



UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE E TECNOLOGIE
Corso di Laurea Triennale in Informatica

Studio sull'incidentalitá stradale tramite dataset aperti

Tesi di Laurea di:
Gabriele Padovani
Matr. 909165

Relatore:
Andrea Trentini

Correlatore:
CORREL

Anno Accademico 2020/2021

Indice

1	Introduzione	4
1.1	Origine dei dati	4
1.1.1	Dati riguardanti incidenti	4
1.1.2	Dati riguardanti autovelox	4
1.1.3	Dati riguardanti il meteo	5
1.1.4	Dati riguardanti trasporti pubblici	5
1.1.5	Dati riguardanti piste ciclabili	5
1.1.6	Dati riguardanti autostrade e manto stradale	5
1.2	Scopo del lavoro	5
1.3	Dati mancanti	5
1.3.1	Dati riguardanti pavè	5
1.3.2	Dati riguardanti traffico stradale	5
2	Dati Geolocalizzati	6
2.1	Incidenti	6
2.2	Incidenti e Linee dei Trasporti Pubblici	6
2.2.1	Il Pavè influisce sull'incidentalità?	6
2.3	Incidenti e Piste Ciclabili	11
2.4	Incidenti e Autovelox	11
2.5	Incidenti e Meteo	11
3	Dati su Incidenti	14
3.1	Dati Istat su veicoli	14
3.1.1	Come cambia il tipo di veicolo al cambiare del tipo di strada?	14
3.2	Dati Istat su conducente	14
3.2.1	Come cambia il sesso del conducente al cambiare della strada?	14
3.2.2	Come cambia l'età del conducente al cambiare della strada?	14
3.2.3	Il numero di passeggeri influisce sull'incidentalità?	17
3.2.4	Il conducente se da solo si distrae con il telefono cellulare?	17
3.3	Dati Istat su orari e mesi	17
3.3.1	Quanto influiscono le ore di punta sull'incidentalità?	17
3.3.2	è possibile accentuare le ore di punta mattutine?	18
3.3.3	Si ha la stessa tendenza di notte?	18
3.3.4	Quanto influiscono le vacanze estive sull'incidentalità?	18
3.3.5	è possibile individuare la tendenza inversa in località di mare?	20

3.4	Dati Istat su tipi di incidenti e incroci	24
3.4.1	Quali incidenti avvengono con più frequenza?	24
3.4.2	Quali tipi di incroci provocano più incidenti?	24
3.4.3	Esistono tipi di incidenti che provocano più feriti?	24
3.4.4	Esistono incroci che favoriscono incidenti con pedoni?	24
3.4.5	Ci sono aspetti salienti sui pedoni coinvolti?	28
3.5	Dati ACI	28
3.5.1	Esiste correlazione tra incidenti e feriti?	28
3.5.2	Quali sono le autostrade con più incidenti?	29
3.5.3	Autostrade pericolose a Milano?	29
3.5.4	In quali mesi avvengono più incidenti?	29
3.5.5	In Quali orari avvengono incidenti sulle autostrade?	29
3.5.6	Quali autostrade sono utilizzate di più per viaggiare in Agosto?	33
3.5.7	Le autostrade più utilizzate cambiano a seconda dell'anno?	33

4 Dati su Meteo **37**

Todo list

Metto i link per citare fonti online?	4
Package per visualizzare codice?	4
Codice utilizzato per autovelox_2014.csv	4
non so ancora se usarli	5
devo guardare il dataset del manto stradale che ho trovato	5
cercare dati su manto stradale	5
devo ancora cercare se c'è un dataset sul traffico	5
reference a immagine fig:week_weekend	18
Aggiungo tabella per trimestre?	20
Affianco i tre grafici dei trimestri?	21
Trovare un modo per dividere pedoni_feriti e tipo_incrocio	24
orari di altri luoghi, ho fatto solo milano	29

Capitolo 1

Introduzione

1.1 Origine dei dati

1.1.1 Dati riguardanti incidenti

La maggior parte dei dati provengono dall'Istituto nazionale di statistica, in particolare dall'archivio di dati non geolocalizzati su incidenti in Italia. Questo dataset contiene campi di vari tipo, come ora, mese e giorno della settimana in cui è avvenuto l'incidente, ma anche dati sui passeggeri, la natura dell'incidente e il tipo di strada.

Per quanto riguarda i dati geolocalizzati, la fonte è il sito web TheSubmarine, che in un articolo riguardante l'incidentalità a Milano, ha ottenuto da Istat la pubblicazione di una sezione degli incidenti avvenuti nella città nel 2016, con la relativa posizione.

Il dataset ACI è stato consigliato da parte di Istat, in quanto questi dati, specifici a autostrade e strade statali, indicano anche il nome della rispettiva via.

Metto i link per citare fonti online?

1.1.2 Dati riguardanti autovelox

Un primo dataset delle posizioni degli autovelox fissi individuato era una mappa GoogleMyMaps creata da un utente anonimo, dunque senza un modo di provare l'autenticità dei dati.

Il dataset utilizzato invece, è stato ottenuto tramite OpenStreetMaps, ricavando solo gli autovelox in Milano. Per restringere il campo di ricerca a solo Milano, si è fatto uso delle Overpass API, specifiche per OpenStreetMaps, eseguendo la seguente query:

```
[out:json];
node(bbox)["highway"="speed_camera"];
out meta;
```

Package per visualizzare codice?

Il dataset ricavato non permette di capire quando gli autovelox siano stati installati. Tuttavia il sito web ztlmilano contiene un articolo nel quale specifica una lista di autovelox installati nel 2014, considerando che il dataset di OpenStreetMaps è aggiornato, è possibile ricavare la posizione precisa di questi ultimi.

Codice utilizzato per autovelox_2014.csv

1.1.3 Dati riguardanti il meteo

non so ancora se usarli

1.1.4 Dati riguardanti trasporti pubblici

I dati riguardanti i trasporti pubblici trovati hanno due provenienze, i primi sono dati riferiti alle linee ATM, scaricati sul sito del comune di milano. Dallo stesso sito proviene anche il dataset riguardante gli autobus turistici, che contiene in particolare le aree di sosta di questi ultimi.

1.1.5 Dati riguardanti piste ciclabili

Oltre al sito principale del comune di Milano, esiste un secondo sito web dedicato unicamente ai dati geolocalizzati. Su questo sito è disponibile un dataset contenente i percorsi delle piste ciclabili di Milano.

1.1.6 Dati riguardanti autostrade e manto stradale

devo guardare il dataset del manto stradale che ho trovato

1.2 Scopo del lavoro

1.3 Dati mancanti

1.3.1 Dati riguardanti pavè

cercare dati su manto stradale

1.3.2 Dati riguardanti traffico stradale

devo ancora cercare se c'è un dataset sul traffico

Capitolo 2

Dati Geolocalizzati

2.1 Incidenti

Controllando i dati trovati, in particolare come questi siano distribuiti, si nota subito che gli incidenti a Milano sono per buona parte uniformemente sparsi in tutta la città, con più alta concentrazione in alcuni punti di interesse, come Piazzale Loreto, Zona Navigli e Monumentale, e Corso Venticinque Marzo.

2.2 Incidenti e Linee dei Trasporti Pubblici

Il dataset dei tragitti dei trasporti pubblici copre molta più superficie rispetto a quello degli incidenti. Dopo aver eliminato alcune linee di autobus che risultavano troppo in periferia, si nota comunque che i trasporti pubblici coprono la maggior parte di Milano.

Se a questi ultimi vengono sovrapposti i dati sugli incidenti, si può notare che la maggior parte dei luoghi con alta concentrazione di incidenti sono attraversati da linee di autobus. Nel caso di Corso Venticinque Marzo, si ha anche una linea di tram.

Dalla sovrapposizione delle mappe, si può notare anche che, alcune strade con alta incidentalità sono parallele a linee di autobus. Un esempio è quello di zona Navigli, dove le vie interessate sono: Viale Gian Galeazzo e Viale Beatrice D'Este, parallele a Viale Col di Lana e Viale Bligny. La stessa cosa si può notare su Viale Gabriele D'Annunzio e Viale Gorizia e Coni Zugna.

Anche vicino a corso Venticinque Marzo si può notare lo stesso fenomeno, tra Viale Bianca Maria e Viale Premuda.

2.2.1 Il Pavè influisce sull'incidentalità?

Spesso le linee di tram coincidono con strade in pavè. Servirebbe una mappa delle strade in pavè a Milano..

