

# Analisi della correlazione tra incidenti e tempo atmosferico a New York

Periodo temporale: 2015-2017

Danilo Deponti - 737838 Mattia Curatitoli - 735722

## Introduzione

introduzione

## Analisi dei dati - Csv Incidenti

### Sorgente dati: Incidenti a New York

Il dataset contenente gli incidenti stradali avvenuti a New York nel periodo temporale da noi analizzato è stato ottenuto attraverso la piattaforma *NYC Open Data* nella sezione *Sicurezza pubblica*. Il sito mette a disposizione numerosi sistemi per filtrare i dati sugli incidenti stradali, come quello per lasso temporale, utilizzato nel progetto. E' possibile esportare questi dati in formato *CSV*, *TSV*, *XML*, *RSS* e *RDF*

### Informazioni raccolte

Analizzando il massiccio dataset ricavato, sono state rilevate le seguenti informazioni che lo caratterizzano:

#### Incidente

L'incidente è l' entità caratterizzante del dataset, questa entità fornisce informazioni come la chiave univoca dell' incidente, data, ora, posizione (latitudine, longitudine) e numero di morti/feriti coinvolti. Sia per i morti che per i feriti è possibile sapere se erano pedoni, ciclisti o conducenti. Inoltre, vi sono anche informazioni riguardanti il tipo di veicolo e in quale modo il relativo veicolo può avere provocato l' incidente.

## Anagrafica Tipologia veicolo

Analizzando attentamente il dataset è stato riscontrato che la descrizione della tipologia del veicolo è identica e si ripete per moltissimi incidenti. Su circa 700000 incidenti analizzati sono state rilevate solamente 200 tipologie di veicolo distinte, tra le quali più della metà sono comprensibili. Molto probabilmente, per la maggior parte degli incidenti è stato compilato un form di reportistica dove la tipologia di veicolo era definita tra un elenco.

## Anagrafica Scatenante

Come già illustrato nel precedente paragrafo, anche per lo scatenante è possibile distinguere un numero finito e molto ridotto (circa 60) di dati che caratterizzano quasi costantemente gli incidenti. Addirittura, per questa informazione, il numero di scatenanti rilevate che non hanno significato è ridotto allo 0.

## Schema logico relazionale

I dati recuperati sono stati forniti in un unico e decisamente ricco csv. I dati, per ogni incidente, erano riportati su una sola riga e non vi era alcuna logica di riutilizzo dei dati. Inoltre, appunto per come sono stati archiviati i dati, non vi era alcuna logica di normalizzazione. E' stato quindi deciso di scomporre i dati ricavati nelle seguenti entità:

*incidente* (id, unique\_key, numero\_morti, numero\_feriti, pedoni\_morti, pedoni\_feriti, ciclisti\_morti, ciclisti\_feriti, conducenti\_morti, conducenti\_feriti, *id\_luogo*, data\_incidente, ora\_incidente, created\_at, updated\_at)

La relazione *incidenti* contiene tutte le informazioni che contraddistinguono l'incidente: data, ora, numero feriti, numero morti ed una distinzione di essi in pedoni, ciclisti e conducenti.

*scatenante* (id, descrizione)

Questa relazione contiene tutti i possibili motivi che hanno provocato l'incidente. Questa relazione, per mezzo del suo identificativo univoco intero, permette di associare il motivo scatenante all'incidente per mezzo di un' apposita tabella di collegamento.

*tipoVeicolo* (id, tipologia)

La relazione *tipoVeicolo* definisce tutti i tipi di veicolo che possono caratterizzare l'incidente. Si collegano all'incidente, attraverso un identificativo univoco, per mezzo di un' apposita tabella di collegamento.

*dettagli* (*id\_incidente*, *id\_tipoveicolo*, *id\_scatenante*)

Questa relazione è ciò che collega l'incidente con i veicoli ed, eventualmente, in che modo il suddetto veicolo ha provocato l'incidente. Il collegamento è creato tramite la tripletta di attributi *id\_incidente*, *id\_tipoveicolo*, *id\_scatenante*.

*luogo* (id, quartiere, zip, latitudine, longitudine)

*strada* (id, strada)

*link\_luogo\_strada* (id\_luogo, id\_strada)

