



# UNIVERSITÀ DEGLI STUDI DI MILANO

## FACOLTÀ DI SCIENZE E TECNOLOGIE

### Studio sull'incidentalità stradale tramite dataset aperti

Gabriele Padovani - Matr. 909165

Termine: 25 gennaio 2021

## 1 Introduzione

Nel documento di tesi si sono esplorate quali tipi di analisi è possibile realizzare, avendo a disposizione una sufficiente quantità di dati liberi. Gli argomenti di questo lavoro si concentrano, in particolare, sull'ambito dell'incidentalità stradale, dando enfasi alla posizione in cui questi ultimi avvengono, 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 la quantità di incidenti e svariati altri fattori, tra cui, la presenza di passeggeri, l'età del conducente, la pavimentazione della direttrice, o la tipologia di incrocio.

## 2 Origine dei dati

Le informazioni utilizzate, generalmente raggruppate in dataset, sono state reperite nei principali siti di open data, come quello del comune di Milano<sup>1</sup> per i dati sugli accessi in area C, o quello del ministero dei trasporti<sup>2</sup>, nel caso delle patenti.

D'altra parte, i file principali, contenenti i dati riguardanti gli incidenti, sono stati reperiti sul sito dell'Istituto nazionale di statistica<sup>3</sup>, e su quello dell'Automobile club Italia<sup>4</sup>.

I dataset contenenti le coordinate degli incidenti avvenuti a Milano, provengono invece dal giornale online TheSubmarine<sup>5</sup>, che ha ottenuto, da parte di Istat, il rilascio di una frazione di questi dati, normalmente oscurati per rispettare la privacy degli individui coinvolti nei sinistri.

Di fondamentale importanza è stato anche il portale di dati geolocalizzati del comune di Milano<sup>6</sup>, tramite cui si sono ottenuti i percorsi dei trasporti pubblici, e le informazioni sul manto stradale.

Infine, per poter rappresentare i risultati ricavati, in molti casi si è fatto uso di alcune mappe, riguardanti le regioni e le province italiane, trovate su un *repository* Github<sup>7</sup>.

---

<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>

<sup>4</sup>[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)

<sup>5</sup><https://thesubmarine.it/2018/06/20/mappe-incidenti-stradali-milano>

<sup>6</sup><https://geoportale.comune.milano.it/>

<sup>7</sup><https://github.com/openpolis/geojson-italy>

### 3 Analisi

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

Questi calcoli prendono in considerazione solamente la città di Milano, in quanto le informazioni rilasciate da Istat sono limitate al capoluogo lombardo.

Nelle analisi realizzate sui dataset provenienti dal sito Istat, ci si è soffermati soprattutto sull'evoluzione dell'incidentalità nei diversi orari della giornata, e allo stesso modo al passare dei mesi.

Si è inoltre controllata l'esistenza di 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.

Altre analisi riguardanti queste informazioni, si sono concentrate su differenze nella guida di uomini e donne, e allo stesso modo, nel numero di sinistri al cambiare della tratta stradale.

Infine, si è tentato di individuare l'esistenza di fattori che potessero distrarre il conducente, in particolare nella forma dei telefoni cellulari o nella presenza di altri passeggeri a bordo.

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 le diverse regioni italiane, per le province di Lombardia e Lazio, e per le principali autostrade, sia in prossimità di Milano, sia più in generale in Italia.

Infine, riprendendo alcune delle analisi realizzate su dati Istat, si sono ripetuti i calcoli incentrati, soprattutto, sull'evolversi del numero di sinistri al cambiare dell'orario, o del mese dell'anno.

I dataset riguardanti le informazioni meteo, sono stati utilizzati per un'analisi sulla correlazione tra i fattori atmosferici e l'incidentalità per le strade in prossimità della città di Milano. In particolare, ci si è concentrati sui dati riguardanti la temperatura, l'umidità e la velocità del vento.

Causa la bassa precisione dei dataset utilizzati, tuttavia, questo capitolo è da intendersi come una dimostrazione di quanto, un calcolo realizzato tra due campioni scollegati, possa fruttare un risultato plausibile ma comunque infondato.

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

### 4 Conclusioni

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

Se per analisi realizzate su dataset Istat, i campi sono già parzialmente aggregati, 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 delle informazioni necessarie come, ad esempio, le date del posizionamento delle telecamere tutor a Milano.

Un altro problema, questa volta presente prevalentemente nei calcoli su dati già aggregati, è stata la difficoltà nel creare un contesto valido attorno alle informazioni ricavate. È il caso dei sinistri divisi in base alla fascia di età del conducente, dove è possibile ottenere esiti marcatamente differenti a seconda del denominatore utilizzato per stimare la popolazione per età.