



PROVINCIA DI LIVORNO
Osservatorio trasporti

PROGETTO SIRSS

Analisi multivariata degli incidenti su tutte le strade della Provincia di Livorno nell'anno 2008, sulla base dei dati pervenuti al 16/11/2009.

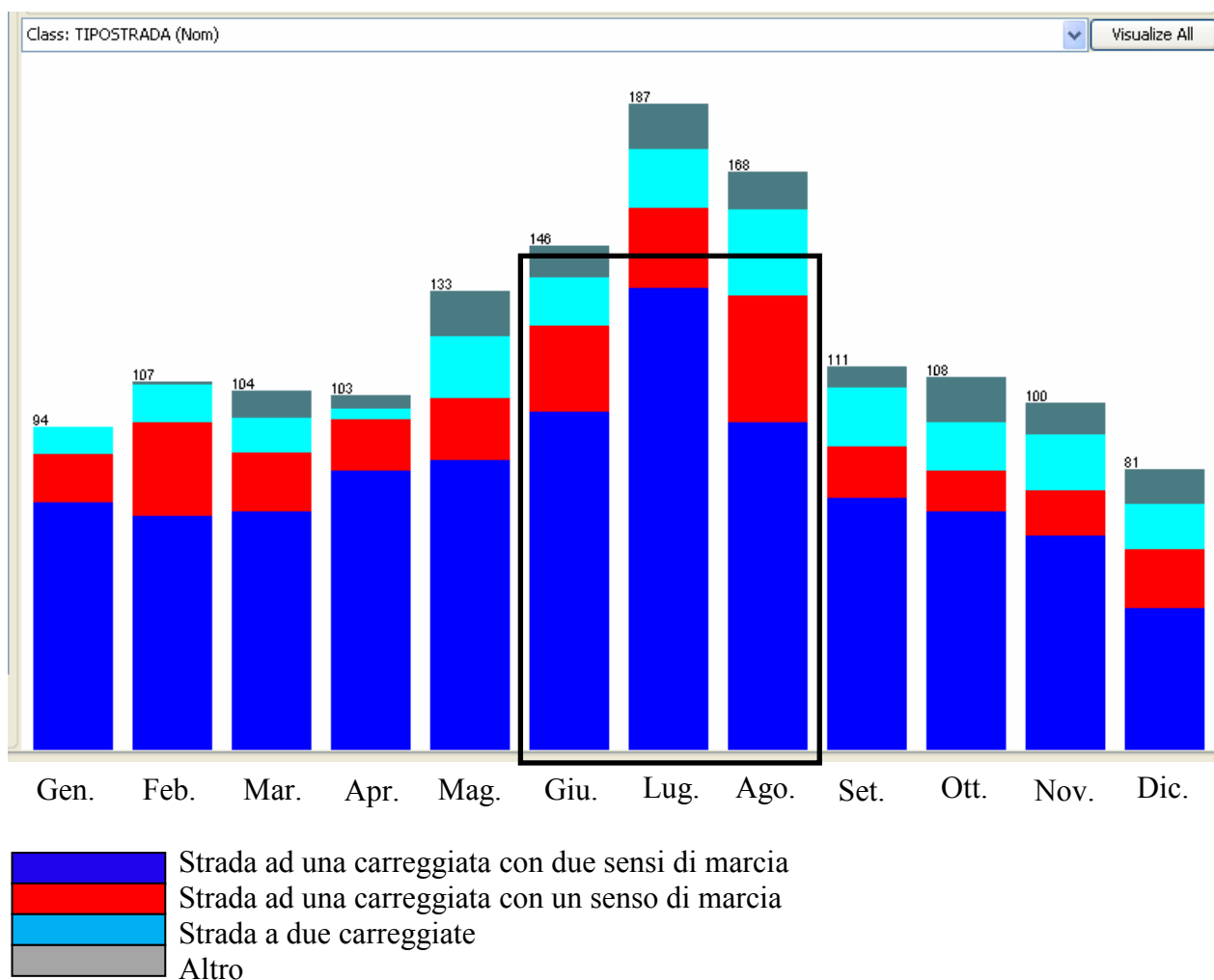
Elementi per la previsione del rischio di incidentalità

Referente del progetto: dott.ssa Irene Nicotra
Operatrice del progetto: Claudia Leandrini
Collaboratore tecnico: Ing. Massimiliano Petri

ANALISI MULTIVARIATA

Al fine di estrarre dagli incidenti la maggior quantità di informazione è stata effettuata un'analisi in un primo momento bivariata e successivamente multivariata.

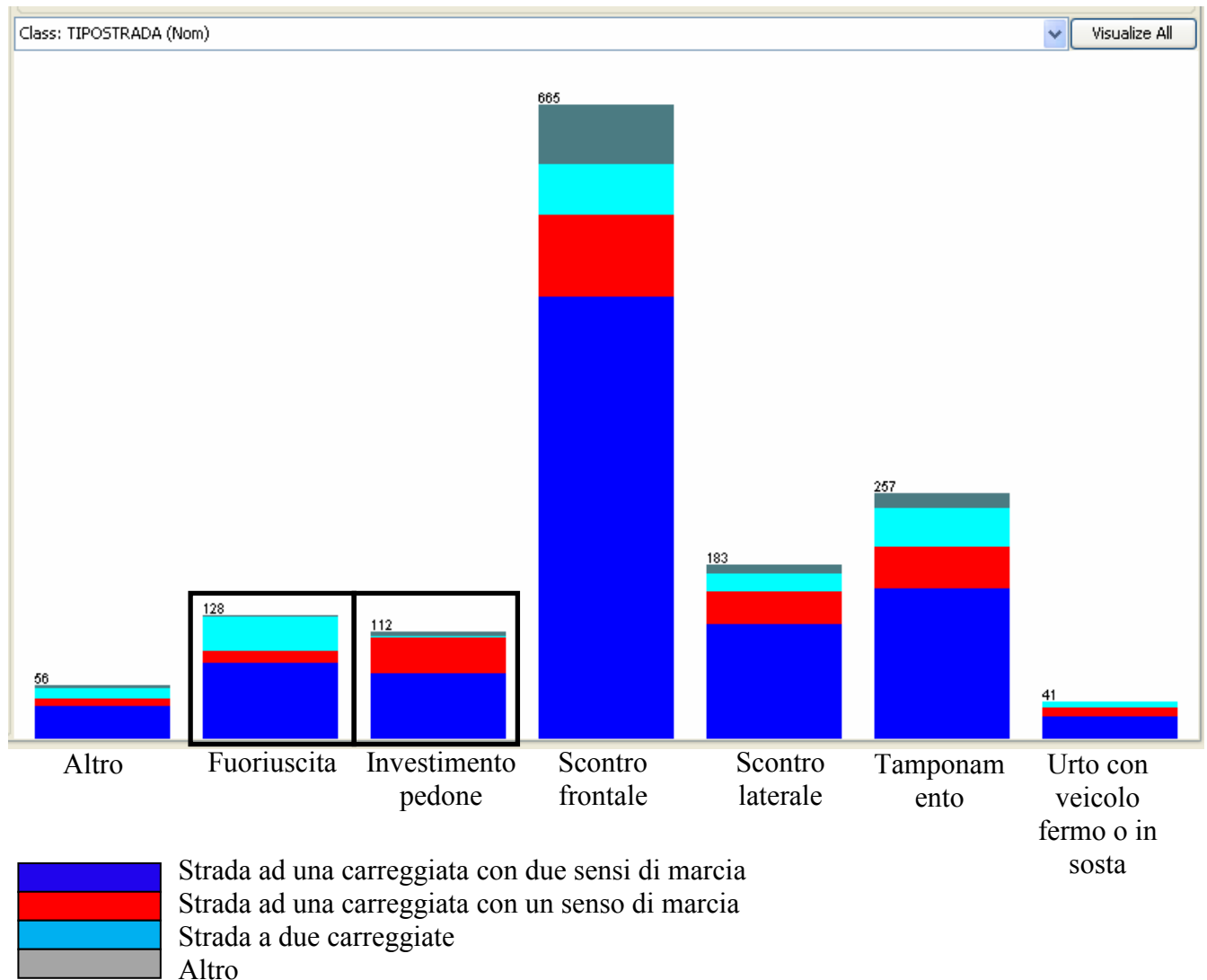
INCIDENTI NEI MESI PER TIPOLOGIA DI STRADA



Commento:

Risulta evidente come **nei mesi estivi (giugno, luglio e agosto) si abbia un aumento di incidenti su strade ad unica carreggiata e due sensi di marcia.**

