



Data Visualization

AA 2021/2022

Potertì Daniele 844892, DS

Sanvito Alessio 844785, DS

Sanvito Simone 844794, DS

Indice

Introduzione	2
Descrizione del database	3
Kaggle	3
Web scraping	4
Realizzazione dell'infografica	5
Obiettivo	5
Target	5
Presentazione delle infografiche finali	6
Scatterplot con istogrammi marginali (visualizzazione interattiva)	6
Lollipop	7
Linechart	9
Alluvial plot (visualizzazione interattiva)	10
Valutazione dell'infografica	12
Valutazione euristica	12
Test utente	15
Questionario	18
Conclusioni e sviluppi futuri	23

Introduzione

L'automobilismo agonistico rappresenta uno degli sport più spettacolari dell'epoca moderna. In particolare, la Formula 1 rappresenta una delle maggiori corse che raduna, per i vari eventi nei circuiti intorno al mondo, centinaia di migliaia di tifosi e tiene incollati allo schermo della tv milioni di tifosi e/o appassionati.

Nelle varie epoche ci sono stati duelli leggendari, piloti che hanno ispirato generazioni, gare che hanno fatto la storia di questo sport. La contesa forse più gettonata, quando si pensa ai vari piloti, è quella per decidere **chi sia il pilota più forte della storia della Formula 1**.

Quando si stila una classifica, c'è sempre una sottile linea che divide soggettività e oggettività e spesso si finisce per non accontentare qualcuno. L'automobilismo, forse più che gli altri sport, genera disorientamento sulla percezione del talento dei piloti, perché sono diverse le condizioni di partenza e gli strumenti a disposizione di ognuno per arrivare al successo. Trovare una soluzione a questa contesa è forse impossibile perché, mai come nella Formula 1, paragonare le varie epoche è esercizio davvero difficile: le variabili sono molteplici, dal numero di Gran Premi a stagione, i motori, le piste, il sistema di punteggio, i regolamenti, le auto più difficili da portare al limite, eccetera, eccetera.

La cosa difficile nel confrontare tutti i conducenti (di epoche diverse) è il fatto che la sicurezza generale dei conducenti è cambiata molto nelle corse ed è un fattore determinante per il successo totale di un pilota. All'inizio molti piloti sono morti prima di poter diventare campioni del mondo (o di riconfermarsi più volte come campioni) o addirittura prima di salire sul podio per la prima volta.

Dunque questa è la nostra research question iniziale, da cui è partita tutta l'analisi e la visualizzazione dei dati: *"Chi è il miglior pilota di Formula 1 di tutti i tempi?"*.

In una prima fase sono stati, quindi, raccolti dati, analizzato il dataset e integrato con altre fonti di dati utili; successivamente sono state progettate e realizzate diverse visualizzazioni che rispondessero ognuna ad una domanda specifica; infine, è stata fatta una valutazione di qualità del lavoro svolto. Sono stati effettuati una valutazione euristica che evidenziasse eventuali problematiche, un test utente che permettesse di valutare efficacia ed efficienza e un questionario per valutare la soddisfazione generale dell'utente riguardo alle infografiche.

1. Descrizione del database

Le fonti di dati che sono state utilizzate per la realizzazione di questo progetto sono principalmente due:

1.1. Kaggle

All'interno di Kaggle, una piattaforma che consente di trovare e pubblicare dati e di usarli in diversi ambiti per costruire modelli, è stato selezionato un dataset

(<https://www.kaggle.com/rohanrao/formula-1-world-championship-1950-2020>) molto utile in quanto presenta molte informazioni in diversi file:

- circuits → circuiti dove sono state svolte delle gare di F1; ogni pista ha associato un id (circuitId) con il suo nome, la località, le coordinate geografiche e un link alla pagina wikipedia del circuito;
- constructor_results → numero dei punti fatti per ogni team in ogni gara;
- constructor_standings → classifica dei costruttori per ogni anno con il relativo numero di punti e la posizione relativa a quell'anno;
- constructors → elenco dei team nella storia della formula 1; di ogni team viene specificato id, nome, nazionalità e indirizzo alla pagina wikipedia;
- driver_standings → elenco dei punti ottenuti ogni anno da ogni pilota, quindi ogni pilota avrà un numero di righe pari al numero di gare corse in ogni campionato con il relativo numero di punti ottenuto, il piazzamento in gara e nella classifica finale dei piloti;
- drivers → elenco dei piloti che hanno corso in formula 1; ogni pilota ha associato il suo id, oltre a nome, cognome, data di nascita, nazionalità, numero;
- lap_times → tempi sul giro di ogni gara di ogni pilota;
- pit_stops → analisi sui pitstop effettuati; contiene l'id della gara, l'id del pilota; che ha effettuato il pitstop, il numero di pitstops nella gara, il numero del giro a cui si è fermato il pilota, l'orario a cui è stato effettuato e la durata (anche in millisecondi);
- qualifying → descrizione di tutte le sessioni di qualifica dal 2003 in poi; comprende id della gara, id del pilota, id del team, numero della macchina, la posizione ottenuta e i tempi migliori ottenuti nelle sessioni di q1, q2 e q3;
- races → informazioni sulle gare effettuate, contiene id della gara, anno, numero della gara in quell'anno, nome della gara, data e ora di inizio della gara;
- results → contiene id della gara, id del pilota, id del team, numero della gara, numero della macchina, posizione in griglia di partenza, posizione finale, punti ottenuti, numero di giri effettuati in gara, numero del giro più veloce in gara (con il tempo associato);

- seasons → elenco delle stagioni;
- status → ogni numero associato a una condizione diversa (ad esempio "finished" indica che il pilota ha finito la gara, "Disqualified" indica che il pilota è stato squalificato, "+1 Lap" indica che il pilota è stato doppiato ecc).

I dati presenti in questi file sono stati raccolti da <http://ergast.com/mrd/>.

Per la nostra analisi sono stati usati principalmente i dati presenti nei file results, drivers, races e status.

1.2. Web scraping

Sono stati raccolti i dati tramite web scraping dal sito

<https://motorsportstats.com/series/formula-one/records/2021>.

In particolare, sono state considerate le voci riguardanti il numero di vittorie, il numero di pole positions e il win percentage (% di gare vinte su gare disputate) associate ad ogni pilota che avesse fatto almeno una vittoria e/o almeno una pole position; inoltre, è stato raccolto il numero dei mondiali vinti da ogni pilota (con un minimo di un mondiale vinto) ed è stato associato al nome del pilota stesso.

