

UNIVERSITÀ POLITECNICA DELLE MARCHE



Progetto Business Intelligence

Analisi e Visualizzazione degli Incidenti Stradali nel Regno Unito (2005-2016)

AUTORI

Davide Ticchiarelli - S1121687

Giampaolo Marino - S1121678

Andrea Pizzuto - S1122741

Anno accademico 2024/2025

Indice

1	Introduzione	1
1.1	Dataset	1
1.2	Pulizia dei dati	5
2	Qlik	7
2.1	Introduzione	7
2.2	Caricamento dati	7
2.3	Data analysis	9
2.3.1	Foglio 1 - Analisi di incidenti e vittime in relazione ai distretti	9
2.3.2	Foglio 2 - Analisi di incidenti e vittime in relazione a gravità, tipologie di veicoli, fasce d'età e zone di provenienza	13
2.3.3	Foglio 3 - Analisi di incidenti e numero di vittime in relazione a sesso del conducente e altri parametri	16
3	Tableau	23
3.1	Introduzione	23
3.2	Caricamento dati	23
3.2.1	Relazione dei dataset	24
3.3	Data analysis	24
3.3.1	Dashboard 1 - Analisi temporale di incidenti per anno, giorno e ora . . .	24
3.3.2	Dashboard 2 - Distribuzione degli incidenti per tipologia di veicolo e carburante	30
3.3.3	Dashboard 3 - Analisi degli incidenti in base a caratteristiche stradali . .	32
3.3.4	Dashboard 4 - Previsione del numero di vittime e severità media degli incidenti	34
4	Power BI	39
4.1	Introduzione	39
4.2	Caricamento dati	39
4.2.1	Relazione dei dataset	40
4.3	Data analysis	40
4.3.1	Foglio 1 - Analisi degli incidenti in base alle condizioni ambientali e temporali	41
4.3.2	Foglio 2 - Confronto degli incidenti tra automobili e motociclette	45
4.3.3	Utilizzo DAX e M	47
4.3.4	Foglio 3 - Analisi delle dinamiche degli incidenti	49
4.3.5	Foglio 4 - Predizione del numero di incidenti in base alla gravità	54
	Elenco delle figure	56

Elenco delle tabelle

58

1 Introduzione

In questa tesina saranno presentate diverse analisi effettuate su un dataset riguardante gli incidenti avvenuti nel Regno Unito nel periodo compreso tra il 2005 e il 2017. Le analisi saranno realizzate mediante l'utilizzo di tre software per l'analisi dei dati: Qlik, Tableau e Power BI.

1.1 Dataset

Il dataset considerato contiene dati statistici sugli incidenti stradali avvenuti nel Regno Unito nel periodo tra il 2005 e il 2017. Questi dati sono stati raccolti direttamente dal governo e successivamente pubblicati online sulle piattaforme ufficiali. Tuttavia, il dataset utilizzato è stato reperito dal seguente link: <https://www.kaggle.com/datasets/tsiaras/uk-road-safety-accidents-and-vehicles/data>, anziché dai canali ufficiali, poiché è stato elaborato per essere più leggibile e comprensibile.

Inoltre, il dataset si compone di due file CSV distinti, collegati dall'attributo *Accident_index*. I dataset in questione sono:

- *Accident_Information.csv*: in questa raccolta sono presenti tutte le informazioni generali su ogni incidente rilevato nel periodo compreso tra il 2005 e il 2017.
- *Vehicle_Information.csv*: in quest'altra raccolta, invece, sono riportate le informazioni sui passeggeri e sui veicoli coinvolti negli incidenti nel periodo compreso tra il 2004 e il 2016.

Tra le ragioni che hanno determinato la scelta di questo dataset, rispetto ad altri, vi sono le seguenti:

- *Quantità di dati*: un fattore molto importante perché un maggior numero di righe spesso aumenta la qualità delle analisi eseguite, poiché consente di avere un campione più ampio e diversificato, aumentando la probabilità che i dati riflettano accuratamente la realtà. In questo caso, il CSV degli incidenti contiene poco più di due milioni di righe, mentre il CSV sui passeggeri e sui veicoli coinvolti poco meno di 1,5 milioni di righe.
- *Numero di colonne*: un fattore molto importante poiché un maggior numero di colonne consente di essere più selettivi su quali colonne tenere e quali rimuovere; inoltre, consente un maggior numero di analisi rispetto a dataset con un numero di colonne ristretto. In questo caso, il CSV degli incidenti contiene 34 colonne, mentre il CSV sui passeggeri e sui veicoli coinvolti contiene 24 colonne.
- *Usabilità*: una valutazione calcolata direttamente da Kaggle su parametri come completezza, credibilità e compatibilità. L'usabilità di un dataset è importante perché un valore elevato indica che il dataset è stato progettato in modo da essere facilmente utilizzabile e,

quindi, ridurre il tempo per eseguire le operazioni di preprocessing dei dati. In questo caso, l'usabilità è pari a 8.24.

Inizialmente era stato selezionato un altro dataset relativo agli incidenti nel Regno Unito. Tuttavia, durante l'utilizzo, è emerso che il dataset presentava problemi di leggibilità e conteneva numerosi dati errati. Per questi motivi, è stata adottata la decisione di optare per un secondo dataset, che, pur mantenendo le stesse colonne, offre una maggiore leggibilità e include un file CSV aggiuntivo con informazioni sui passeggeri e sui veicoli coinvolti.

Di seguito, nelle Tabelle 1 e 2 sono riportate rispettivamente le colonne che compongono i due dataset.

Attributo	Descrizione
Accident_Index	Identificatore univoco per ogni incidente
1st_Road_Class	Classe della prima strada
1st_Road_Number	Numero della prima strada
2nd_Road_Class	Classe della seconda strada
2nd_Road_Number	Numero della seconda strada
Accident_Severity	Severità dell'incidente
Carriageway_Hazards	Pericoli sulla carreggiata
Date	Data dell'incidente
Day_of_Week	Giorno della settimana dell'incidente
Did_Police_Officer_Attend_Scene_of_Accident	Se un ufficiale di polizia ha partecipato alla scena dell'incidente
Junction_Control	Tipo di controllo dell'incrocio
Junction_Detail	Dettaglio sull'incrocio
Latitude	Latitudine dell'incidente
Light_Conditions	Condizioni di luce al momento dell'incidente
Local_Authority_(District)	Autorità locale (distretto)
Local_Authority_(Highway)	Autorità locale (strada).
Location_Easting_OSGR	Coordinate OSGR Est
Location_Northing_OSGR	Coordinate OSGR Nord
Longitude	Longitudine dell'incidente.
LSOA_of_Accident_Location	LSOA della posizione dell'incidente
Number_of_Casualties	Numero di vittime
Number_of_Vehicles	Numero di veicoli coinvolti
Pedestrian_Crossing-Human_Control	Passaggio pedonale - controllo umano
Pedestrian_Crossing-Physical_Facilities	Passaggio pedonale - strutture fisiche
Police_Force	Forza di polizia che ha gestito l'incidente
Road_Surface_Conditions	Condizioni della superficie stradale
Road_Type	Tipo di strada
Special_Conditions_at_Site	Condizioni speciali sul sito dell'incidente
Speed_limit	Limite di velocità (mph)
Time	Orario dell'incidente
Urban_or_Rural_Area	Area urbana o rurale
Weather_Conditions	Condizioni meteo al momento dell'incidente
Year	Anno dell'incidente
InScotland	Se l'incidente è avvenuto in Scozia

Tabella 1: Colonne del dataset *Accident_Information.csv*

Attributo	Descrizione
Accident_Index	Identificatore univoco per ogni incidente
Age_Band_of_Driver	Fascia di età del conducente
Age_of_Vehicle	Età del veicolo coinvolto
Driver_Home_Area_Type	Tipo di area di residenza del conducente
Driver_IMD_Decile	Decile dell'Indice di Deprivazione Multipla (IMD) del Conducente
Engine_Capacity_CC.	Capacità del motore (in CC)
Hit_Object_in_Carriageway	Oggetto colpito nella carreggiata dal veicolo
Hit_Object_off_Carriageway	Oggetto colpito fuori dalla carreggiata dal veicolo
Journey_Purpose_of_Driver	Scopo del viaggio del conducente
Junction_Location	Posizione dell'incrocio
Make	Marca del veicolo
Model	Modello del veicolo
Propulsion_Code	Codice di propulsione del veicolo
Sex_of_Driver	Sesso del conducente
Skidding_and_Overturning	Se il veicolo ha slittato o si è ribaltato
Towing_and_Articulation	Se il veicolo è a traino o articolato
Vehicle_Leaving_Carriageway	Come il veicolo ha lasciato la carreggiata
Vehicle_Location.Restricted_Lane	Se il veicolo si trovava in una corsia limitata
Vehicle_Manoeuvre	Manovra del veicolo al momento dell'incidente
Vehicle_Reference	Riferimento univoco per ogni veicolo coinvolto
Vehicle_Type	Tipo di veicolo coinvolto
Was_Vehicle_Left_Hand_Drive	Se il veicolo era a guida a sinistra
X1st_Point_of_Impact	Primo punto di impatto del veicolo
Year	Anno del veicolo

Tabella 2: Colonne del dataset *Vehicle_Information.csv*

1.2 Pulizia dei dati

Una volta individuato il dataset, è stato necessario eseguire un primo processo di pulizia in modo da garantire che non ci siano incongruenze e/o valori mancanti, assicurando così risultati più precisi. Tale processo di pulizia è stato eseguito utilizzando la libreria Pandas in Python e si articola nelle seguenti fasi:

- **Caricamento dei file CSV:** In questa fase, sono stati caricati i file CSV.

```
1     import pandas as pd
2
3     # Caricamento dei file CSV
4     accident = pd.read_csv(r"..\\UK_Accident.csv")
5     vehicles = pd.read_csv(r"..\\UK_Accident_Vehicles.csv")
```

- **Rimozione delle colonne non necessarie:** Dopo il caricamento dei file, sono state rimosse alcune colonne ritenute poco utili ai fini dell'analisi. Ad esempio, sono stati eliminati il numero e la classe delle strade, il campo Year (già implicitamente presente nella colonna Date), le coordinate OSGR ed altre colonne.

```
1     # Rimozione delle colonne non necessarie nel dataset degli incidenti
2     accident = accident.drop(columns=[
3         "Carriageway_Hazards",
4         "1st_Road_Class", "1st_Road_Number",
5         "2nd_Road_Class", "2nd_Road_Number",
6         "Did_Police_Officer_Attend_Scene_of_Accident",
7         "Special_Conditions_at_Site",
8         "Location_Easting_OSGR", "Location_Northing_OSGR",
9         "LSOA_of_Accident_Location", "Year", "InScotland"
10    ])
11
12     # Rimozione delle colonne non necessarie nel dataset dei veicoli
13     vehicles = vehicles.drop(columns=[
14         "Driver_IMD_Decile",
15         "Hit_Object_in_Carriageway",
16         "Hit_Object_off_Carriageway",
17         "Journey_Purpose_of_Driver",
18         "Vehicle_Location.Restricted_Lane",
19         "Vehicle_Reference",
20         "Was_Vehicle_Left_Hand_Drive"
21    ])
```


- **Filtraggio dei record:** In questa fase, poiché i due dataset non coprivano esattamente lo stesso periodo temporale, per ottenere analisi più coerenti è stato deciso di mantenere soltanto le righe corrispondenti in entrambi i dataset. Grazie a questa operazione, inoltre, è stato possibile ridurre la dimensione dei dataset, permettendo un'elaborazione più agile e senza rallentamenti dovuti al loro grande peso.

```
1      # Rimuovi gli incidenti che non hanno un veicolo associato
2      accident_filtered = accident[accident["Accident_Index"]
3                                  .isin(vehicles["Accident_Index"])]
4
5      # Rimuovi i veicoli che non hanno un incidente associato
6      vehicles_filtered = vehicles[vehicles["Accident_Index"]
7                                  .isin(accident_filtered["Accident_Index"])]
```

- **Salvataggio dei nuovi file CSV:** Una volta eseguite le operazioni di pulizia e filtraggio, i dataset risultanti sono stati salvati in nuovi file CSV per facilitare le analisi successive.

```
1      # Salva i nuovi file CSV
2      accident_filtered.to_csv(r"..\\UK_Accident_ETL.csv", index=False)
3      vehicles_filtered.to_csv(r"..\\UK_Accident_Vehicles_ETL.csv", index=False)
```

Per quanto riguarda la gestione dei dati, abbiamo valutato che quelli a nostra disposizione fossero già sufficientemente validi e completi, decidendo quindi di procedere con la pulizia solo quando strettamente necessario. Inoltre, sebbene fosse possibile eseguire il JOIN dei due dataset in un unico file tramite Python, abbiamo scelto di non farlo per evitare di appesantire ulteriormente il file CSV. In entrambi i casi, la pulizia e la gestione delle JOIN saranno eseguite, quando necessario, tramite le funzionalità offerte da Qlik, Tableau e PowerBi, per garantire un'elaborazione più efficiente e mirata.

2 Qlik

2.1 Introduzione

Qlik è una piattaforma di business intelligence (BI) che offre strumenti per l'analisi dei dati e la visualizzazione interattiva. Fondata nel 1993, l'azienda è conosciuta per il suo approccio innovativo, che permette agli utenti di esplorare, scoprire e condividere informazioni in modo intuitivo. La soluzione principale, Qlik Sense, consente di creare dashboard e report interattivi in tempo reale, con un forte focus sull'auto-servizio e sull'interattività. Grazie alla tecnologia di associazione in-memory, Qlik permette risposte rapide anche su set di dati complessi. In aggiunta, Qlik View è un'altra soluzione più orientata a utenti avanzati e sviluppatori, che permette una personalizzazione più approfondita delle dashboard e report, ideali per scenari BI complessi. Entrambe le piattaforme supportano l'integrazione con strumenti avanzati come R e Python per l'analisi predittiva, offrendo flessibilità nella gestione dei dati e nella creazione di applicazioni BI su misura.



Figura 1: Logo di Qlik

2.2 Caricamento dati

Qlik semplifica il caricamento e l'analisi dei dati grazie a strumenti ETL avanzati, che raccolgono e organizzano informazioni da diverse fonti. La sua architettura in-memory garantisce accesso rapido ai dati, permettendo analisi immediate. Inoltre, la piattaforma consente trasformazioni per uniformare le informazioni e si integra facilmente con svariate fonti, supportando decisioni basate su dati aggiornati e completi. Innanzitutto per procedere alle analisi e alle successive attività di ETL (Extract, Transform and Load) si è proceduto con il caricamento dei dati da una sorgente in locale, essendo Qlik Sense una piattaforma online. I file sorgenti sono due: `UK_Accident.csv` e `UK_Accident_Vehicles.csv`. Una volta caricati i dataset si è proceduto con alcune operazioni di trasformazione. Per il successivo svolgimento di analisi geografiche e per sfruttare al meglio le informazioni in merito alle località, è stato modificato il tipo di dato individuato da Qlik per la colonna chiamata `Local_Authority_(Highway)` da *'generico dato'* a *'dato geografico'* come si può osservare nella Figura 2.

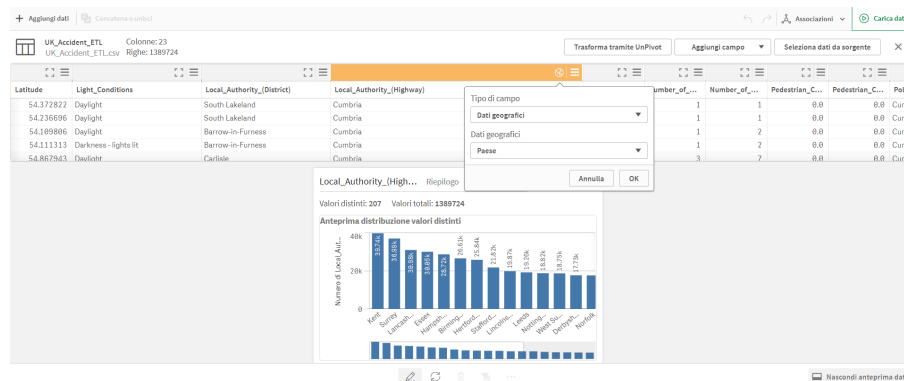


Figura 2: Trasformazione tipo di dato per la colonna Local_Authority_(Highway)

In seguito sono stati modificati i valori della colonna **Accident_Severity** che originariamente conteneva tre valori di tipo stringa, ovvero Slight, Serious e Fatal, successivamente sostituiti rispettivamente con i valori numerici 1, 2 e 3, al fine di poter svolgere delle analisi statistiche utilizzando questi parametri (Figura 3).

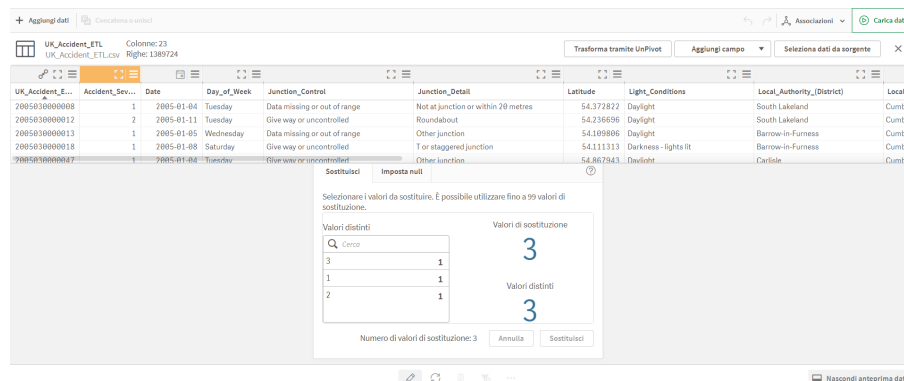


Figura 3: Sostituzione dei valori di Accident_Severity

Infine, sono stati associati i due dataset sulla base del parametro **Accident_Index** al fine di riuscire a collegare e completare le informazioni relative agli incidenti con le rispettive relative alle persone e ai veicoli coinvolti in essi (Figura 4).

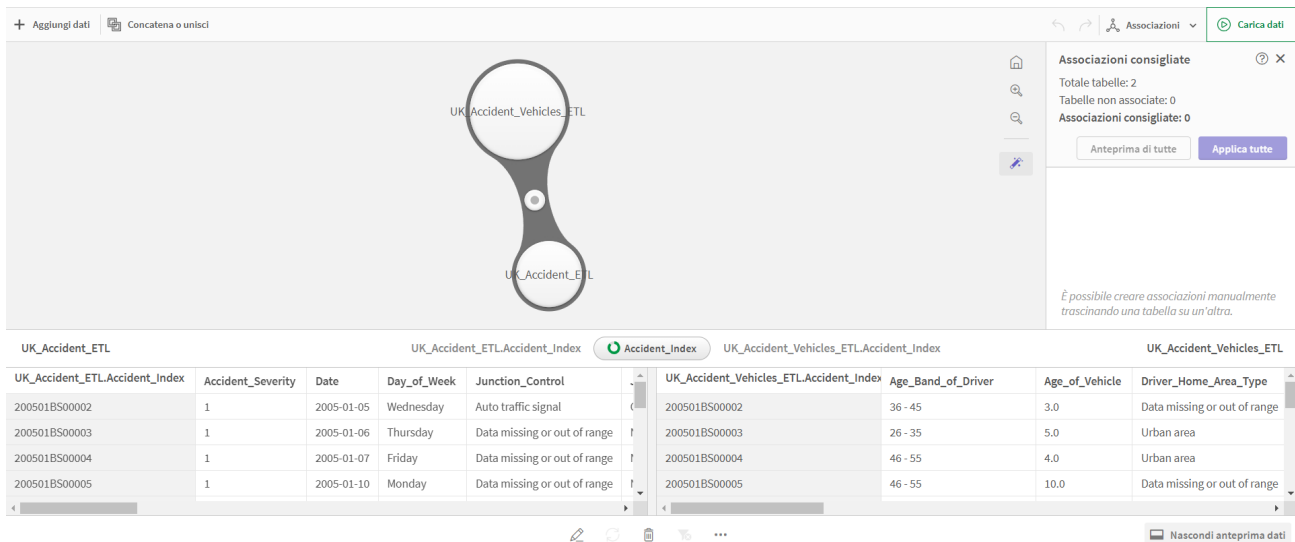


Figura 4: Collegamento dei dataset

Con quest'ultima operazione si è conclusa la fase di caricamento dei dati e si può procedere con la definizione delle analisi svolte e le considerazioni svolte in merito.

2.3 Data analysis

Per eseguire le analisi su Qlik, si utilizzano i fogli di lavoro, ognuno dei quali può ospitare diversi oggetti visivi. Ogni foglio consente di esplorare un determinato aspetto del lavoro complessivo svolto, di seguito sono elencati:

- *Foglio 1*: Analisi di incidenti e vittime in relazione ai distretti.
- *Foglio 2*: Analisi di incidenti e vittime in relazione a gravità, tipologie di veicoli, fasce d'età e zone di provenienza.
- *Foglio 3*: Analisi di incidenti e numero di vittime in relazione a sesso del conducente e altri parametri

2.3.1 Foglio 1 - Analisi di incidenti e vittime in relazione ai distretti

Nel secondo foglio di analisi (Figura 5) si possono osservare le analisi svolte in merito agli incidenti e al numero di vittime in relazione ai distretti del Regno Unito.