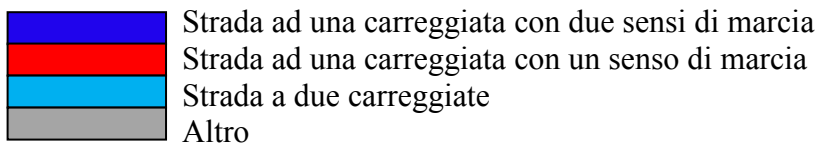
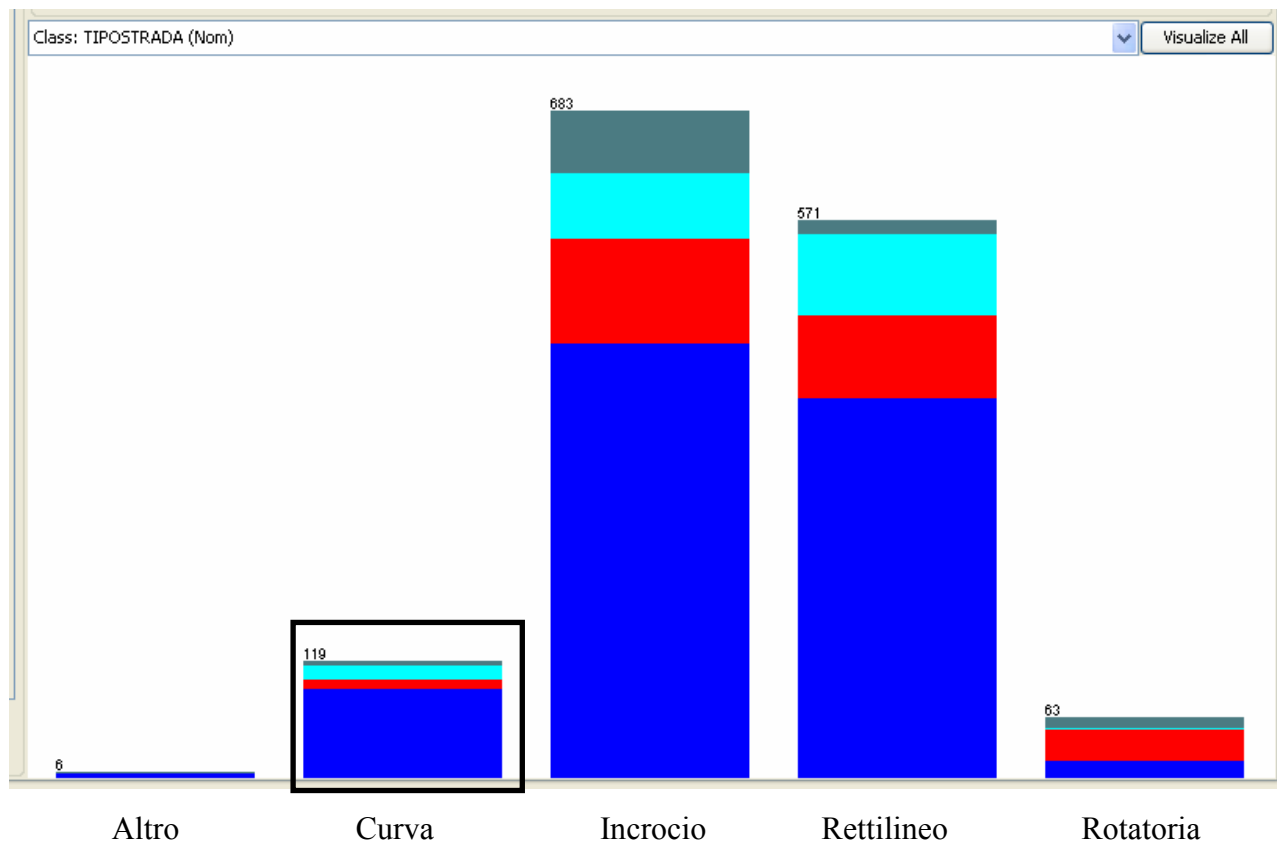
GENERE DI INCIDENTE PER TIPOLOGIA DI STRADA



Commento:

Valutando la percentuale di incidenti di una tipologia sul totale di incidenti per ogni genere, si denota come **la fuoriuscita sia prevalente per le strade a due carreggiate mentre l'investimento del pedone abbia un valore elevato per le strade ad una carreggiata con un senso di marcia.**

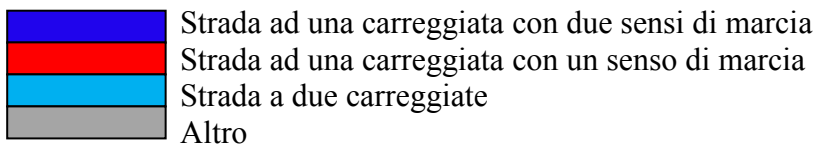
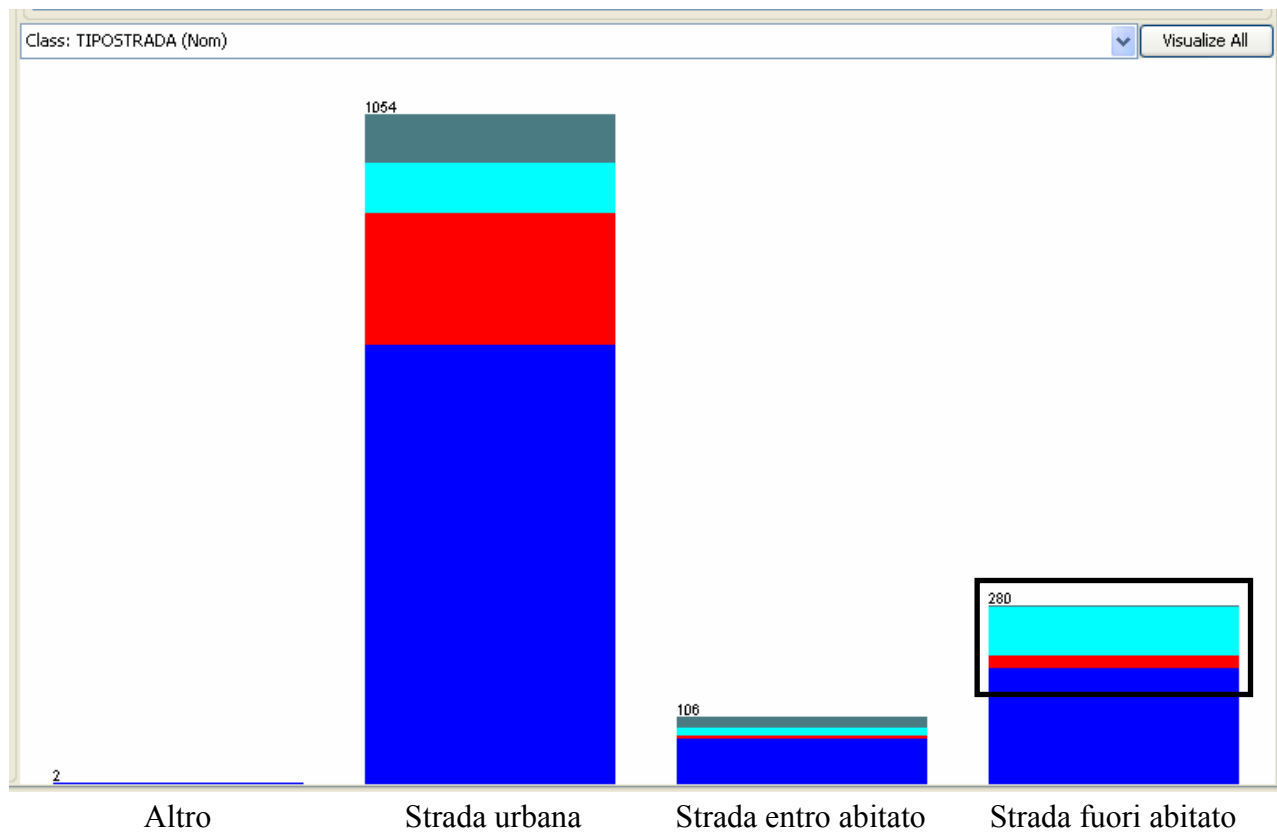
TIPO DI TRACCIATO DELL'INCIDENTE PER TIPOLOGIA DI STRADA



Commento:

Valutando la percentuale di incidenti di una tipologia sul totale di incidenti per ogni tipo di tracciato si denota come **gli incidenti in curva siano prevalenti in caso di strada ad una carreggiata con due sensi di marcia.**

