

PROGETTO DATA SCIENCE

NASA WILDFIRES

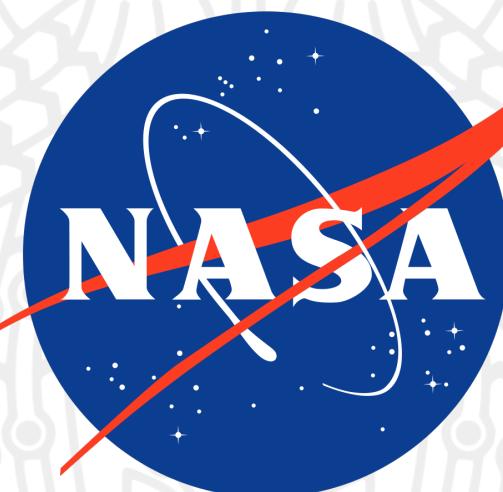
Professore:

Domenico Ursino

Dipartimento di Ingegneria dell'Informazione

Corso: Data Science

Università Politecnica delle Marche, Ancona



Di:

Claudio Menotta

Mattia Ippoliti

Italo Cervigni

Indice

1	Introduzione	1
1.1	Storia degli Incendi negli USA	1
1.2	Dataset	3
1.3	Operazioni di ETL	5
1.3.1	Operazioni ETL in Python	6
1.3.2	Operazioni ETL in Power BI	7
2	Qlik Sense	9
2.1	Gestione dati	9
2.2	Panoramica sulle dimensioni dell'incendio	12
2.2.1	KPI	13
2.2.2	Mappa delle dimensioni	13
2.2.3	Espansione per Stato	14
2.2.4	Numero di incendi per Stato	15
2.3	Analisi sulla Potenza irradiata dall'incendio	15
2.3.1	KPI	16
2.3.2	Analisi della Potenza irradiata per singolo Stato	16
2.3.3	Stati con gli incendi peggiori degli ultimi 10 anni	17
2.3.4	Mappa della Potenza irradiata negli USA	17
2.3.5	Luminosità degli incendi divisa per Stato	19

INDICE

2.4	Analisi per Stato	19
2.4.1	KPI	20
2.4.2	Analisi Espansione per Mese	20
2.4.3	Mappa della brillantezza degli incendi negli USA	21
2.4.4	Analisi della Potenza irradiata media	22
2.4.5	Analisi dell'Espansione negli anni	23
2.5	Analisi sul numero di incendi per Anno e per Mese	24
2.5.1	KPI	24
2.5.2	Numero di incendi per anni	25
2.5.3	Numero di incendi per Anni/Mese	25
2.5.4	Numero di incendi per mesi	26
2.6	Analisi Numero di incendi per Stato e per Mese	27
2.6.1	Mappa di calore sul numero di incendi	27
2.7	Analisi Trimestri con Medie	28
2.7.1	Numero di incendi per Anno e per Trimestre	29
2.7.2	Numero di incendi per Trimestre	30
2.7.3	Media della Brillantezza, Espansione, Potenza irradiata	31
2.8	Incendi peggiori negli ultimi dieci anni	31
2.8.1	Tabella peggiori incendi	32
2.8.2	Diagramma a dispersione peggiori incendi	33
2.8.3	Mappe di Potenza irradiata e di Espansione dei peggiori incendi	33
2.9	Analisi per Tipo	35
2.9.1	Analisi del numero di incendi, della brillantezza media e dell'espansione media per Stato	35
2.9.2	Parametri a confronto per Tipo	36
2.10	Analisi per Satellite	37
2.11	Analisi Giorno/Notte e Terra/Aqua	37

INDICE

2.12 Analisi Giorno/Notte, Terra/Aqua e Tipo	38
2.13 Analisi sulle Confidenze	39
3 Tableau	41
3.1 Gestione dati	41
3.2 Città con maggiori incendi	42
3.2.1 Città con maggior numero di incendi	42
3.2.2 Tipi di incendi per le 15 città con maggior numero di incendi	43
3.2.3 Mappa delle città con maggior incendi	44
3.2.4 Città con maggior numero di incendi negli anni	44
3.3 Incendi durante il periodo del Covid	45
3.3.1 Confronto numero medio di incendi nel periodo del Covid e negli anni precedenti	46
3.3.2 Confronto numero medio di incendi nel periodo del Covid e negli anni precedenti con istogrammi	46
3.3.3 Confronto mappe incendi tra il periodo del Covid e gli anni precedenti	47
3.4 Analisi del trimestre con più incendi	48
3.4.1 Trimestre negli ultimi 10 anni con più incendi	49
3.4.2 Analisi per Stati del terzo trimestre del 2017	49
3.4.3 Numero di incendi per città del terzo trimestre del 2017	50
3.4.4 Andamento nelle 3 località più colpite del 3° trimestre del 2017	51
3.5 Analisi per tipo	51
3.5.1 Numero di incendi per tipo	52
3.5.2 Tabellen per tipo	52
3.5.3 Caratteristiche dei tipi di incendi	54
3.5.4 Incendi peggiori per tipo	55
3.6 Previsione numero di incendi fino al 2020	55

INDICE

3.6.1	KPI	56
3.6.2	Previsione numero di incendi fino al 2020	57
3.6.3	Numero di incendi reali nel terzo trimestre del 2020	57
3.6.4	Previsione numero di incendi per trimestre fino al 2021	58
3.7	Previsioni incendi nel 2016 e nel 2019	58
3.7.1	Confronto tra previsione del numero di incendi nel 2016 e dati reali	59
3.7.2	Confronto tra previsione del numero di incendi nel 2019 e dati reali	60
3.8	Previsione scan, frp e brightness fino al 2020	61
3.8.1	KPI previsioni	61
3.8.2	Previsione scan, frp e brightness fino al 2021	62
3.8.3	Scan, frp e brightness nel terzo trimestre del 2020	62
3.9	Cluster degli Stati	63
3.9.1	Cluster per Stato, istogrammi	63
3.9.2	Cluster per Stato, mappa	65
3.10	Cluster per mese e città	65
3.10.1	Cluster per mese	66
3.10.2	Cluster per città	67
3.10.3	Cluster sulla gravità degli incendi	67
4	Power BI	69
4.1	Introduzione al tool	69
4.2	Preparazione del dataset	69
4.3	Dashboard di tipo descrittivo e diagnostico	70
4.3.1	Panoramica per stato	70
4.3.2	Analisi per città	74
4.3.3	Analisi per ora	77
4.4	Dashboard R-Based	78

INDICE

4.4.1	Forecasting with ARIMA	79
4.4.2	Forecasting TBATS	80
4.4.3	Correlazione degli attributi	81
4.4.4	Albero decisionale	83
4.4.5	Clustering per state	83
4.4.6	Clustering per city	84
4.4.7	Clustering per mese	85
5	Conclusioni	87
	Bibliografia	89

1. Introduzione

Il progetto su cui è incentrata questa relazione consiste nell’analisi di un dataset relativo agli incendi avvenuti negli USA negli ultimi 10 anni, ovvero da settembre 2010 fino a settembre 2020. Le analisi sono avvenute sotto tutti gli aspetti, cioè quello descrittivo, diagnostico, predittivo e prescrittivo, e sono state effettuate con i tool di business intelligence messi a disposizione dal nostro ateneo e leader nel mercato, come riportato dal *Magic Quadrant Gartner*, che sono Qlik, Tableau, e Power BI.

Il progetto si è sviluppato in diverse fasi, partendo dalla scelta del dataset, che è stato trovato all’interno di Kaggle, una comunità di supporto per ricercatori e data scientists che mette a disposizione una grande varietà di dati per diversi usi, che vanno dalla classificazione, regressione, alla sentiment analysis e molto altro. Una volta scelto il dataset, si sono analizzate le informazioni che esso metteva a disposizione, e si è cercato di definire gli obiettivi della nostra analisi. Poi si è passato alla fase di ETL dei nostri dati, che ha comportato l’inserimento di alcune colonne e la modifica di alcuni valori. Infine, una volta che il dataset è stato reso utilizzabile e ripulito da dati sporchi, si è passati alle analisi vere e proprie, utilizzando prima Qlik per analisi descrittive e diagnostiche, poi Tableau e Power BI per aggiungere anche analisi predittive e prescrittive.

1.1 Storia degli Incendi negli USA

Gli incendi possono verificarsi in molti luoghi negli Stati Uniti, specialmente durante la siccità, ma sono più comuni negli Stati Uniti occidentali e in Florida. Possono essere innescati naturalmente, più comunemente da fulmini o da attività umane come materiali da fumo non estinti, apparecchiature elettriche difettose, automobili surriscaldate o incendio doloso (incendi di tipo 3 del dataset). La politica di gestione degli incendi ha favorito la soppressione aggressiva degli incendi a partire dall’inizio del XX secolo. Nel

CAPITOLO 1: INTRODUZIONE

21 ° secolo, l'aumento della temperatura e la siccità causati dal riscaldamento globale sono diventati più preoccupanti.

In particolare, la California d'estate ha un clima particolarmente caldo e secco, che spesso provoca vasti incendi, che sono spesso aggravati da forti venti secchi. Questi incendi, che si propagano principalmente nelle foreste e nelle riserve, possono provocare decine di morti durando anche diversi mesi e arrivando a minacciare i centri abitati più vicini. Le contee californiane più colpite risultano essere Orange, Riverside, San Bernardino e Los Angeles.

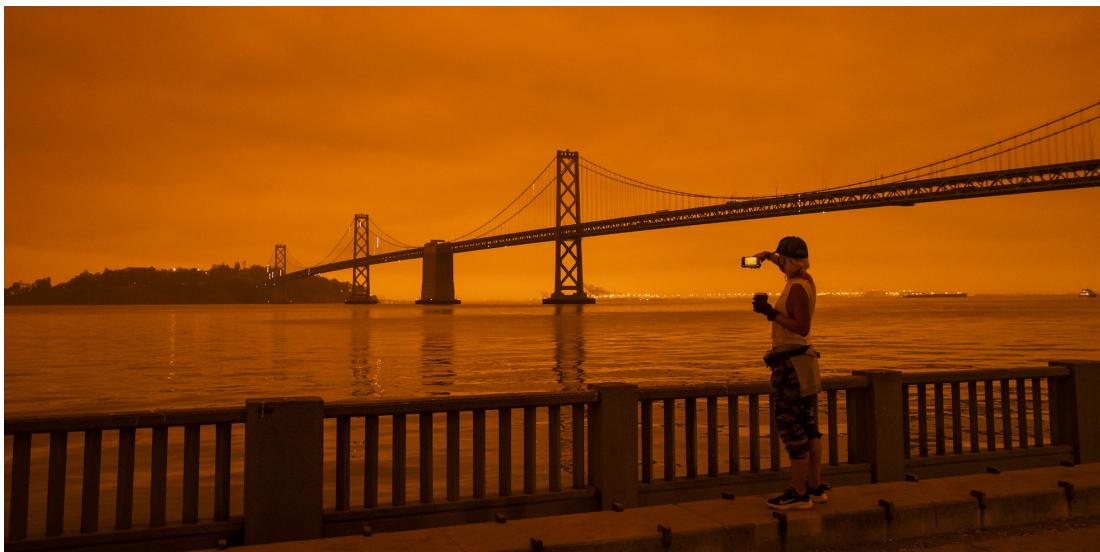


Figura 1.1: Incendi del 16 settembre 2020 nella città di San Francisco, California

Anche se all'apparenza può sembrare assurdo, l'Alaska è uno degli stati più colpiti, dove, secondo i rapporti di Berkeley Earth, la temperatura media in Alaska è aumentata di circa 2 °C tra il 1970 e il 2005. L'Alaska è perciò una delle regioni del mondo in cui il riscaldamento globale è più evidente. A causa della crisi climatica, gli incendi boschivi nelle aree forestali boreali hanno raggiunto livelli di scala e frequenza in Alaska che non sono stati raggiunti per almeno 10 000 anni. Nel 2019 l'Alaska è stata colpita, al pari di Siberia e Groenlandia, da vasti incendi boschivi, causati un riscaldamento anomalo delle zone artiche.

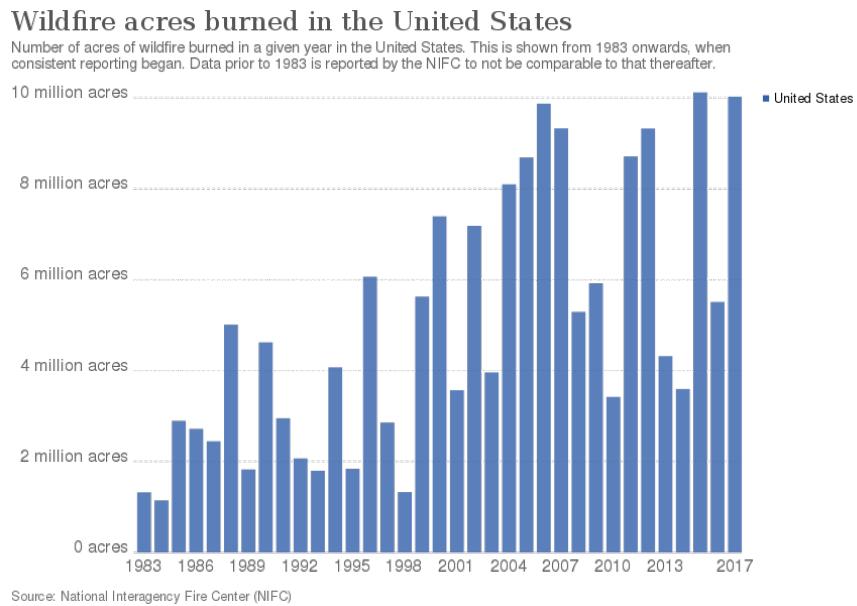
Così come California ed Alaska, gli stati più colpiti dagli incendi sono il Texas, l'Oregon, la Georgia e il Florida, dove negli ultimi anni si sono viste decine e decine di morti.

I cambiamenti climatici sono già in atto e stanno comportando un aumento costante delle temperature globali, con variazioni dei modelli di precipitazioni con la conseguenza di

CAPITOLO 1: INTRODUZIONE

un aumento del rischio legato ad eventi meteorologici e climatici estremi. Gli effetti varieranno in base alle regioni e alle località, ma in combinazione con gravi siccità e ondate di calore, gli incendi boschivi hanno già conseguenze catastrofiche.

Nella seguente foto si può vedere come sta cambiando il fenomeno degli incendi negli ultimi quarant'anni circa [1].



1.2 Dataset

Come già detto in precedenza il dataset scelto riguarda gli incendi avvenuti negli USA tra settembre 2010 e settembre 2020, raccolti dalla NASA attraverso i propri satelliti, ovvero MODIS e VIIRS. In particolare l'intero dataset è composto da due file *CSV* di circa 95MB, il primo con i dati riguardanti gli incendi rilevati dal 26 Settembre 2010 fino al 30 Giugno 2020, mentre il secondo con i dati dal 1 Luglio 2020 fino al 27 Settembre 2020. I dati sono divisi in due file differenti proprio perché sono stati raccolti con due tecnologie diverse, e nello specifico nel primo file i dati sono relativi al satellite MODIS, mentre nel secondo al satellite VIIRS. Gli attributi raccolti dai due presentano alcune differenze. Nel dataset ogni riga corrisponde ad un incendio rilevato, che può essere di varia natura, ovvero può essere un incendio di vegetazione, eruzione vulcanica, esplosione avvenuta in mare e incendio di altra natura (come incendi di case), e per ciascuno di esso sono rilevati 13 attributi diversi. Di seguito sono riportati gli attributi dei due datset

CAPITOLO 1: INTRODUZIONE

sopra descritti:

MODIS - VIIRS	
Caratteristiche	Descrizione
Latitude	Centro di 1 km del pixel di incendio ma non necessariamente la posizione effettiva dell'incendio poiché uno o più incendi possono essere rilevati entro il pixel di 1 km.
Longitude	Centro di 1 km del pixel di incendio ma non necessariamente la posizione effettiva dell'incendio poiché uno o più incendi possono essere rilevati entro il pixel di 1 km.
Brigthness (Kelvin)	La temperatura di colore (o luminosità) è una grandezza fisica associata alla tonalità della luce: a ogni temperatura corrisponde una sola tonalità ed è misurata in Kelvin.
Scan e Track	Pixel di incendio di 1 km ma i pixel MODIS diventano più grandi verso il bordo della scansione. Queste riflettono le dimensioni effettive dei pixel.
Acq Date	Data di acquisizione MODIS.
Acq Time	Ora di acquisizione del satellite (in UTC).
Satellite	A = Aqua e T = Terra.

MODIS - VIIRS	
Caratteristiche	Descrizione
Confidence	Le stime di confidenza variano tra 0 e 100% e vengono assegnate a una delle tre classi di incendio (fuoco a bassa confidenza, fuoco a confidenza nominale o fuoco a confidenza alta)
Bright T31 (Kelvin)	Temperatura di calore o luminosità del canale 31 del pixel dell'incendio misurata in Kelvin.
FRP (MW)	Raffigura la potenza radiativa dell'incendio integrata nei pixel in MW (megawatt).
Type	0 = presunto incendio della vegetazione 1 = vulcano attivo 2 = altra fonte di terra statica 3 = offshore
DayNight	D = fuoco diurno, N = fuoco notturno

1.3 Operazioni di ETL

Prima di passare all'analisi vera e propria si è svolta una fase di ETL con l'obiettivo di rendere i dati coerenti e aggiungere alcuni valori che saranno poi utili per la rappresentazione su mappa e per l'analisi spaziale degli incendi. In particolare la fase di ETL è stata svolta in due modi diversi, dove una prima parte è stata fatta in Python e una seconda parte è stata fatta direttamente in Power BI.

1.3.1 Operazioni ETL in Python

Innanzitutto si sono caricati i due dataset su Python, e si è subito verificato la presenza di valori nulli al loro interno, con le funzioni `isnull.sum()` e `info()`.

Dal risultato si nota che all'interno dei due dataframe non vi sono valori nulli, e perciò non è stato necessario andare ad inserire nessun nuovo valore. Invece si è notato che per ciascun incendio vi erano le informazioni di Latitudine e Longitudine esatta del pixel individuato, ma non c'erano alcuna indicazione sullo stato e sulla città in cui esso è avvenuto. Perciò, anche in ottica delle future analisi nel quale ci si potrà concentrare su stati o città particolari, si è deciso di aggiungere ai dataset due colonne: una che indichi lo stato e una la città, all'interno degli USA, nel quale l'incendio è avvenuto, utilizzando appunto i campi di Latitudine e Longitudine, con l'ausilio della libreria *reverse geocoder*. Ciò è stato fatto utilizzando il codice sotto riportato:

Listing 1.1: Add City and State

```
import pandas as pd
import reverse_geocoder as rg
import pprint

df = pd.read_csv("data2.csv")

def reverseGeocode(coordinates):
    result = rg.search(coordinates)
    return (result)

if __name__ == "__main__":
    # Coordinates tuple. Can contain more than one pair.
    coordinates = list(zip(df['latitude'], df['longitude']))
    data = reverseGeocode(coordinates)

    df['State'] = [i['State'] for i in data]
    df['City'] = [i['City'] for i in data]

df.to_csv("fire_archive.csv")
```

Infine l'ultima operazione svolta sui due dataset è stata aggiungere una colonna che

sembra essere superflua, ma in realtà è molto utile ai fini dell'analisi. E' stata infatti aggiunta una colonna di soli 1, denominata *wildfire_presence*, che indica appunto la presenza dell'incendio. Ciò è stato indispensabile per poter effettuare il conteggio degli incendi, magari anche con filtri per luoghi o periodi particolari, poiché i tool di business intelligence utilizzati non prevedono il conteggio delle righe del dataset che soddisfano particolari caratteristiche. Ciò è stato fatto utilizzando il seguente codice:

Listing 1.2: Add the wildfire presence

```
import pandas as pd

df1 = pd.read_csv("fire_2020.csv")

df1['wildfire_presence'] = 1

df1.to_csv("2020.csv") # write to csv
```

1.3.2 Operazioni ETL in Power BI

In Power BI è stato necessario svolgere un'ulteriore fase di ETL, che non era stata fatta precedentemente perché in Qlik e Tableau non è stata utilizzata la colonna *acq_time*, che indica appunto l'ora di acquisizione dei dati relativi all'incendio. Questo perché nei due tools non è presente il tipo di dato Ora, e perciò il suo utilizzo rimaneva molto complicato, cosa che invece è presente in Power BI.

Il primo problema che si è presentato è stata la conversione da tipo *Stringa* a tipo *Ora*. Infatti non tutte le stringhe erano uguali e della stessa lunghezza, poiché l'ora di acquisizione poteva essere formata da uno o due caratteri, a seconda che facesse parte del range 0 – 9 o 10 – 24, proprio perché nel primo caso non viene indicato lo zero più significativo. Ciò non rende possibile la conversione nel tipo *Ora*, che trasforma in tale formato solo stringhe di 4 caratteri. Perciò si è costruita una nuova colonna condizionale, che inserisce uno 0 come primo carattere della stringa nel caso la sua lunghezza fosse uguale a tre. Tale colonna è stata poi convertita nel tipo *Ora* e sostituita alla vecchia *acq_time*.

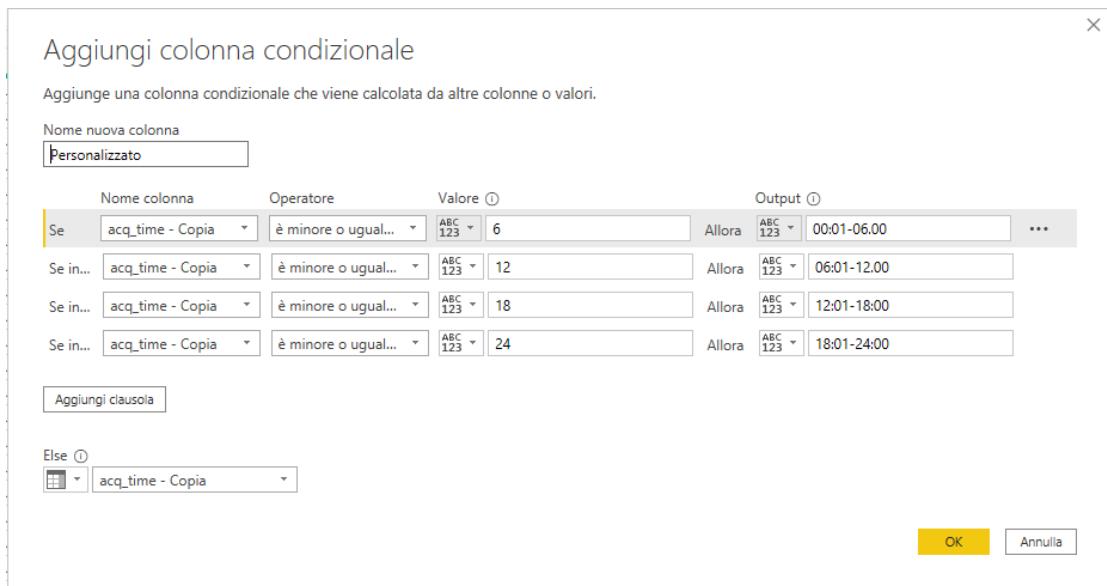
Inoltre, sempre considerando le analisi future, si è deciso costruire una nuova colonna, denominata *time_window*, nella quale si va ad inserire per ciascun incendio a quale

CAPITOLO 1: INTRODUZIONE

fascia oraria della giornata appartiene. La giornata è stata divisa in 4 fasce orarie, di ampiezza 6 ore:

- 00 : 01 – 06 : 00
- 06 : 01 – 12 : 00
- 12 : 01 – 18 : 00
- 18 : 01 – 24 : 00

e ciò è stato fatto creando una nuova colonna condizionale, composta da 4 condizioni, una per fascia oraria, come riportato nella seguente immagine.



Infine, sempre in ottica di future analisi, era interessante ottenere semplicemente l'ora dell'incendio, non includendo quindi i minuti. Perciò si è costruita una colonna *hour* duplicando la colonna di *acq_time*, la quale poi è stata convertita in *Stringa* e per ogni riga è stato sostituito il carattere ":" con ",", ed infine la colonna è stata convertita in *Intero*, applicando così un troncamento alla parte decimale dei valori di ciascuna riga.

2. Qlik Sense

Qlik è un software di Business Intelligence che consente di poter usufruire della totalità delle informazioni presenti nei database, grazie a potenti operazioni di calcolo ed aggregazione che aggiornano istantaneamente i dati durante la navigazione dei report.

Grazie a questa funzionalità i dati caricati da molteplici e differenti sorgenti (file xls, datawarehouse, query sql etc.) vengono immediatamente integrati in Qlik e sono sempre disponibili all'interno dell'applicazione qualsiasi interrogazione l'utente performi, senza necessità di avere delle porzioni di dati precaricati o delle query specifiche salvate a seconda delle esigenze di navigazione dell'utilizzatore finale.

In Qlik l'interesse principale è nel fare un analisi descrittiva e diagnostica del problema, mentre è solo attraverso software come Tableau e PowerBI che si introdurrà anche una parte di analisi predittiva.

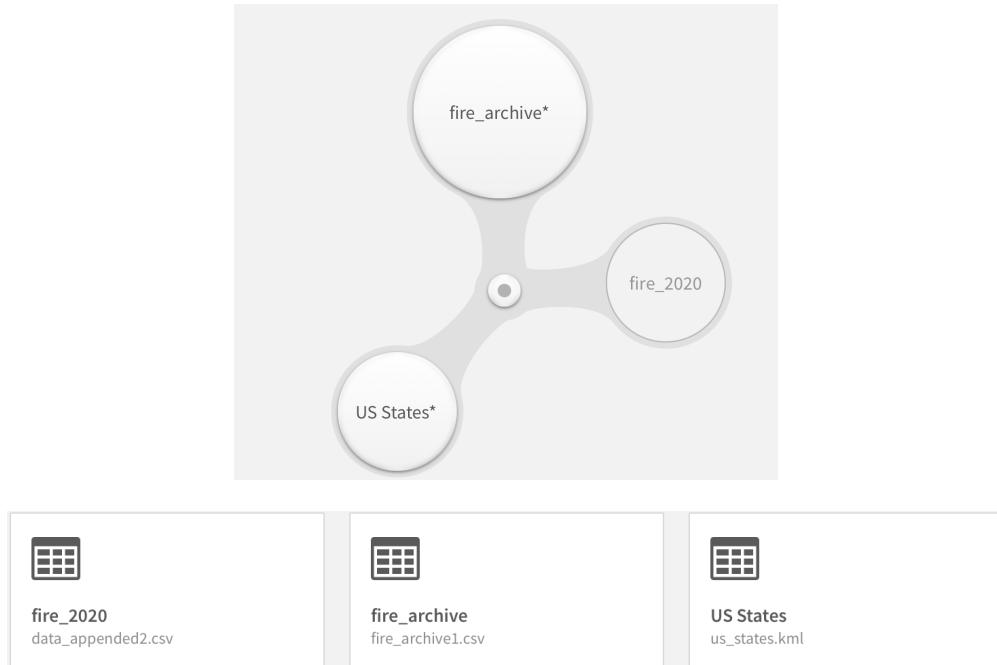
2.1 Gestione dati

La prima parte nell'utilizzo di Qlik Sense consiste nel caricare i dati all'intero del "Gestore di Dati", dove esistono due viste:

- **Associazioni:** È possibile creare e modificare l'associazione tra le tabelle.
- **Tabelle:** Presenta una panoramica di tutte le tabelle dati nell'app, sia che siano state aggiunte mediante Aggiungi dati o che siano state caricate con lo script di caricamento dei dati. Ciascuna tabella viene visualizzata con il nome della tabella, il numero dei campi di dati e il nome della sorgente dati.

I tre dataset sono strutturati nel seguente modo all'interno di Qlik Sense:

CAPITOLO 2: QLIK SENSE



- US States:

- **Fire Archive:**

CAPITOLO 2: QLIK SENSE

fire_archive.latitude	fire_archive.longitude	fire_archive.brightness	fire_archive.scan	fire_archive.track	fire_archive.acq_date	fire_archive.acq_time	fire_archive.satellite	fire_archive.instrument
18.9823	-155.6211	318.9	1.4	1.2	2014-08-01	2047	Terra	MODIS
18.9823	-155.6112	317.8	1.9	1.3	2015-09-02	2323	Aqua	MODIS
18.9838	-155.6408	347.5	1.3	1.1	2014-06-06	0	Aqua	MODIS
18.9854	-155.6281	339.6	1.3	1.1	2014-06-06	0	Aqua	MODIS
19.0141	-155.5965	326.8	1.1	1.1	2017-09-22	2112	Terra	MODIS
19.015	-155.8226	315.6	1.0	1.0	2019-06-26	2100	Terra	MODIS
19.0158	-155.6072	330.4	1.1	1.1	2017-09-22	2112	Terra	MODIS
19.0218	-155.5844	327.4	1.1	1.1	2017-09-22	2112	Terra	MODIS
19.0235	-155.5951	333.3	1.1	1.1	2017-09-22	2112	Terra	MODIS
19.0238	-155.6024	368.3	1.7	1.3	2017-09-22	981	Terra	MODIS
19.0244	-155.5874	307.5	1.8	1.3	2017-09-22	1146	Aqua	MODIS
19.0256	-155.5864	367.8	1.7	1.3	2017-09-22	981	Terra	MODIS
19.0258	-155.5966	338.8	3.1	1.7	2017-09-21	2029	Terra	MODIS
19.0269	-155.6002	321.3	3.1	1.7	2017-09-21	2029	Terra	MODIS
19.0279	-155.5879	329.8	1.1	1.1	2017-09-22	2112	Terra	MODIS
19.0356	-155.5759	342.8	1.1	1.1	2017-09-22	2112	Terra	MODIS
19.0642	-155.5722	310.1	1.2	1.1	2017-02-07	830	Terra	MODIS
19.0659	-155.5614	307.2	1.2	1.1	2017-02-07	830	Terra	MODIS
19.1352	-155.5252	324.6	1.4	1.2	2015-05-04	2331	Aqua	MODIS
19.1374	-155.5122	357.7	1.4	1.2	2015-05-04	2331	Aqua	MODIS
19.1571	-155.8903	315.9	2.5	1.5	2015-08-22	2036	Terra	MODIS
19.1683	-155.4702	378.0	1.8	1.3	2012-06-19	817	Terra	MODIS
19.1713	-155.4563	370.3	1.8	1.3	2012-06-19	817	Terra	MODIS
19.1801	-155.4721	348.6	1.8	1.3	2012-06-19	817	Terra	MODIS
19.1831	-155.4549	359.7	1.8	1.3	2012-06-19	817	Terra	MODIS

fire_archive.confidence	fire_archive1.bright_t31	fire_archive1.frp	daynight	fire_archive.type	City	State	wildfire_pres...	fire_archive.longitude,...	fire_archive.id
34	295.4	11.0	D	0	Hawaiian Ocean View	Hawaii	1	[-155.6211,18.9823]	447808
0	300.2	20.6	D	0	Hawaiian Ocean View	Hawaii	1	[-155.6112,18.9823]	605973
94	304.7	65.2	D	0	Hawaiian Ocean View	Hawaii	1	[-155.6408,18.9838]	434206
85	303.9	43.9	D	0	Hawaiian Ocean View	Hawaii	1	[-155.6281,18.9854]	434201
42	301.7	19.4	D	0	Hawaiian Ocean View	Hawaii	1	[-155.5965,19.0141]	868143
48	308.2	9.4	D	3	Hawaiian Ocean View	Hawaii	1	[-155.8226,19.0151]	1671837
69	301.3	24.7	D	0	Hawaiian Ocean View	Hawaii	1	[-155.6072,19.0158]	868142
23	305.8	28.7	D	0	Hawaiian Ocean View	Hawaii	1	[-155.5844,19.0218]	868137
74	305.8	31.0	D	0	Hawaiian Ocean View	Hawaii	1	[-155.5951,19.0235]	868141
68	284.6	19.6	N	0	Hawaiian Ocean View	Hawaii	1	[-155.6024,19.0238]	867910
55	283.8	22.3	N	0	Hawaiian Ocean View	Hawaii	1	[-155.5874,19.0244]	867924
56	286.9	18.6	N	0	Hawaiian Ocean View	Hawaii	1	[-155.5864,19.0256]	867907
0	291.5	122.7	D	0	Hawaiian Ocean View	Hawaii	1	[-155.5966,19.0258]	867897
53	298.3	63.2	D	0	Hawaiian Ocean View	Hawaii	1	[-155.6021,19.0269]	867898
70	305.5	21.3	D	0	Hawaiian Ocean View	Hawaii	1	[-155.5879,19.0279]	868140
0	305.7	42.0	D	0	Hawaiian Ocean View	Hawaii	1	[-155.5759,19.0356]	868139
69	292.8	11.7	N	0	Pahala	Hawaii	1	[-155.5722,19.0642]	755875
53	293.4	9.0	N	0	Pahala	Hawaii	1	[-155.5614,19.0659]	755873
47	295.4	26.1	D	0	Pahala	Hawaii	1	[-155.5252,19.1352]	523542
0	301.6	120.8	D	0	Pahala	Hawaii	1	[-155.5122,19.1374]	523543
25	295.3	22.9	D	0	Hawaiian Ocean View	Hawaii	1	[-155.8903,19.1571]	593582
100	305.9	347.9	N	0	Pahala	Hawaii	1	[-155.4702,19.1683]	198642
100	302.5	277.5	N	0	Pahala	Hawaii	1	[-155.4531,19.1713]	198645
100	305.9	148.0	N	0	Pahala	Hawaii	1	[-155.4721,19.1801]	198644
100	306.9	198.7	N	0	Pahala	Hawaii	1	[-155.4549,19.1831]	198643

• Fire 2020:

Quando si utilizzano più tabelle che devono essere associate, la situazione ottimale è che l'associazione venga effettuata in base ai campi chiave che presentano nomi identici nelle diverse tabelle. In questo caso è possibile aggiungerle a Qlik Sense con l'opzione di disabilitazione profilatura dei dati di Aggiungi dati e il risultato sarà una struttura di dati con le tabelle associate correttamente.