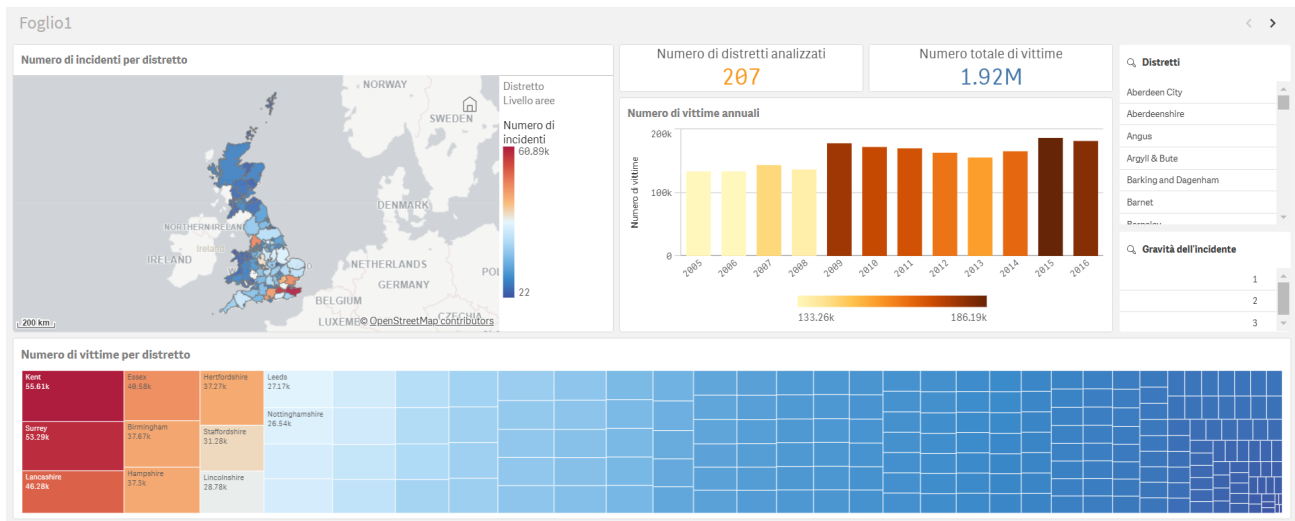


Figura 5: Foglio di analisi numero 1

In questo foglio sono presenti cinque oggetti visivi e due filtri:

- *Mappa ad aree*: descrive la relazione tra gli incidenti commessi e i distretti del Regno Unito.
- *Heatmap*: descrive la relazione tra il numero di vittime e i distretti del Regno Unito.
- *Grafico a barre verticali*: descrive l'andamento annuale del numero di vittime nel Regno Unito.
- *Indicatore di parametro*: indica il valore totale di vittime.
- *Indicatore di parametro*: indica il numero totale di distretti analizzati.
- *Filtro distretti*: filtra gli oggetti visivi in base a uno o più distretti selezionati.
- *Filtro gravità*: filtra gli oggetti visivi in base a uno o più livelli di gravità degli incidenti selezionati.

Innanzitutto, è possibile analizzare la *mappa ad aree*, che rappresenta il numero di incidenti per ciascun distretto attraverso una scala di colori che va dal blu (minore numero di incidenti) al rosso scuro (maggiore numero di incidenti). Dal grafico emerge chiaramente che i distretti con il maggior numero di incidenti sono "Kent" (circa 61.000 incidenti), "Surrey" (circa 57.500 incidenti) e "Lancashire" (circa 46.000 incidenti). Questa distribuzione può essere attribuita alla maggiore densità abitativa e al traffico più intenso in questi distretti, che aumentano la probabilità di sinistri stradali. Al contrario, distretti come "Gwynedd" o "Isle of Anglesey" registrano un numero significativamente inferiore di incidenti. Ciò potrebbe essere dovuto alla

minore popolazione e, di conseguenza, a un volume di traffico ridotto rispetto ai centri più urbanizzati. Questo suggerisce una possibile correlazione tra densità di popolazione, traffico e incidenza degli incidenti stradali.

Successivamente, è possibile analizzare il *grafico a barre verticali*, che rappresenta l'andamento del numero di vittime nel corso degli anni. Ogni barra corrisponde a un anno, dal 2005 al 2016, ed è colorata con un gradiente di rosso in base al numero di vittime registrate in quell'anno. Dal grafico si osserva che gli anni con il maggior numero di vittime sono stati il *2015* (circa 186.000 vittime), il *2016* (circa 181.000 vittime) e il *2009* (quasi 178.000 vittime). Questi picchi potrebbero essere legati a diversi fattori, tra cui condizioni meteorologiche particolarmente avverse che hanno reso la circolazione più pericolosa, un aumento del traffico dovuto a eventi economici o sociali, oppure variazioni nella regolamentazione stradale o nei controlli sulla sicurezza. Inoltre, è possibile che l'evoluzione delle infrastrutture e la diffusione di nuove tecnologie nei veicoli abbiano avuto un impatto differente sulla sicurezza stradale nel corso degli anni.

Infine, è possibile osservare la *heatmap* che rappresenta il numero di vittime per distretto. Ogni porzione del grafico corrisponde a un distretto e le sue caratteristiche visive, come dimensione e colore, sono determinate dal numero totale di vittime registrate in quella specifica area. Dal grafico si evince che i distretti con il maggior numero di vittime sono *Kent* (55.610 vittime), *Surrey* (53.300) e *Lancashire* (46.300), mentre quelli con il numero più basso risultano essere *Orkney Island* (262 vittime) e *Shetland Island* (405 vittime). Tuttavia, analizzando il rapporto tra il numero di incidenti e il numero di vittime, emergono dati interessanti: il *Kent*, pur essendo il distretto con il maggior numero di incidenti e vittime, presenta un rapporto vittime/incidenti pari a 0.91, mentre in *Orkney Island*, dove si registra un numero inferiore di incidenti, il rapporto è più alto, pari a 1.03. Questa differenza potrebbe essere legata a diversi fattori. Nelle aree più urbanizzate, come il Kent, la presenza di infrastrutture più efficienti e di un pronto intervento tempestivo potrebbe ridurre la letalità degli incidenti, mentre nelle zone più remote, come Orkney Island, i tempi di soccorso potrebbero essere più lunghi, aumentando la gravità delle conseguenze. Anche la tipologia degli incidenti può incidere su questa disparità: nelle aree urbane gli incidenti potrebbero essere più numerosi ma meno gravi, come tamponamenti o scontri a bassa velocità, mentre nelle zone rurali, dove le strade sono spesso più pericolose e le velocità più elevate, il numero di vittime per incidente tende a essere maggiore. Anche la densità del traffico gioca un ruolo fondamentale, poiché in aree più congestionate la velocità media è inferiore, riducendo l'impatto degli scontri e, di conseguenza, la loro letalità. Questa analisi suggerisce quindi che un numero maggiore di incidenti non implica necessariamente un numero proporzionale di vittime, evidenziando come fattori infrastrutturali, ambientali e logistici possano influenzare in modo significativo la gravità degli incidenti stradali.

Di seguito si può osservare cosa accade applicando il filtro per gravità degli incidenti nel seguente foglio (Figure 6, 7 e 8).

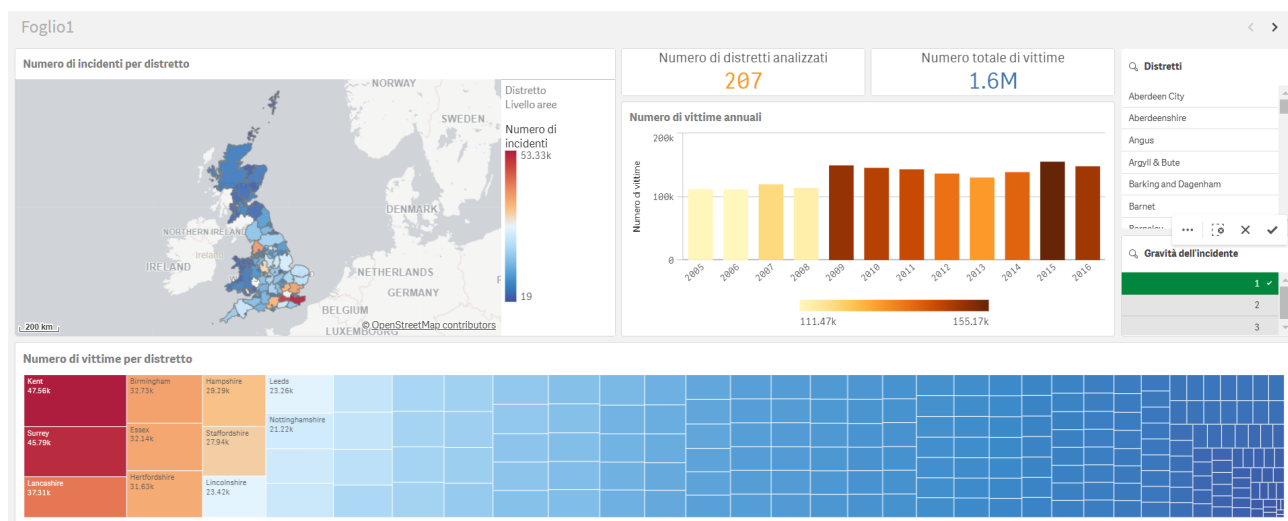


Figura 6: Foglio di analisi numero 1 con filtro per gravità

Nella Figura 6 si osserva che, applicando il filtro di gravità 1 (lieve), si nota che i valori restano pressoché invariati. Questo indica che la maggior parte degli incidenti rientra in questa categoria, evidenziando fortunatamente una prevalenza di sinistri di lieve entità.

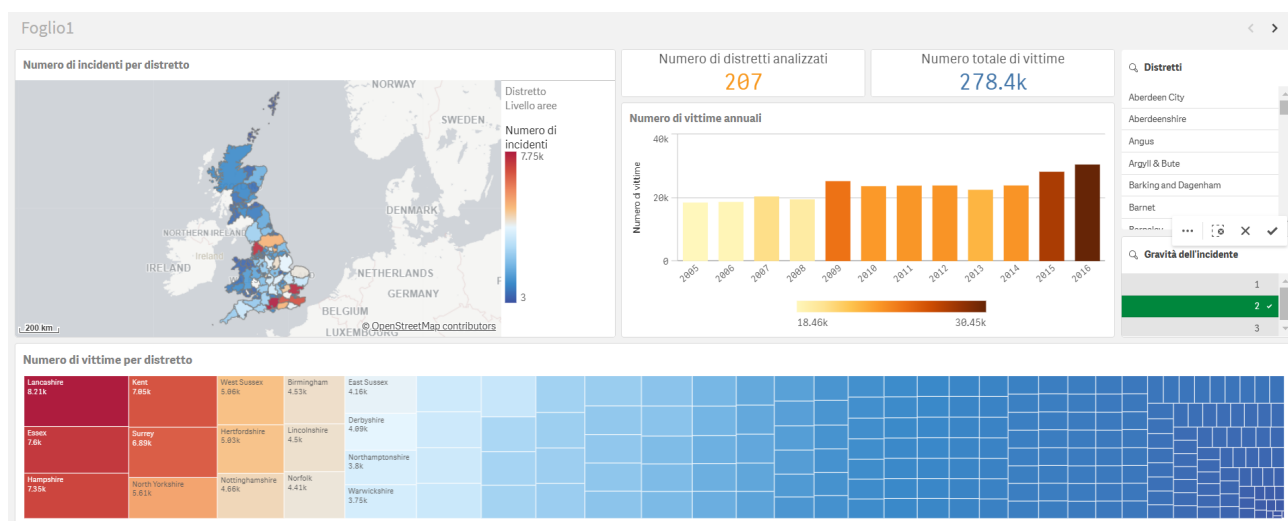


Figura 7: Foglio di analisi numero 1 con filtro per gravità

Nella Figura 7, applicando il filtro di gravità 2 (media), si osservano alcuni cambiamenti significativi. Ad esempio, sulla mappa, la distribuzione degli incidenti varia, come nel caso del distretto di "Lancashire". In questo caso, l'andamento annuale rimane pressoché proporzionato a quello del totale, mentre sulla heatmap si evidenzia che Lancashire è il distretto con il maggior

numero di vittime. Inoltre, nella mappa ad aree, si posiziona al terzo posto con circa 7.380 incidenti, simile a distretti come "Essex" (7.510 incidenti) e "Hampshire" (7.750 incidenti).

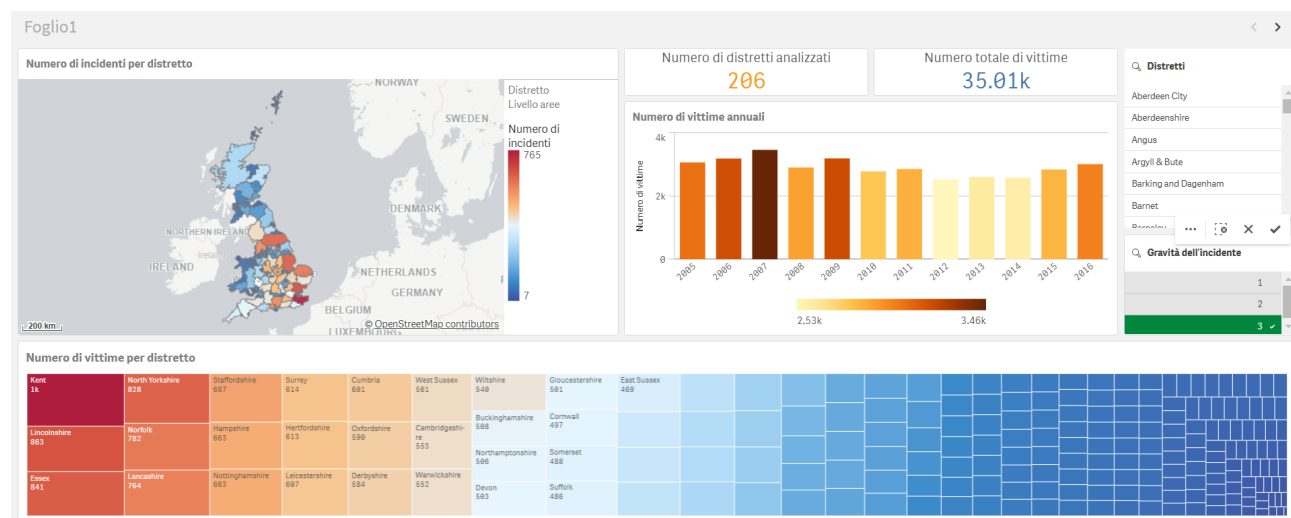


Figura 8: Foglio di analisi numero 1 con filtro per gravità

Nella Figura 8, applicando il filtro di gravità 3 (fatale), si osservano significativi cambiamenti nella distribuzione degli incidenti. In particolare, gli incidenti fatali non sono concentrati esclusivamente nelle aree urbane, ma si verificano anche in zone meno densamente popolate. Il distretto di "Kent" rimane al primo posto con 765 incidenti fatali, seguito da "Essex" con 676 incidenti. Nella heatmap, il Kent si conferma come il distretto con il numero più alto di vittime, circa 1.000, seguito da Lincolnshire con 863 vittime e Essex con 841. Un dato interessante riguarda il distretto "London Airport", che registra solo 6 vittime nonostante l'intenso traffico. Questo potrebbe essere attribuito a una combinazione di fattori, tra cui una maggiore presenza di controlli sulla sicurezza e infrastrutture stradali più sicure intorno agli aeroporti.

2.3.2 Foglio 2 - Analisi di incidenti e vittime in relazione a gravità, tipologie di veicoli, fasce d'età e zone di provenienza

Nel secondo foglio di analisi (Figura 9) si possono osservare le analisi svolte in merito agli incidenti, la loro gravità e la relazione che esiste tra questi due parametri e le fasce d'età dei guidatori e le tipologie di veicolo coinvolte.



In seguito, è possibile analizzare il *grafico a barre verticali* situato al di sotto del grafico a torta. Questo oggetto visivo mostra la relazione tra il numero di incidenti e le fasce d'età dei guidatori, fornendo una panoramica sulla distribuzione dei sinistri in base all'età del conducente. Si osserva chiaramente che la maggior parte degli incidenti è stata commessa da guidatori di età compresa tra i 26 e i 35 anni, con un totale di circa 388.800 incidenti. Seguono i conducenti di età compresa tra i 36 e i 45 anni, con circa 377.460 incidenti, e successivamente quelli nella fascia 46-55 anni, con un totale di circa 308.870 sinistri. Di particolare rilevanza è la netta differenza rispetto ai conducenti minorenni (0-15 anni), che fortunatamente hanno registrato un numero di incidenti molto basso, pari a circa 4.000.

Successivamente, è possibile analizzare il *grafico a linee* che rappresenta la gravità media degli incidenti in relazione alle fasce d'età dei conducenti e alla tipologia di veicolo. Questo oggetto visivo prende in considerazione le categorie di veicoli più comunemente utilizzate, ovvero: "Car", "Motorcycle 50cc and under", "Motorcycle 125cc and under", "Motorcycle over 125cc up to 500cc", "Motorcycle over 500cc" e "Pedal cycle". Analizzando l'andamento generale dei trend relativi alla gravità media degli incidenti, si nota che per tutte e sei le categorie vi è un incremento con l'aumentare dell'età dei conducenti. Questo fenomeno potrebbe essere attribuito a un naturale declino dei riflessi e delle capacità motorie con l'avanzare dell'età. Un'osservazione particolarmente rilevante riguarda il comportamento dei trend nelle fasce d'età più basse. In particolare, si registra la presenza di alcuni sinistri con conducenti di automobili di età compresa tra 0 e 10 anni. Questo potrebbe essere dovuto a errori nei dati oppure a casi eccezionali di incidenti con guidatori così giovani. Un altro aspetto significativo riguarda le categorie di veicoli a due ruote: gli incidenti iniziano a essere registrati a partire dalla fascia d'età "11-15", probabilmente perché è in questo periodo che si comincia generalmente ad approcciarsi a questo tipo di mezzi. Gli *spike* osservati nei trend potrebbero essere dovuti all'inesperienza dei giovani alla guida. Anche in questo caso, la presenza di incidenti in fasce d'età inaspettate potrebbe essere spiegata da errori nel dataset o dal fatto che nella categoria "Car" possano essere incluse vetture di piccola cilindrata, destinate anche alla guida dei minori.

In seguito, è possibile osservare il *grafico Mekko*, situato al di sotto del grafico a linee, che descrive la relazione tra le fasce d'età dei conducenti, la zona degli incidenti e il numero massimo di vittime coinvolte. L'ampiezza delle colonne, rappresentative delle fasce d'età, è determinata dal numero massimo totale di vittime registrate per ciascuna fascia nelle due zone considerate: "Urban" e "Rural". Si nota che le fasce d'età maggiormente evidenziate sono quelle comprese tra i 16 e i 75 anni, con un numero massimo di vittime che varia tra 70 e 93 nella zona rurale e tra 19 e 63 nella zona urbana. Confrontando questi dati con il *grafico a torta* precedentemente analizzato, emerge un possibile collegamento con i limiti medi di velocità delle diverse aree. In particolare, nelle zone rurali, dove i limiti di velocità tendono a essere più elevati rispetto alle aree urbane, il numero massimo di vittime per incidente risulta generalmente più alto. Questo suggerisce una possibile correlazione tra velocità e gravità degli incidenti: all'aumentare del limite di velocità medio, e quindi della velocità a cui i veicoli viaggiano, cresce anche il rischio di incidenti con un maggior numero di vittime. Ciò potrebbe essere dovuto a diversi fattori, tra

cui il maggiore impatto derivante da velocità elevate, minori possibilità di frenata tempestiva e la più bassa densità di infrastrutture di sicurezza nelle aree rurali rispetto a quelle urbane.

2.3.3 Foglio 3 - Analisi di incidenti e numero di vittime in relazione a sesso del conducente e altri parametri

Nel terzo foglio di analisi (*Figura 13*) si possono osservare le analisi svolte in merito agli incidenti e vittime coinvolte in relazione ad altri parametri come fasce d'età, condizioni di luce, condizioni della strada tutti in relazione al sesso del guidatore.

