

# Studio sull'incidentalità stradale tramite dataset aperti

Gabriele Padovani - Matr. 909165

Termine:

## 1 Introduzione

Nel documento di tesi si è esplorato che tipo di analisi è possibile realizzare, avendo a disposizione sufficiente quantità di dati liberi. Gli argomenti del lavoro, si concentrano, in particolare, sull'ambito dell'incidentalità stradale, dando enfasi alle posizioni di questi ultimi, sia dal punto di vista delle coordinate geografiche, sia da quello delle strade con maggiore numero di sinistri.

Si è inoltre tentato di ricavare il legame tra il numero di incidenti e svariati fattori, tra cui, la presenza di passeggeri, l'età del conducente, la pavimentazione della direttrice, o la tipologia di incrocio.

## 2 Origine dei dati

La maggior parte delle informazioni utilizzate, generalmente raggruppate in dataset, sono state reperite nei principali siti di open data, come quello del comune di Milano<sup>1</sup>, o quello del ministero dei trasporti<sup>2</sup>.

D'altra parte, i file principali, contenenti i dati riguardanti gli incidenti, sono stati reperiti sul sito dell'Istat<sup>3</sup>, e su quello dell'Acì [1].

Le informazioni riguardanti le coordinate degli incidenti avvenuti a Milano, provengono invece dal giornale online TheSubmarine [2], che ha ottenuto il rilascio di una parte di questi dati, normalmente oscurati per rispettare la privacy degli individui coinvolti nei sinistri.

## 3 Analisi dei dati geolocalizzati

Per quanto le analisi riguardanti gli eventi che dispongono di localizzazione, si siano incentrate soprattutto sulla relazione tra incidentalità e pavimentazione della strada, è stata posta particolare enfasi anche sul rapporto tra numero di incidenti e presenza di linee di trasporti pubblici, e allo stesso modo sulla differenza di sinistri nelle vicinanze di autovelox.

Queste analisi si concentrano sulla città di Milano, in quanto le informazioni rilasciate da Istat sono limitate al capoluogo lombardo.

---

<sup>1</sup><https://dati.comune.milano.it/>

<sup>2</sup><https://www.mit.gov.it/>

<sup>3</sup><https://www.istat.it/it/archivio/87539>

## 4 Analisi con dati Istat

Per quanto riguarda le analisi realizzate sui dataset provenienti dal sito Istat, ci si è concentrati sull'evoluzione dell'incidentalità nei diversi orari della giornata, e allo stesso modo al passare dei mesi.

Si è inoltre controllato se esistessero tipologie di strade e incroci che favorissero un numero maggiore di sinistri, e allo stesso modo, se alcuni di questi presentassero un valore anomalo di pedoni coinvolti.

Infine, si è tentato di individuare l'esistenza di fattori di distrazione del conducente, in particolare nella forma dei telefoni cellulari e nella presenza di altri passeggeri a bordo.

## 5 Analisi con dati Aci

I dataset provenienti dal sito Aci, d'altra parte, sono stati oggetto di varie analisi per quanto riguarda il luogo dell'incidente, inteso come il nome della strada o la provincia in cui questo è avvenuto.

Più nello specifico, si sono create varie rappresentazioni, incentrate sugli incidenti per regione, per le province di Lombardia e Lazio, e per le principali autostrade, sia in prossimità di Milano sia in generale in Italia.

Infine, riprendendo alcune delle analisi realizzate su dati Istat, si sono calcolate alcune informazioni, incentrate soprattutto sull'evoluzione degli incidenti al cambiare dell'orario o del mese dell'anno.

## 6 Analisi con dati Meteo

I dataset riguardanti le informazioni meteo, sono stati utilizzati per un'analisi sulla correlazione tra i fattori atmosferici e l'incidentalità in prossimità della città di Milano. In particolare, ci si è concentrati sulle informazioni riguardanti la temperatura, l'umidità e la velocità del vento. Causa la bassa precisione dei dati utilizzati, tuttavia, questo capitolo è da intendersi come una dimostrazione di quanto, un calcolo realizzato tra due dataset scollegati, possa comunque fruttare un risultato plausibile ma infondato.

D'altra parte, non è da escludere che, tramite analisi più approfondite, sia possibile ottenere risultati validi.

## 7 Conclusioni

È opportuno evidenziare che la maggior parte delle analisi realizzate, è stata portata a termine facendo compromessi tra la disponibilità del dato e la precisione di questo.

Se per analisi realizzate su dataset Istat, le informazioni sono già parzialmente aggregate, e quindi di facile elaborazione, per i dati riguardanti le coordinate degli incidenti, o quelli sugli autovelox, uno dei maggiori problemi è stata l'assenza dei dati necessari come, ad esempio, le date di posizionamento delle telecamere a Milano.

Un altro problema, questa volta presente prevalentemente nelle analisi di dati già aggregati, è stata la difficoltà nel creare un contesto valido attorno alle

informazioni ricavate. È il caso dei sinistri divisi a seconda della fascia di età del conducente, dove è possibile ottenere esiti marcatamente differenti a seconda del denominatore utilizzato per stimare la popolazione per età.

## Riferimenti bibliografici

- [1] Automobile Club Italia. *LOCALIZZAZIONE DEGLI INCIDENTI STRADALI*. URL: [http://www.aci.it/fileadmin/documenti/ACI/Trasparenza/Open\\_Data/Catalogo\\_localizzazione\\_in\\_formato\\_Open\\_2019.pdf](http://www.aci.it/fileadmin/documenti/ACI/Trasparenza/Open_Data/Catalogo_localizzazione_in_formato_Open_2019.pdf).
- [2] Tommaso Sansone. «Dove avvengono più incidenti stradali a Milano?» In: *thesubmarine* (2018). URL: <https://thesubmarine.it/2018/06/20/mappa-incidenti-stradali-milano>.