Questi dati sono stati raccolti in due file csv e inseriti all'interno del database.

- File WinsPoles
 - Driver: nome del pilota;
 - Wins: numero di gare vinte in carriera;
 - Poles: numero di pole positions ottenute in carriera (1° posto durante la sessione di qualifica);
 - WinPerc: percentuale di gare vinte sul totale delle gare disputate.
- File worldChampions
 - Driver: nome del pilota;
 - WorldTitles: numero di titoli mondiali vinti dal pilota.

2. Realizzazione dell'infografica

Per provare a rispondere alla research question, seppur in maniera non universale, sono stati considerati tutti i piloti che hanno vinto una gara e/o fatto una pole position durante le qualifiche di un Gran Premio. Successivamente, per realizzare le successive analisi, sono stati presi in considerazione degli ambiti specifici per andare ad approfondire alcuni aspetti in particolare più di altri.

Per farci un'idea di ciò che pensa la community dei tifosi della Formula 1 abbiamo ricercato informazioni e opinioni su chi fosse il migliore pilota nella storia. Ad esempio sono state trovate queste classifiche:

<https://www.f1-fansite.com/it/f1-drivers/top-10-best-f1-drivers-of-all-time/>

<https://www.autoappassionati.it/chi-e-il-pilota-di-formula-1-piu-forte-della-storia-ecco-la-classifica-definitiva/>. Non ci siamo trovati, però, d'accordo sui criteri di valutazione di queste classifiche, in quanto non vanno a tenere conto della costanza o del numero di gare corse da ogni pilota. Abbiamo perciò deciso di portare una nostra analisi per decretare chi sia, a nostro avviso, il GOAT della Formula 1.

2.1. Obiettivo

L'obiettivo delle visualizzazioni è, come detto in precedenza, stabilire chi sia il miglior pilota della storia della Formula 1. Trovare una risposta che metta d'accordo tutti è alquanto difficile, ma vorremmo provare a trovare una risposta convincente con il sostegno delle infografiche realizzate. Il nostro focus iniziale è ampio, va a considerare tutti i piloti con almeno una pole position o una vittoria in gara in carriera; durante lo storytelling il focus va via via restringendosi in funzione di alcune caratteristiche, in modo da poter trovare la risposta più adatta alla research question iniziale.

2.2. Target

Il target di queste visualizzazioni è quello di sensibilizzare gli appassionati circa i risultati ottenuti dai più grandi piloti. Secondo la nostra opinione, infatti, non si riesce ad avere sempre una visione oggettiva quando bisogna paragonare due piloti, soprattutto se hanno gareggiato in due epoche diverse. Nelle discussioni prevale quindi sempre più l'affetto per il pilota che una reale oggettività.

3. Presentazione delle infografiche finali

In questa sezione sono presentati i risultati finali delle infografiche; le modifiche fatte durante la valutazione di qualità sono mostrate nella sezione successiva, ma qui vengono riportate le **versioni finali** di ogni infografica.

3.1. Scatterplot con istogrammi marginali (*visualizzazione interattiva*)

L'obiettivo di un pilota di Formula 1 è essere il più veloce in qualifica e arrivare primo in gara.

Il primo passo per arrivare primo in gara è posizionarsi tra le prime file alla partenza effettuando una buona qualifica. È infatti il pilota più veloce in qualifica colui che conquista la pole position e che quindi ha più possibilità di vincere la gara.

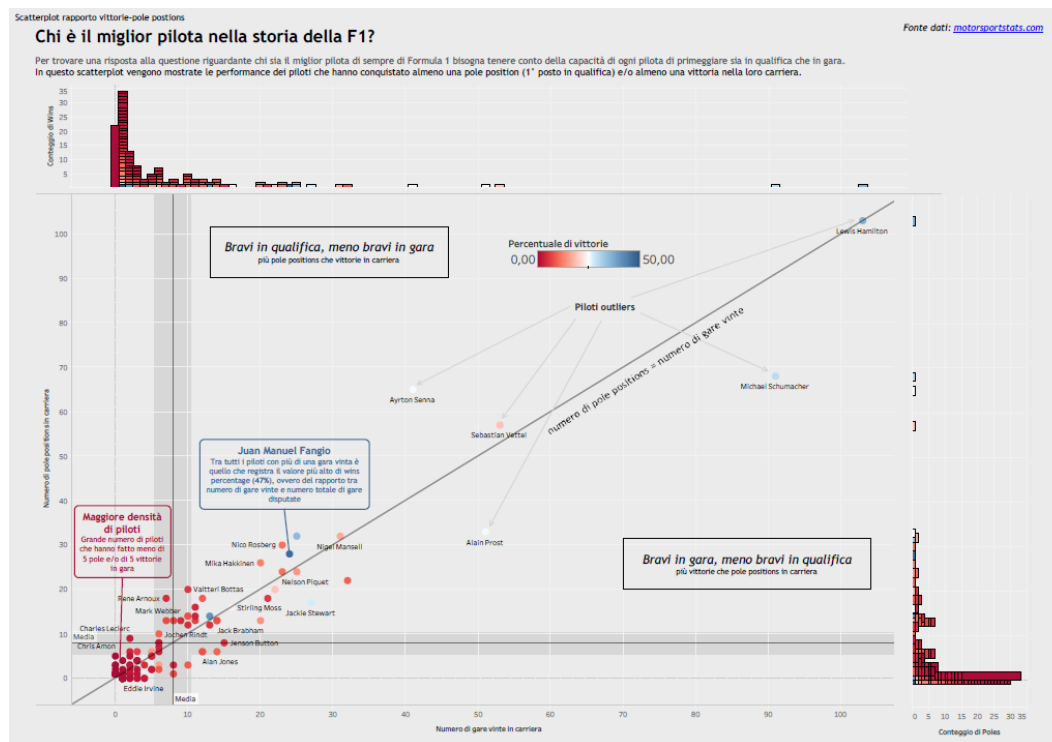
È evidente il fatto che, per determinare la bravura di un pilota, è necessario valutarlo per quante gare ha vinto e per quante pole position ha conquistato.

Ci è sembrato quindi utile mostrare in funzione di queste due dimensioni la distribuzione dei piloti attraverso uno scatterplot, realizzato tramite *Tableau*. All'interno del grafico i piloti sono rappresentati tramite un pallino colorato: non è da sottovalutare la questione del colore, infatti ogni pilota ha il suo pallino che è colorato con una gradazione associata alla sua percentuale di vittorie, ovvero il rapporto tra gare vinte e gare disputate in carriera.