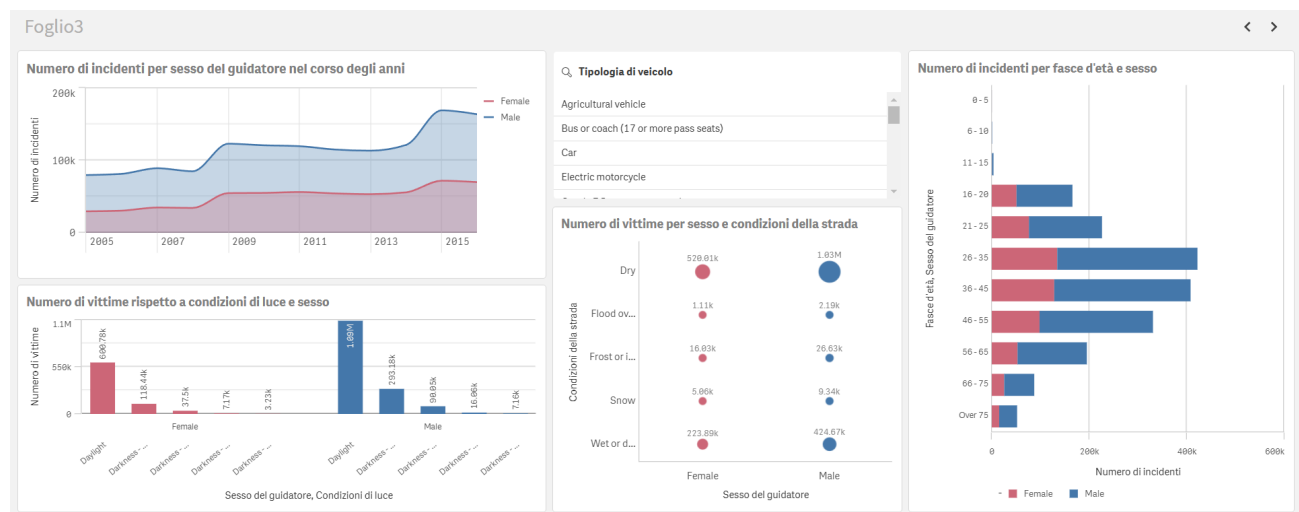


Figura 10: Foglio di analisi numero 3

Innanzitutto possiamo osservare che questo foglio è composto di 4 oggetti visivi ed un filtro per tipologia di veicolo:

- *Grafico ad aree*: descrive l'andamento temporale del numero di incidenti in base al sesso.
- *Grafico a barre raggruppate*: descrive il numero di vittime per sesso in relazione alle condizioni di luce.
- *Grafico a bolle*: descrive il numero di vittime per sesso in relazione alle condizioni stradali.
- *Grafico a barre orizzontali in pila*: descrive il numero di incidenti per fasce d'età in relazione al sesso del conducente.

In alto a sinistra è possibile osservare un *grafico ad aree* che rappresenta il numero totale di incidenti commessi ogni anno da guidatori di sesso maschile (in blu) e femminile (in rosa). Il grafico mette in evidenza una marcata differenza tra le due categorie, mostrando come il

numero di incidenti annuali causati dagli uomini sia circa il doppio rispetto a quello delle donne. Analizzando l'andamento temporale, si possono individuare due picchi significativi: il primo nel 2009, con circa 123.170 incidenti per gli uomini e 55.000 per le donne; il secondo nel 2015, con un aumento rispettivamente a circa 170.000 e 72.000. Questi "spike" indicano periodi in cui si è verificato un incremento anomalo del numero di incidenti, suggerendo possibili correlazioni con fattori esterni come cambiamenti nelle normative stradali, variazioni nel traffico o condizioni socio-economiche. Questa disparità potrebbe essere dovuta a diversi fattori. Ad esempio gli uomini, tendono ad adottare uno stile di guida più aggressivo e propenso al rischio, con una maggiore incidenza di eccessi di velocità e manovre azzardate. Al contrario, le donne mostrano generalmente un comportamento più prudente alla guida, riducendo la probabilità di incidenti.

In basso a sinistra è possibile osservare un *grafico a barre raggruppate*, in cui ogni barra rappresenta una specifica condizione di luce, mentre la sua altezza indica il numero di vittime. Il grafico distingue due gruppi: uno per i conducenti di sesso femminile e uno per quelli di sesso maschile. Dall'analisi del grafico emerge che il numero di vittime tra le donne è circa la metà rispetto agli uomini, in linea con la proporzionalità già osservata nel numero totale di incidenti. Inoltre, si nota che la maggior parte degli incidenti con vittime si verifica in condizioni di luce diurna ("*Daylight*"), con valori che si attestano intorno ai 600.000 per le donne e 1.100.000 per gli uomini. Le altre condizioni di luce, ovvero "*Darkness light lit*" (buio con illuminazione artificiale) e "*Darkness no lighting*" (buio senza illuminazione), presentano un numero inferiore di vittime. Questo potrebbe essere dovuto a una minore esposizione al rischio nelle ore notturne, poiché la quantità complessiva di veicoli in circolazione è inferiore rispetto al giorno. Tuttavia, la gravità degli incidenti notturni potrebbe essere maggiore a causa di una ridotta visibilità, della stanchezza dei conducenti e di una possibile maggiore incidenza di guida sotto l'effetto di alcol o sostanze stupefacenti.

Successivamente, è possibile analizzare il *grafico a bolle* situato al centro della pagina, il quale rappresenta il numero di vittime in base al sesso del conducente e alle condizioni stradali. Analogamente all'oggetto visivo descritto in precedenza, anche in questo caso il numero di vittime è proporzionale al numero di incidenti, ma si possono osservare variazioni significative in relazione al parametro "*Condizioni della strada*". Si nota che la maggior parte delle vittime si registra in condizioni di strada asciutta ("*Dry*"), con valori pari a circa 520.000 per le donne e 1.000.000 per gli uomini. Questo è prevedibile, poiché la maggior parte del traffico si verifica in condizioni normali di viabilità, rendendo più frequenti gli incidenti su strade asciutte. Il secondo scenario più rilevante è quello delle strade bagnate o umide ("*Wet or damp*"), in cui si registrano circa 224.000 vittime tra le donne e 425.000 tra gli uomini. La presenza di acqua sull'asfalto può ridurre l'aderenza dei veicoli e aumentare il rischio di aquaplaning, contribuendo all'incremento degli incidenti rispetto alle condizioni di guida ideali. Gli altri tre casi considerati, ovvero "*Frost or ice*", "*Flood over 3cm deep*" e "*Snow*", presentano un numero significativamente inferiore di vittime. Ciò è plausibile, poiché tali condizioni meteorologiche avverse si verificano con minore frequenza rispetto a pioggia e asfalto asciutto. Tuttavia, quando presenti, possono avere un impatto notevole sulla sicurezza stradale. Approfondendo l'analisi attraverso il grafico ad aree,

si possono osservare alcune tendenze specifiche:

- Nel caso di *"Frost or ice"* (Figura 11), si evidenziano due picchi, uno nel 2010 e uno nel 2012. Questo potrebbe suggerire la presenza di inverni particolarmente rigidi in quegli anni, caratterizzati da forti gelate o nevicate che hanno reso le strade estremamente scivolose e pericolose.

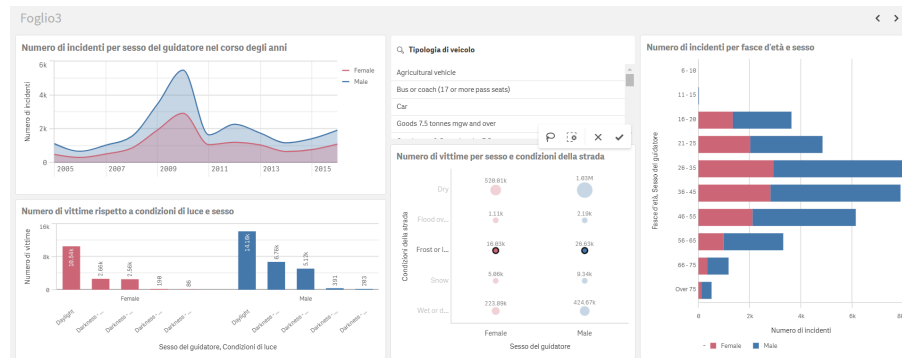


Figura 11: Foglio di analisi numero 3 - Condizioni della strada "frost or ice"

- Nel caso di *"Flood over 3cm deep"* (Figura 12), sono stati registrati tre picchi nel 2009, 2012 e 2014. Ciò potrebbe indicare che in questi anni si sono verificate piogge eccezionalmente abbondanti o eventi alluvionali, aumentando il rischio di incidenti a causa di strade sommerse o rese impraticabili dall'acqua.

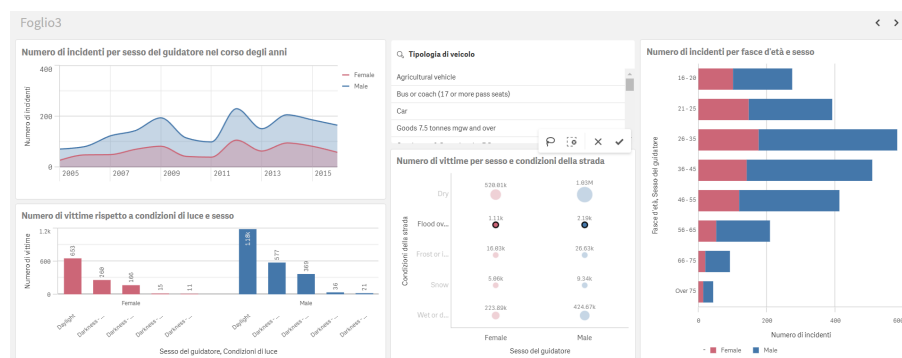


Figura 12: Foglio di analisi numero 3 - Condizioni della strada "flood"

- Per quanto riguarda *"Snow"* (Figura 13), si notano degli aumenti significativi nel numero di vittime nel 2010 e nel 2013. Questo potrebbe essere dovuto a stagioni invernali particolarmente nevose, durante le quali l'accumulo di neve sulle strade ha reso la guida più difficile, aumentando il numero di sinistri.

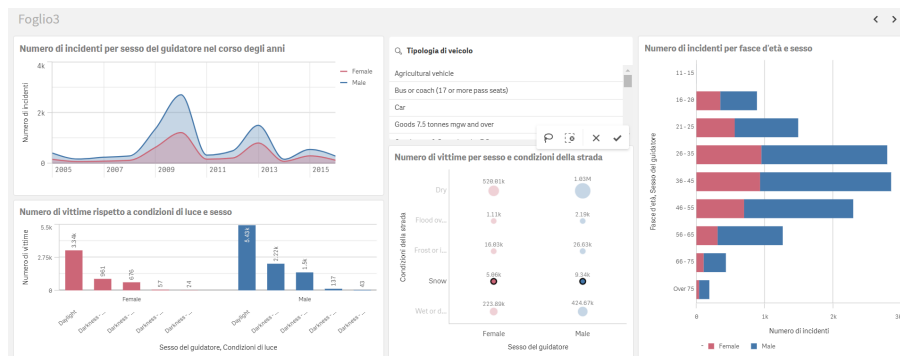


Figura 13: Foglio di analisi numero 3 - Condizioni della strada "snow"

Queste osservazioni suggeriscono che, sebbene le condizioni meteorologiche estreme siano meno frequenti, quando si verificano possono avere un impatto significativo sulla sicurezza stradale. La presenza di picchi nei dati potrebbe riflettere l'influenza del clima sulle abitudini di guida e sull'efficacia delle misure di sicurezza adottate in quegli anni.

Infine, è possibile analizzare l'ultimo grafico, ovvero il *grafico a barre in pila orizzontali*, situato sulla sinistra del foglio. Questo oggetto visivo rappresenta il numero di incidenti in relazione alle fasce d'età dei conducenti e al loro sesso. Ogni barra corrisponde a una specifica fascia d'età ed è suddivisa al suo interno in base al sesso del conducente. L'ampiezza complessiva della barra è proporzionale al numero totale di incidenti registrati per quella fascia. Anche in questo caso, si osserva chiaramente che le donne hanno un'incidenza significativamente inferiore rispetto agli uomini nel numero totale di incidenti. Un aspetto rilevante evidenziato dal grafico è che la maggior parte degli incidenti è concentrata nelle fasce d'età "26-35", "36-45" e "46-55". Si nota inoltre un aumento significativo del numero di incidenti rispetto alla fascia "21-25", con un incremento che risulta circa il doppio sia per gli uomini che per le donne. Questo "gap" potrebbe essere spiegato dal fatto che la fascia "21-25" comprende in gran parte giovani studenti, che utilizzano meno frequentemente l'auto rispetto alle fasce di età successive. Al contrario, la fascia "26-35" e le successive includono principalmente persone inserite nel mondo del lavoro, che fanno un uso più intenso del mezzo privato per spostamenti quotidiani, aumentando così la loro esposizione al rischio di incidenti. Inoltre, con l'aumentare dell'età, possono incidere anche altri fattori, come la maggiore esperienza alla guida (che potrebbe ridurre il numero di incidenti) e l'aumento del tempo trascorso su strada (che invece ne accresce la probabilità).

All'interno di questo foglio è stato inserito anche un filtro per tipologia di veicolo al fine di effettuare delle analisi che possano mostrare altri "insight" nascosti nei dati. Filtrando i dati includendo soltanto quelli relativi a tutte le tipologie di motocicli, possiamo osservare il foglio cambiare nel seguente modo (Figura 14).

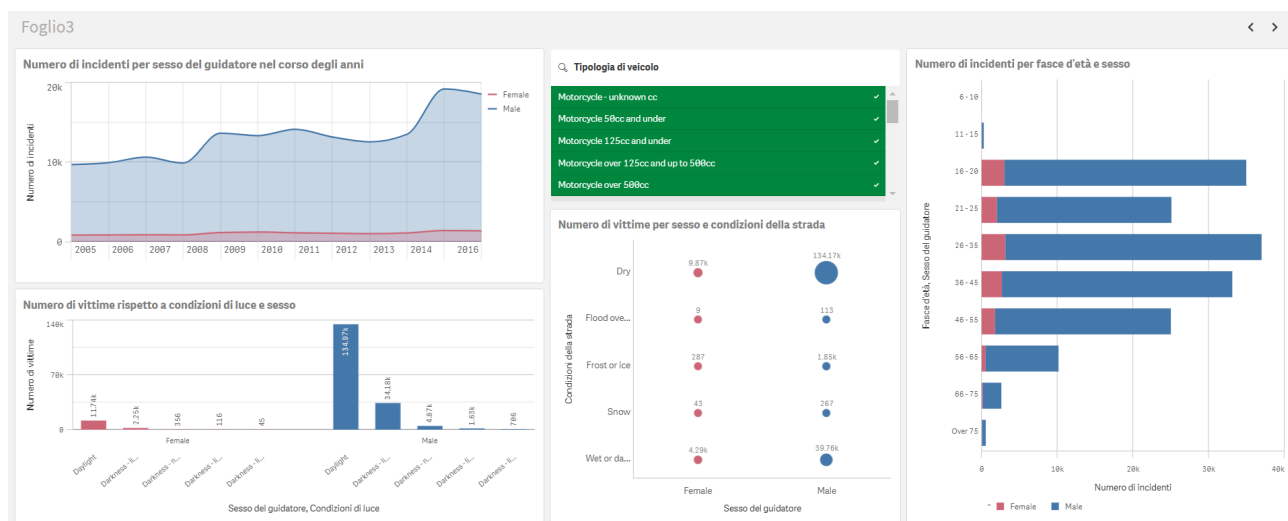


Figura 14: Foglio di analisi numero 3 - Applicazione filtro per motocicli

Con l'applicazione di questi filtri emerge chiaramente il divario nell'utilizzo dei motocicli tra uomini e donne: i dati mostrano infatti che il numero di incidenti che coinvolgono motociclisti di sesso maschile è mediamente quindici volte superiore rispetto a quello delle donne. In tutti gli oggetti visivi si conferma che il numero di incidenti e vittime è nettamente più elevato tra gli uomini. Un aspetto particolarmente interessante si osserva nel *grafico a barre in pila*, dove, a differenza della versione non filtrata, la fascia d'età "16-20" registra un numero di incidenti superiore rispetto alle altre, fatta eccezione per la fascia "26-35", che rimane quella con il maggior numero complessivo di incidenti su motocicli. Questo dato suggerisce che una parte significativa degli incidenti tra i più giovani avviene proprio a bordo di veicoli a due ruote, probabilmente perché rappresentano il mezzo di trasporto più diffuso tra i neopatentati, specialmente nelle aree urbane. In effetti, gli incidenti che coinvolgono motociclisti tra i 16 e i 20 anni costituiscono oltre un quarto del totale, con circa 32.000 incidenti su un totale di 114.000. Questo potrebbe essere legato a fattori come l'inesperienza alla guida, la maggiore esposizione al rischio tipica di questa categoria di veicoli e una minore percezione del pericolo tra i più giovani.

Invece, filtrando i dati soltanto per le automobili si può osservare come variano i grafici all'interno del foglio nella Figura 15.

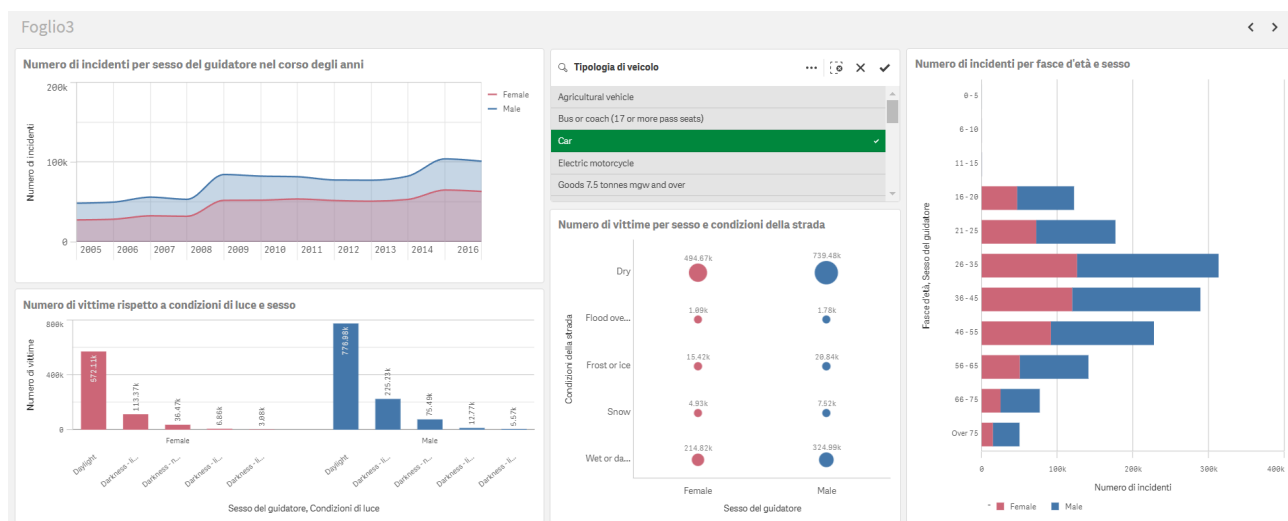


Figura 15: Foglio di analisi numero 3 - Applicazione filtro per automobili

Si osserva che, a differenza del caso precedente, le automobili rappresentano il mezzo di gran lunga più utilizzato dalle donne, costituendo quasi la totalità degli incidenti registrati per le guidatrici. Un esempio evidente si trova confrontando i dati con il foglio senza filtri: nel 2008, su circa 55.000 incidenti che hanno coinvolto donne, ben 52.000 sono avvenuti a bordo di automobili. Anche analizzando gli altri oggetti visivi emerge chiaramente come la stragrande maggioranza degli incidenti che coinvolgono donne avvenga con questo tipo di veicolo, mentre per gli uomini si registra una distribuzione molto più eterogenea, con un coinvolgimento significativo anche in incidenti su motocicli, veicoli commerciali e altri mezzi. Questo potrebbe essere legato a una maggiore propensione delle donne a scegliere l'auto come principale mezzo di trasporto mentre gli uomini tendono a diversificare maggiormente i mezzi utilizzati.