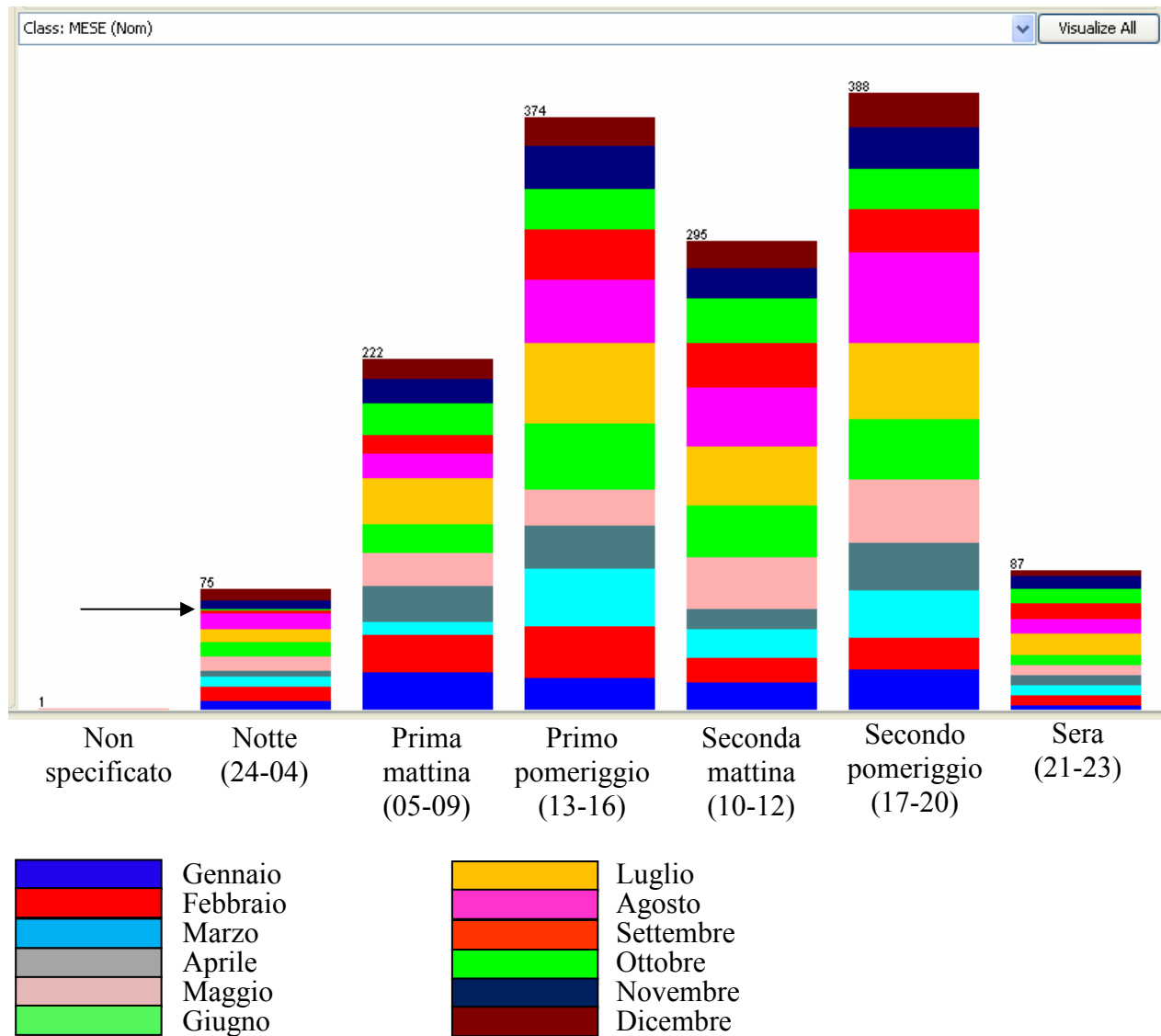
LOCALIZZAZIONE DELL'INCIDENTE PER TIPOLOGIA DI STRADA



Commento:

Gli incidenti fuori dall'abitato avvengono maggiormente su strade a due carreggiate.

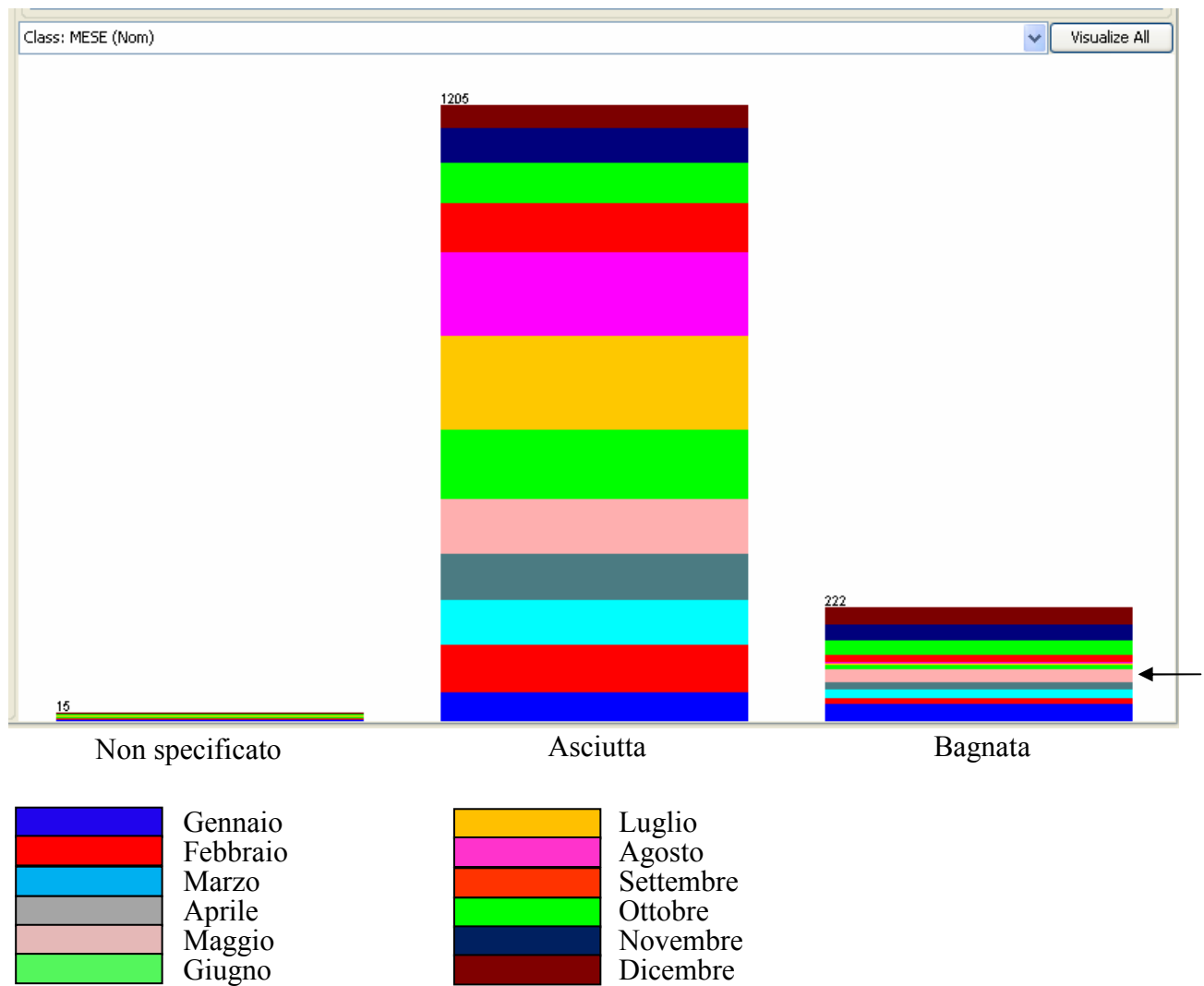
PERIODO GIORNALIERO DELL'INCIDENTE PER MESE



Commento:

Gli incidenti notturni **sono molto pochi** nei mesi di **Settembre ed Ottobre**.