In questo grafico, però, mancava una componente che mostrasse la distribuzione numerica del numero di vittorie (sull'asse x) e del numero delle pole positions (sull'asse y). Sono stati perciò inseriti degli istogrammi marginalmente allo scatterplot, la media delle due misure (wins e poles) e le bande dell'intervallo di confidenza al 95%, in modo da colmare questa mancanza.

Visualizzazione interattiva:

<https://public.tableau.com/app/profile/alessio.sanvito/viz/ScatterPlotF1/StoriaHist?publish=yes>



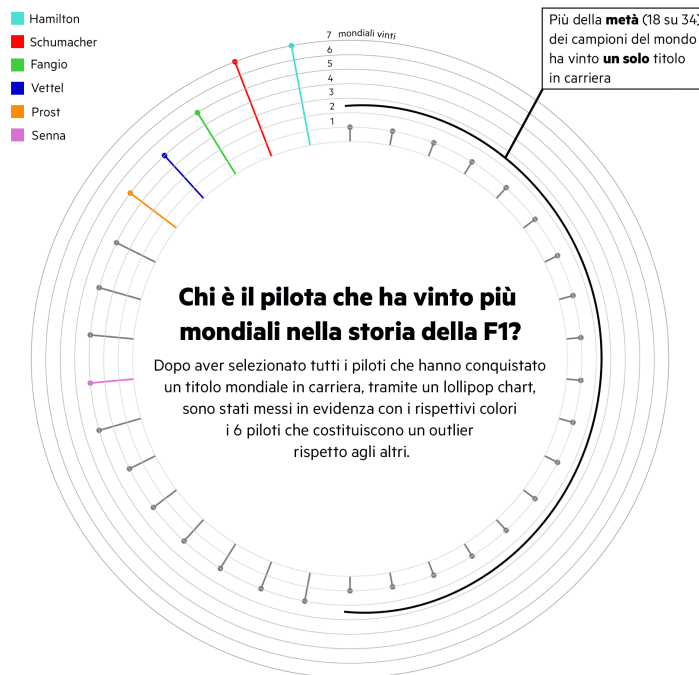
Da questo grafico risulta in maniera ben evidente che il pilota che ha fatto registrare più pole positions e più vittorie è Lewis Hamilton, il quale si trova perfettamente in parità tra numero di vittorie e numero di poles. Oltre ad Hamilton ci sono altri 4 piloti outliers, che si distaccano molto dagli altri: Michael Schumacher, Sebastian Vettel, Alain Prost e Ayrton Senna. Per quanto riguarda il valore di wins percentage il pilota con il valore più alto è Lee Wallard, il quale però ha disputato solo due gare, vincendone una: risulta perciò poco interessante per la nostra domanda. Più interessante è Juan Fangio, il quale ha vinto più di una gara (24 gare vinte), a differenza di Wallard, e ha una wins percentage del 47%.

Per le analisi successive abbiamo perciò deciso di focalizzarci su piloti campioni del mondo, con un focus particolare su questi 6 piloti appena descritti: Hamilton, Schumacher, Vettel, Prost, Senna e Fangio.

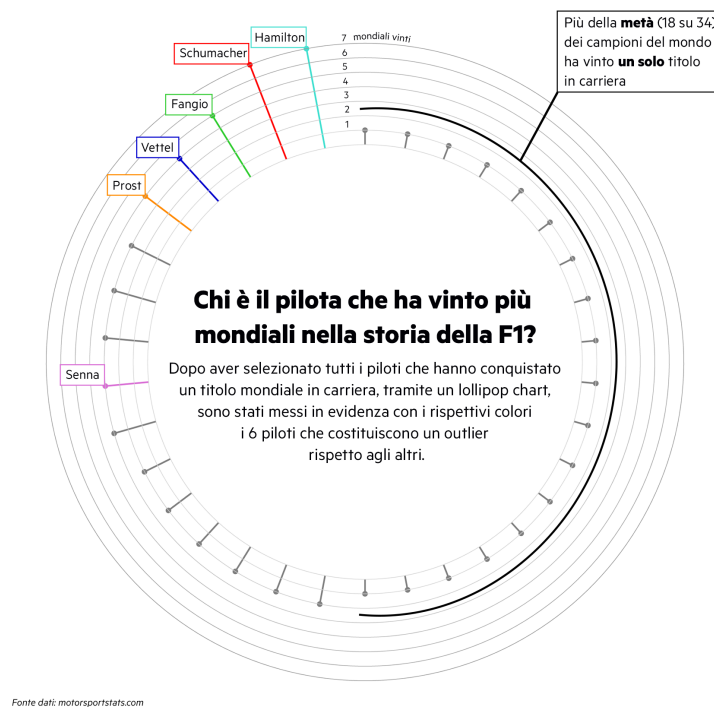
3.2. Lollipop

Il lollipop è stato scelto per andare a mostrare la relazione tra la variabile categorica corrispondente al nome del pilota con la variabile numerica corrispondente al numero di mondiali vinti da ogni pilota.

Questo grafico, realizzato tramite *Python*, serve per mostrare quale è il pilota che ha vinto più titoli mondiali; sono stati considerati soltanto i piloti che avessero vinto almeno un campionato del mondo.



La legenda di questo grafico, però, non era molto soddisfacente; abbiamo dunque modificato questo aspetto del grafico in modo da rendere ancora più rapida la comprensione dello stesso.



Link per foto con migliore risoluzione:

<https://drive.google.com/file/d/1kHOwcCxtiCAGw5A9UfBgIrAoyS-l7Jq5/view?usp=sharing>

Ne risulta che, tra i piloti outliers identificati nello scatterplot, Hamilton e Schumacher sono i due piloti più titolati con 7 titoli a testa, seguiti da Fangio con 5 titoli, da Vettel e Prost con 4 titoli e da Senna con 3. Bisogna considerare però che Hamilton e Vettel sono piloti ancora in attività, Schumacher, Prost e Fangio si sono ritirati e Senna ha avuto una carriera più breve del previsto, a causa della sua prematura scomparsa in un incidente nelle qualifiche del GP di Imola (1994). Inoltre, bisogna considerare il numero di gare corse dai piloti e, per definire un pilota migliore di un altro, bisogna vedere anche quanto un pilota è costante in termini di risultati. Per questo è stato realizzato il grafico successivo.