Infine si può analizzare un caso rilevante che è quello dei mezzi di trasporto pubblici o mezzi di trasporto che hanno molti posti a sedere. Filtrando i dati per "Bus or coach (17 or more post seats)", "Minibus (8 to 16 post seats)" e "Tram" si può osservare che (Figura 16):

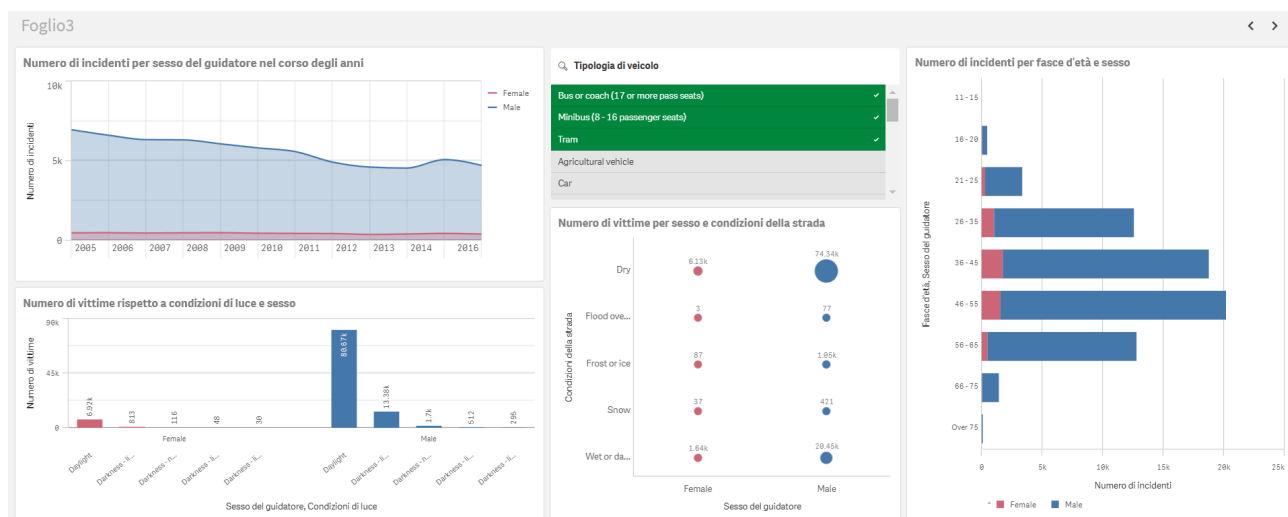


Figura 16: Foglio di analisi numero 3 - Applicazione filtro per mezzi di trasporto a grande capienza

Anche in questo caso si può osservare che il numero di conducenti di sesso maschile è nettamente superiore rispetto a quello femminile, ma l'aspetto più rilevante è che il numero di vittime per ogni incidente risulta significativamente più alto rispetto, ad esempio, ai motocicli. Svolgendo dei calcoli approssimativi, si stima che il numero di vittime per incidente sia pari a circa 1,5, mentre nel caso dei motocicli il valore si attesta intorno a 1,1. Un altro elemento su cui porre l'attenzione è il trend decrescente del numero di incidenti tra gli uomini nel periodo analizzato (2005-2016), che passa da circa 6.950 a circa 4.710, mentre per le donne il valore si mantiene stabile, con una media di circa 400 incidenti all'anno. Questo calo potrebbe essere attribuito a diversi fattori, tra cui il miglioramento delle misure di sicurezza stradale e l'evoluzione delle normative, che hanno reso più severi i controlli e più efficaci le campagne di sensibilizzazione. Inoltre, l'introduzione di nuove tecnologie nei veicoli, come i sistemi di assistenza alla guida e il miglioramento delle dotazioni di sicurezza, potrebbe aver contribuito a ridurre la gravità e la frequenza degli incidenti.

3 Tableau

3.1 Introduzione

Tableau è uno dei software di data visualization più potenti e diffusi, progettato per aiutare le aziende e gli analisti a esplorare, visualizzare e comprendere i dati in modo interattivo. Fondato nel 2003 da Christian Chabot, Pat Hanrahan e Chris Stolte, il software ha rapidamente guadagnato popolarità grazie alla sua interfaccia intuitiva e alle sue potenti funzionalità di analisi. Nel 2019, Tableau è stato acquisito da Salesforce, consolidando ulteriormente la sua posizione nel settore della business intelligence.

Tableau permette di connettersi a diverse fonti di dati, tra cui database, fogli di calcolo e piattaforme cloud, facilitando l'elaborazione e la visualizzazione delle informazioni in tempo reale. Grazie alle sue funzionalità di drag-and-drop, gli utenti possono creare dashboard e report interattivi senza necessitare di competenze avanzate in programmazione.



Figura 17: Logo di Tableau

3.2 Caricamento dati

Per il caricamento del dataset su Tableau, è stata creata una nuova cartella di lavoro, dopodiché sono stati caricati due dataset in formato CSV: `UK_Accident.csv` e `UK_Accident_Vehicles.csv`. Per ciascun dataset, sono stati definiti i separatori del file CSV e, dove necessario, sono state eseguite operazioni di trasformazione dei dati.

Alcuni esempi di tali trasformazioni includono:

- Conversione delle date dal formato stringa al formato data.

- Creazione di una nuova colonna contenente i tre livelli di severità dell'incidente, assegnando i valori numerici: 1 per *Slight*, 2 per *Serious* e 3 per *Fatal*.

Queste operazioni si sono rivelate necessarie per eseguire alcune analisi, come il calcolo della media della severità, che non sarebbe stato possibile con valori testuali.

3.2.1 Relazione dei dataset

Dopo aver caricato i due dataset ed eseguito tutte le operazioni di pulizia, il passo successivo è stato definire la relazione tra i due dataset.

Questa attività è risultata molto semplice, poiché è stata gestita direttamente dal software. Le uniche operazioni eseguite sono state la definizione della chiave di collegamento, `accident_index`, e del tipo di relazione, in questo caso è una relazione di uguaglianza ("=").



Figura 18: Relazione dei due dataset

3.3 Data analysis

Per eseguire le analisi su Tableau, si utilizzano le dashboard, ognuna delle quali può ospitare diversi fogli. Ogni foglio consente di esaminare un aspetto specifico e, combinati tra loro, offrono una visione complessiva più dettagliata su un tema centrale. Le dashboard create su Tableau sono le seguenti:

- *Dashboard 1*: Analisi temporale di incidenti per anno, giorno e ora
- *Dashboard 2*: Distribuzione degli incidenti per tipologia di veicolo e carburante
- *Dashboard 3*: Gravità degli incidenti in base ai limiti di velocità
- *Dashboard 4*: Previsione del numero di vittime e severità media degli incidenti

3.3.1 Dashboard 1 - Analisi temporale di incidenti per anno, giorno e ora

La prima dashboard, mostrata in Figura 19, fornisce un'analisi temporale degli incidenti, suddivisa per anno, giorno della settimana e ora del giorno. Questo tipo di visualizzazione consente di individuare pattern ricorrenti e tendenze nel tempo, facilitando l'identificazione dei periodi con una maggiore concentrazione di incidenti. L'analisi può essere utile per sviluppare strategie di prevenzione e ottimizzare le risorse destinate alla gestione della sicurezza stradale.

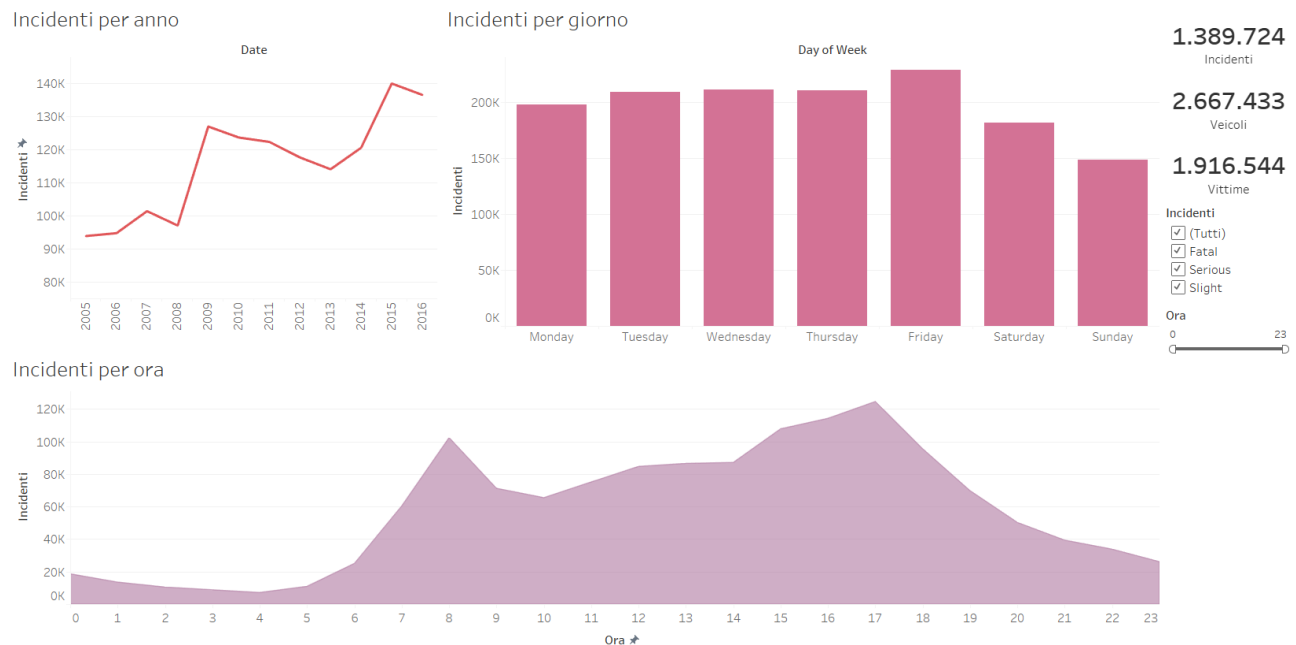


Figura 19: Dashboard di analisi numero 1

In particolare, la dashboard presenta i seguenti oggetti visivi:

- *Grafico a linee*: mostra l'andamento del numero di incidenti nel corso degli anni.
- *Grafico a barre orizzontali*: rappresenta la distribuzione degli incidenti in base al giorno della settimana.
- *Grafico ad area*: illustra la frequenza degli incidenti in base all'ora del giorno.
- *Indicatori di valore*: sono presenti tre differenti indicatori di valore, che riportano i dati sul numero totale di incidenti, il numero totale di veicoli coinvolti e il numero totale di vittime.

Inoltre, sono stati inseriti dei filtri per eseguire analisi più dettagliate:

- *Filtro per severità*: consente di selezionare una o più severità degli incidenti.
- *Filtro per fascia oraria*: consente di selezionare un range orario.

Senza applicare filtri, partendo dal grafico a linee in alto a sinistra, è possibile analizzare l'andamento del numero di incidenti nel corso degli anni. Si può osservare un trend crescente: infatti, a partire dal 2005 si registrano 93.531 incidenti, con un primo picco nel 2009, che raggiunge 127.059 incidenti. Negli anni successivi, si nota una diminuzione del numero di incidenti, seguita

da un nuovo aumento nel 2013, che culmina nel 2015 con il picco massimo, con 140.056 incidenti registrati. Infine, si osserva una leggera discesa nel 2016, con 136.621 incidenti. Questo trend crescente potrebbe suggerire un aumento del traffico e dei veicoli in circolazione, ma anche una possibile insufficienza delle infrastrutture stradali o cambiamenti nei comportamenti di guida.

Passando al grafico a barre verticali in alto a destra, che mostra la distribuzione settimanale degli incidenti, emerge un chiaro pattern: i giorni feriali sono caratterizzati da un numero significativamente maggiore di incidenti, con un picco che si verifica in particolare il giovedì e il venerdì. Questa tendenza potrebbe essere legata a diversi fattori, come l'aumento del traffico causato dalle attività lavorative e dai preparativi per il weekend. In particolare, il venerdì rappresenta una giornata di transizione, in cui la conclusione della settimana lavorativa si unisce al traffico maggiore dovuto agli spostamenti per il fine settimana, creando una situazione di congestione che favorisce il verificarsi di incidenti. Al contrario, durante il weekend si nota una riduzione degli incidenti, con una diminuzione marcata proprio la domenica, che segna il picco di bassa incidenza. Questo potrebbe essere dovuto alla minore attività lavorativa e a una riduzione dei viaggi quotidiani. È interessante notare anche la differenza tra il lunedì e la domenica: il lunedì, infatti, può essere associato alla ripresa della routine settimanale, con l'aumento dello stress e una possibile disattenzione da parte dei conducenti, che porta a un incremento degli incidenti, sebbene non paragonabile a quello di giovedì e venerdì. In generale, i giorni lavorativi mostrano un rischio maggiore rispetto al weekend, con un significativo calo degli incidenti durante le giornate di riposo.

Infine, nel grafico ad area in basso, esaminando i dati, emerge chiaramente che la maggior parte degli incidenti si verifica nella fascia oraria tra le 8:00 e le 18:00, con picchi significativi alle 8:00 e alle 17:00. Questo andamento suggerisce che il traffico legato alle attività lavorative e scolastiche contribuisce in modo significativo alla frequenza degli incidenti, in particolare negli orari di punta, quando si verifica un'elevata densità di veicoli sulla strada. La concentrazione di incidenti in questi orari potrebbe essere attribuita a una combinazione di fattori, tra cui l'intensificarsi degli spostamenti legati al lavoro e la maggiore pressione e stanchezza che i conducenti possono sperimentare.

Applicando alla dashboard i filtri relativi al tipo di incidente (Slight, Serious, Fatal) e in base all'orario, è possibile analizzare in modo più approfondito le differenze orarie in relazione alla severità dell'incidente.

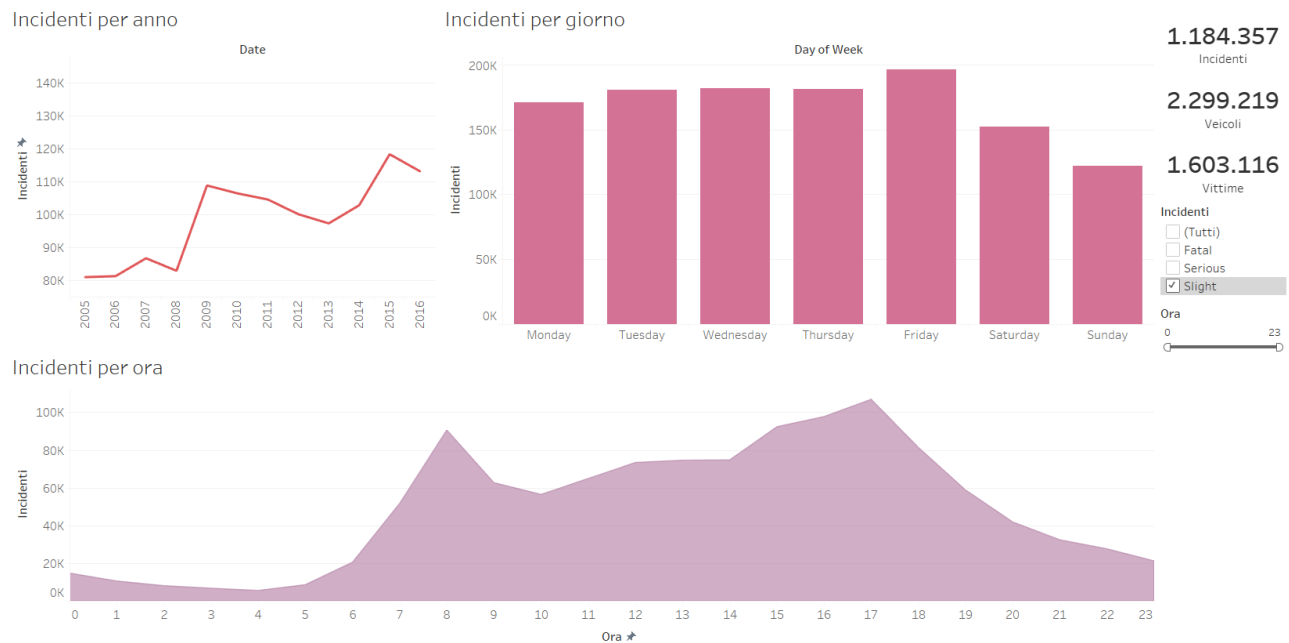


Figura 20: Dashboard di analisi temporale filtrata per incidenti "slight"

Osservando i dati filtrati per severità di incidente, emerge che gli incidenti di lieve entità ("slight") seguono un andamento simile a quello generale, senza variazioni particolarmente marcate. Mostrando quindi che sono gli incidenti che accadono più spesso rispetto a quelli di entità maggiore e che quindi influenzano maggiormente il grafico generale.

Tuttavia, quando si considerano gli incidenti gravi ("serious"), si nota, in *Figura 21* un incremento del numero alla fine del weekend, mentre nei primi giorni della settimana il numero di incidenti rimane più stabile e vicino al trend degli incidenti lievi. L'aumento del numero di incidenti gravi nel fine settimana potrebbe essere associato a una maggiore propensione a comportamenti di guida rischiosi, come l'eccesso di velocità o la guida sotto l'influenza di alcol e sostanze, che tendono ad aumentare nelle ore notturne del venerdì e del sabato. Tuttavia l'andamento durante i primi quattro giorni della settimana rimane coerente con quello filtrato per incidenti lievi. Il numero di incidenti decresce notevolmente infatti abbiamo 187.146 incidenti gravi rispetto ai 1.184.357 di entità lieve.

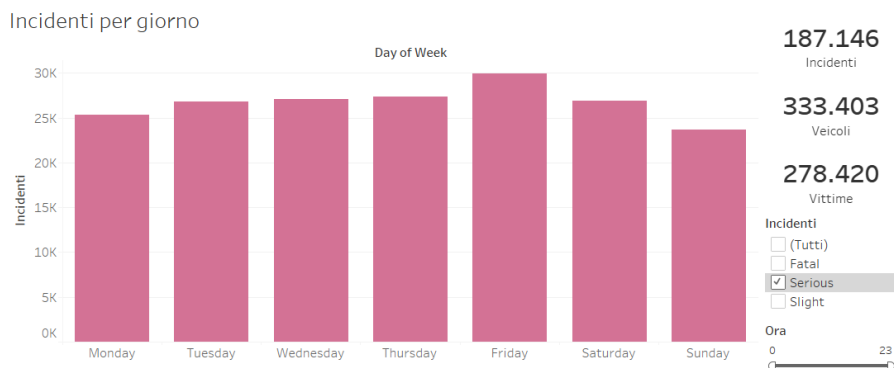


Figura 21: Distribuzione settimanale degli incidenti con severità "serious"

Per quanto riguarda gli incidenti mortali ("fatal"), la variazione presente in *Figura 22* rispetto agli altri due gruppi è ancora più accentuata. Il numero di incidenti fatali aumenta notevolmente durante il fine settimana, per poi diminuire sensibilmente nei giorni feriali. Questo dato indica una correlazione significativa tra la gravità dell'incidente e il giorno della settimana, suggerendo che il rischio di eventi mortali sia più elevato in periodi in cui la guida diventa più rischiosa, probabilmente a causa di una maggiore velocità media dei veicoli, una minore congestione del traffico, un aumento della guida notturna e, potenzialmente, una maggiore incidenza di guida sotto effetto di alcol o sostanze. Durante la settimana, invece, il numero di incidenti mortali è sensibilmente più basso, il che potrebbe essere dovuto alla presenza di traffico più regolato e a una minore incidenza di guida pericolosa.

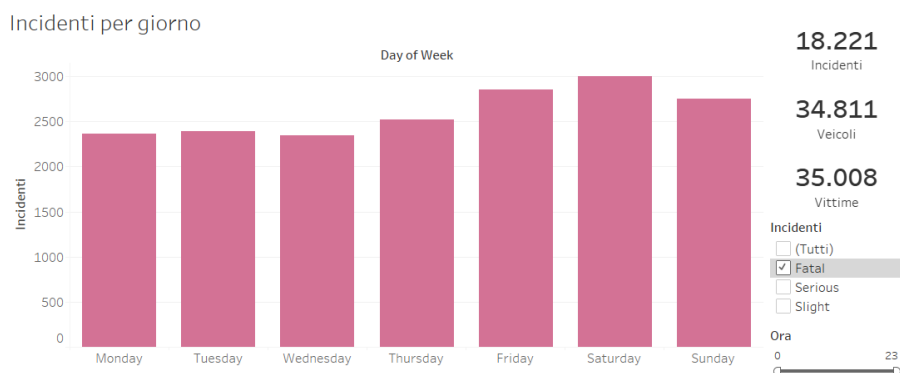


Figura 22: Distribuzione settimanale degli incidenti con severità "fatal"

L'analisi della distribuzione oraria ha permesso di evidenziare ulteriori differenze tra le categorie di severità. Per gli incidenti lievi ("slight"), in *Figura 20*, emergono due picchi molto accentuati nelle fasce orarie 8:00-9:00 e 17:00-18:00. Questo andamento è fortemente correlato ai flussi di traffico nei momenti di punta della giornata, ovvero le ore in cui le persone si spostano

per recarsi al lavoro o a scuola e per rientrare a casa nel pomeriggio. In queste fasce orarie, l'aumento del traffico comporta un incremento del numero totale di incidenti, sebbene la gravità di questi eventi rimanga prevalentemente bassa, probabilmente a causa della velocità ridotta e della maggiore attenzione dei conducenti nei contesti di traffico intenso.

Per gli incidenti gravi ("serious") e mortali ("fatal"), la distribuzione oraria, mostrata rispettivamente in *Figura 23* e *Figura 24*, non presenta picchi così evidenti nelle stesse fasce. Questo suggerisce che la severità di un incidente non sia necessariamente legata alla densità del traffico, ma piuttosto a fattori diversi come la velocità, la stanchezza o la visibilità. Gli incidenti più gravi tendono infatti a verificarsi in orari meno congestionati, come la notte o le prime ore del mattino, quando le strade sono più libere e i conducenti possono essere più inclini a superare i limiti di velocità o a distrarsi.