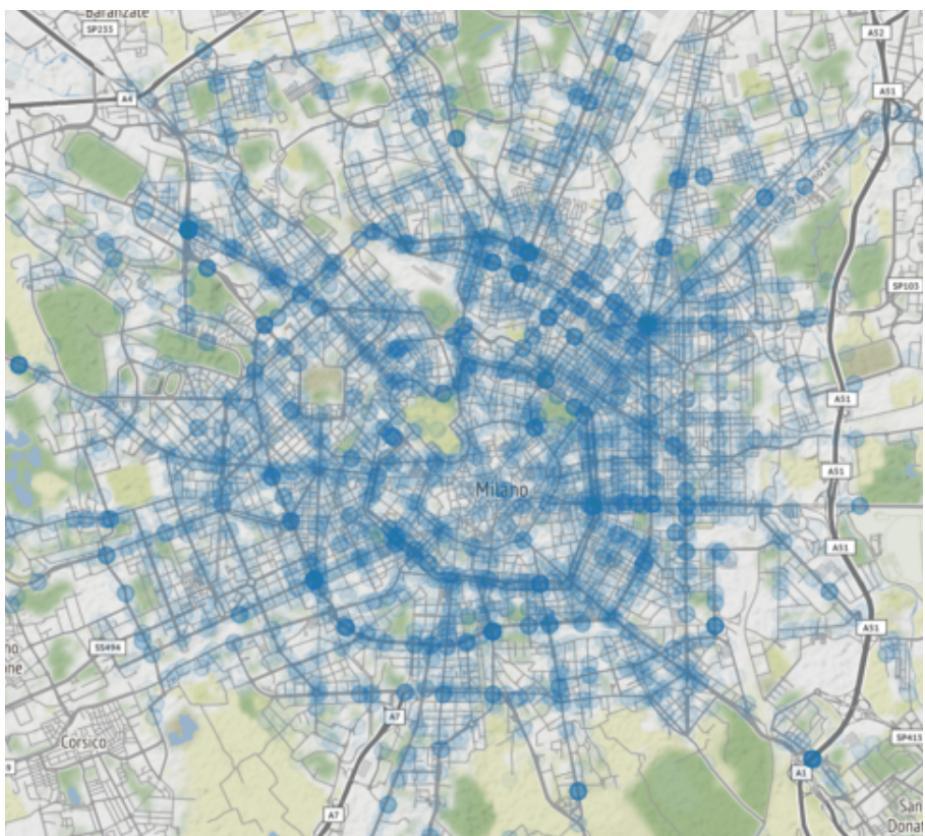


Figura 2.1: Distribuzione di incidenti a Milano

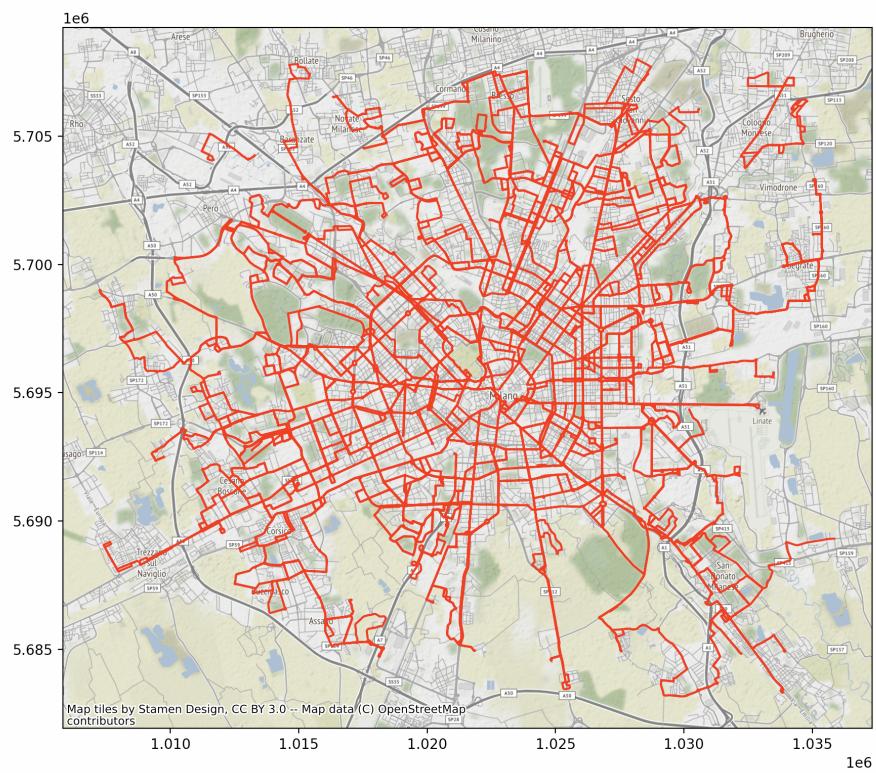


Figura 2.2: Linee Autobus e Tram a Milano



Figura 2.3: Linee Autobus e Tram a Milano

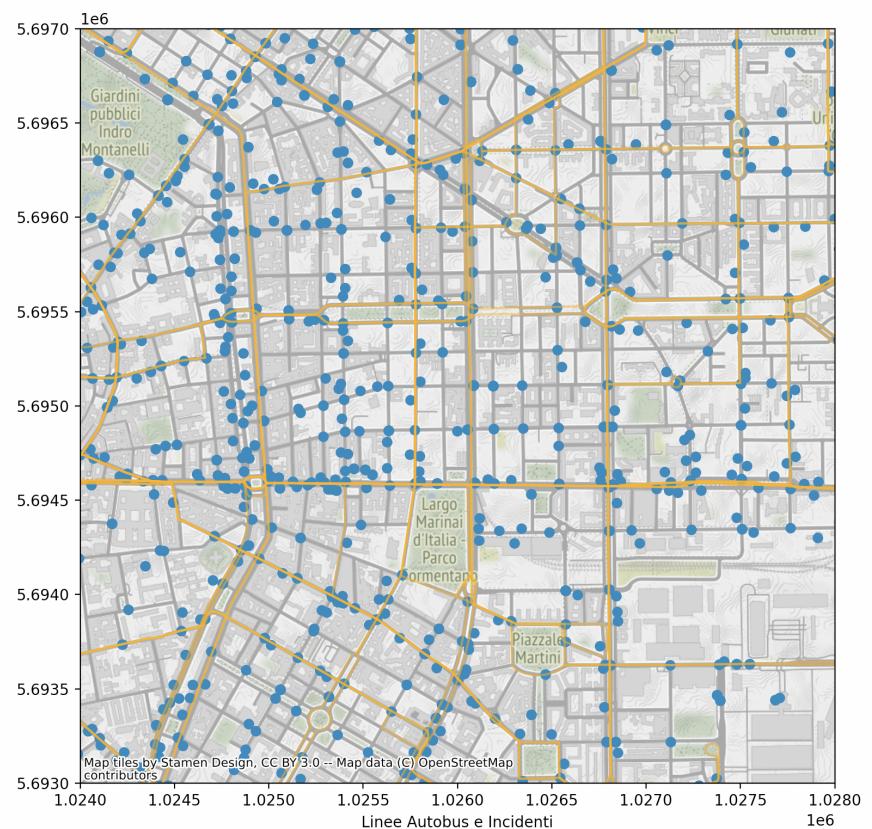


Figura 2.4: Linee Autobus e Tram a Milano

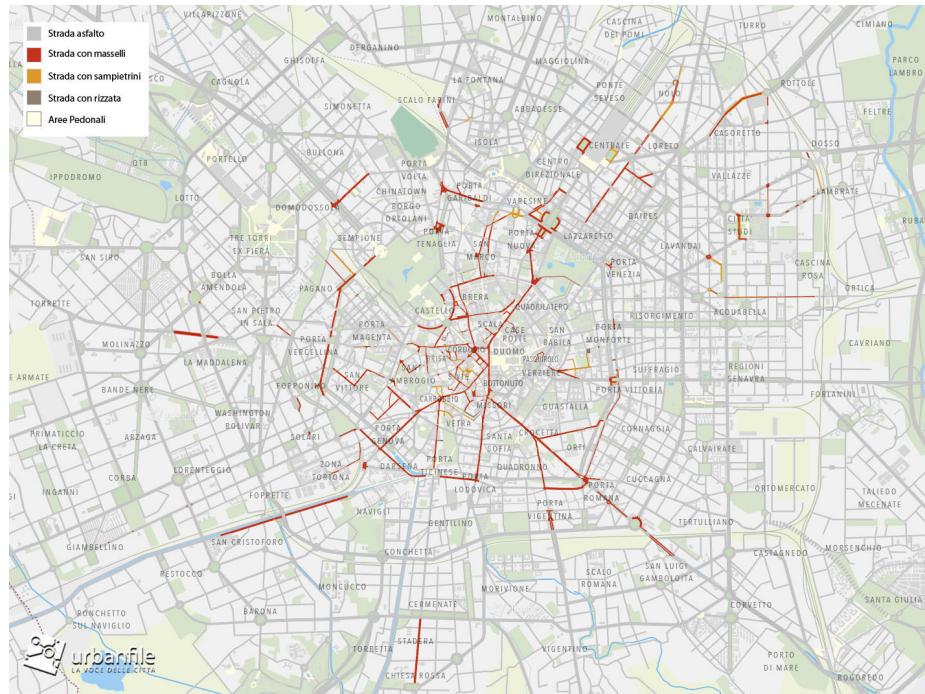


Figura 2.5: Cartina con strade in pavè a Milano

2.3 Incidenti e Piste Ciclabili

2.4 Incidenti e Autovelox

Per sapere se gli autovelox hanno influenza sull'incidentalità, bisognerebbe innanzitutto sapere quando sono stati posizionati i dispositivi, e solo a quel punto, avendo dati su incidenti prima e dopo l'installazione, sarebbe possibile trarre conclusioni.

Alcuni dati sull'installazione di autovelox esistono per l'anno 2014, tuttavia i dati riguardo agli incidenti sono solo riguardanti l'anno 2016, in quanto Istat non ha rilasciato le posizioni degli incidenti in altre annate.

Gli autovelox installati nel 2014, presenti nel dataset, sono rappresentati nella seguente mappa.

è comunque possibile sovrapporre i dataset, per vedere se gli autovelox hanno un qualche tipo di effetto sugli incidenti.

2.5 Incidenti e Meteo

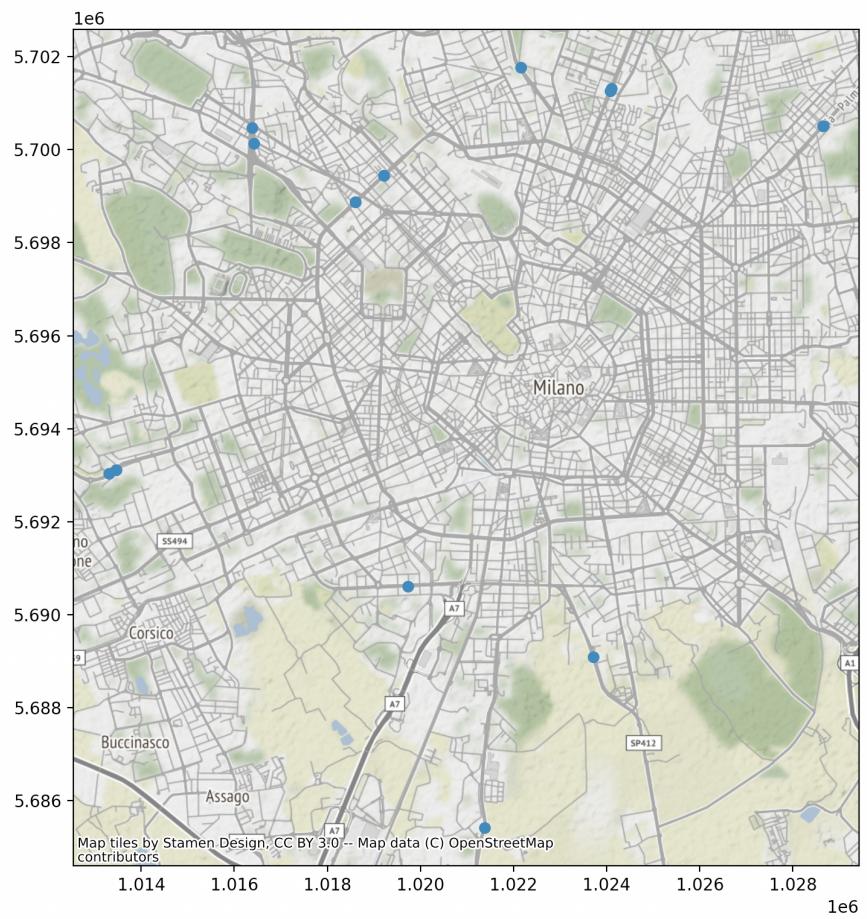


Figura 2.6: Autovelox installati nel 2014

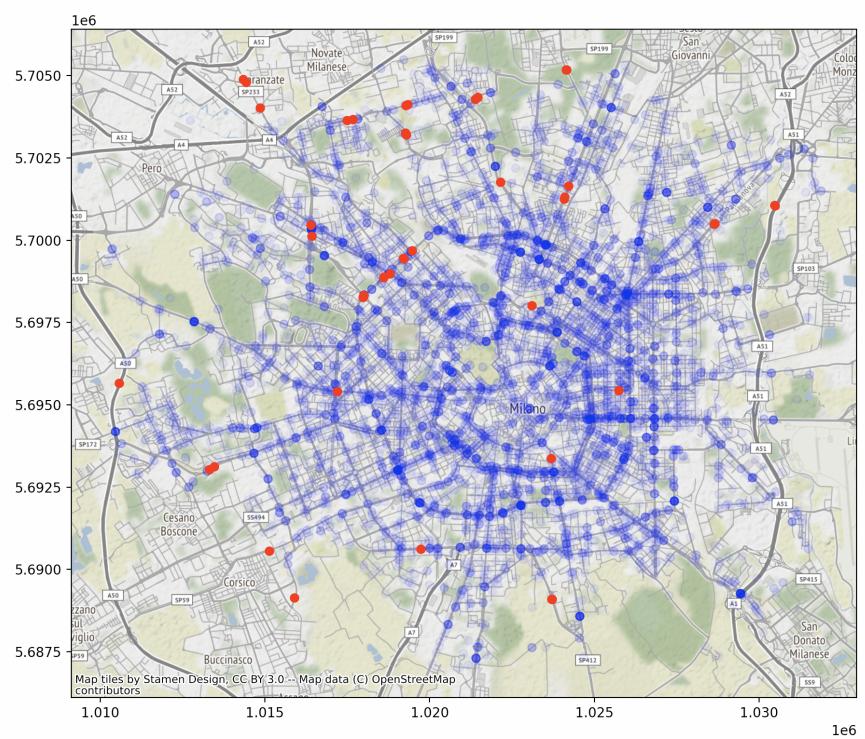


Figura 2.7: Autovelox e Incidenti a Milano

Capitolo 3

Dati su Incidenti

Per quanto riguarda dati generali su incidenti in Italia, sono disponibili due dataset molto ampi, il primo, rilasciato da Istat, contiene dati dal 2010 al 2018 che riguardano campi come data, ora, numero di persone a bordo, tipo di incrocio, tipo di veicolo, ecc.. Il secondo è invece messo a disposizione da Automobile Club D'Italia (ACI) che contiene dati simili, ma in più mette a disposizione il luogo dell'incidente, come autostrada o strada provinciale.