3.3. Linechart

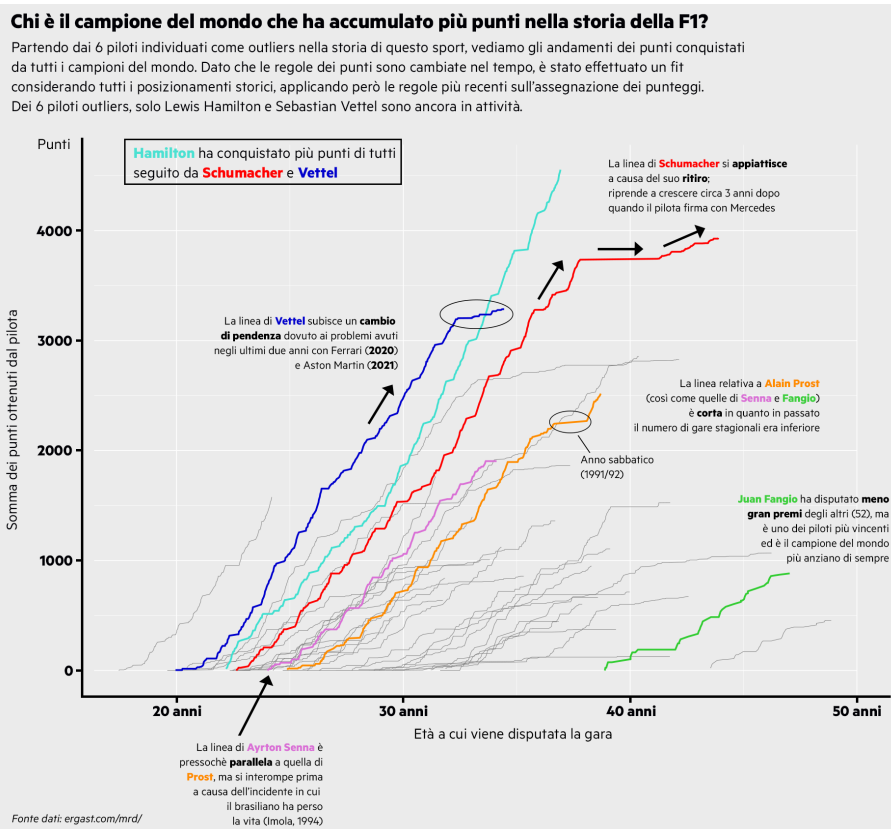
Un'altra metrica per la quale può essere utile paragonare i diversi piloti è relativa a quanti punti hanno accumulato nella loro carriera. Per valutare l'andamento di ogni pilota durante gli anni della sua carriera è stata fatta un'analisi sui punti realizzati in funzione dell'età del pilota; questo va ad evidenziare la performance nel tempo di ogni pilota e la costanza nell'ottenere punti (quindi la costanza nel piazzarsi nelle posizioni di testa).

Il line chart (realizzato con *Python*) è stato scelto per mostrare l'evoluzione nel tempo di alcune variabili numeriche, in questo caso particolare delle variabili che mostrano i punti realizzati in carriera da ogni pilota che abbia vinto almeno un campionato del mondo.

Per la realizzazione di questo grafico è servita una procedura di uniformazione dei valori dei punteggi assegnati in base all'ordine di arrivo di ciascuna gara: infatti, dato che i piloti considerati appartengono a diverse epoche della storia della F1, le regole sull'assegnazione dei punteggi non sono mai state costanti nel tempo. È stato perciò effettuato un fit considerando i posizionamenti storici a cui è stata applicata la regolamentazione dei punteggi attuale, esclusi punteggi aggiuntivi (come quello per il giro più veloce in gara).

Si riportano gli standard di assegnazione dei punti attuale:

Posizionamento	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11° in poi
Punti	25	18	15	12	10	8	6	4	2	1	0



Link per foto con migliore risoluzione:

<https://drive.google.com/file/d/1j1mKXYtj1frgpVqs-eL8-cp5uKftMFen/view?usp=sharing>

Da questo Line Chart si può notare che i tre piloti più “moderni”, ovvero Vettel, Schumacher e Hamilton sono quelli che hanno realizzato più punti di tutti; in particolare, gli ultimi due hanno guadagnato almeno 500 punti in più del primo.

Il confronto quindi può essere limitato a Hamilton e Schumacher; il primo è l'unico pilota nella storia ad aver superato la soglia dei 4000 punti. Il secondo, nonostante una carriera (al momento) più lunga, vanta una quantità di punti di poco inferiore, dovuta anche a un periodo di inattività di circa 3 anni.

3.4. Alluvial plot (visualizzazione interattiva)

In seguito, come ultimo passo della nostra analisi, si è voluto realizzare un confronto tra Hamilton e Schumacher, i due piloti che, secondo le analisi precedenti, si giocano il titolo di miglior pilota di sempre.

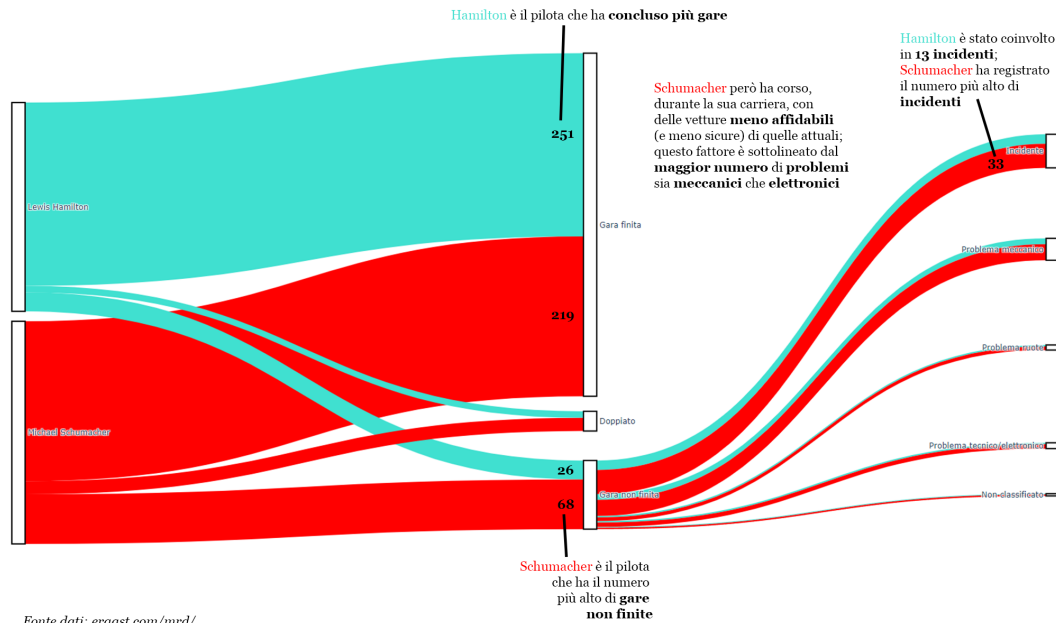
È stata quindi fatta un'analisi sull'andamento delle gare dei due piloti tramite un alluvial plot (realizzato tramite *Python*): è stato mostrato come i fatti legati ai due piloti sono legati alle dimensioni categoriche riguardanti la fine delle loro gare. Vengono mostrati i collegamenti che indicano come sono state terminate le gare dai due. Le categorie presenti sono: gara finita, quindi la gara è stata terminata, senza essere stato doppiato; doppiato,