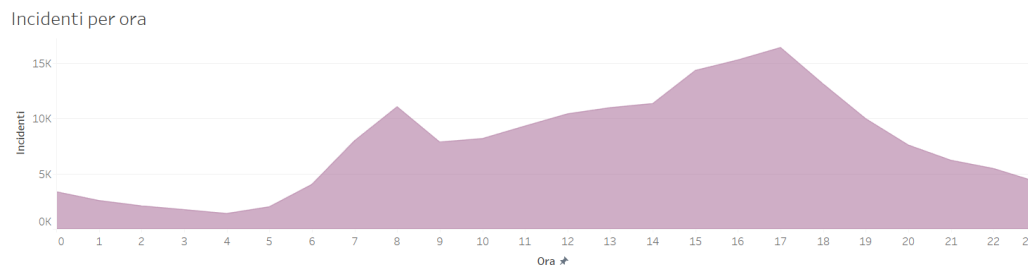


Figura 23: Distribuzione oraria degli incidenti "serious"

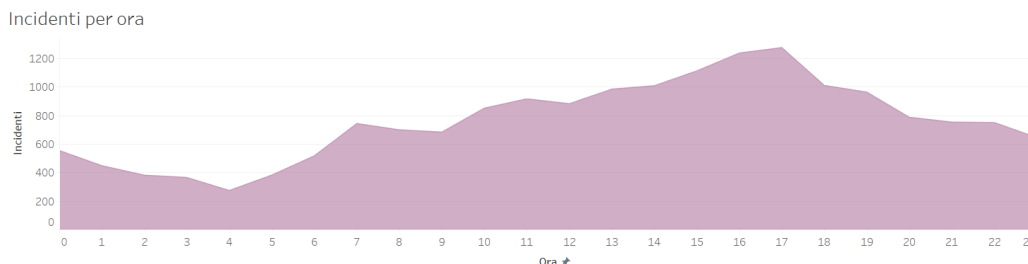


Figura 24: Distribuzione oraria degli incidenti "fatal"

Concludendo, gli incidenti lievi sembrano essere fortemente correlati alla densità del traffico, con picchi ben definiti durante le ore di punta. Gli incidenti gravi, invece, presentano un andamento più irregolare e mostrano un incremento significativo nel weekend, indicando che la gravità degli incidenti potrebbe dipendere più da fattori comportamentali che dalla semplice esposizione al traffico. Infine, gli incidenti mortali seguono un pattern ancora più specifico, con una netta prevalenza nei giorni del fine settimana, probabilmente a causa di una combinazione di velocità più elevate, condizioni stradali diverse e una maggiore incidenza di guida pericolosa.

3.3.2 Dashboard 2 - Distribuzione degli incidenti per tipologia di veicolo e carburante

La seconda dashboard, mostrata in *Figura 25*, offre un'analisi approfondita degli incidenti stradali, suddivisi in base a vari fattori che caratterizzano i veicoli coinvolti. In particolare, vengono analizzati diversi aspetti, come la casa produttrice del veicolo, la sua tipologia, il tipo di carburante utilizzato e l'età del veicolo. Questo consente di ottenere una visione più dettagliata sui veicoli coinvolti negli incidenti.

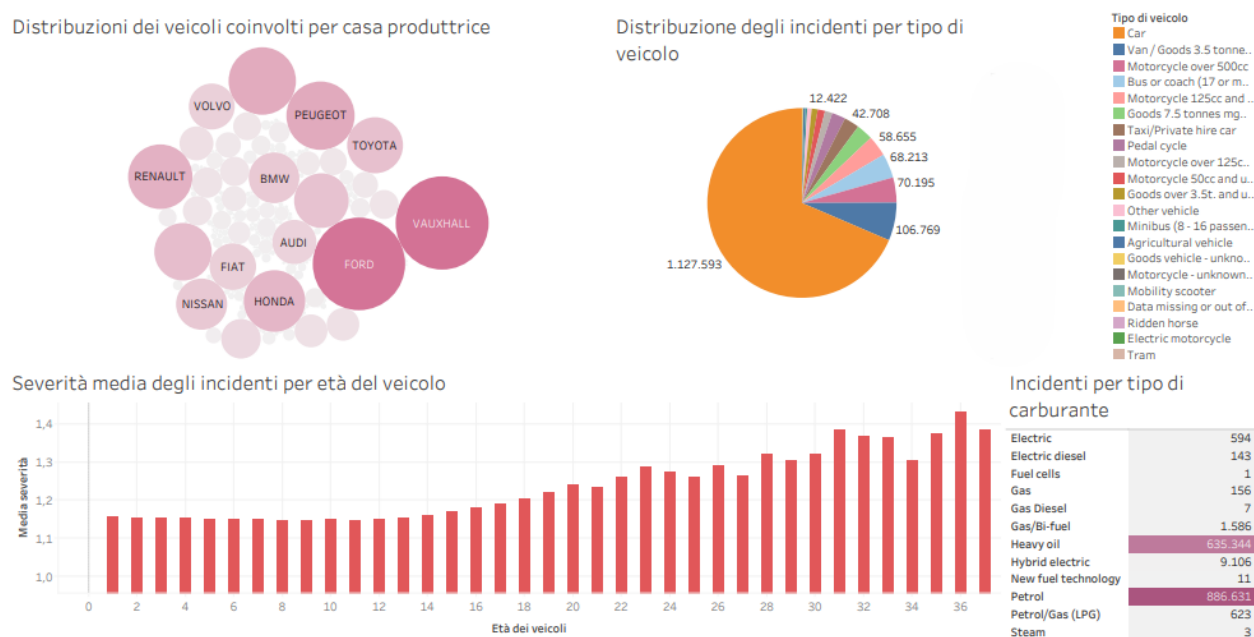


Figura 25: Dashboard di analisi numero 2

Nella dashboard sono presenti quattro oggetti visivi:

- *Grafico a bolle*: mostra la distribuzione delle case produttrici in relazione al numero di veicoli coinvolti in incidenti.
- *Grafico a torta*: rappresenta la distribuzione degli incidenti in base alle diverse tipologie di veicoli.
- *Grafico a barre verticali*: illustra la severità media degli incidenti in relazione all'età del veicolo.
- *Tabella di valori*: riporta il numero di incidenti per ciascun tipo di carburante, evidenziando quelli con i valori più elevati.

Partendo dal grafico a bolle in alto a sinistra, si osserva la distribuzione dei veicoli coinvolti negli incidenti per casa produttrice. Dai dati emerge che i marchi con il maggior numero di sinistri sono Vauxhall (217.649) e Ford (214.822), seguiti da Peugeot, Renault, Honda, BMW, Audi, Toyota e Nissan, con valori inferiori. Questa tendenza può dipendere da diversi fattori, come il numero di veicoli di quel marchio in circolazione, le caratteristiche di sicurezza dei modelli e il profilo dei conducenti che prediligono determinati marchi.

Spostandoci sul grafico a torta in alto a destra, emerge che la maggior parte degli incidenti coinvolge autovetture, seguite da furgoni con un peso di 3.5 tonnellate e moto con cilindrata superiore a 500cc. Questo dato è prevedibile considerando che le autovetture rappresentano il mezzo di trasporto più diffuso, i furgoni sono ampiamente utilizzati per il trasporto merci, e le moto di grossa cilindrata sono spesso coinvolte in incidenti a causa delle alte velocità e della loro minore stabilità rispetto agli altri veicoli. In particolare, le motociclette presentano una minore protezione rispetto alle auto, una maggiore esposizione a condizioni meteorologiche avverse ed inoltre la tendenza dei motociclisti a viaggiare a velocità più elevate porta un numero di incidenti maggiore in proporzione. Inoltre, i motociclisti sono più vulnerabili agli errori di altri conducenti, come manovre improvvise o mancata visibilità negli angoli ciechi. Per quanto riguarda i veicoli commerciali, i furgoni e i mezzi pesanti sono coinvolti in una quota inferiore di incidenti rispetto alle autovetture. Tuttavia, è un dato da tenere in forte considerazione, poiché quando un incidente coinvolge un mezzo pesante, il peso e le dimensioni di questi veicoli contribuiscono a un impatto più distruttivo, soprattutto in aree urbane o su strade ad alto scorrimento.

Spostandoci nel grafico a barre verticali in basso a sinistra, un altro aspetto rilevante riguarda la severità media degli incidenti in base all'età del veicolo. Dai dati emerge che i veicoli più vecchi tendono ad essere coinvolti in incidenti con una gravità maggiore. Questo potrebbe essere dovuto a diversi fattori, tra cui l'assenza di tecnologie avanzate di sicurezza nei modelli più datati, la minore efficienza dei sistemi frenanti e la probabilità che i veicoli più vecchi siano meno mantenuti rispetto a quelli più nuovi. Al contrario, i veicoli più recenti, dotati di sistemi avanzati di assistenza alla guida, risultano coinvolti in incidenti meno gravi, evidenziando l'importanza dell'innovazione tecnologica nella riduzione degli impatti degli incidenti stradali.

Infine, nella tabella in basso a destra è possibile analizzare il tipo di carburante utilizzato. Dai dati riportati si osserva che i veicoli alimentati a benzina e diesel sono i più coinvolti negli incidenti, il che è coerente con la loro maggiore diffusione sul mercato. È interessante notare, inoltre, che i veicoli elettrici e ibridi sono coinvolti in un numero significativamente inferiore di incidenti. Questo fenomeno è attribuibile a diversi fattori, tra cui la minore diffusione di questi veicoli rispetto a quelli tradizionali e il loro impiego prevalente in contesti urbani, dove la velocità è generalmente più contenuta e i rischi di incidenti gravi sono ridotti. Queste informazioni potrebbero essere analizzate con dati attuali per verificare la possibile esistenza di un trend emergente: con l'aumento della quota di mercato dei veicoli elettrici, potrebbe essere interessante monitorare se la loro percentuale di incidenti subirà un incremento nei prossimi anni. Ciò potrebbe dipendere non solo dalla crescente presenza di questi veicoli sulle strade,

ma anche da fattori come la silenziosità dei motori elettrici, che potrebbe ridurre la percezione acustica del veicolo da parte di pedoni e altri utenti della strada, aumentando il rischio di incidenti in alcuni contesti urbani.

Questi dati offrono spunti utili per implementare strategie di prevenzione, come campagne di sensibilizzazione rivolte ai motociclisti, miglioramenti nella sicurezza dei veicoli commerciali e adeguamenti normativi per la gestione della crescente diffusione dei veicoli elettrici sulle strade.

3.3.3 Dashboard 3 - Analisi degli incidenti in base a caratteristiche stradali

La terza dashboard, mostrata in *Figura 26*, offre un'analisi generale degli incidenti stradali in relazione alle caratteristiche delle strade. In particolare, vengono esplorati vari aspetti, come la distribuzione degli incidenti in base ai limiti di velocità, la distinzione tra aree urbane e rurali, e la tipologia di strada. Questa dashboard fornisce una visione più completa dei fattori stradali che influenzano gli incidenti, permettendo di identificare possibili correlazioni tra le caratteristiche della strada e la frequenza degli incidenti.

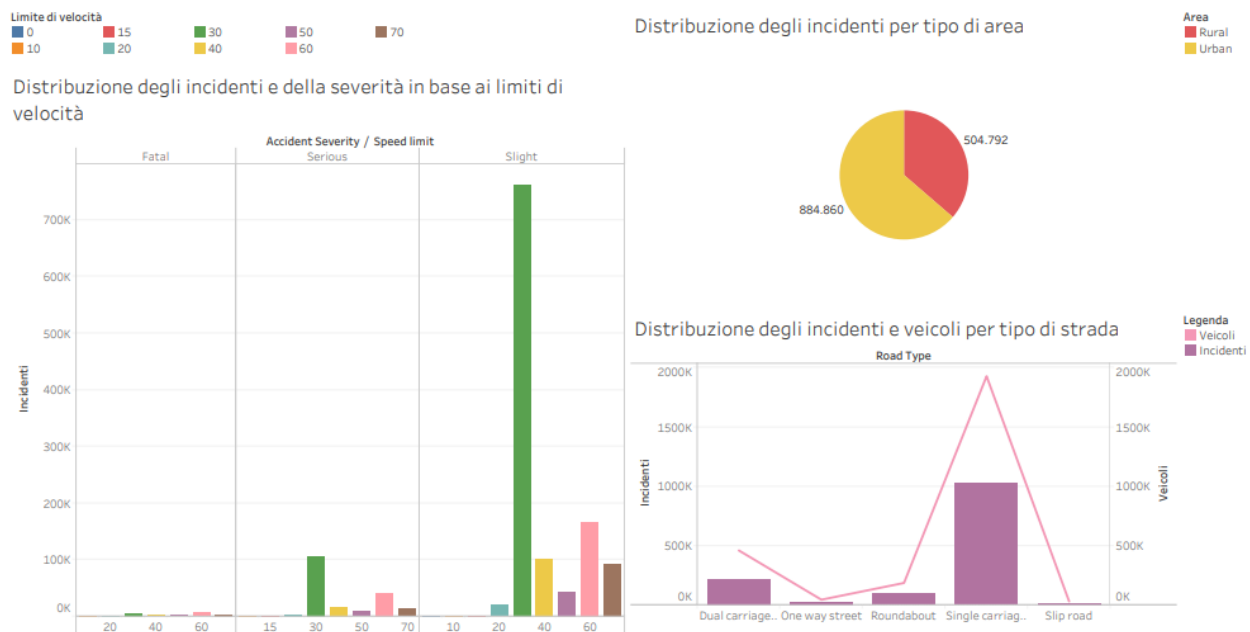


Figura 26: Dashboard di analisi numero 3

Nella dashboard sono presenti tre oggetti visivi:

- *Grafico a barre affiancate*: mostra la distribuzione degli incidenti in base ai limiti di velocità e alla severità.

- *Grafico a torta*: mostra la distribuzione degli incidenti tra aree urbane e rurali.
- *Grafico combinato a linee e barre verticali*: mostra il numero di incidenti e di veicoli coinvolti in base alla tipologia di strada.

L'analisi dei dati del primo grafico a sinistra mostra una chiara relazione tra il limite di velocità e la gravità degli incidenti. Le strade con limiti di velocità più bassi (fino a 30 mph) registrano un numero maggiore di incidenti, ma la percentuale di quelli gravi e fatali è inferiore rispetto alle strade con limiti di velocità superiori. Questo può essere dovuto al fatto che, a velocità più basse, l'energia dell'impatto è minore, riducendo il rischio di lesioni gravi per gli occupanti del veicolo e per gli altri utenti della strada. D'altra parte, sulle strade con limiti superiori a 50 mph, la probabilità di incidenti fatali aumenta considerevolmente. In particolare, i limiti di 70 mph e superiori sono associati a un incremento significativo della severità media degli incidenti. Questo è attribuibile a diversi fattori, tra cui maggiore velocità d'impatto, tempi di reazione più ridotti e minore possibilità di evitare ostacoli improvvisi.

Nel grafico a torta in alto a destra viene analizzata un'altra variabile rilevante: la distinzione tra aree urbane e rurali. I dati mostrano che la maggior parte degli incidenti, circa il 63,7% (884.860), avviene in contesti urbani, mentre il 36,3% (504.792) si verifica in aree rurali. Questo riflette l'impatto della maggiore densità di traffico, delle numerose intersezioni e della presenza di pedoni e ciclisti nelle città, fattori che aumentano il rischio di incidenti. Al contrario, nelle aree rurali, sebbene il traffico sia meno intenso, possono entrare in gioco altri pericoli, come velocità più elevate e condizioni stradali variabili.

Infine, analizzando il grafico combinato in basso a destra, emerge che la maggior parte degli incidenti avviene sulle strade a carreggiata singola, seguite dalle strade a carreggiata doppia e dalle rotonde. Queste ultime sono particolarmente interessanti dal punto di vista della sicurezza stradale: sebbene vengano introdotte per ridurre il numero complessivo di incidenti rispetto alle intersezioni tradizionali, rappresentano ancora un punto critico. Il comportamento degli automobilisti all'interno delle rotonde, la scarsa segnaletica e le manovre improvvise contribuiscono a generare situazioni di pericolo. Un altro aspetto rilevante riguarda il numero di veicoli coinvolti in relazione agli incidenti per tipo di strada. Nelle strade a carreggiata singola, dove i veicoli circolano nei due sensi senza separatori, il numero di veicoli coinvolti è quasi il doppio rispetto al numero di incidenti, suggerendo che molti coinvolgano più di un veicolo. Questo potrebbe indicare che si tratta spesso di scontri frontali o tamponamenti multipli, che tendono a coinvolgere più mezzi. Tuttavia, in media, il numero di veicoli coinvolti non risulta esattamente il doppio, ma leggermente inferiore a causa della presenza degli incidenti con un solo veicolo (come quelli fuori strada) o con più di due (come quelli a catena). Al contrario, sulle strade a senso unico (One Way Street), il rapporto tra il numero di veicoli coinvolti e gli incidenti è circa pari a 1, il che suggerisce che la maggior parte riguardi un solo veicolo, spesso a causa di perdita di controllo, distrazione o condizioni stradali sfavorevoli.

Questi risultati evidenziano l'importanza di politiche di riduzione della velocità, in particolare nelle aree urbane e nelle strade a carreggiata singola, dove avviene la maggior parte

degli incidenti. Misure come l'installazione di dossi rallentatori, segnaletica migliorata e l'uso di tecnologie avanzate di monitoraggio del traffico possono contribuire a ridurre il numero di incidenti e la loro gravità. Allo stesso tempo, nelle aree rurali, interventi mirati come una migliore illuminazione stradale, manutenzione del manto stradale e maggiori controlli possono aiutare a mitigare i rischi legati alle alte velocità e alle condizioni stradali meno favorevoli.

3.3.4 Dashboard 4 - Previsione del numero di vittime e severità media degli incidenti

La quarta dashboard, mostrata in *Figura 27*, presenta un'analisi basata su modelli predittivi forniti da Tableau, con l'obiettivo di stimare il numero di vittime e la severità media degli incidenti nel periodo dal 2017 al 2020.

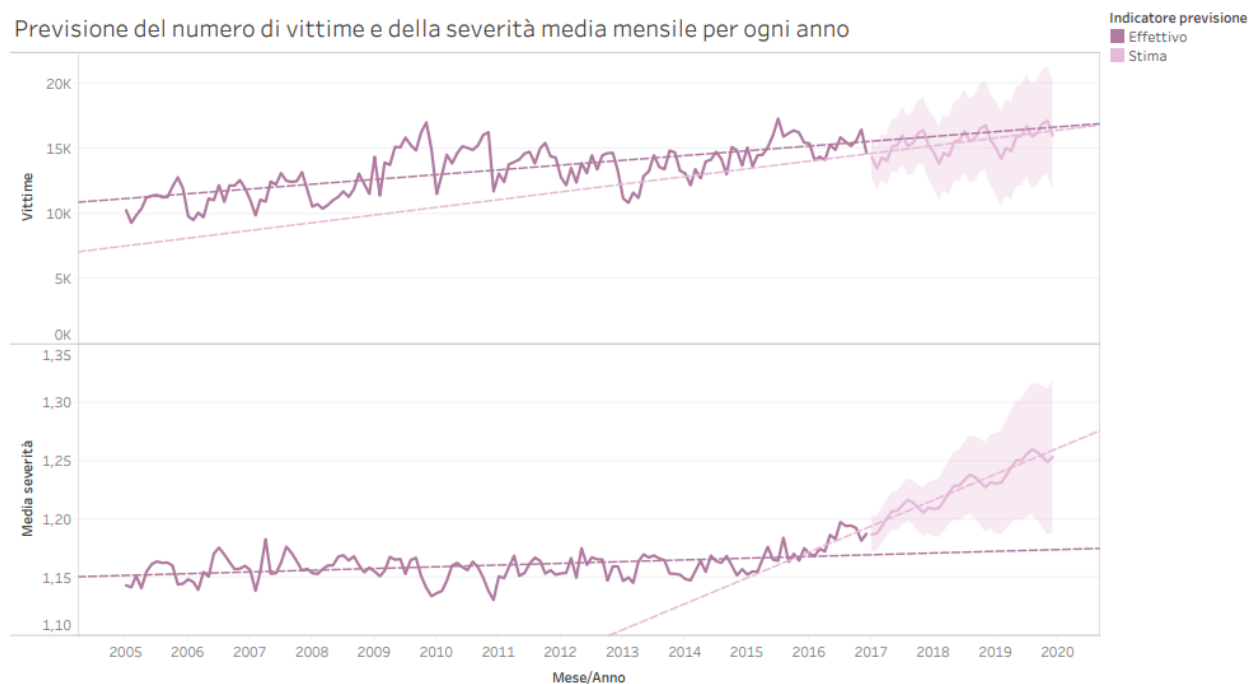


Figura 27: Dashboard di analisi numero 4

All'interno di questa dashboard sono, dunque, presenti due grafici a linee che mostrano l'andamento nel tempo del numero di vittime e della severità media degli incidenti, rispettivamente. Il primo grafico visualizza il numero totale di vittime mensile per anno, evidenziando eventuali tendenze e picchi di incidenti. Il secondo grafico, invece, rappresenta la severità media degli incidenti e permette di osservare eventuali cambiamenti nel livello di pericolosità degli incidenti nel periodo considerato.

