

UNIVERSITÀ POLITECNICA DELLE MARCHE

INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE

PROGETTO DATA SCIENCE 2021-2022

**Analisi delle Olimpiadi nel periodo
1896-2016 mediante l'uso dei tools Qlik e
Tableau**



Studenti:

MASSIMO CIAFFONI

SIMONE CAPPANERA

Indice

1	Introduzione	2
1.1	Dataset	2
1.2	ETL	4
2	Qlik	5
2.1	Caricamento dei dati	5
2.2	Analisi e creazione Dashboard	6
2.3	Fogli	6
2.3.1	Prospetto Generale	6
2.3.2	Analisi Atleti	10
2.3.3	Analisi Nazioni	15
2.3.4	Analisi Eventi Sportivi	18
3	Tableau	23
3.1	Caricamento Dati	23
3.2	Analisi e creazione delle Storie	24
3.3	Analisi Generale	24
3.3.1	Prima dashboard	24
3.3.2	Seconda dashboard	27
3.4	Analisi Atletica	31
3.4.1	Distribuzione medaglie ed eventi	31
3.4.2	Caratteristiche fisiche delle atlete femminili	36
3.5	Analisi Italia	37
3.5.1	Distribuzione medaglie e sport	37
3.5.2	Caratteristiche Atleti	41

Capitolo 1

Introduzione

1.1 Dataset

Il dataset utilizzato per l'analisi descrive i 120 anni delle Olimpiadi moderne a partire dalla prima edizione di Atene del 1896 fino all'edizione di Rio del 2016. In particolare il dataset descrive la partecipazione degli atleti ai diversi eventi svolti durante le Olimpiadi; per ogni atleta partecipante ad un determinato evento sono descritte, se disponibili, le caratteristiche fisiche ed anagrafiche, il team d'appartenenza, l'eventuale medaglia vinta e altre informazioni di carattere generale di quel particolare evento. Inoltre, si noti che i giochi invernali ed estivi si sono svolti nello stesso anno fino al 1992. Successivamente, le due edizioni sono state divise in modo tale che i Giochi invernali si svolgessero in un ciclo di quattro anni a partire dal 1994 mentre quelli estivi a partire dal 1996. Questa discrepanza temporale delle varie edizioni deve essere considerata nella descrizione e nella diagnosi dei vari grafici che si sono generati. Il dataset è disponibile al seguente link:

<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

Il dataset è composto da due tabelle:

- **athlete_events.csv**: descrive tutte le partecipazioni degli atleti alle Olimpiadi.
- **dictionary.csv**: descrive i diversi comitati olimpici esistenti e le nazioni di appartenenza.

Di seguito sono descritti i vari campi che compongono le tabelle:

Athlete Events

Nome Campo	Tipo di dato	Descrizione
ID	Number	Identificativo numerico dell'atleta partecipante.
Name	Text	Nome dell'atleta partecipante.
Sex	Text	Sesso dell'atleta partecipante che può assumere il valore 'M' o 'F'.
Age	Text	Età dell'atleta partecipante. Se il dato non è disponibile il campo sarà dato dalla stringa 'NA'.
Height	Text	Altezza dell'atleta partecipante. Se il dato non è disponibile il campo sarà dato dalla stringa 'NA'.
Weight	Text	Peso dell'atleta partecipante. Se il dato non è disponibile il campo sarà dato dalla stringa 'NA'.
Team	Text	Team dell'atleta.
NOC	Text	Codice identificativo a 3 caratteri del comitato olimpico rappresentato dall'atleta.
Games	Text	Edizione dell'Olimpiade in cui si è svolto l'evento.
Year	Number	Anno di svolgimento dell'evento.
Season	Text	Stagione di svolgimento dell'evento che può assumere i valori 'Winter' o 'Summer'.
City	Text	Città in cui si è svolto l'evento (in generale l'edizione dell'Olimpiade)
Sport	Text	Sport di riferimento dell'evento. Rappresenta una supercategoria degli eventi (Es. Atletica ->Maratona Uomini)
Event	Text	Evento a cui l'atleta ha partecipato
Medal	Text	Medaglia vinta dall'atleta. Se l'atleta non ha vinto alcuna medaglia in quell'evento il campo avrà valore 'NA'.

Dictionary

Nome Campo	Tipo	Descrizione
NOC	Text	Codice identificativo a 3 caratteri del comitato olimpico
Country	Text	Nome della Nazione
Notes	Text	Informazioni extra

1.2 ETL

Il dataset relativo agli eventi olimpici è stato ripulito tramite l'uso dello strumento di caricamento dei dati messo a disposizione da Qlik:

- I campi *Age*, *Height*, *Weight*, *Medal* sono stati modificati da tipo testuale a numerico andando a sostituire i valori non disponibili con valori nulli. In questo modo è stato possibile fare un'analisi numerica per questi campi.

Per quanto riguarda il dataset *dictionary* alcuni campi sono stati modificati:

- Alcuni valori del campo *Country* erano segnati con valore NA (Es. Il codice NOC UNK con notes Unknown). In tutti questi casi, i vari NOC non erano presenti nel dataset degli eventi olimpici, quindi è stato sufficiente eliminare i valori NA in modo che, nell'analisi, l'etichetta non apparisse come una nazione.
- Due record avevano un valore diverso di country (*Russia e Soviet Union*). In questo caso, volendo fare un'analisi storica della nazione, il valore Soviet Union è stato sostituito con Russia, in modo tale che le medaglie vinte dall'Unione Sovietica apparissero nel conteggio aggiunte a quelle della Russia.

Capitolo 2

Qlik

Qlik è un software di visualizzazione e business intelligence che permette di svolgere analisi, prettamente di tipo descrittivo e diagnostico, riuscendo a presentare i dati in maniera intuitiva e facilmente comprensibile tramite l'utilizzo di dashboard completamente personalizzabili. Una caratteristica molto interessante è la possibilità di cambiare dinamicamente le visualizzazioni sulla base dei filtri.

2.1 Caricamento dei dati

Creata una applicazione abbiamo caricato le tabelle tramite l'interfaccia grafica messa a disposizione. Al termine del caricamento è stata creata l'associazione consigliata da Qlik tra le due tabelle tramite l'attributo comune "NOC" della tabella `athletes_events` e della tabella `dictionary`.



Figura 2.1: Associazione visualizzata in Qlik

2.2 Analisi e creazione Dashboard

E' stato deciso di dividere l'applicazione in quattro dashboard contenenti una serie di grafici relativi ad uno specifico campo di analisi. Le dashboard sono:

- **Prospetto generale:** panoramica generale delle Olimpiadi.
- **Analisi Atleti:** descrizione degli atleti che hanno partecipato alle Olimpiadi.
- **Analisi Nazioni:** analisi relativa alle varie nazioni che hanno partecipato ai giochi.
- **Analisi Eventi Sportivi:** analisi relativa agli eventi sportivi che si sono svolti durante i giochi.

Nel successivo paragrafo i diversi fogli sono descritti in dettaglio.

2.3 Fogli

2.3.1 Prospetto Generale

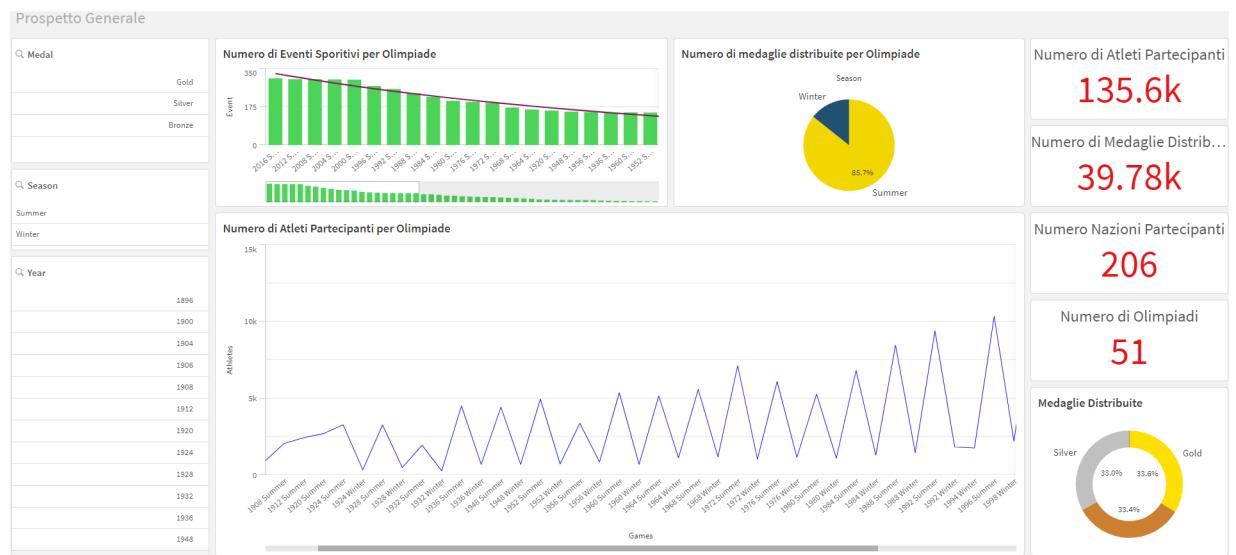


Figura 2.2: Dashboard sul prospetto generale delle Olimpiadi

In questo foglio sono descritte in generale le Olimpiadi in relazione al numero di eventi, di nazioni ed atleti partecipanti e di medaglie distribuite nel corso delle varie edizioni. Sulla sinistra sono presenti tre filtri principali, i quali sono riutilizzati

nelle diverse dashboard prodotte. I filtri vanno a discriminare tra la tipologia di medaglia distribuita ("Medal"), tra la tipologia di olimpiade ("Season") e l'anno di svolgimento ("Year"). La dashboard mostra 4 grafici e 4 **KPI** in alto a destra i quali conteggiano il numero di atleti e di nazioni che hanno partecipato ed il numero di medaglie ed edizioni delle Olimpiadi del periodo 1896-2016. In particolare il numero di atleti e nazioni è stato conteggiato tramite un *count distinct* in modo da non contare più volte lo stesso atleta (che ha partecipato a più gare e/o più edizioni) o una nazione. Dall'analisi dei KPI è possibile dedurre che in 51 edizioni delle Olimpiadi hanno partecipato più di 135.000 atleti provenienti da 206 nazioni e che sono state distribuite più di 39.000 medaglie.

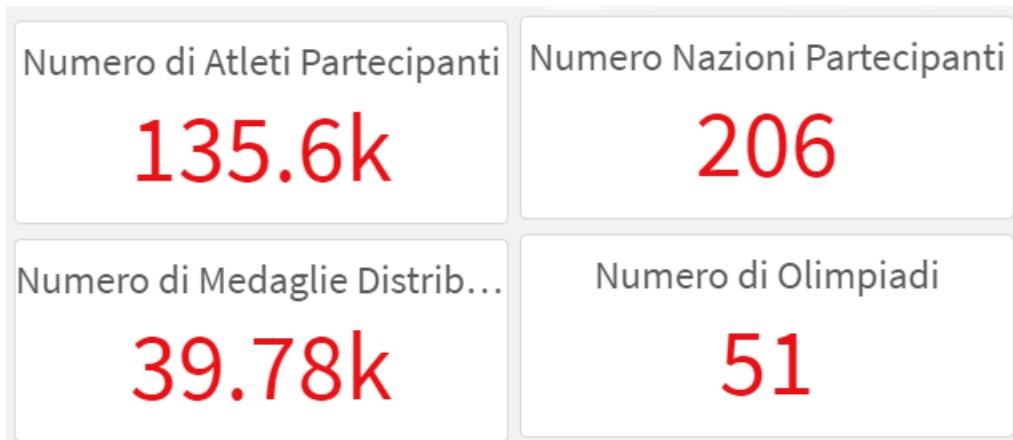
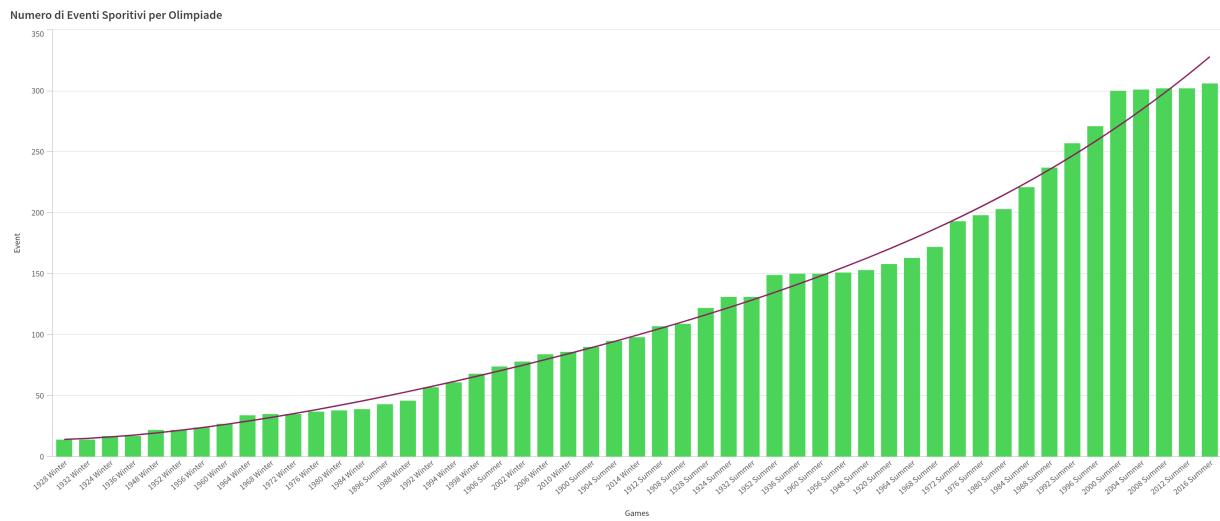


Figura 2.3: KPI generali delle Olimpiadi

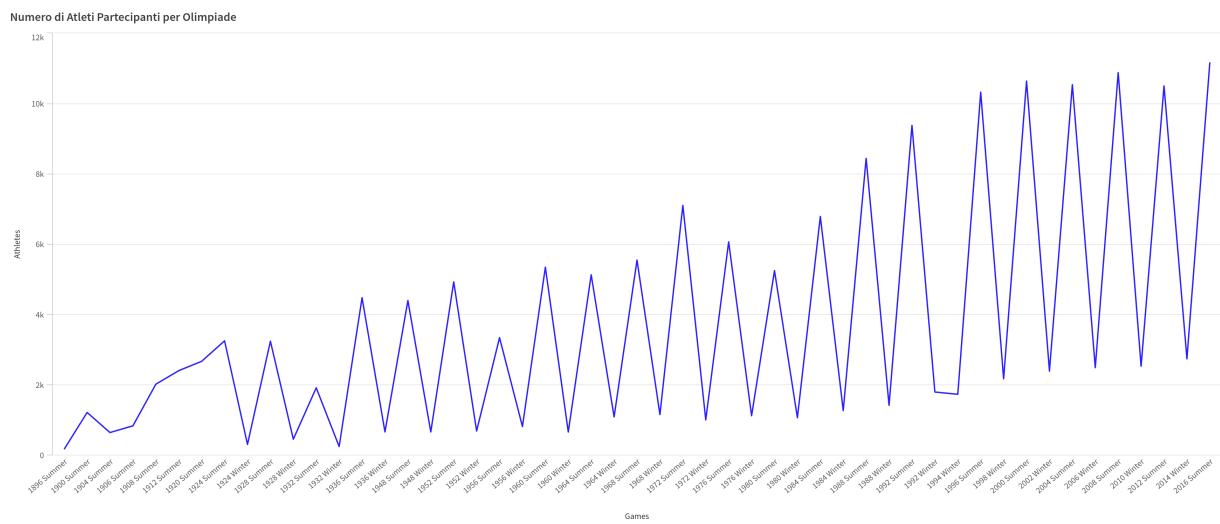
Numero di Eventi Sportivi per Olimpiade

Il grafico a barre mostra il numero di eventi per ogni edizione delle Olimpiadi. Il grafico è stato generato scegliendo come dimensione l'attributo **Games** e come misura è stato effettuato un *count distinct* degli **Event**. E' interessante notare come in generale con il passare degli anni il numero di eventi sportivi (e quindi anche il numero di discipline) presenti in un'edizione sia aumentato e che generalmente l'edizione estiva delle olimpiadi presenti un numero di eventi superiore rispetto a quelle invernali.