In particolare si può notare come oltre ai due dataset riguardati gli incendi dal 2010 al 2020 negli Stati Uniti, si è anche aggiunto il file .kml "Us States" contenente le aree geografiche corrispondenti ai vari stati americani, in modo da riuscire a visualizzare nelle mappe tali stati. A questo proposito si sono collegati il campo "usstates.Name" con il campo State sia della "FireArchive" che della "Fire2020" (ricavato tramite la libreria *reversegeocoder* di *python*). Per quanto riguarda le associazioni tra i due file riguardanti gli incendi, sono stati associati tra loro i campi: "Instrument", "confidence", "city" e "state".

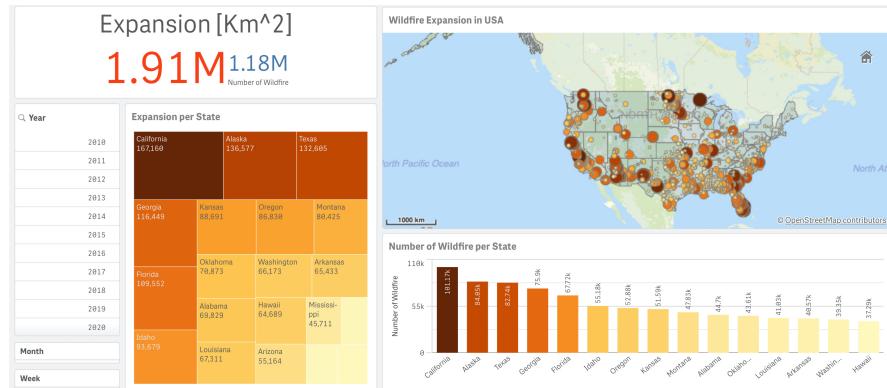
CAPITOLO 2: QLIK SENSE

fire2020.id	latitude	longitude	brightness	scan	track	acq_date	acq_time	fire_2020.sat...	instrument	confidence	bright_t31
0	37.656	-94.51799999999999	385.1	1.0	1.0	2020-07-01	425	Terra	MODIS	63	294.7
1	36.78	-92.806	387.8	1.1	1.1	2020-07-01	425	Terra	MODIS	70	293.6
2	36.664	-98.821	344.8	1.1	1.1	2020-07-01	425	Terra	MODIS	100	299.3
3	36.073	-98.823	322.6	1.1	1.1	2020-07-01	425	Terra	MODIS	100	298.6
4	36.275	-97.921	315.8	1.1	1.1	2020-07-01	425	Terra	MODIS	96	298.7
5	33.443000000000005	-88.76	385.7	1.7	1.3	2020-07-01	425	Terra	MODIS	65	298.1
6	47.696999999999996	-95.345	389.9	1.2	1.1	2020-07-01	438	Terra	MODIS	73	284.6
7	47.695	-95.361	388.8	1.2	1.1	2020-07-01	438	Terra	MODIS	6	286.1
8	47.698	-95.352	310.7	1.2	1.1	2020-07-01	438	Terra	MODIS	71	285.7
9	32.393	-110.7670000000001	300.7	2.5	1.5	2020-07-01	605	Terra	MODIS	16	289.1
10	32.395	-110.743	389.3	2.5	1.5	2020-07-01	605	Terra	MODIS	77	288.8
11	35.849000000000004	-109.32	304.6	3.4	1.7	2020-07-01	605	Terra	MODIS	61	284.9
12	35.864000000000004	-109.321	304.7	3.4	1.7	2020-07-01	605	Terra	MODIS	54	285.2
13	35.926	-109.258	303.5	3.4	1.7	2020-07-01	605	Terra	MODIS	33	283.9
14	35.924	-109.295	307.3	3.4	1.7	2020-07-01	605	Terra	MODIS	60	285.2
15	35.921	-109.26299999999999	318.9	3.4	1.7	2020-07-01	605	Terra	MODIS	25	285.8
16	35.919000000000004	-109.29899999999999	311.5	3.4	1.7	2020-07-01	605	Terra	MODIS	26	285.8
17	35.861	-109.291	300.1	3.4	1.7	2020-07-01	605	Terra	MODIS	13	284.1
18	35.859	-109.32779999999999	312.2	3.4	1.7	2020-07-01	605	Terra	MODIS	84	285.6
19	35.908	-109.294	309.8	3.4	1.7	2020-07-01	605	Terra	MODIS	57	286.0
20	29.124000000000002	-100.697	305.9	1.1	1.0	2020-07-01	840	Aqua	MODIS	63	295.3
21	35.923	-109.26299999999999	308.9	2.4	1.5	2020-07-01	840	Aqua	MODIS	39	285.7
22	35.926	-109.67999999999999	308.4	2.4	1.5	2020-07-01	840	Aqua	MODIS	54	285.8
23	35.915	-109.2970000000001	302.4	2.4	1.5	2020-07-01	840	Aqua	MODIS	12	284.7
74	35.881	-104.307	304.8	2.4	1.5	2020-07-01	840	Aqua	MODIS	53	284.3

frp	fire_2020.da...	fire_2020.City	fire_2020.State	longitude_latitude
4.7	N	Arma	Kansas	[+94.518,37.656]
7.8	N	Ava	Missouri	[+92.886,36.78]
58.3	N	Hennessey	Oklahoma	[+98.021,36.064]
20.3	N	Hennessey	Oklahoma	[+98.023,36.073]
12.7	N	Waukomis	Oklahoma	[+97.921,36.275]
12.0	N	Starkville	Mississippi	[+88.763,33.443]
14.7	N	Bagley	Minnesota	[+95.345,47.697]
6.2	N	Bagley	Minnesota	[+95.361,47.695]
13.9	N	Bagley	Minnesota	[+95.352,47.698]
15.2	N	Tanque Verde	Arizona	[+118.767,32.393]
35.5	N	Tanque Verde	Arizona	[+118.743,32.395]
49.2	N	Ganado	Arizona	[+109.32,35.849]
48.5	N	Ganado	Arizona	[+109.321,35.864]
44.3	N	Navajo	New Mexico	[+109.258,35.926]
57.4	N	Navajo	New Mexico	[+109.295,35.924]
69.0	N	Navajo	New Mexico	[+109.263,35.921]
70.6	N	Navajo	New Mexico	[+109.299,35.919]
36.1	N	Fort Defiance	Arizona	[+109.291,35.861]
81.0	N	Ganado	Arizona	[+109.328,35.859]
63.1	N	Navajo	New Mexico	[+109.294,35.998]
4.7	N	Laughlin Air Force Base	Texas	[+106.697,29.124]
38.3	N	Navajo	New Mexico	[+109.263,35.923]
37.7	N	Navajo	New Mexico	[+109.268,35.926]
24.3	N	Navajo	New Mexico	[+109.297,35.915]
79.4	N	Fort Defiance	Arizona	[+109.302,35.861]

2.2 Panoramica sulle dimensioni dell'incendio

In questa Dashboard si è andati a costruire dei grafici che permettono di avere una visione generale riguardo le dimensioni dei vari incendi negli ultimi dieci anni.



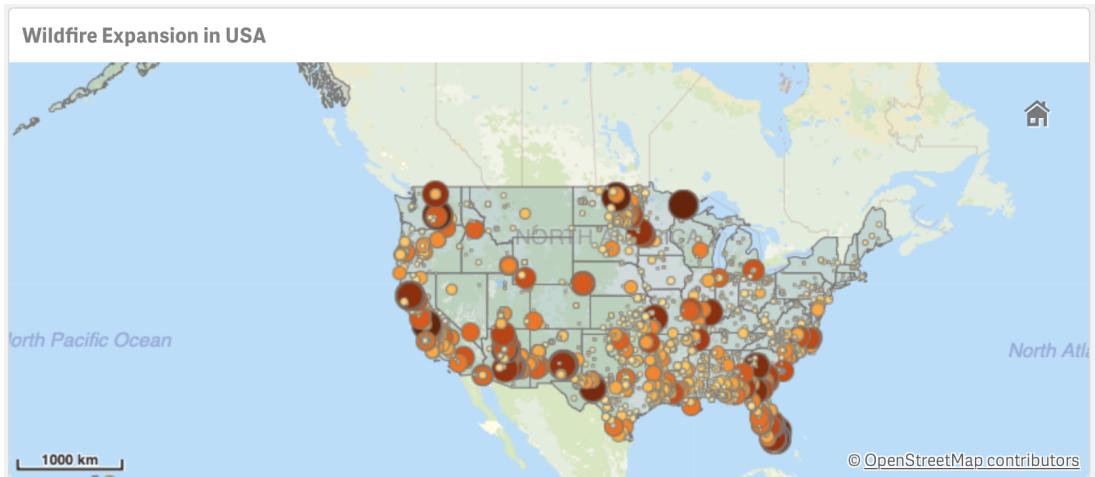
2.2.1 KPI

Come primo grafico come panoramica sul dataset a nostra disposizione, si è deciso di presentare alcuni numeri sull'espansione totale dell'incendio in Km^2 e sul numero totali di incendi avvenuti nell'ultima decade. Questi sono stati calcolati tramite una misura di Sum sul numero di incendi e sull'espansione di questi.



Il grafico utilizzato per entrambi i valori è quello del KPI (Key Performance Indicator), che riesce a mostrare un numero rilevante in modo chiaro e conciso.

2.2.2 Mappa delle dimensioni



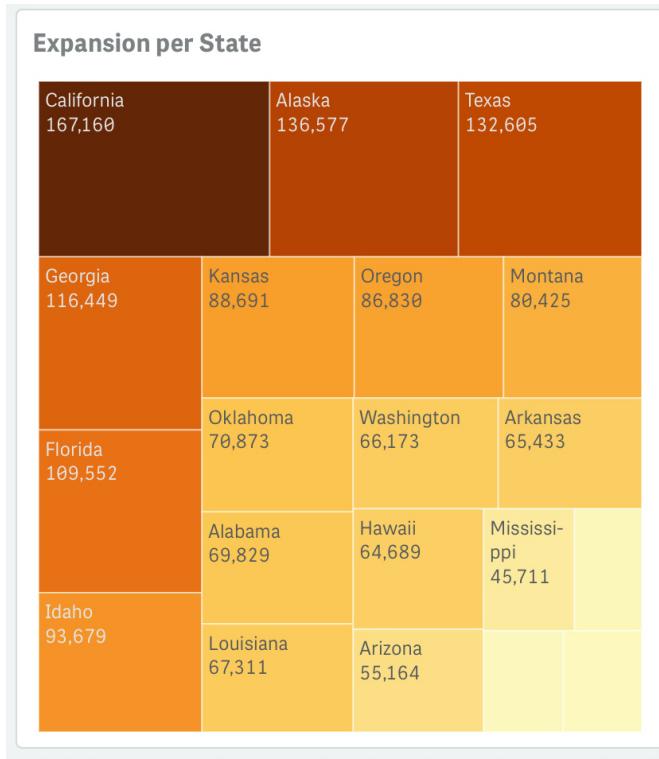
Al fine di mostrare geograficamente le aree degli interessate dagli incendi, si è andati a realizzare questa mappa. Si è deciso di adottare una mappa di punti per questa particolare visualizzazione.

Il campo usato per identificare le posizioni è “fire archive.longitude fire archive.latitude” che indica appunto la latitudine e la longitudine precisa di quel particolare incendio. In particolare tanto più grande è la sua dimensione e la sua intensità di colore, tanto maggiore è la dimensione dell'incendio.

Il grafico permette di capire che l'espansione maggiore si è verificata negli stati della California e del Texas negli ultimi 10 anni. In realtà, spostandosi con il cursore tale mappa mostra anche gli incendi presenti nell'Alaska e nelle Hawaii.

Un problema riscontrato nelle mappe di Qlik è che queste riescono a lavorare con fino un certo numero di dati, oltre il quale non visualizzano più correttamente eventuali altri dati.

2.2.3 Espansione per Stato



Questo grafico ha l'obiettivo di fornire informazioni sugli stati che hanno avuto maggior espansione in Km² di incendi negli ultimi 10 anni. Il grafico utilizzato è la treemap in cui ogni quadrato rappresenta uno stato e tanto quanto più grande è la sua dimensione tanto maggiore sarà la somma delle espansioni degli incendi, oltretutto anche l'intensità del colore dei vari quadrati sta ad indicare la somma delle espansioni degli incendi.

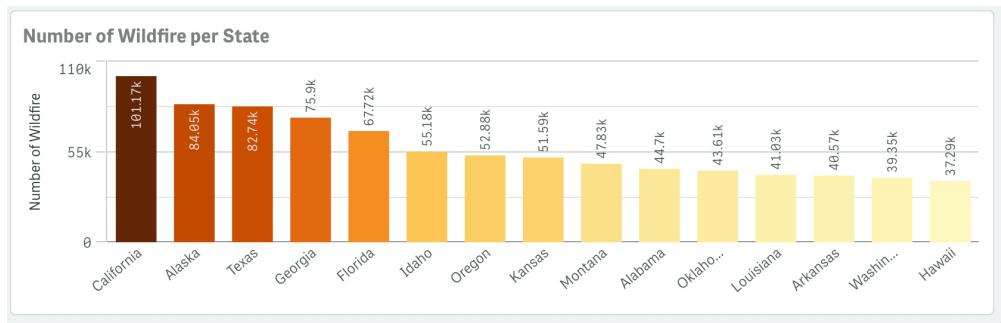
Questo grafico mostra che la maggior parte degli incendi del dataset con maggiore espansione è avvenuta principalmente in California, Alaska, Texas, Georgia e Florida.

2.2.4 Numero di incendi per Stato

Questo grafico ha l'obiettivo di fornire informazioni sugli stati che hanno avuto un maggior numero di incendi negli ultimi dieci anni.

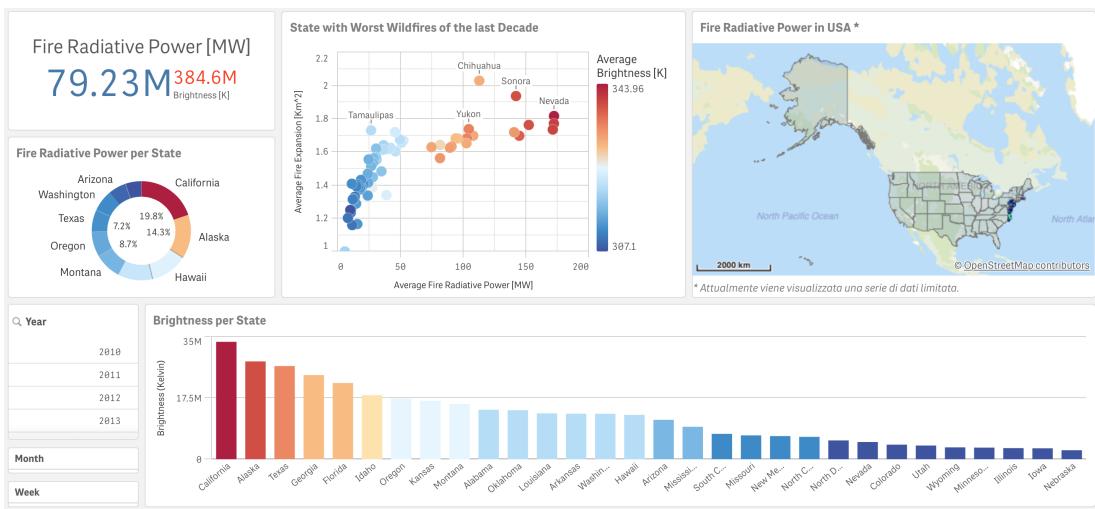
Si è andati ad utilizzare un grafico a barre per la sua rappresentazione in cui ogni barra rappresenta il numero di incendi per ogni singolo stato, dove l'altezza della barra e la sua intesità di colore mi sta ad indicare gli stati più colpiti.

Si nota facilmente come questo grafico mostra che la maggior parte degli incendi del dataset è avvenuta principalmente in California, Alaska, Texas, Georgia e Florida (similmente a quanto era avvenuto per l'espansione).



2.3 Analisi sulla Potenza irradiata dall'incendio

In questa Dashboard si è voluto dare una panoramica sulla Potenza irradiata in Mega-Watt degli incendi.

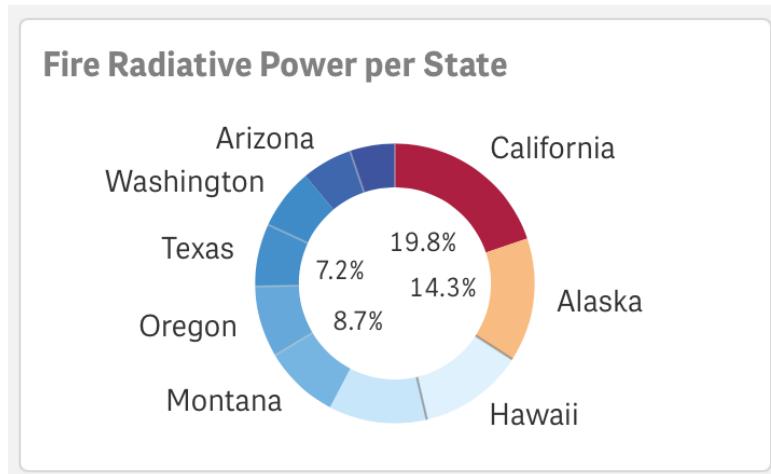


2.3.1 KPI



Come primo grafico come panoramica sul dataset a nostra disposizione, si è deciso di presentare alcuni numeri sulla Potenza irradiata sul numero totali di incendi avvenuti nell'ultima decade e sulla luminosità in Kelvin di questi. Il grafico utilizzato per entrambi i valori è quello del KPI (Key Performance Indicator), che riesce a mostrare un numero rilevante in modo chiaro e conciso.

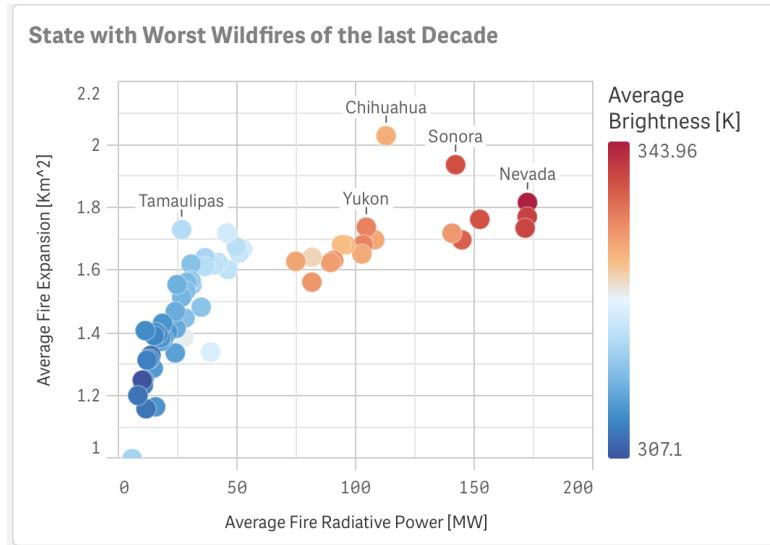
2.3.2 Analisi della Potenza irradiata per singolo Stato



In questa visualizzazione si possono vedere quali sono gli stati degli Stati Uniti in cui ci sono più incendi con potenza irradiata maggiore. Il grafico a torta permette di individuare gli stati più rischiosi. Viene mostrato in particolare, come oltre gli stati che ci aspettavamo avessero maggiori incendi come California e Alaska, ci sono anche le Hawaii, dato che principalmente gli incendi che avvengono in queste zone sono dovuti a fenomeni vulcanici, i quali incendi hanno potenza irradiata maggiore rispetto ad incendi di vegetazione/boschivi.

Dal punto di vista tecnico, per realizzare questo grafico, si è preso di nuovo come misura la Sum della "frp" e per indicare gli stati il campo "State" del dataset "Fire Archive".

2.3.3 Stati con gli incendi peggiori degli ultimi 10 anni

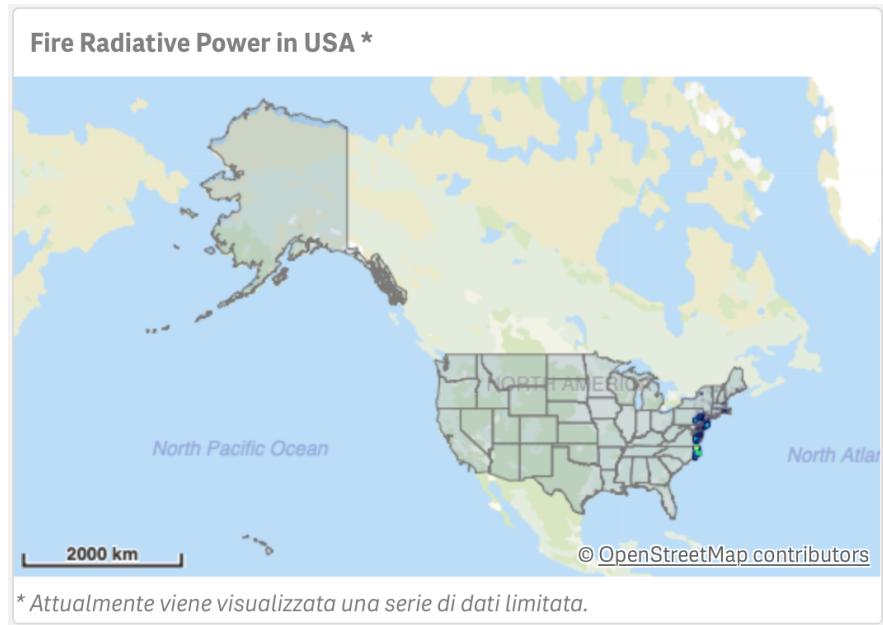


In questo grafico si possono notare quali sono gli stati che hanno avuto incendi peggiori in termini di Potenza irradiata (asse delle x) e di espansione dell'incendio (asse delle y), dove la scala di colore dipende dalla luminosità in Kelvin degli incendi. Lo scatter plot (grafico a dispersione) o scatter graph è un tipo di grafico in cui due variabili di un set di dati sono riportate su uno spazio cartesiano. I dati sono visualizzati tramite una collezione di punti ciascuno con una posizione sull'asse orizzontale determinato da una variabile e sull'asse verticale determinato dall'altra. In particolare ogni punto rappresenta uno stato e più è alta la potenza irradiata, più è grande l'espansione e più la scala di colore della luminosità tende al rosso scuro, più l'incendi sono considerati peggiori per stato.

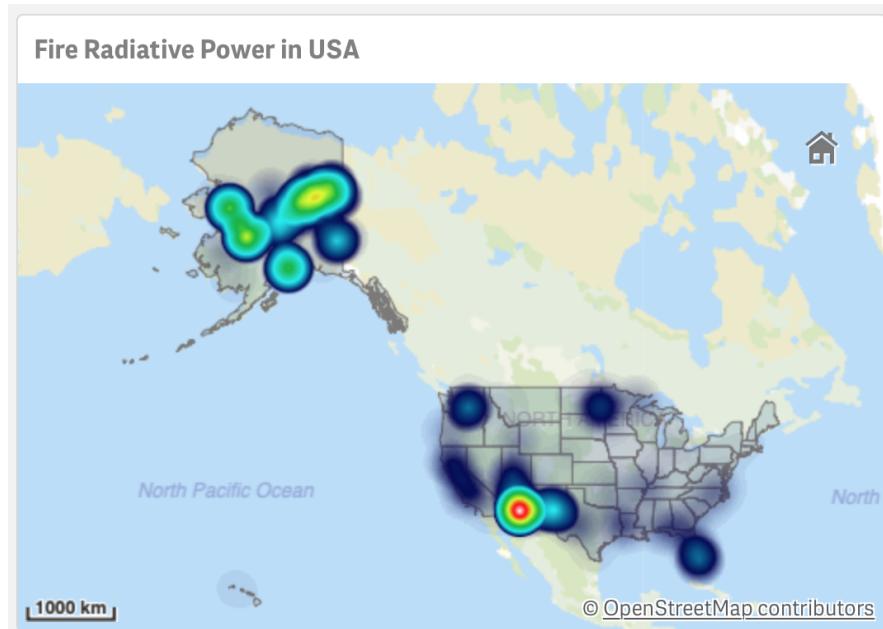
In questo caso infatti l'analisi porta a pensare che, studiato l'insieme di questi tre fattori, il Nevada sembra essere lo stato più colpito non per numero di incendi, bensì per questi altri tre fattori, dove sull'asse delle x si è presa come misura la media della "frp", sulle y la media della "expansion" e sulla scala di colore la media della "brightness".

2.3.4 Mappa della Potenza irradiata negli USA

Al fine di mostrare geograficamente dove la potenza irradiata fosse maggiore negli Stati Uniti è stata graficata tale mappa. In particolare l'intensità in Mega Watt dell'incendio è stata plottata attraverso un livello di densità, passandogli la latitudine e la longitudine come dimensione e la "frp" come peso e raggio. Tuttavia come possiamo leggere sotto la

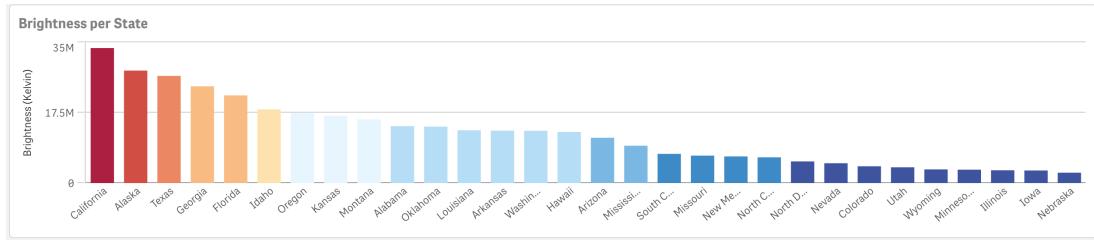


figura, Qlik non riesce a visualizzare un quantitativo ingente di dati nella stessa porzione di mappa. Questo problema è dovuto appunto al fatto che, essendo un dataset di circa 150 megabyte, non riesce a processare graficamente troppi dati nella mappa. Quello che si può andare a fare è utilizzare dei filtri per fare delle analisi più puntuali in un determinato anno e mese (per esempio Luglio del 2019 nella figura in basso) e in questi casi il grafico riesce ad elaborare tutti i dati all'interno della stessa mappa e graficare correttamente la potenza irradiata in tutti gli Stati Uniti.



2.3.5 Luminosità degli incendi divisa per Stato

Questo grafico permette di avere informazioni sulla quantità e sulla gravità della luminosità degli incendi in Kelvin, divisa per stato. In particolare, l'altezza delle barre indica



la somma delle brightness degli incendi in quello stato e lo stesso avviene per il colore che diventa più tendente al rosso di pari passo alla gravità dell'incendio. Dei 50 stati appartenenti agli USA, abbiamo rappresentato i primi 30 con luminosità più elevata e quindi più rilevanti per il grafico preso in esame.

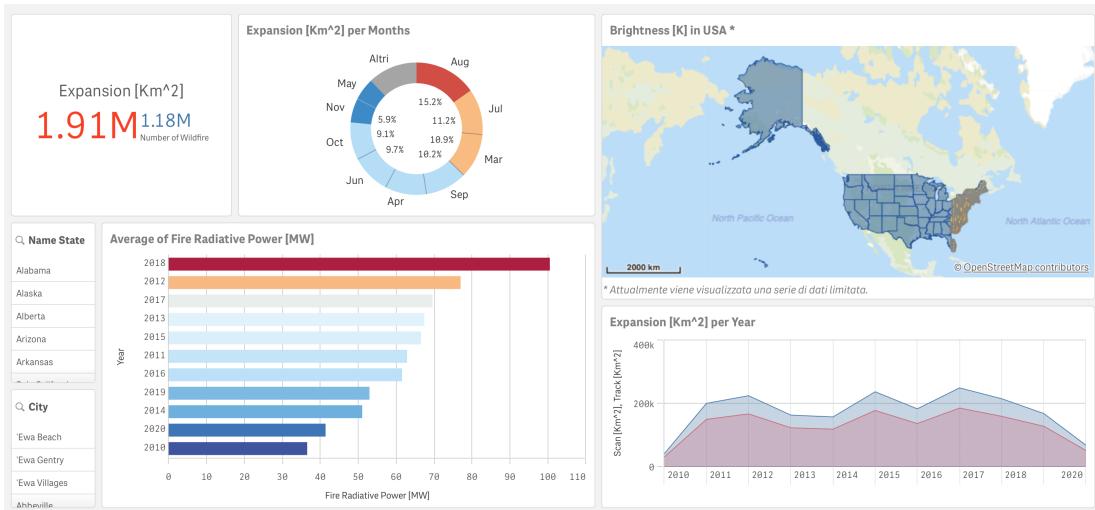
A questo punto si può evincere, per esempio, che come anche per l'espansione in chilometri quadri e il numero di incendi, gli stati più colpiti sono rispettivamente: la California, l'Alaska, il Texas e a seguire la Georgia e la Florida.

La dimensione scelta è il campo "US State", la misura è la consueta Sum, ora però sulla brillantezza dell'incendio: "brightness" ed il colore è implementato con un gradiente sequenziale che dipende anch'esso dal campo "brightness".

2.4 Analisi per Stato

In questa Dashboard si è andati a costruire dei grafici che permettono di avere una visione generale riguardo gli incendi divisi per paese negli ultimi 10 anni. In basso a sinistra sono stati inseriti dei filtri che permettono di concentrare l'analisi appunto su un determinato stato piuttosto che una certa città degli Stati Uniti. Nel seguito della trattazione vengono descritti i grafici che compongono questa dashboard.

CAPITOLO 2: QLIK SENSE



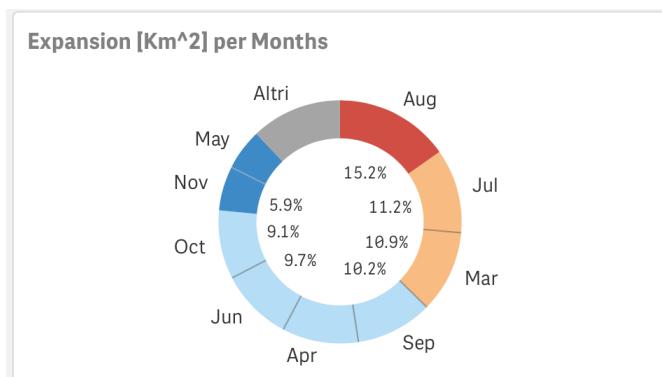
2.4.1 KPI



Come primo grafico come panoramica della dashboard a nostra disposizione, si è deciso di presentare alcuni numeri sull'espansione totale dell'incendio in Km^2 e sul numero totali di incendi avvenuti nell'ultima decade. Questi sono stati calcolati tramite una misura di Sum sul numero di incendi e sull'espansione di questi.

Il grafico utilizzato per entrambi i valori è quello del KPI (Key Performance Indicator), che riesce a mostrare un numero rilevante in modo chiaro e conciso.

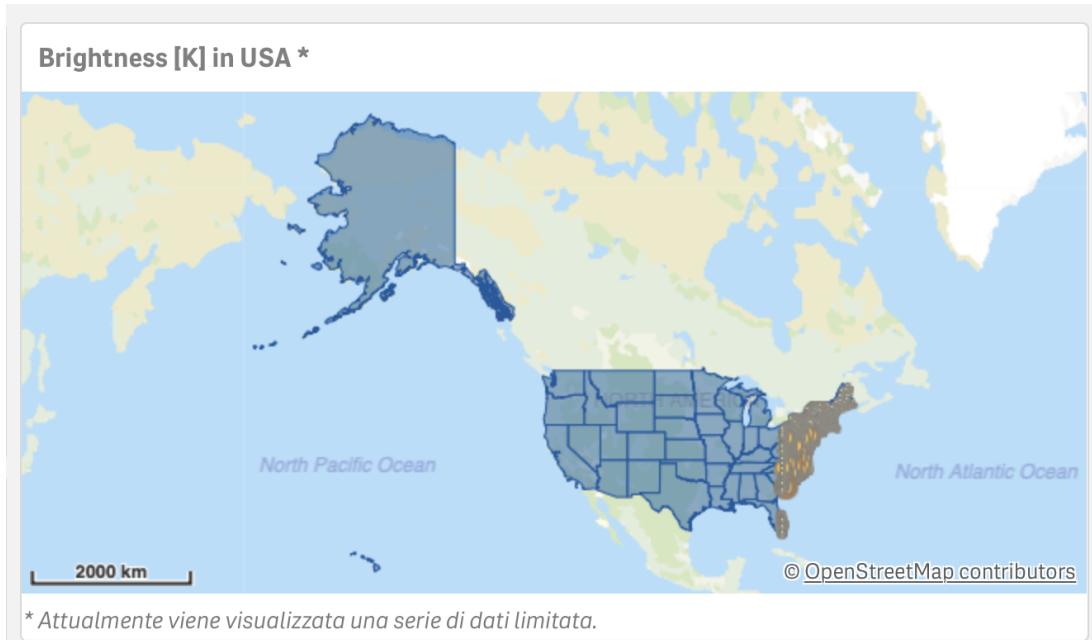
2.4.2 Analisi Espansione per Mese



In questa visualizzazione si possono vedere quali sono i mesi in cui negli anni ci sono stati più incendi con espansione in chilometri quadrati maggiore. Il grafico a torta permette di verificare ciò che ci si aspettava: nei mesi estivi c'è una prevalenza di espansione degli incendi, con un picco del 15.2% nel mese di Agosto. Particolare è il terzo mese con più incendi: Marzo con ben il 10.9% dell'espansione totale. Questo è dovuto al fatto che in stati come il Texas o l'Idaho, il mese con più incendi è appunto quello di Marzo (lo si può verificare selezionando ad esempio nei filtri questi due paesi).