Queste analisi non solo permettono di comprendere l'andamento storico degli incidenti, ma sono anche utili per effettuare analisi previsionali. L'analisi previsionale delle vittime stradali e della severità degli incidenti è un aspetto cruciale per la pianificazione della sicurezza stradale e per l'identificazione di trend emergenti. Tableau, grazie ai suoi strumenti di previsione avanzati, consente di generare modelli basati su serie temporali e metodi di livellamento esponenziale per stimare il numero futuro di incidenti e la loro gravità.

Nelle previsioni fornite, i dati sono stati elaborati considerando una serie storica dal gennaio 2005 al dicembre 2016, con una previsione avanzata di 36 mesi, ovvero dal gennaio 2017 al dicembre 2019. La scelta di Tableau per la modellazione previsionale è dovuta alla sua capacità di applicare tecniche statistiche avanzate in modo intuitivo, rendendo possibile una valutazione immediata dei trend e delle fluttuazioni stagionali.

Opzioni utilizzate per creare previsioni

Serie temporali: Mese di Date
Misure: Med. Accident Severity Numeric, Somma di Number of Casualties
Previsione avanti: 36 mesi (gennaio 2017 – dicembre 2019)
Previsioni basate su: gennaio 2005 – dicembre 2016
Ignora ultimo: Nessun periodo ignorato
Modello stagione: 12 ciclo mensile

Med. Accident Severity Numeric

Iniziale gennaio 2017	Cambia dall'iniziale gennaio 2017 – dicembre 2019	Effetto stagione		Contributo		Qualità
		Elevato	Basso	Trend	Stagione	
1,1871 ± 0,0149	0,0664	agosto 2019 0,0104	febbraio 2019 -0,0073	91,2%	8,8%	Scarso

Somma di Number of Casualties

Iniziale gennaio 2017	Cambia dall'iniziale gennaio 2017 – dicembre 2019	Effetto stagione		Contributo		Qualità
		Elevato	Basso	Trend	Stagione	
14.322 ± 1.488	1.672	novembre 2019 1.184	febbraio 2019 -1.465	14,4%	85,6%	Ok

Figura 28: Descrizione della previsione fornita da Tableau

La severità media degli incidenti presenta una tendenza crescente nel periodo di previsione. Il valore iniziale della previsione a gennaio 2017 è stato stimato in $1,1871 \pm 0,0149$, con una variazione complessiva fino a dicembre 2019 di $+0,0664$.

Tutte le previsioni sono state calcolate utilizzando un livellamento esponenziale.

Med. Accident Severity Numeric

Livello	Modello		Metriche della qualità					Coefficienti di livellamento		
	Trend	Stagione	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Aggiuntivo	Aggiuntivo	Aggiuntivo	0,0076	0,0057	0,80	0,5%	-1.372	0,140	0,201	0,200

Somma di Number of Casualties

Livello	Modello		Metriche della qualità					Coefficienti di livellamento		
	Trend	Stagione	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Aggiuntivo	Aggiuntivo	Aggiuntivo	759	494	0,59	3,8%	1.944	0,440	0,000	0,095

Figura 29: Descrizione dei modelli utilizzati per la previsione fornita da Tableau

Dai coefficienti di livellamento esponenziale riportati nell'analisi, si nota che:

- **Alpha** = 0,140: indica che il modello si aggiorna lentamente, dando ancora un peso significativo ai dati storici rispetto ai più recenti.
- **Beta** = 0,201: suggerisce che il modello aggiorna il trend in modo moderato, senza però specificare se tale trend sia in crescita o in calo.
- **Gamma** = 0,200: indica che il modello aggiorna il pattern stagionale in modo moderato, senza variazioni eccessivamente rapide.

L'effetto stagionale è più evidente in alcuni mesi: agosto 2019 ha mostrato il picco più alto (+0,0104), mentre febbraio 2019 il valore più basso (-0,0073). Questa variazione suggerisce che la severità media degli incidenti possa essere influenzata da fattori stagionali, come condizioni meteorologiche, traffico vacanziero e altri elementi specifici. Tuttavia, il contributo della tendenza generale è molto elevato (91,2%), mentre la stagionalità incide per l'8,8%, confermando che l'aumento della severità è principalmente dovuto a fattori a lungo termine piuttosto che a fluttuazioni stagionali.

L'analisi previsionale della severità media degli incidenti è stata valutata attraverso diverse metriche di errore e qualità del modello:

- **RMSE** (Root Mean Squared Error) = 0,0076: misura l'errore quadratico medio tra i valori osservati e quelli previsti. Un valore basso indica che le previsioni sono abbastanza precise.
- **MAE** (Mean Absolute Error) = 0,0057: rappresenta l'errore medio assoluto tra i valori reali e quelli stimati. Anche in questo caso, un valore basso suggerisce una buona affidabilità del modello.

- **MASE** (Mean Absolute Scaled Error) = 0,80: indica che l'errore medio del modello è leggermente inferiore rispetto a un metodo di riferimento, come una semplice media mobile. Un valore inferiore a 1 è positivo, ma 0,80 suggerisce che il miglioramento rispetto al metodo base non è particolarmente elevato.
- **MAPE** (Mean Absolute Percentage Error) = 0,5%: indica che l'errore medio relativo nelle previsioni è molto basso, suggerendo un'ottima capacità del modello di adattarsi ai dati.
- **AIC** (Akaike Information Criterion) = -1372: un valore negativo suggerisce un buon adattamento del modello ai dati. Tuttavia, il valore deve essere confrontato con altri modelli per valutare l'effettiva superiorità dell'approccio utilizzato.

Nonostante questi buoni risultati in termini di errore, la qualità complessiva della previsione è stata valutata come scarsa. Questo potrebbe suggerire che vi siano fattori esterni non inclusi nel modello che influenzano la severità media degli incidenti. Di conseguenza, potrebbe essere utile considerare variabili aggiuntive per migliorare la capacità predittiva del modello.

L'analisi previsionale della somma del numero di vittime (Number of Casualties) mostra un incremento tra il 2017 e il 2019. Il valore iniziale a gennaio 2017 è stato stimato in 14.322 ± 1.488 , con un aumento di 1.672 vittime nel periodo di previsione.

I coefficienti di livellamento esponenziale rivelano:

- **Alpha** = 0,440: indica che il modello è più reattivo ai dati più recenti, dando loro un peso maggiore rispetto ai dati storici.
- **Beta** = 0,000: suggerisce l'assenza di una chiara tendenza crescente o decrescente nel numero di vittime.
- **Gamma** = 0,095: indica che la componente stagionale è presente ma con un effetto limitato sulla previsione.

L'effetto stagionale sui dati evidenzia che il picco massimo di vittime si è registrato a novembre 2019 (+1.184 vittime), mentre il minimo è stato a febbraio 2019 (-1.465 vittime).

La stagionalità incide per l'85,6%, mentre il trend generale contribuisce solo per il 14,4%. Questo suggerisce che il numero di vittime è fortemente influenzato da fattori stagionali, come le condizioni meteorologiche, il traffico durante le festività e le variazioni nei comportamenti degli utenti della strada.

Dall'analisi delle metriche di qualità del modello:

- **RMSE** = 759 e **MAE** = 494 indicano che l'errore assoluto medio nelle previsioni è elevato.
- **MASE** = 0,59 e **MAPE** = 3,8%, che mostrano un'accuratezza discreta nelle previsioni, ma con una variabilità significativa.

- $AIC = 1.944$, un valore positivo che indica che il modello potrebbe essere migliorato con ulteriori variabili esplicative.

Dai risultati emerge che, mentre la severità degli incidenti segue una tendenza crescente con un basso impatto stagionale, il numero di vittime è fortemente influenzato da fattori stagionali. Questo significa che le misure di sicurezza stradale dovrebbero essere implementate con un focus maggiore sui periodi critici dell'anno per ridurre il numero di vittime.

In conclusione, l'analisi previsionale della severità media degli incidenti e del numero di vittime fornisce indicazioni cruciali per la gestione della sicurezza stradale. I risultati evidenziano una tendenza crescente nella severità degli incidenti, con un impatto stagionale limitato, mentre il numero di vittime è fortemente influenzato da fattori stagionali. Queste informazioni suggeriscono che le strategie di prevenzione dovrebbero concentrarsi sia su interventi a lungo termine per ridurre la gravità degli incidenti sia su misure mirate nei periodi critici per limitare il numero di vittime.

Nonostante le previsioni fornite da Tableau siano utili per individuare trend e schemi ricorrenti, la qualità complessiva dei modelli suggerisce la necessità di integrare ulteriori variabili esplicative per migliorare la capacità predittiva. Fattori esterni, come le condizioni meteorologiche dettagliate, la densità del traffico e le politiche di sicurezza stradale, potrebbero fornire un quadro più accurato e completo. Future analisi potranno quindi beneficiare di modelli più complessi e integrati, contribuendo a una pianificazione più efficace delle misure di sicurezza stradale.

4 Power BI

4.1 Introduzione

Power BI è una suite di strumenti di business intelligence sviluppata da Microsoft che consente agli utenti di connettersi a una vasta gamma di origini di dati, visualizzare i dati in modo interattivo e condividere le visualizzazioni con altri. Inoltre, Power BI offre numerose funzionalità intuitive per la trasformazione, l'analisi e la presentazione dei dati, semplificando il processo per le organizzazioni nell'estrarre informazioni significative dalle loro fonti e nel prendere decisioni informate. Tutte queste caratteristiche rendono Power BI uno strumento altamente valido per l'analisi e la visualizzazione dei dati, posizionandosi come una delle soluzioni migliori nel mercato, spesso preferito rispetto a software come Qlik e Tableau.



Figura 30: Logo di Power BI

4.2 Caricamento dati

Una volta avviato Power BI e creato un nuovo progetto, il primo passo per eseguire le analisi è connettersi ai due dataset. Per farlo, è necessario caricare in Power BI i due file CSV: `UK_Accident.csv` e `UK_Accident_Vehicles.csv`. Durante il caricamento, tramite l'utilizzo di Power Query, sono state eseguite delle operazioni di trasformazione, dovute a riconoscimenti errati dei campi numerici.

Inizialmente, è stato caricato il file senza alcuna distinzione sul numero di colonne. Successivamente, utilizzando la "virgola" come delimitatore, la tabella è stata suddivisa in più colonne. Il risultato ottenuto è stato una tabella con le intestazioni delle colonne come prima riga, problema risolto velocemente tramite la funzionalità "Usa la prima riga come intestazione".

Per evitare problemi con i valori numerici, tramite l'apposito tool di sostituzione, sono stati sostituiti i "punti" con le "virgole" nelle colonne interessate (come ad esempio le colonne delle coordinate o le colonne con il limite di velocità).

Inoltre, altre operazioni eseguite includono la conversione del campo "Time" nella corrispondente ora, e l'aggiunta di una colonna per convertire il livello di severità dell'incidente da

testo al corrispondente valore numerico (in particolare: Fatal = 3, Serious = 2, Slight = 1), al fine di facilitare calcoli futuri.

Una volta eseguite tutte queste operazioni, la fase di caricamento termina, si può chiudere Power Query e si possono salvare i cambiamenti fatti tramite il pulsante "Chiudi e Applica".

4.2.1 Relazione dei dataset

Per quanto riguarda il collegamento dei due dataset, esso è stato gestito automaticamente da Power BI tramite l'utilizzo della chiave comune **Accident_Index**, presente in entrambi i dataset. Il tipo di relazione impostata è una relazione *uno-a-molti*, poiché a un incidente possono essere associati più veicoli. L'unica modifica effettuata è stata quella di impostare la direzione del filtro incrociato su "Entrambi", in modo tale da poter utilizzare i campi delle due tabelle in modo indistinto tra loro.

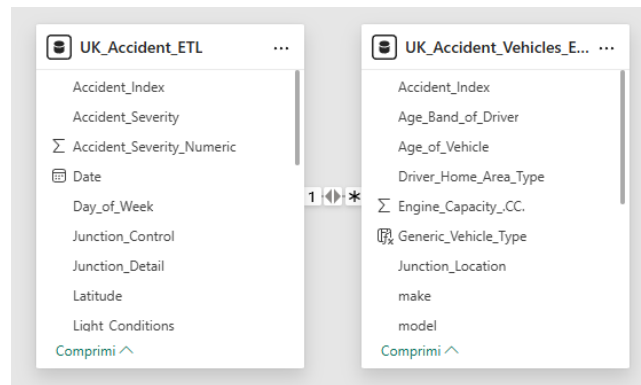


Figura 31: Relazione dataset

4.3 Data analysis

Per eseguire le analisi su Power BI, si utilizzano i fogli di lavoro, ognuno dei quali può ospitare diversi report. Ogni report consente di esaminare un aspetto specifico e, combinati tra loro, offrono una visione complessiva più dettagliata su un tema centrale. I fogli creati in Power BI sono i seguenti:

- *Foglio 1*: Analisi degli incidenti stradali in base alle condizioni ambientali e temporali
- *Foglio 2*: Confronto degli incidenti tra automobili e motociclette
- *Foglio 3*: Analisi delle dinamiche degli incidenti
- *Foglio 4*: Predizione del numero di incidenti in base alla gravità degli incidenti

4.3.1 Foglio 1 - Analisi degli incidenti in base alle condizioni ambientali e temporali

La prima dashboard realizzata con Power BI, mostrata in *Figura 32*, fornisce un'analisi dettagliata degli incidenti stradali in relazione alle condizioni ambientali e meteorologiche. L'obiettivo è identificare tendenze e correlazioni che possano contribuire a una migliore comprensione dei fattori che influenzano la frequenza e la gravità degli incidenti. Le condizioni meteorologiche e le variabili ambientali, come le condizioni stradali e le condizioni di luce, sono essenziali per prevedere le situazioni di maggiore rischio e ottimizzare gli interventi preventivi, come la gestione del traffico e la sicurezza stradale.

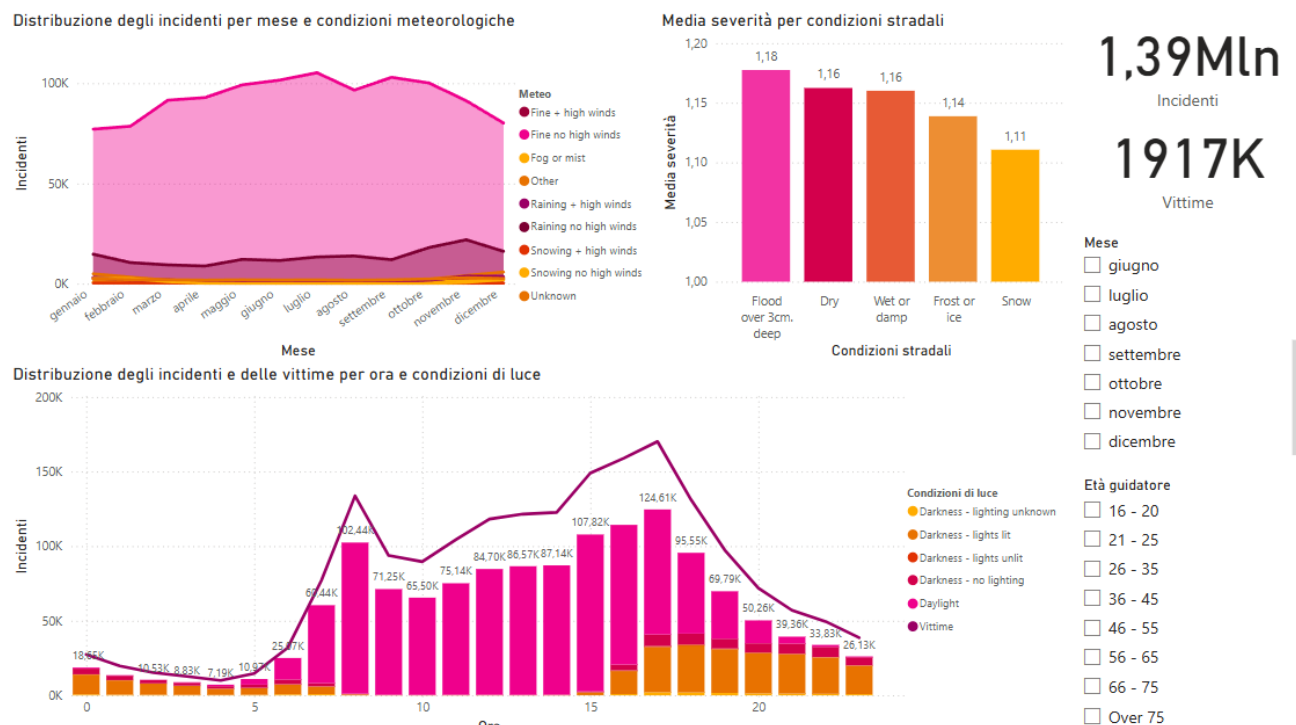


Figura 32: Dashboard 1

Il foglio contiene cinque diversi oggetti visivi:

- Un *grafico ad aree* che rappresenta la distribuzione degli incidenti in base alle condizioni meteorologiche nei vari mesi dell'anno.
- Un *istogramma a colonne in pila* che illustra, per ogni condizione stradale, la media della severità degli incidenti.
- Un *grafico combinato a linee e istogramma a colonne in pila* che mostra il numero di incidenti, colorati in base alla condizione di luce, e il numero di vittime suddivise per ora.

- Un *indicatore di valore* che riporta il numero totale di incidenti.
- Un *indicatore di valore* che riporta il numero totale di vittime.

A supporto di questi oggetti visivi, sulla destra sono stati introdotti i seguenti filtri:

- Mese dell'anno.
- Età del guidatore.

Senza applicare alcun filtro, osservando il grafico ad aree in alto a sinistra, si nota che la maggior parte degli incidenti avvengono principalmente con due condizioni meteorologiche, ossia: "Tempo buono senza vento forte" ("Fine no high winds") e "Pioggia senza vento forte" ("Raining no high winds"). Tuttavia, si può osservare che la distribuzione degli incidenti ha un forte fattore stagionale. Infatti, la condizione "Tempo buono senza vento forte" presenta picchi soprattutto nei mesi estivi, come giugno con 101.569 incidenti e luglio con 105.254 incidenti, per poi diminuire nei periodi autunnali/invernali, quando il maltempo aumenta, con valori più bassi nei mesi di dicembre, gennaio e febbraio, dove gli incidenti si stabilizzano rispettivamente intorno a 80.143, 78.560 e 77.098 incidenti. Al contrario, per quanto riguarda "Pioggia senza vento forte", i picchi si verificano nei periodi autunnali/invernali, come a novembre con 21.927 incidenti, dicembre con 16.148 e gennaio con 14.737 incidenti.

Spostandoci a destra, sull'istogramma a colonne in pila, è possibile osservare come varia la severità degli incidenti in base alle condizioni stradali, le quali dipendono generalmente dalle condizioni meteorologiche. Si nota che la severità media più alta si verifica quando l'asfalto è "Allagato oltre 3 cm di profondità" ("Flood over 3 cm deep"), con un valore medio di 1,18, seguito dall'asfalto "Asciutto" ("Dry") e "Bagnato o umido" ("Wet or damp"), entrambi con una media di 1,16. L'asfalto con "Ghiaccio o brina" ("Frost or ice") registra una severità media di 1,14, mentre l'asfalto con "Neve" ("Snow") presenta il valore più basso, pari a 1,11. Questi dati suggeriscono che, le condizioni meteorologiche più estreme, come la neve o il ghiaccio, non sempre portano a una severità maggiore rispetto a condizioni meno critiche, come l'asfalto asciutto o umido/bagnato. Questo potrebbe essere dovuto al fatto che, in situazioni estreme come quelle con strada innevata o ghiacciata, le persone tendono a circolare meno e a uscire solo se strettamente necessario, riducendo quindi il numero di veicoli in circolazione e, al contempo, adottando maggiore cautela alla guida.

Infine, dal grafico combinato a linee e istogramma a colonne in pila in basso, si osserva che la distribuzione degli incidenti si concentra principalmente tra le 7:00 e le 20:00, con un picco significativo durante le ore diurne. In particolare, si registrano 102.436 incidenti alle 8:00, 107.820 alle 15:00, 114.280 alle 16:00 e 124.612 alle 17:00, che rappresentano gli orari di punta. Al contrario, il numero più basso di incidenti si registra alle 3:00 e alle 4:00 del mattino, con rispettivamente 8.830 e 7.190 incidenti. Inoltre, è possibile osservare come la maggior parte degli incidenti si verifica in condizioni di luce naturale, mentre le categorie "Darkness - no lighting" e "Darkness - lights lit" emergono principalmente nelle ore serali e notturne, quando la luce del

sole non è più presente. Infine, il rapporto tra il numero totale di vittime e il numero totale di incidenti, calcolato per ogni ora, risulta proporzionato, con valori compresi tra 1,3 e 1,5 vittime per incidente.