quindi la gara è stata terminata dopo essere stato doppiato; gara non finita, ovvero gara terminata con un ritiro (dovuto a diversi problemi, come problemi meccanici, elettronici, problemi alle ruote ecc.).

Visualizzazione interattiva: danielepoterti.github.io

Come si sono concluse le gare dei due migliori piloti?

Attraverso l'analisi dei risultati di ogni gara disputata da ognuno dei piloti considerati, è stato realizzato un alluvial plot per rappresentare la quantità di gare concluse, concluse dopo essere stati doppiati o non concluse (con le varie motivazioni).



Link per foto con migliore risoluzione:

https://drive.google.com/file/d/1pnrsOq-iHcXM-ktf91rvnP3MzP7_uO8u/view?usp=sharing

Ne risulta che Hamilton ha subito meno problemi di affidabilità (dovuto all'evoluzione delle tecnologie), ma è stato anche capace di portare a termine più gare e di totalizzare meno incidenti rispetto al pilota tedesco, denotando una maggiore costanza nelle performances e una maggiore abilità nell'evitare e/o gestire problemi.

4. Valutazione dell'infografica

È stata poi eseguita una valutazione dell'infografica, un assessment di tipo qualitativo-quantitativo per controllare le visualizzazioni realizzate. Come prima cosa è stata fatta una valutazione euristica con il fine di raccogliere ed elencare alcuni problemi (di usabilità e comprensione) delle visualizzazioni create; successivamente è stato creato un test utente, in modo da testare efficacia ed efficienza delle visualizzazioni; infine, è stato fatto compilare un questionario psicometrico ad alcuni utenti per raccoglierne dei pareri soggettivi riguardo la qualità delle infoviz.

4.1. Valutazione euristica

Nella valutazione euristica sono stati scelti 3 utenti e gli è stato chiesto di interagire con le nostre visualizzazioni. È stata utilizzata la metodologia "think aloud", dunque ogni utente ha commentato a voce alta quanto stava vedendo. I risultati sono i seguenti:

Utenti	Valutazione
U1	(Scatterplot) Non conosce il termine outlier, non conosce i boxplot e non comprende il range della scala dei colori (0-50); con le dovute spiegazioni ritiene chiaro il tutto. Fatica ad interagire con l'interfaccia di Tableau Public. (Line Chart) Confuso a primo impatto, tende a ignorare il contesto; una volta esplicitata l'importanza di questo ne fa uso e comprende quanto rappresentato. Troppa confusione nell'alluvial plot. Capisce il lollipop intuitivamente.
U2	Riesce a comprendere ed interpretare tutti e 4 i grafici con leggera difficoltà per l'alluvial plot; necessita di essere leggermente guidato nell'interpretazione delle visualizzazioni, ma comprende praticamente tutto da sé a intuito.
U3	Inizialmente non coglie la possibilità di interagire con il grafico (scatterplot); una volta illustratagli questa possibilità la sfrutta e analizza correttamente quanto visibile. Non conosce i boxplot e il loro significato, dopo una breve spiegazione riesce a percepirne la finalità e l'importanza. Critica l'alluvial che presenta i 6 piloti, risulta di difficile comprensione; suggerisce di ridurre il numero dei piloti e sottolinea l'importanza dell'interattività in questo grafico per la piena comprensione. Comprende gli ultimi due grafici subito.

Dal momento che le maggiori critiche sono state portate nei confronti dell'alluvial plot e dello scatterplot si è deciso di cercare di portare un miglioramento a questi grafici. Nell'alluvial sono stati considerati solo i piloti Hamilton e Schumacher, in modo da fare un

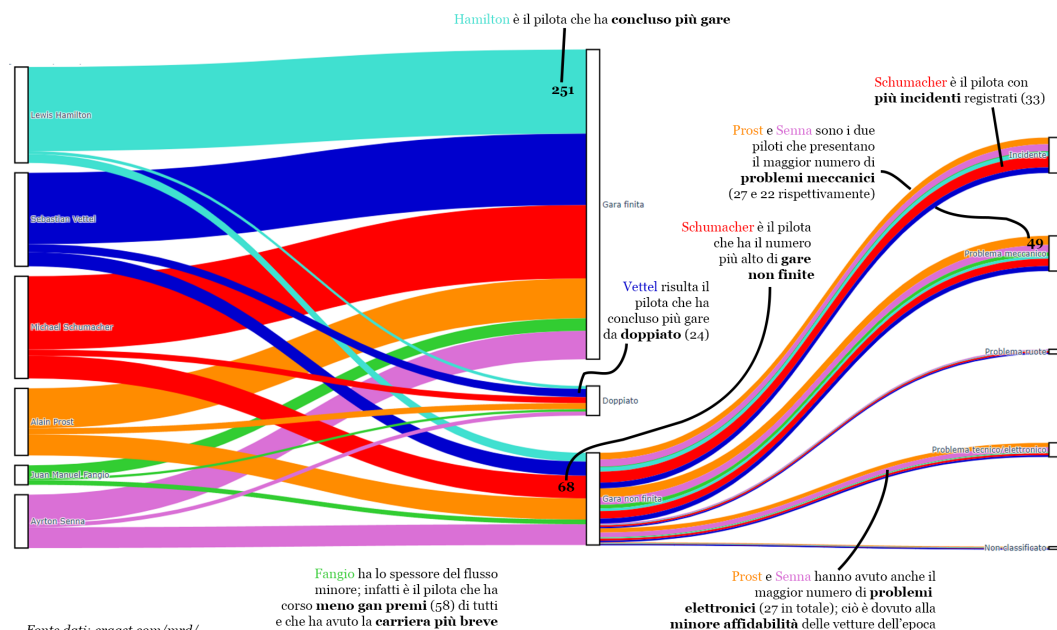
confronto diretto tra i due piloti in questione e da poter raccogliere ancora più informazioni sui due, informazioni necessarie a rispondere alla research question.