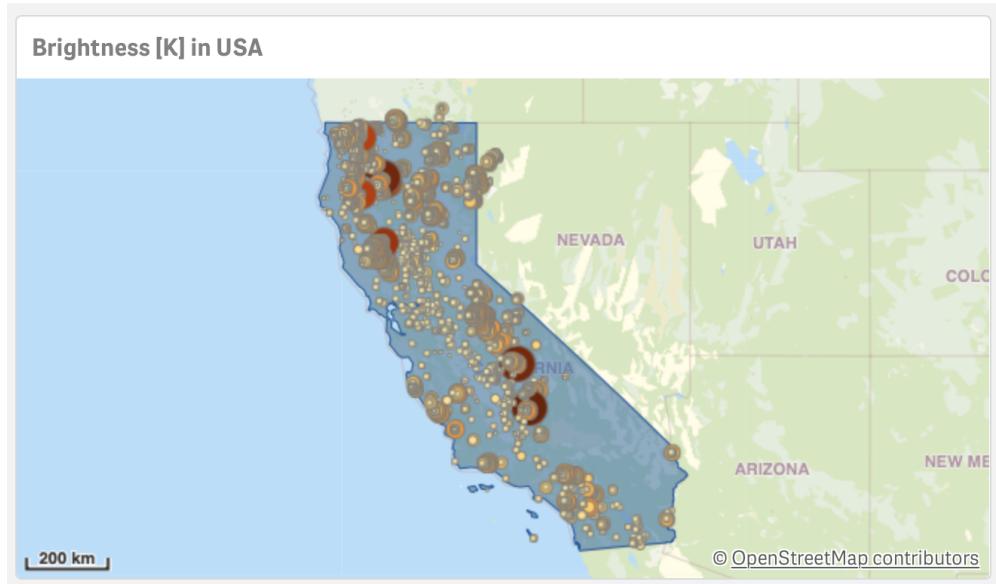
Il diagramma circolare viene costruito dividendo un cerchio in spicchi le cui ampiezze angolari come anche le aree sono proporzionali alle classi di frequenza, data come Sum del campo "Scan" presente nel dataset "Fire Archive". Il colore è implementato con classi di colore diverse che partono dal blu scuro fino ad arrivare al rosso per i mesi più colpiti da incendi, rappresentati anche qui dalla Sum della "Scan". Infine si è preso come dimensione il campo "Months".

2.4.3 Mappa della brillantezza degli incendi negli USA

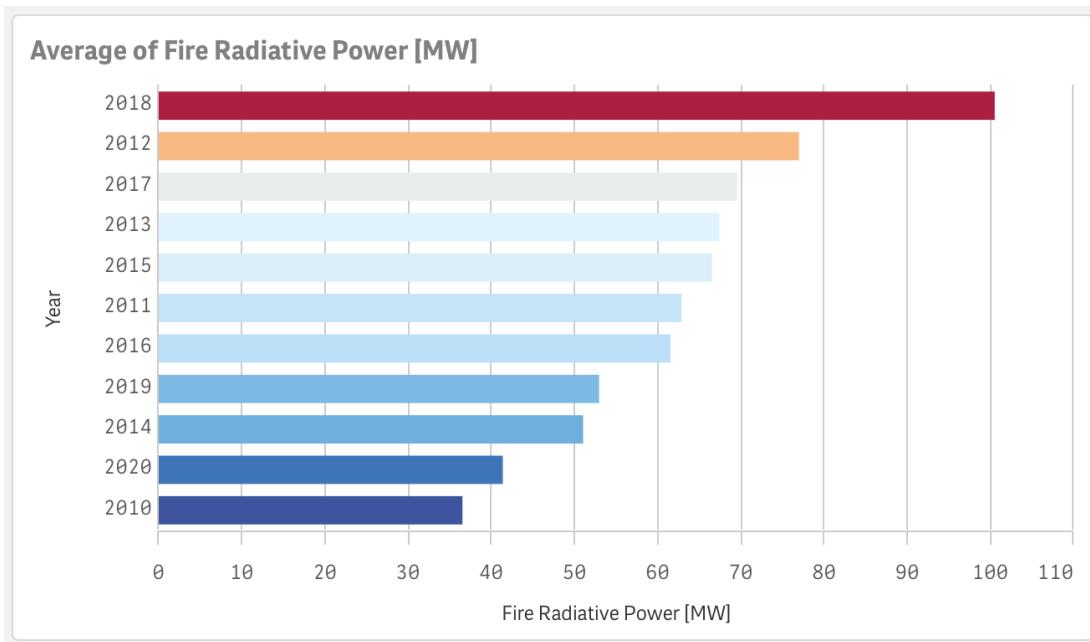


Tale mappa è stata costruita al fine di mostrare geograficamente dove le luminosità in Kelvin degli incendi fossero maggiori negli Stati Uniti. In particolare il numero di incendi è stata graficato attraverso un livello ad aree, passandogli la latitudine e la longitudine come dimensione e la "brightness" come peso e raggio, così come anche per il colore. Tuttavia come possiamo leggere sotto la figura, Qlik non riesce a visualizzare un

quantitativo ingente di dati nella stessa porzione di mappa. Quello che si può fare però è utilizzare i filtri per fare delle analisi più puntuale in un determinato luogo dei Stati Uniti (per esempio scegliendo la California attraverso il filtro) e in questi casi il grafico riesce ad elaborare tutti i dati all'interno della stessa mappa e graficare correttamente la luminosità nella porzione di mappa.



2.4.4 Analisi della Potenza irradiata media



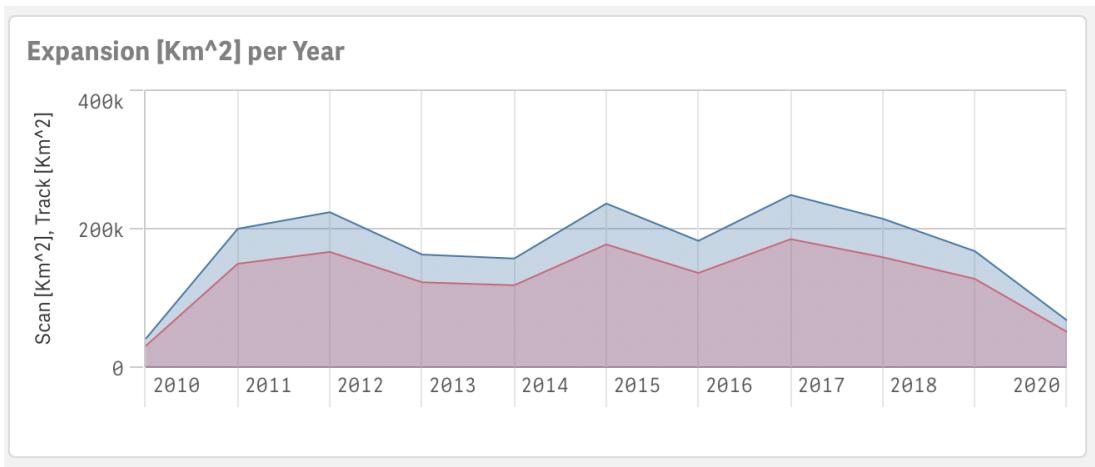
In questo grafico si vogliono far risaltare quella che è la media di Potenza irradiata in [MW] dall'incendio nel corso degli ultimi dieci anni.

Si tratta di un grafico a barre sulla cui ascissa si trova il numero medio di Potenza irradiata, mentre sulle ordinate il particolare anno nel quale si è rilevata in media tale potenza. In particolare la potenza irradiata è stata calcolata come misura facendo una Avg della "frp" mentre la dimensione è l'anno che è stato calcolato facendo: = $Dual(Year([fire_archive.acq_date]), YearStart([fire_archive.acq_date]))$.

Per quanto riguarda l'anno con una media di potenza irradiata maggiore si può notare dal grafico come questo sia il 2018. Si può infatti ricordare ad esempio l'incendio del Nord della California dell'28 luglio 2018, definito il peggiore nella storia di tale stato, dove risultarono inceneriti 114.850 ettari di terreno, una zona grande quanto l'intera Los Angeles e più di 60 morti.

Mentre bisogna far attenzione ai due anni con meno potenza media irradiata che sono il 2010 e 2020, nel quale mancano diversi mesi che possono aver influito sul calcolo della media.

2.4.5 Analisi dell'Espansione negli anni



In questo grafico si vuole far risaltare quella che è l'espansione degli incendi nel corso degli ultimi dieci anni.

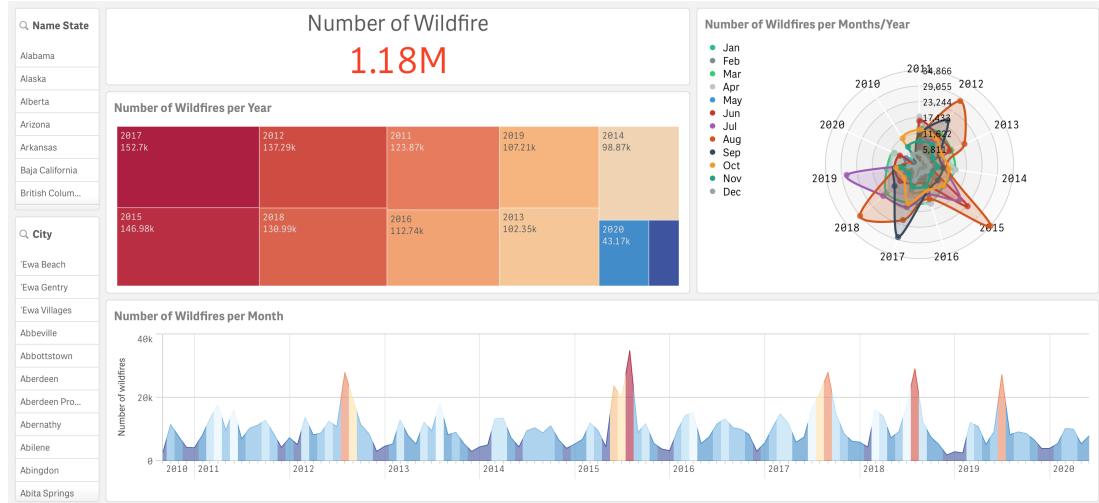
Si tratta di un grafico lineare dove determinati i punti sul grafico grazie alle due variabili, è possibile unirli se si vuole rappresentare un andamento temporale. In questo caso, per convenzione si rappresenterà lo scorrere del tempo in anni sull'asse orizzontale (x).

La dimensione scelta è il campo "Year", mentre le misure sono due: la Sum della "Scan" e la Sum della "Track". Queste due misure rappresentano entrambe l'espansione dell'incendio calcolata attraverso due metodi differenti e per questo si è voluto graficarle entrambe nella stessa porzione di piano per vedere le differenze tra le due. Per quanto riguarda i colori abbiamo: rosso per la "Track" e blu per la "Scan".

2.5 Analisi sul numero di incendi per Anno e per Mese

Se fino ad ora l'analisi è stata concentrata sui parametri di luminosità, espansione e potenza irradiata dall'incendio, questa dashboard vuole fornire nuove analisi su un parametro altrettanto importante: il numero di incendi.

Sulla parte sinistra della dashboard si possono trovare due tipi di filtri: per Stati e per Città, in modo da fare un analisi più puntuale anche in determinati luoghi geografici a cui si è interessati.



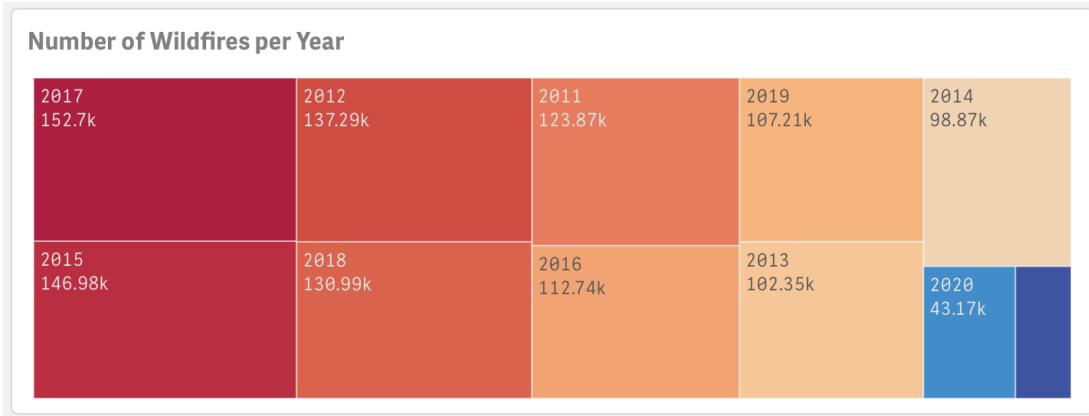
2.5.1 KPI



Anche qui come primo grafico come panoramica sul dataset a nostra disposizione, si è deciso di presentare il numero totale di incendi avvenuti nell'ultima decade. Questi sono

stati calcolati tramite una misura di Sum sulla *wildfire_presence*. Il grafico utilizzato per tale valore è quello del KPI (Key Performance Indicator), che riesce a mostrare un numero rilevante in modo chiaro e conciso.

2.5.2 Numero di incendi per anni



Il grafico utilizzato per questa rappresentazione è una mappa ad albero dove vengono ordinati in ordine crescente gli anni che hanno avuto un numero maggiore di incendi.

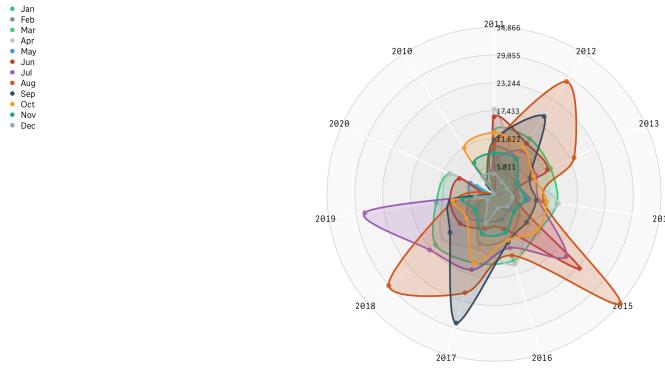
Se nella dashboard precedente al punto [2.4.4](#) l'anno con maggior potenza media è stato il 2018, stessa cosa non si può dire sul numero, dove l'anno peggiore è stato il 2017 con più di 152 mila incendi in un solo anno. Questo porta a pensare come gli incendi del 2018 siano stati in numero minore ma con conseguenze peggiori.

Da un punto di vista più tecnico, per trovare tale grafico si è utilizzato come dimensione la *Dual(Year([fire_archive.acq_date]), YearStart([fire_archive.acq_date]))* mentre come misura la *wildfire_presence*. Anche per quanto riguarda la scala di colori abbiamo inserito come variabile la *wildfire_presence*, dove a colori tendenti alla scala del rosso corrisponde un maggior numero di incendi.

2.5.3 Numero di incendi per Anni/Mese

Il grafico utilizzato nella figura è un Radar (diagramma di Kiviat) che è un metodo grafico per mostrare dati su variabili multiple in forma di un grafico bidimensionale di tre variabili (due dimensioni e una misura), rappresentate su assi con la stessa origine.

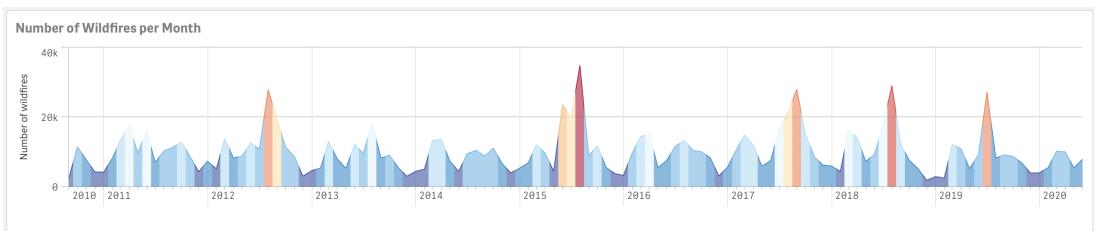
Questa rappresentazione grafica consiste di una sequenza di raggi che hanno origine da un centro e formano angoli uguali tra loro; ogni raggio rappresenta una delle variabili.



La distanza dal centro del punto marcato sul raggio è proporzionale al valore della variabile rispetto al valore massimo raggiungibile. I punti sui raggi vengono congiunti con segmenti arrotondati, così che il grafico ha la forma di una ragnatela smussata. Come misura si è andata a graficare una Sum della *wildfire_presence*, mentre come dimensioni usiamo i mesi *[fire_archive.acqdate.autoCalendar.Month]* mentre per l'anno il campo *Dual(Year([fire_archive.acq_date])*.

In particolare notiamo come ci siano 12 differenti curve all'interno del radar che rappresentano i 12 mesi di un anno, mentre lungo la circonferenza del radar sono presenti gli anni nei quali si vuole studiare l'andamento. Come potevamo già immaginare i mesi con maggiori incendi sono Agosto, Settembre, Luglio e Giugno e gli anni peggiori sono stati il 2017, 2015, 2012 e il 2018 (come era già risultato dal grafico precedente).

2.5.4 Numero di incendi per mesi



Come ultima analisi di questa dashboard, dopo aver studiato il numero di incendi per anni, trimestri nei vari anni, è interessante vedere anche qual'è l'andamento nei vari mesi dell'anno.

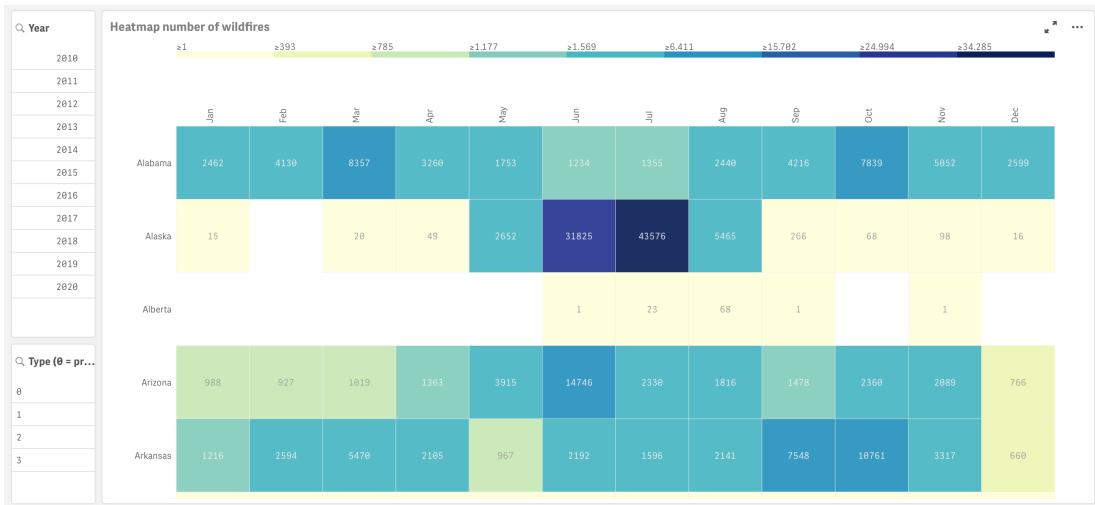
Notiamo come i picchi maggiori li abbiamo in corrispondenza dei mesi di Agosto e Settembre mentre nei mesi più freddi come Dicembre e Gennaio i numeri sono decisamente minori. A livello tecnico, è stato utilizzato un grafico lineare, dove l'andamento sull'asse

delle y è rappresentato da un area indicata dalla misura Sum della *wildfire_presence*, mentre sull'asse delle x sono graficati come dimensione i vari mesi dei singoli anni.

2.6 Analisi Numero di incendi per Stato e per Mese

Dopo aver fatto un analisi sul numero di incendi divisa per anni e mesi, ci si è incentrati su un analisi basata sull'andamento degli incendi per i vari stati oltre che in base ai vari mesi dell'anno (heat-map).

Per quanto riguarda il lato sinistro della dashboard si notano due filtri: Anni e Tipo. In questo modo si potrà specializzare l'analisi in un determinato arco temporale e per un determinato tipo, oltre che si potranno selezionare gli stati interessati direttamente dal grafico.

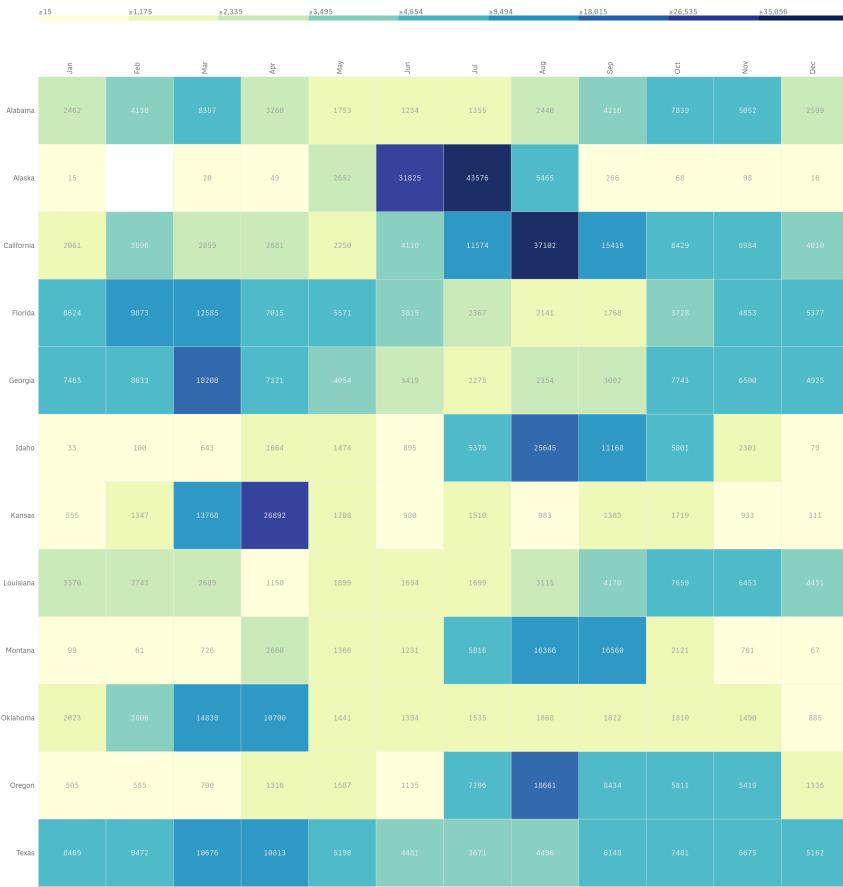


2.6.1 Mappa di calore sul numero di incendi

Una Heat-Map (o mappa termica) è una tecnica di visualizzazione dei dati che mostra l'entità di un fenomeno sotto forma di colore in due dimensioni.

Da notare che la variazione di colore è sia di tonalità o intensità, fornendo al lettore evidenti indizi visivi su come il fenomeno è raggruppato o varia nello spazio. In particolare nella presente mappa di calore, le magnitudini sono disposte in una matrice di dimensione cella fissa le cui righe e colonne sono fenomeni e categorie distinti, e l'ordinamento di righe e colonne è intenzionale per quanto riguarda i mesi, mentre è alfabetico per gli stati.

CAPITOLO 2: QLIK SENSE



In particolare si è voluto incentrare l’analisi negli stati con il maggior numero di incendi raffiguranti in [2.2.4](#), andando a prendere i peggiori 12 stati per numero di incendi. Si può notare come i mesi di Giugno, Luglio e Agosto sono quelli dove si incentra il maggior numero di incendi, come era già stato notato nella [2.5.4](#). Inoltre studiando per righe, si nota come regioni come il Texas, la Georgia e la Florida presentano una quasi assenza di stagionalità dato che gli incendi sono mediamente dello stesso numero durante il corso dell’anno. Al contrario stati come l’Alaska, la California o il Kansas mostrano una forte stagionalità, dove in particolare l’Alaska è caratterizzata da un numero elevatissimo di incendi nei mesi di Giugno e Luglio per poi esser quasi nulli nei vari altri mesi dell’anno.

2.7 Analisi Trimestri con Medie

Un’ulteriore analisi che si può fare è quella sull’andamento del numero di incendi sia nei trimestri di ogni anno, che unicamente nei vari trimestri.

Oltre questo nei tre grafici in basso a destra sono rappresentate le medie di Brillantezza,

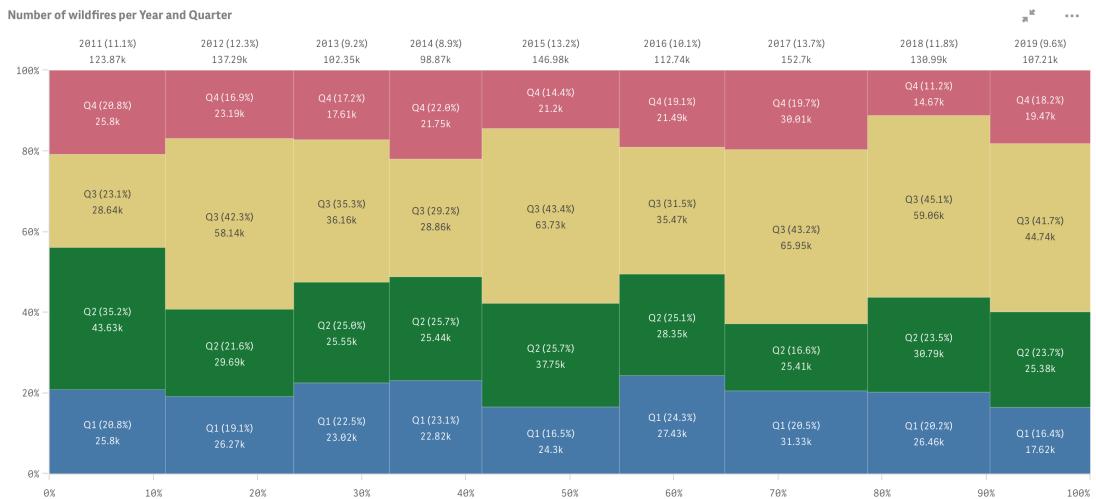
CAPITOLO 2: QLIK SENSE



Espansione e Potenza irradiata. In questa rappresentazione sono stati esclusi gli anni 2010 e 2020 dato che i dati riguardanti questi due anni non sono ben comparabili con i restanti poiché non completi.

Nella parte sinistra della dashboard notiamo un filtro per stato e un secondo filtro per Anno/Mese/Settimana, in modo da fare analisi puntali in un determinato periodo e luogo.

2.7.1 Numero di incendi per Anno e per Trimestre



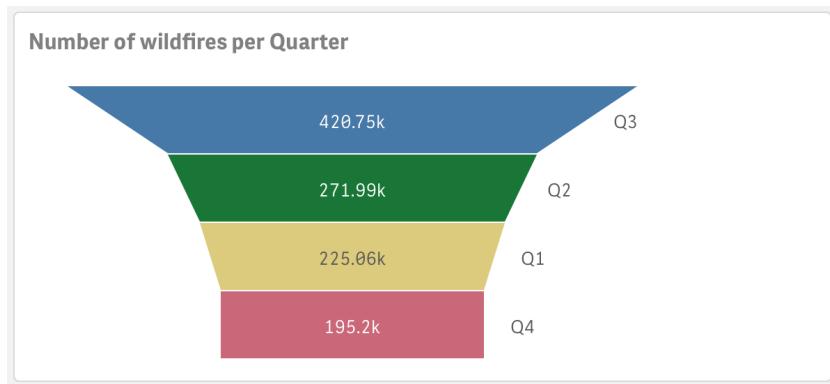
Per rappresentare l'andamento del numero degli incendi al variare dei vari trimestri dell'ultimo decennio, si è deciso di utilizzare un grafico Mekko (a mosaico), metodo grafico per visualizzare i dati da due o più variabili qualitative. Fornisce una panoramica dei dati e consente di riconoscere le relazioni tra le diverse variabili.

Da un punto di vista tecnico, come dimensioni sono state scelte il campo e il campo "Quarter", mentre come misura la Sum della *wildfire_presence*. Le diverse celle e i loro relativi colori sono rappresentate dai quattro trimestri:

- **Q1:** Gennaio, Febbraio, Marzo;
- **Q2:** Aprile, Maggio, Giugno;
- **Q3:** Luglio, Agosto, Settembre;
- **Q4:** Ottobre, Novembre, Dicembre.

Come si può vedere dalla figura, all'aumentare del numero di incendi in quel particolare anno l'area di quest'ultimo sarà più grande rispetto le altre, stessa cosa vale per i vari trimestri, dove saranno maggiori in altezza a valori elevati del numero di incendi. Se già in precedenza da [2.5.2](#) si notava come il 2017 fosse stato l'anno con maggior numero di incendi, da questa analisi notiamo come in particolare sia il terzo trimestre di quell'anno quello ad essere più colpito. In seguito nello studio in Tableau faremo un analisi più approfondita precisamente su tale trimestre.

2.7.2 Numero di incendi per Trimestre

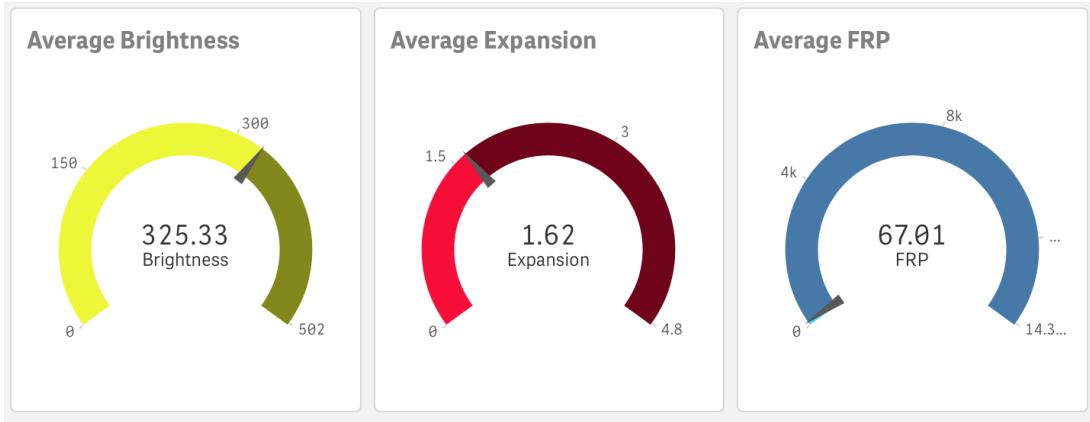


I grafici a imbuto sono un tipo di grafico, spesso utilizzato per rappresentare i trimestri e mostrano l'ammontare del numero di incendi per ciascuna fase. Questo tipo di grafico può essere utile per identificare il trimestre peggiore. Come è facile immaginare è il terzo trimestre (Q3) ad essere quello con maggior numero di incendi, dato che anche in [2.5.4](#) e [2.5.3](#) si notava come i mesi con più incendi fossero Agosto e Settembre.

Per quanto riguarda le dimensioni (rappresentata dai vari segmenti), si è inserito il "Quarter" mentre la misura è rappresentata dalla *wildfire_number*.

2.7.3 Media della Brillantezza, Espansione, Potenza irradiata

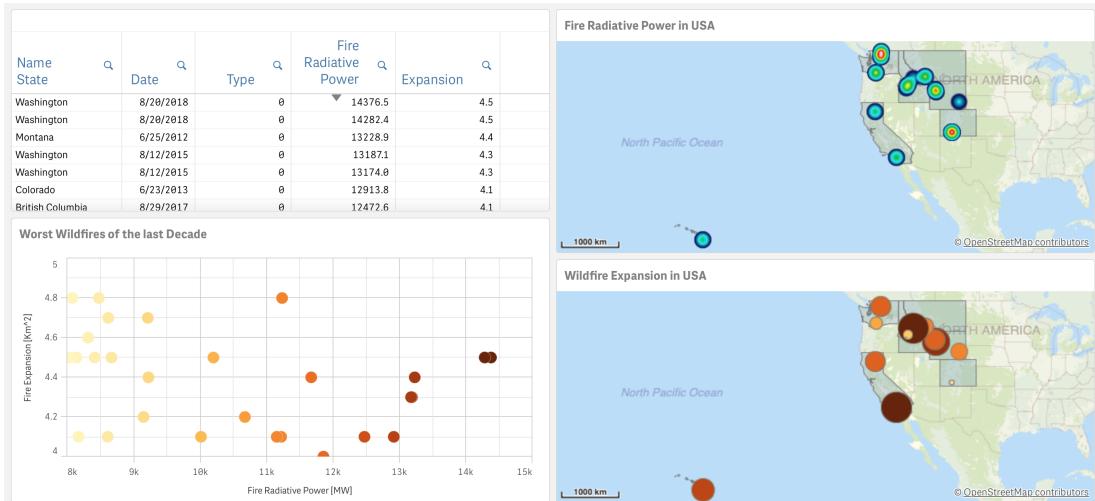
Nei seguenti grafici è stata rappresentata la media della brillantezza in Kelvin, espansione dell'incendio in chilometri quadri e potenza irradiata in mega watt. In particolare il puntatore presente nel grafico mostra la media su tutti e 10 gli anni e in tutti gli USA. Perciò quando si selezionerà un filtro per una determinata area o un determinato arco temporale si potrà vedere la differenza rispetto alla media totale. I misuratori partono



tutti da zero e arrivano al massimo valore, diverso per ciascuno dei tre parametri dato che le unità di misura sono differenti.