Numero di Atleti Partecipanti per Olimpiade

Il grafico a linea mostra il numero di atleti partecipanti per ogni edizione delle Olimpiadi. Il grafico è generato scegliendo come dimensione l'attributo **Games** e come misura è stato aggregato l'attributo **Name** tramite un *count distinct*. Anche in questo caso si può notare una crescita del numero di partecipanti dovuta sicuramente anche all'aumentare del numero di eventi sportivi svolti ed, in generale, una maggiore partecipazione nelle edizioni estive rispetto a quelle invernali.

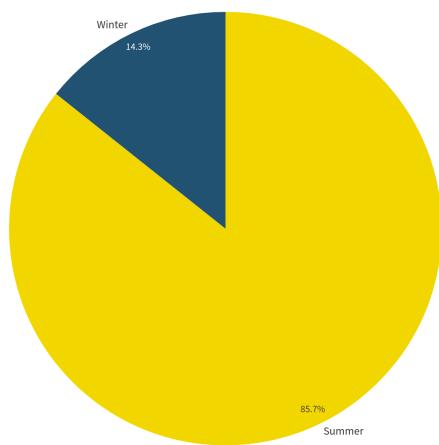


Medaglie Distribuite

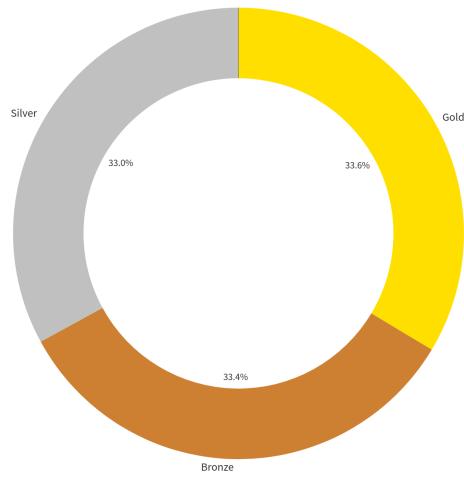
Infine è stata prodotta un'analisi sulle medaglie distribuite tramite la generazione di due grafici a torta: il primo mostra come il numero di medaglie distribuite avvenga maggiormente nelle edizioni estive (come previsto, in quanto secondo i grafici precedenti il numero di eventi sportivi nelle edizioni estive è più del doppio rispetto a quelle invernali), mentre il secondo mostra una distribuzione abbastanza uniforme delle tre tipologie di medaglie attorno al 33% (le minime differenze sono sicuramente dovute ad eventi in cui due o più atleti sono arrivati nella medesima posizione).

Numero di medaglie distribuite per Olimpiade

Season



Medaglie Distribuite



2.3.2 Analisi Atleti

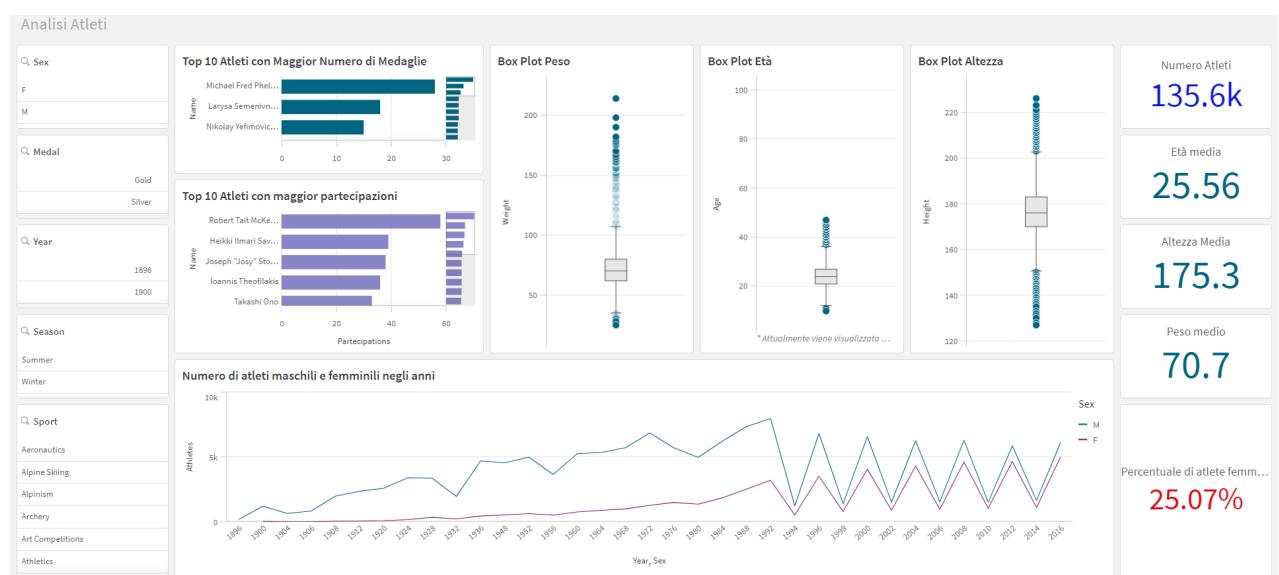


Figura 2.4: Dashboard sull'Analisi Atleti

Nella seconda dashboard è stata svolta un'analisi degli atleti che hanno partecipato alle varie edizioni delle Olimpiadi, descrivendo in particolare il numero degli atleti, la differenza tra il numero di atleti maschili e quelle femminili e il loro andamento negli anni, le Top 10 degli atleti rispettivamente con maggior numero

di medaglie vinte e con il maggior numero di partecipazioni ai giochi olimpici e i box plot dell'altezza, dell'età e del peso dei vari atleti. Sulla sinistra sono stati ripresi i tre filtri già elencati nel paragrafo precedente, ma oltre a questi ne sono stati aggiunti altri due, ossia quello che filtra gli atleti in base al sesso ("Sex") e quello che li filtra in base alla disciplina sportiva in cui gareggiano ("Sport"). La dashboard contiene 6 grafici e 5 **KPI**, i quali descrivono rispettivamente il numero complessivo di atleti; l'altezza, il peso e l'età media dei partecipanti e la percentuale di atlete femminili che hanno partecipato rispetto al totale (utile per capire nel corso degli anni quanto è diminuito il divario di genere nelle discipline olimpiche). In particolare l'ultimo KPI è stato realizzato utilizzando un colore condizionale il quale cambia da rosso a verde se la percentuale indicata supera il 40% (in questo modo è possibile evidenziare come negli ultimi anni il divario sia diminuito, infatti una volta filtrata la dashboard il valore sale oltre il 40%).

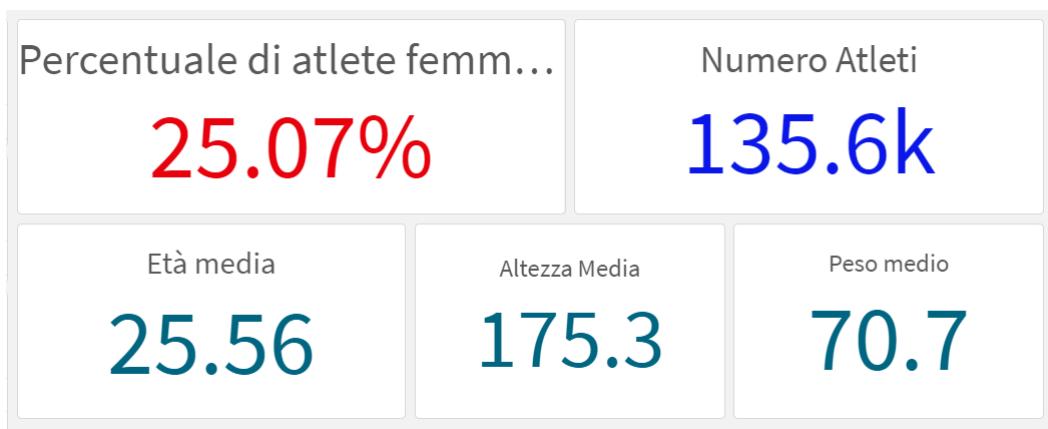
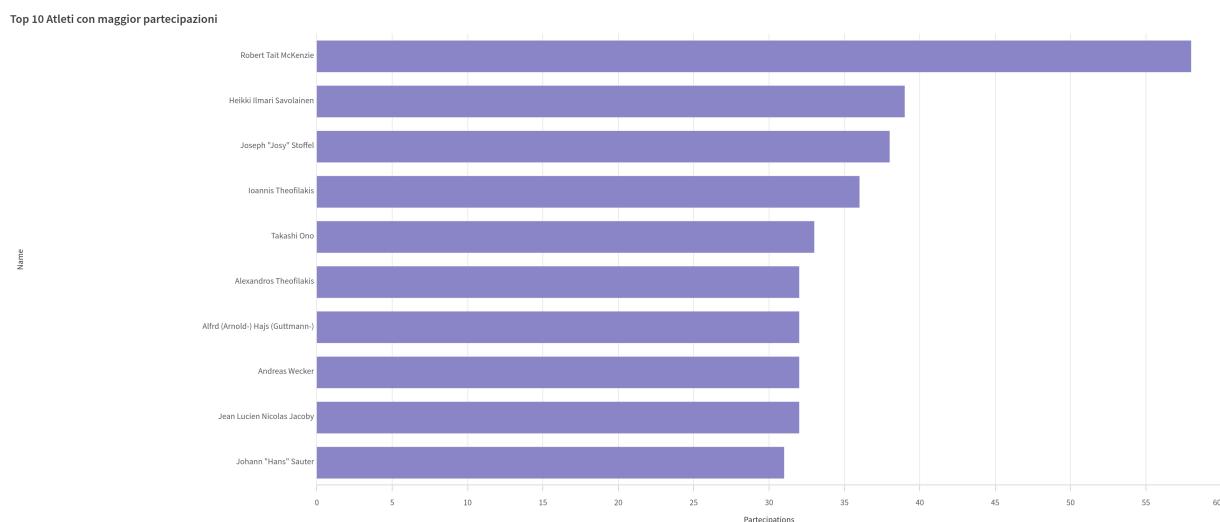
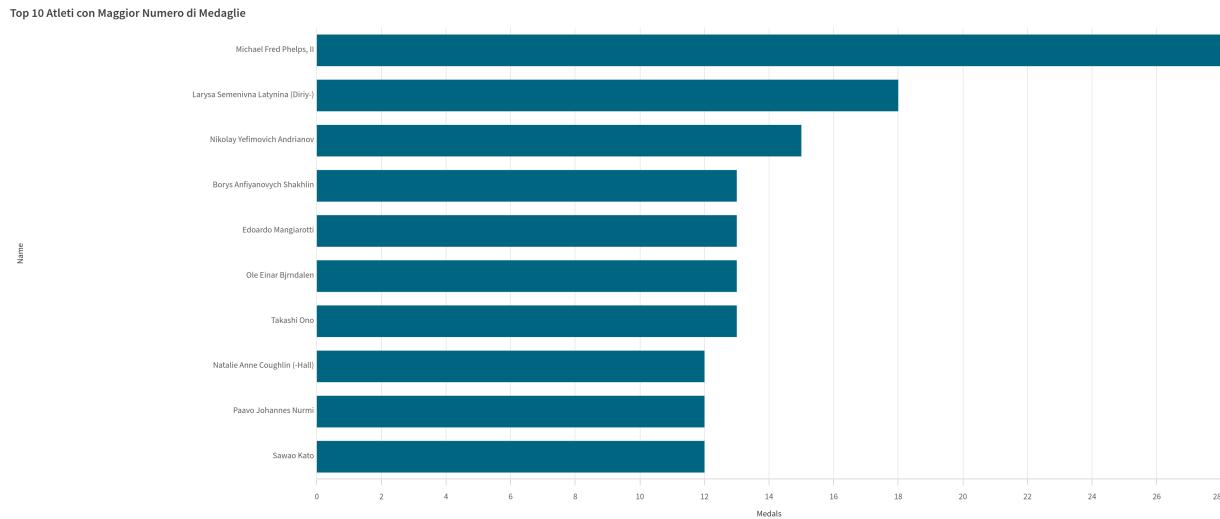


Figura 2.5: KPI relativi agli atleti e al divario di genere nelle discipline olimpiche

Top 10 Atleti con maggior numero di medaglie e maggior numero di partecipazioni

Una prima analisi effettuata è quella relativa agli atleti più famosi delle Olimpiadi. In particolare, sono stati scelti due grafici per mostrare quali sono gli atleti che hanno vinto più medaglie in assoluto e quali atleti hanno partecipato a più eventi in assoluto (in un'edizione un atleta può partecipare a più eventi che rappresentano una determinata specialità in uno sport). I due grafici a barre sono stati generati scegliendo come dimensione l'attributo **Name** e come misura il *count* del campo **Medal** nel primo caso e del campo **Events** nel secondo. In entrambi i casi sono stati selezionati i primi 10 elementi tramite un filtro in modo da non appesantire la visualizzazione. Il primo grafico mostra un relativo distacco del famoso campione olimpico "Michael Phelps" con un numero di medaglie pari a 28 rispet-

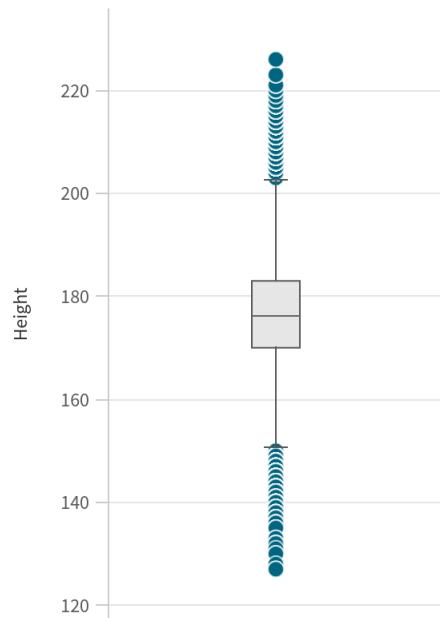
to alle 18 di "Larysa Semenivna Latynina". Anche il secondo grafico mostra un distacco di "Robert Tait McKenzie" con un ben 58 partecipazioni a diversi eventi olimpici rispetto alle 38 di "Heikki Savolainen".