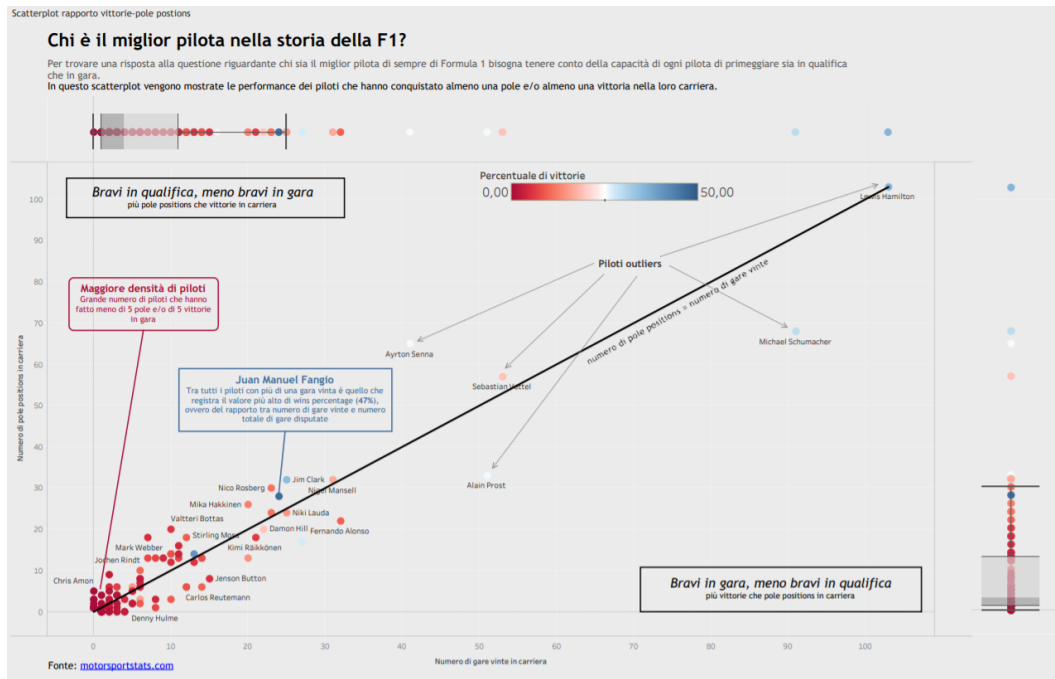
Sono state apportate anche delle modifiche allo scatterplot: al posto dei boxplot sono stati inseriti degli istogrammi, con le fasce che rappresentano l'intervallo di confidenza per la media. I boxplot sono stati ritenuti poco informativi in quanto nascondono dei dettagli rilevanti come la numerosità di ogni occorrenza; inoltre, è stata rappresentata la linea della media per evidenziare quali siano i piloti che si distaccano dagli altri in termini di pole position e vittorie.

Grafici prima:

Come si sono concluse le gare dei migliori 6 piloti?

Attraverso l'analisi dei risultati di ogni gara disputata da ognuno dei piloti considerati, è stato realizzato un alluvial plot per rappresentare la quantità di gare concluse, concluse dopo essere stati doppiati o non concluse (con le varie motivazioni).