2.8 Incendi peggiori negli ultimi dieci anni

Un aspetto di notevole importanza su cui focalizzarsi è quali sono stati gli incendi peggiori in termini di espansione e di potenza irradiata nell'ultima decade.



2.8.1 Tabella peggiori incendi

Nell'analisi degli incendi peggiori degli ultimi dieci anni, questa tabella fornisce diverse informazioni: quale è stato lo stato nel quale è avvenuto tale incendio, data dell'avvenimento, tipo di incendio, potenza irradiata ed espansione.

Name State	Date	Type	Fire Radiative Power	Expansion
Washington	8/20/2018	0	14376.5	4.5
Washington	8/20/2018	0	14282.4	4.5
Montana	6/25/2012	0	13228.9	4.4
Washington	8/12/2015	0	13187.1	4.3
Washington	8/12/2015	0	13174.0	4.3
Colorado	6/23/2013	0	12913.8	4.1
British Columbia	8/29/2017	0	12472.6	4.1
Idaho	8/21/2016	0	11855.1	4.0
Montana	6/25/2012	0	11671.0	4.4
California	8/8/2018	0	11233.2	4.8
Colorado	6/23/2013	0	11218.4	4.1
Colorado	6/23/2013	0	11152.6	4.1
Idaho	9/28/2018	0	10672.9	4.2
Wyoming	9/22/2012	0	10198.9	4.5
British Columbia	8/29/2017	0	10012.0	4.1
Wyoming	6/29/2012	0	9220.8	4.4
Wyoming	7/26/2016	0	9211.3	4.7
Idaho	9/28/2018	0	9148.7	4.2
California	8/3/2014	0	8665.3	4.5
Wyoming	7/26/2016	0	8616.6	4.7
Colorado	6/23/2013	0	8607.3	4.1
Idaho	8/24/2012	0	8474.5	4.8
Wyoming	9/22/2012	0	8412.9	4.5

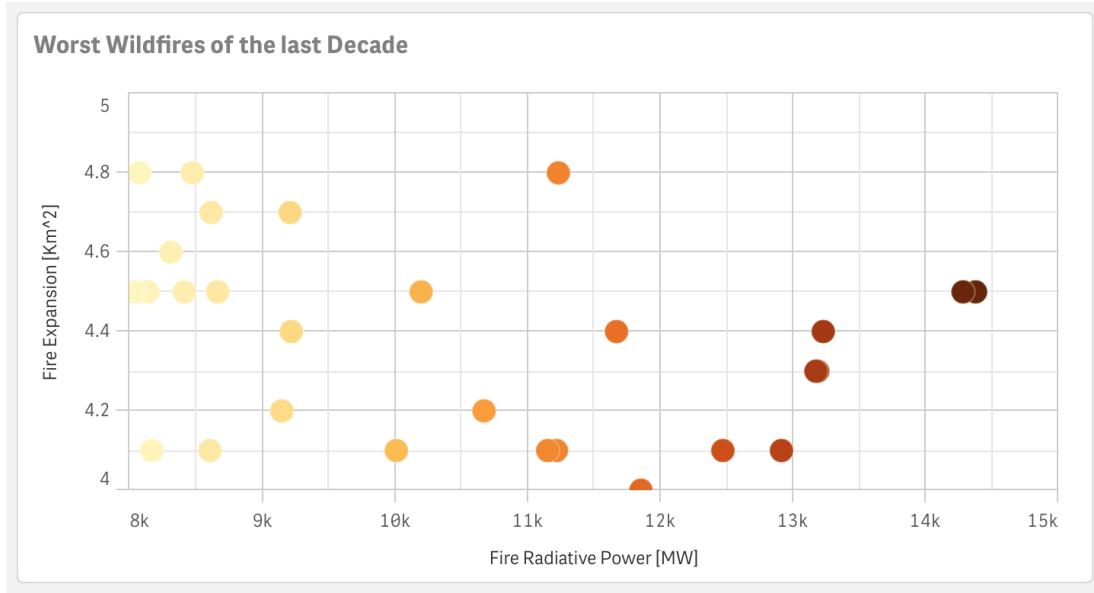
La tabella è ordinata in ordine decrescente per potenza irradiata, dove tuttavia a stessi incendi corrispondono anche "id" dell'incendio differenti. Questo avviene perché lo stesso incendio può avere più fasi iniziali di propagazione, per poi confluire in un unico incendio. A questo proposito il satellite può rilevare incendi diversi che in realtà sarebbero da considerarsi unico.

Notiamo come l'incendio peggiore in termini di potenza irradiata ("frp") ed espansione ("Scan") è stato l'incendio dello stato di Washington iniziato il primo di luglio del 2018 e che ha avuto il culmine nella giornata del 20 Agosto del 2018. Per quanto riguarda il secondo incendio peggiore, questo è avvenuto nello stato del Montana dove gli 1,36 milioni di acri bruciati a causa di tali incendi.

Tra gli incendi peggiori c'è anche quello della California del 8/8/2018 già trattato nel [2.4.4](#).

2.8.2 Diagramma a dispersione peggiori incendi

Nel seguente grafico si può notare come in base alla potenza irradiata in Mega Watt e in base all'espansione dell'incendio in chilometri quadri si hanno diversi punti che rappresentano i singoli incendi. Da un punto di vista tecnico, è stato utilizzato uno

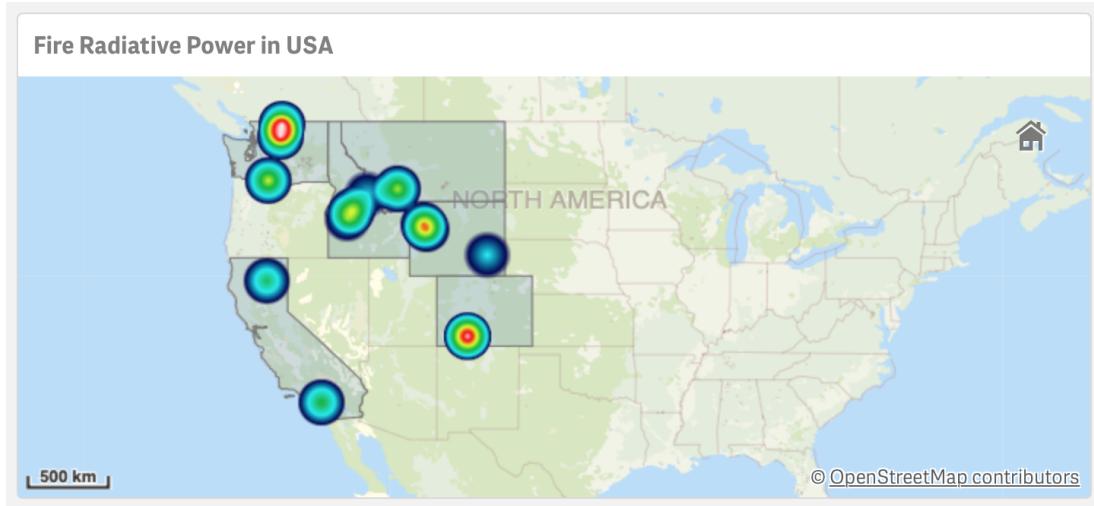


scatter plot, dove sull'asse delle ascisse x si ha la Sum della "frp" che parte da 8.000MW ed arriva a 15.000MW in modo da escludere gli incendi con minor rilevanza. Al contrario sull'asse delle ordinate y si nota la Sum della "Scan" che parte in questo caso da 4Km^2 ed arriva a 5Km^2 . Per quanto riguarda i colori è stato usato un gradiente sequenziale che prende come misura la Sum della "frp" (si è voluto dare più peso alla potenza irradiata che all'espansione dell'incendio).

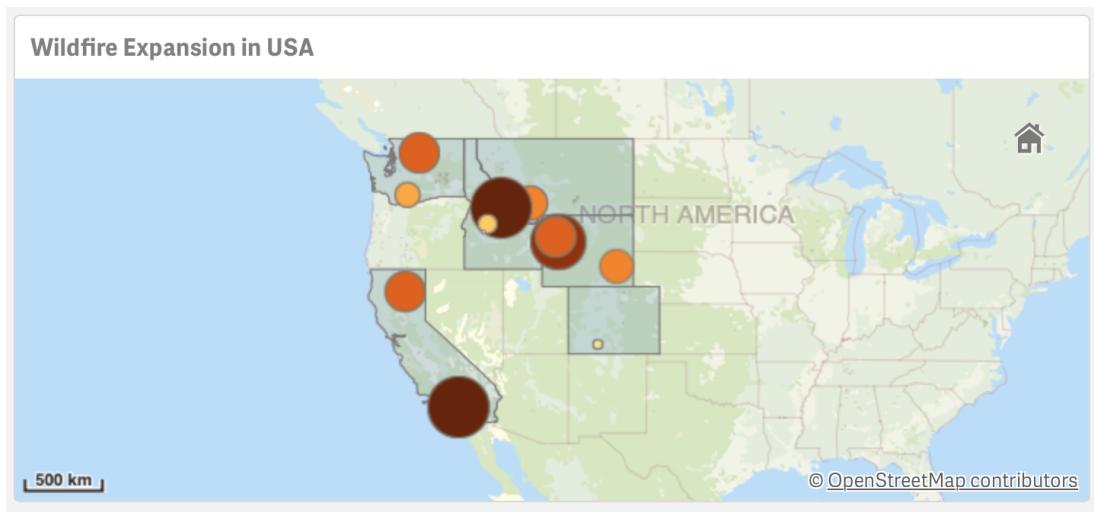
2.8.3 Mappe di Potenza irradiata e di Espansione dei peggiori incendi

Per mostrare dove e con che intensità si è verificato un determinato incendio (che si può selezionare dalla tabella) o più valori, si è pensato di inserire due mappe:

- **Mappa di Potenza irradiata:** Da un punto di vista tecnico, il grafico è costituito da due livelli: livello densità dove come dimensione sono state inserite la latitudine e la longitudine dei vari incendi, mentre come misura della densità di peso e di raggio è stata inserita la Sum della "frp"; il secondo livello è quello delle aree dove come dimensione è stata inserita la *us_states.Area*.



- **Mappa di Espansione:** Da un punto di vista tecnico, il grafico è costituito da



due livelli: livello punti dove come dimensione sono state inserite la latitudine e la longitudine dei vari incendi, mentre come misura della densità di peso e di raggio è stata inserita la Sum della "Scan"; il secondo livello è quello delle aree dove come dimensione è stata inserita la *us_states.Area*.

Possiamo notare che a zone con forte potenza irradiata (in rosso) possono corrispondere delle espansioni non elevatissime. Stessa cosa possiamo dire per gli incendi con espansione maggiore che spesso non hanno potenza irradiata elevatissima, anche se comunque di valore non sicuramente bassa.

Inoltre dalla mappa si nota come i peggiori incendi sono concentrati soprattutto nella West Coast negli stati del: Washington, Montana, California, Colorado, Nevada, Idaho

e British Columbia.

2.9 Analisi per Tipo

Come si può vedere da in alto a destra nella dashboard, la NASA distingue nel suo dataset 4 tipi di incendi:

- **0:** Incendi di vegetazione/boschivo;
- **1:** Incendio ti tipo vulcanico;
- **2:** Incendi causati da altre fonti sulla terra;
- **3:** A largo.

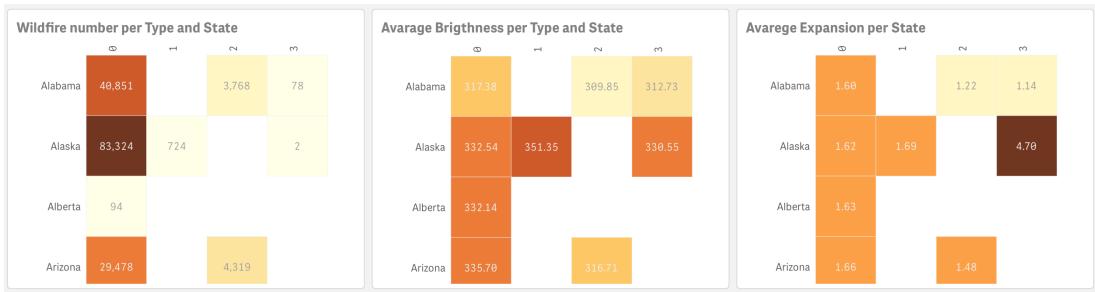
In questa dashboard si vuole fare appunto una analisi per tipi, andando a studiare innanzitutto il numero di incendi, la brillantezza media e l'espansione media divisa per i vari stati, per poi andare a studiare un confronto di questi ultimi due e della potenza irradiata media, divisa per tipo. In basso a sinistra ci sono infine i vari filtri per specializzare l'analisi sia in termini di luogo che in termini di arco temporale.



2.9.1 Analisi del numero di incendi, della brillantezza media e dell'espansione media per Stato

Il grafico utilizzato è una Mappa di calore (in inglese: heatmap) è una rappresentazione grafica dei dati dove i singoli valori contenuti in una matrice sono rappresentati da colori.

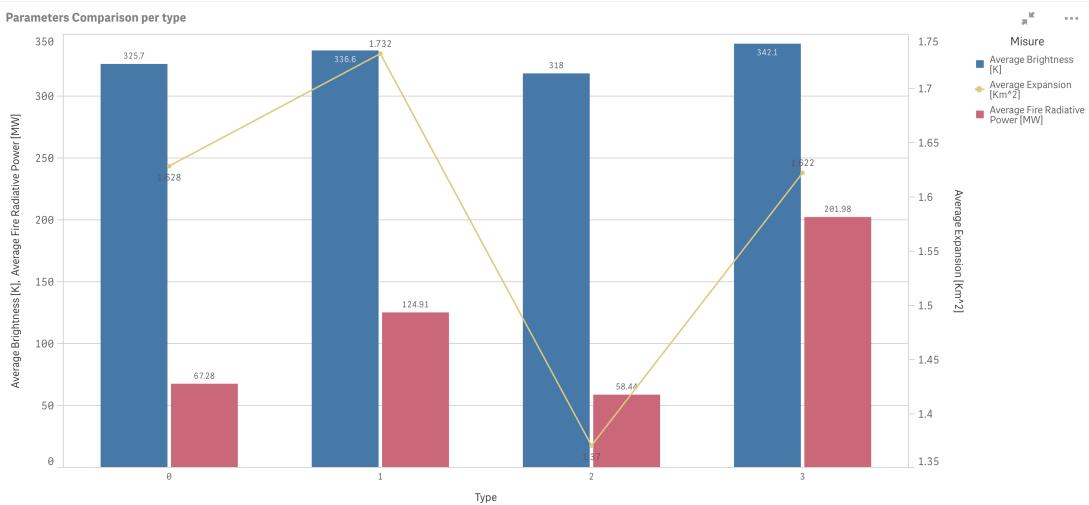
CAPITOLO 2: QLIK SENSE



In particolare la scala di colori usata per i tre grafici è la stessa, dove più il valore tende al marrone scuro, più quel valore è alto. Scorrendo verso il basso verranno inoltre visualizzati tutti gli altri stati in ordine alfabetico o, eventualmente, si può cercare lo stato al quale siamo interessati utilizzando il filtro sulla sinistra.

Attraverso l'utilizzo dei filtri presenti infatti si può fare un analisi più puntale sullo stato che interessa analizzare e vedere quale dei quattro tipi di incendi è più presente e quale meno.

2.9.2 Parametri a confronto per Tipo



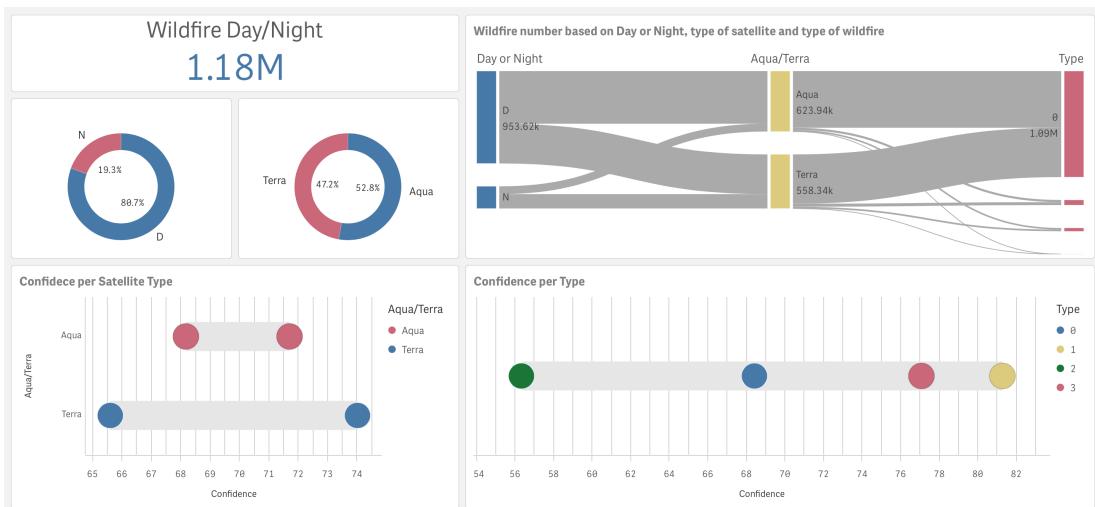
Riprendendo i tre grafici presenti in [2.7.3](#), dove si rappresentava la brillantezza, l'espansione e la potenza irradiata media, si vuole fare un'ulteriore analisi dividendo per tipo.

Il grafico utilizzato è un grafico multiplo dove si può notare che, per quanto riguarda la potenza media irradiata dall'incendio, i valori più alti si trovano per incendi di tipo 1 (vulcanici) e di tipo 3 (a largo). Questo riflettendo è plausibile dato che la potenza sprigionata da uno scoppio di una nave a largo è elevatissima, come d'altronde lo è

anche l'esplosione di un vulcano attivo, mentre risulta essere minore in un incendio di tipo vegetativo.

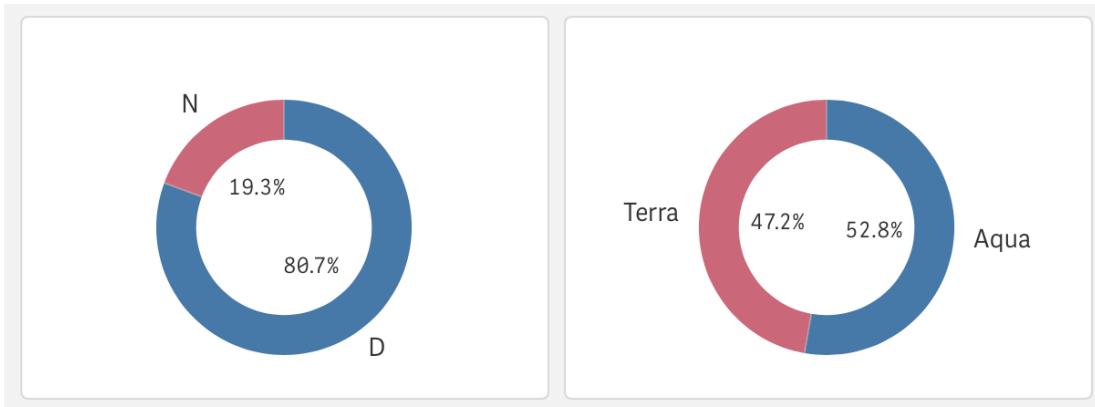
2.10 Analisi per Satellite

Quest'ultima dashboard ha come obiettivo lo studio di alcuni parametri fin qui non ancora considerati, come gli incendi che avvengono di giorno o di notte, gli incendi che sono stati rilevati dal satellite Aqua piuttosto che Terra e la confidenza sulla reale presenza di tale rilevazione di incendio. Questo valore si basa su una raccolta di quantità di algoritmi intermedi utilizzati nel processo di rilevamento. Ha lo scopo di aiutare gli utenti a valutare la qualità dei singoli pixel di hotspot / fuoco. Le stime di fiducia variano tra 0 e 100% e vengono assegnate a una delle tre classi di fuoco (fuoco a bassa confidenza, fuoco a confidenza nominale o fuoco a confidenza alta). In alto a sinistra si può notare una KPI che ci indica il valore della Sum della *wildfire_presence*.



2.11 Analisi Giorno/Notte e Terra/Aqua

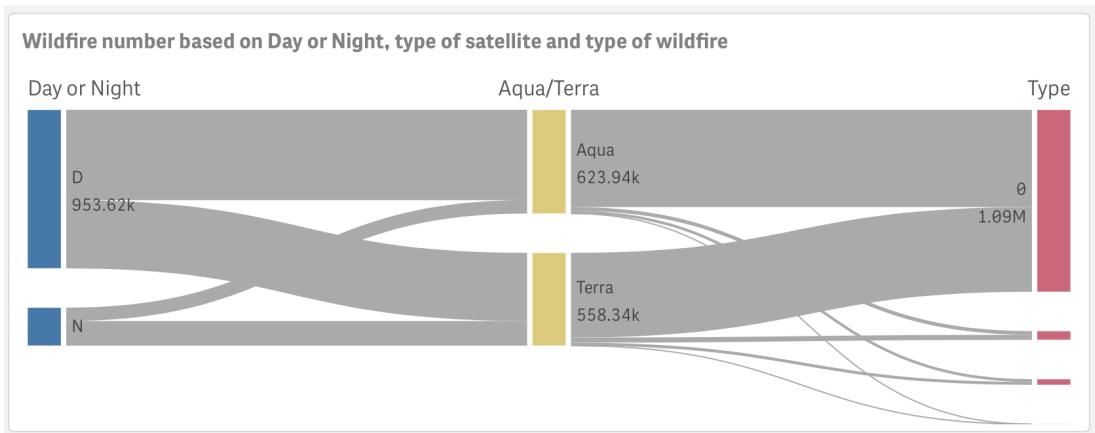
Come primi veri grafici nella dashboard si notano i due diagrammi a torta che rappresentano rispettivamente: la percentuale di incendi avvenuti di notte rispetto che di giorno e la percentuale di incendi rilevati attraverso il satellite "Aqua" rispetto che il satellite "Terra". Come dimensione (fetta) è stata usata la "daynight" nel primo caso mentre



la *fire.archive_satellite* nel secondo. Per quanto riguarda le misure sono entrambe calcolate attraverso la Sum della *wildfire_presence*.

Si può notare come previsto che la maggior parte degli incendi venga rilevata di giorno, ovvero quando le temperature sono più alte e c'è una probabilità più alta del formarsi di un incendio.

2.12 Analisi Giorno/Notte, Terra/Aqua e Tipo



Il diagramma di Sankey è un particolare tipo di diagramma di flusso in cui l'ampiezza delle frecce è disegnata in maniera proporzionale alla quantità di flusso. In generale i diagrammi di Sankey accentuano visivamente i grandi trasferimenti o flussi all'interno di un sistema: sono perciò utili in questo caso per individuare i contributi dominanti in un flusso complessivo.

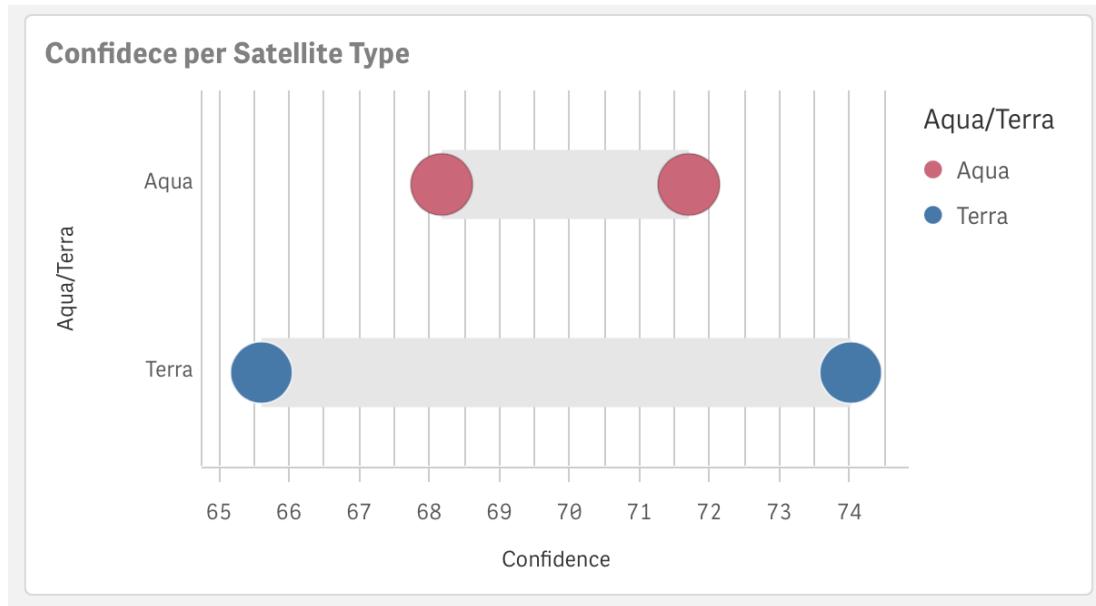
In particolare tale diagramma si dirama due volte: dalla distinzione tra incendi avvenuti di notte o di giorno, si dirama in incendi registrati tramite il satellite Aqua o Terra e

da qui avviene una seconda diramazione in base al tipo di incendio. Dal punto di vista tecnico, come dimensioni sono presenti la *daynight*, la *fire_archive.satellite* e la *fire_archive.type*, mentre come misura la Sum della *wildfire_presence*.

2.13 Analisi sulle Confidenze

Infine come ultimi grafici della dashboard sono presenti le confidenze sia per i due tipi di satellite Aqua o Terra e le confidenze in base al tipo di incendio.

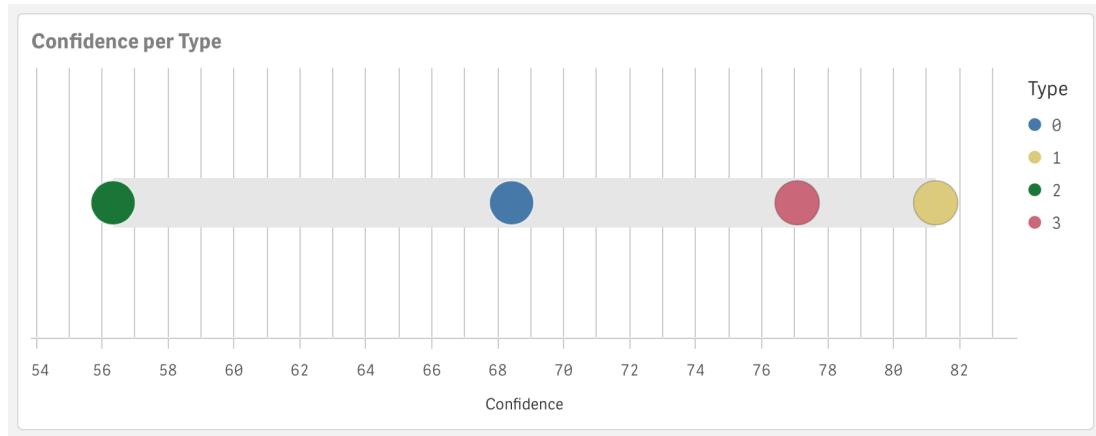
- **Confidenza per tipo di satellite:** Notiamo come per entrambi i satelliti la con-



fidenza della presenza di un incendio di giorno o di notte sia leggermente diversa, e si abbia una confidenza maggiore proprio di notte. Infine si osserva anche il fatto che nel satellite Terra si abbia una differenza di confidenza maggiore tra giorno e notte.

Dal punto di vista tecnico come punto è stata usata la dimensione *daynight* e come asse Y la dimensione *fire_archive.satellite*. Al contrario come asse X è stato utilizzato la Avg della *fire_archive.confidence*.

- **Confidenza per tipo di incendio:** Esattamente da quanto ci si poteva aspettare, gli incendi con confidenza maggiori sono quelli di tipo vulcanico dato che facilmente rilevabili dal satellite, al contrario quelli meno rilevabili con unicamente il 56% di confidenza sono quelli causate da fonti vario genere sulla terra ma non boschivi.



Da un punto di vista tecnico è stata utilizzata come dimensione per i vari punti la *fire_archive.type*, mentre come misura la Avg della *fire_archive.confidence*. Per quanto riguarda i colori, questi sono differenti in base alla misura (tipo di incendio) come si può vedere dalla legenda.

3. Tableau

Dopo una prima analisi puramente descrittiva e diagnostica realizzata utilizzando Qlik è stata svolta un'analisi analoga con l'aggiunta di una parte predittiva usando Tableau, altro tool di business intelligence leader nel mercato. Si è cercato di analizzare aspetti simili a quelli visti in precedenza, essendo i dati a disposizione sempre gli stessi, ma usando strumenti diversi e considerando contesti differenti. Sono state realizzate diverse dashboard, ognuna con obiettivo differente.

Tableau come Qlik è un software di Business Intelligence (BI) molto utilizzato che permette di realizzare analisi sia di tipo descrittivo che diagnostico tramite un'illustrazione intuitiva e facile da comprendere dei dati, come Qlik, ma in più Tableau permette anche la realizzazione di analisi predittive. Anche la realizzazione dei Grafici contenuti negli Sheet, poi combinati per la realizzazione delle Dashboard, risulta facile. Anche Tableau presenta la capacità di cambiare dinamicamente le visualizzazioni in base all'uso dei filtri e alle selezioni che vengono fatte dagli utenti. Questo permette quindi non solo di rispondere alle domande di coloro che hanno chiesto la realizzazione delle visualizzazioni, ma anche a possibili domande di follow-up, ovvero domande che qualcuno potrebbe porsi a seguito dell'interazione con una Dashboard piuttosto che con un Sheet specifico. La caratteristica che più distingue Tableau da Qlik è la già accennata possibilità di realizzare delle analisi predittive dei dati temporali, permettendo di predire con un certo intervallo di confidenza l'andamento temporale e le linee di trend dei nostri dati. La versione utilizzata è Tableau Desktop ed è la versione Desktop del software

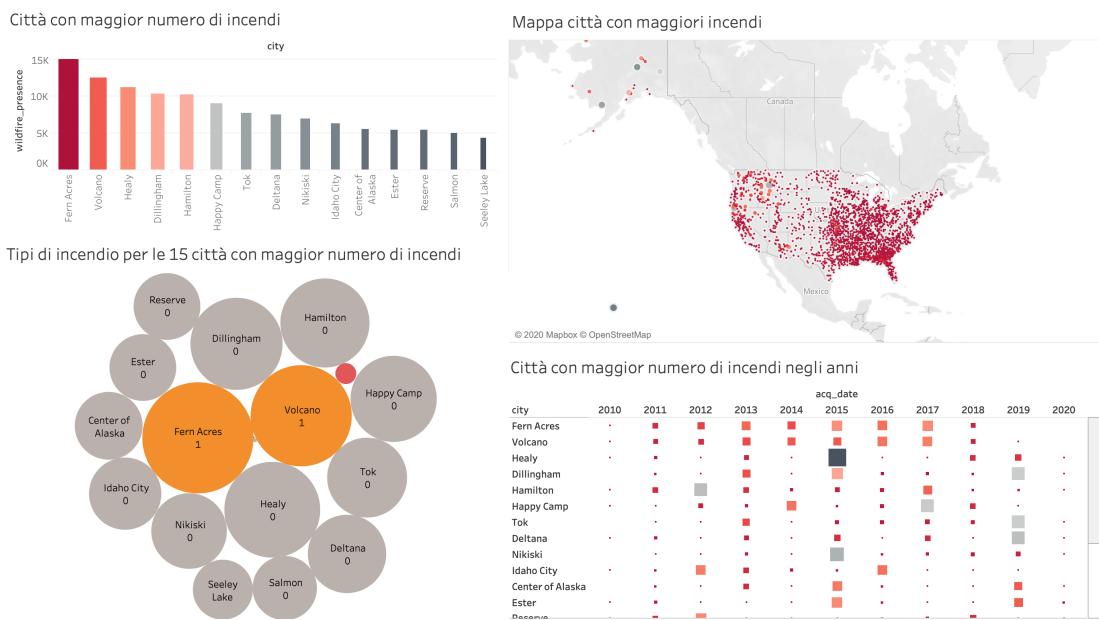
3.1 Gestione dati

Per prima cosa, una volta caricato il data set sono state necessarie delle operazioni preliminari, ovvero è stato necessario indicare i tipi delle varie colonne che, in alcuni casi,

CAPITOLO 3: TABLEAU

Tableau non è riuscito ad individuare correttamente. In particolare latitude e longitude non sono stati riconosciuti subito come tali ma come numeri, poi è stato assegnato il corrispettivo ruolo geografico a city e state; l'acquisition date (data) è stata settata come data, satellite come stringa e le restanti colonne come valori numerici. Poi, essendo state le colonne inizialmente non caricate correttamente è stato necessario cambiare il separatore campo da punto e virgola a virgola. Infine per leggere correttamente i valori numerici è stato necessario settare come impostazioni locali la voce “inglese (Stati Uniti)”.