3.1 Dati Istat su veicoli

Il dataset Istat contiene molte informazioni riguardanti i conducenti dei veicoli coinvolti nell'incidente, oltre al tipo di veicoli.

3.1.1 Come cambia il tipo di veicolo al cambiare del tipo di strada?

è possibile notare che, nonostante le auto private siano di gran lunga il tipo di veicolo più coinvolto in incidenti, nelle autostrade non sono presenti incidenti con velocipedi, anche il numero di incidenti con motocicli è ridotto, mentre cresce molto nelle strade urbane.

3.2 Dati Istat su conducente

3.2.1 Come cambia il sesso del conducente al cambiare della strada?

Il numero di incidenti per genere è tende ad essere 75% circa uomini e 25% donne. Nelle strade urbane, la percentuale di incidenti con conducente donna aumenta leggermente nel 2010, questo vale per tutti gli anni?

3.2.2 Come cambia l'età del conducente al cambiare della strada?

Risultati non troppo interessanti...

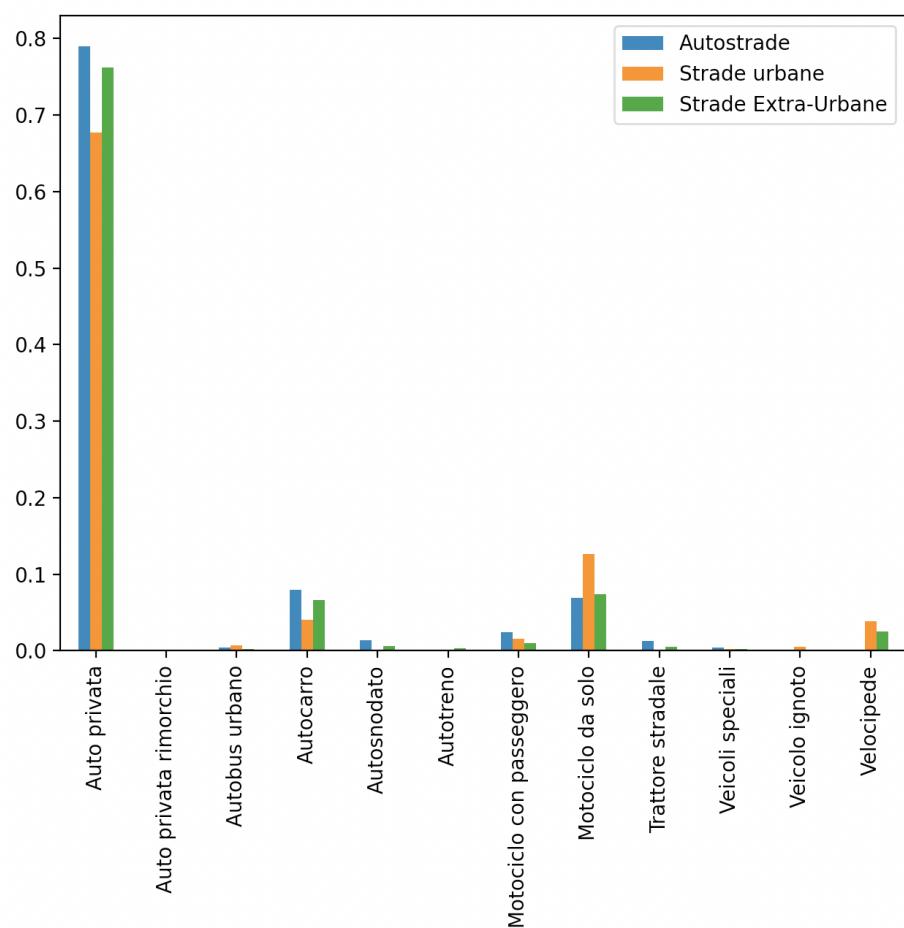


Figura 3.1: Incidenti per tipo di veicolo nel 2010

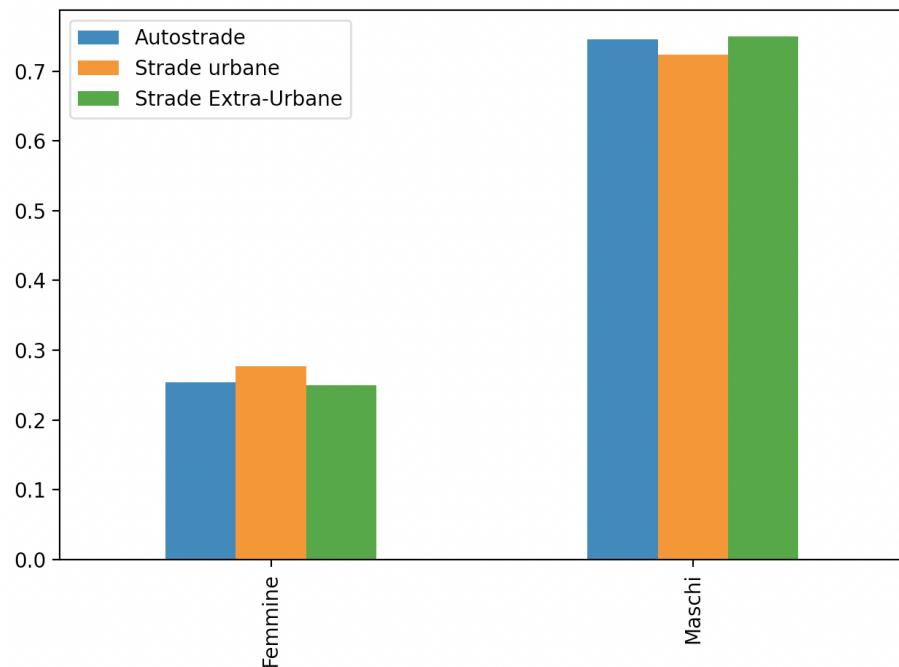


Figura 3.2: Sesso del conducente per tipo di veicolo nel 2010

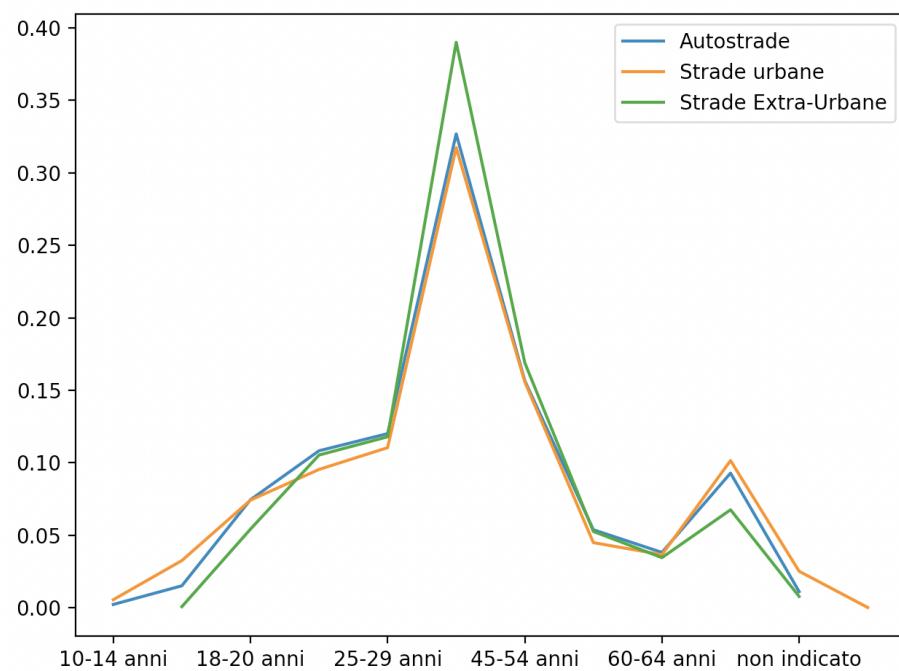


Figura 3.3: Fascia di età del conducente per tipo di veicolo nel 2010

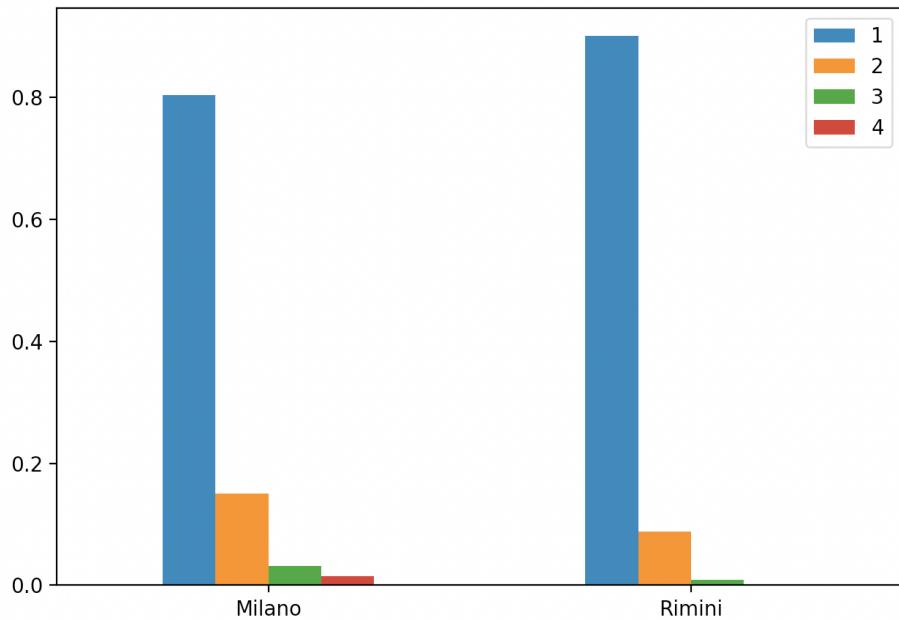


Figura 3.4: Numero di passeggeri in incidenti per Milano e Rimini

3.2.3 Il numero di passeggeri influisce sull'incidentalità?

La maggior parte degli incidenti sembra avvenire quando in macchina è presente solo il conducente. Si sono prese in considerazione le provincie di Milano e Rimini, per controllare se la località marittima influisse sul numero di incidenti, ma sembra che quest'ultima abbia una percentuale ancora più alta di incidenti in cui è presente solo il conducente, rispetto a Milano.