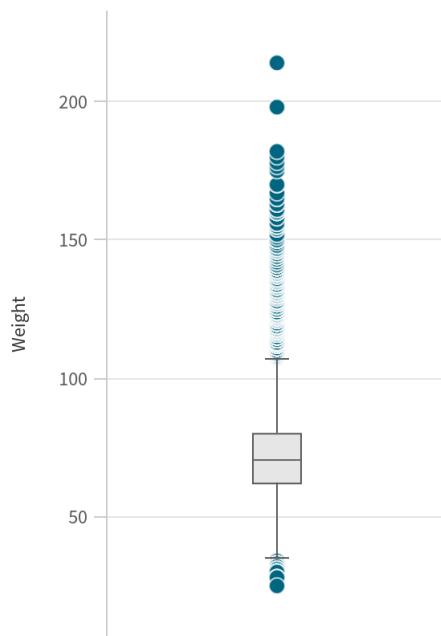
Box Plot Atleti

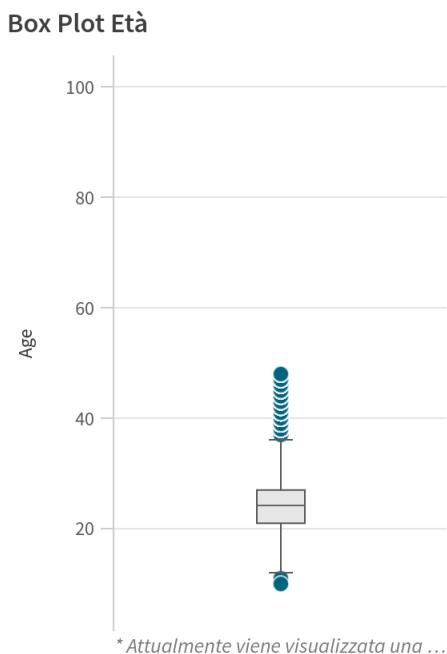
I tre diversi box plot mostrano la distribuzione delle caratteristiche fisiche (altezza e peso) ed anagrafiche (età) dei diversi partecipanti alle Olimpiadi e sono stati realizzati scegliendo come dimensione l'attributo **ID** e come misura il relativo campo che si vuole analizzare tra **Age**, **Weight** ed **Height**. I box plot inoltre mostrano i diversi outlier, cioè quali atleti hanno caratteristiche anomale rispetto alla mediana.

Box Plot Altezza



Box Plot Peso

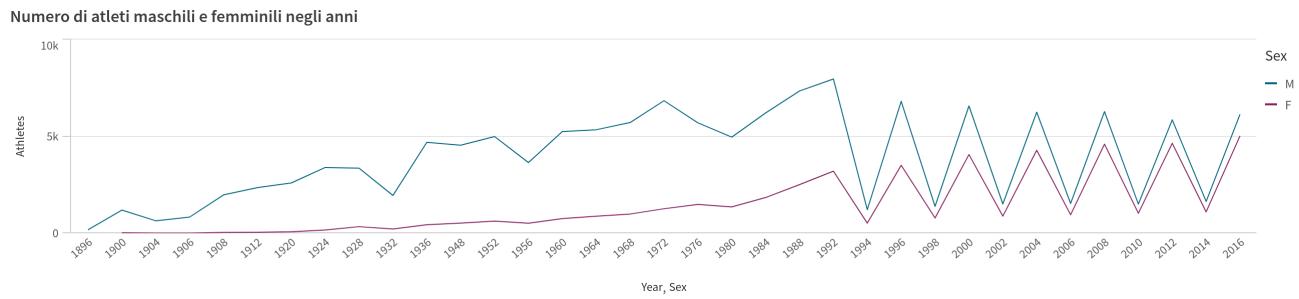




Numero di Atleti maschili e femminili negli anni

Il grafico a linea mostra come la disparità di genere nelle Olimpiadi sia diminuita nel corso degli anni. La visualizzazione è stata generata scegliendo come dimensioni il campo **Year** e il campo **Sex** (in modo da generare le due diverse linee, una per gli atleti maschili e una per quelle femminili) e come misura è stato effettuato un *count distinct* degli atleti tramite il campo **ID**. Il grafico mostra come negli anni il numero delle atleti femminili che partecipano alle Olimpiadi sia aumentato e come, in particolare, sia sempre minore la disparità di genere (è possibile utilizzare il KPI creato insieme al filtro per Year per avere un'idea della percentuale del numero di partecipanti femminili rispetto al totale).

N.B. Il grafico mostra un andamento diverso a partire dall'edizione del 1994. Questo perché le edizioni invernale ed estiva, che fino a quel momento si svolgevano nello stesso anno, dall'edizione estiva del 1992 sono state divise in anni diversi. Inoltre, come si evince dal grafico, l'edizione invernale ha un numero di partecipanti nettamente minore delle edizioni estive.



2.3.3 Analisi Nazioni

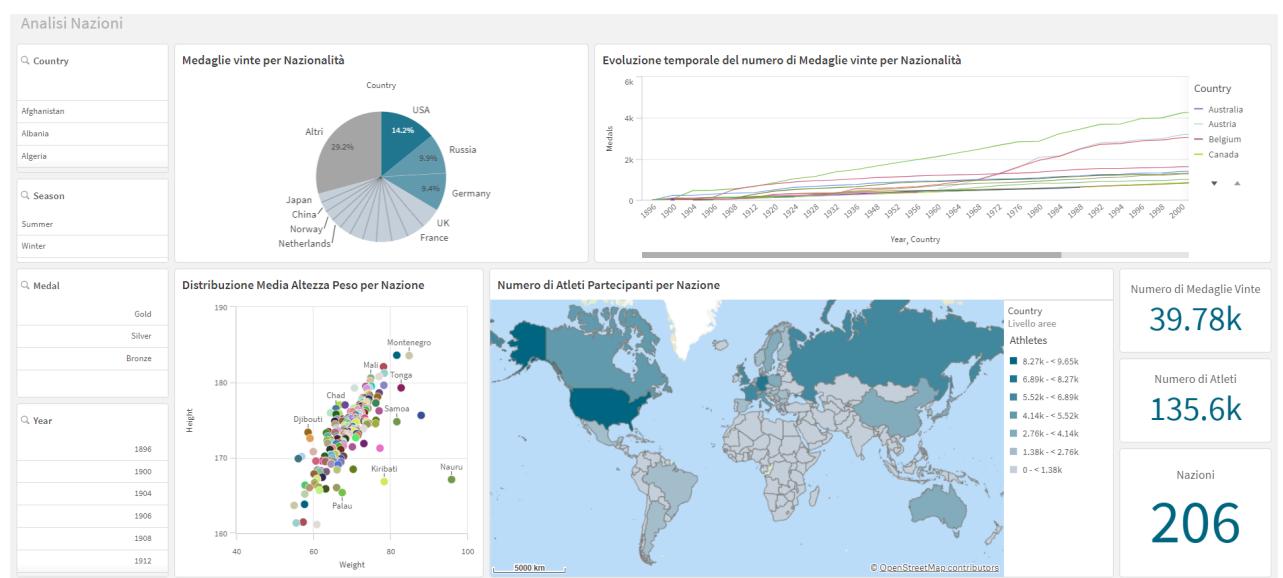
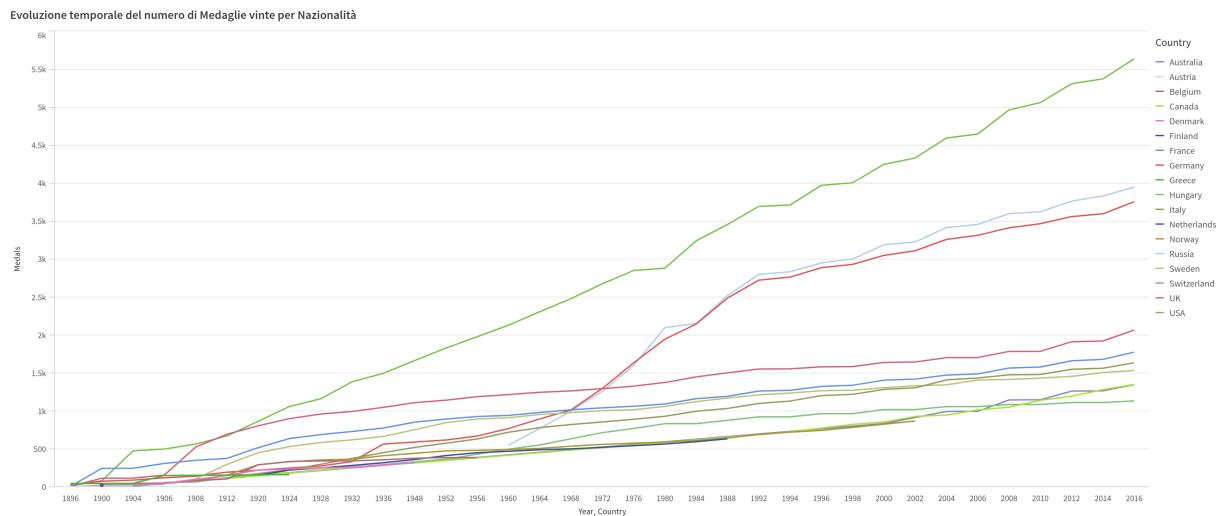
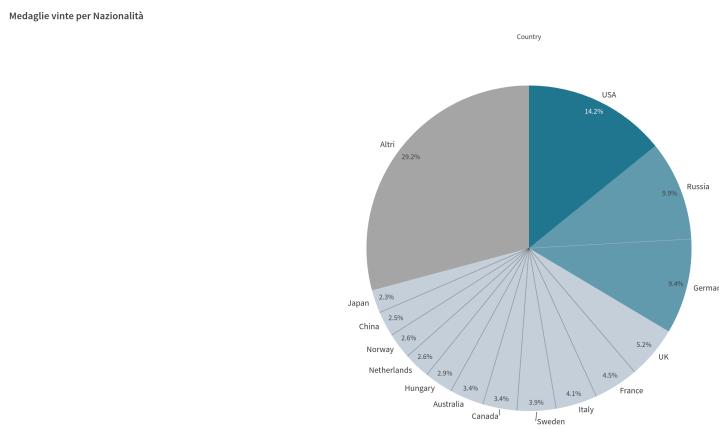


Figura 2.6: Dashboard sull'Analisi Nazioni

La terza dashboard mostra un'analisi delle Olimpiadi focalizzandosi sulle Nazioni, mostrando il numero e l'andamento temporale di acquisizione delle medaglie vinte e il numero di partecipanti, descrivendo con particolare attenzione le caratteristiche fisiche medie di quest'ultimi. Anche in questo caso sono stati utilizzati diversi filtri, i quali eseguono una selezione in base alla nazione considerata ("Country"), l'anno di svolgimento ("Year"), la tipologia di Olimpiadi ("Season") e il tipo di medaglia ottenuta ("Medal"). La dashboard è composta da 4 grafici e da 3 **KPI**. Quest'ultimi, ripresi dalle precedenti analisi, in questo caso ci descrivono il numero delle nazioni selezionate nella visualizzazione corrente, oltre al numero di medaglie vinte e di atleti partecipanti dalle nazioni prese in considerazione.

Medaglie vinte per nazionalità ed evoluzione temporale

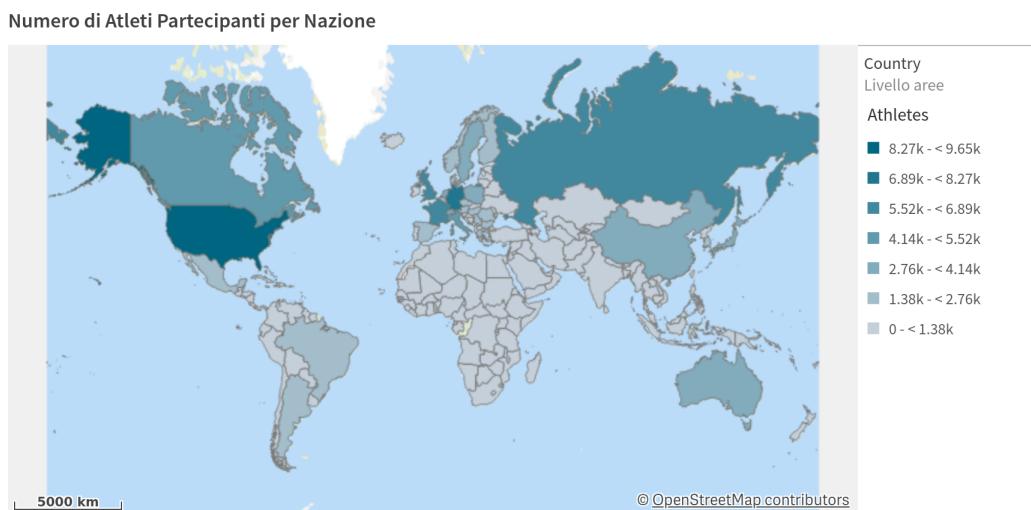
In questi due grafici si sono caratterizzate le medaglie vinte dalle nazioni. Nel primo si è realizzato un grafo a torta, il quale mostra la Top 14 delle nazioni per numero di medaglie, con l'ultima fetta della torta che rappresenta la percentuale di tutte le nazioni rimanenti fuori dalle prime 14. Il grafico è stato realizzato scegliendo come dimensione l'attributo **Country** e come misura l'attributo **Medal** aggregata tramite un *count* e mostra come circa il 70% delle medaglie vinte alle Olimpiadi siano state aggiudicate da 14 nazioni. Il secondo grafico, invece, mostra l'andamento temporale delle medaglie vinte dalle 10 nazioni con il più alto numero di medaglie; il grafo a linea è stato realizzato scegliendo come dimensioni i campi **Year** (come gruppo) e **Country** (come linea), mentre come misura si è utilizzato il campo **Medal** aggregato tramite *count* con un modificatore di accumulo.