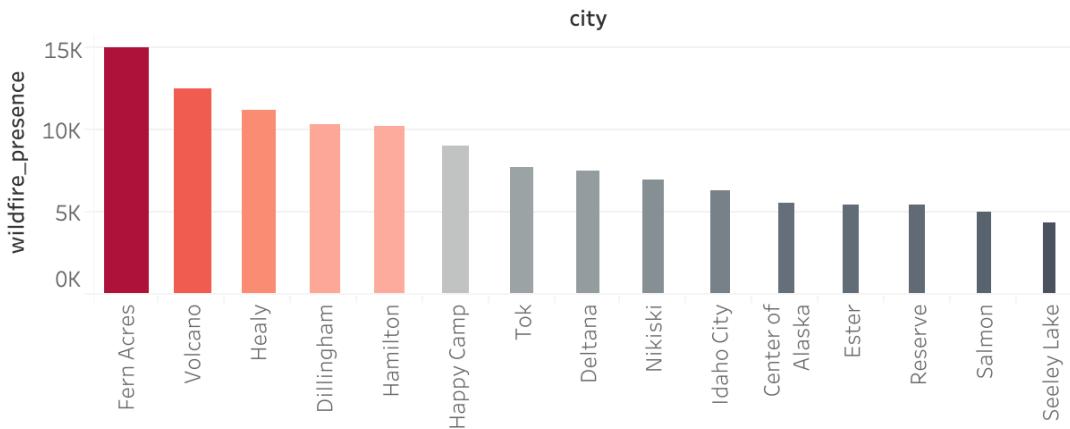
3.2 Città con maggiori incendi



3.2.1 Città con maggior numero di incendi

La prima dashboard si concentra su un’analisi per città utilizzando quattro grafici per visualizzare i dati. Il primo è un istogramma in cui sono state rappresentate le 15 città in cui vi sono stati più incendi in totale nell’arco di tempo complessivo coperto dal dataset. Si è scelto, per accentuare le differenze dei valori visualizzati, di utilizzare delle scale di colori e di grandezze per le varie colonne, basate sui valori delle ordinate. Si può notare che la località che ha registrato il maggior numero di incendi è stata Fern Acres, nelle isole Hawaii, con quasi 15.000 incendi complessivi, la seconda è stata Volcano, sempre

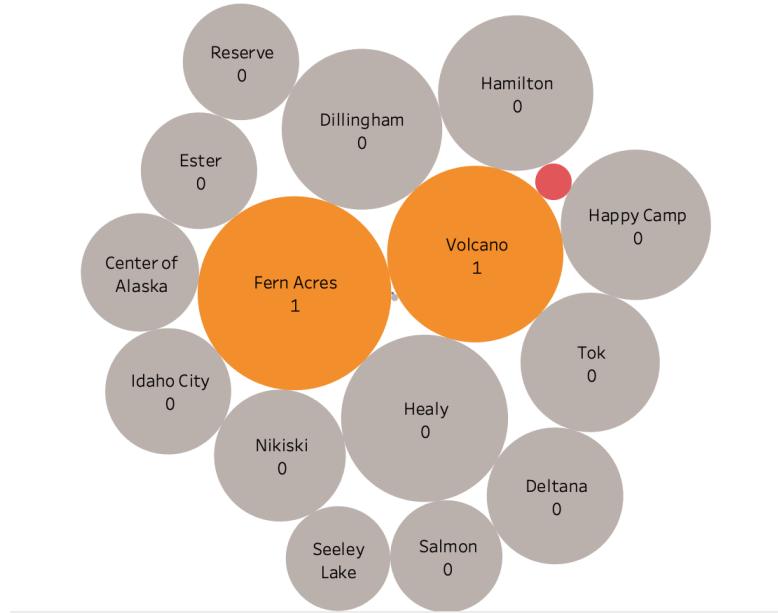
CAPITOLO 3: TABLEAU



nelle Hawaii, con un totale di 12.417 incendi ed infine Healy e Dillingham, in Alaska, con rispettivamente 11.122 e 10.153 incendi. Si nota subito che le Hawaii e l'Alaska sono paesi che presentano molti incendi a causa della vasta presenza di vulcani (nelle Hawaii) e di aree boschive (in Alaska) soggette sempre più ad incendi negli ultimi anni soprattutto dovuti al surriscaldamento globale e alle estati sempre più calde e secche.

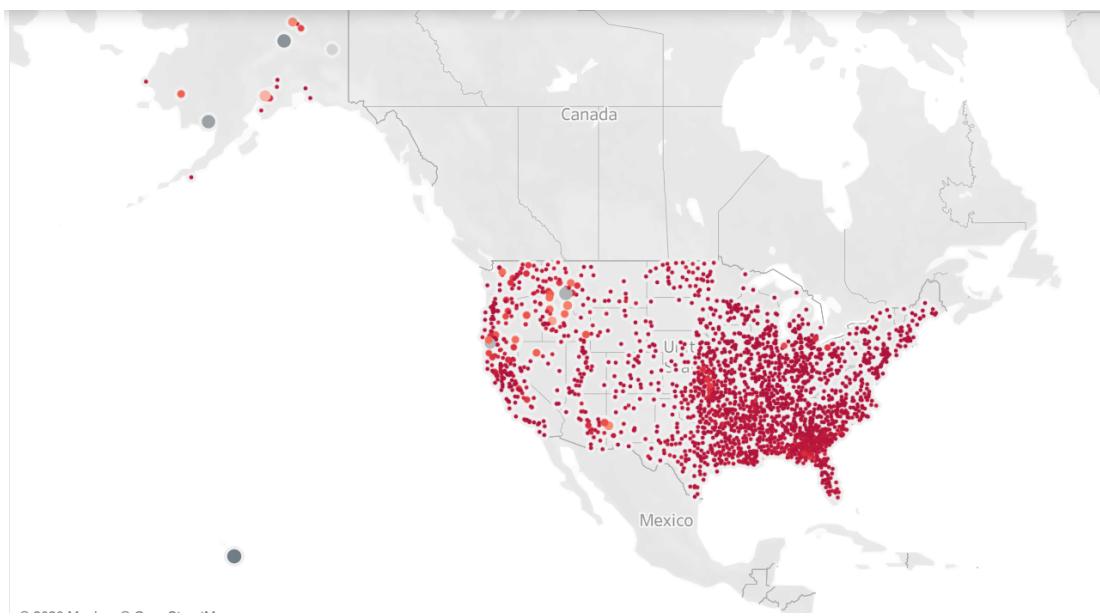
3.2.2 Tipi di incendi per le 15 città con maggior numero di incendi

Queste considerazioni sono confermate dal grafico a bolle che mostra le 15 località con più incendi totali per tipo di incendio dando ad ogni bolla una grandezza relativa al valore corrispondente e un colore che indica i tipi di incendi per città.



Quindi le località Hawaiiane Fern Acres e Volcano presentano prevalentemente incendi di tipo 1, ovvero di origine vulcanica, invece le località dell'Alaska, Healy e Dillingham, presentano soprattutto incendi di tipo 0, cioè dovuti alla vegetazione. Guardando con attenzione si nota che ovviamente anche Volcano e Fern Acres hanno avuto incendi dovuti alla vegetazione, ma molto pochi e trascurabili rispetto a quelli di origine vulcanica. Infine, si osserva che la località di Hamilton, in Montana, oltre a presentare soprattutto incendi di tipo 0 ne ha avuti anche di tipo 2, cioè dovuti ad altre cause.

3.2.3 Mappa delle città con maggior incendi

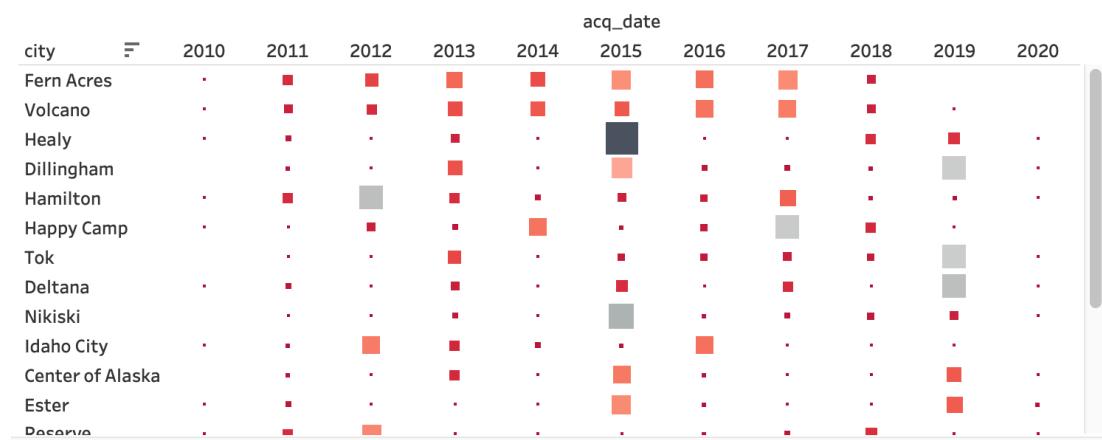


Per localizzare a livello geografico gli incendi nella dashboard si è scelto di mettere una mappa degli USA in cui sono indicati come punti tutti gli incendi, localizzati usando la latitudine e la longitudine. Le dimensioni dei punti sono date dal numero complessivo di incendi che si sono verificati in quell'area.

3.2.4 Città con maggior numero di incendi negli anni

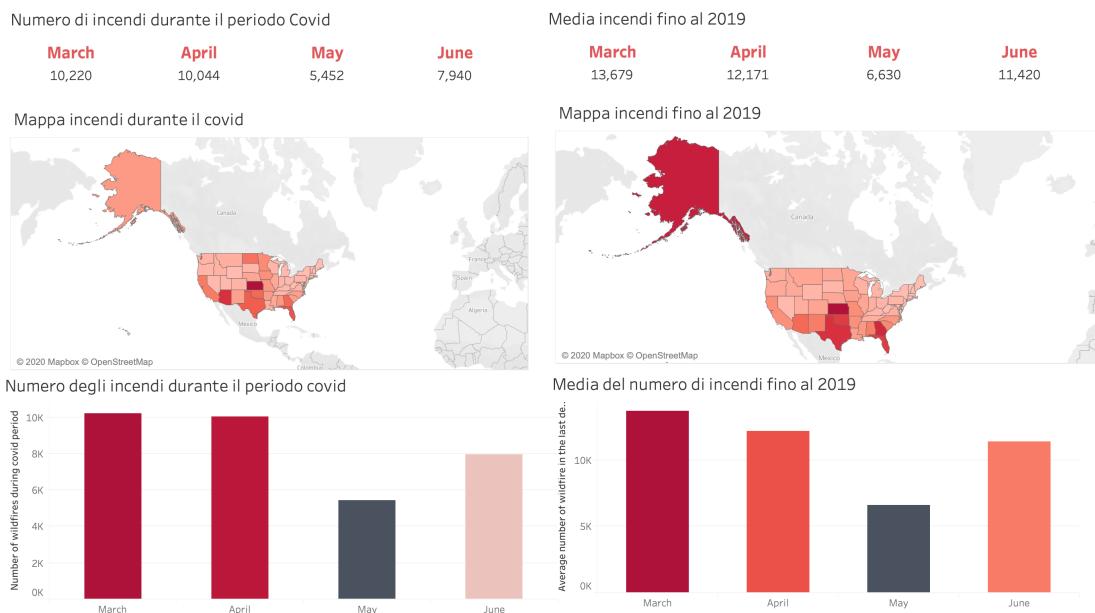
Infine è stata realizzata una heatmap in cui sono state rappresentate le 20 città con più incendi complessivi mostrando gli anni in cui se ne sono verificati di più. Considerando che i dati del 2010 e del 2020 non sono completi, i dati relativi a questi anni non sono veritieri ovviamente. Si nota poi che per quanto riguarda le due località Hawaiiane di Fern Acres e Volcano il numero di incendi è variato negli anni in maniera parallela, ciò

CAPITOLO 3: TABLEAU



può far dedurre che le attività vulcaniche che li ha causati sono correlate tra loro e si verificano in parallelo. Per quanto riguarda la località in Alaska di Healy si nota che nel 2015 vi è stato un forte picco nel numero di incendi rispetto agli altri anni, invece per quanto riguarda Dillingham, il 2019 e il 2015 sono stati gli anni con più incendi. Per concludere la descrizione di questa dashboard si ricorda che l'analisi sul numero di incendi non dice niente sulla loro effettiva intensità (per cui invece sono utili grandezze come la frp, la brightness o la scan) ma è solo un'analisi sulla loro numerosità.

3.3 Incendi durante il periodo del Covid



Consideriamo ora la seconda dashboard in cui viene analizzato l'effetto della prima ondata della pandemia di COVID-19 e dei conseguenti provvedimenti attuati dai vari Stati

CAPITOLO 3: TABLEAU

sul numero di incendi avvenuti negli USA. I vari Stati hanno adottato politiche diverse limitando gli spostamenti, i raduni e in alcuni casi imponendo un vero e proprio lock down nei mesi di Marzo, Aprile, Maggio e Giugno. Di conseguenza è ragionevole aspettarsi un calo negli incendi rispetto agli stessi mesi degli anni precedenti, in particolare dovuto all'ovvio calo di incendi dolosi o dovuti a incidenti stradali per esempio.

3.3.1 Confronto numero medio di incendi nel periodo del Covid e negli anni precedenti

Nella dashboard in alto sono riportati i KPI relativi al numero totale di incendi nel quadrimestre (Marzo, Aprile, Maggio e Giugno) del 2019 a destra e le medie di incendi negli stessi mesi degli anni precedenti a destra.

Numero di incendi durante il periodo Covid

March	April	May	June
10,220	10,044	5,452	7,940

Media incendi fino al 2019

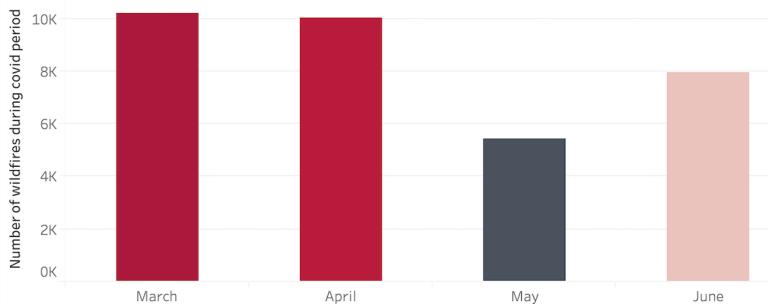
March	April	May	June
13,679	12,171	6,630	11,420

3.3.2 Confronto numero medio di incendi nel periodo del Covid e negli anni precedenti con istogrammi

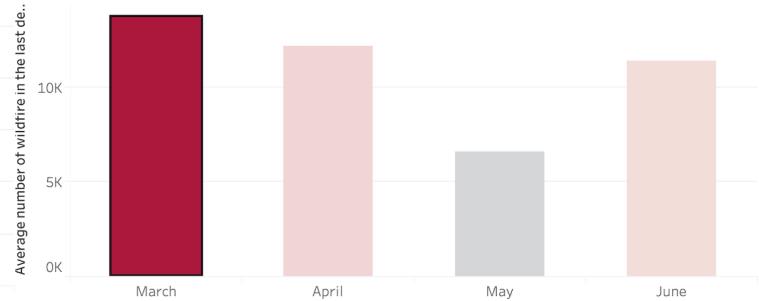
Gli stessi dati sono riportati in due istogrammi nella parte bassa della dashboard in cui sono dati dei colori alle varie barre in base all'entità del valore sulle ordinate. Si nota che già ad inizio epidemia, a Marzo 2020, il numero di incendi che si sono verificati sono stati più di 3000 in meno rispetto a Marzo degli anni precedenti e vi è un netto calo anche nei mesi successivi.

CAPITOLO 3: TABLEAU

Numero degli incendi durante il periodo covid



Media del numero di incendi fino al 2019

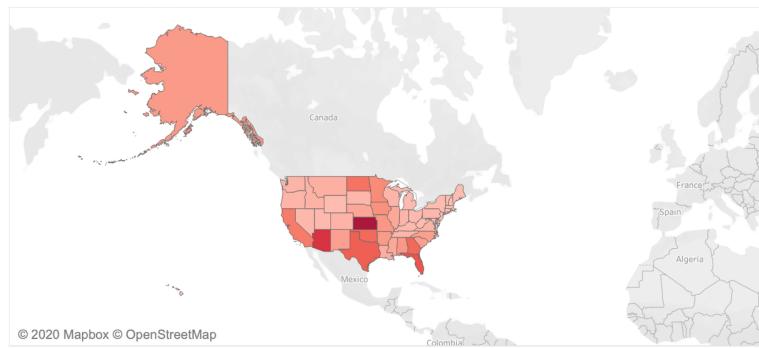


3.3.3 Confronto mappe incendi tra il periodo del Covid e gli anni precedenti

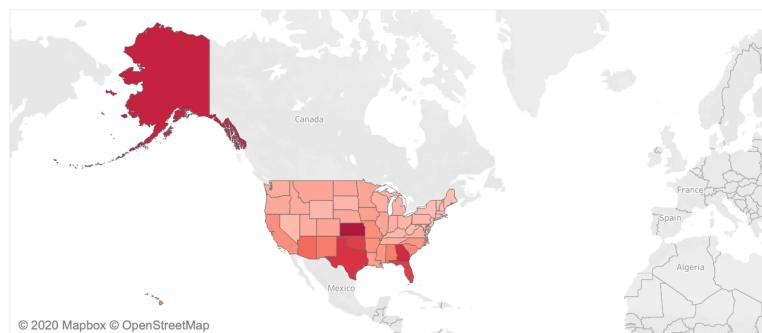
Poi si è voluto mostrare come si sono distribuiti gli incendi nel quadri mestre considerato degli anni precedenti e del 2020, tra i vari Stati degli USA, utilizzando due mappe in cui è associato un colore ad ogni Nazione in relazione al numero di incendi avvenuti. Si nota come in alcuni paesi vi è stato un netto calo degli incendi nel quadri mestre considerato del 2020, ad esempio in Alaska si è passati da un totale di 3739 incendi nel quadri mestre degli anni precedenti a 896 del periodo del Covid (una riduzione di più del 70%), oppure in Georgia in cui si è passati da 3444 incendi a 2007 (riduzione di più del 40%), oppure in Texas da 3217 a 2212 (riduzione di più del 30%). Altri Stati invece hanno stranamente avuto la tendenza opposta, ovvero hanno avuto più incendi nel periodo Covid rispetto allo stesso periodo degli anni precedenti, ad esempio l'Arizona in cui si è passati da 1981 incendi a 3216 (aumento del 160%) oppure in North Dakota in cui si è passati da 563 a 1777 (aumento di più del 300%).

CAPITOLO 3: TABLEAU

Mappa incendi durante il covid



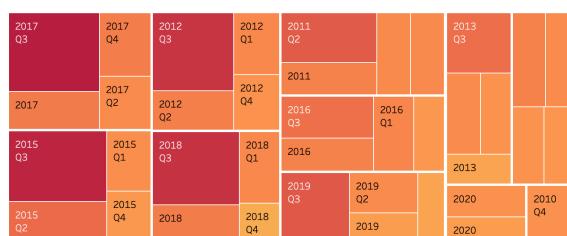
Mappa incendi fino al 2019



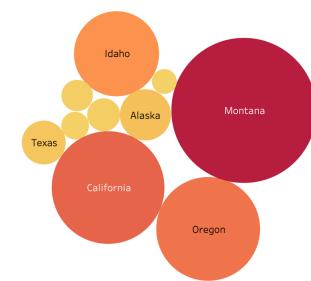
3.4 Analisi del trimestre con più incendi

Un‘aspetto interessante sul quale focalizzarsi è l’andare a studiare quale sia stato il trimestre negli ultimi dieci anni con più incendi.

Trimestri negli ultimi dieci anni con più incendi



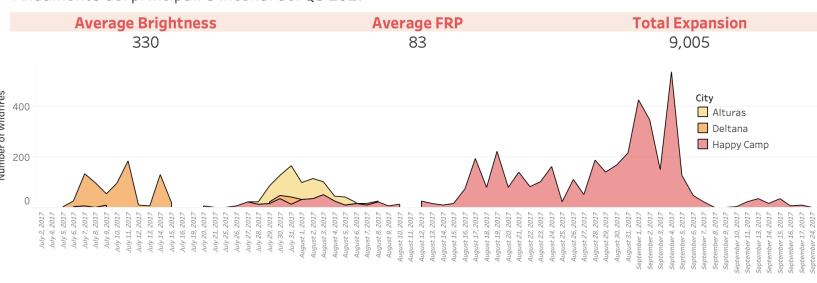
Analisi per stati del terzo trimestre del 2017



Numero di incendi per città del Q3 del 2017

City	Number of Wildfires
Happy Camp	4,242
Deltana	803
Alturas	531
Tok	486
Yosemite Valley	453
East Porterville	420
Mariposa	388
Weaverville	377
College	348
Kernville	286
Oakhurst	220
Nikiski	218

Andamento dei principali 3 incendi del Q3 2017



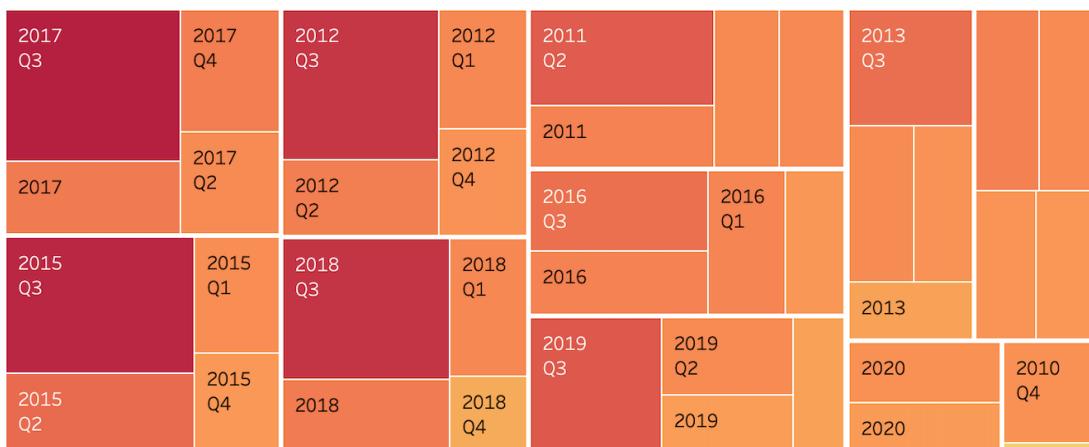
In particolare si è voluto andare a graficare sia gli stati che le città nelle quali si sono

verificati più incendi in quel trimestre.

3.4.1 Trimestre negli ultimi 10 anni con più incendi

Nella terza dashboard si fa una analisi più approfondita a livello temporale, per prima cosa si deve considerare il grafico ad albero in alto a sinistra in cui si analizza quali trimestri sono stati i peggiori a livello di numero di incendi negli anni studiati. Per ogni anno è dato un colore a ogni trimestre in base al numero di incendi avvenuti e gli anni sono ordinati da sinistra a destra per numero di incendi decrescenti. Si deduce che ogni anno il trimestre peggiore a livello di incendi è il terzo, ovvero Luglio, Agosto, Settembre, come ci si poteva aspettare, essendo mesi estivi e più caldi in cui gli incendi si sviluppano più facilmente. Infine, si nota che l'anno che ha avuto più incendi è stato il 2017 sul cui terzo trimestre (con un totale di 65950 incendi) ci si concentra nelle analisi dei successivi grafici della dashboard.

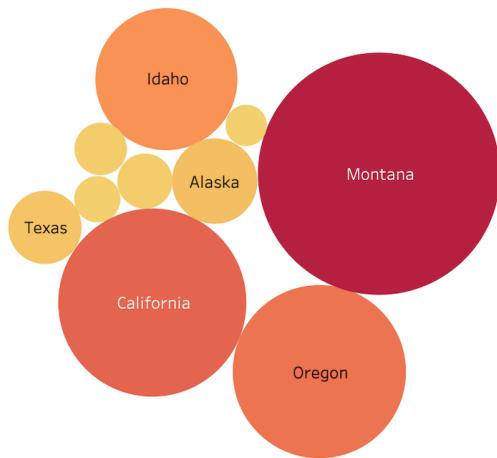
Trimestri negli ultimi dieci anni con più incendi



3.4.2 Analisi per Stati del terzo trimestre del 2017

Nel grafico a bolle si mostrano gli Stati che hanno avuto più incendi in tale trimestre del 2017, rappresentando ogni Stato con una bolla di dimensioni e colori basati sul numero di incendi verificatisi. Come si può vedere il Montana è stato il Paese con più incendi (16693) in tale trimestre seguito dalla California (10068) e Oregon (8537).

Analisi per stati del terzo trimestre del 2017



3.4.3 Numero di incendi per città del terzo trimestre del 2017

Dopo l'analisi per Stato vi è una tabella in cui ci si concentra sulle città che sono state luogo di più incendi nel terzo trimestre del 2017 in cui è associato un colore in base al valore riportato per ogni città.

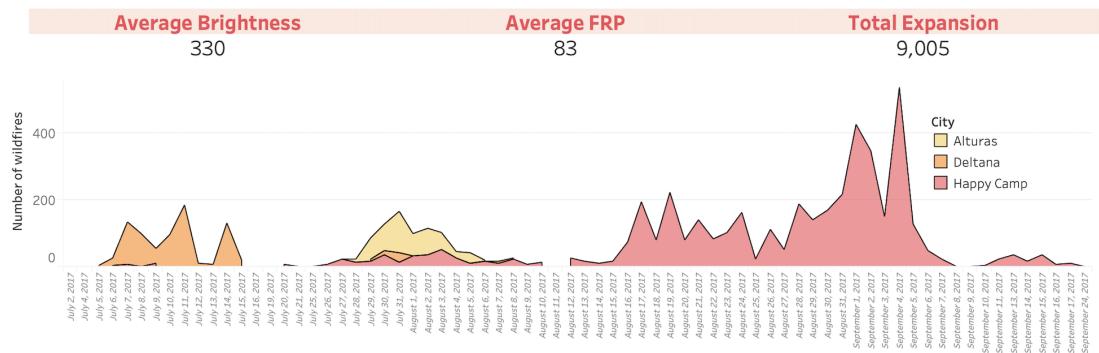
Numero di incendi per città del Q3 del 2017

Happy Camp	4,242
Deltana	803
Alturas	531
Tok	486
Yosemite Valley	453
East Porterville	420
Mariposa	388
Weaverville	377
College	348
Kernville	286
Oakhurst	220
Nikiski	218

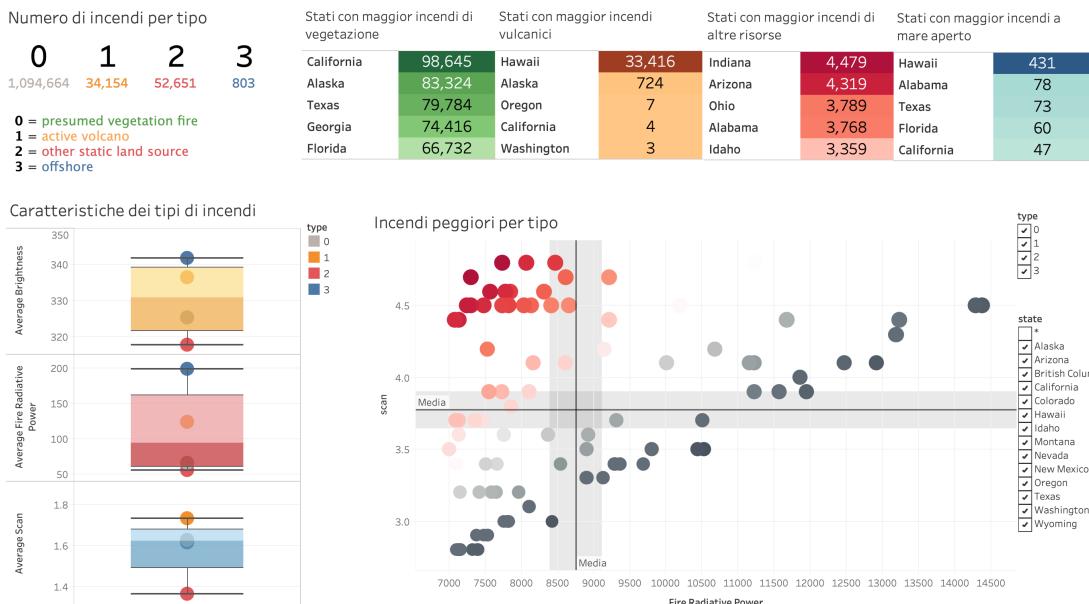
La località che domina su tutte per numero di incendi in questo periodo è Happy Camp con 4242 incendi totali, seguita da Deltana con 803 e Alturas con 531.

3.4.4 Andamento nelle 3 località più colpite del 3° trimestre del 2017

Infine delle 3 località che hanno avuto più incendi in questo trimestre è rappresentato con un grafico a linee l'andamento giorno per giorno del numero di incendi durante tutto il trimestre, in cui si vede che gli incendi avvenuti ad Happy Camp sono iniziati a fine Luglio e durati fino a fine Settembre con i maggiori picchi tra fine Agosto e inizio Settembre. A Deltana invece gli incendi si sono concentrati soprattutto ad inizio Luglio ed infine in Alturas si sono verificati soprattutto tra fine Luglio e inizio Agosto. Sono anche riportate tre KPI che mostrano l'entità della brightness e della frp media e anche la totale espansione degli incendi che si sono avuti in queste tre località più colpite durante il terzo quadrimestre del 2017.



3.5 Analisi per tipo



3.5.1 Numero di incendi per tipo

Nella quarta dashboard invece è fatta un'analisi per tipo degli incendi avuti negli USA nell'arco di tempo considerato dal 2010 al 2020. In alto a sinistra sono riportati quattro KPI che mostrano il totale di incendi per ogni tipo che si sono avuti negli USA. Si nota che predominano gli incendi di tipo 0 negli USA, ovvero quelli dovuti alla vegetazione, essendovi in molte regioni degli Stati Uniti grandi aree boschive a rischio di incendi soprattutto nei periodi caldi e secchi. Questo tipo di incendi vede infatti un totale di 1094664 (più del 90% del numero complessivo). Invece gli incendi del tipo 2 ammontano a un totale di 52651 e sono i secondi più presenti in USA seguiti da quelli del tipo 3, cioè vulcanici con 34154 incendi. Infine gli incendi più rari (ammontano a 803) sono quelli del tipo 3, ovvero quelli che avvengono a largo e non sulla terraferma come ad esempio esplosioni o incendi su petroliere o in stabilimenti presenti per diversi utilizzi in mare.

Numero di incendi per tipo



0 = presumed vegetation fire

1 = active volcano

2 = other static land source

3 = offshore

3.5.2 Tabelle per tipo

Dopo questa analisi generale per tipo di incendio negli USA ci si è voluti concentrare su che Paesi sono colpiti principalmente da ogni tipo di incendio, a tal fine sono state realizzate le quattro tabelle in alto a destra. Si può osservare come lo Stato più colpito

CAPITOLO 3: TABLEAU

da incendi in aree con vegetazione sia la California, come dimostrato dal recente incendio avvenuto a fine estate 2020 (non presente nel nostro dataset) che è durato per settimane e, a detta degli esperti, è stato considerato addirittura come l'incendio peggiore che abbia mai colpito la California inghiottendo un'area di 31000 ettari, pari a due volte la superficie di San Francisco! Per quanto riguarda gli incendi di origine vulcanica il paese che ne è più a rischio è sicuramente quello delle Hawaii, seguito a distanza dall'Alaska. Tale risultato è logico, essendo le Hawaii isole di origine vulcaniche con vari vulcani ancora attivi, tra cui il Mauna Loa, il più grande sulla Terra, e l'Alaska è disseminata di vulcani ancora attivi. I Paesi che hanno avuto più incendi di tipo 2 sono stati invece Indiana e Arizona, mentre la maggior parte degli incendi a largo sono avvenuti ancora nelle Hawaii.