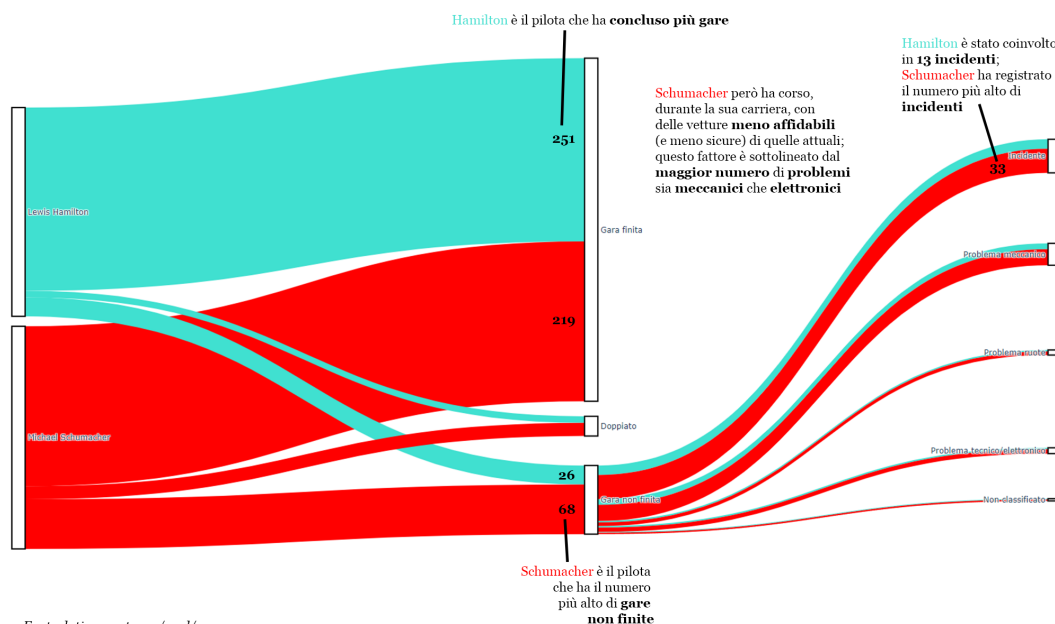


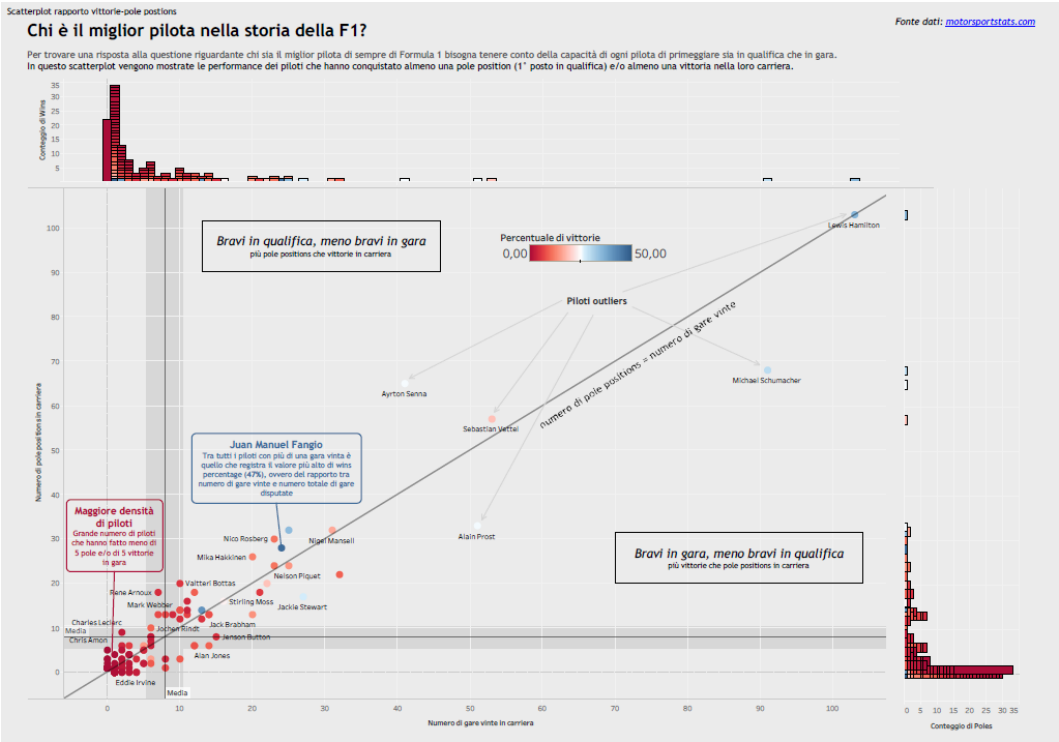


Grafici dopo:

Come si sono concluse le gare dei due migliori piloti?

Attraverso l'analisi dei risultati di ogni gara disputata da ognuno dei piloti considerati, è stato realizzato un alluvial plot per rappresentare la quantità di gare concluse, concluse dopo essere stati doppiati o non concluse (con le varie motivazioni).





Dopo aver modificato l'alluvial plot e lo scatterplot abbiamo deciso di sottoporli nuovamente all'analisi dei 3 utenti.

U1	(Alluvial) Ritiene più chiaro quanto rappresentato; necessita però di tempo per comprendere a fondo la visualizzazione. Sfrutta in autonomia l'interattività del grafico. (Scatterplot) Ritiene migliori gli istogrammi rispetto ai boxplot in quanto sono più conosciuti e facilmente interpretabili. Comprende il significato delle bande.
U2	Ritiene soddisfacenti i cambiamenti apportati e continua a comprendere quanto rappresentato.
U3	(Alluvial) Apprezza il cambio di focus, ma trova ancora qualche difficoltà ad analizzare la visualizzazione in autonomia; vorrebbe che le specifiche riguardanti i "flussi" fossero fisse sullo schermo, senza bisogno di dover interagire con il grafico. (Scatterplot) Preferisce la visualizzazione con gli istogrammi marginali, lo ritiene un buon upgrade della visualizzazione; con un aiuto comprende lo scopo delle bande.

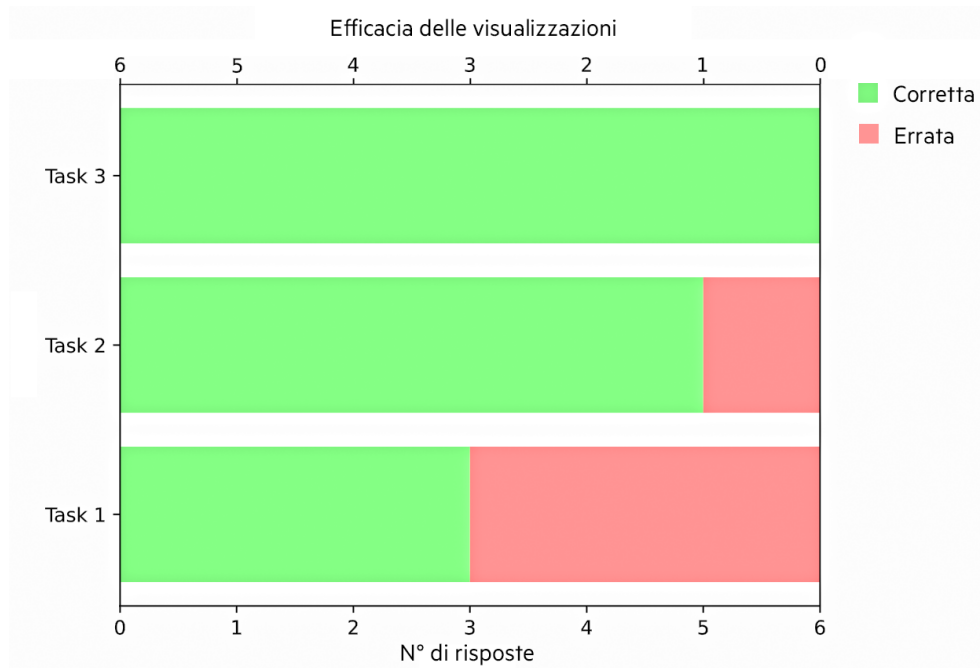
4.2. Test utente

Successivamente alla valutazione euristica è stato fatto un test utente: sono stati selezionati 6 utenti e 3 task da far eseguire ad ognuno. In primis è stata realizzata una stima del tempo medio necessario a completare ogni task; sono stati poi cronometrati gli