Numero di Atleti partecipanti per nazione

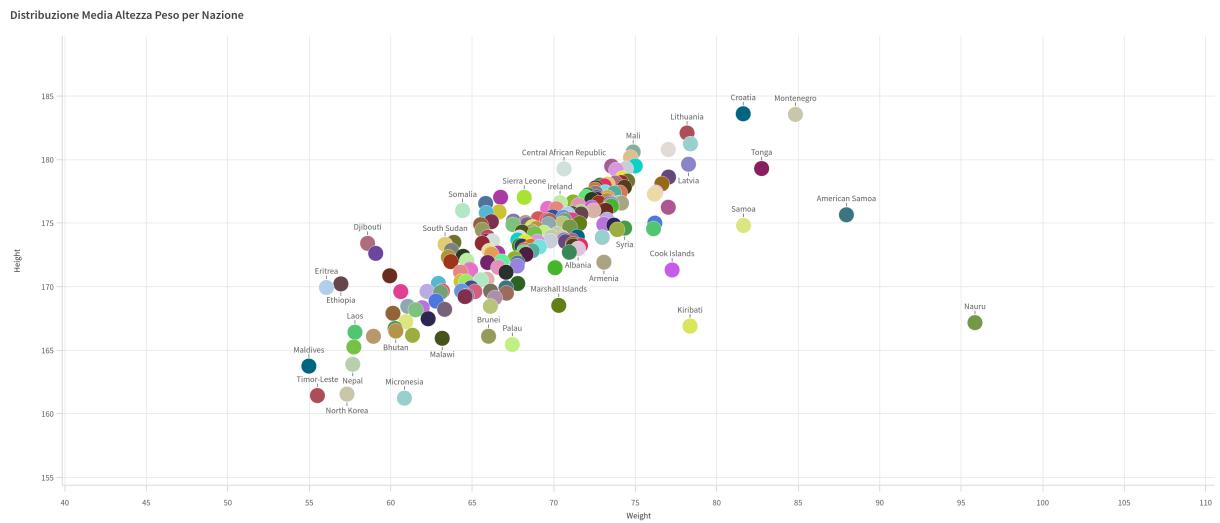
L'heatmap seguente mostra il numero di atleti partecipanti per ogni nazione. Il grafico è stato realizzato andando a definire come livello della mappa l'attributo **Country** e come misura l'aggregato dell'attributo **Name** calcolato tramite un *count distinct*.

N.B. Il numero di atleti partecipanti mostrato dalla mappa è statisticamente influenzato dal numero di abitanti di una nazione (maggiore infatti è la popolazione e maggiore sarà la probabilità che un atleta provenga da quella determinata nazione) e dal numero di volte in cui una determinata nazione ha partecipato alle Olimpiadi.



Distribuzione Peso e Altezza per nazione

Nell'ultimo grafico della dashboard è stato realizzato uno scatterplot il quale mostra le caratteristiche medie di altezza e peso degli atleti delle varie nazioni. Il grafico è stato realizzato usando come dimensione l'attributo **Country** e come misure sono stati scelti gli attributi **Weight** e **Height** calcolati tramite l'aggregato *mean* relativamente per l'asse delle ascisse e delle ordinate.



2.3.4 Analisi Eventi Sportivi

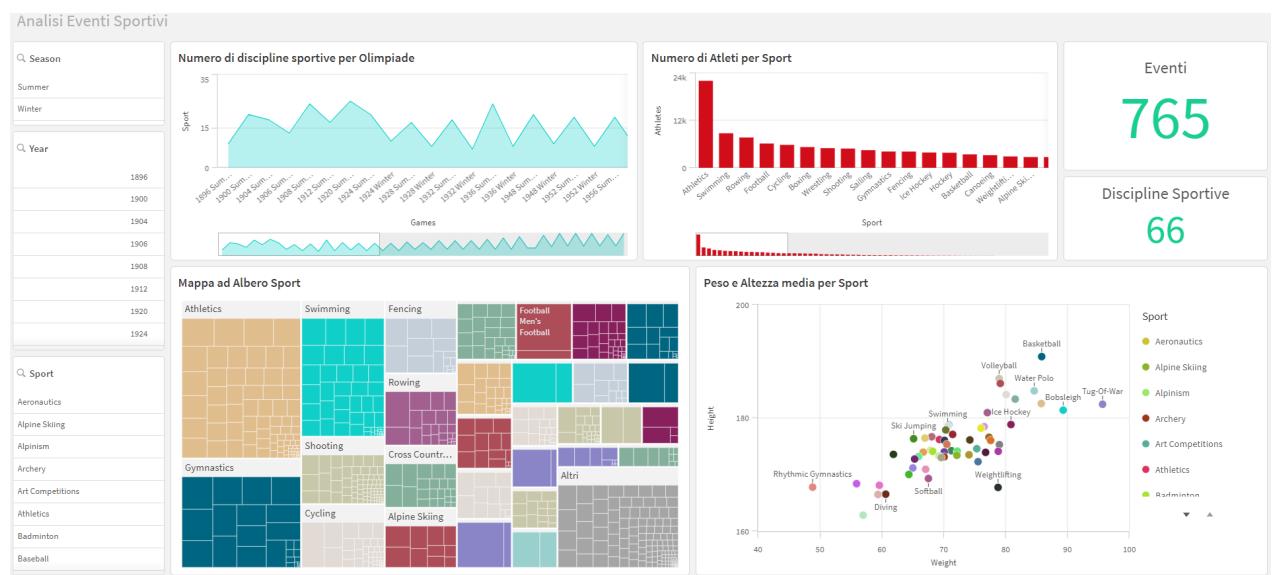


Figura 2.7: Dashboard sull'Analisi degli eventi sportivi

Nella quarta ed ultima dashboard è stata effettuata un'analisi dei diversi eventi sportivi definendo il numero di discipline nelle varie edizioni, il numero degli atleti e delle loro caratteristiche fisiche (altezza e peso) in relazione allo sport di appartenenza ed infine la distribuzione dei vari eventi con i relativi sport. Sulla sinistra sono stati ripresi tre filtri dalle altre dashboard i quali discriminano l'edizione ("Season"), l'anno di svogliamento ("Year") e la disciplina sportiva ("Sport"). La

dashboard contiene 4 grafici e 2 **KPI** i quali descrivono il numero di eventi e il numero di discipline sportive. Tali valori sono particolarmente utili per capire la numerosità di eventi e discipline quando si va a filtrare per anno o per sport.

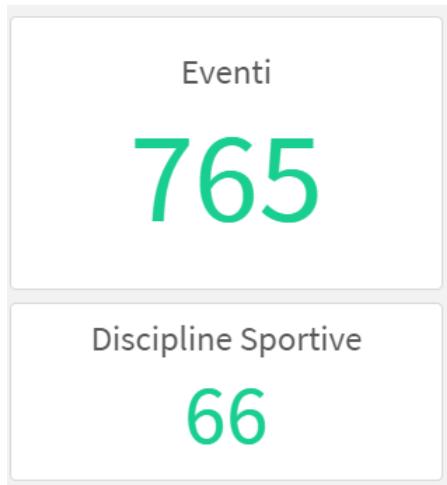
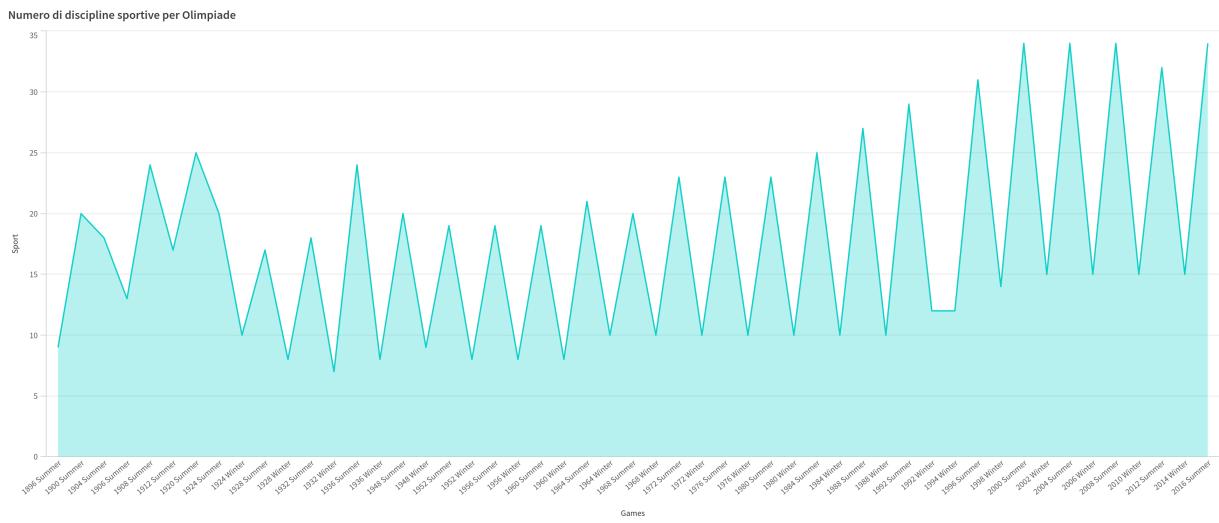


Figura 2.8: KPI relativi alle discipline e agli eventi

Numero di discipline sportive per Olimpiade

Il grafico a linea mostra il numero di discipline sportive per ogni edizione delle Olimpiadi ed è stato generato scegliendo come dimensione **Games** e come misura è stato aggregato il campo **Sport** con un *count distinct*. Come si può facilmente notare, il numero di discipline sportive è aumentato negli anni e risulta sempre esserci una maggiore numerosità nelle edizioni estive rispetto a quelle invernali. In particolare il numero di discipline nel caso delle Olimpiadi estive è passato da 9 a 34, mentre nel caso invernale è passato da 10 a 15 (dunque c'è stata complessivamente una maggiore crescita nell'edizione estiva).



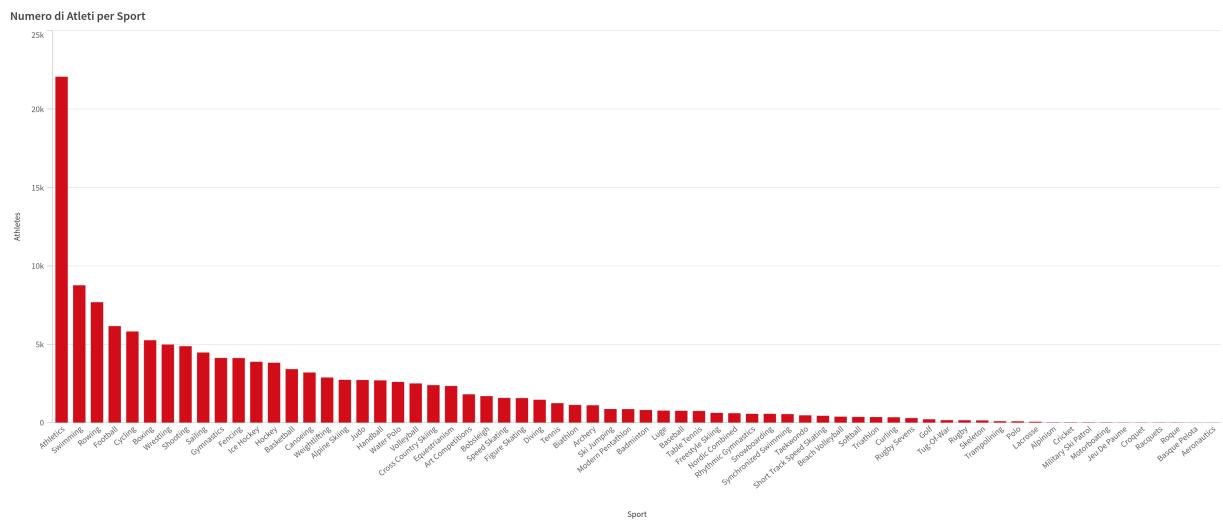
Mappa ad albero discipline sportive

In questa visualizzazione si è deciso di sfruttare il grafico ad albero messo a disposizione da Qlik per mostrare in che modo i vari eventi si distribuissero nelle varie discipline sportive. La mappa è stata generata scegliendo due dimensioni: il campo **Sport** è stato utilizzato per definire i gruppi della mappa, mentre il campo **Event** è stato utilizzato per definire i rettangoli. Per quanto riguarda la misura, era interessante visualizzare quale degli eventi di ogni disciplina sportiva fosse maggiormente numeroso all'interno della stessa, dunque è stato aggregato il campo **Event** tramite un *count*. Il grafico mostra chiaramente come l'atletica, la ginnastica e il nuoto siano caratterizzati da un numero elevato di eventi.



Numero di Atleti per disciplina sportiva

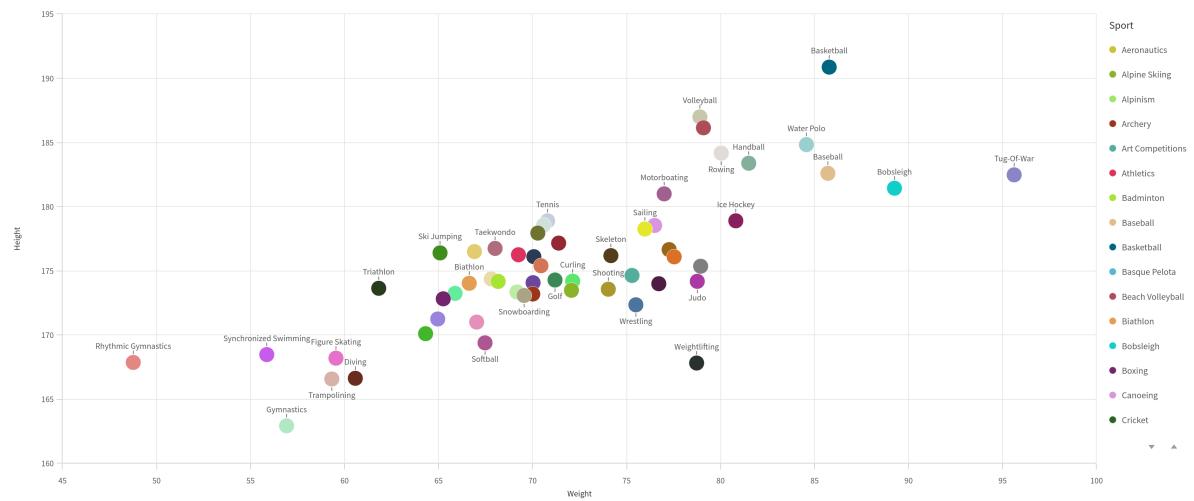
In questo istogramma sono visualizzati il numero di atleti che hanno partecipato per ogni disciplina sportiva. In congruenza ai grafici precedenti, l'atletica ed il nuoto risultano essere gli sport più numerosi grazie alla presenza di diversi eventi sportivi. Il grafico è stato generato scegliendo come dimensione il campo **Sport** e come misura è stato aggregato con un *count distinct* l'attributo **Name**.