Stati con maggior incendi di vegetazione

California	98,645	Hawaii	33,416
Alaska	83,324	Alaska	724
Texas	79,784	Oregon	7
Georgia	74,416	California	4
Florida	66,732	Washington	3

Stati con maggior incendi vulcanici

Stati con maggior incendi di altre risorse

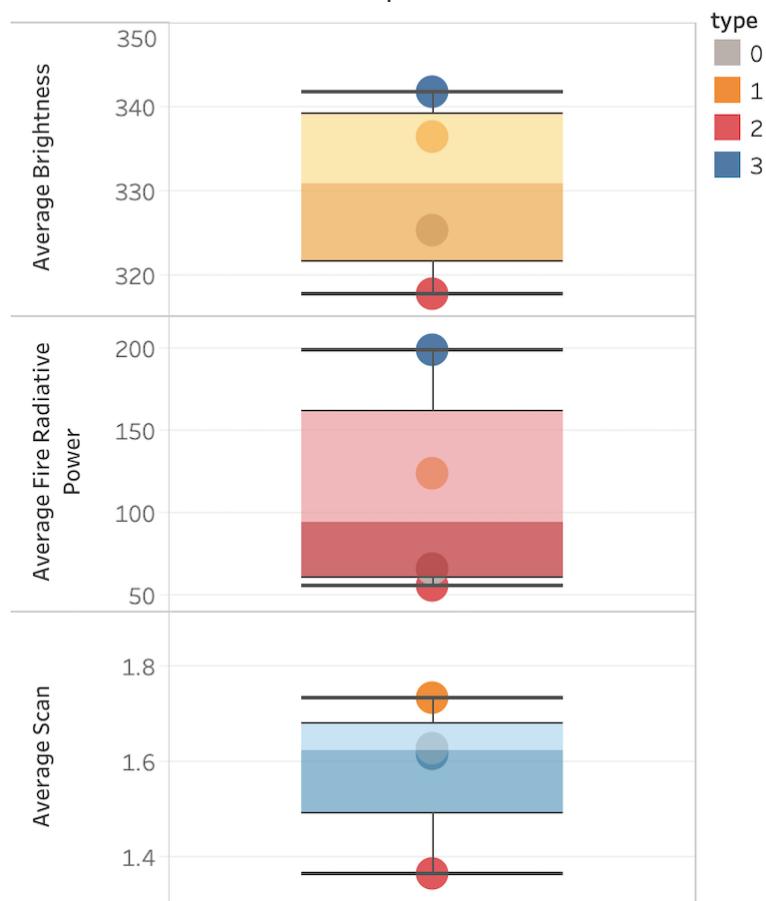
Indiana	4,479	Hawaii	431
Arizona	4,319	Alabama	78
Ohio	3,789	Texas	73
Alabama	3,768	Florida	60
Idaho	3,359	California	47

Stati con maggior incendi a mare aperto

3.5.3 Caratteristiche dei tipi di incendi

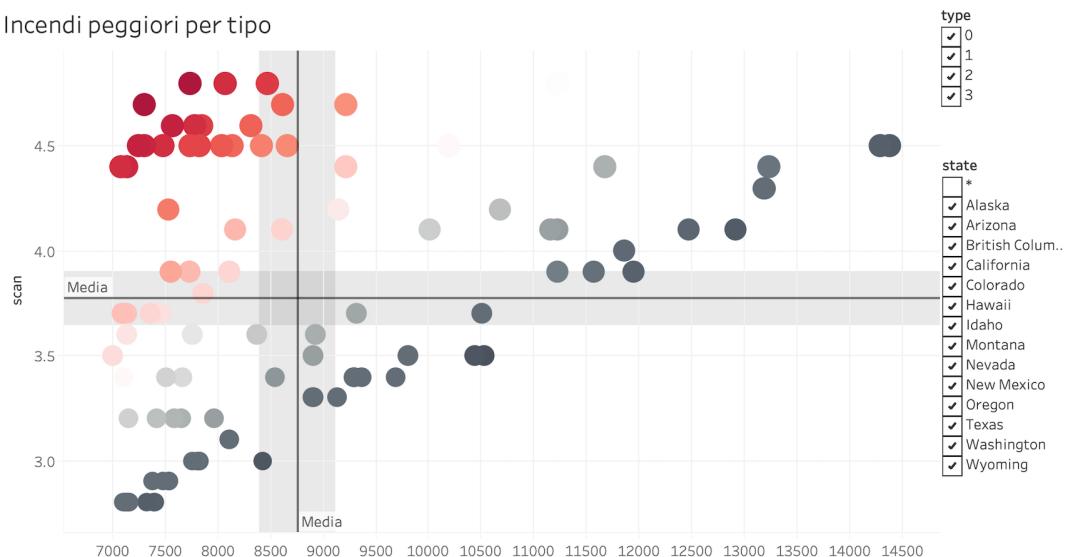
In basso a sinistra vi è una distribuzione in quartili che mostra come si dispongono i vari tipi di incendio in termini di brightness, frp e estensione medie. Si può osservare come gli incendi che presentano maggiori frp, cioè potenze irradiate e brightness, cioè brillantezza in termini di calore irradiato, risultano quelli di tipo 3. Ciò si può spiegare notando che spesso questi consistono in esplosioni a largo. Essi sono seguiti dagli incendi di tipo vulcanico molte volte dovuti a eruzioni che sprigionano quindi grandi potenze nell'ambiente circostante. Per quanto riguarda la scan, indice della dimensione degli incendi, il tipo che prevale è quello vulcanico, in quanto spesso le eruzioni o altri fenomeni di questa origine tendono a espandersi su vaste aree geografiche.

Caratteristiche dei tipi di incendi



3.5.4 Incendi peggiori per tipo

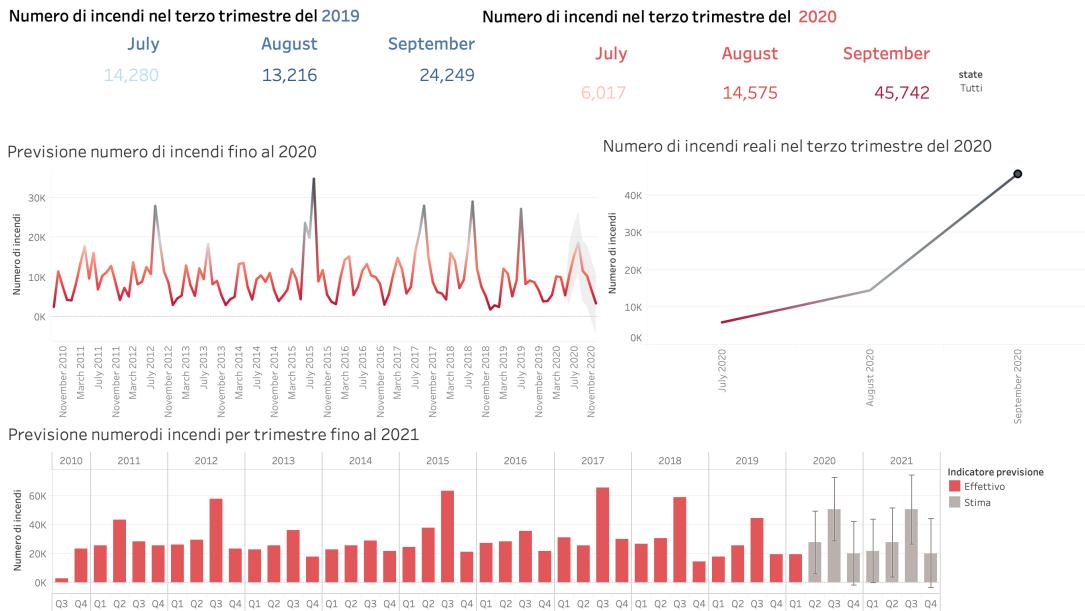
Infine, in basso a destra è stato realizzato un grafico a dispersione in cui sono riportati gli incendi peggiori avvenuti per tipo. Il concetto di peggiore è definito considerando la frp, l'estensione e la brightness (con bolle aventi grandezza in relazione all'estensione e colore in base alla brightness). È presente inoltre un filtro per Stato e uno per tipo per rendere più specifica la visualizzazione e vi sono due linee incrociate che rappresentano l'estensione e la frp medie degli incendi. Considerando gli incendi di tipo 0 si può osservare che i due incendi peggiori per frp sono avvenuti a Washington il 20 Agosto 2018 con una potenza irradiata di più di 14000 MW, questi in realtà coincidono con un unico grande incendio avvenuto quel giorno che però il satellite ha considerato come pixel separati. Considerando invece l'estensione, i peggiori incendi di tipo 0 sono avvenuti in Idaho e in California arrivando tutti a quasi 5 km^2 . Infine, gli incendi di tipo vulcanico peggiori sia per estensione che per frp e brightness sono avvenuti nelle Hawaii, come ci si poteva aspettare.



3.6 Previsione numero di incendi fino al 2020

Con questa dashboard si inizia un'analisi predittiva dei dati a disposizione. Come già introdotto, i dati usati nelle analisi precedenti arrivano fino a Maggio 2020, tuttavia è disponibile un ulteriore dataset separato che contiene i valori dei mesi successivi a

CAPITOLO 3: TABLEAU



partire da Giugno 2020. Perciò si è voluto provare a predire, utilizzando gli strumenti di predizione di Tableau, l’andamento degli incendi nei mesi mancanti del 2020, controllando successivamente l’errore commesso rispetto all’andamento reale degli incendi che si è avuto.

3.6.1 KPI

In alto abbiamo dei KPI che mostrano il numero di incendi avuti nei mesi del terzo trimestre del 2019 e del 2020 per fare un confronto. Si nota che, pur essendoci state differenze tra Luglio e Agosto dei due anni è Settembre che si registra un anormale crescita di incendi nel 2020 rispetto al 2019, passando a 45742 da 24249 incendi! Questo grande aumento purtroppo è stato causato soprattutto dai lunghi e vasti incendi avvenuti in California a fine estate.

Numero di incendi nel terzo trimestre del 2019

July	August	September
14,280	13,216	24,249

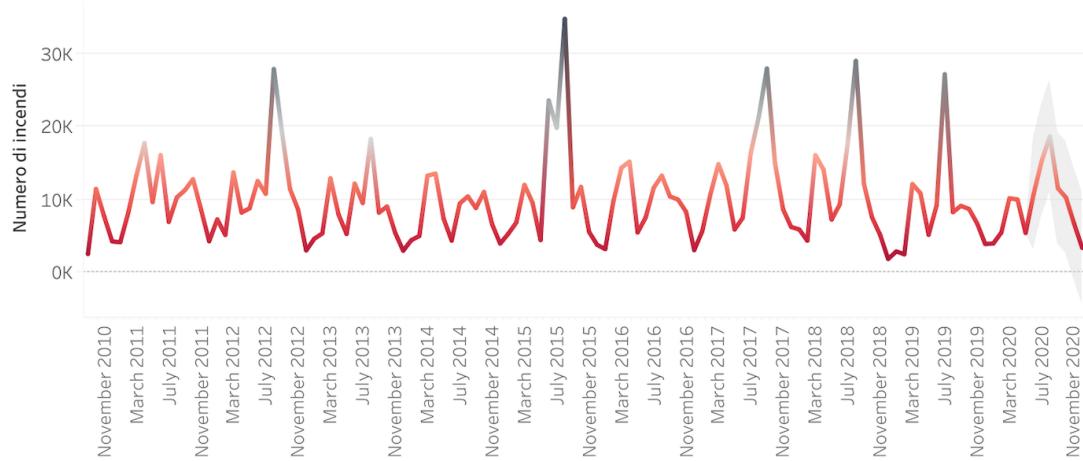
Numero di incendi nel terzo trimestre del 2020

July	August	September
6,017	14,575	45,742

3.6.2 Previsione numero di incendi fino al 2020

Facendo una predizione dell'andamento del numero di incendi nei mesi del 2020 mancati nel dataset iniziale quello che si ottiene è mostrato nel grafico a linee al centro a sinistra, dove la previsione si può riconoscere dall'area grigia che indica l'intervallo di confidenza.

Previsione numero di incendi fino al 2020

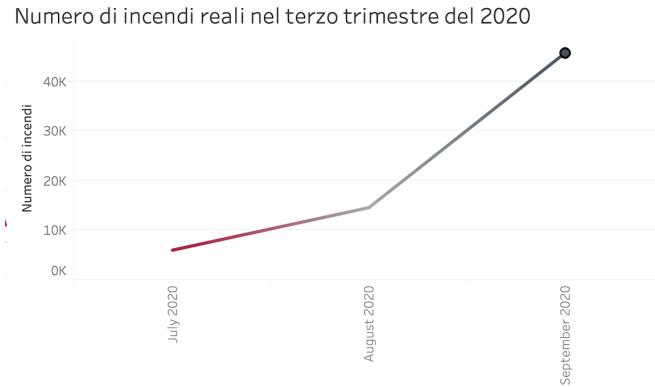


3.6.3 Numero di incendi reali nel terzo trimestre del 2020

Invece a destra sono mostrati, sempre con un grafico a linee gli andamenti reali che si sono registrati in Giugno, Luglio e Settembre 2020. La previsione non è molto buona, soprattutto quella del mese di settembre, questo perché, osservando gli anni precedenti si è quasi sempre avuto un calo netto degli incendi a Settembre rispetto ad Agosto, quindi quello del 2020 è stato un caso eccezionale imprevedibile con questo dataset, non essendoci stato alcun caso simile nel training set considerato, cioè negli anni precedenti al 2020 (presenti nel dataset utilizzato). Infatti il numero di incendi previsto per settembre risulta 11593, in calo rispetto alla previsione di agosto di 18702, mentre i dati

CAPITOLO 3: TABLEAU

dicono che da agosto a settembre 2020 si è passati da 14575 a 45742 incendi. È inoltre possibile scegliere il paese per cui fare la predizione utilizzando un filtro. Selezionando la California si possono confermare ulteriormente le considerazioni fatte. Considerando altri paesi come l'Illinois si può vedere che la qualità della previsione è migliore.



3.6.4 Previsione numero di incendi per trimestre fino al 2021

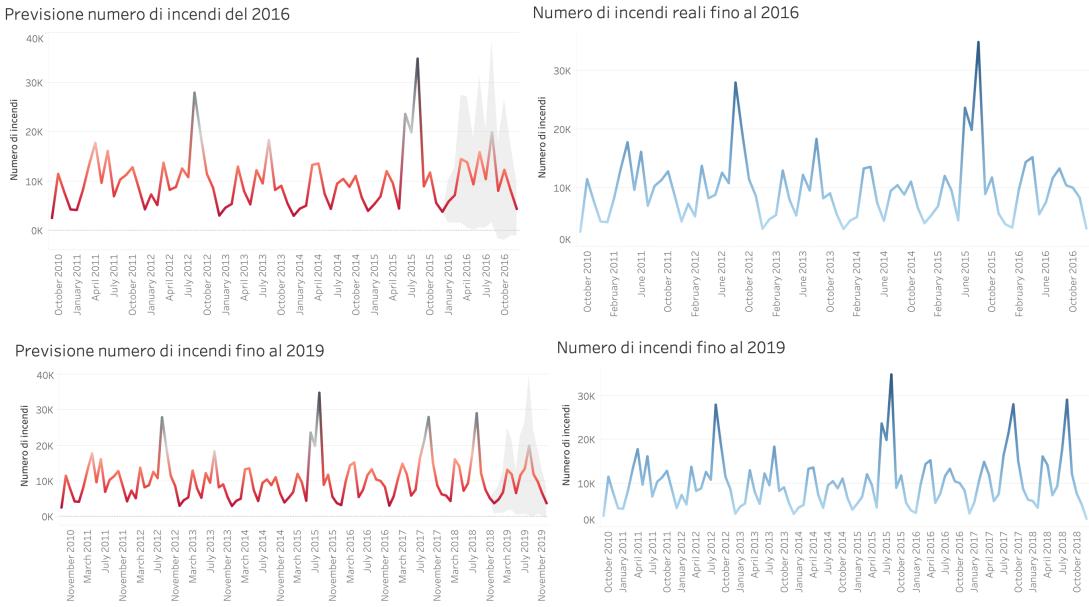
In basso è fatta un'altra previsione del numero di incendi però non più per mesi ma per trimestri, e si prevedono non solo i trimestri della parte mancante del 2020 ma anche del 2021. La visualizzazione non è più tramite un grafico a linee ma è usato un istogramma in cui il range di confidenza è rappresentato con dei segmenti e la previsione è sempre in grigio.



3.7 Previsioni incendi nel 2016 e nel 2019

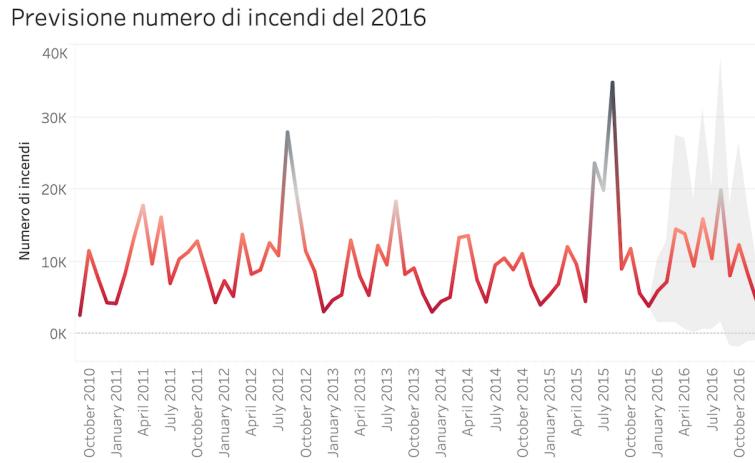
Nella seguente dashboard si è voluto andare a fare previsione andando a campionare il dataset nelle annate del 2016 e del 2019.

CAPITOLO 3: TABLEAU



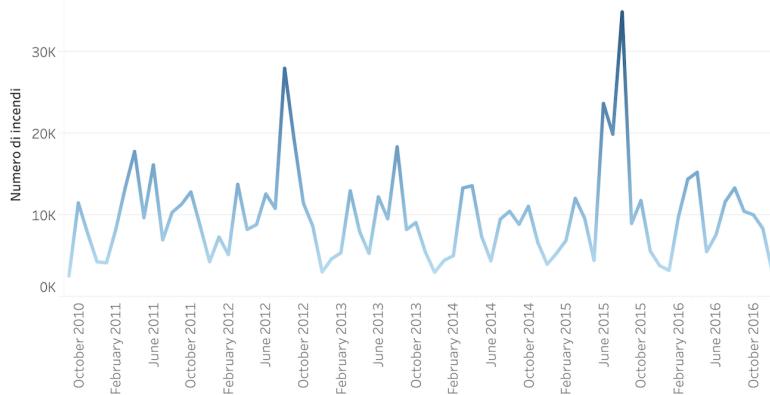
3.7.1 Confronto tra previsione del numero di incendi nel 2016 e dati reali

Si è sottolineato come il 2020 sia stato un anno difficile da predire a livello di incendi, quindi si è voluto utilizzare un troncamento del dataset ad un certo anno per provare a predire l'andamento del numero di incendi nei mesi degli anni successivi. Ad esempio si è troncato il dataset al 2015 per predire l'andamento nel 2016, confrontando poi la previsione con l'andamento reale visualizzato a destra.



Si osserva che la predizione per quanto riguarda il 2016 non è ottimale soprattutto nei mesi estivi che risultano spesso quelli più difficili da predire.

Numero di incendi reali fino al 2016



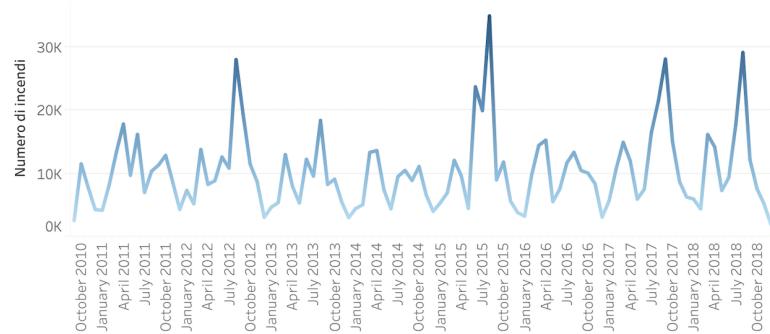
3.7.2 Confronto tra previsione del numero di incendi nel 2019 e dati reali

Una cosa analoga è stata fatta troncando il dataset al 2018, facendo poi una predizione del 2019 e confrontando con l'andamento reale registrato.

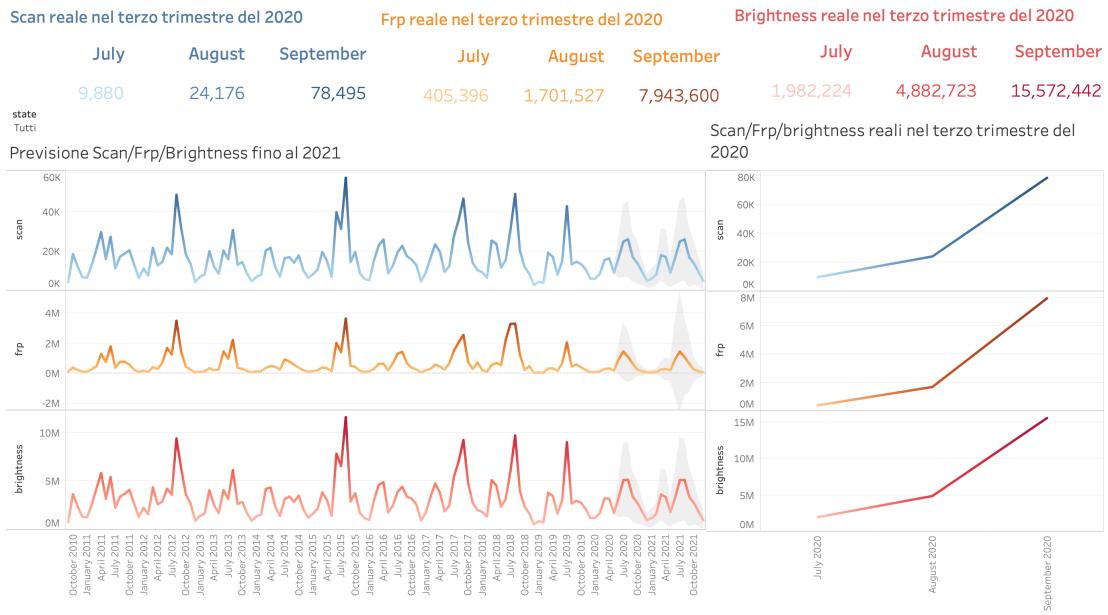
Previsione numero di incendi fino al 2019



Numero di incendi fino al 2019



CAPITOLO 3: TABLEAU



3.8 Previsione scan, frp e brightness fino al 2020

3.8.1 KPI previsioni

In questa dashboard la previsione non è più basata sul numero di incendi ma si estende a scan (estensione), brightness e frp (potenza irradiata) ed è sempre incentrata sul terzo trimestre del 2020 (giugno, luglio e agosto). Queste KPI mostrano le somme di scan, frp e brightness reali nei vari mesi del trimestre considerato. Anche qui si nota che i valori di settembre crescono molto rispetto a quelli di agosto.

Scan reale nel terzo trimestre del 2020

July	August	September
9,880	24,176	78,495

Frp reale nel terzo trimestre del 2020

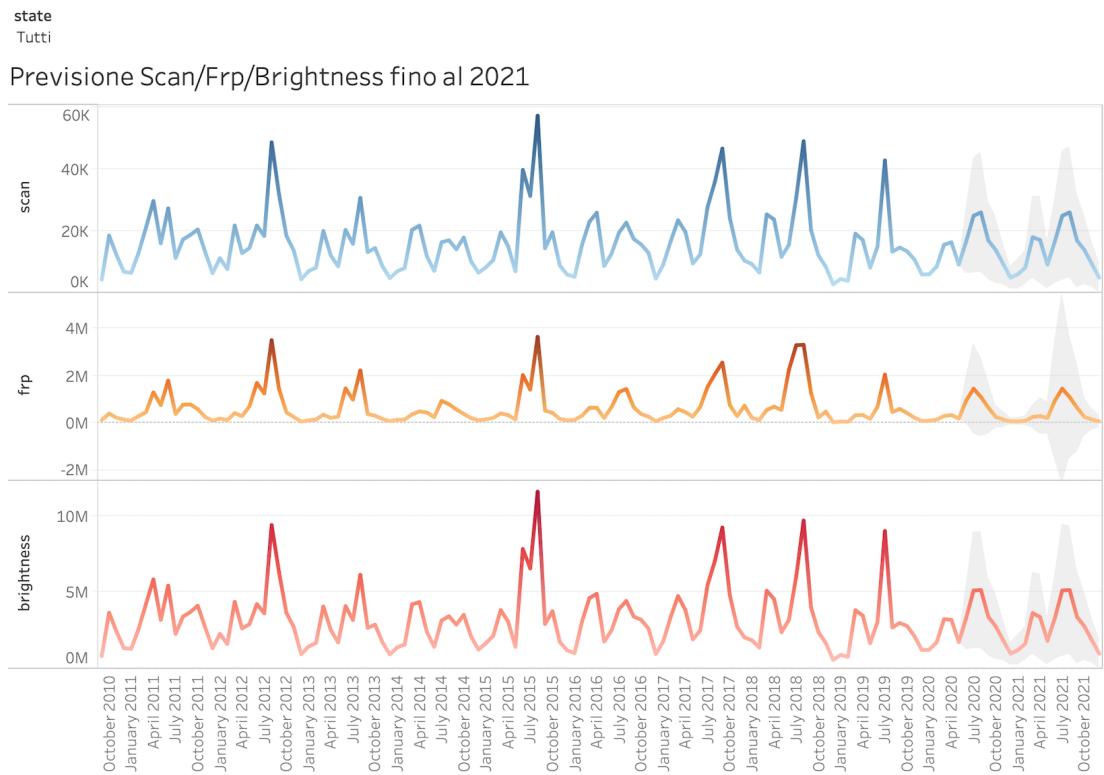
July	August	September
405,396	1,701,527	7,943,600

Brightness reale nel terzo trimestre del 2020

July	August	September
1,982,224	4,882,723	15,572,442

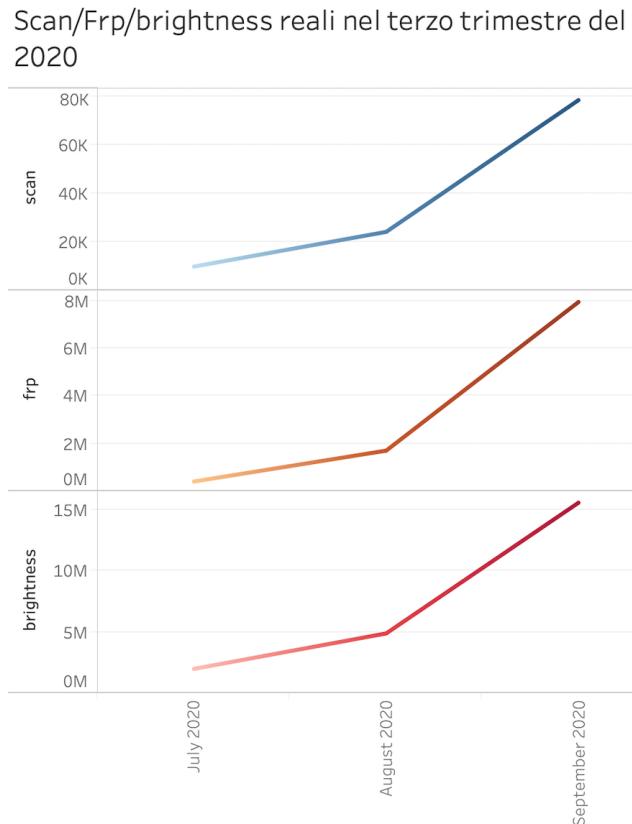
3.8.2 Previsione scan, frp e brightness fino al 2021

Questo grafico a linee riporta le previsioni di scan, frp e brightness fatte per i vari mesi del 2020 e del 2021. Le aree in grigio rappresentano sempre gli intervalli di confidenza delle previsioni. La dashboard presenta anche in tal caso un filtro per Paese che agisce su questo e sul prossimo grafico per ottenere informazioni più specifiche riguardanti le previsioni di un particolare Stato.



3.8.3 Scan, frp e brightness nel terzo trimestre del 2020

Questi grafici a linee invece riportano gli effettivi valori di scan, frp e brightness nel trimestre considerato e come per il numero di incendi si rileva un grande discostamento tra previsione e realtà a settembre in cui vi è un picco anormale mai verificatosi negli anni precedenti.



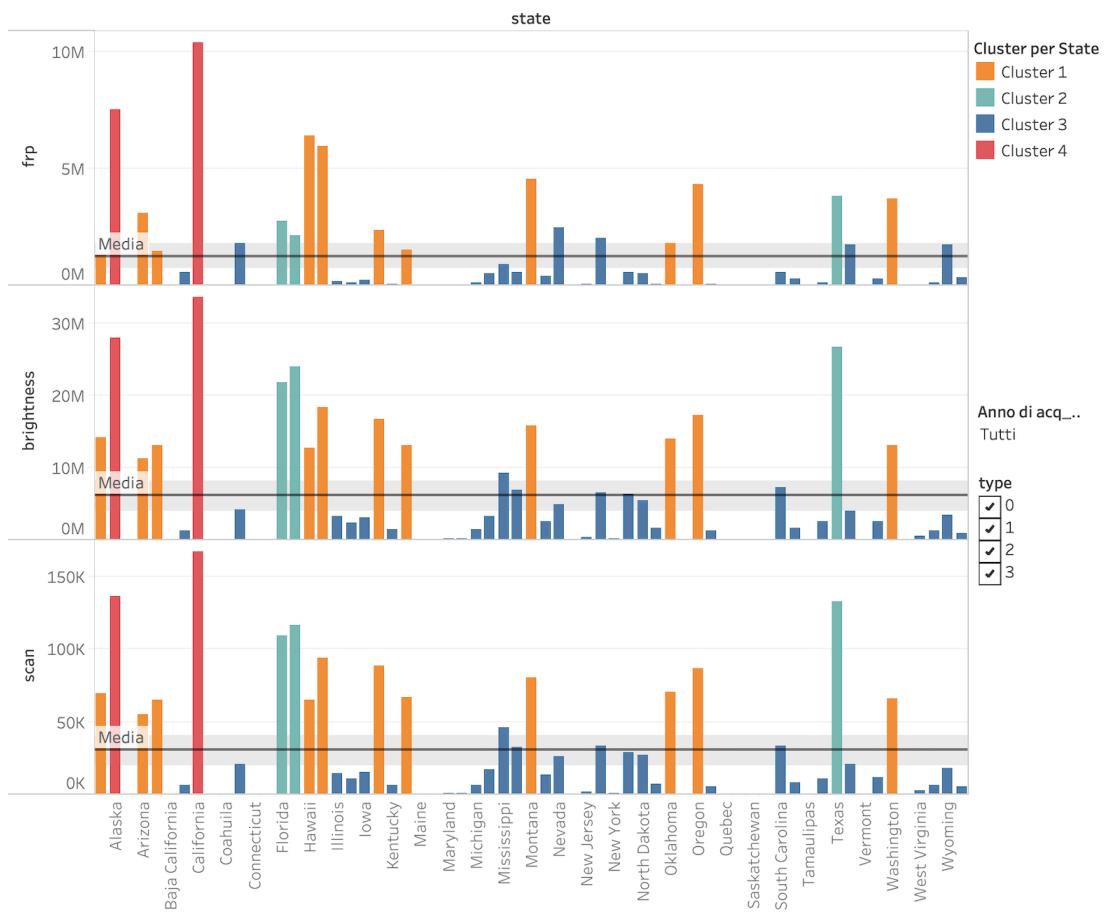
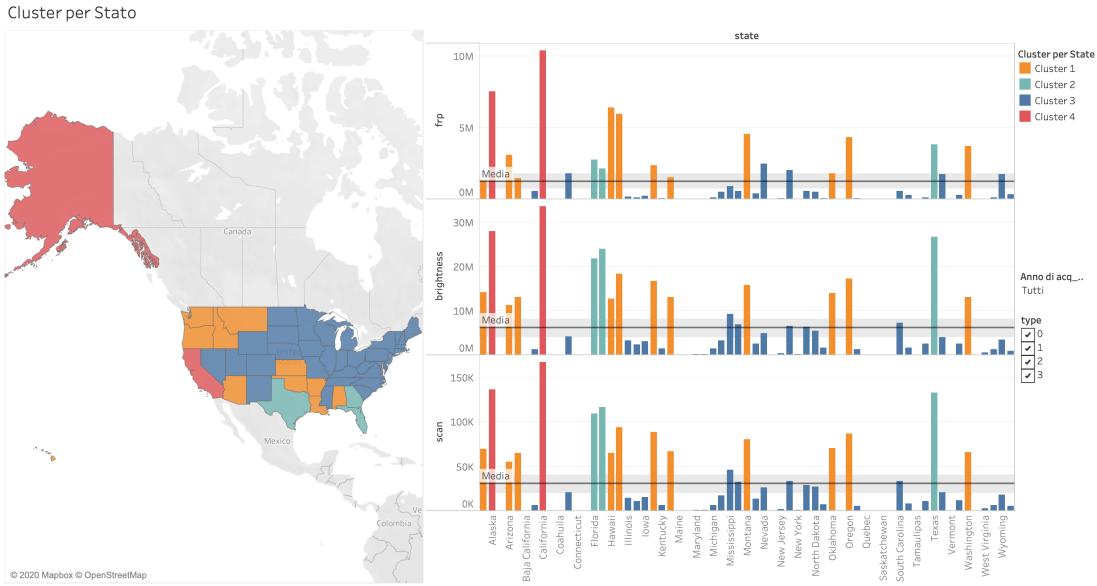
3.9 Cluster degli Stati

Questa dashboard si pone l'obiettivo di ottenere dei cluster di Stati in base alle tre grandezze frp, brightness e scan, di cui si considera la somma.