3.2.4 Il conducente se da solo si distrae con il telefono cellulare?

Per quanto non siano disponibili dati su questo ambito, si potrebbe confrontare gli anni tra 2010 e 2013, in cui l'uso del cellulare in macchina ancora non era frequente, rispetto agli anni più recenti.

3.3 Dati Istat su orari e mesi

3.3.1 Quanto influiscono le ore di punta sull'incidentalità?

Per prima cosa, con un semplice conto degli incidenti durante il weekend e confrontandolo con il numero di quelli avvenuti durante la settimana lavorativa, si osserva che nel weekend avvengono più incidenti durante la sera e la notte, mentre durante la settimana in giornata.

Per quanto riguarda gli orari di punta, sono state prese in considerazione due fasce orarie, la prima, mattutina dalle 7:00 alle 10:00, e la seconda pomeridiana, dalle 17:00 alle 19:00

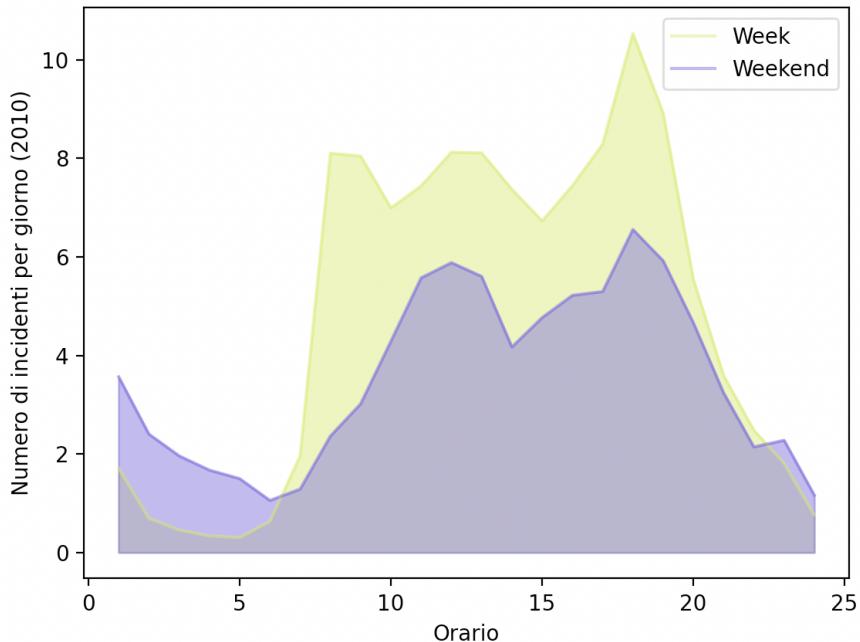


Figura 3.5: Incidenti per ora

Va sottolineato che i grafici indicano gli incidenti normalizzati per numero di giorni, quindi gli incidenti totali della settimana sono divisi per cinque giorni, mentre quelli del weekend per due. Si può osservare che nella fascia oraria delle 18:00, avvengono molti più sinistri, mentre nella fascia mattutina, il numero non sembra variare molto dalla media di incidenti durante il giorno.

3.3.2 è possibile accentuare le ore di punta mattutine?

Se si selezionano solo gli incidenti nella provincia di Milano, è possibile individuare il secondo picco di incidenti, quello durante le ore di punta mattutine

3.3.3 Si ha la stessa tendenza di notte?

Come individuato dal primo grafo in *week_weekend*, durante le ore notturne si ha la tendenza opposta.

reference a
immagine
fig:week_weekend

3.3.4 Quanto influiscono le vacanze estive sull'incidentalità?

Si nota un chiaro calo di incidenti nel mese di Agosto in provincia di Milano. È possibile controllare il decremento del numero di incidenti, anche per gli altri anni, in cui sono disponibili dati.

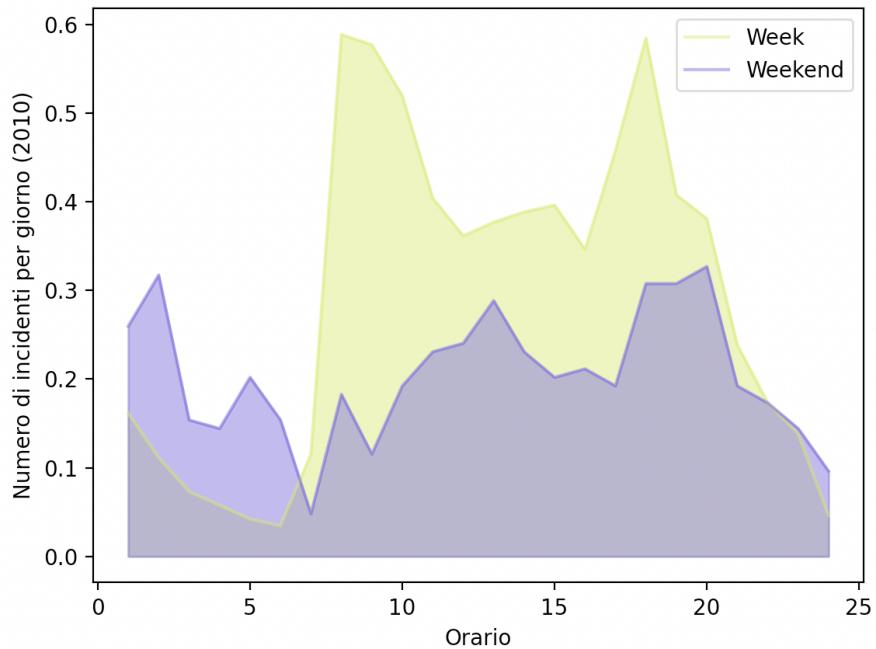


Figura 3.6: Incidenti per ora a Milano

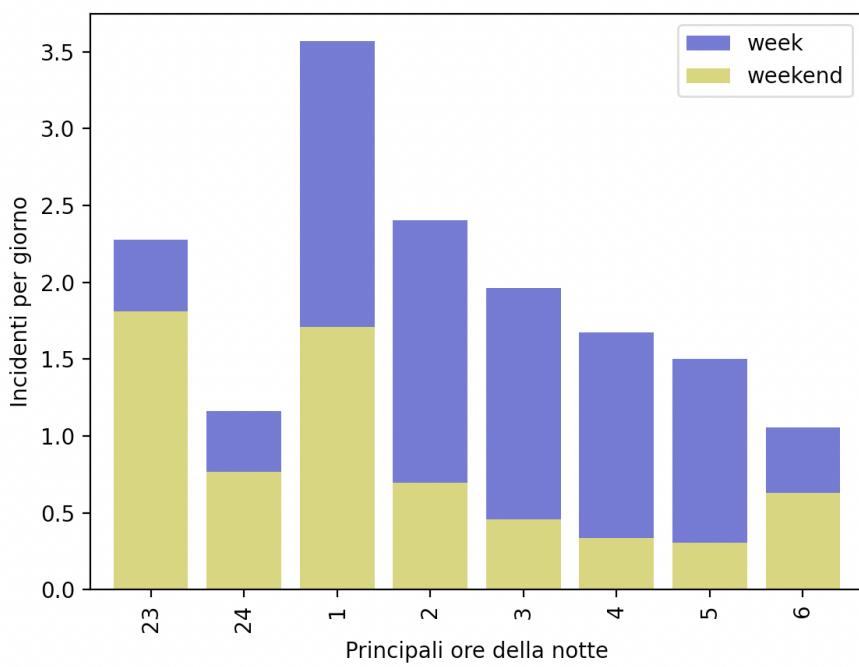


Figura 3.7: Incidenti durante ore serali o notturne

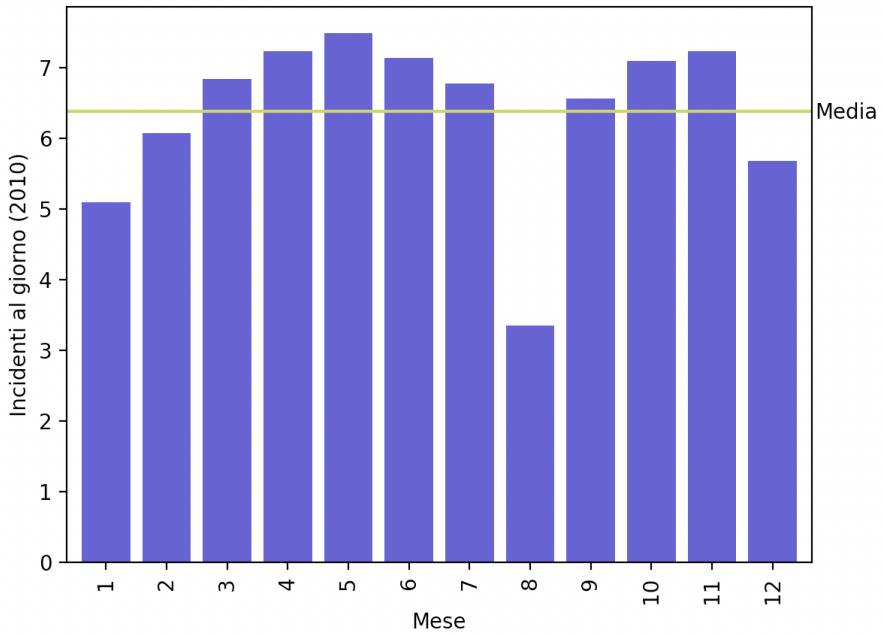


Figura 3.8: Incidenti per mese in Milano

Anno	Incremento in Agosto
2010	-47.4%
2011	-35.14%
2012	-45.46%
2013	-41.37 %

Per quanto riguarda gli anni successivi, il dataset non dispone più di un campo che indichi il mese dell'incidente, ma è disponibile soltanto il trimestre.

Per quanto possano esserci molti fattori che contribuiscono a questa tendenza, quello che influisce di più devono essere le partenze per le vacanze.

Aggiungo tabella per trimestre?

3.3.5 è possibile individuare la tendenza inversa in località di mare?

L'unica località in cui si è riscontrata la tendenza inversa è in provincia di Rimini.