Peso e Altezza media per disciplina sportiva

In quest'ultimo diagramma si è effettuata un'analisi del peso e dell'altezza media degli atleti in base allo sport di appartenenza. Il grafico a dispersione è stato generato scegliendo come dimensione il campo **Sport** e come misure sono stati aggregati con un *avg* i campi **Weight** e **Height**, rispettivamente per l'asse delle ascisse e delle ordinate. In particolare, il grafico mette a confronto le caratteristiche fisiche degli atleti in base alla propria disciplina: ad esempio, gli atleti della ginnastica artistica risultano essere molto più bassi e meno pesanti rispetto a quelli del basket.

Peso e Altezza media per Sport



Capitolo 3

Tableau

Tableau, proprio come Qlik, è un software di business intelligence (BI) molto diffuso ed utilizzato. Permette di realizzare analisi di tipo descrittivo e diagnostico tramite un'illustrazione dei dati facile ed intuitiva da comprendere. Inoltre, permette di svolgere analisi predittive, anche se solo parzialmente e con risultati molto basici ed elementari. Anche la realizzazione dei grafici contenuti negli Sheet, successivamente combinati per la realizzazione delle Dashboard, risulta facile. Anche Tableau, come Qlik, presenta la capacità di cambiare dinamicamente le visualizzazioni in base all'uso dei filtri e alle selezioni che vengono fatte dagli utenti. Per tutti questi motivi, Tableau risulta essere più performante e più diffuso di Qlik. La versione utilizzata in questo progetto è Tableau Desktop.

3.1 Caricamento Dati

In maniera analoga a come è avvenuto in Qlik, i due file *csv* sono stati caricati nell'applicazione di Tableau e poi successivamente associati tra loro utilizzando il campo comune **NOC** come campo di join.

⊕ dictionary.csv+ (Connessioni multiple)



Figura 3.1: Collegamento dati in Tableau

3.2 Analisi e creazione delle Storie

L'analisi degli eventi olimpici in Tableau è stata fatta andando a creare sia nuove visualizzazioni, sia riprendendo alcune visualizzazioni fatte in Qlik. Quest'ultime sono state migliorate ed aggiornate grazie all'uso degli strumenti messi a disposizione da Tableau. In particolare sono state realizzate tre storie composte da due dashboard ciascuna:

- **Analisi Generale:** Analisi generale degli eventi olimpici e degli atleti partecipanti;
- **Analisi Atletica:** Analisi della disciplina olimpica dell'atletica, che è risultata la più seguita e partecipata nelle analisi precedenti in Qlik;
- **Analisi Italia:** Analisi dell'Italia alle Olimpiadi e degli atleti italiani.

3.3 Analisi Generale

Nella prima storia è stata svolta un'analisi generale delle Olimpiadi, simile a quella svolta in Qlik, ma con nuove visualizzazioni. In particolare, le visualizzazioni nelle dashboard si concentrano principalmente sulle medaglie vinte e sulle partecipazioni agli eventi olimpici.

3.3.1 Prima dashboard

La prima dashboard si concentra sul fare una panoramica generale delle Olimpiadi e sulle medaglie vinte. In particolare, vengono mostrati i 10 atleti che hanno vinto il maggior numero di medaglie nella storia delle Olimpiadi e la distribuzione del numero di medaglie vinte rispetto all'età degli atleti. Inoltre, sono presenti due filtri che permettono di visualizzare i grafici selezionando il sesso degli atleti e/o un sottoinsieme delle varie edizioni delle Olimpiadi. Per quanto riguarda la panoramica generale delle Olimpiadi, la dashboard presenta 5 **KPI** che erano presenti anche nell'analisi in Qlik, ma rivisti in una grafica nuova che usa al massimo l'espressività di Tableau. Infatti, questi 5 **KPI** sono inseriti ognuno all'interno di uno dei 5 anelli che formano la celebre bandiera olimpica. Nel cerchio blu viene mostrato il numero di nazioni partecipanti alle Olimpiadi nel periodo 1896-2016; in quello giallo viene descritto il numero di discipline sportive olimpiche, sia invernali che estive, sempre nel medesimo periodo; in quello nero viene conteggiato il numero di medaglie distribuite nelle varie edizioni delle Olimpiadi; in quello verde viene rappresentato il numero totale di eventi sportivi olimpici, sia invernali che estivi, ed infine in quello rosso viene indicato il numero complessivo di atleti partecipanti alle Olimpiadi dal 1896 al 2016.

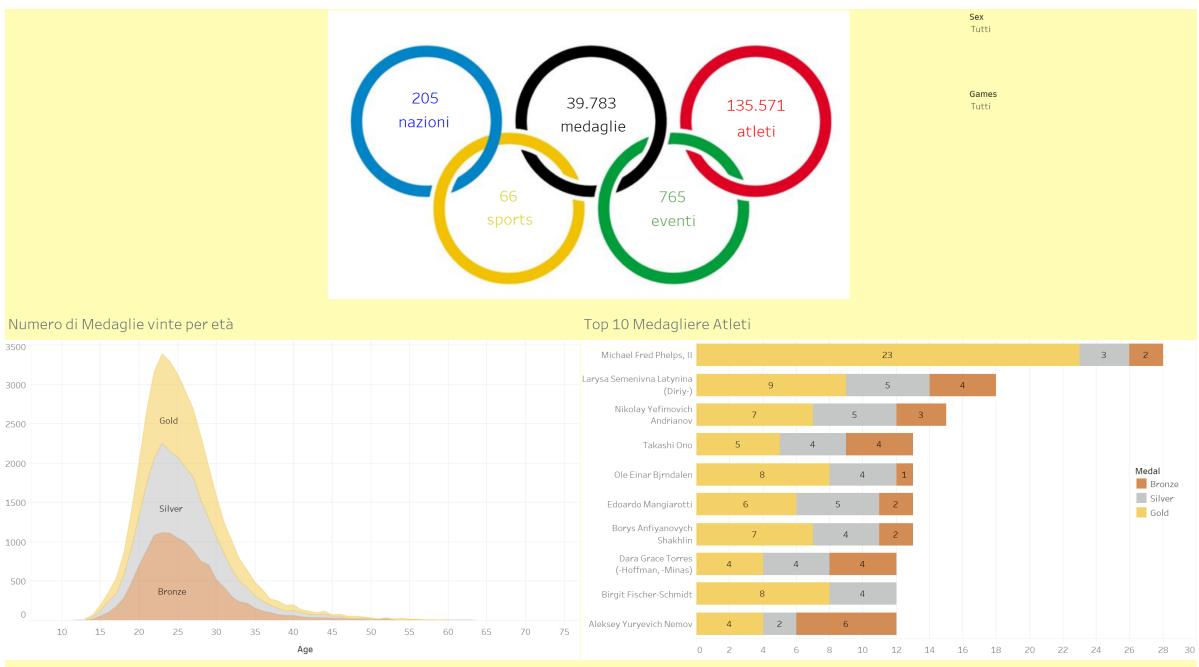
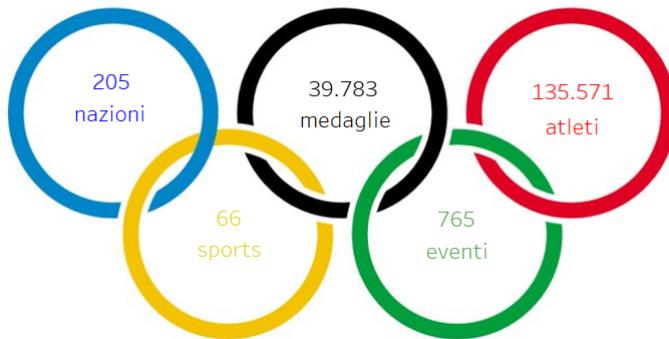


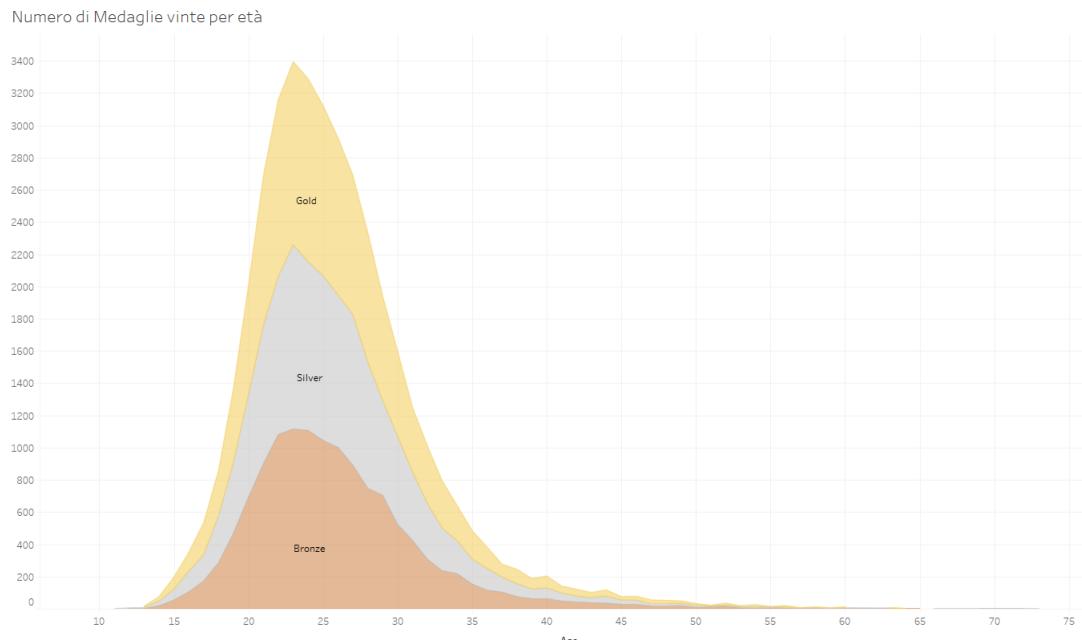
Figura 3.2: Prima dashboard generale sulle Olimpiadi



Distribuzione del numero di medaglie vinte per età

La prima visualizzazione mostra la distribuzione delle medaglie vinte rispetto all'età degli atleti. Come si può facilmente notare, il picco totale si ha per un'età pari a 23 anni, risultato che rispecchia perfettamente gli andamenti delle distribuzioni di tutte e tre le tipologie di medaglie (Oro, Argento e Bronzo), in cui si ha un picco sempre per quella particolare età. Questa visualizzazione è stata realizzata grazie alle funzionalità avanzate di Tableau, che permette di distinguere in una distribuzione le singole componenti tramite etichette e colori. In particolare,

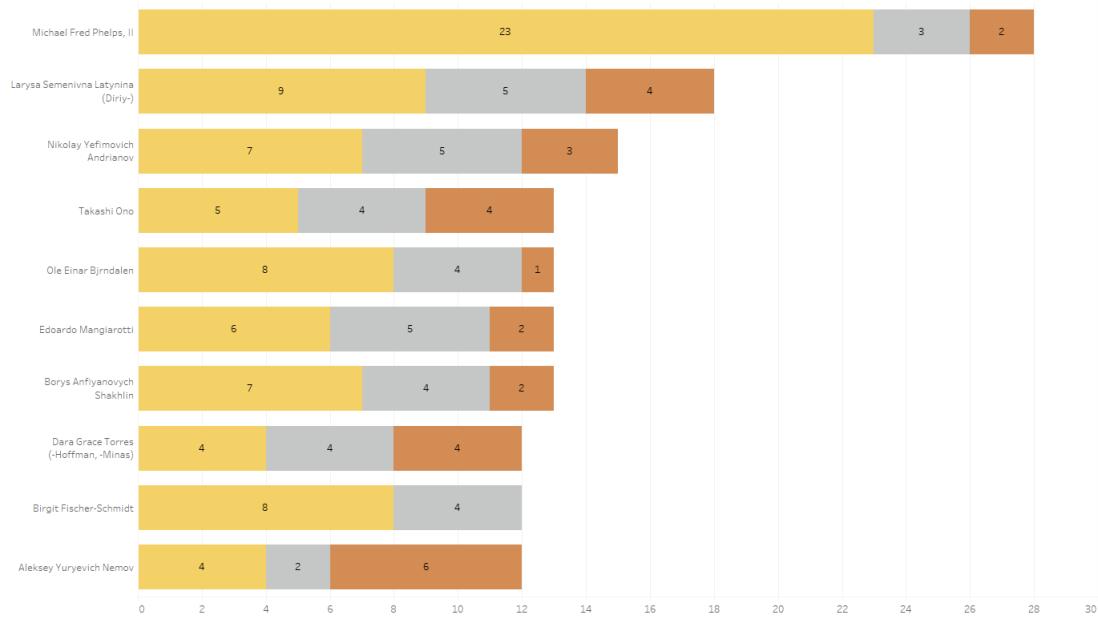
il grafico è stato realizzato scegliendo, appunto, come dimensione il campo **Age** e come misura il campo **Medal** (calcolato come conteggio). Inoltre, il campo **Medal** è stato utilizzato, come precedentemente accennato, anche come etichetta e nella selezione dei colori per distinguere, nella distribuzione totale, le componenti relative alla tipologia di medaglia.



Top 10 Medagliere Atleti

La seconda ed ultima visualizzazione della prima dashboard mostra la Top 10 degli atleti mondiali per numero di medaglie, al quale si possono poi applicare i filtri descritti precedentemente. Anche in questo caso, grazie agli strumenti messi a disposizione da Tableau, è possibile distinguere la tipologia di medaglia utilizzando il campo **Medal** come dimensione del colore. Nello specifico, il grafico a barre è stato realizzato scegliendo, invece, come dimensione il campo **Name** e come misura, anche in questo caso, il conteggio del campo **Medal**.

Top 10 Medagliere Atleti



3.3.2 Seconda dashboard

La seconda dashboard dell'analisi generale conclude il focus sulle medaglie per poi concentrarsi sulle partecipazioni agli eventi olimpici delle varie nazioni e delle atlete femminili. Anche qui, come nella dashboard precedente, sono presenti due filtri che, però, permettono di visualizzare i grafici andando a selezionare un sottoinsieme degli anni in cui si sono svolte le Olimpiadi e/o un sottinsieme delle nazioni che complessivamente hanno partecipato ai giochi olimpici.