3.9.1 Cluster per Stato, istogrammi

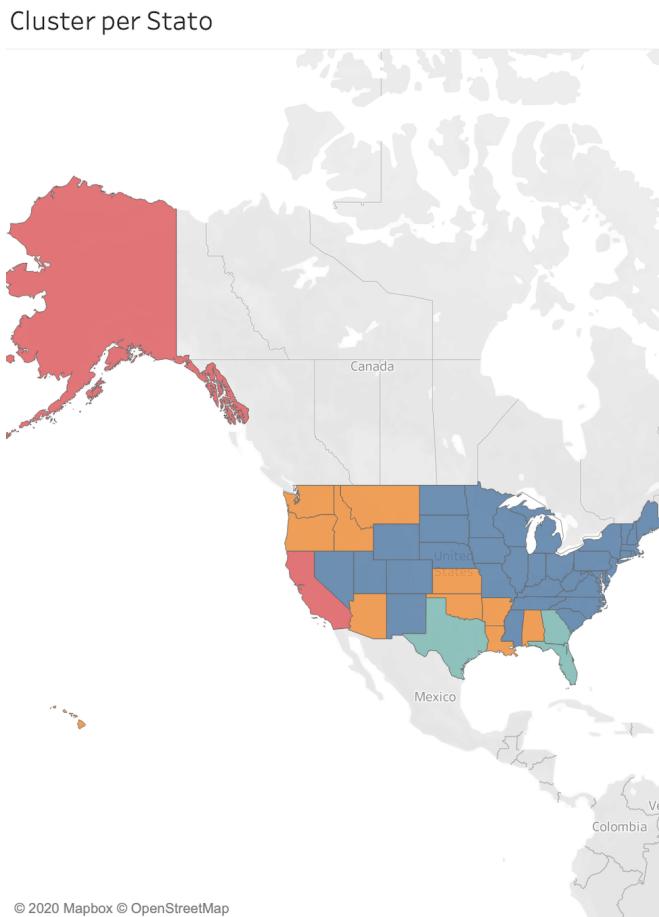
A destra, utilizzando tre histogrammi sono rappresentati i cluster ottenuti. Essi sono quattro e ad ognuno di esso è assegnato un colore specifico. Il cluster 4 è quello dei paesi che hanno avuto incendi più intensi, cioè con valori di scan, frp e brightness maggiori, il cluster 3 è invece quello con gli stati colpiti da incendi meno forti e infine i cluster 1 e 2 sono costituiti da Paesi che hanno registrato incendi di medio-alta intensità. Infine, con una retta orizzontale è rappresentata la media di ogni valore. Con un'area grigia attorno alla media è rappresentato l'intervallo di confidenza medio.

CAPITOLO 3: TABLEAU



3.9.2 Cluster per Stato, mappa

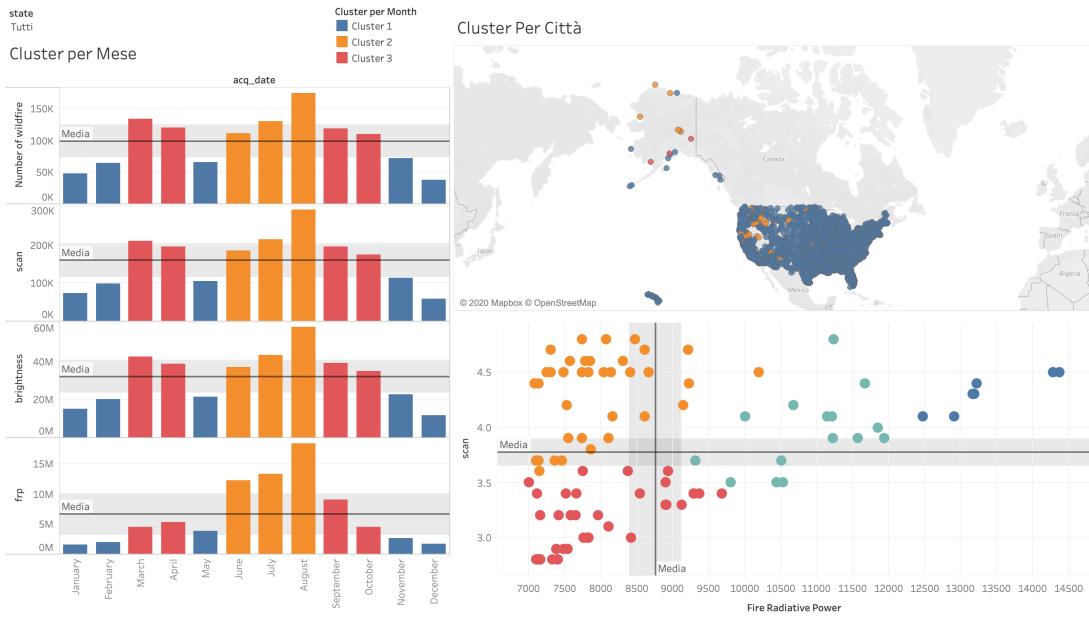
Nella mappa si è voluto rappresentare i vari Stati con gli stessi colori dell'istogramma per indicare il cluster di appartenenza. Come si nota i paesi appartenenti al cluster 4, cioè quelli più colpiti, fortunatamente sono solamente due, la California e l'Alaska, invece la maggior parte dei Paesi degli USA appartengono al cluster 3 (sono Paesi soprattutto della costa atlantica o della zona continentale), cioè hanno registrato incendi di minore entità. Invece i Paesi appartenenti ai cluster intermedi 1 e 2, appartengono soprattutto alle coste meridionali o a quelle nord-occidentali.



3.10 Cluster per mese e città

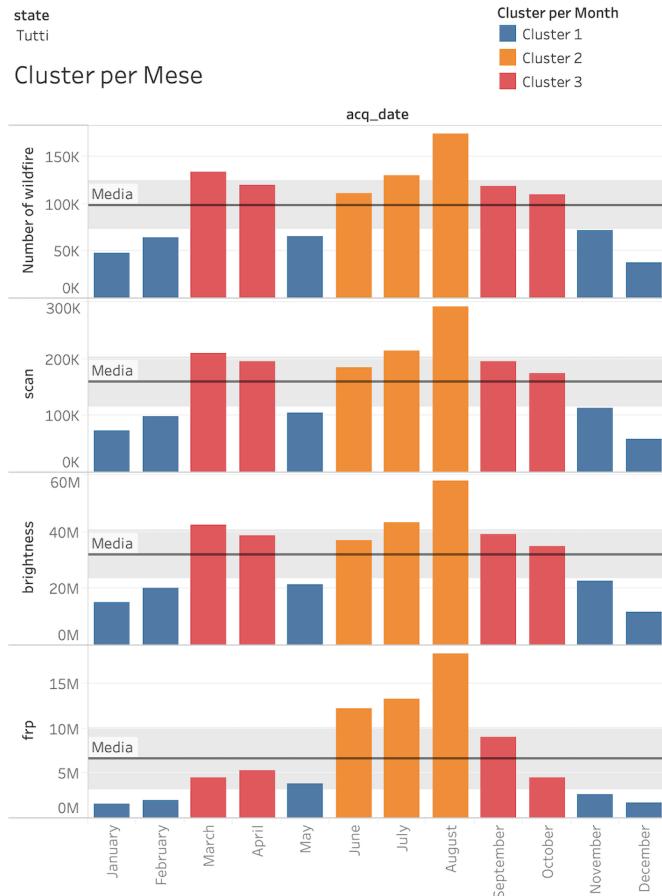
In questa dashboard sono stati realizzati ulteriori cluster.

CAPITOLO 3: TABLEAU



3.10.1 Cluster per mese

I cluster per mesi sono stati realizzati considerando le somme di frp, brightness, scan e numero di incendi. Sono visualizzati con un istogramma simile a quello ottenuto nel

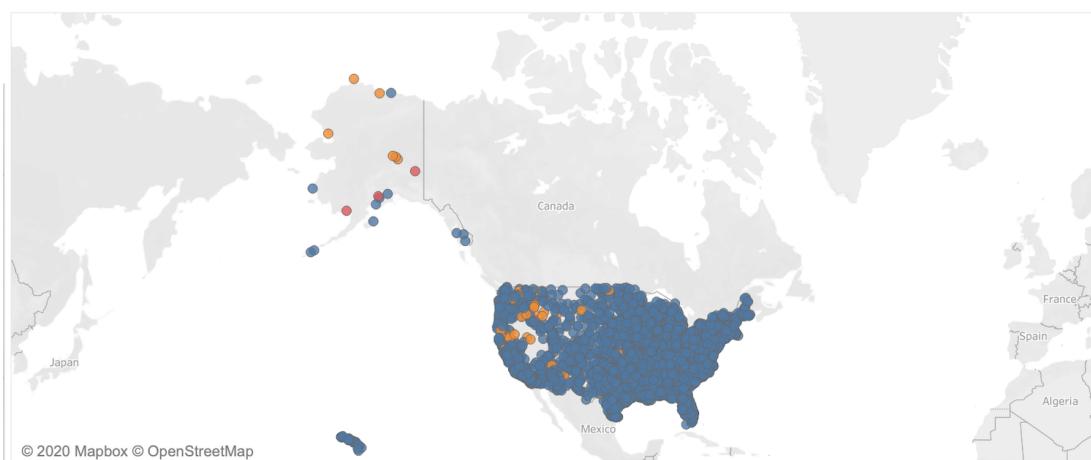


caso della clusterizzazione per Stato, tuttavia in tal caso i cluster risultanti sono tre. Il cluster 2, associato al colore arancione è costituito dai mesi più caldi dell'anno, giugno, luglio e agosto, in cui si hanno più incendi e sono anche più intensi e vasti, il cluster 1 è costituito invece dai mesi di novembre, dicembre, gennaio e febbraio, che sono anche mesi freddi in cui quindi si tende ad avere meno incendi e essi sono anche di minore entità. Infine, vi è il cluster 3, che è quello intermedio a cui appartengono ottobre, settembre, marzo e aprile. Infine, con una retta orizzontale è rappresentata la media di ciascuna grandezza e si nota che i mesi del cluster 2 sono quelli sopra i valori medi, quelli del cluster 1 si trovano al disotto ed infine quelli del cluster 3 si trovano attorno al valore medio. Con un'area grigia attorno alla media è rappresentato l'intervallo di confidenza medio.

3.10.2 Cluster per città

I cluster per città sono stati realizzati considerando le somme di frp, brightness e scan e vengono visualizzati con una mappa degli USA in cui ogni città è rappresentata da un punto con il colore del proprio cluster di appartenenza. Sono stati trovati 3 cluster di cui l'1, associato al colore blu, è quello predominante mentre il 2 è costituito da molte meno città e il 3 ancora di meno.

Cluster Per Città

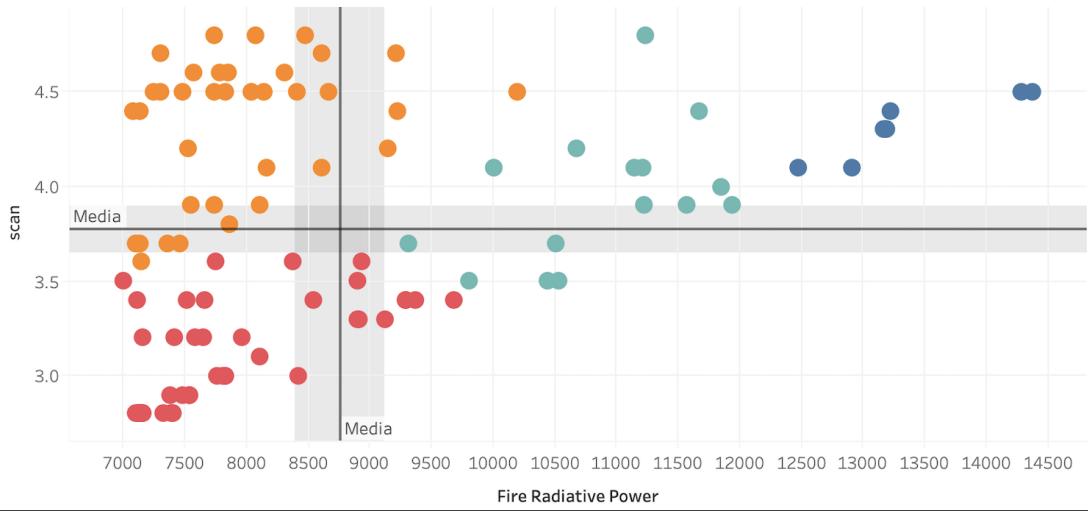


3.10.3 Cluster sulla gravità degli incendi

Infine l'ultimo cluster è fatto per frp e scan basandosi sulle somme di scan, frp e brightness ed è visualizzato tramite un grafo a dispersione in cui ogni bolla rappresenta un

CAPITOLO 3: TABLEAU

incendio e ha il colore del cluster di appartenenza. Sono stati ricavati in tal modo 4 cluster differenti. Con due rette sono state rappresentate le medie di scan e frp con delle aree grige circostanti che rappresentano gli intervalli di confidenza.



Tali rette individuano nel grafo quattro quadranti e si può osservare che vi è un cluster, quello rosso che ricopre gran parte del quadrante in basso a sinistra (sono gli incendi caratterizzati da scan e frp non altissime), un altro, cioè quello arancione, che copre gran parte del quadrante in alto a sinistra (sono gli incendi caratterizzati da scan alta e frp bassa). Invece il cluster verde è caratterizzato da valori di scan e di frp medio-alti e quello blu da valori di frp molto alti e di scan medio-alti.

4. Power BI

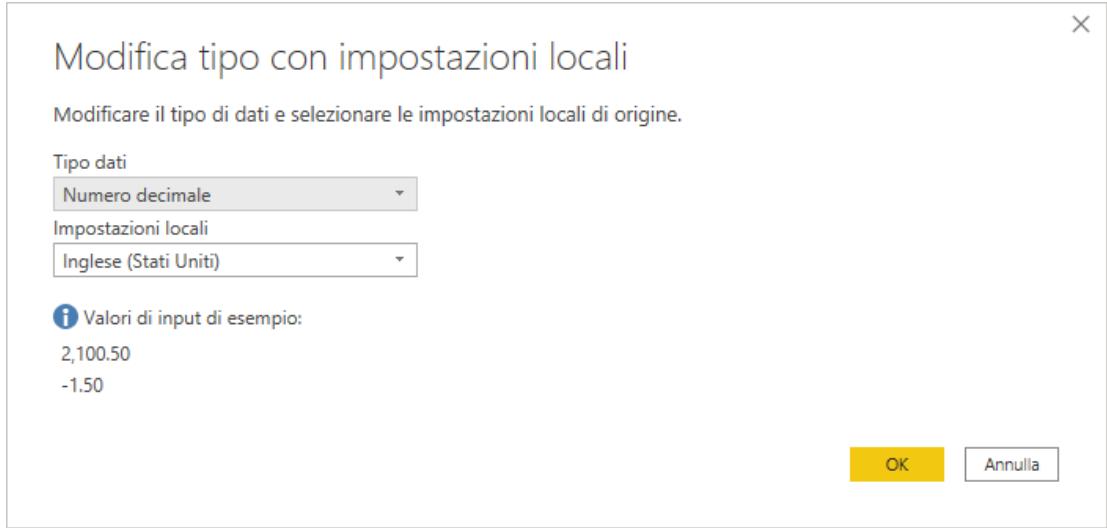
4.1 Introduzione al tool

Power BI è il tool di business intelligence svilupato da Microsoft, che offre ulteriori funzionalità rispetto a quelle di Qlik e Tableau, ed infatti come riportato sul *Magic Quadrant Gartner* è lui il vero leader del mercato. Tra le funzioni più importanti che integra troviamo sicuramente gli scripting *R* e *Python*, entrambi linguaggi di programmazione specifici per l'analisi dei dati. L'utente può utilizzare oggetti visivi basati su questi due elementi senza essere a conoscenza del relativo linguaggio di programmazione, il che rende possibile realizzare analisi avanzate anche a chi non ha competenze informatiche su tali campi. Inoltre l'integrazione tra *PowerBI* e *R* ci consente di fondere l'efficace visualizzazione dei dati del primo con l'implementazione degli algoritmi di Machine Learning del secondo portando l'analisi ad uno step superiore.

4.2 Preparazione del dataset

Prima di iniziare l'analisi è stato necessario fare una fase di ETL come descritto nel Par.1.3.2, e parallelamente si sono svolte altre operazioni di ETL molto simili a quelle già descritte per Qlik e Tableau, nelle quali si sono correttamente definiti il tipo di dato per ogni colonna e impostato la prima riga come intestazione della colonna stessa.

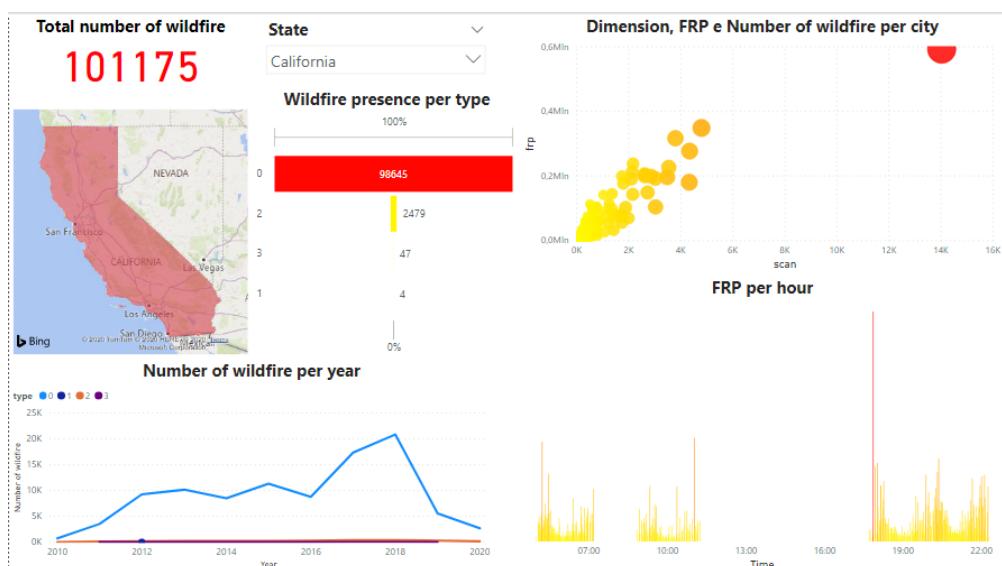
Nel momento del caricamento dei dati nell'Editor di Power Query, tutte le colonne con valori di tipo *Numero decimale* sono state lette come colonne di tipo *Stringa*. Per il casting nel tipo corretto, è stato necessario scegliere sulla voce *Impostazioni locali* il valore *Inglese(Stati Uniti)*, poichè i valori contenuti nel DataFrame hanno come indicatore della parte decimale il *"."* invece che la *","* come da impostazioni standard.



4.3 Dashboard di tipo descrittivo e diagnostico

La parte di analisi descrittiva e diagnostica è stata già trattata in maniera molto approfondita in Qlik e Tableau, quindi anche le pagine create in Power BI saranno molto simili a quelle già viste in precedenza. Però per evitare di analizzare sempre le stesse cose e ottenere sempre le stesse considerazioni, si è deciso di eseguire un'analisi completa per stato e per città che comprenda i principali attributi visti fin qui, ed è stata aggiunta anche un'analisi per ora e per fascia oraria, cosa che negli altri tools non era stata fatta per mancanza di valori di tipo *Ora*.

4.3.1 Panoramica per stato



CAPITOLO 4: POWER BI

Nella prima Pagina creata si è fatta un'analisi per Stato, ovvero è stato inserito un filtro che permette la selezione di uno o più stati (nella figura è selezionato lo stato della California) nella quale andare a fare l'analisi degli incendi negli ultimi dieci anni.



L'analisi avviene attraverso diverse visualizzazioni di grafici. Il primo da osservare è sicuramente l'*indicatore KPI*, nel quale si legge il valore degli incendi totali avvenuti nello stato selezionato.

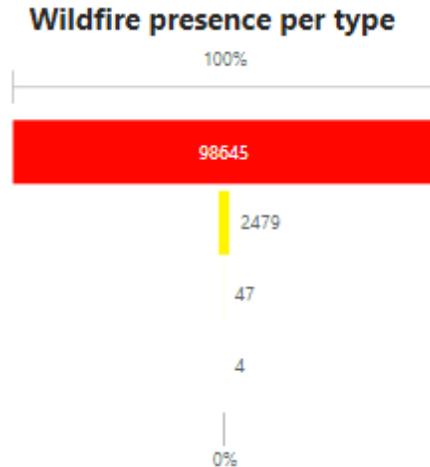
Total number of wildfire
101175

Il secondo da osservare è la mappa che Power BI mette disposizione (*Mappa colorata*), nella quale viene evidenziata la zona, che può essere uno stato o una città, che gli viene passata semplicemente inserendo il suo nome nella sezione Località, evitando quindi di utilizzare coordinate di latitudine e longitudine come fatto in Qlik.

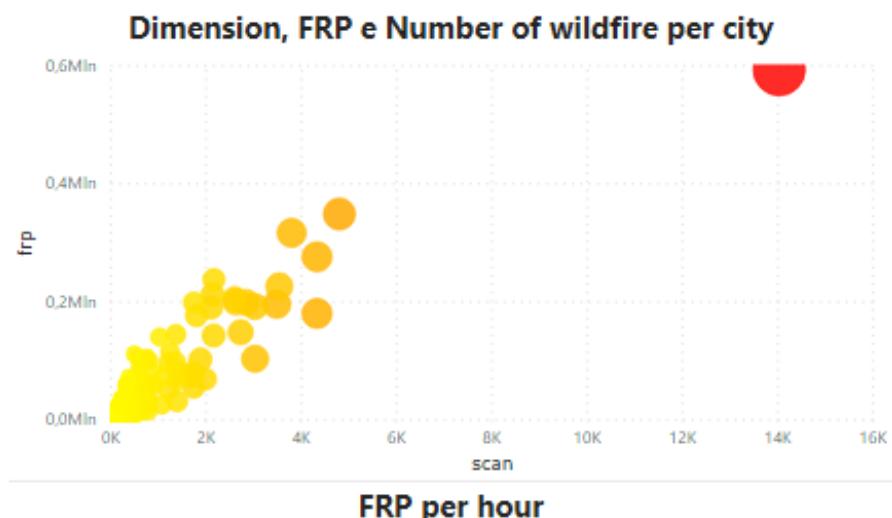


Osservati i primi indicatori del contesto che si sta visualizzando, si può passare alla let-

tura dei grafici più importanti. In particolare si può subito notare il grafico ad *Imbuto* che descrive il numero di incendi avvenuto nei dieci anni, diviso per tipo. Si può notare che nel caso della California la quasi totalità degli incendi sia di tipo 0, ovvero di tipo vegetativo, ma comunque ci sono stati incendi anche per tutti gli altri tipi.

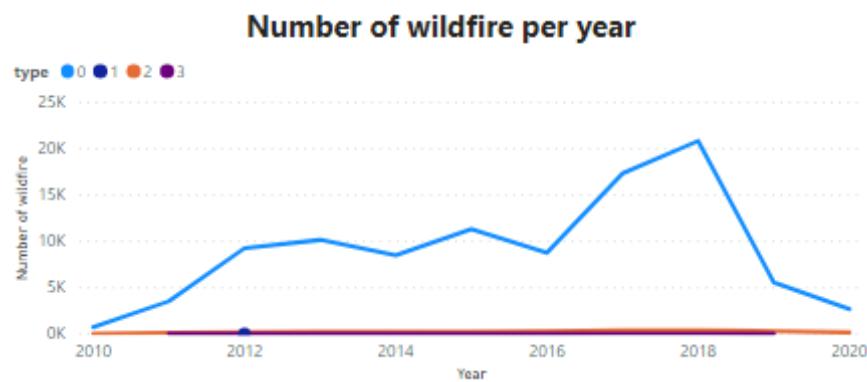


Si passa poi al *Grafico a dispersione*, dove ogni punto del grafo rappresenta una città all'interno del stato selezionato. Poichè negli assi *x* e *y* si hanno rispettivamente *scan* (quindi dimensione dell'incendio) e *frp* (quindi potenza irradiata), più il punto si troverà verso l'angolo superiore destro, maggiore sarà la gravità degli incendi totali avvenuti in quella specifica città. Le dimensioni e il colore del punto dipendono invece dal numero degli incendi. Scorrendo con il puntatore sopra ciascun punto, viene visualizzata un'etichetta informativa nella quale vi sono tutte le informazioni descritte nel grafico.

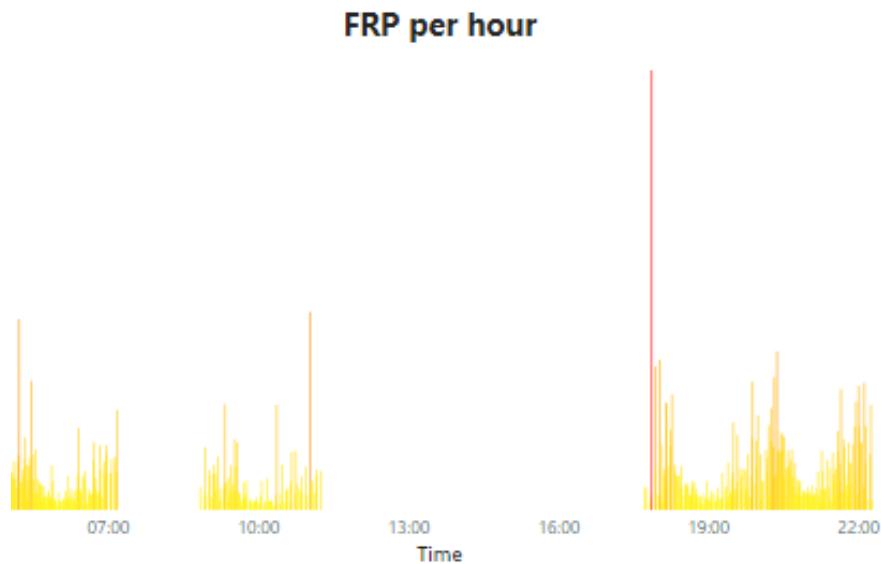


Nel caso della California si osserva che la città che ha avuto più incendi, con un'esten-

sione e una potenza irradiata totale maggiore è Happy Camp, villaggio per campeggi turistici immerso nella natura. Successivamente si analizzano gli andamenti degli incendi distribuiti nei 10 anni, divisi per tipo. Si nota che il 2010 e il 2020 sono gli anni che hanno un numero di incendi minore, e ciò è dovuto, come già ribadito, al fatto che non si hanno dati disponibili per l'intero anno. Inoltre si può evidenziare anche il fatto che il 2018 sia stato il peggiore per quanto riguarda gli incendi di tipo 0. Per gli altri tipi, l'andamento è stato circa lineare negli anni, con numeri sempre molto minori rispetto al tipo vegetativo.

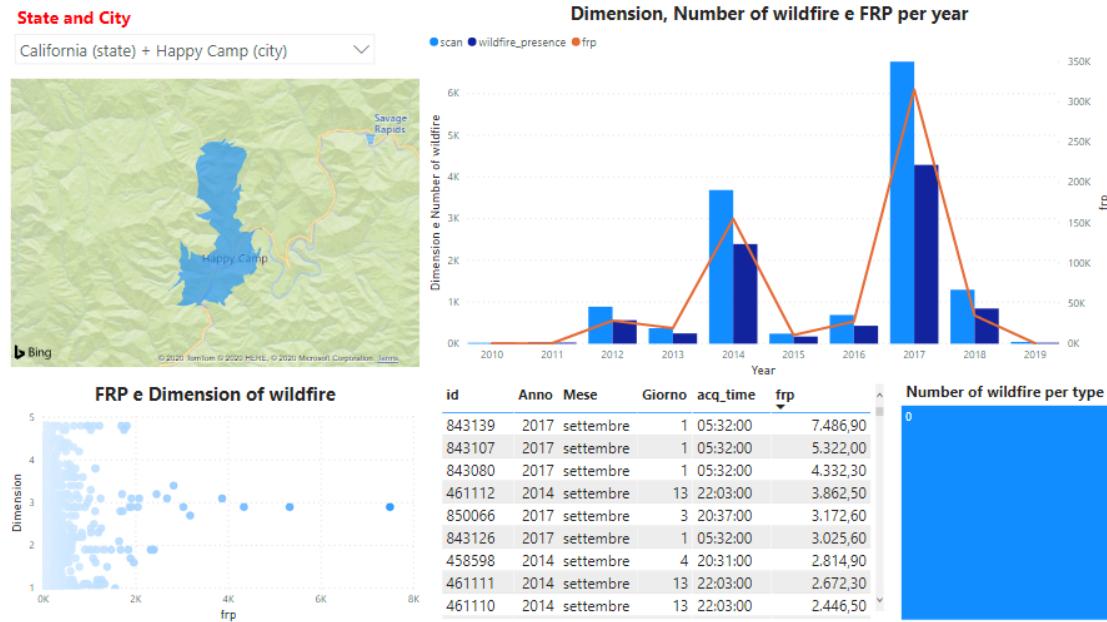


Infine in questa pagina è stata analizzata la potenza irradiata media dagli incendi, distribuita nell'arco della giornata, evidenziando nel caso della California che l'ora di picco è intorno alle 18:00.

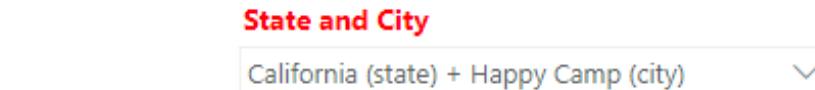


4.3.2 Analisi per città

Nella seconda pagina creata si fa un'analisi dettagliata per città, analizzando gli andamenti negli anni e quali e quando sono stati gli incendi peggiori.



Anche in questo caso per partire nell'analisi bisogna prima di tutto selezionare la città all'interno del filtro, che sono divise per stato per facilitare la ricerca. In questo caso si decide di scegliere Happy Camp, ovvero la città che nella pagina precedente si era vista come quella con il più grande numero di incendi all'interno della California.



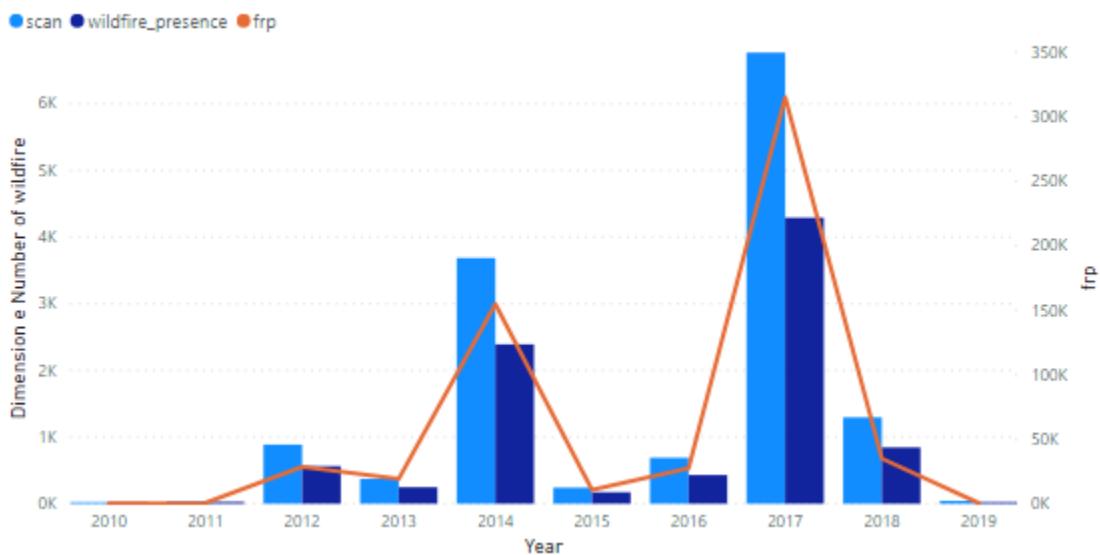
Selezionata la città, la si può osservare all'interno della mappa messa a disposizione da Power BI.

Il primo grafico in cui viene fatta una vera analisi è il *Grafico a linee e istogramma a colonne raggruppate*, nel quale vengono visualizzate tutte insieme l'andamento delle dimensioni, del numero (i primi due tramite istogramma) e della potenza irradiata (tramite linea) degli incendi nell'arco dei 10 anni nella città in analisi. Si nota immediatamente che gli anni peggiori sono stati il 2014 e il 2017, mentre in tutti gli altri anni i numeri sono molto più contenuti.