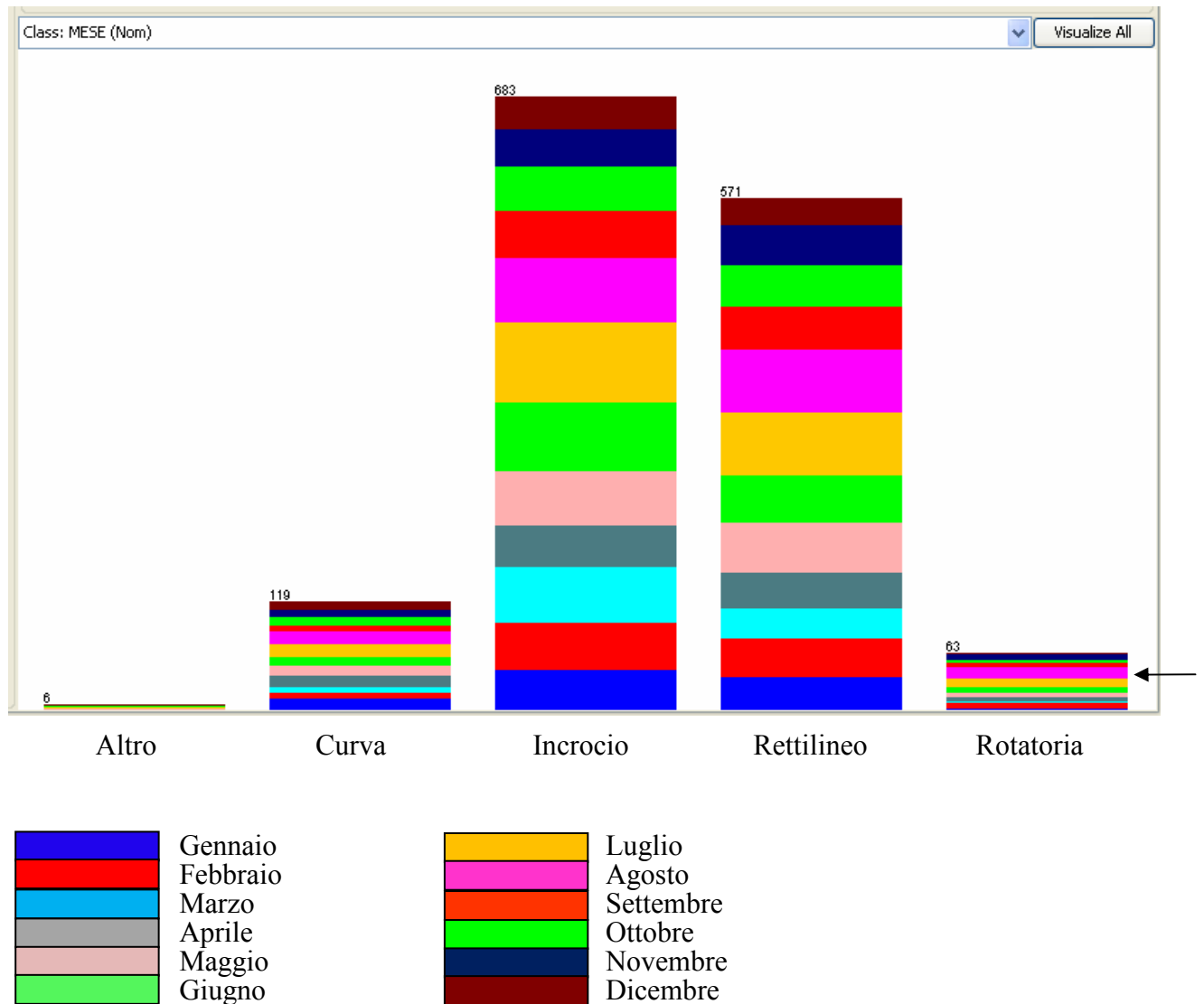
STATO DELLA STRADA DELL'INCIDENTE PER MESE



Commento:

Gli incidenti avvenuti in condizioni di strada bagnata denotano un **aumento nei mesi di Ottobre, Novembre, Dicembre e Gennaio** ed un alto valore anche nel mese di **Maggio**.

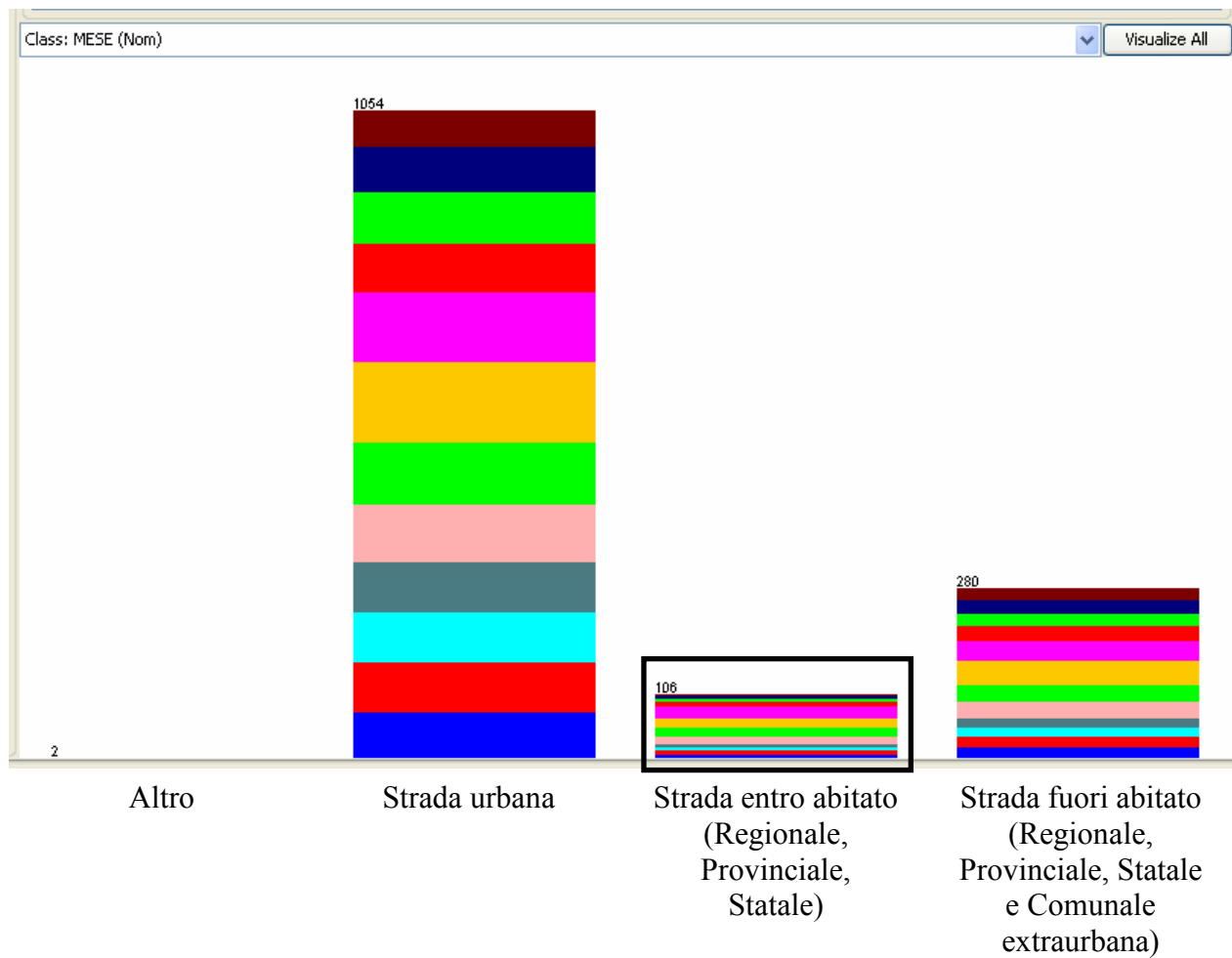
LUOGO DELL'INCIDENTE PER MESE DI ACCADIMENTO



Commento:

Gli incidenti avvenuti all'interno di **rotatorie** denotano un aumento nei mesi di **Luglio ed Agosto**.

TIPOLOGIA DI STRADA PER MESE DI ACCADIMENTO



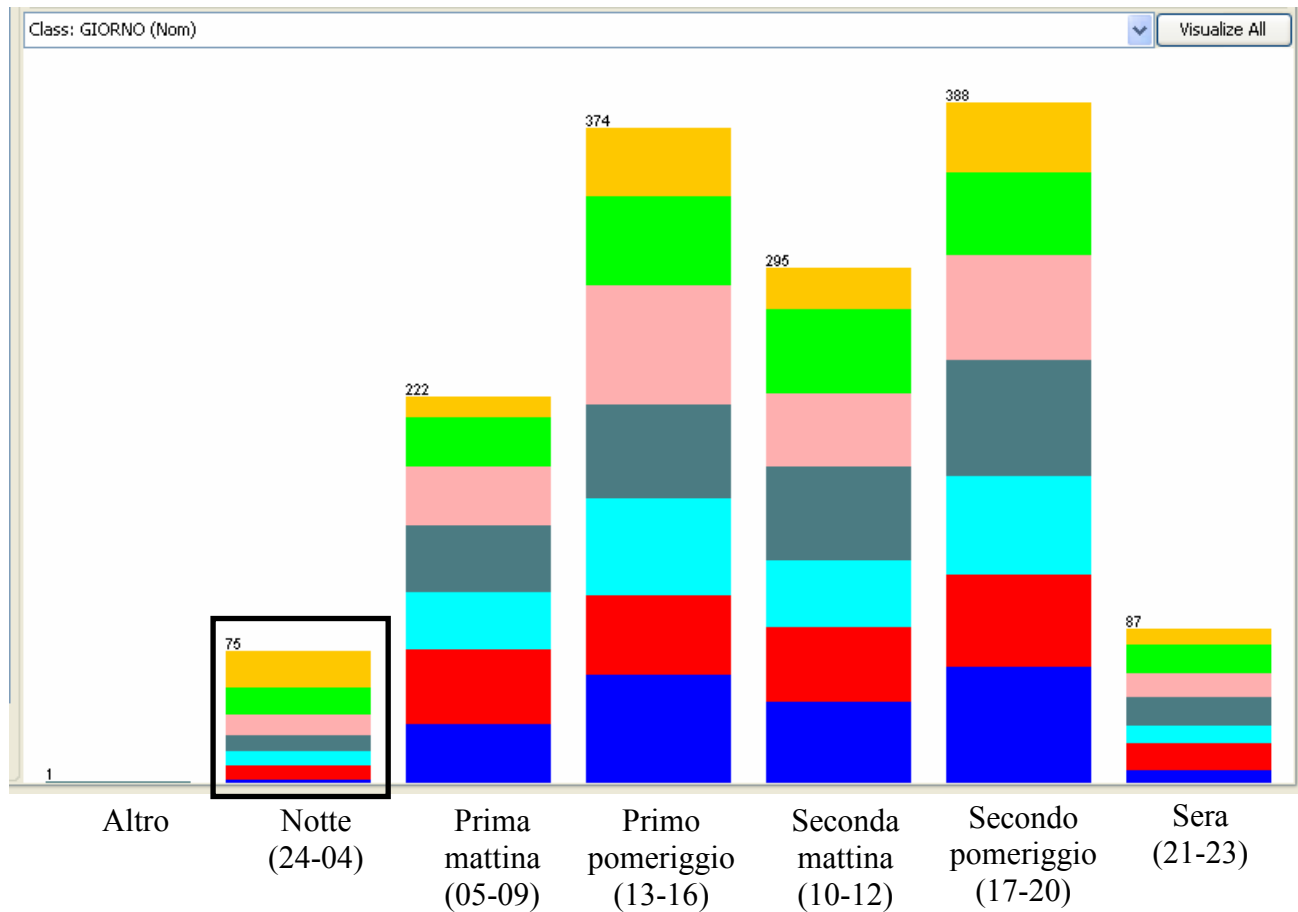
	Gennaio
	Febbraio
	Marzo
	Aprile
	Maggio
	Giugno