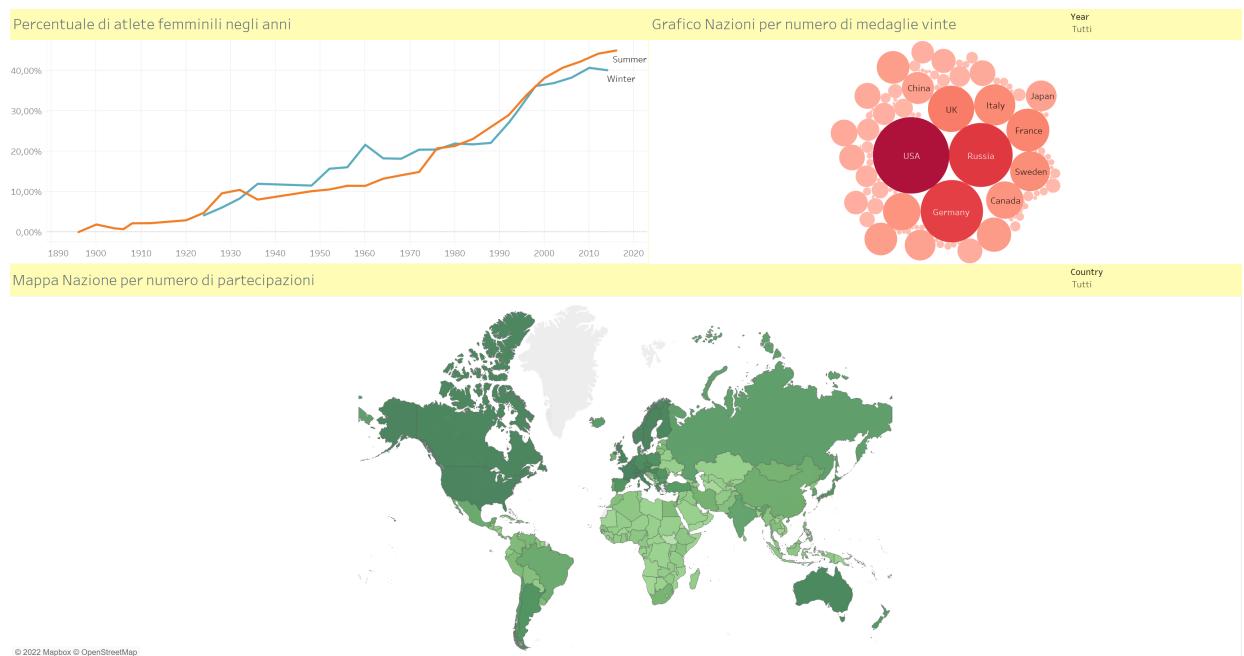
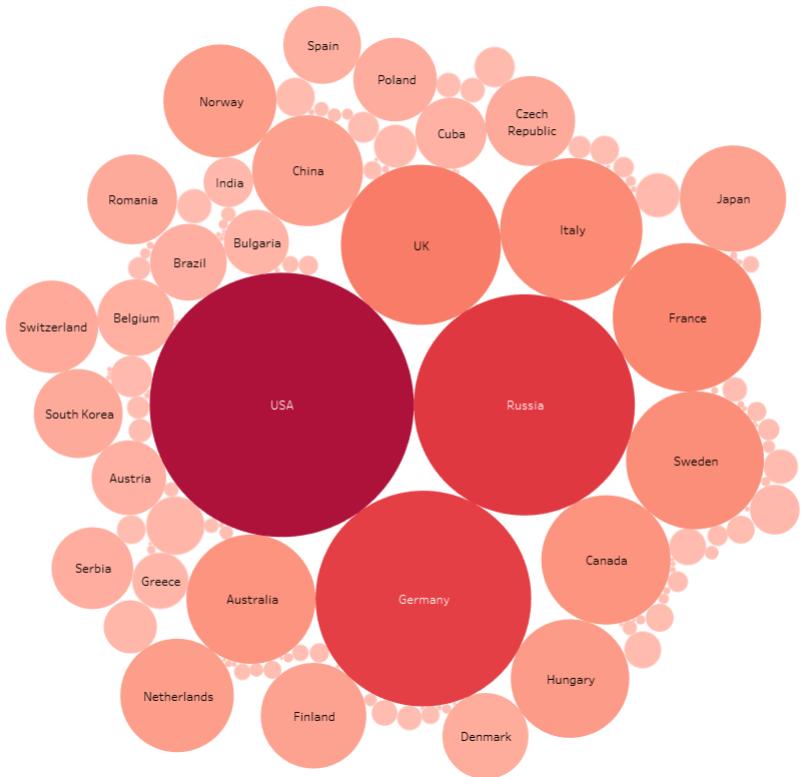


Figura 3.3: Seconda dashboard generale sulle Olimpiadi

Numero di Medaglie vinte per Nazione

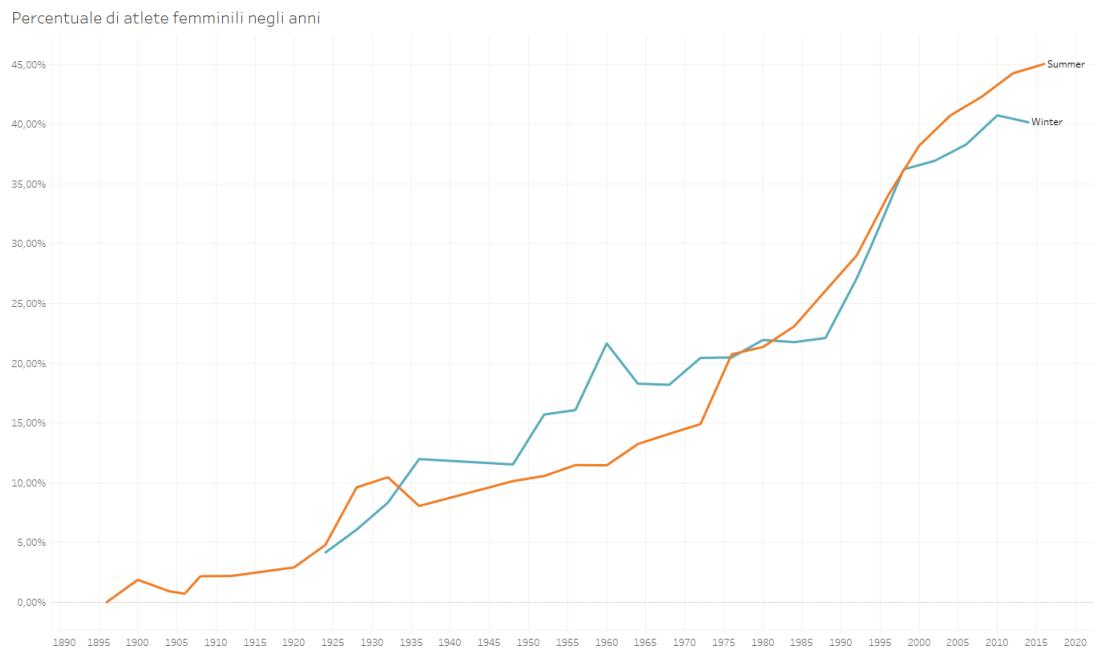
La prima visualizzazione della seconda dashboard mostra, tramite l'uso di un bubble chart, il numero di medaglie vinte dalle varie nazioni. Andando nello specifico, il grafico è stato ottenuto scegliendo per le bolle e come colore la misura del conteggio del campo **Medal**, mentre per le etichette la dimensione scelta è stata ovviamente **Country**. La visualizzazione mostra chiaramente e senza ombra di dubbio che gli Stati Uniti, la Russia e la Germania sono le nazioni che hanno vinto complessivamente, dal 1896 al 2016, il maggior numero di medaglie nelle varie edizioni delle Olimpiadi, tenendo in considerazione sia quelle invernali che estive.

Grafico Nazioni per numero di medaglie vinte



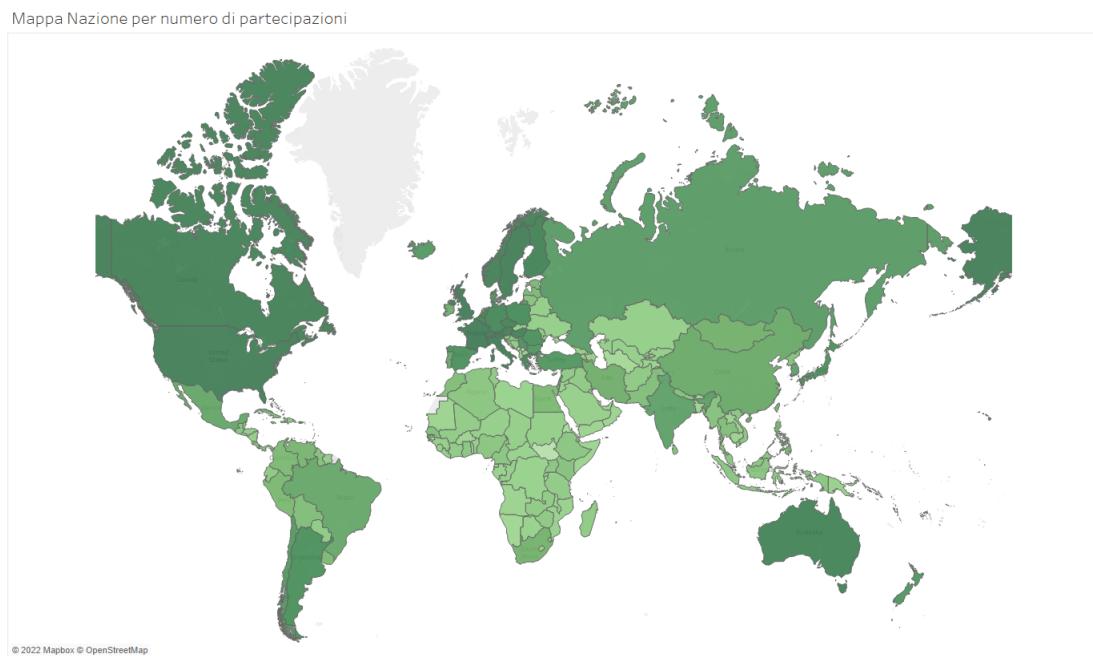
Percentuale di partecipazione delle atlete femminili negli anni

In questa visualizzazione viene mostrato l'andamento della percentuale di partecipazione alle Olimpiadi delle atlete femminili negli anni dal 1896 al 2016. In particolare, il grafico a linee è stato realizzato scegliendo come dimensione il campo **Year** e come misura la percentuale delle atlete femminili. Questa misura è stata ricavata facendo un *count distinct* sul campo **ID**, mettendo come condizione che il campo **Sex** fosse uguale a 'F', per poi dividere il risultato ottenuto per il numero di atleti totali, ricavabile sempre tramite un *count distinct* sul campo **ID**, ma in questo caso senza alcuna condizione. Inoltre, è stato utilizzato il campo **Season** nella selezione dei colori e come etichetta, così da distinguere l'andamento della percentuale di atlete femminili tra Olimpiadi estive e invernali. La visualizzazione mostra come l'andamento della percentuale di partecipazione delle atlete femminili sia andato sempre più a crescere negli anni arrivando nel 2016, ultimo anno preso in esame nella nostra analisi, al 45% nelle Olimpiadi estive e al 40% in quelle invernali.



Numero di partecipazioni alle Olimpiadi per Nazione

L'ultima visualizzazione della seconda dashboard è simile a un grafico che era stato già ricavato in Qlik, ossia l'heatmap con il numero di atleti partecipanti per nazione. In questo caso, invece, l'heatmap analizza il numero di partecipazioni alle Olimpiadi per nazione, ossia a quante edizioni delle Olimpiadi, sia estive che invernali, ha partecipato ogni nazione. Nel dettaglio, il grafico è stato realizzato andando a definire come livello della mappa il campo **Country** e come misura l'aggregato sul campo **Games**, calcolato tramite un semplice *count*.



3.4 Analisi Atletica

Nella seconda storia ci si è concentrati sull'analisi della disciplina olimpica con più eventi in assoluto, ossia l'Atletica (quindi tutte le visualizzazioni sono state realizzate usando un filtro sul campo *Sport*, assegnandogli il valore *Athletics*). Le dashboard si focalizzano principalmente sugli eventi e le medaglie di questa grande disciplina sportiva, oltre che sulle caratteristiche fisiche delle atlete femminili che la praticano.

3.4.1 Distribuzione medaglie ed eventi

La prima dashboard mostra la distribuzione degli eventi sportivi e delle medaglie sportive all'interno della disciplina olimpica dell'Atletica. Anche in questo caso sono presenti dei filtri che permettono di cambiare le visualizzazioni all'interno della dashboard. In particolare, i filtri danno la possibilità di selezionare un sottoinsieme delle varie edizioni delle Olimpiadi e/o un sottoinsieme dei numerosi eventi appartenenti alla disciplina olimpica dell'Atletica. In questa storia, poi, a differenza delle altre, non sono presenti dei **KPI** all'interno delle dashboard che la compongono.

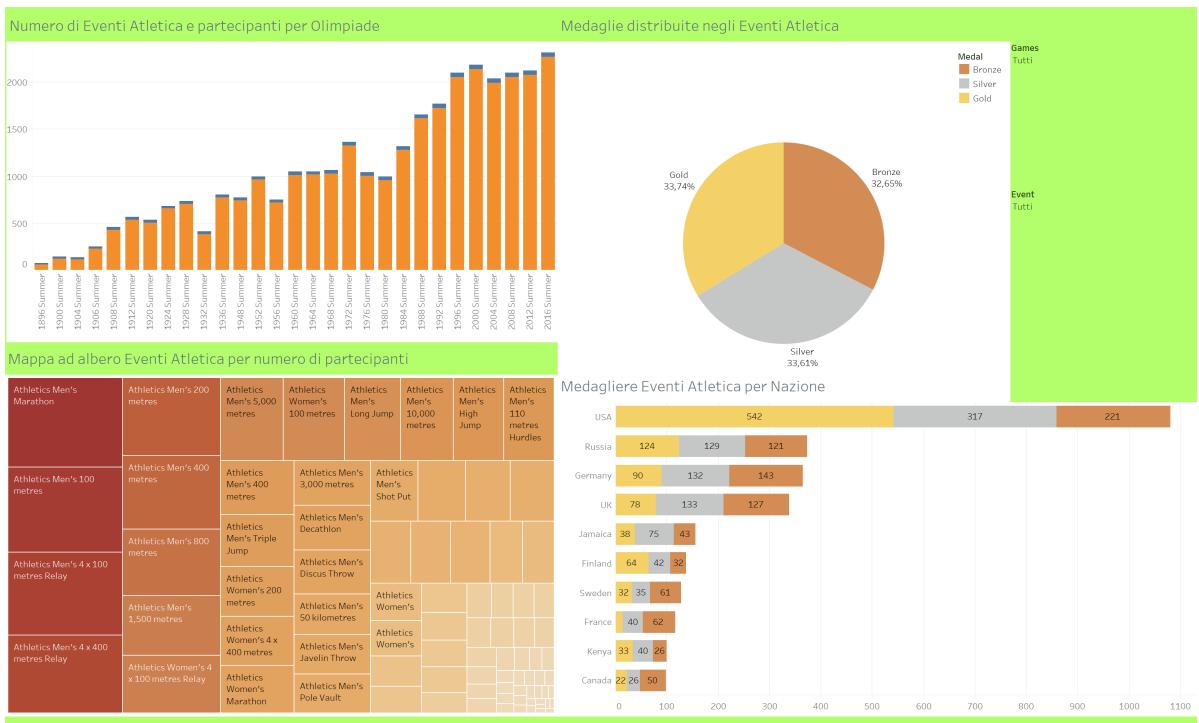
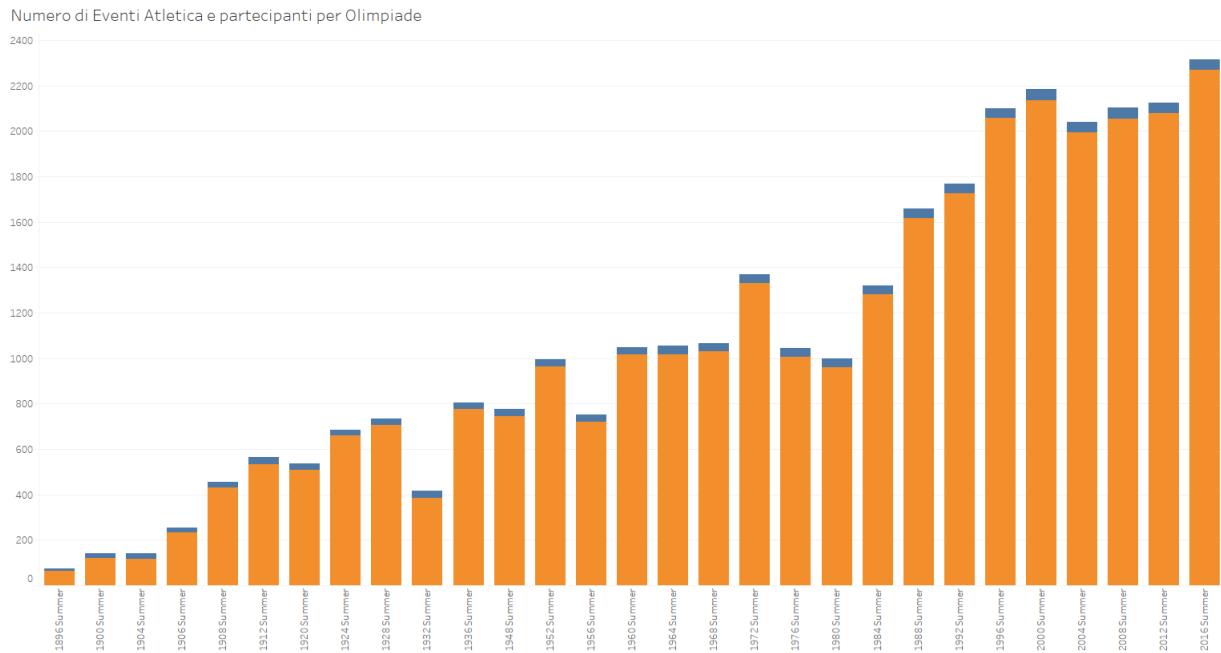