Anno	Agosto	Luglio
2010	63.75%	101.72%
2011	60.49%	54.6 %
2012	45.93%	75.97%
2013	51.82%	116.37%

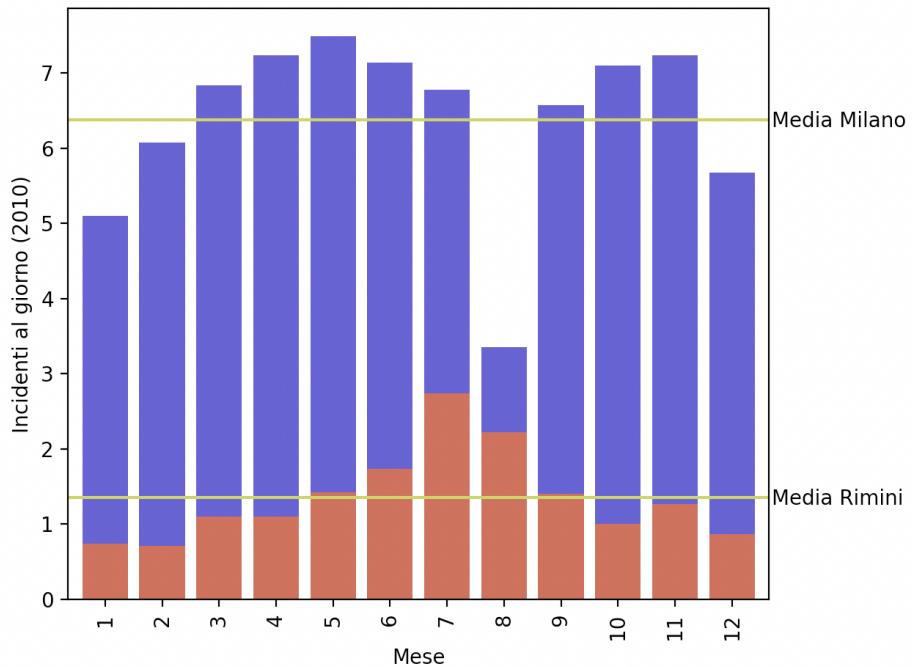


Figura 3.9: Incidenti per mese in Milano e Rimini

Tuttavia, osservando il grafo equivalente della Valle d'Aosta, è possibile notare un notevole incremento di incidenti sia in Gennaio che in Agosto, possibili conseguenze, rispettivamente, dell'inizio della stagione sciistica ed estiva.

Questa tendenza avviene ogni anno?

Anno	Agosto	Gennaio
2010	-2.93 %	78.48%
2011	124.0 %	-32.12%
2012	52.25 %	-41.44%
2013	-15.21%	-24.63%

Il comportamento degli incidenti in Valle d'Aosta è molto inconsistente, soprattutto in Agosto. Va specificato che la taglia del campione con cui si sta lavorando, per questa provincia, è molto piccolo, e sicuramente influisce sulle percentuali di incremento e decremento degli incidenti.

Nonostante ciò, se si assume che Gennaio 2010 sia un outlier, il numero di incidenti in questo mese è sempre in decremento, in modo molto consistente, rispetto alla media.

La prima cosa che è possibile notare, è che il picco di Gennaio del 2010 è in linea con la tendenza del trimestre invernale. Tuttavia, si osserva anche che dall'anno 2015 c'è un ampio gap nel numero di incidenti. È un cambio di metro di misurazione?

Affianco i tre grafici dei trimestri?