	Luglio
	Agosto
	Settembre
	Ottobre
	Novembre
	Dicembre

Commento:

Gli incidenti avvenuti nei mesi di **Maggio, Giugno, Luglio ed Agosto** coprono più di metà degli incidenti avvenuti in strade non urbane interne all'abitato.

FASCIA TEMPORALE GIORNALIERA DELL'INCIDENTE PER GIORNO

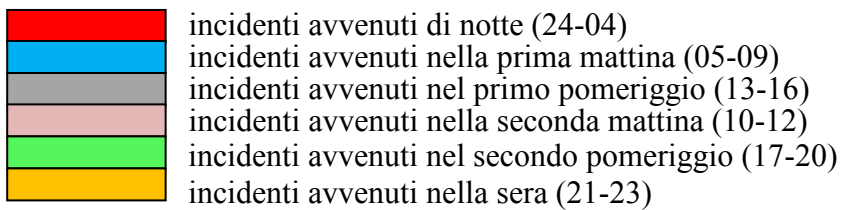
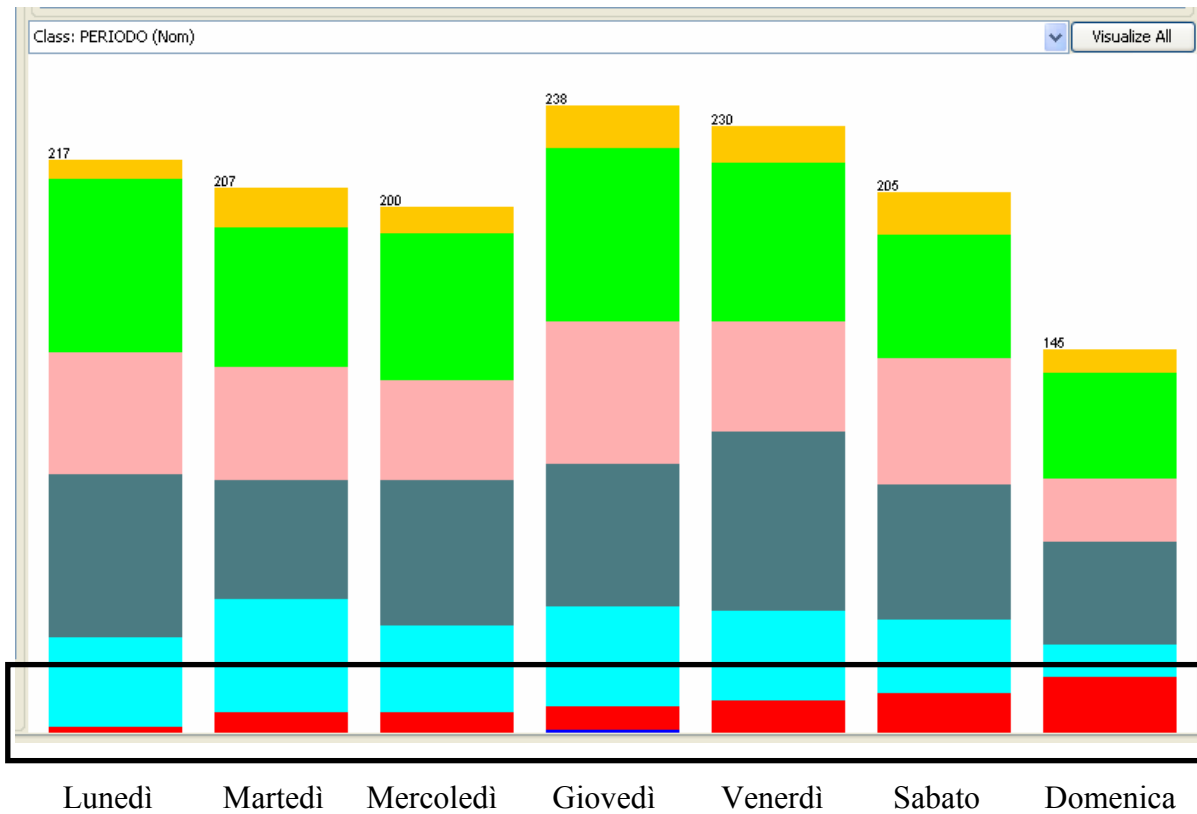


Commento:

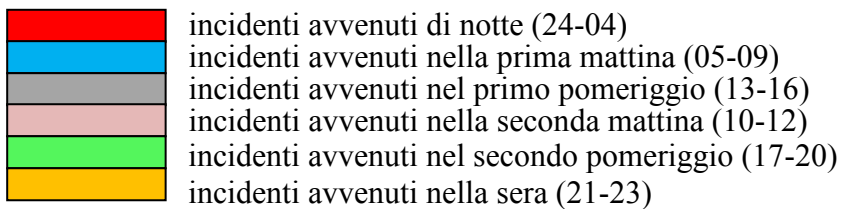
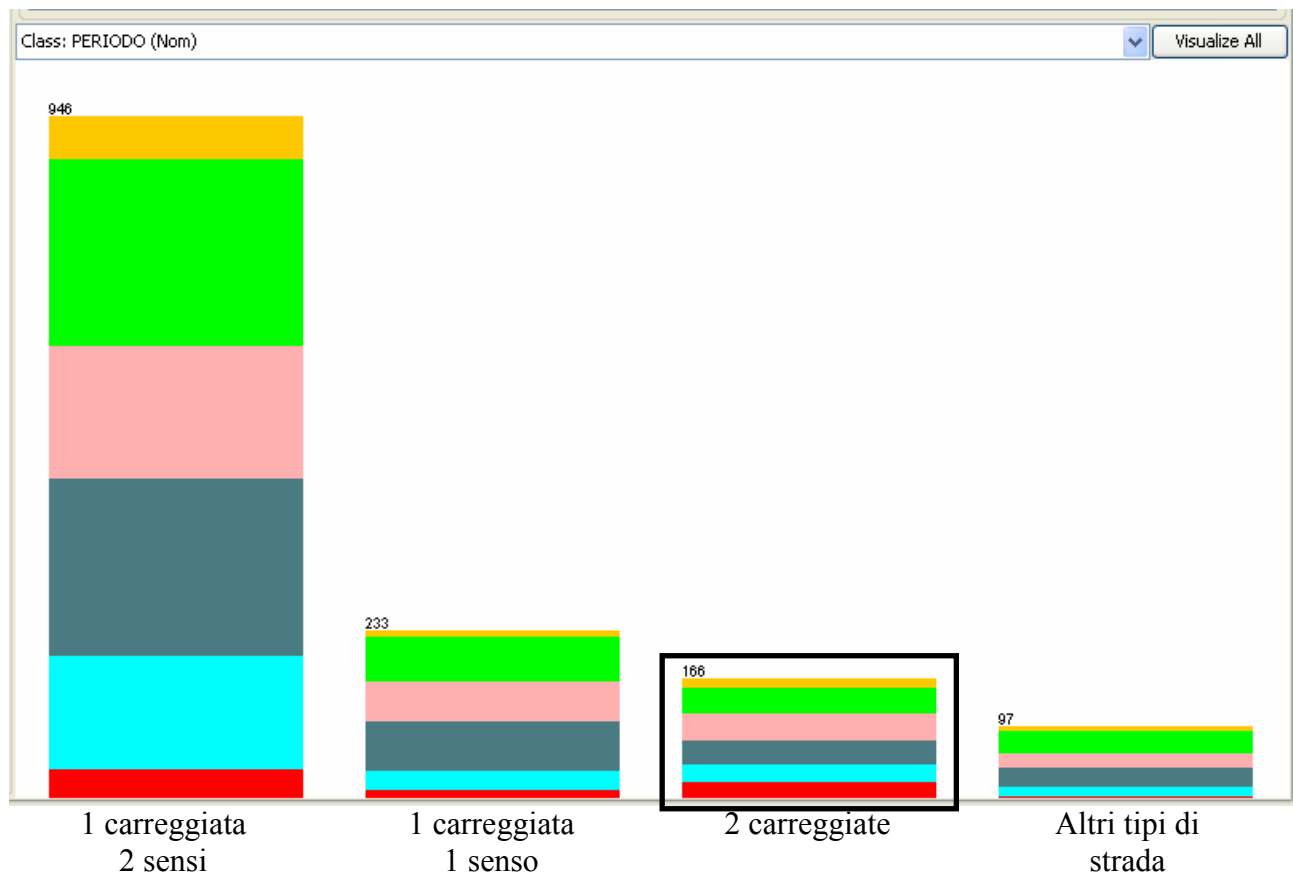
Gli **incidenti avvenuti di notte** rilevano un valore praticamente nullo al lunedì e si **incrementano progressivamente durante tutta la settimana** fino ad un valore massimo alla domenica.