Figura 3.4: Dashboard generale sugli Eventi Atletica

Numero di partecipanti ed eventi di Atletica per Olimpiade

La prima visualizzazione mostra, tramite l'utilizzo di un grafico a barre in pila, la distribuzione dei partecipanti e degli eventi di Atletica per edizione delle Olimpiadi. Come è facile notare, l'andamento del grafico è crescente negli anni, anche se non in modo uniforme, sia per il numero di partecipanti che per quello degli eventi sportivi. Andando nel dettaglio, il grafico è stato realizzato scegliendo come dimensione il campo **Games**, ossia il campo che esprime le edizioni delle Olimpiadi, e come misure i conteggi univoci effettuati sui due campi **Event** e **Name**.

N.B. Avendo applicato a tutte le visualizzazioni della dashboard il filtro *Sport* con valore *Athletics*, ovviamente i partecipanti e gli eventi mostrati nei dati sono quelli della disciplina dell'Atletica. Quindi, questo comporta che nel grafico le colonne, che esprimono la dimensione (**Games**), abbiano come valori soltanto le edizioni estive delle Olimpiadi, perché l'Atletica è una disciplina che non è presente nelle Olimpiadi invernali, ma solo in quelle estive.



Mappa ad albero degli eventi di Atletica

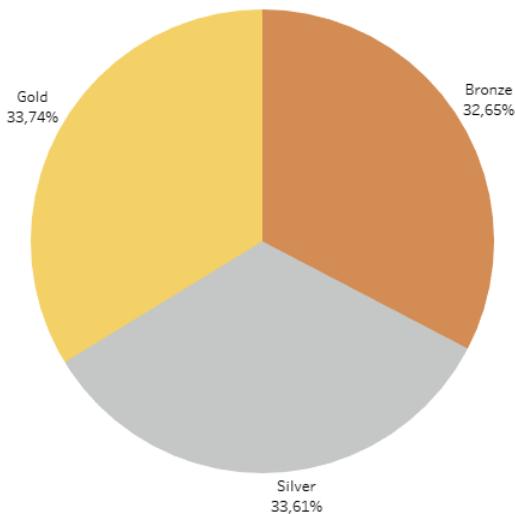
Questa visualizzazione riprende un grafico che era stato precedentemente realizzato in Qlik, ovvero la mappa ad albero delle discipline sportive per numero di eventi olimpici associati ad ognuna. In questo caso, invece, la mappa ad albero esegue un focus su una specifica disciplina sportiva olimpica, l'Atletica, e su di questa esegue un'analisi dei vari eventi che la compongono raggruppandoli in base al numero di atleti partecipanti. In particolare, il grafico è stato generato scegliendo un'unica dimensione, il campo **Event**, mentre come misura il conteggio univoco sul campo **Name**, che identifica il numero degli atleti partecipanti. La mappa mostra chiaramente come gli eventi di atletica con il maggior numero di partecipanti siano la maratona maschile, i 100 metri piani maschili e la staffetta 4x100 metri maschile.

Mappa ad albero Eventi Atletica per numero di partecipanti

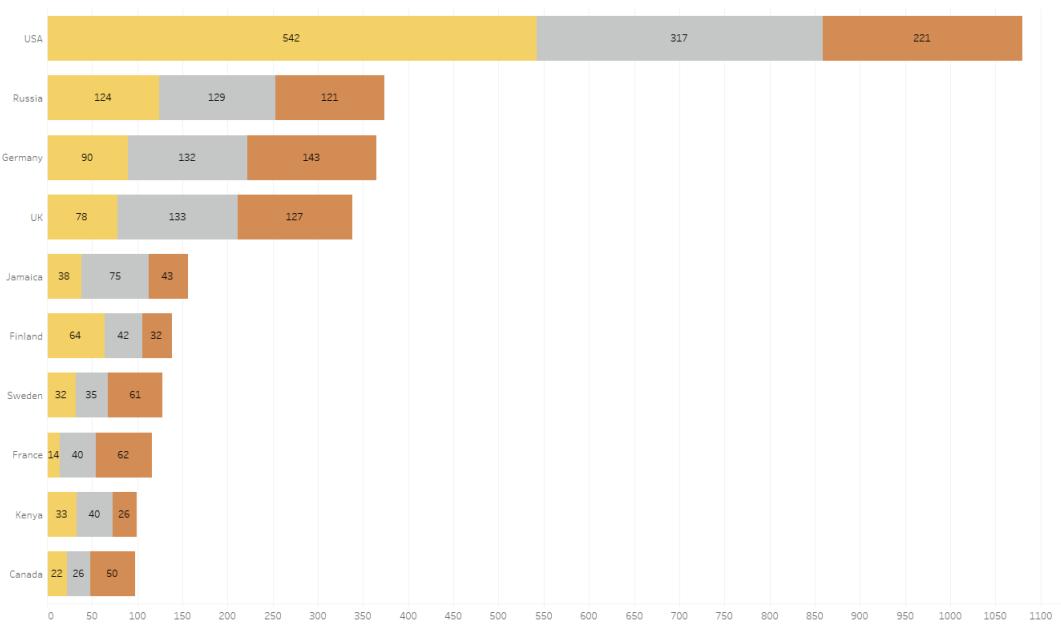


Medaglie negli eventi di Atletica

Le ultime due visualizzazioni di questa dashboard mostrano rispettivamente la distribuzione delle medaglie negli eventi di Atletica tramite un grafico a torta e il medagliere delle nazioni che hanno vinto il maggior numero di medaglie in questa disciplina attraverso un grafico a barre. Il diagramma a torta mostra un certo equilibrio nella distribuzione di medaglie vinte in questa disciplina olimpica, con una percentuale leggermente inferiore di medaglie di bronzo. La seconda visualizzazione, invece, che rappresenta la Top 10 delle nazioni con il maggior numero di medaglie vinte negli eventi di Atletica non è affatto equilibrata. Infatti, come è possibile vedere, gli Stati Uniti primeggiano incontrastati la classifica avendo quasi il triplo delle medaglie della Russia, che si trova al secondo posto. Grazie alle funzionalità avanzate di Tableau, anche per questa visualizzazione è possibile distinguere nella classifica la tipologia di medaglie utilizzando il campo **Medal** come dimensione del colore. Andando nello specifico, il grafico a barre è stato generato scegliendo come dimensione il campo **Country** e come misura ovviamente il conteggio del campo **Medal**. Inoltre, il conteggio delle medaglie è stato usato anche come etichetta, così da visualizzare il numero di medaglie totali per tipologia vinte da ogni nazione.



Medagliere Eventi Atletica per Nazione



3.4.2 Caratteristiche fisiche delle atlete femminili

La seconda dashboard illustra le caratteristiche fisiche delle atlete, suddivise per i vari eventi sportivi femminili di Atletica (quindi entrambe le visualizzazioni sono state generate usando un filtro sul campo *Sex* degli atleti, così da prendere solo quelle femminili). In questo caso, la dashboard non presenta **KPI**, ma soltanto le due visualizzazioni con i rispettivi e vari box plot delle atlete. Queste sono state realizzate scegliendo come dimensione il campo **Event** nelle righe e rispettivamente **Height** e **Weight** nelle colonne, ossia le due diverse caratteristiche fisiche che si vogliono analizzare nelle due visualizzazioni. In questo modo, per ogni evento femminile di Atletica, viene generato un box plot il quale mostra, sia la mediana del valore preso in considerazione per quello specifico evento sportivo femminile, sia i valori dei baffi e dei quartili superiore ed inferiore.



Figura 3.5: Dashboard Box Plot Atlete femminili negli Eventi Atletica

3.5 Analisi Italia

Nella terza storia si è effettuata un' analisi dell'Italia alle Olimpiadi (dunque tutte le visualizzazioni sono state realizzate utilizzando un filtro per il campo *Country*). Le dashboard si concentrano in particolare sulle medaglie e gli sport dei vari atleti italiani e sulle loro caratteristiche fisiche ed anagrafiche.

3.5.1 Distribuzione medaglie e sport

La prima dashboard mostra la distribuzione delle medaglie e degli sport per gli atleti italiani. Sono presenti due filtri che permettono di visualizzare i grafici andando a selezionare atleti maschili/femminili e/o un sottoinsieme di sport. Inoltre, la dashboard presenta un **KPI** che mostra il numero di medaglie vinte dagli atleti italiani nelle diverse edizioni delle Olimpiadi. In particolare, come si può notare, le medaglie sono state maggiormente aggiudicate nelle edizioni estive, grazie alla presenza di un maggior numero di eventi sportivi e quindi, conseguentemente, di un maggior numero di partecipazioni.

N.B. Il numero segnalato dal **KPI** per quanto riguarda le medaglie vinte non coincide con il medagliere ufficiale, in quanto, in questo caso, le medaglie per gli sport di squadra sono considerate più volte (una per ogni atleta componente), rispetto al medagliere ufficiale che considera le medaglie di squadra vinte come uniche.

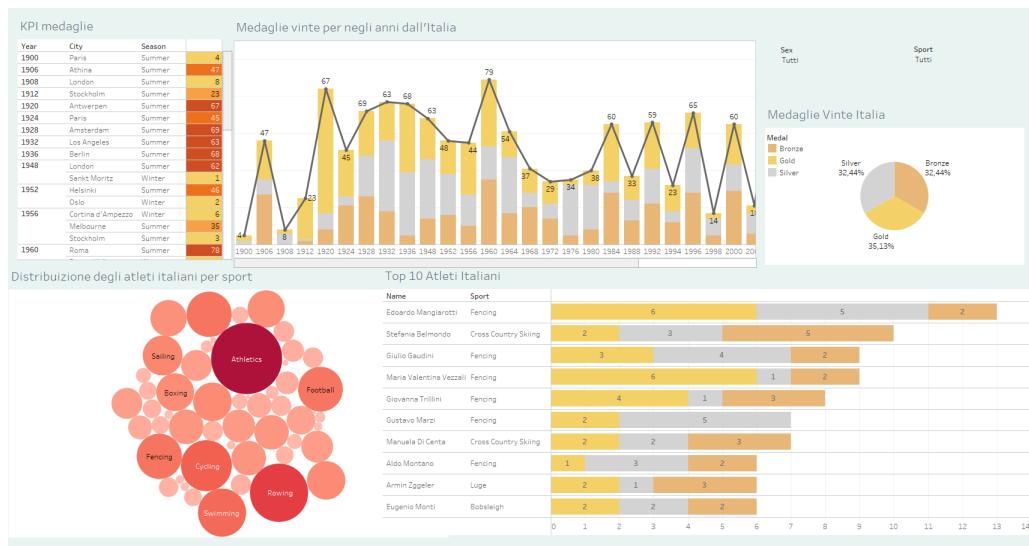


Figura 3.6: Dashboard distribuzione medaglie e sport

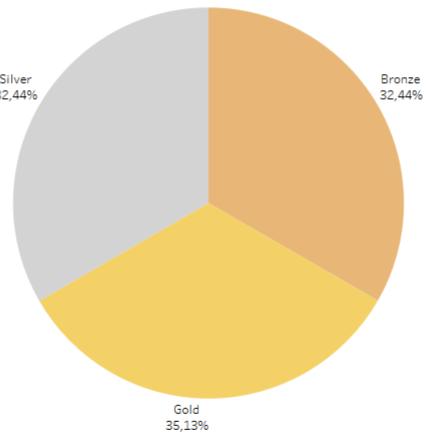
KPI medaglie

Year	City	Season	
1900	Paris	Summer	4
1906	Athina	Summer	47
1908	London	Summer	8
1912	Stockholm	Summer	23
1920	Antwerpen	Summer	67
1924	Paris	Summer	45
1928	Amsterdam	Summer	69
1932	Los Angeles	Summer	63
1936	Berlin	Summer	68
1948	London	Summer	62
	Sankt Moritz	Winter	1
1952	Helsinki	Summer	46
	Oslo	Winter	2
1956	Cortina d'Ampezzo	Winter	6
	Melbourne	Summer	35
	Stockholm	Summer	3
1960	Roma	Summer	78
	Squaw Valley	Winter	1
1964	Innsbruck	Winter	8
	Tokyo	Summer	46
1968	Grenoble	Winter	6
	Mexico City	Summer	31
1972	Munich	Summer	21
	Sapporo	Winter	8
1976	Innsbruck	Winter	4
	Montreal	Summer	30
1980	Lake Placid	Winter	3
	Moskva	Summer	35
1984	Los Angeles	Summer	58
	Sarajevo	Winter	2
1988	Calgary	Winter	6
	Seoul	Summer	27
1992	Albertville	Winter	15
	Barcelona	Summer	44
1994	Lillehammer	Winter	23
1996	Atlanta	Summer	65
1998	Nagano	Winter	14
2000	Sydney	Summer	60
2002	Salt Lake City	Winter	18
2004	Athina	Summer	101
2006	Torino	Winter	21
2008	Beijing	Summer	39
2010	Vancouver	Winter	5
2012	London	Summer	63
2014	Sochi	Winter	11
2016	Rio de Janeiro	Summer	69

