobus di linea	1727.992227003443		{1675909, incidente} {1675909, autobus... {autobus urbano, ...
autovettura di po...	1726.2462047087581		{1722395, incidente} {1722395, autocar... {autocarro superi...
autocarro superio...	1664.5457589705754		{1722907, incidente} {1722907, autocar... {autocarro inferi...
autobus urbano	1414.2978793836626		
velocipedo	1328.091857579713		
quadriciclo leggero	1100.8026395694899		
veicolo speciale	507.95914859030984		
autobus turistico	466.6825689044868		
autoarticolato	453.89372139515086		
macchina operatrice	316.2795958006678		
altro	263.40115744619123		
autobus in extra...	241.57587876259083		
vettura tranviaria	220.28665955542903		
ciclomotore con p...	180.08219882248162		

	id	component
	1680782	8589935636
	1680881	25769804833
	1680906	25769804835
	1680920	17179870188
	1680986	8589935652
	1681126	25769804848
	1681352	25769804865

	id	label
	1680782	8589935636
	1680881	25769804833
	1680906	25769804835
	1680920	17179870188
	1680986	8589935652
	1681126	25769804848
	1681352	25769804865

Risultati

analisi su grafi -

grafo delle strade coinvolte negli incidenti

- Via Casilina e Via Cristoforo Colombo sono le strade con il più alto PageRank, che indicano la loro rilevanza nel grafo degli incidenti stradali.
- Queste stesse strade sono anche le più frequentemente coinvolte negli incidenti, come indicato dal loro alto Degree Totale.

	id	pagerank
via casilina	4626.551944541163	
via cristoforo co...	4625.624238487799	
via prenestina	3982.723943506452	
via tiburtina	3864.441421702525	
via tuscolana	3301.787700337176	
via nomentana	2763.7181893862817	
viale palmiro tog...	2577.713125686914	
via cassia	2425.105479908629	
via di boccea	2362.485321306597	
via appia nuova	2272.033981103663	
via aurelia	2189.468142354317	
via ostiense	2174.1609924738204	
via salaria	2143.0828396861452	
via portuense	2011.3485801085378	
via laurentina	1858.2770813035713	
via trionfale	1659.2841328571149	
via flaminia	1611.9711241355797	
via della magliana	1396.743319755263	
via ardeatina	1375.4060805279041	
via collatina	1272.8945616312446	

	a	e	b
{1681351, incidente}	{1681351, via cas...}{via cassia nuova...}		
{1681592, incidente}	{1681592, via del...}{via della maglia...}		
{1681748, incidente}	{1681748, via del...}{via della giusti...}		
{1681793, incidente}	{1681793, via ant...}{via antonio paci...}		
{1682117, incidente}	{1682117, via ema...}{via emanuele fil...}		
{1682338, incidente}	{1682338, via del...}{via della maglia...}		
{1682528, incidente}	{1682528, via cav...}{via cavour, strada}		
{1682549, incidente}	{1682549, via pin...}{via pinciana, st...}		
{1682835, incidente}	{1682835, via art...}{via arturo graf,...}		

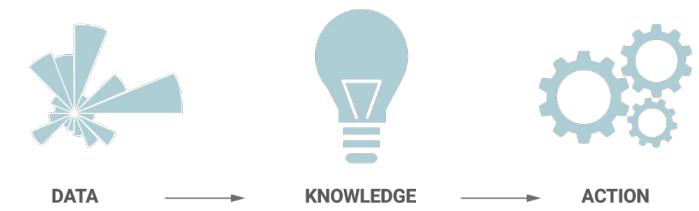
	id outDegree		id component
5250596	2		1680194 8589935591
5250595	2	1680194 8589935591	1680920 17179870246
1680192	1	1680920 17179870246	1681009 25769804816
1680096	1	1681009 25769804816	1681115 17179870260
1680246	1	1681115 17179870260	1681195 8589935653
1680163	1	1681195 17179869247	1681292 25769804837
1680285	1	1681292 25769804837	1681737 8589935698
1680238	1	1681737 8589935698	1681843 17179870315
1680403	1	1681843 17179870315	1681934 1076
1680251	1	1681934 1076	1681989 17179870326
1680471	1	1681989 17179870326	1682080 17179870333
1680287	1	1682080 17179870333	

Conclusioni

Il lavoro di analisi dei dati presentato in questo progetto ha consentito di evidenziare tendenze e aspetti significativi che emergono nell'ambito dei sinistri stradali avvenuti nel periodo temporale dal 2006 al 2021 nel territorio di Roma capitale.

Nel dettaglio sono state individuate le strade più pericolose, ovvero quelle in cui si è registrato un tasso di incidenti più alto e con maggiori vittime, sono state mappate geograficamente le aree a più alto rischio di incidenti, è stato rappresentato l'andamento degli incidenti nelle varie fasce orarie giornaliere ecc...

Tutte queste informazioni possono essere preziose per il decisore politico per intraprendere azioni utili a migliorare la sicurezza stradale, ad esempio, riducendo il traffico nelle ore di punta (nelle quali si registrano più incidenti), intervenendo sulle strade più pericolose e così via



Possibili sviluppi futuri

Questo lavoro di analisi rappresenta solo un primo passo per la comprensione delle dinamiche che sono alla base degli incidenti stradali. Si tratta di un'analisi sostanzialmente descrittiva di un fenomeno che permette di acquisire conoscenza. Ulteriori sviluppi di questo lavoro potrebbero venire dall'utilizzo di altri dataset, laddove disponibili, che fornissero informazioni circa i lavori sulle infrastrutture stradali. Inoltre, l'uso di tecniche di machine learning potrebbe consentire di sviluppare modelli predittivi per identificare aree e condizioni di rischio. In tal modo si potrebbero pianificare interventi mirati e adottare misure di sicurezza preventive più efficaci.