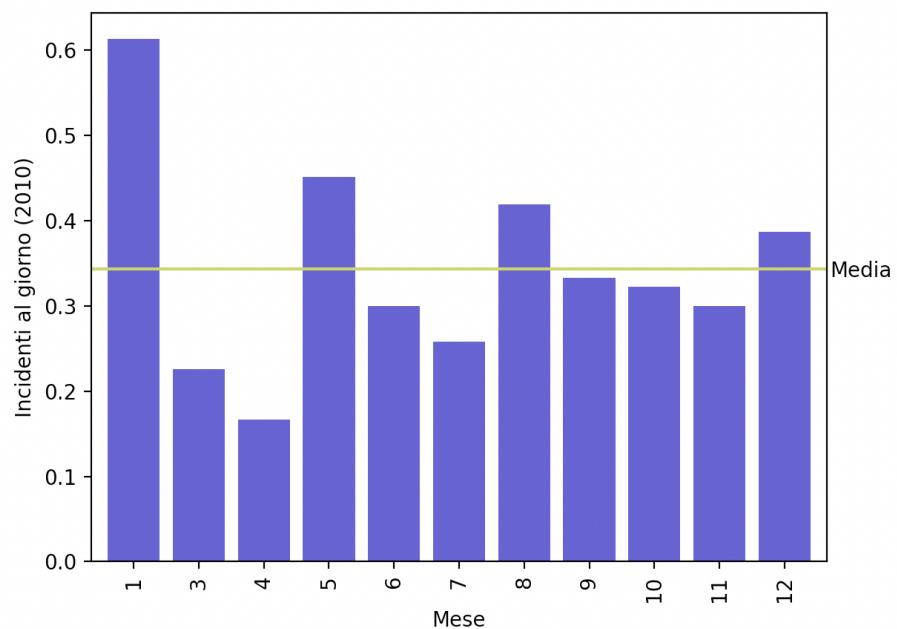


Figura 3.10: Incidenti per mese Valle d'Aosta

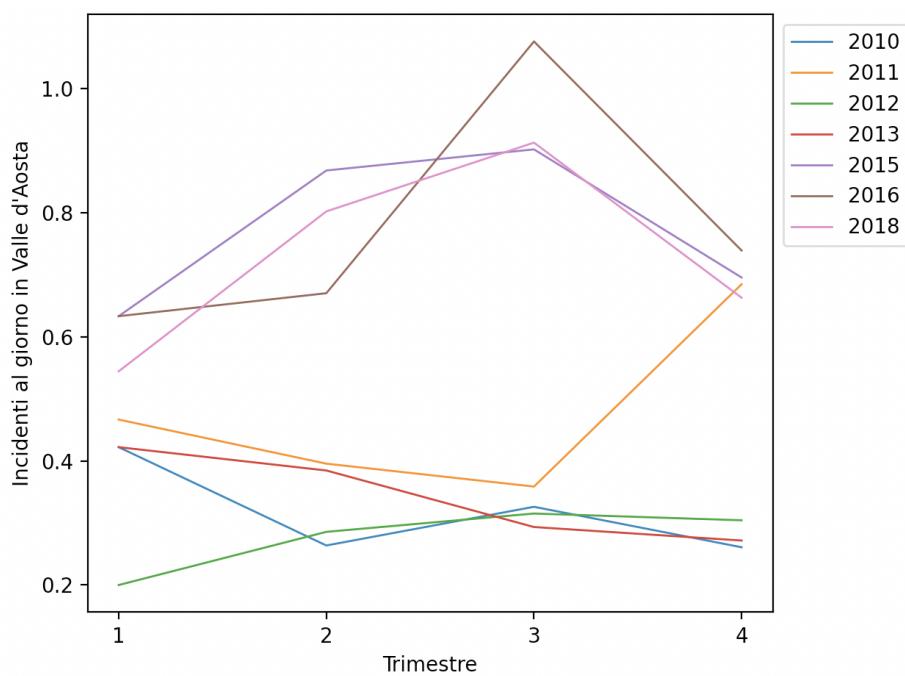


Figura 3.11: Incidenti per trimestre in Valle d'Aosta

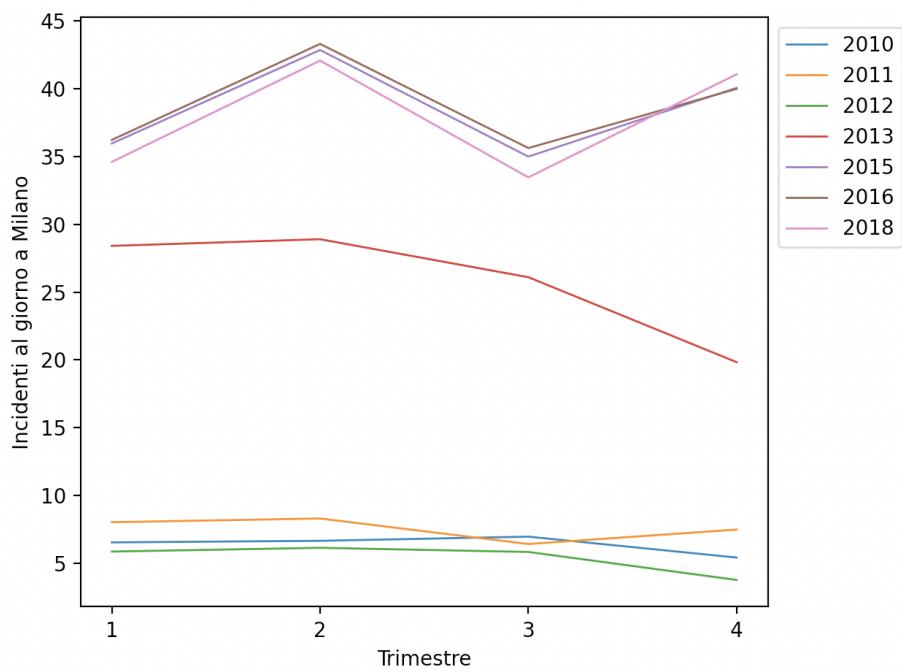


Figura 3.12: Incidenti per trimestre a Milano

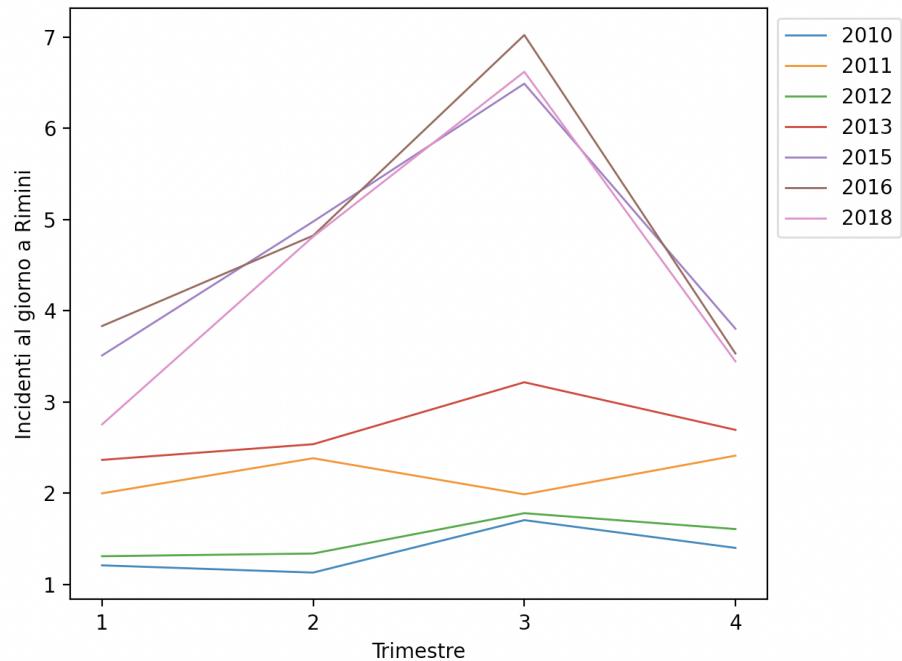


Figura 3.13: Incidenti per trimestre a Rimini

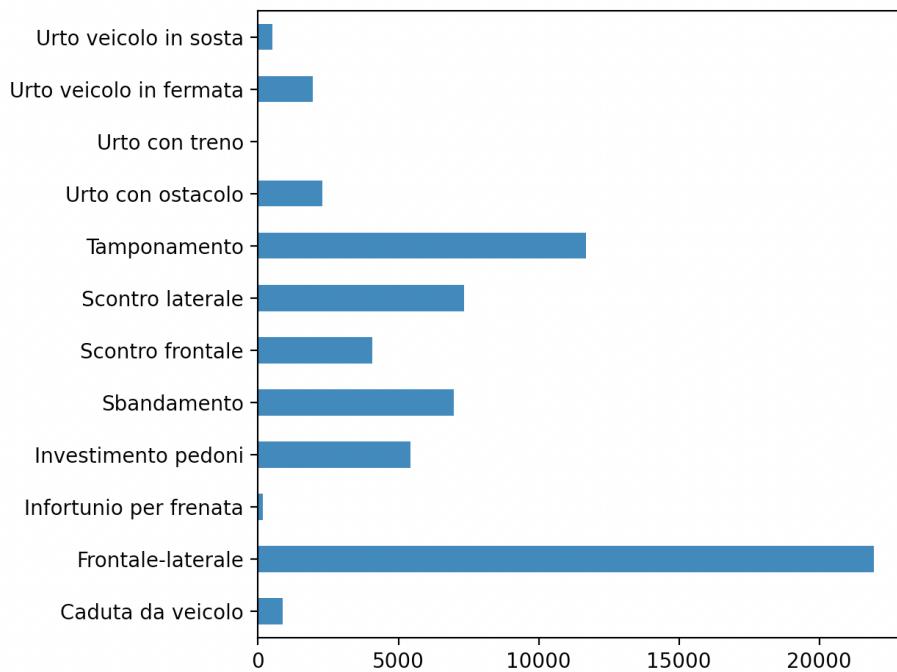


Figura 3.14: Tipologia di incidente

I grafici equivalenti di Milano e Rimini mostrano la stessa tendenza di incremento del numero di incidenti, ma a Milano, in particolare, il gap sembra essersi 'creato' nel 2013.

3.4 Dati Istat su tipi di incidenti e incroci

3.4.1 Quali incidenti avvengono con più frequenza?

Sono molto frequenti scontri frontali, laterali e tamponamenti.

3.4.2 Quali tipi di incroci provocano più incidenti?

La maggior parte degli incidenti avviene nei rettilinei e negli incroci.

3.4.3 Esistono tipi di incidenti che provocano più feriti?

Si nota che esistono tipologie di sinistri che favoriscono la presenza di un solo ferito, come gli incidenti con pedone. Al contrario, incidenti come il tamponamento e frontale, hanno un'alta percentuale di situazioni con due o più feriti.

3.4.4 Esistono incroci che favoriscono incidenti con pedoni?

Trovare un modo per dividere pedoni_feriti e tipo_incrocio

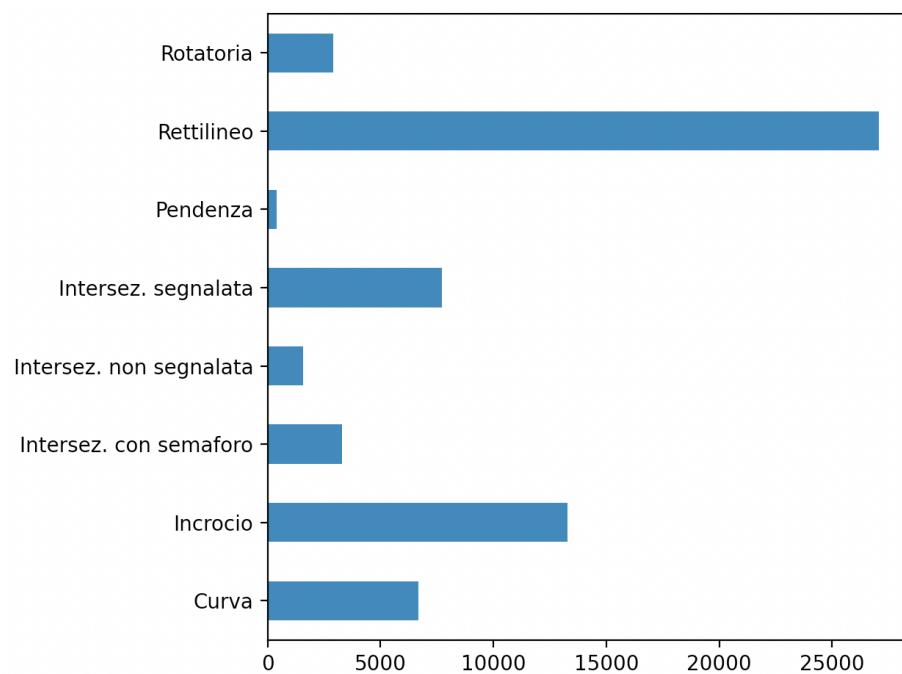


Figura 3.15: Tipologia di intersezioni

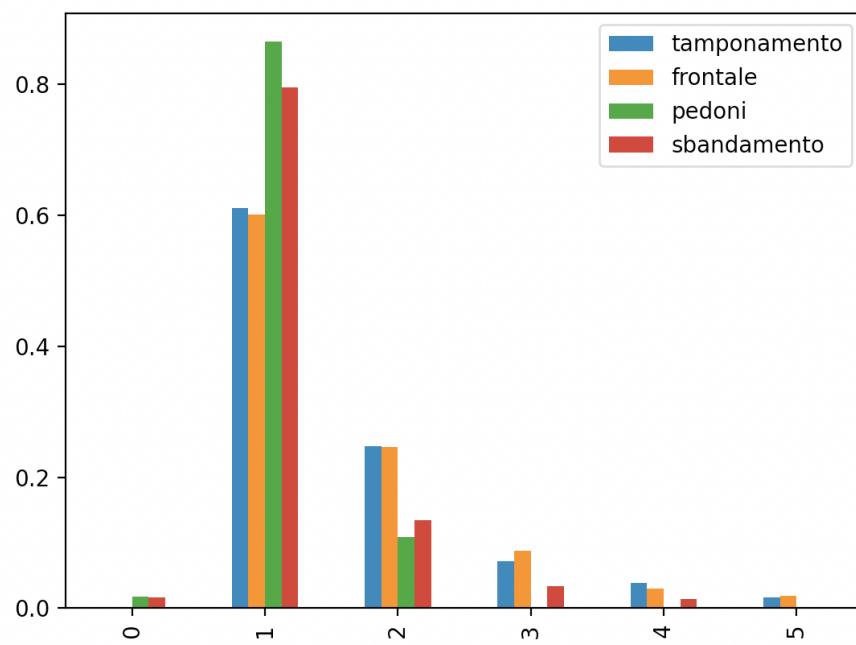


Figura 3.16: Numero di feriti in base alla natura dell'incidente

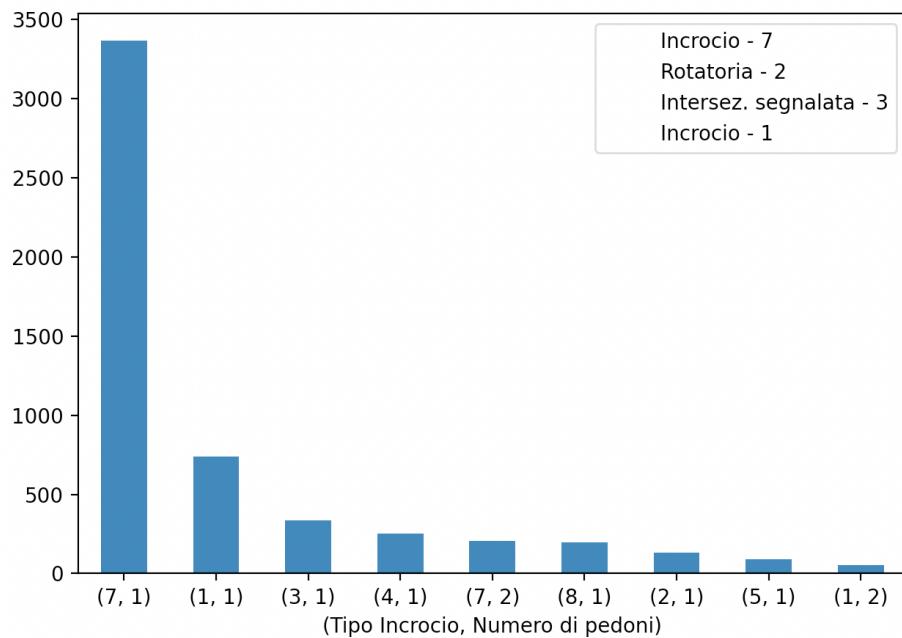


Figura 3.17: Tipologia di intersezioni e pedoni coinvolti

pedoni_feriti tipo_incrocio	1	2	3	4
Curva	195	15	1	0
Dosso	17	0	0	0
Incrocio	740	55	2	0
Intersez. con semaforo	250	20	1	0
Intersez. non segnalata	89	5	0	0
Intersez. segnalata	335	19	2	1
Passaggio a livello	1	0	0	0
Pendenza	42	4	1	1
Rettilineo	3369	206	18	6
Rotatoria	134	10	0	0

Tra i tipi di strada che favoriscono incidenti con pedoni spiccano i rettilinei, probabilmente in parte per l'alta velocità dei veicoli, ma anche per l'alto volume di tratti di strada di questo tipo.