Per approfondire la distribuzione degli incidenti e individuare eventuali variazioni significative rispetto all'analisi generale, sono stati applicati diversi filtri ai dati. In particolare, sono stati considerati differenti scenari selezionando specifici intervalli temporali e fasce d'età dei conducenti. Ogni filtro ha consentito di evidenziare dinamiche particolari che non emergevano dall'analisi complessiva.

Di seguito vengono riportate le principali osservazioni ottenute dall'analisi dei vari casi:

- Selezionando i mesi da *novembre* a *gennaio*, il numero totale di incidenti scende a 352.620, mentre il numero totale di vittime si riduce a 482.000. Risulta particolarmente interessante la distribuzione degli incidenti in base alle condizioni di luce per ora: essa è simile a quella del caso generale, ma si osserva un incremento degli incidenti nelle ore successive alle 16 quando le condizioni di luce sono "Darkness - light lit" e "Darkness - no lighting", a causa del tramonto anticipato in quel periodo dell'anno (Figura 33).

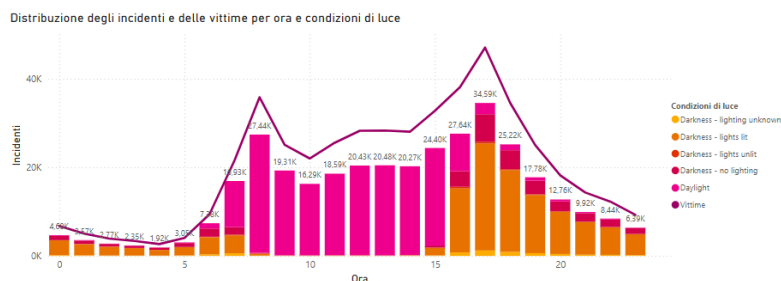


Figura 33: Distribuzione incidenti: novembre - gennaio

Al contrario, selezionando i mesi da *giugno* a *settembre*, naturalmente, si verifica l'effetto opposto: il numero di incidenti in condizioni di oscurità diminuisce drasticamente, poiché in questi mesi il sole è presente fino alle 20/21 di sera (Figura 34).

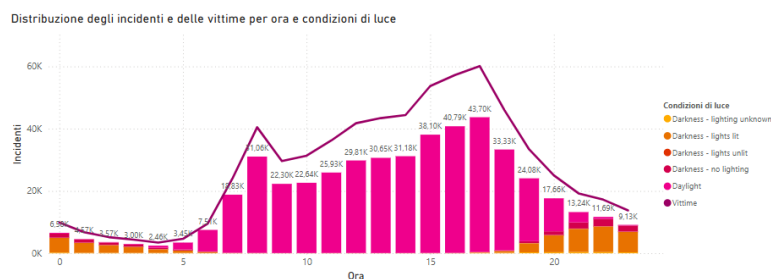


Figura 34: Distribuzione incidenti: giugno - settembre

- Selezionando come filtro l'età del conducente compresa tra *16 e 20 anni*, il numero di incidenti scende a 160.880, con 237.000 vittime. Un aspetto interessante riguarda la distribuzione oraria degli incidenti: in proporzione, si registra una riduzione degli incidenti tra le 9:00 e le 11:00 di mattina, mentre si osserva un aumento tra le 19:00 e le 2:00 di notte, probabilmente a causa di una maggiore attività notturna da parte dei giovani.

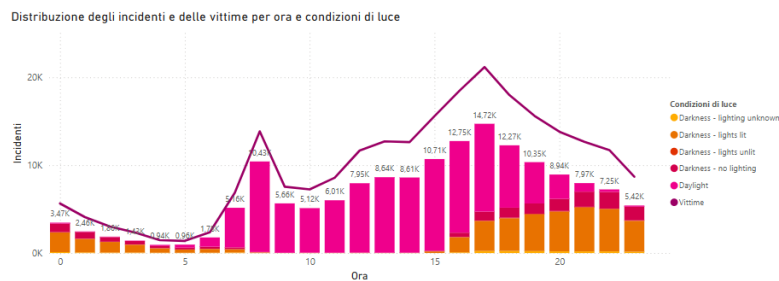


Figura 35: Distribuzione incidenti: 16 - 20 anni

Al contrario, selezionando la fascia d'età tra *66 e 75 anni*, il numero di incidenti e vittime scendono rispettivamente a 85k e 126k, con una distribuzione concentrata prevalentemente nelle ore diurne, tra le 7:00 del mattino e le 20:00 di sera.

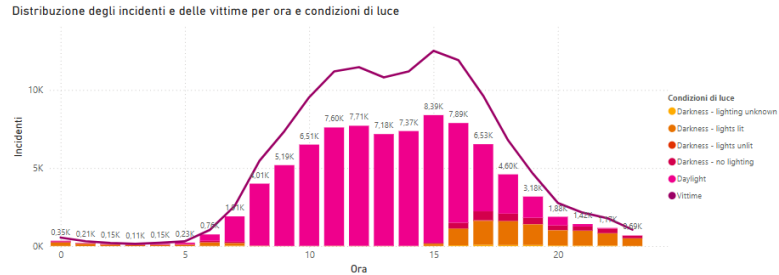


Figura 36: Distribuzione incidenti: 66 - 70 anni

4.3.2 Foglio 2 - Confronto degli incidenti tra automobili e motociclette

La seconda dashboard creata con Power BI e mostrata in *Figura 37*, presenta un confronto tra due tipologie di veicoli: automobili e motociclette. La visualizzazione è suddivisa in due sezioni: la parte sinistra è dedicata agli incidenti che coinvolgono le automobili, mentre la parte destra è focalizzata sulle motociclette. Ogni sezione include diversi oggetti visivi che forniscono informazioni dettagliate sugli incidenti, con lo scopo di evidenziare le differenze principali tra i due veicoli.

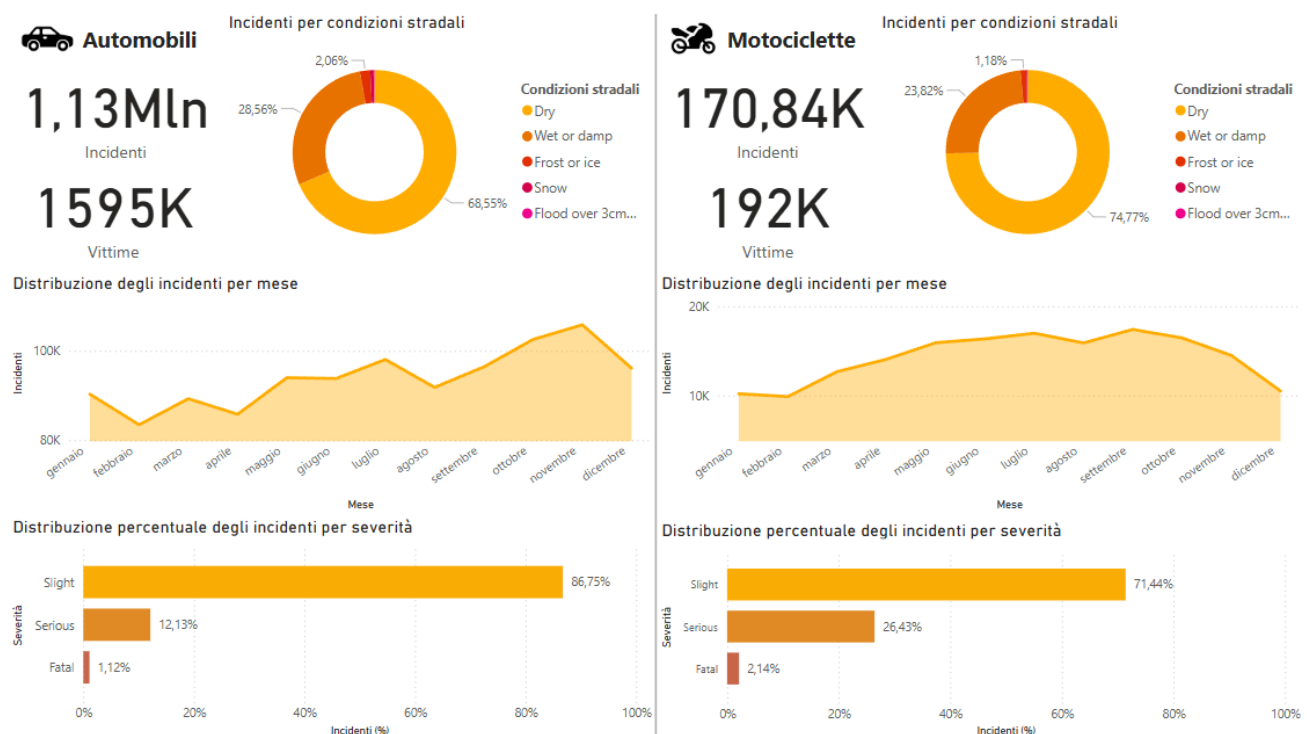


Figura 37: Dashboard 2

Gli oggetti visivi utilizzati sono:

- Un *indicatore di valore* che indica il numero totale di incidenti;
- Un *indicatore di valore* che mostra il numero di vittime coinvolte negli incidenti;
- Un *grafico ad anello* che rappresenta la distribuzione percentuale degli incidenti in base alle condizioni stradali (ad esempio: asciutto, bagnato, nebbioso, ecc.);
- Un *grafico ad aree* che visualizza la distribuzione degli incidenti mensili, evidenziando eventuali tendenze stagionali o periodi di maggiore frequenza;

- Un *grafico a barre in pila* che illustra la divisione degli incidenti in base alla severità (leggeri, gravi, mortali).

Nel dettaglio, partendo dai due indicatori di valore in alto a sinistra, si osserva che le auto registrano 1,13 milioni di incidenti con circa 1.6 milioni di vittime, mentre le motociclette hanno 170.840 incidenti con 192.000 vittime. Il rapporto vittime/incidenti è pari a 1,41 per le auto e a 1,12 per le moto. Da questi dati emerge una netta differenza, sia nei numeri che nel tasso di vittime/incidenti. Tale differenza può essere spiegata dal fatto che le automobili sono più diffuse e, tipicamente, trasportano un numero maggiore di passeggeri, il che contribuisce a un numero maggiore di vittime. Inoltre, le auto sono più esposte a situazioni di traffico intenso, aumentando la probabilità di incidenti, anche di tipo collettivo, come i tamponamenti a catena.

Analizzando il grafico ad anello in alto a destra, si nota che la maggior parte degli incidenti avviene su strada asciutta ("Dry"), con il 68,55% per le automobili e il 74,77% per le motociclette. In seconda posizione troviamo l'asfalto bagnato o umido ("Wet or damp"), che rappresenta rispettivamente il 28,56% e il 23,82% degli incidenti. Gli incidenti su strade ghiacciate o con presenza di brina ("Frost or ice") sono molto meno frequenti, con il 2,16% per le automobili e l'1,18% per le motociclette. Sono poi riportate anche altre condizioni, come la neve ("Snow") e gli allagamenti superiori ai 3 cm ("Flood over 3 cm"), ma con percentuali marginali in entrambi i casi.

Scendendo nel grafico ad aree, si osserva che gli incidenti mensili evidenziano picchi a novembre e dicembre per le auto, probabilmente dovuti alle condizioni invernali più difficili. Per le motociclette, invece, il picco si verifica nel periodo che va da marzo a settembre, con una diminuzione nei mesi più freddi.

Nel complesso, sia nel grafico a torta che in quello a aree, emerge chiaramente che le condizioni climatiche influenzano la presenza di veicoli in circolazione, la distribuzione degli incidenti in base alle condizioni dell'asfalto e la distribuzione mensile degli stessi. Le strade asciutte, tipiche dei mesi primaverili ed estivi, registrano, in proporzione, il maggior numero di incidenti per le motociclette, poiché in questi periodi le moto sono più numerose sulle strade. Per le automobili, invece, il numero di incidenti raggiunge dei picchi nei mesi che vanno da settembre a dicembre, quando aumentano le strade bagnate o con altre condizioni meteorologiche sfavorevoli, poiché, a differenza delle motociclette, le auto sono utilizzate anche in condizioni climatiche meno favorevoli.

Infine, nel grafico a barre in pila sulla distribuzione degli incidenti per severità, si può osservare come le auto sono maggiormente coinvolte in incidenti leggeri, con una percentuale dell'86,75%, probabilmente grazie alla maggiore protezione offerta dalla carrozzeria e dai sistemi di sicurezza. Le motociclette, invece, mostrano una percentuale più alta di incidenti gravi o mortali, pari rispettivamente al 26,43% e al 2,14%, valori circa doppi rispetto alle automobili, che registrano il 12,13% e l'1,12%. Questo è dovuto al fatto che i motociclisti sono più esposti ai pericoli e meno protetti, aumentando significativamente il rischio di conseguenze gravi in caso di incidente.

4.3.3 Utilizzo DAX e M

Per facilitare l'analisi 4.3.2 e consentire un confronto diretto tra automobili e motociclette, è stata utilizzata una funzione in DAX per categorizzare i veicoli in gruppi generici. La funzione implementata assegna il valore "Car" a tutti i veicoli classificati come automobili e "Motorcycle" a tutte le tipologie di motociclette presenti nel dataset. Se il veicolo non rientra in queste categorie, viene mantenuto il valore originale. Questo approccio consente di semplificare la visualizzazione e migliorare l'interpretazione dei dati, raggruppando i veicoli in classi più significative per il confronto tra incidenti.

Il codice DAX utilizzato è il seguente:

```
Generic_Vehicle_Type =
IF(
    'UK_Accident_Vehicles_ETL'[Vehicle_Type] = "Car",
    "Car",
    IF(
        'UK_Accident_Vehicles_ETL'[Vehicle_Type] IN {
            "Electric motorcycle",
            "Motorcycle - unknown cc",
            "Motorcycle 125cc and under",
            "Motorcycle 50cc and under",
            "Motorcycle over 125cc and up to 500cc",
            "Motorcycle over 500cc"
        },
        "Motorcycle",
        'UK_Accident_Vehicles_ETL'[Vehicle_Type] // Mantiene il valore originale
    )
)
```

In alternativa, questa categorizzazione sarebbe potuta essere implementata anche utilizzando il linguaggio M all'interno di Power Query. L'uso di M sarebbe stato utile per trasformare e pulire i dati prima di caricarli in Power BI, permettendo di gestire la classificazione direttamente nella fase di preparazione del dataset.

Il codice equivalente in M per Power Query è il seguente:

```
let
    Source = UK_Accident_Vehicles_ETL,
    AddGenericVehicleType = Table.AddColumn(Source, "Generic_Vehicle_Type", each
        if [Vehicle_Type] = "Car" then "Car"
        else if List.Contains({"Electric motorcycle",
                                "Motorcycle - unknown cc",
                                "Motorcycle 125cc and under",
```

```
        "Motorcycle 50cc and under",  
        "Motorcycle over 125cc and up to 500cc",  
        "Motorcycle over 500cc"}, [Vehicle_Type])  
    then "Motorcycle"  
    else [Vehicle_Type]  
)  
in  
    AddGenericVehicleType
```

L'implementazione in M utilizza la funzione `Table.AddColumn` per aggiungere una nuova colonna `Generic_Vehicle_Type` basata su una logica `if-else`. Questo approccio permette di pre-elaborare i dati prima di caricarli nel modello di Power BI, rendendo il calcolo più efficiente. Tuttavia, l'utilizzo di DAX offre maggiore flessibilità per modifiche in fase di creazione della dashboard.

4.3.4 Foglio 3 - Analisi delle dinamiche degli incidenti

La terza dashboard creata con Power BI e mostrata in *Figura 38* fornisce un'analisi approfondita delle dinamiche degli incidenti. In particolare, vengono esaminati diversi aspetti, come le manovre effettuate, il punto di impatto, l'eventuale uscita del veicolo dalla carreggiata e, in tal caso, in che modo, e se il veicolo è stato soggetto a movimenti come sbandamenti o capovolgimenti.

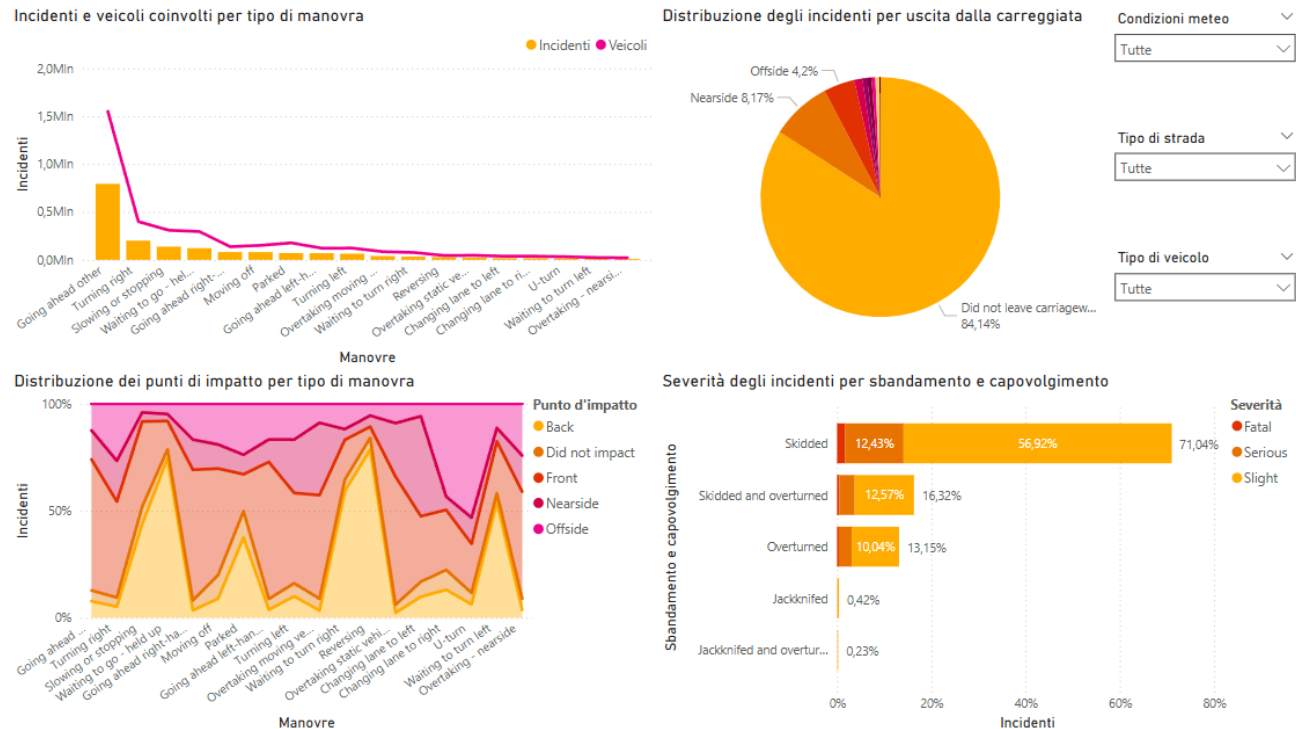


Figura 38: Dashboard 3

La dashboard si compone di quattro oggetti visivi:

- Un *grafico combinato a linee e istogramma*, che mostra rispettivamente il numero di incidenti e il numero di veicoli coinvolti per ogni tipo di manovra;
- Un *grafico a torta*, che rappresenta la distribuzione percentuale di tutte le possibili fuoriuscite dei veicoli dalla carreggiata;
- Un *grafico ad aree in pila al 100%*, che illustra la distribuzione percentuale dei punti di impatto per ogni manovra effettuata;
- un grafico a barre in pila, che mostra la distribuzione percentuale degli incidenti in cui il veicolo, in seguito all'impatto, ha subito un movimento (come sbandamento o capovolgimento), combinata con la percentuale della severità dell'incidente.

Inoltre, all'interno del foglio, per poter eseguire delle analisi più approfondite, sono stati inseriti tre filtri:

- Condizioni meteo;
- Tipo di strada;
- Tipo di veicolo.

