

Esercizio: Analisi e raccolta dei requisiti per un progetto di analisi con Open Data

1. Descrizione del caso di studio scelto

Si intende realizzare un sistema per l'analisi di dati relativi agli incidenti stradali avvenuti a Roma nell'anno 2021. Lo scopo del progetto è la prevenzione e la diminuzione di tali sinistri, tramite l'individualizzazione delle strade più pericolose della città. L'obiettivo è quindi il miglioramento della viabilità metropolitana da parte del Comune di Roma.

2. Scelta degli Open Data, fonti, descrizione, classificazione e classificazione metadati, specifica della licenza e quale licenza applicare al sistema

Sono stati scelti gli Open Data pubblicati dal Comune di Roma sul sito [Open Data Roma](#). Il dataset ([Incidenti stradali nel territorio di Roma Capitale - Anno 2021 - Dataset - CKAN](#)) è costituito da un elenco di tutti gli incidenti stradali in cui è intervenuta la Polizia Locale avvenuti nel territorio di Roma Capitale nell'anno 2021. Gli incidenti verificatisi sul Grande Raccordo Anulare non sono compresi all'interno del dataset in oggetto.

Il dataset è costituito da un totale di 108 file in formato CSV, XML, JSON, contenenti i dati relativi ai pedoni, persone e veicoli coinvolti nei sinistri e suddivisi per mese. Le varie tabelle possono essere unite grazie all'ID protocollo, presente in ogni file.

La classificazione di questi formati rientra nella categoria delle 3 stelle, secondo la classificazione per valutare la qualità degli Open Data; la presenza dei metadati all'esterno del dataset fa sì che essi rientrino nella categoria 2, secondo la classificazione per valutare la qualità dei metadati.

La licenza del database è CC BY 4.0 (Creative Commons Attribuzione 4.0 Internazionale), che permette di condividere ed adattare il materiale per qualsiasi scopo, indicando la paternità della fonte. Questa licenza sarà applicata anche al sistema di dati in oggetto ed ai relativi dati prodotti.

3. Analisi e raccolta dei requisiti (generici e poi più specifici)

Si hanno a disposizione i dati relativi agli incidenti, agli automobilisti, i pedoni coinvolti ed ai veicoli.

4. Strutturazione dei requisiti in gruppi di frasi omogenee

Per gli **incidenti**, identificati attraverso codice univoco, si hanno a disposizione ID protocollo, gruppo, data/ora incidente, localizzazione strada, ID strada, civico della strada, natura incidente, particolarità della strada, tipo di strada, fondo stradale, pavimentazione stradale, segnaletica, condizione atmosferica, traffico, visibilità, illuminazione, numero feriti, numero morti, numero illesi, longitudine e latitudine.

Per le **persone a bordo di veicolo**, identificati attraverso codice univoco, si hanno a disposizione ID protocollo, ID veicolo, tipo persona, sesso, tipo lesione, decesso, utilizzo di casco/cintura, funzionamento airbag ed eventuale decesso a seguito del sinistro.

Per i **pedoni**, identificati attraverso codice univoco, si hanno a disposizione ID protocollo, tipo persona, sesso, tipo lesione, decesso ed eventuale decesso a seguito del sinistro..

Per i **veicoli**, identificati attraverso codice univoco, si hanno a disposizione ID protocollo, ID veicolo, tipo di veicolo e stato del veicolo.

Dizionario dei dati:

Colonna	Tipo
ID Protocollo	numeric
Gruppo	numeric
Data/ora incidente	timestamp
Localizzazione strada	text
ID strada	text
Civico della strada	text
Natura incidente	text
Particolarità della strada	text
Tipo di strada	text
Fondo stradale	text
Pavimentazione stradale	text
Segnaletica	text
Condizione atmosferica	text
Traffico	text
Visibilità	text
Illuminazione	text
Numero feriti	numeric
Numero morti	numeric
Numero illesi	numeric
Longitudine	numeric

Latitudine	numeric
Tipo veicolo	text
Stato veicolo	text
Tipo persona	text
Sesso	text
Tipo lesione	text
Deceduto	boolean
Deceduto dopo	boolean
Cintura/casco utilizzato	boolean
Airbag	boolean