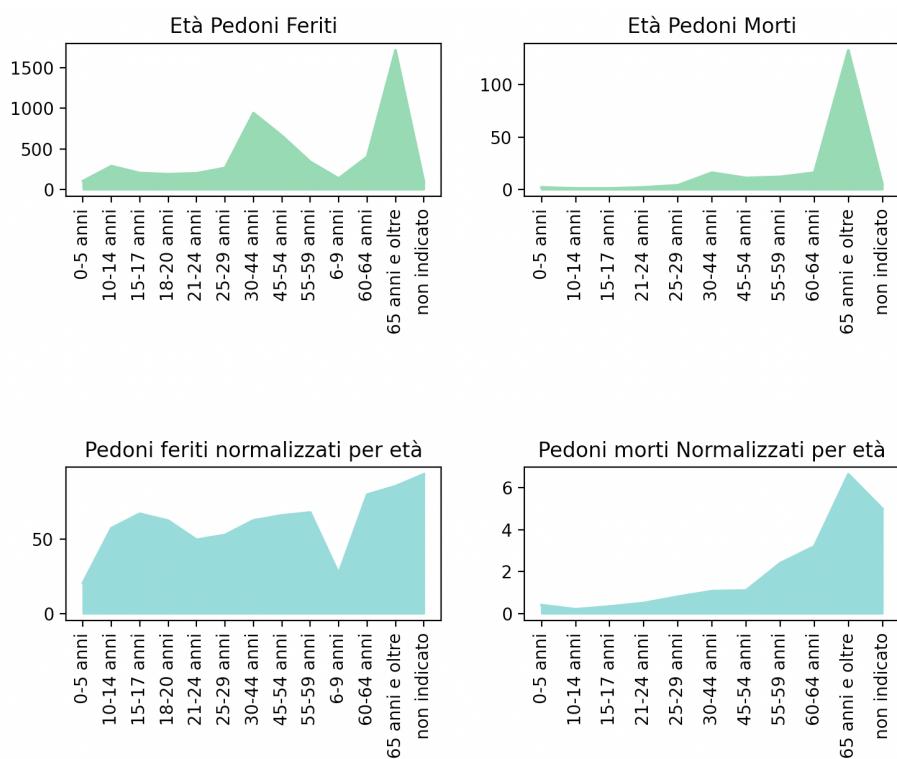


Figura 3.18: Fasce di età dei pedoni coinvolti in incidenti

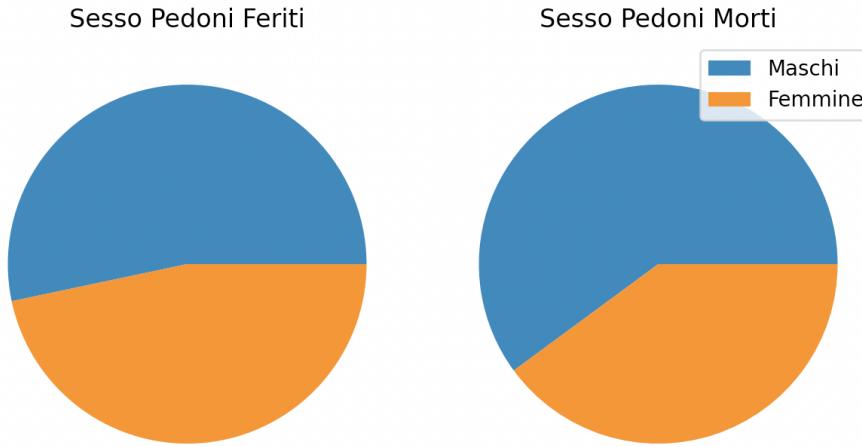


Figura 3.19: Pedoni coinvolti in incidenti per genere

3.4.5 Ci sono aspetti salienti sui pedoni coinvolti?

La fascia di età più colpita dagli incidenti è quella dei 65 anni, va comunque detto che questo gruppo probabilmente contiene la maggior parte degli individui. Se si normalizza per anni contenuti in ogni fascia, ipotizzando un numero costante di persone di ogni età, si ottiene un grafo, per quanto riguarda i pedoni feriti, molto differente.

La normalizzazione tiene conto che la fascia di età '65 anni e oltre' vale venti anni.

Dati non sorprendenti...

3.5 Dati ACI

3.5.1 Esiste correlazione tra incidenti e feriti?

E allo stesso modo, esiste correlazione tra morti e incidenti, o tra morti e feriti? Ovviamente si attende un esito positivo a queste domande. L'indice di correlazione utilizzato è il coefficiente di Pearson.

Incidenti Feriti	Incidenti Morti	Feriti Morti
0.9827	0.8205	0.8332

Il coefficiente di Pearson, per quanto riguarda Incidenti e Feriti, è molto vicino a uno, quindi i due campioni sono strettamente correlati. I coefficienti riguardanti i morti sono meno vicini uno, probabilmente per il minor numero di incidenti mortali, che rendono il campione più ristretto

Tracciando il grafo dei primi trenta valori del dataset ACI, è chiaramente visibile la correlazione tra numero di incidenti e numero di feriti.

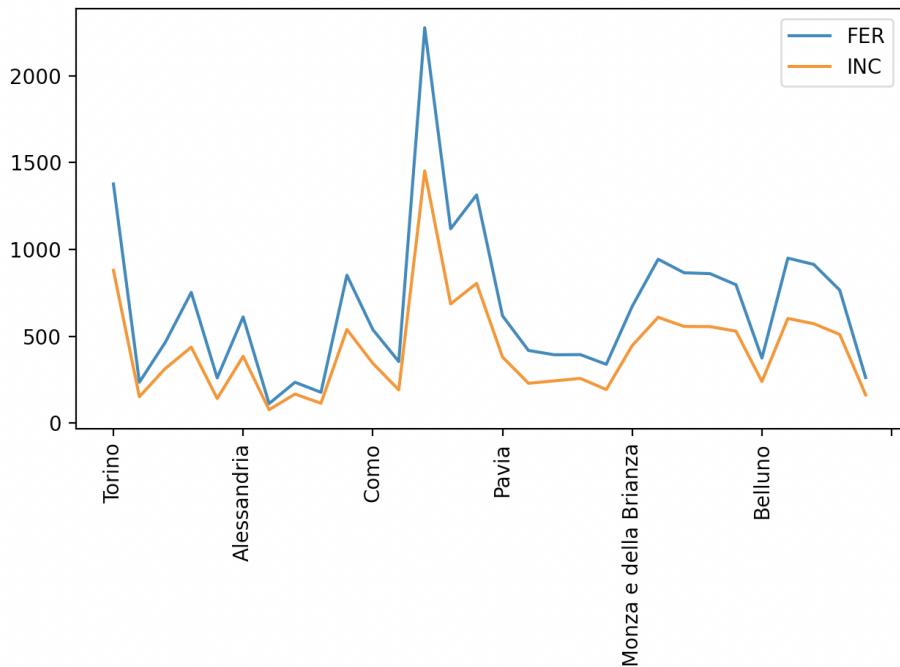


Figura 3.20: Correlazione tra numero di feriti e incidenti

3.5.2 Quali sono le autostrade con più incidenti?

Si può notare subito che le autostrade con più incidenti sono anche quelle più trafficate, come l'Autostrada del Sole e l'Adriatica.

3.5.3 Autostrade pericolose a Milano?

3.5.4 In quali mesi avvengono più incidenti?

Le curve sono state normalizzate per bilanciare il volume di incidenti maggiori per l'autostrada Adriatica. L'Adriatica e l'Aurelia, autostrade utilizzate molto durante i mesi caldi, hanno un picco di incidenti in Luglio e Agosto, mentre l'A1, ha picco più basso in Agosto, probabilmente perché bilanciato dagli incidenti in inverno intorno a Milano.

Infatti se si prendono solo i dati della A1 nella provincia di Milano, si nota che in Maggio e Settembre avvengono la maggior parte degli incidenti. Invece sull'Adriatica prevalgono incidenti in Luglio e Agosto.

3.5.5 In Quali orari avvengono incidenti sulle autostrade?

Il grafo conferma per alcune autostrade i picchi di incidenti per il traffico durante orari di punta, in particolare è molto visibile per la Torino–Trieste e per la Tangenziale ovest.