Tale dato risulta ancora più evidente ribaltando l'ortogramma (vedi pagina seguente)

GIORNO DELL'INCIDENTE PER FASCIA TEMPORALE GIORNALIERA



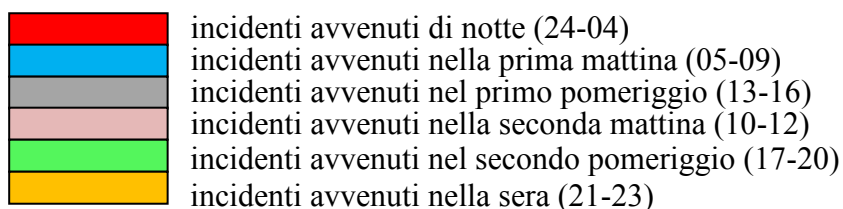
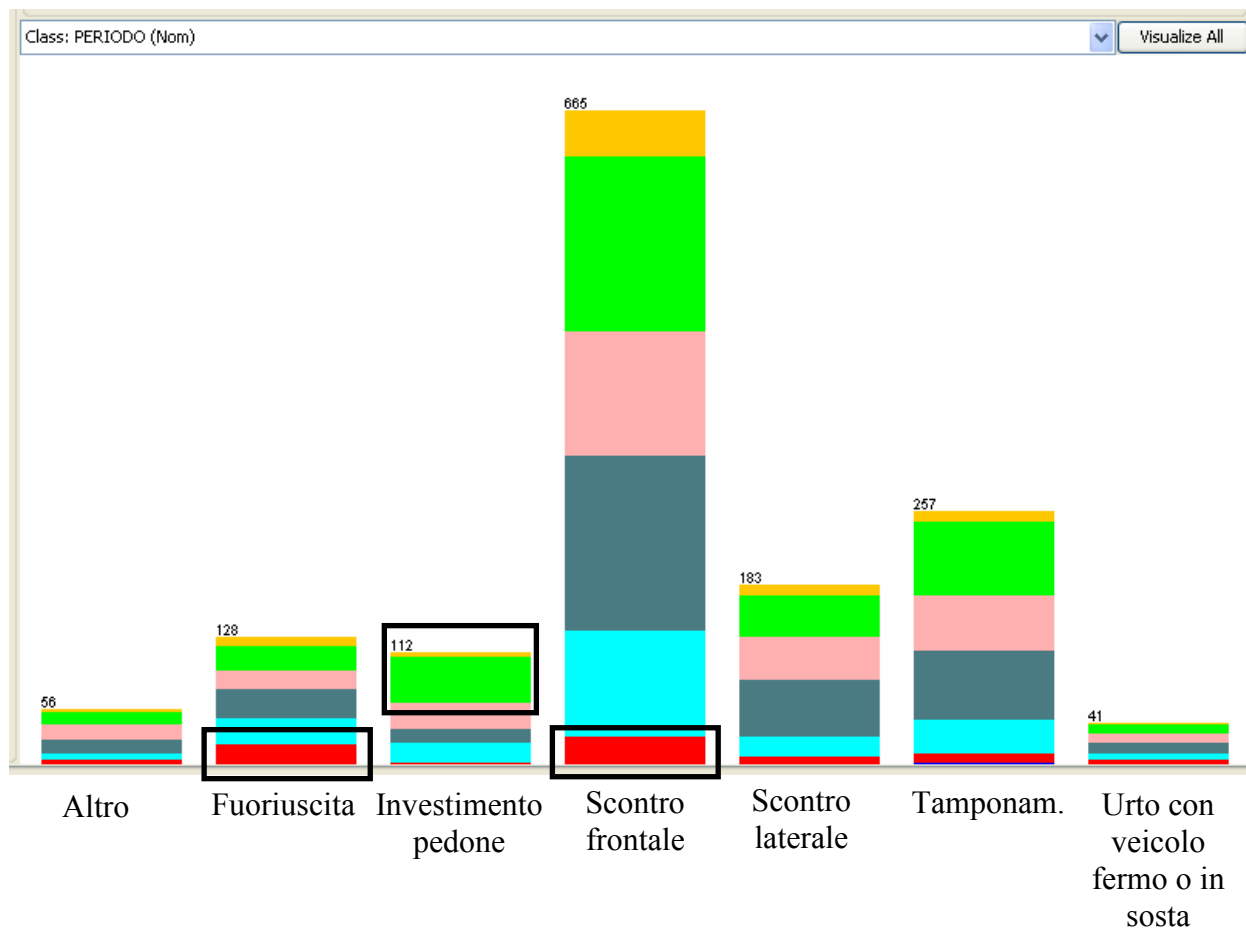
PERIODO QUOTIDIANO DELL'INCIDENTE PER TIPO DI STRADA



Commento:

Gli incidenti in strade a due carreggiate hanno una maggiore densità **nella notte e nella prima mattina.**

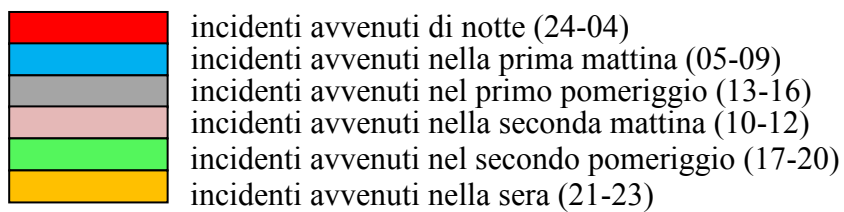
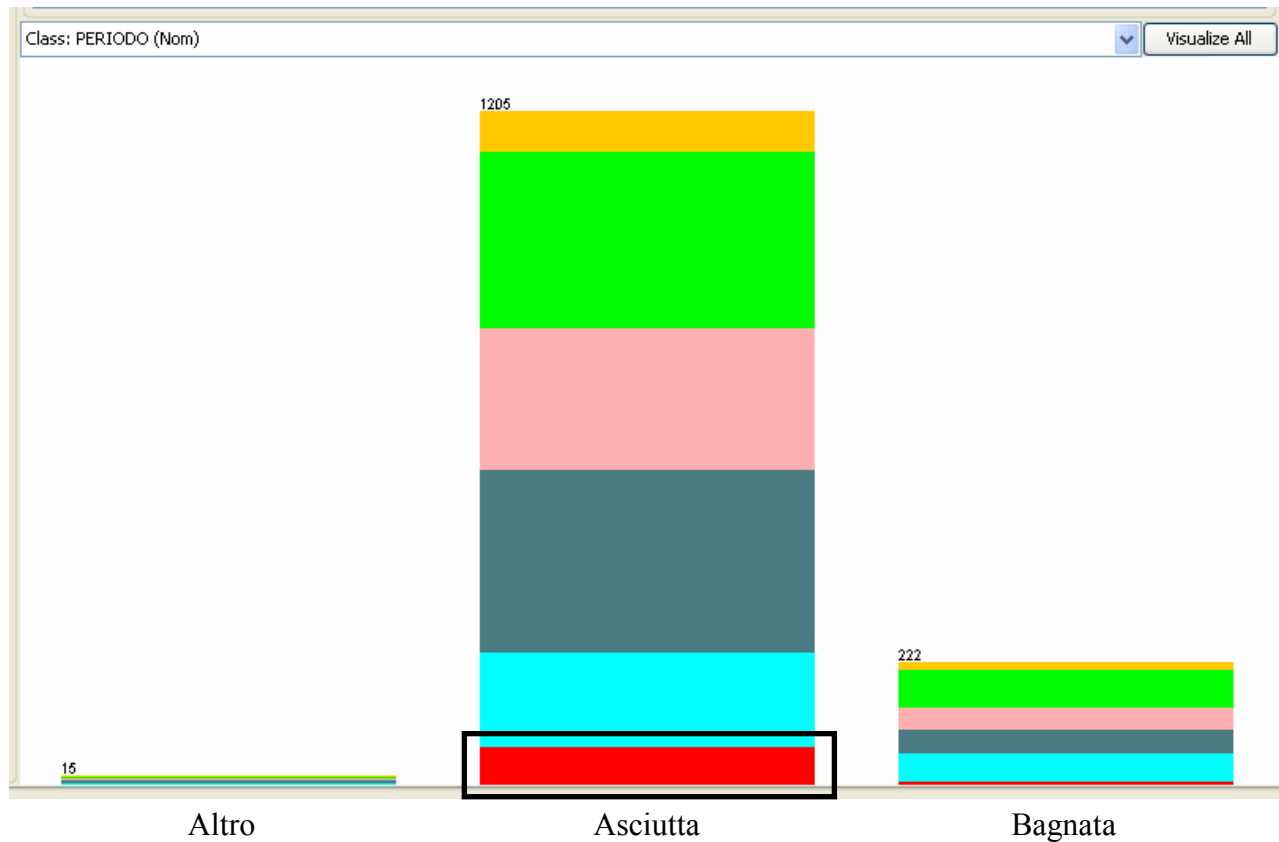
PERIODO QUOTIDIANO DELL'INCIDENTE PER GENERE