Senza applicare alcun filtro, osservando il grafico combinato in alto a sinistra, si nota che, tra i vari tipi di manovre, le più problematiche in termini di incidenti e veicoli coinvolti sono, in prima posizione, "Procedere davanti ad altri veicoli" ("Going ahead other") con 794.214 incidenti e 1.551.825 veicoli coinvolti. A seguire, a distanza, "Svoltare a destra" ("Turning right") con 201.046 incidenti e 399.442 veicoli, e "Rallentare o fermarsi" ("Slowing or stopping") con 138.055 incidenti e 309.202 veicoli. Successivamente, vi sono altre manovre che, seppur con valori inferiori, contribuiscono comunque ad aumentare il numero complessivo di incidenti e veicoli coinvolti, come ad esempio "Aspettare di partire - bloccato" ("Waiting to go - held up"), "Procedere lungo una curva a destra" ("Going ahead right-hand bend"), e così via, fino a raggiungere l'ultimo valore, "Sorpasso - lato passeggero" ("Overtaking - nearside"). Questi dati, oltre a fornirci informazioni sulle principali manovre che generano incidenti e sul numero di veicoli coinvolti, risultano molto utili per studiare il rapporto tra tali informazioni. Infatti, se si calcola il rapporto veicoli/incidenti per ogni manovra, è possibile osservare alcune differenze significative. In particolare, manovre come "Procedere davanti ad altri" (1,95), "Svoltare a destra" (1,99), "Aspettare di partire - bloccato" (2,00) e "Rallentare o fermarsi" (2,24) risultano coinvolgere, per ogni incidente, almeno due veicoli, cosa che non succede con manovre come "Procedere lungo una curva a destra" (1,73), "Procedere lungo una curva a sinistra" (1,68) o "Retromarcia" ("Reversing") (1,58), che generalmente coinvolgono meno di due veicoli per incidente. Queste differenze suggeriscono che alcune manovre, come "Procedere davanti ad altri" e "Rallentare o fermarsi", comportano il coinvolgimento di un numero maggiore di veicoli, probabilmente a causa della loro natura o delle condizioni del traffico. Al contrario, manovre come "Retromarcia" o "Proseguire lungo una curva a sinistra o a destra" tendono a causare incidenti con un minor numero di veicoli coinvolti, poiché si verificano più spesso in situazioni particolari, come la perdita di controllo di un singolo veicolo, collisioni in spazi ristretti o urti contro ostacoli.

Spostandoci nel grafico a torta in alto a destra, possiamo notare come la stragrande maggioranza degli incidenti sia caratterizzata principalmente da veicoli che non escono dalla carreggiata, con l'84,4%. Le uscite dalla carreggiata costituiscono il resto del grafico: tra le principali, troviamo l'8,17% dei veicoli che finisce sul margine sinistro ("Nearside"), mentre il 4,2% esce sul lato destro ("Offside"), spesso a causa di motivi come errori di guida, distrazioni del conducente o condizioni stradali sfavorevoli. Vi sono poi altre uscite dalla carreggiata con valori molto ridotti: Nearside and rebounded, Offside to central reservation, Offside and rebounded,

Straight ahead at junction, Offside onto central reservation + rebounded e Offside - crossed central reservation, tutte con percentuali al di sotto dell'1%.

Nel grafico in basso a sinistra, è possibile osservare la distribuzione dei punti di impatto per le varie manovre. Ad esempio, considerando la manovra "Procedere davanti ad altri veicoli" ("Going ahead other"), si nota che il 61,19% degli incidenti causati da questa manovra ha come punto di impatto la parte frontale dei veicoli ("Front"), seguito dal lato del passeggero ("Nearside") con il 13,55%, dal lato del guidatore ("Offside") con il 12,43% e, infine, nessun impatto ("Did not impact") con il 5,03%. È interessante confrontare questa situazione con quella della manovra "Rallentare o fermarsi" ("Slowing or stopping"), dove le percentuali sono molto diverse. In questo caso, la maggior parte degli incidenti ha come punto di impatto la parte posteriore del veicolo ("Back") con il 43,71%, seguita dalla parte frontale ("Front") con il 39,79%, da nessun impatto ("Did not impact") con l'8,21%, dal lato del passeggero ("Nearside") con il 4,23% e dal lato del guidatore ("Offside") con il 4,06%. Da quanto emerge, risulta evidente che la distribuzione dei punti di impatto varia sensibilmente in base alla manovra effettuata. Questa differenza può essere attribuita alla dinamica specifica di ciascuna manovra e alle circostanze in cui avvengono gli incidenti. Nel caso della manovra "Procedere davanti ad altri" ("Going ahead other"), la preponderanza degli impatti frontali suggerisce che molti di questi incidenti avvengano in situazioni di avanzamento senza ostacoli immediati. Al contrario, nella manovra "Rallentare o fermarsi" ("Slowing or stopping"), la maggiore incidenza di impatti nella parte posteriore è coerente con tamponamenti dovuti a frenate improvvise o rallentamenti in condizioni di traffico intenso.

Infine, esaminando il grafico a barre impilato in basso a destra, si può osservare come la maggior parte degli incidenti in cui i veicoli hanno subito un movimento a seguito di un impatto sia caratterizzata da uno slittamento ("skidded"), con un totale del 71,04%. Seguono gli incidenti "slittato e capottato" ("skidded and overturned") con il 16,32%, "capottato" ("overturned") con il 13,15%, e infine, con valori inferiori all'1%, "accartocciato" ("jackknifed") e "accartocciato e capottato" ("jackknifed and overturned"). Inoltre, per quanto riguarda la gravità degli incidenti, la maggior parte di essi è risultata di lieve entità, seguita da incidenti gravi e, infine, da incidenti fatali, in percentuali più basse. Questo è particolarmente evidente per gli incidenti con slittamento, dove si registra il 56,92% di incidenti lievi, il 12,43% di incidenti gravi e l'1,69% di incidenti fatali.

Utilizzando i filtri, l'analisi diventa ancora più interessante, poiché permette di approfondire lo studio tenendo conto di diverse condizioni meteorologiche, del tipo di strada e del tipo di veicolo. Ad esempio:

- Filtrando per tipo di veicolo e selezionando "Autobus o pullman con almeno 17 posti per passeggeri" (Bus or coach - 17 or more passenger seats), si può notare come il numero di incidenti diminuisca drasticamente, con un picco di 29.940 incidenti per la manovra "Going ahead other" (Figura 39). Tuttavia il rapporto veicoli/incidenti per tipo di manovra rimane però proporzionato rispetto al caso generale.

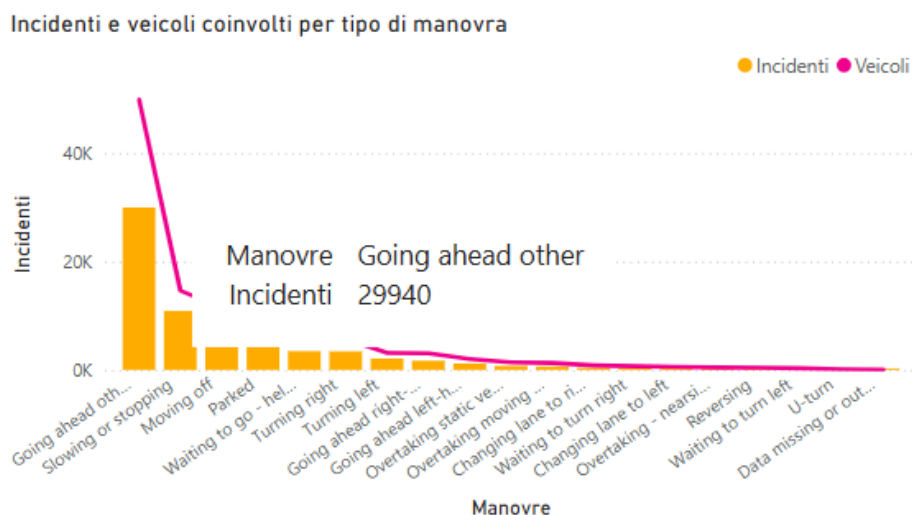


Figura 39: Incidenti e veicoli per manovra - Autobus o pullman

Inoltre, risulta interessante osservare il cambiamento nel grafico a torta in alto a destra (Figura 40), dove la sezione relativa a "Did not leave carriageway" passa dall'84,4% del caso generale al 97,74% in questo caso specifico, seguita da "Nearside" con l'1,47%. Questa variazione può essere spiegata dal fatto che il caso generale include tutti i tipi di veicoli, compresi quelli di piccole dimensioni, più inclini a lasciare la carreggiata in seguito a un urto. Al contrario, gli autobus o pullman, avendo dimensioni maggiori, risultano meno inclini a uscire dalla carreggiata.

Incidenti per uscita dalla carreggiata

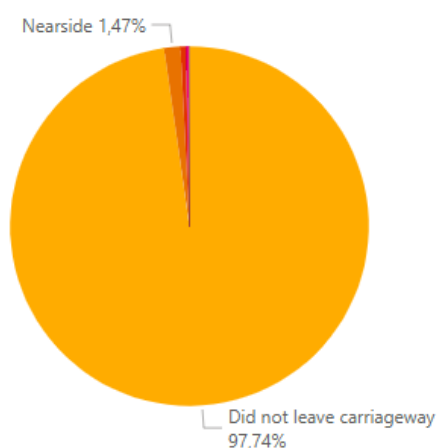


Figura 40: Distribuzione tipo di uscita per incidente - Autobus o pullman

Per quanto riguarda il grafico a barre orizzontali (*Figura 41*), è interessante osservare come la colonna relativa a "slittamento" ("Skidded") aumenti fino al 96,78%, rispetto al 71,04% del caso generale. Questo fenomeno è spiegabile con lo stesso motivo legato alla diversa tipologia di veicoli coinvolti.

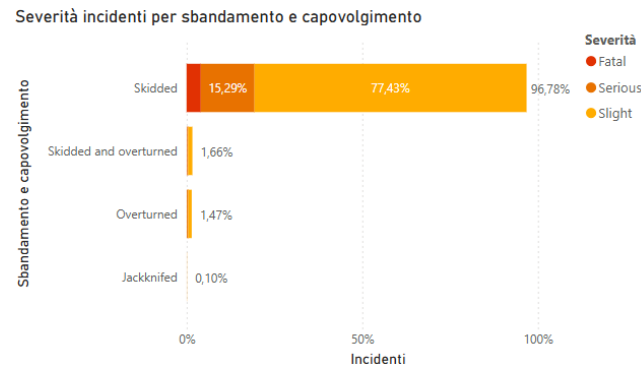


Figura 41: Distribuzione degli incidenti per slittamento nei veicoli tipo autobus

- Se, invece, applichiamo un filtro sulle condizioni meteo (*Figura 42*), selezionando "Neve e alto vento" ("Snowing + High Wind"), si osserva un'importante redistribuzione del diagramma a torta. Il numero di incidenti in cui i veicoli non escono dalla carreggiata si riduce al 59,92%, mentre aumenta la percentuale di incidenti con fuoriuscita: "Lato passeggero" ("Nearside") sale al 21,08%, "Lato del guidatore" ("Offside") al 10,46% e diventa rilevante anche il caso "Lato passeggero e rimbalzo" ("Nearside and rebounded"), che raggiunge il 2,95%.

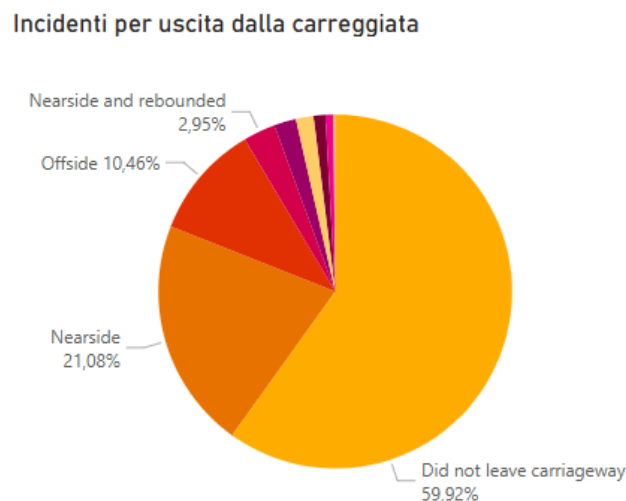


Figura 42: Distribuzione degli incidenti sotto le condizioni di neve e vento forte

4.3.5 Foglio 4 - Predizione del numero di incidenti in base alla gravità

La quarta dashboard creata con Power BI, mostrata nella *Figura 43*, presenta tre analisi predittive relative al numero di incidenti, distribuiti mensilmente per ogni anno e suddivisi in base alla gravità. La dashboard è articolata in tre sezioni, ognuna dedicata a un livello di severità. Oltre a fornire previsioni, consente anche un confronto tra i diversi gradi di gravità. In ciascuna sezione, è presente un *grafico a linee* che include una previsione sull'andamento futuro degli incidenti per i tre anni successivi, da gennaio 2017 a dicembre 2019.

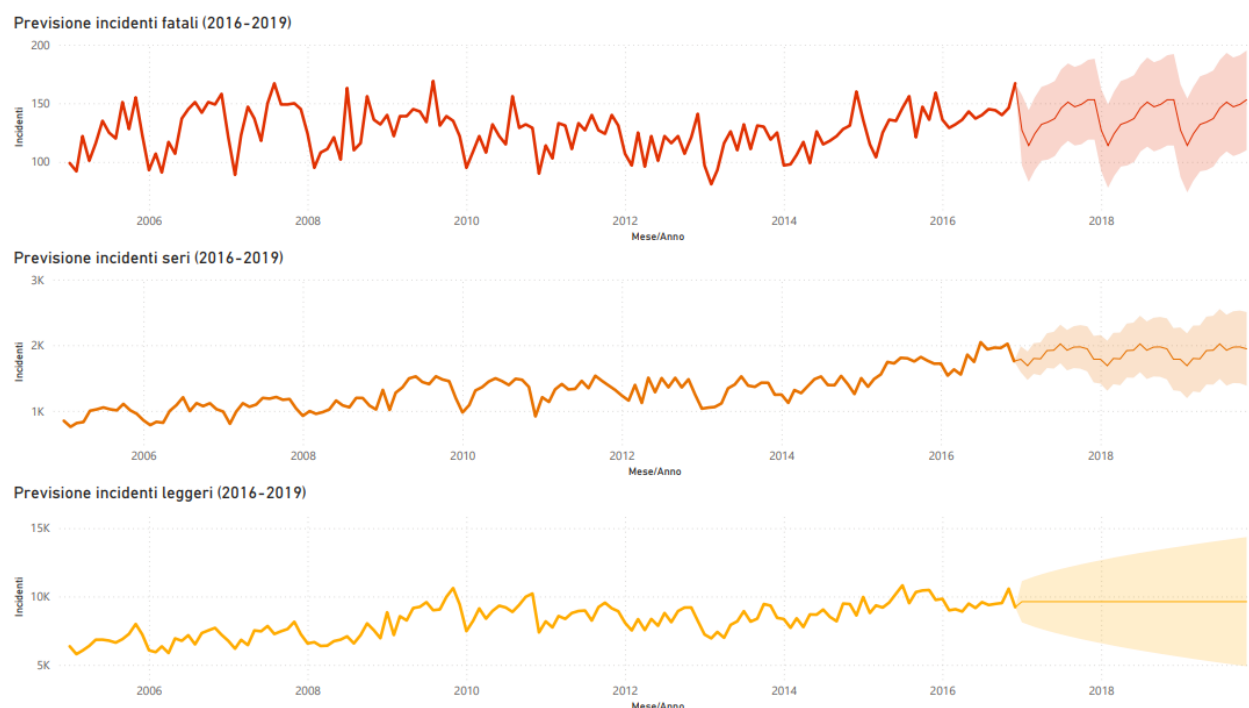


Figura 43: Dashboard 4

Partendo dal grafico in alto, viene mostrata una previsione degli incidenti fatali, ovvero con severità "Fatal". Il grafico evidenzia un trend in cui il numero di incidenti fatali risulta più frequente nei mesi di giugno, luglio e dicembre, mentre è minore nei mesi di febbraio e marzo. Il picco più alto si verifica ad agosto 2009, con 169 incidenti fatali, mentre il picco più basso si osserva a febbraio 2013, con 81 incidenti.

Il grafico centrale mostra la previsione degli incidenti gravi, classificati con severità "Serious". Si osserva un pattern ricorrente: ogni anno, il numero di incidenti inizia con un valore più basso a gennaio, per poi crescere gradualmente fino a raggiungere il picco a dicembre. A partire da gennaio dell'anno successivo, il numero di incidenti tende a diminuire, ma sempre a livelli superiori rispetto al gennaio dell'anno precedente. Questo andamento continua fino al

2016, con un aumento dal numero di incidenti iniziale a gennaio 2006, pari a 849, a quello finale a dicembre 2016, pari a 1.757, con una crescita del 107%. Il picco massimo si registra a luglio 2016, con 2.046 incidenti gravi.

Infine, nel grafico in basso, viene mostrata la previsione degli incidenti lievi, ovvero con severità "Slight". Anche in questo caso, il trend risulta crescente nel corso degli anni, partendo da 6.360 incidenti a gennaio 2006 per arrivare a 9.222 a dicembre 2016, con un aumento del 45%. Il picco massimo si registra a novembre 2009, con 10.615 incidenti.

Da questi tre grafici possiamo osservare come le tre tipologie di severità siano piuttosto sbilanciate: gli incidenti fatali variano tra gli 81 e i 169 incidenti al mese, quelli gravi tra 757 e 2.046, e, infine, quelli lievi tra 5.790 e 10.615. Ciò evidenzia una netta disparità nella distribuzione delle diverse categorie di incidenti.

Analizzando le previsioni per ogni severità, possiamo osservare come i trend identificati negli anni precedenti vengano mantenuti anche nelle stime future. In particolare:

- Per quanto riguarda gli *incidenti fatali*, la previsione indica un andamento simile a quello storico, con picchi durante i mesi estivi e nel periodo invernale, in particolare dicembre.
- Nel caso degli *incidenti seri*, la previsione evidenzia un trend crescente che rispecchia l'andamento osservato negli anni passati. Si nota un incremento graduale del numero di incidenti, con un picco sempre più elevato nei mesi estivi.
- Infine, gli *incidenti lievi* mostrano una previsione più stabile rispetto agli anni precedenti. Sebbene il numero di incidenti abbia seguito una crescita costante dal 2006 al 2016, la proiezione futura indica una sorta di plateau, con una minore variazione dei dati. Tuttavia, l'intervallo di confidenza della previsione si allarga progressivamente, suggerendo un'incertezza crescente negli anni successivi.

Nel complesso, le previsioni evidenziano come gli incidenti seri e fatali seguano trend più definiti e stagionali, mentre gli incidenti lievi tendano a stabilizzarsi nel lungo periodo, sebbene continuino a crescere. Inoltre, se le condizioni attuali dovessero rimanere invariate, il numero di incidenti seri e fatali continuerà ad aumentare nel tempo. Ciò sottolinea la necessità di interventi mirati per contenerne l'aumento, con un focus specifico sui periodi più critici individuati.

Elenco delle figure

1	Logo di Qlik	7
2	Trasformazione tipo di dato per la colonna Local_Authority_(Highway)	8
3	Sostituzione dei valori di Accident_Severity	8
4	Collegamento dei dataset	9
5	Foglio di analisi numero 1	10
6	Foglio di analisi numero 1 con filtro per gravità	12
7	Foglio di analisi numero 1 con filtro per gravità	12
8	Foglio di analisi numero 1 con filtro per gravità	13
9	Foglio di analisi numero 2	14
10	Foglio di analisi numero 3	16
11	Foglio di analisi numero 3 - Condizioni della strada "frost or ice"	18
12	Foglio di analisi numero 3 - Condizioni della strada "flood"	18
13	Foglio di analisi numero 3 - Condizioni della strada "snow"	19
14	Foglio di analisi numero 3 - Applicazione filtro per motocicli	20
15	Foglio di analisi numero 3 - Applicazione filtro per automobili	21
16	Foglio di analisi numero 3 - Applicazione filtro per mezzi di trasporto a grande capienza	22
17	Logo di Tableau	23
18	Relazione dei due dataset	24
19	Dashboard di analisi numero 1	25
20	Dashboard di analisi temporale filtrata per incidenti "slight"	27
21	Distribuzione settimanale degli incidenti con severità "serious"	28
22	Distribuzione settimanale degli incidenti con severità "fatal"	28
23	Distribuzione oraria degli incidenti "serious"	29
24	Distribuzione oraria degli incidenti "fatal"	29
25	Dashboard di analisi numero 2	30
26	Dashboard di analisi numero 3	32
27	Dashboard di analisi numero 4	34
28	Descrizione della previsione fornita da Tableau	35
29	Descrizione dei modelli utilizzati per la previsione fornita da Tableau	36
30	Logo di Power BI	39
31	Relazione dataset	40
32	Dashboard 1	41
33	Distribuzione incidenti: novembre - gennaio	43
34	Distribuzione incidenti: giugno - settembre	43
35	Distribuzione incidenti: 16 - 20 anni	44
36	Distribuzione incidenti: 66 - 70 anni	44
37	Dashboard 2	45

38	Dashboard 3	49
39	Incidenti e veicoli per manovra - Autobus o pullman	52
40	Distribuzione tipo di uscita per incidente - Autobus o pullman	52
41	Distribuzione degli incidenti per slittamento nei veicoli tipo autobus	53
42	Distribuzione degli incidenti sotto le condizioni di neve e vento forte	53
43	Dashboard 4	54

Elenco delle tabelle

1	Colonne del dataset <i>Accident_Information.csv</i>	3
2	Colonne del dataset <i>Vehicle_Information.csv</i>	4