orari di altri luoghi, ho fatto solo milano

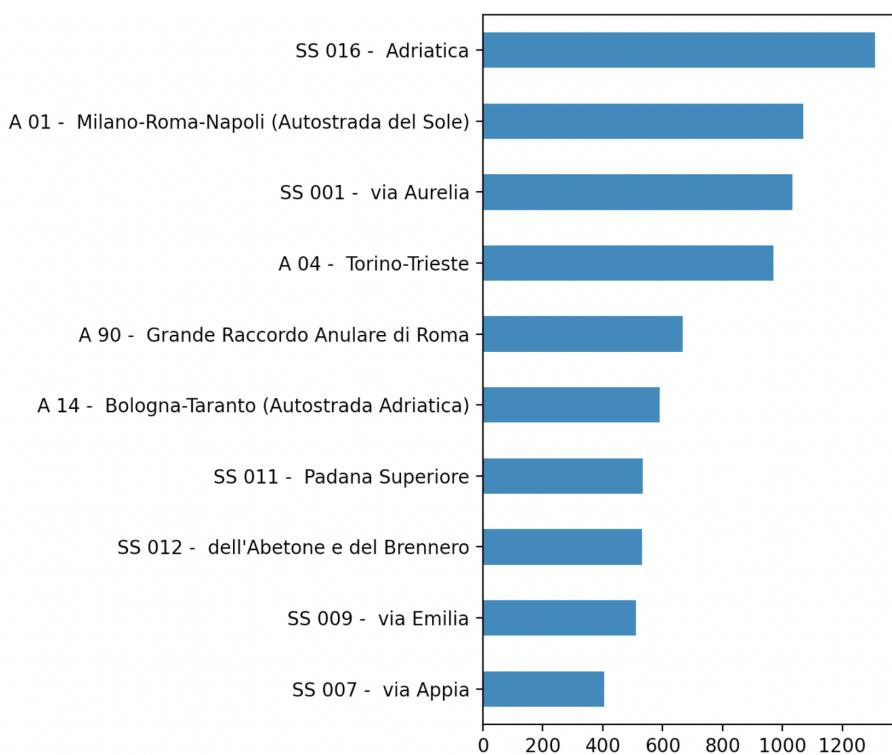


Figura 3.21: Autostrade con più incidenti nel 2018

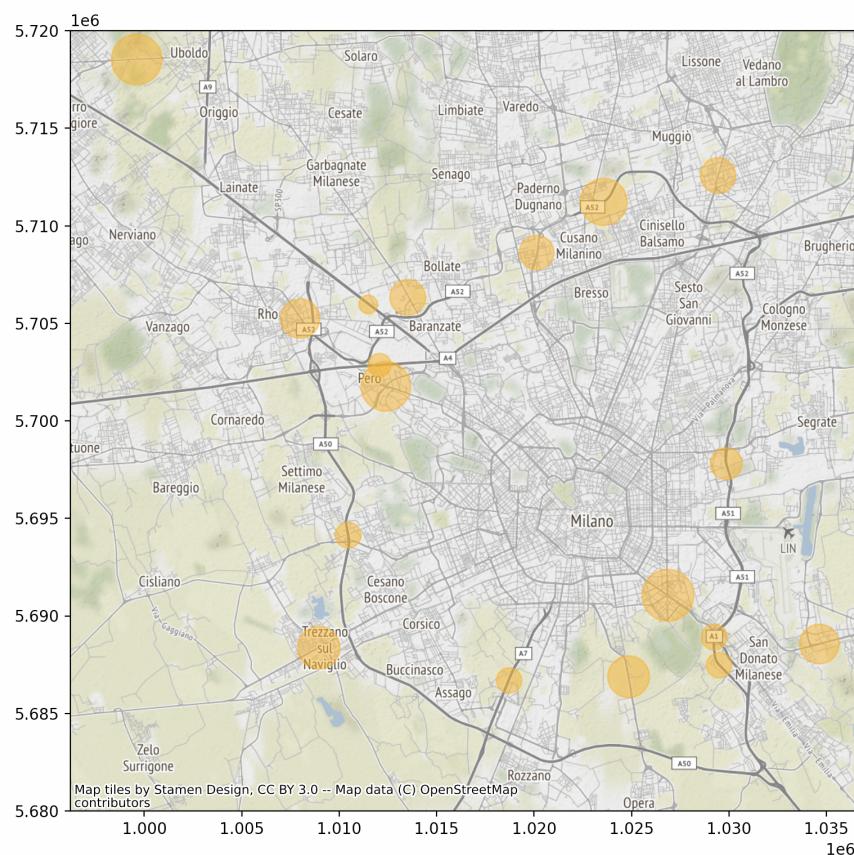


Figura 3.22: Autotrade con più incidenti nel 2012

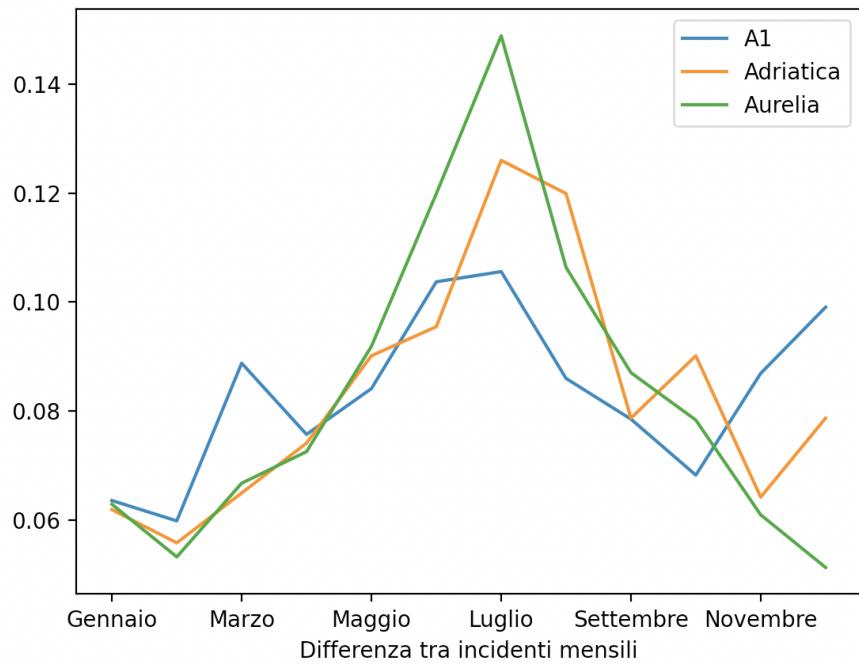


Figura 3.23: Incidenti per mese nel 2018

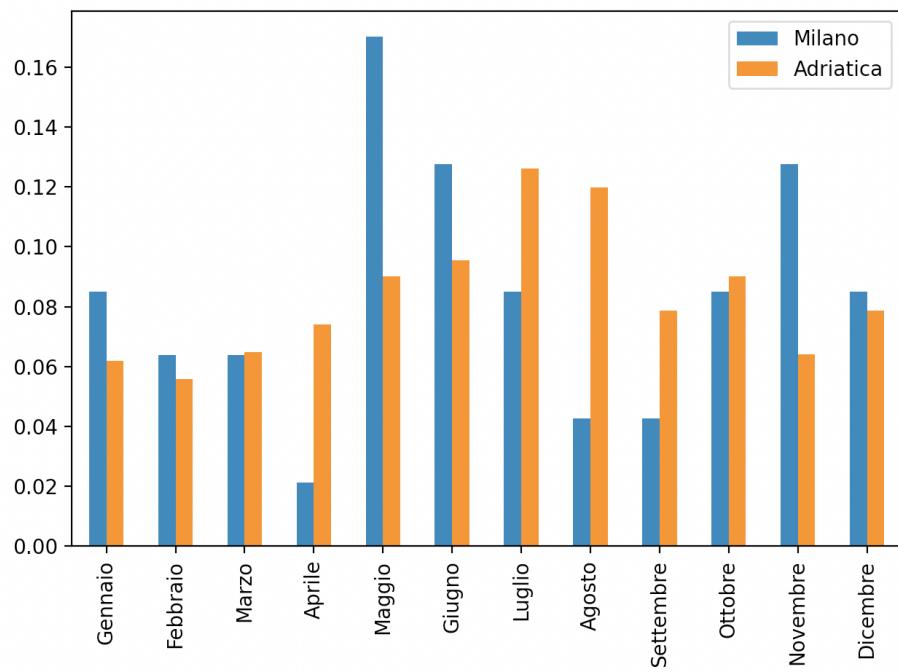


Figura 3.24: Incidenti in provincia di Milano e sull'Adriatica

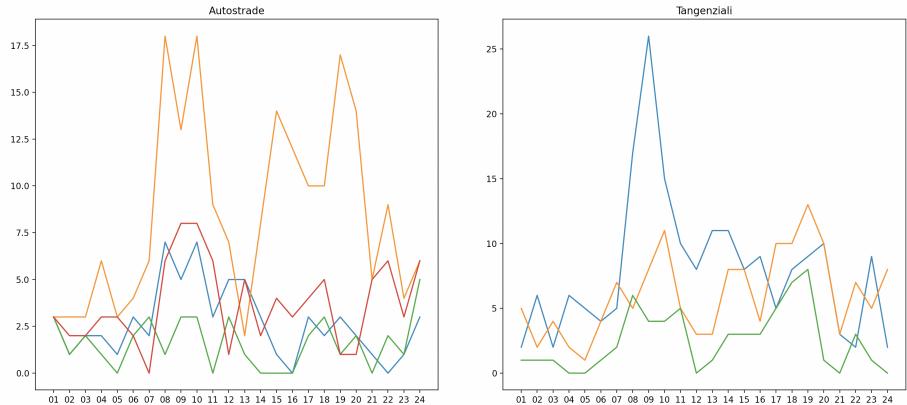
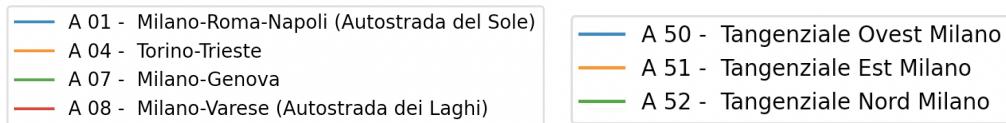


Figura 3.25: Incidenti nelle principali autostrade di Milano



3.5.6 Quali autostrade sono utilizzate di più per viaggiare in Agosto?

Il grafo rappresenta in quali autostrade sono avvenuti più incidenti in base al mese di Agosto del rispettivo anno. Non è sorprendente che le autostrade caratterizzate da un alto numero di bollini rossi e neri siano anche le più pericolose. In particolare spiccano la SS16 Adriatica e SS1 Aurelia, non è difficile trovare il motivo dell'alto numero di incidenti, queste ultime infatti sono Strade Statali molto lunghe e con corsie di marcia non separate, per non parlare del fatto che entrambe le strade passano attraverso un alto numero di centri abitati,

3.5.7 Le autostrade più utilizzate cambiano a seconda dell'anno?

Il numero di persone in vacanza nelle località del centro e sud Italia cambiano a seconda dell'anno, il numero di incidenti rispecchia questo cambiamento?

Controllando le prime cinque strade per incidentalità ogni anno, si ottiene un'immagine abbastanza stabile, in quanto gli itinerari in testa alla classifica sono sempre gli stessi. Si nota, che la strada statale Adriatica è sempre in prima posizione, con una media di 171.5 incidenti nel mese di Agosto. Le posizioni successive invece cambiano a seconda dell'anno, in particolare però, l'autostrada A4 (Torino-Trieste) ha progressivamente aumentato il numero di incidenti, evento probabilmente dovuto all'aumento di popolarità delle vacanze in montagna.

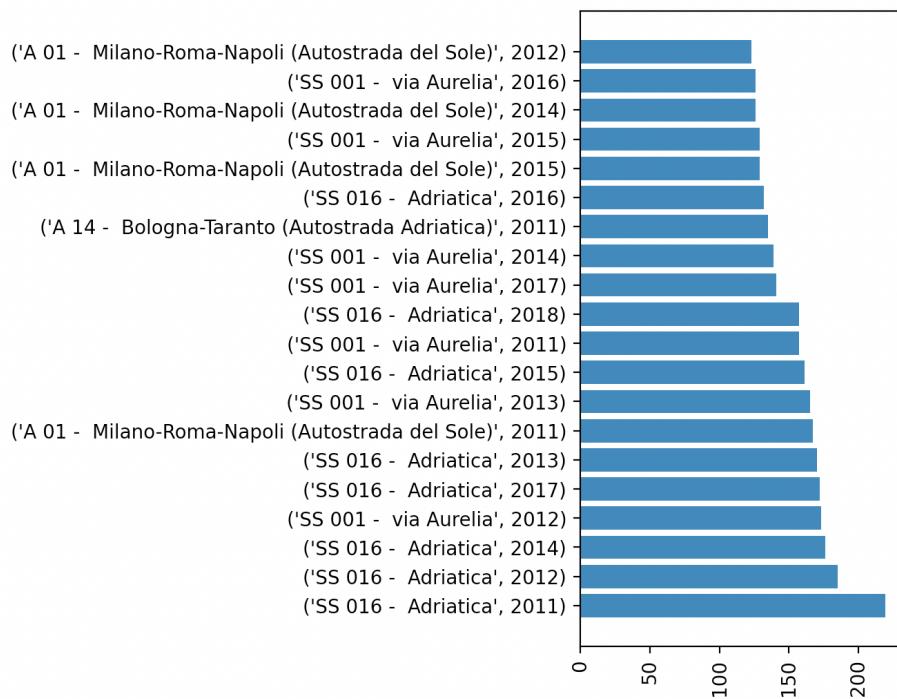


Figura 3.26: Autostrade con più incidenti per anno, in Agosto



Figura 3.27: SS16 Adriatica vicino Foggia

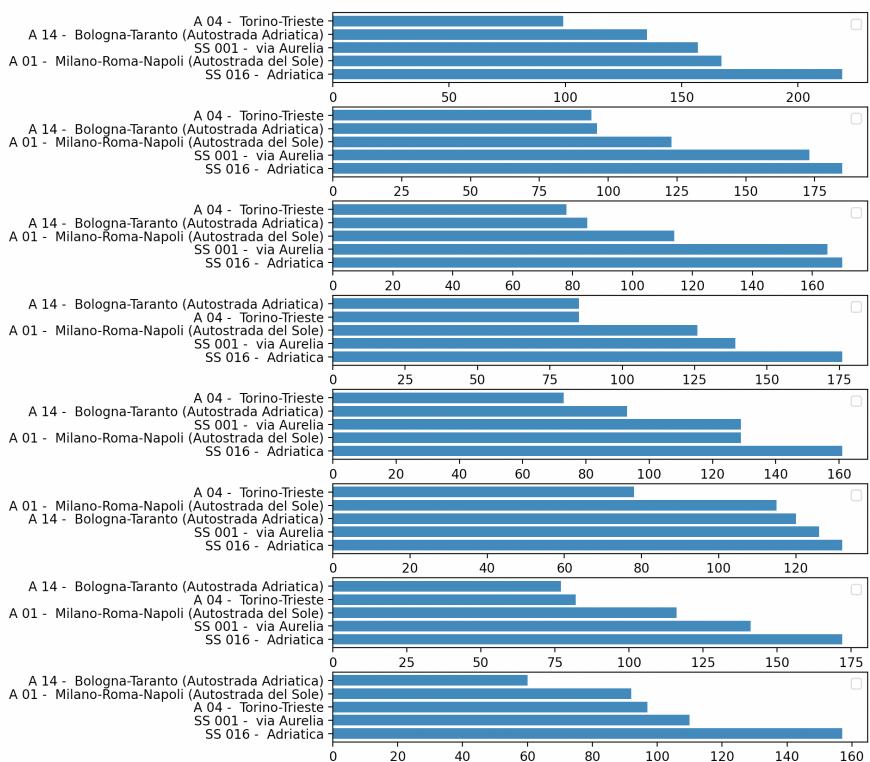


Figura 3.28: Incidenti in Strade per anno

Capitolo 4

Dati su Meteo

Bibliografia