Successivamente con un *Grafo a dispersione* si visualizzano i "peggiori" incendi avve-



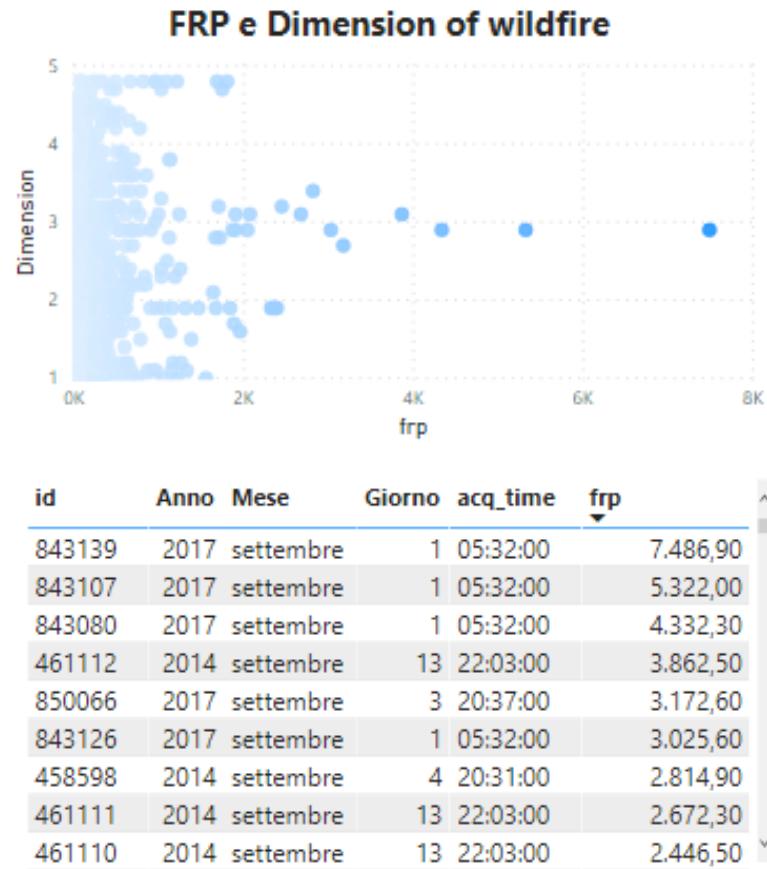
Dimension, Number of wildfire e FRP per year



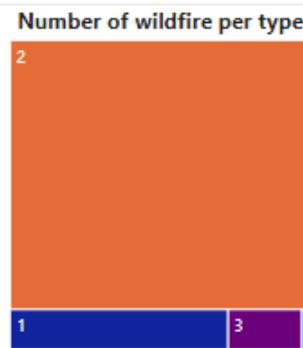
nuti nella città, dove per definire l’aggettivo peggiore si utilizzano i parametri di *frp* e *scan*, che appunto definiscono la potenza irradiata e la dimensione dell’incendio. Scorrendo con il puntatore sopra i vari punti del grafico comparirà una nota informativa con i valori di *scan* e *frp* dell’incendio corrente, ed in più il numero dell’*id* che identifica l’incendio stesso.

Se si vogliono avere altre informazioni sull’incendio, basta spostare lo sguardo sulla tabella di fianco, nella quale si può trovare la corrispondenza con l’incendio che si sta analizzando attraverso il suo valore dell’*id* (inoltre sono ordinati per *frp* decrescente). Le informazioni aggiuntive all’interno della tabella riguardano la data esatta dell’incendio, specificando anno, mese, giorno e ora dell’acquisizione del dato da parte del satellite.

Infine, come ultimo grafico è stata inserita una *Mappa ad albero*, dove si visualizzano il numero di incendi totali nell’arco dei 10 anni divisi per tipo. Nel caso di Happy Camp

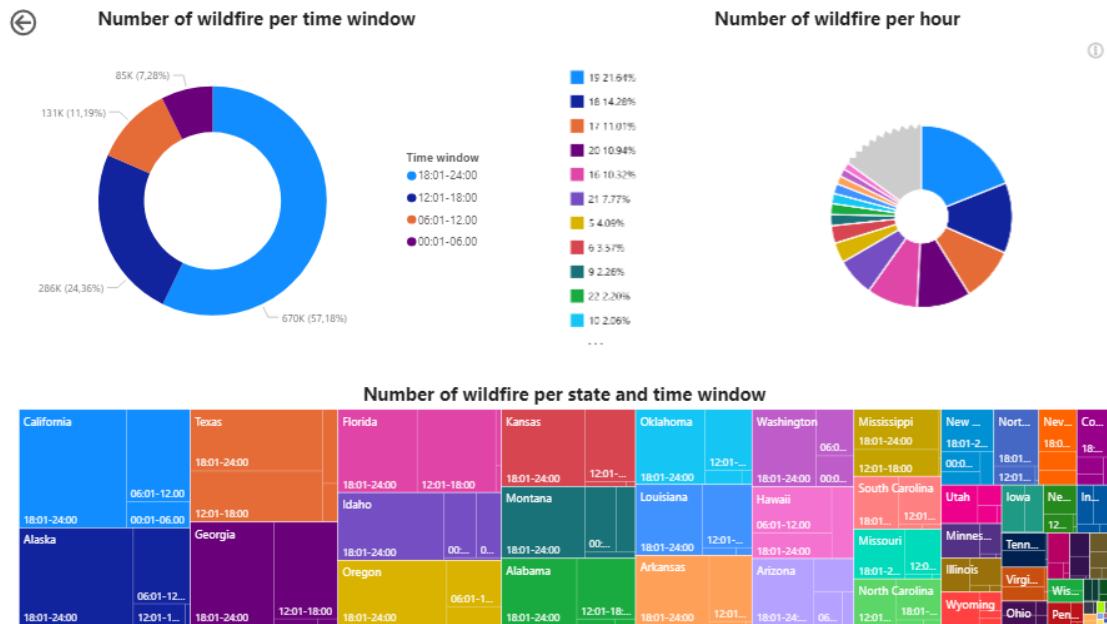


questo grafo non è molto significativo, poichè la totalità degli incendi è di tipo vegetativo. Perciò per far concepire a pieno l'utilità del grafo, qui di seguito viene visualizzato quello che si ottiene ad esempio facendo l'analisi su Nanawale Estates, situata nelle Hawaii.

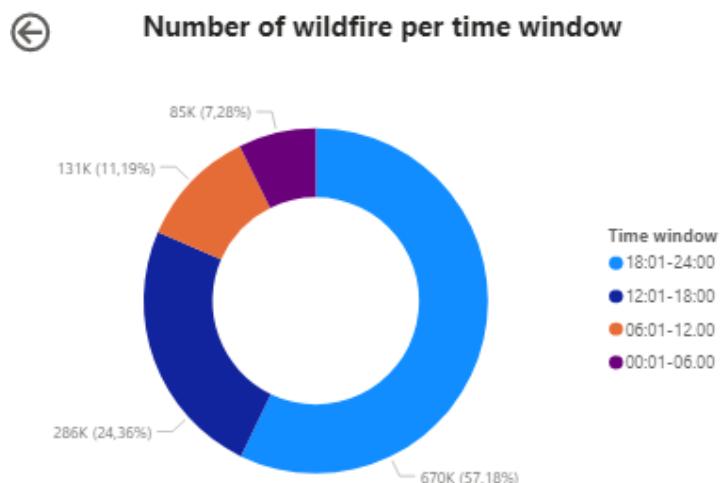


4.3.3 Analisi per ora

In questa ultima pagina descrittiva si è analizzata la distribuzione degli incendi all'interno della giornata, considerando sia gli incendi totali degli USA, sia il numero di incendi divisi per fascia oraria all'interno di ciascuno stato.

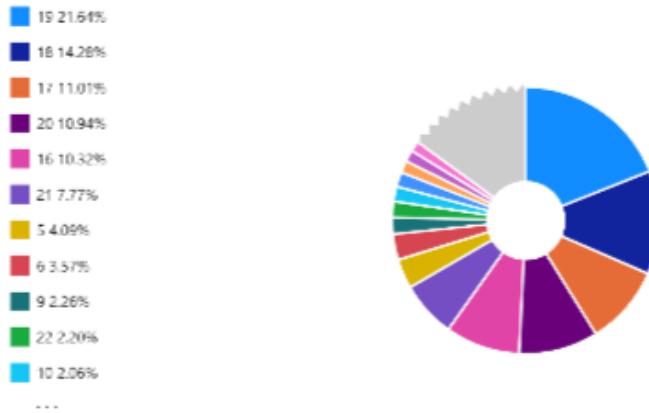


Il primo grafo da osservare è la distribuzione degli incendi all'interno delle 4 fasce orarie create, considerando tutti gli stati degli USA. Si nota immediatamente che la fascia con più incendi è quella serale, che va dalle 18 alle 24, fascia in cui probabilmente si hanno più atti vandalici e accensione di fuochi nei boschi.

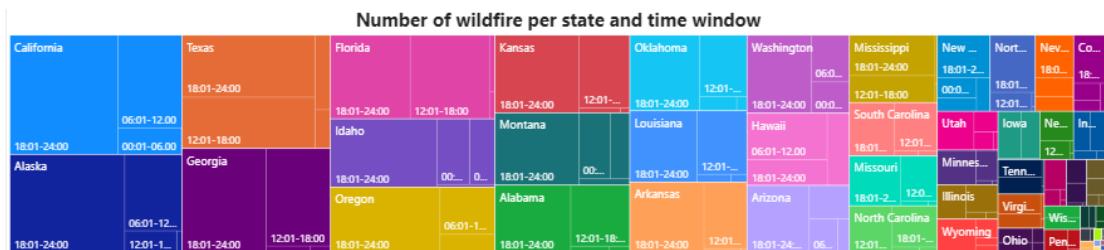


La seconda torta, ovvero quella sulla sinistra, rappresenta in maniera ordinata rispetto al numero degli incendi, tutte le ore della giornata e la relativa percentuale di incendi in quello specifico orario. Si nota immediatamente che le ore con più incendi ricadono effettivamente nella fascia oraria determinata precedentemente, e sono appunto le ore 18 e 19.

Number of wildfire per hour



Infine nella parte inferiore della pagina si può notare la distribuzione del numero degli incendi, stato per stato, nelle 4 fasce orarie determinate, ottenendo così una visualizzazione immediata degli stati con più incendi e su quali fasce orarie essi sono distribuiti.



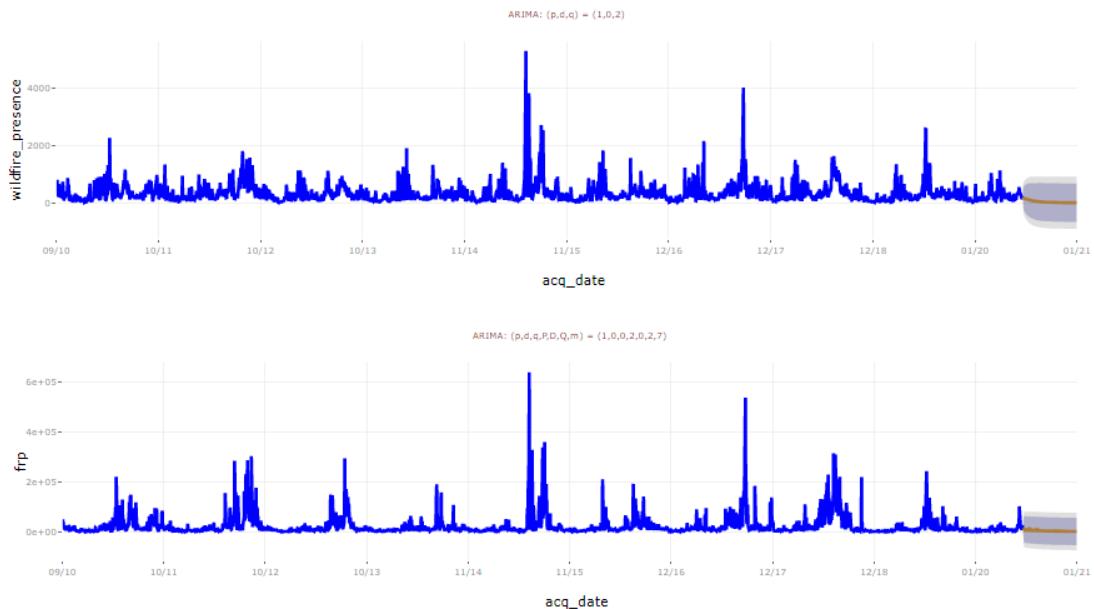
4.4 Dashboard R-Based

In queste pagine verrà svolta un'analisi di tipo predittivo sul numero di incendi e sull'andamento dei principali parametri. Per fare ciò si utilizzano elementi grafici costruiti attraverso il linguaggio *R*, che sono *Forecasting with ARIMA* e *Forecasting TBATS*.

Inoltre verrà fatta un'analisi delle correlazioni tra i vari attributi del dataset, ed infine verranno creati alcuni cluster su dei dati più significativi.

4.4.1 Forecasting with ARIMA

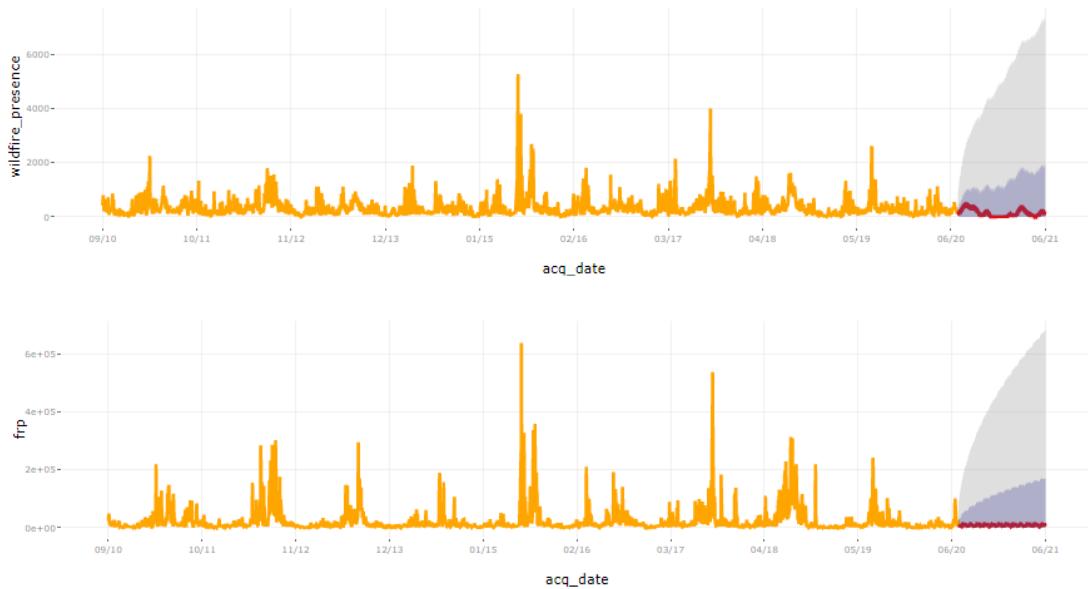
In questa pagina viene fatta un'analisi predittiva sull'andamento del numero di incendi e della *frp* totale giornaliera irradiata, per l'intero Paese.



Nel primo grafico viene rappresentato l'andamento del numero totale di incendi nei 10 anni, e poi attraverso *ARIMA* viene predetto l'andamento nei successivi 200 giorni, impostando una stagionalità dei dati di tipo annua. Nel secondo grafico invece viene rappresentato l'andamento della potenza irradiata giorno per giorno, e anche in questo caso attraverso *ARIMA* viene predetto l'andamento nei successivi 200 giorni, impostando una stagionalità dei dati di tipo settimanale. I risultati ottenuti non sono evidentemente soddisfacenti, e ciò è probabilmente dovuto all'alta complessità dei dati e alla poca correlazione nel tempo che essi hanno, perciò si è optato nell'utilizzare un diverso strumento visivo disponibile sul market place, ovvero *Forecasting TBATS*

4.4.2 Forecasting TBATS

BATS (Box-Cox transformation, ARMA errors, Trend and Seasonal components), basato su un modello di exponential smoothing, utilizza la trasformazione Box-Cox per trattare dati non lineari ed ARMA per decorrelare i dati dalle serie temporali. Tuttavia, *BATS* non funziona bene quando la seasonality è complessa e ad alta frequenza. Per risolvere questo problema è stato introdotto il modello *TBATS*(Trigonometric Seasonal + BATS). L'espressione trigonometrica dei termini di stagionalità fornisce al modello una maggiore flessibilità per affrontare la stagionalità complessa e riduce drasticamente i parametri del modello nel caso di stagionalità con frequenze elevate.



Nel primo grafico viene fatta una previsione dell'andamento del numero degli incendi nei prossimi 365 giorni, e si nota subito che la predizione sembra molto più significativa rispetto a quella fatta con *ARIMA*. Ciò probabilmente è dovuto alla possibilità di definire una doppia stagionalità, e in questo caso viene posto:

- Seasonal factor 1 = year
- Seasonal factor 2 = week

Nel secondo grafico, analogo a quello della pagina precedente, viene rappresentata la previsione dell'andamento della potenza irradiata dagli incendi nei prossimi 365 giorni.

ni. In questo caso però, nonostante la doppia stagionalità, non si ottengono i risultati sperati. I parametri utilizzati per le stagionalità sono:

- Seasonal factor 1 = month
- Seasonal factor 2 = week

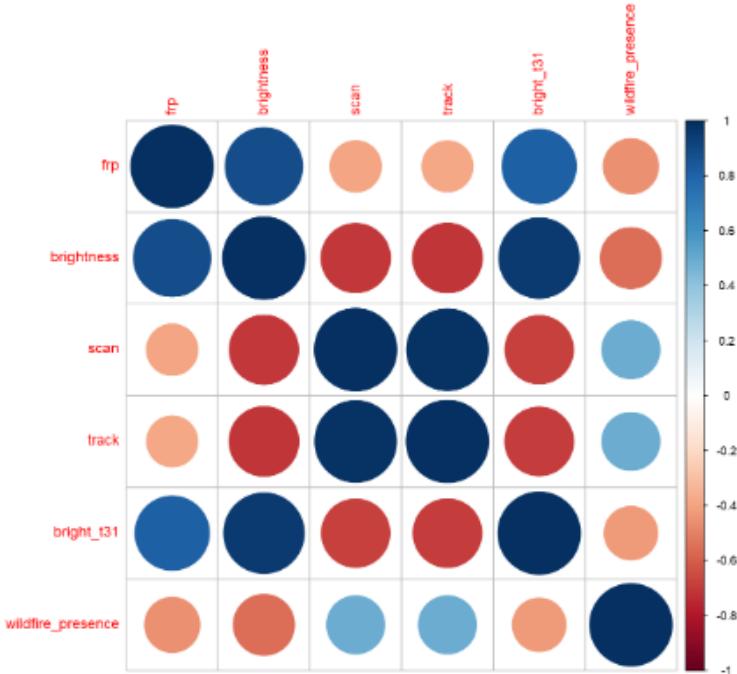
4.4.3 Correlazione degli attributi

In questa pagina, creata sfruttando l'integrazione di R in PowerBI, si cerca di fare emergere la correlazione tra alcune grandezze principali. Si è utilizzata un'estensione di PowerBI scaricabile dal Marketplace chiamata *Correlation Plot*, che fornisce una semplice interfaccia grafica per analizzare la presenza o meno delle correlazioni. Gli attributi analizzati sono in totale 5:

- frp
- brightness
- scan
- track
- bright_t31

Si vuole quindi mettere in relazione tutte quelle grandezze che forniscono una misura quantitativa dei principali attributi che identificano un incendio.

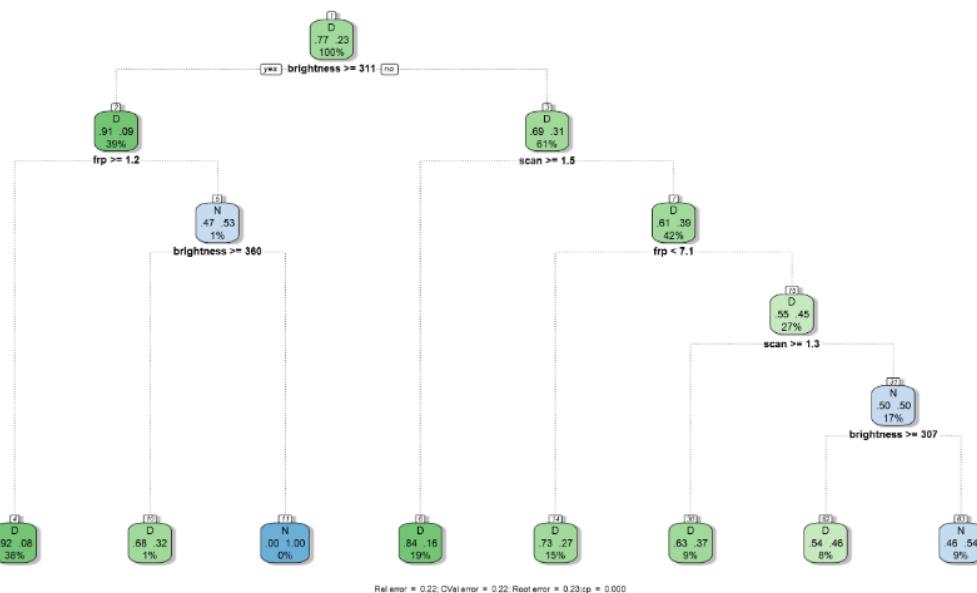
Il grafico ha nella parte superiore i nomi degli attributi e nella parte laterale sinistra il nome di quelli di cui si verifica la correlazione. Ogni quadrato quindi sarà associato a due grandezze e il cerchio all'interno di esso rappresenta il grado di correlazione: più un cerchio è grande e tendente al blu più avrà una correlazione elevata, più il cerchio è grande e tendente al rosso e più la correlazione è inversa, infine più è tendente al bianco e più la correlazione è assente. La diagonale principale della matrice che si viene a creare ha quadrati contenenti cerchi grandi e blu scuro in quanto indica che una grandezza è in correlazione con sé stessa: è una caratteristica di ogni matrice di correlazione. Per giunta si tratta di una matrice simmetrica: per questa ragione limitiamo le considerazioni alla parte superiore, in quanto le altre sono speculari. Esistono diverse correlazioni tra le variabili. Infatti si ha che *scan* e *track* sono in forte correlazione tra loro, indicando infatti



entrambe la dimensione dell’incendio, ma la prima è normalizzata rispetto all’angolo che si forma tra la direzione incendio-satellite e la perpendicolare al satellite stesso. Inoltre si nota una forte correlazione tra la misura *brightness* e *bright_t31*, dato che anche in questo caso entrambe esprimono la brillantezza, ovvero la luce emessa dall’incendio, ma raccolta sotto forma di dato attraverso due tecnologie diverse. In aggiunta si vede una forte correlazione tra *frp* e *brightness* (e di conseguenza con *bright_t31*), essendo entrambe indicatore dell’energia sprigionata dall’incendio, mentre si nota una correlazione inversa con la *scan*. Ciò probabilmente è dovuto al fatto che più l’incendio è esteso, e minore è la potenza irradiata per ciascun pixel captato dal satellite. Stesso discorso per la correlazione tra *frp* e *wildfire_presence*, poiché la potenza irradiata per pixel è inversamente proporzionale al numero di pixel che uno stesso incendio occupa, poiché si ha la potenza più distribuita e non concentrata in un unico punto. Infine si nota una correlazione diretta tra la *scan* (e quindi anche la *track*) con la *wildfire_presence*, poiché più l’incendio è grande e maggiore è la probabilità che esso possa essere visto come formato da più incendi.

4.4.4 Albero decisionale

In questa pagina è stato sfruttato l'oggetto visivo scaricabile dal marketplace chiamato *Decision Tree*. Come dice il nome esso consente di costruire alberi di decisione a partire da un dataset. Si è cercato di predire il valore dell'attributo *daynight* che indica appunto se l'incendio è avvenuto di giorno o di notte, a partire dai valori di *brightness*, *frp* e *scan*.



L'albero creato è un albero binario con profondità massima di 5. Gli split ad ogni livello vengono fatti sugli attributi numerici, creando intervalli di valori che cercano di massimizzare l'information gain. I risultati che otteniamo hanno un errore relativo pari al 22%: è un risultato buono ma non eccellente.

4.4.5 Clustering per state

L'oggetto visivo utilizzato in questa pagina è chiamato *Clustering* e anche esso può essere scaricato dal marketplace. La funzione di questo grafico è quello di clusterizzare gli eventi rispetto ad un insieme di attributi. In particolare si è deciso di creare dei cluster per stato, in base ai valori di *frp*, *wildfire_presence*, *scan*, e *brightness*, andando così ad identificare diversi insiemi di stati accomunati dalla stessa gravità di incendi.

Nella figura è possibile osservare come i cluster creati siano effettivamente 3. In quello rosso, posizionato nell'angolo più alto a destra, sono rappresentati gli stati che hanno

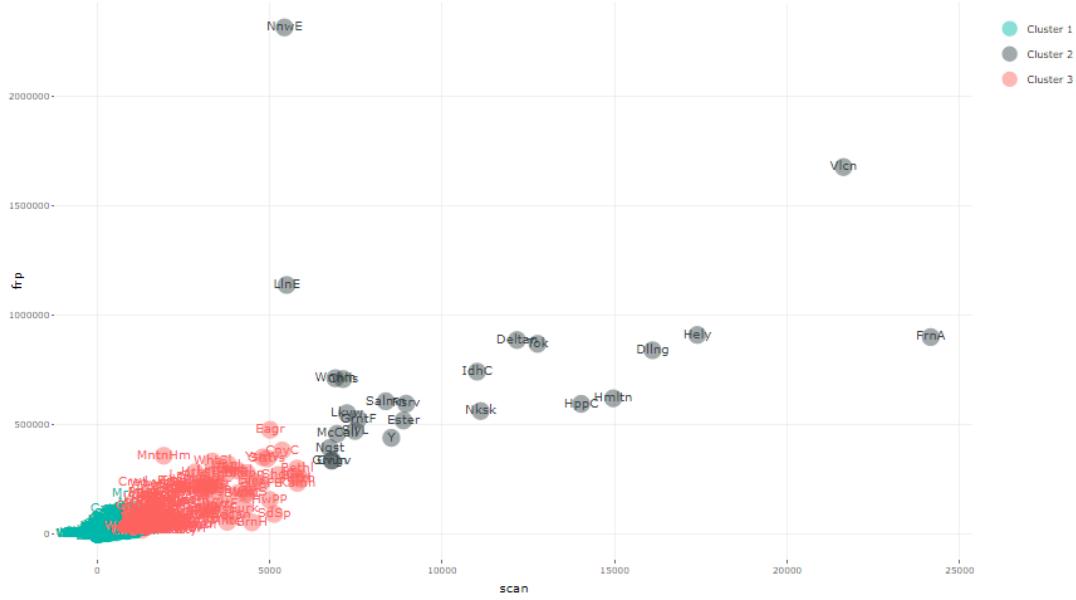


avuto un numero di incendi maggiori, e quindi con una potenza irradiata e un'espansione complessiva maggiore, e si vede immediatamente come la California domini su tutti gli altri stati. Si hanno poi altri due cluster, che indicano gli stati con un numero di incendi (e di conseguenza di potenza irradiata e dimensione totale) medio e basso.

4.4.6 Clustering per city

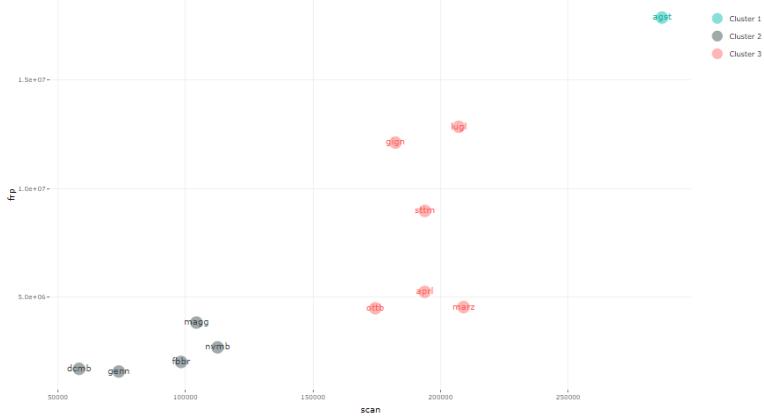
In questo caso si è deciso di creare dei cluster per città, in base ai valori di *frp*, *wildfire_presence*, *scan*, e *brightness*, andando così ad identificare diversi insiemi di città accomunati dalla stessa gravità di incendi.

Nella figura è possibile osservare come i cluster creati siano effettivamente 3. In quello grigio, posizionato nell'angolo più alto a destra, sono rappresentati le città che hanno avuto un numero di incendi maggiori, e quindi con una potenza irradiata e un'espansione complessiva maggiore. Se si scorre con il puntatore sopra ai punti del grafo, compare una nota con tutti i parametri registrati in quella città e di quale stato fa parte. Si notano che le città dove sono avvenuti gli incendi peggiori sono quelle delle Hawaii, come Volcano e Fern Acres, probabilmente dovuti alle tante eruzioni vulcaniche. Si hanno poi altri due cluster, che indicano le città con un numero di incendi (e di conseguenza di potenza irradiata e dimensione totale) medio e basso.



4.4.7 Clustering per mese

In questa ultima pagina si è deciso di creare dei cluster per mese, in base ai valori di *frp*, *wildfire_presence*, *scan*, e *brightness*, andando così ad identificare diversi insiemi di mesi accomunati dalla stessa gravità di incendi. Nella figura è possibile osservare come



i cluster creati siano effettivamente 3. In quello azzurro, posizionato nell'angolo più alto a destra, sono rappresentati i mesi che hanno avuto un numero di incendi maggiori, e quindi con una potenza irradiata e un'espansione complessiva maggiore. Si nota subito che questo cluster è formato solamente da un mese, ovvero Agosto, che sarà appunto il mese con gli incendi peggiori dell'anno, e ciò molto probabilmente è dovuto alle alte temperature che solitamente si registrano. Si hanno poi altri due cluster, che indicano le città con un numero di incendi (e di conseguenza di potenza irradiata e dimensione

CAPITOLO 4: POWER BI

totale) medio e basso: i primi sono i mesi primaverili e autunnali, mentre i secondi quelli invernali (ad esclusione di Maggio).

5. Conclusioni

Dopo aver esaminato il lavoro svolto in Qlik, Tableau e Power BI, si è in grado di valutare pro e contro di ciascun tool, mettendo in risalto le differenze sostanziali che li caratterizzano. Come già detto nei capitoli precedenti tutti i tools ci hanno permesso di condurre un'analisi descrittiva e diagnostica. Inoltre Tableau ci ha permesso di condurre una minima parte di analisi predittiva attraverso le serie temporali e di effettuare una prima parte di clusterizzazione e classificazione. Questa parte è stata poi approfondita in Power BI, grazie alla sua integrazione con R. Quest'ultimo risulta infatti essere quello che offre le maggiori possibilità espressive a livello di analisi.

Anche dal punto di vista del caricamento e della gestione dei dati si ritiene che il tool più affidabile sia Power BI, dato che mette a disposizione un'interfaccia molto simile ad Excel e permette di effettuare numerose operazioni di ETL in maniera semplice. Su Tableau non è stato possibile effettuare tutte le operazioni di ETL fatte in Power BI dato che la sua interfaccia non è così intuitiva come quella di Power BI e le funzioni messe a disposizione sono più limitate. Qlik infine permette il caricamento dei dati in maniera semplice, ma le operazioni che possono essere effettuate sui dati sono molto limitate rispetto agli altri tools.

Dal punto di vista della creazione delle Dashboard, possiamo dire che Qlik e Power BI risultano concettualmente molto simili, in quanto a partire da una Dashboard è possibile inserire al suo interno delle tipologie di grafico messe a disposizione dai tools. Tableau invece permette di sviluppare ogni grafico su un foglio di lavoro distinto e poi raggruppare un insieme di fogli per formare una Dashboard.

Uno svantaggio di Qlik riguarda la gestione dei filtri: quando si seleziona un filtro in una Dashboard, esso influenza sia i grafici della stessa Dashboard sia i grafici presenti nelle altre. In Tableau e Power BI invece i filtri influenzano esclusivamente la Dashboard di

CAPITOLO 5: CONCLUSIONI

appartenenza e in particolare si può anche applicare un filtro diverso per ogni singolo grafico all'intero della Dashboard: questo permette di avere un focus diverso per ogni grafico ed effettuare un'analisi più variegata. Il numero di grafici messi a disposizione è elevato in tutti e tre i tools, anche se va detto che Power BI, grazie all'integrazione con R, estende di molto le sue funzionalità. Le tipologie di grafico presenti sono più o meno le stesse e ciò permette di replicare l'analisi sviluppata inizialmente in Qlik anche negli altri tools, ma si è cercato comunque di estrarre informazioni diverse.

Una caratteristica importante da considerare riguarda la gestione delle mappe nei vari tools. Il tool che si comporta meglio sotto questo punto di vista è Tableau perché riesce a gestire gli attributi come le aree geografiche e le città ma anche attributi più specifici come latitudine e longitudine. Power BI si comporta comunque bene, mentre Qlik risulta essere il peggiore dal momento che non consente di rappresentare correttamente i valori delle coordinate geografiche, infatti come già detto è stato necessario integrare un file con le coordinate di tutti gli stati e le città degli Stati Uniti.

A livello di utilizzo si può affermare che il tool più semplice da usare è Qlik, in quanto consente di sfruttare le sue funzionalità in maniera davvero immediata per l'utente, pur disponendo di un discreto numero di strumenti. Inoltre è sviluppato in Cloud e ciò evita di fare installazioni sulla propria macchina e di aver accesso ai file da qualunque dispositivo. Anche Power BI risulta essere abbastanza semplice nell'utilizzo, mentre Tableau sembra essere quello un po' più macchinoso e meno istantaneo da capire.

A livello grafico e di interfaccia si può dire che tutti e tre i tools sono piuttosto gradevoli. Power BI è quello che rimane un po' più indietro rispetto agli altri in termini di grafica, in quanto Tableau e Qlik hanno una grafica più accattivante e moderna, sia a livello di colori che di aspetto dei grafici. Tuttavia il tool più maturo è Power BI, in quanto combina numerose funzionalità, un utilizzo semplice e una grafica comunque gradevole. Il tool più innovativo dei tre che consente di sviluppare analisi descrittiva, diagnostica, predittiva e prescrittiva è Power Bi, ed infatti secondo il *Magic Quadrant* di Gartner è lui il leader nel settore.

Bibliografia

- [1] Our World Data. Wildfires in the united states, 2020. URL https://en.wikipedia.org/wiki/Wildfires_in_the_United_States.