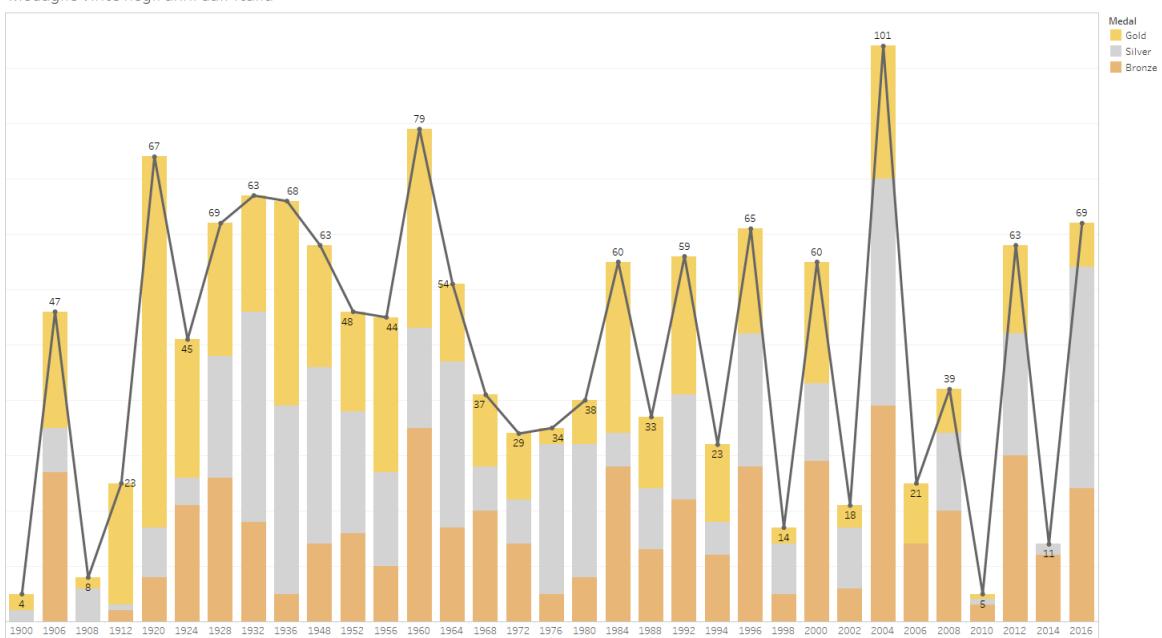
Medaglie Vinte Italia

Le prime due visualizzazioni mostrano la distribuzione delle medaglie vinte attraverso un grafico a torta e l'andamento temporale delle medaglie vinte. Il diagramma a torta mostra un certo equilibrio nella distribuzione delle medaglie vinte, con una vittoria leggermente maggiore di medaglie d'oro. La seconda visualizzazione, invece, è stata realizzata grazie alle funzionalità messe a disposizione da Tableau, che permette di unire un grafo a barre e uno a linea. In particolare, il grafico è stato realizzato scegliendo come dimensione il campo **Year** e come misura il campo **Medal** (calcolato come conteggio), il quale è stato anche utilizzato nella sezione colori per distinguere il tipo di medaglia. Inoltre, il conteggio delle medaglie è stato utilizzato anche come etichetta, in modo da visualizzare il numero totale di medaglie vinte in un determinato anno nel grafo a linea.

Medaglie Vinte Italia



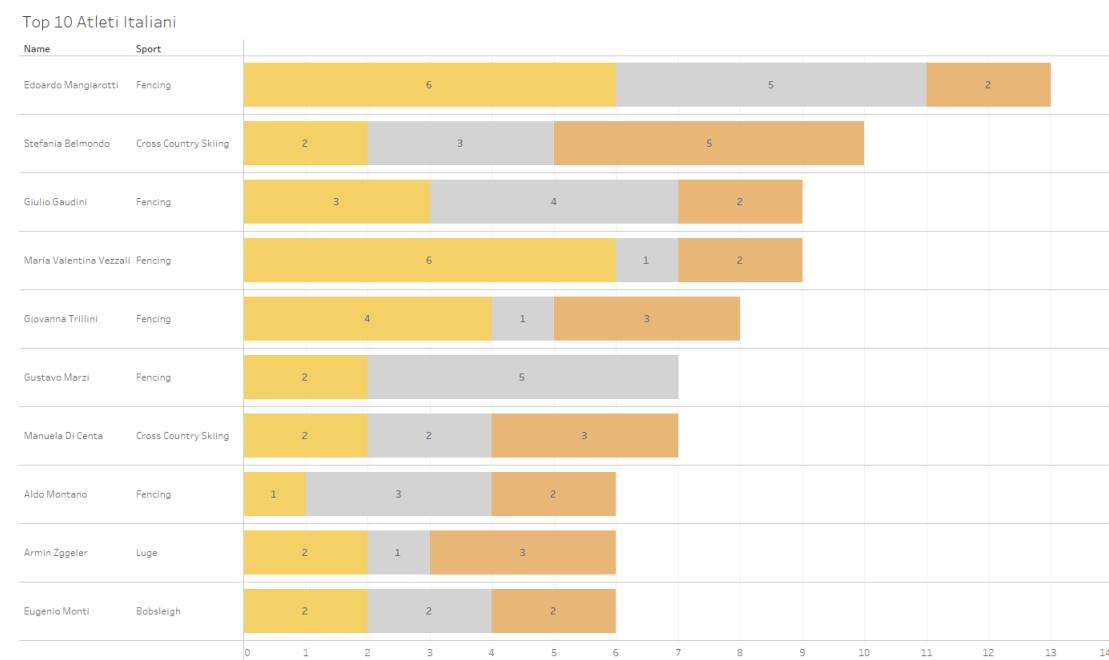
Medaglie vinte negli anni dall'Italia



Top 10 Atleti Italiani

La visualizzazione mostra la Top 10 degli atleti italiani per numero di medaglie e il loro sport di riferimento. Anche in questo caso, grazie agli strumenti messi a disposizione da Tableau, è stato possibile distinguere la tipologia di medaglia usando il campo **Medal** come colore. In particolare, il grafo a barre è stato realizzato scegliendo come dimensione il campo **Name** e come misura il conteggio di **Medal**.

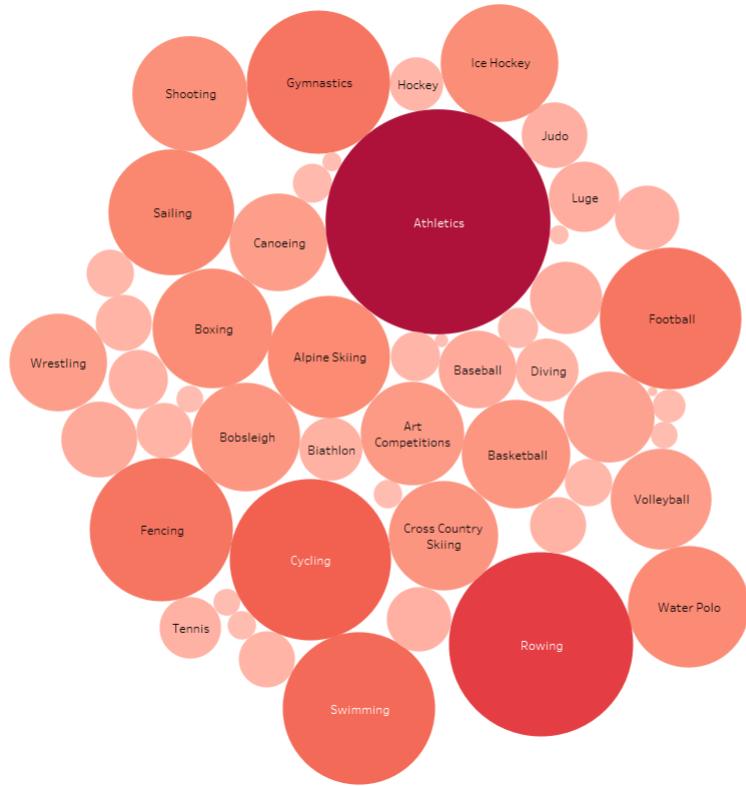
N.B. Grazie all'uso del filtro *Sport* è possibile visualizzare la Top 10 degli atleti italiani in una particolare disciplina sportiva, se presenti; altrimenti, se non ne vengono trovati almeno dieci, verrà visualizzato un numero minore di atleti.



Numero di Atleti per Sport

L'ultima visualizzazione della dashboard mostra, utilizzando un bubble chart, il numero di atleti italiani che hanno partecipato alle Olimpiadi, divisi per disciplina sportiva. Il numero di partecipanti è stato caratterizzato scegliendo come colore la misura del conteggio del campo **Name**, mentre per le bolle è stata scelta la dimensione **Sport** come anche per l'etichetta. La visualizzazione mostra come l'atletica, il canottaggio, il nuoto, il ciclismo, la ginnastica, il calcio e la scherma siano le discipline maggiormente numerose, che infatti sono le specialità sportive caratteristiche del nostro paese.

Distribuzione degli atleti italiani per sport

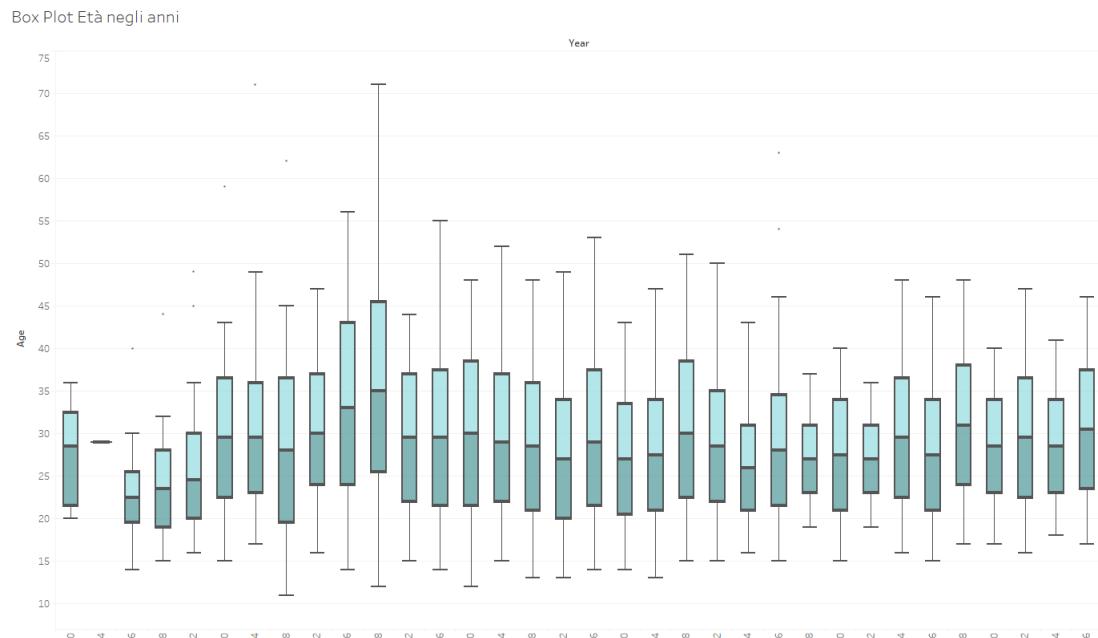


3.5.2 Caratteristiche Atleti

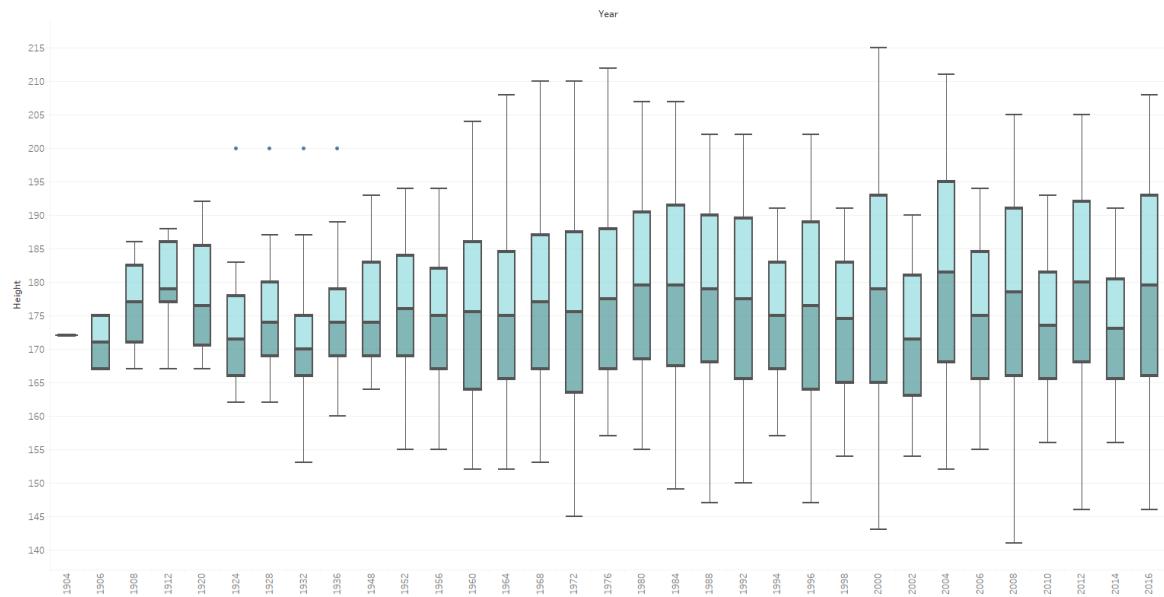
La seconda dashboard realizzata mostra la distribuzione delle caratteristiche fisiche ed anagrafiche negli anni degli atleti italiani che hanno partecipato alle Olimpiadi. La dashboard presenta un filtro per modificare le visualizzazioni dei box plot andando a selezionare valori specifici del campo **Year**, ed un **KPI** il quale mostra l'altezza, il peso e l'età media degli atleti nel periodo considerato dal filtro. Le tre diverse visualizzazioni temporali sono state realizzate scegliendo come dimensione il campo **Year** nelle righe e nelle colonne la relativa caratteristica che si vuole analizzare. In questo modo, per ogni anno, viene generato un box plot il quale mostra, sia la mediana del valore preso in considerazione in quello specifico anno, sia i valori dei baffi e dei quartili superiore ed inferiore.



Figura 3.7: Dashboard caratteristiche atleti



Box Plot Altezza negli anni



Box Plot Peso negli anni