Commento:

Gli **incidenti notturni** sono dovuti principalmente a **fuoriuscite e scontri frontali** probabilmente in circostanze di stato personale 'alterato' o in presenza di colpi di sonno mentre gli investimenti di pedoni avvengono principalmente nel secondo pomeriggio, periodo di maggior flusso pedonale.

PERIODO QUOTIDIANO DELL'INCIDENTE PER STATO DELLA STRADA



Commento:

Gli incidenti notturni sono avvenuti quasi totalmente con strade asciutte.

ANALISI MULTIVARIATA

Visto l'alto numero di variabili e la diversità e casualità dei fenomeni coinvolti risulta difficile ipotizzare a priori un legame di qualche tipologia fra le variabili stesse (per esempio un legame lineare, esponenziale, poissoniano o altro) e quindi utilizzare regressioni ed analisi di correlazione multivariata di qualsiasi tipo (vedi figura seguente relativa al caso di due variabili di indipendenti o di input, X e Y e di una variabile target o dipendente Z) risulta soggettivo ed erroneo.

$Z = aX + bY + C$ (legame lineare)

$Z = aX^2 + bY^2 + C$ (legame quadratico)

$Z = a \log(X) + b \log(Y) + C$ (legame logaritmico)

$Z = aX + b \log(Y) + C$ (legame misto)

$Z = \dots\dots$



Quale relazione scegliere ??

Per quanto suddetto gli strumenti più idonei ad “estrarre conoscenza” da questa grande quantità di dati appartengono al campo di applicazioni del KDD- **Knowledge Discovery in Databases** o **Datamining**.

Nel Datamining è il computer che si occupa di trovare modelli per i dati in modo induttivo, identificando regole e caratteristiche che li legano senza una struttura/equazione di correlazione ipotizzata a priori, eliminando qualsiasi arbitrarietà e forzatura esterna.

Infatti, il **Datamining** è un processo atto a scoprire correlazioni, relazioni, tendenze nuove e significative, setacciando grandi quantità di dati immagazzinati nei Databases, usando tecniche di riconoscimento delle relazioni e tecniche statistiche e matematiche.

In pratica, si applica un processo di estrazione di



conoscenza da banche dati di grandi dimensioni tramite l'applicazione di algoritmi che individuano le associazioni "nascoste" tra le informazioni e le rendono visibili.

L'estrazione della conoscenza avviene tramite l'individuazione delle associazioni, o "patterns", o sequenze ripetute, o regolarità, nascoste nei dati.

In particolare vista l'analisi prettamente interpretativa del fenomeno dell'incidentalità (ovvero lontano da un fine classificatorio o modellistico di tipo previsionale) si è pensato di utilizzare lo strumento degli **Alberi Decisionali** (Decision Trees Induction) che fornisce risultati molto leggibili in quanto in forma di sentenze condizionali del tipo:

```
IF (Xa AND Yb AND Zc)
    THEN Ta
```

dove Xa, Yb, Zc sono le categorie delle variabili X, Y e Z per le quali si ha un'alta probabilità di avere la classe Ta della variabile T detta variabile target.

Per esempio, relativamente al fenomeno dell'incidentalità, potremmo investigare le variabili che influenzano il numero di feriti negli incidenti (variabile target) e quindi trovare regole del tipo:

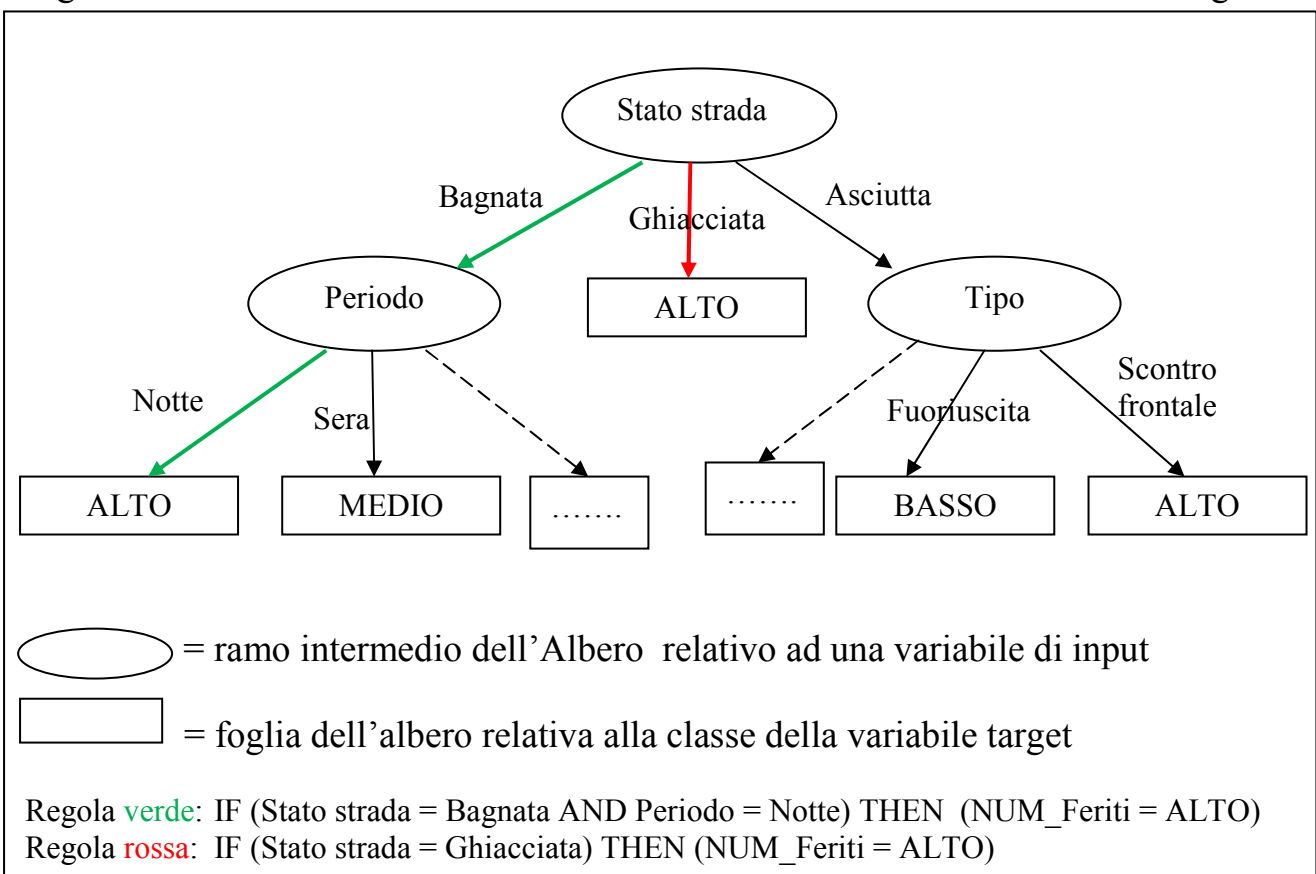
```
IF (Strada = Bagnata AND Periodo = Notte)
    THEN Numero_Feriti = Alto
```

Nel caso in esame è stato utilizzato il software open-source Weka, l'algoritmo C4.5 per la costruzione delle regole classificadorie (che utilizza il concetto di entropia per la costruzione dell'albero) e le variabili inserite nell'analisi sono le seguenti.