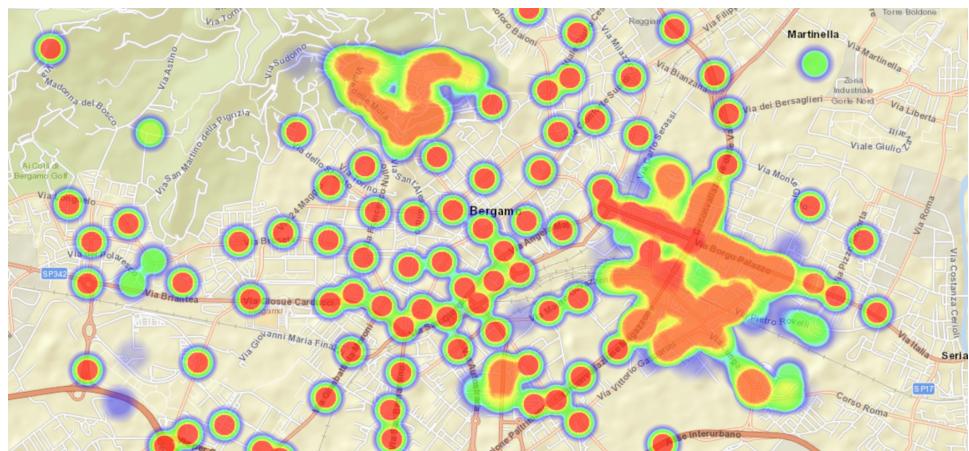
Il dizionario dei dati riguarda gli attributi utilizzati in tutti e quattro i dataset.

5. Glossario dei termini

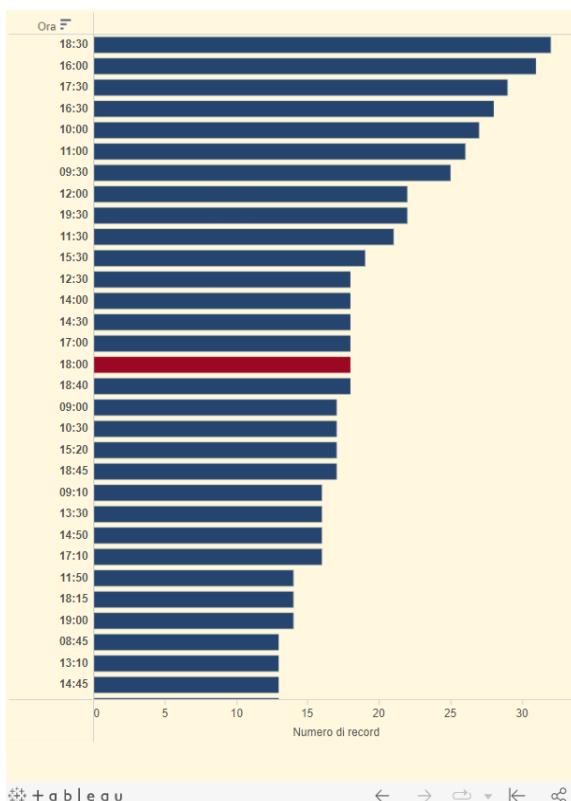
Termine	Descrizione	Sinonimi	Collegamenti
Incidente	Evento accidentale tra veicoli e persone. Può avere conseguenze sulla salute di chi vi è coinvolto	Sinistro	Persona, Pedone
Persona	Individuo all'interno del veicolo coinvolto in un incidente	Automobilista	Incidente
Pedone	Individuo a piedi per strada coinvolto in un incidente	//	Incidente
Veicolo	Mezzo di trasporto coinvolto in un incidente	Autovettura, vettura	Persona

6. Outcome atteso dal sistema (analisi di business / analisi predittive), come potrebbe essere di supporto nelle decisioni ed eventualmente quali ‘storie’ è possibile ricavare