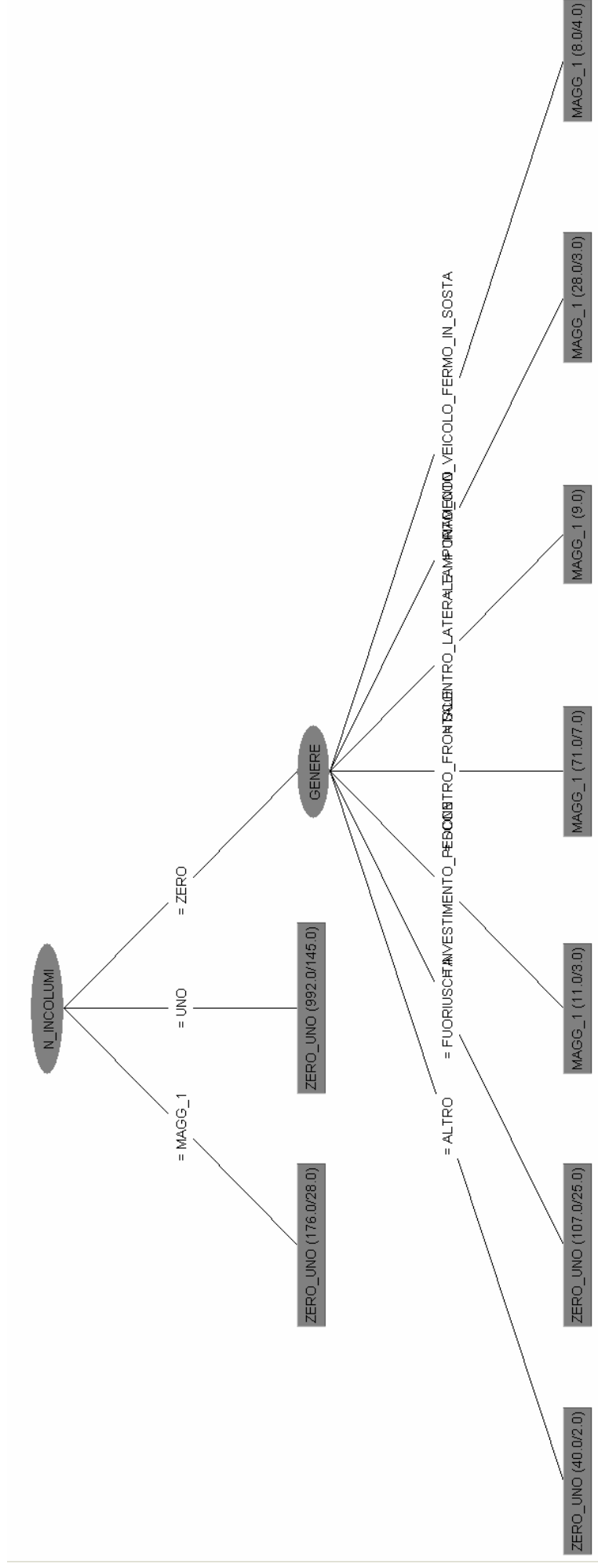
- Tipo di Strada
- Mese

- Giorno
- Periodo
- Genere;
- Condizioni stradali
- Stato segnaletica
- Numero di incolumi
- Numero di feriti
- Numero di morti
- Tipo di tracciato
- Localizzazione

Il grafico ad albero deve essere letto dall'alto verso il basso come indicato di seguito.



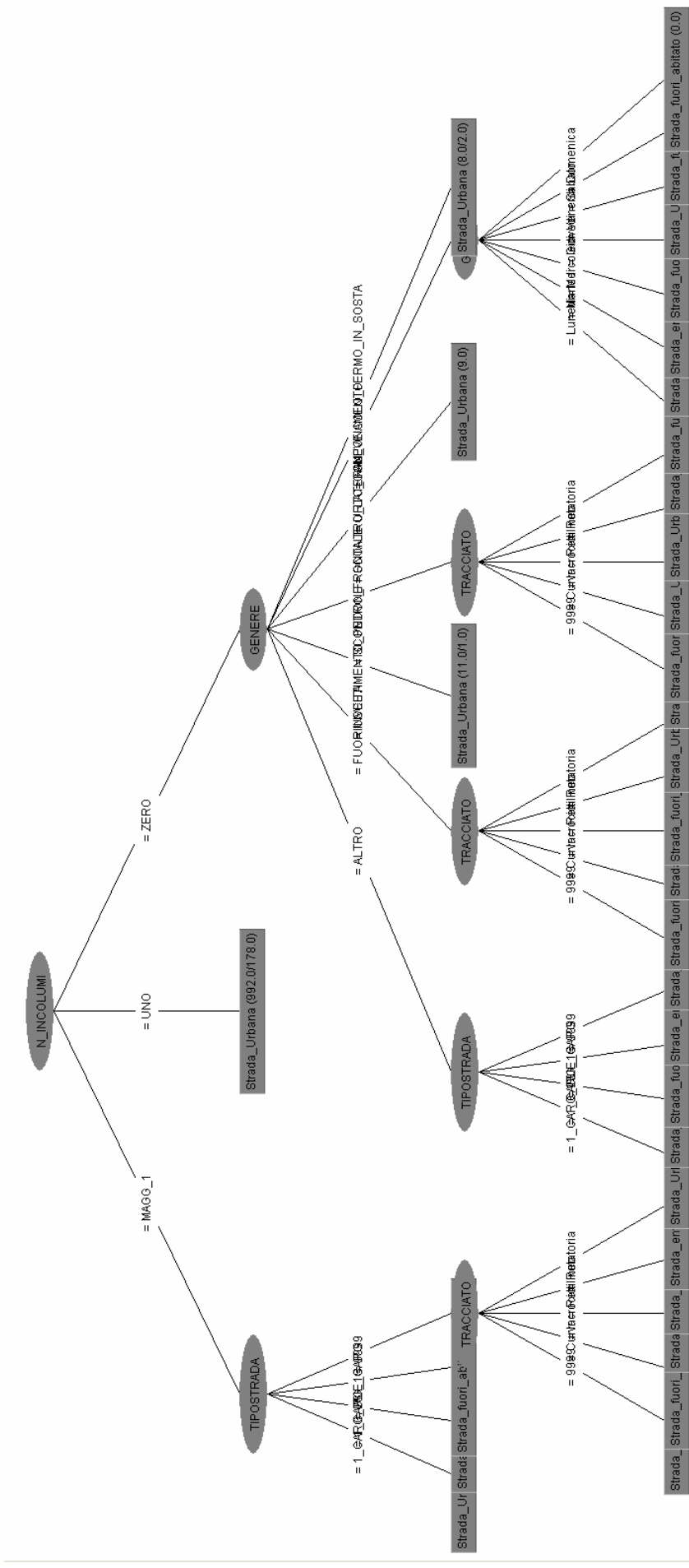
Nella figura seguente è rappresentato l'Albero Decisionale relativo alla variabile target "Numero di Feriti". Esso ha una percentuale di incidenti correttamente classificati dell'85%, ovvero che per 1225 su un totale di 1442 valgono le regole da esso estraibili.



Infatti sono subito estraibili due regole:

- 1) Se il numero di incolumi è maggiore di uno si ha probabilità dell'84% di avere uno o zero feriti (regola che vale per 176 incidenti)
- 2) Se il numero di incolumi è uno si ha probabilità dell'85% di avere uno o zero feriti (regola che vale per 992 incidenti)
- 3) Se non ci sono incolumi e l'incidente è per fuoriuscita o altri motivi, si ha una probabilità dell'81% di avere uno o zero feriti (accaduto per 147 incidenti) mentre se l'incidente è di altro genere si ha la probabilità del 87% di avere un numero di feriti superiore (accaduto per 127 incidenti).

Nella figura seguente è rappresentato l'Albero Decisionale relativo alla variabile target "Localizzazione incidente".



Esso ha una percentuale di incidenti correttamente classificati dell'77%, ovvero che per 1110 su un totale di 1442 valgono le regole da esso estraibili e riportate di seguito.

Dalle regole risulta chiaro che nel caso di un **numero di incolumi maggiore o uguale ad uno si ha un'alta probabilità (81%) di trovarsi in una strada urbana**, dove gli incidenti hanno quindi danni alle persone di gravità inferiore (regola avvenuta per 1133 incidenti).

Gli incidenti senza incolumi, ovvero con una **gravità maggiore**, avvengono principalmente in **quattro situazioni**:

- strada urbana ad una carreggiata a due sensi di marcia;
- strada fuori dall'abitato ed in curva con fuoriuscita dalla carreggiata;
- strada fuori dall'abitato ed in rettilineo con fuoriuscita dalla carreggiata;
- incrocio stradale in strada urbana con scontro frontale.

Relativamente a quanto estratto da questi due primi capitoli risulterà particolarmente utile verificare nei prossimi anni il ripetersi delle condizioni evidenziate ed il confronto multi temporale dei dati stessi.