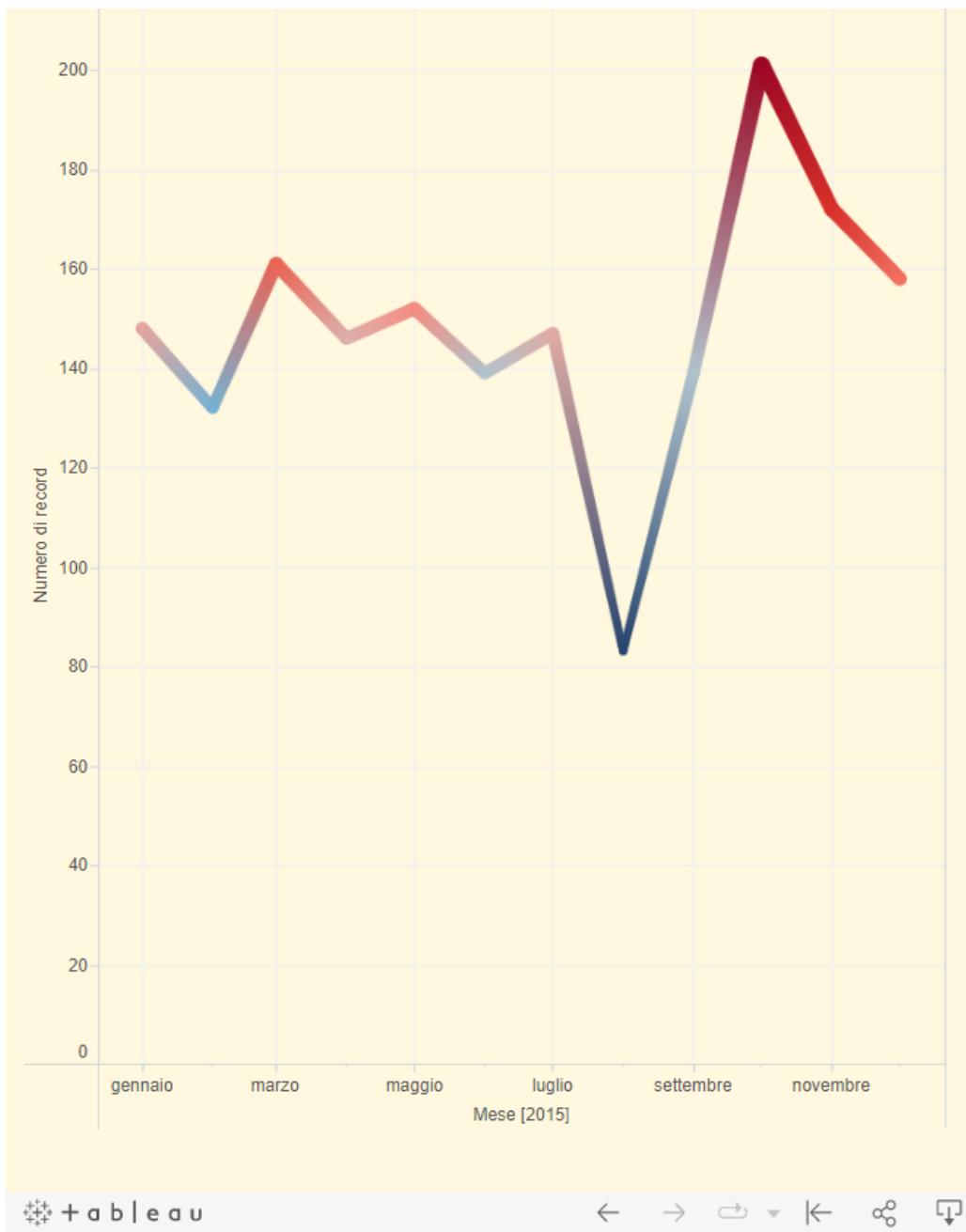
L'analisi dei dati della densità degli incidenti, incrociata con il dato di localizzazione, rappresentata su una “mappa di calore” permette di individuare le strade di Roma “più pericolose”, dove cioè è avvenuto il maggior numero di incidenti durante l'anno 2021.



L'analisi dei dati della frequenza degli incidenti, incrociata con l'ora in cui tali sinistri sono avvenuti, permette di individuare le “ore più pericolose” della metropoli.



L'analisi dei dati della frequenza degli incidenti, incrociata con i mesi dell'anno in cui tali sinistri sono avvenuti, permette di individuare il numero di incidenti per mese



Incrociando i diversi dati disponibili nel dataset è possibile estrapolare altri tipi di informazioni, quali “giorno con maggior numero di incidenti”, “sesso maggiormente coinvolto in incidenti”, “tipo di veicolo maggiormente incidentato”, “in che entità la visibilità impatta sulla gravità degli incidenti”, “sondare la relazione, se esistente, tra presenza/assenza di segnaletica e frequenza degli incidenti”.

Analisi di questo tipo permetterebbero all'Amministrazione del Comune di Roma di migliorare la viabilità della metropoli del Lazio.

Ruoli:

- **Data Engineer**: si assicura che i dati necessari per le analisi siano raccolti nel formato e nel modo corretto.
- **Data Analyst**: si occupa di individuare trend e pattern esaminando i dataset del Comune di Roma.
- **Data Scientist**: si occupa di elaborare modelli predittivi sulla frequenza, modalità e gravità degli incidenti stradali.
- **Data Journalist**: si occupa di rendere fruibili e facilmente comprensibili ai cittadini le informazioni derivate da dati complessi grazie a storytelling e dashboard interattive. Il suo è un ruolo chiave nel sensibilizzare i cittadini sulla tematica degli incidenti stradali e fare prevenzione grazie all'utilizzo degli Open Data.

C. Nardelli
F. Testa