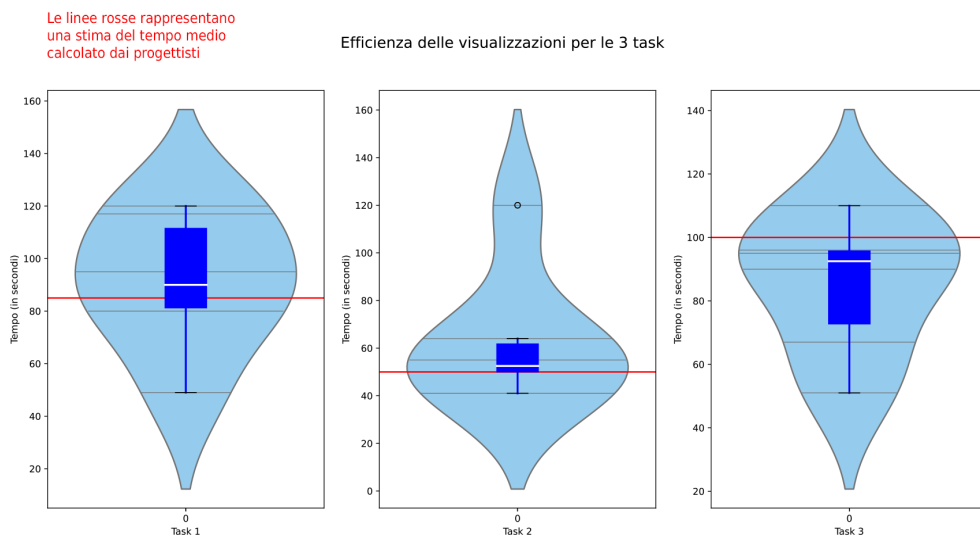
utenti per valutare il tempo impiegato per dare una risposta ad ogni task, quindi per valutare l'efficienza delle nostre visualizzazioni. Infine, sono state registrate le risposte fornite dagli utenti per valutare l'efficacia delle visualizzazioni. Le risposte verdi sono quelle completate autonomamente e senza errori, quelle rosse sono le risposte sbagliate o quelle fornite grazie al supporto offerto agli utenti.

Task	Compito
Task 1	Quanti e quali sono i piloti che hanno una percentuale di vittorie maggiore del 25% e che hanno realizzato più pole position e più vittorie rispetto alla media?
Task 2	Evidenziare, tra Hamilton e Schumacher, il pilota che ha finito più gare, che ha fatto più incidenti, che ha avuto più problemi meccanici e che ha avuto più problemi elettronici. Motivare la risposta confrontando i dati numerici.
Task 3	Elencare chi sia, tra i 6 piloti evidenziati, il più giovane pilota a raggiungere soglia 1000, 2000, 3000, 4000 punti e chi sia il più vecchio a raggiungere le stesse soglie. A che età (approssimativamente) sono stati raggiunti questi traguardi dai piloti in questione?

Utenti	Task 1 (tempo)	Task 2 (tempo)	Task 3 (tempo)
Tempo Medio	85"	50"	100"
U1	80"	120"	67"
U2	49"	50"	90"
U3	85"	50"	110"
U4	95"	55"	95"
U5	120"	41"	51"
U6	117"	64"	96"



Si nota che la task 3 non presenta problemi, tutti e 6 gli utenti hanno risposto correttamente alla domanda posta; per quanto riguarda la task 2 viene registrato un solo errore; la task 1, invece, risulta la più ostica, generando metà risposte corrette e metà risposte errate o assistite.



Per analizzare i tempi di esecuzione delle 3 task sono stati creati 3 violin plot, uno per ogni task, con, al loro interno, il relativo boxplot. Tramite il violin plot si può osservare la distribuzione dei dati, tramite il boxplot i valori massimo, minimo, la mediana (linea bianca) e gli outlier, tramite gli stick del violin plot vengono mostrati i dati veri e propri. Inoltre,

tramite la linea rossa, è stato rappresentato il tempo medio da noi stabilito necessario al completamento di ogni task.

La prima task è maggiormente distribuita nell'intervallo tra 80/85 e 110/120 secondi; la seconda task è maggiormente distribuita tra i 40 e i 65 secondi, con un outlier a 120 secondi; la terza task è maggiormente distribuita tra 90 e 110 secondi.

4.3. Questionario

Infine, per quanto riguarda la valutazione psicometrica, è stato fornito un questionario a degli utenti in modo che potessero valutare la nostra data visualization utilizzando la scala Cabitza-Locoro. Di seguito viene mostrata la scala:

Valuta la qualità dell'infografica riportata in questa pagina dando un valore da 1 (pochissimo) a 6 (moltissimo) a ciascuno dei seguenti aggettivi:

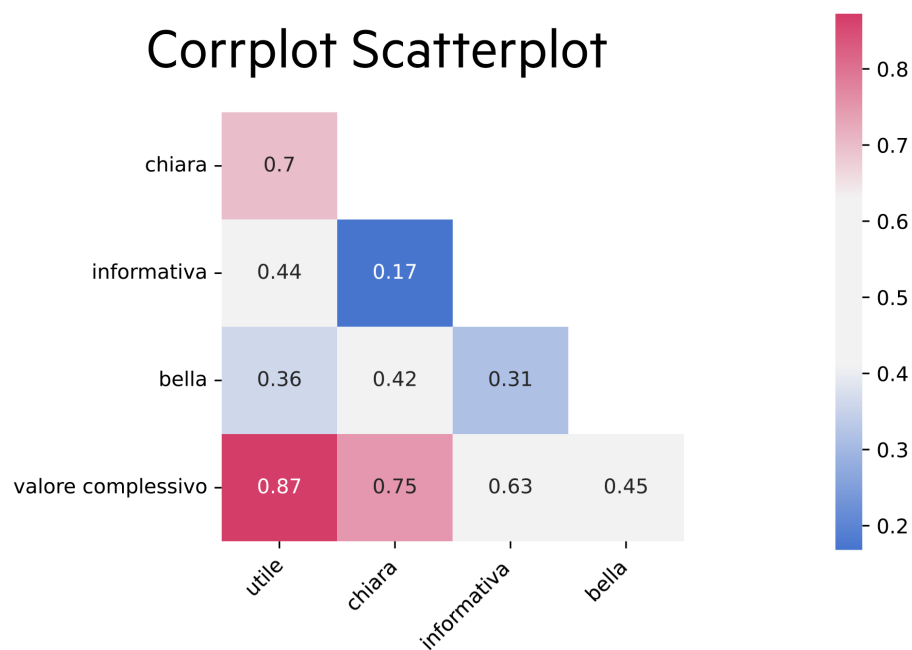
	1 - pochissimo					6 - moltissimo
Utile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Chiara	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Informativa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bella	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Valuta infine l'infografica indicando un valore di qualità complessivo da te percepito:

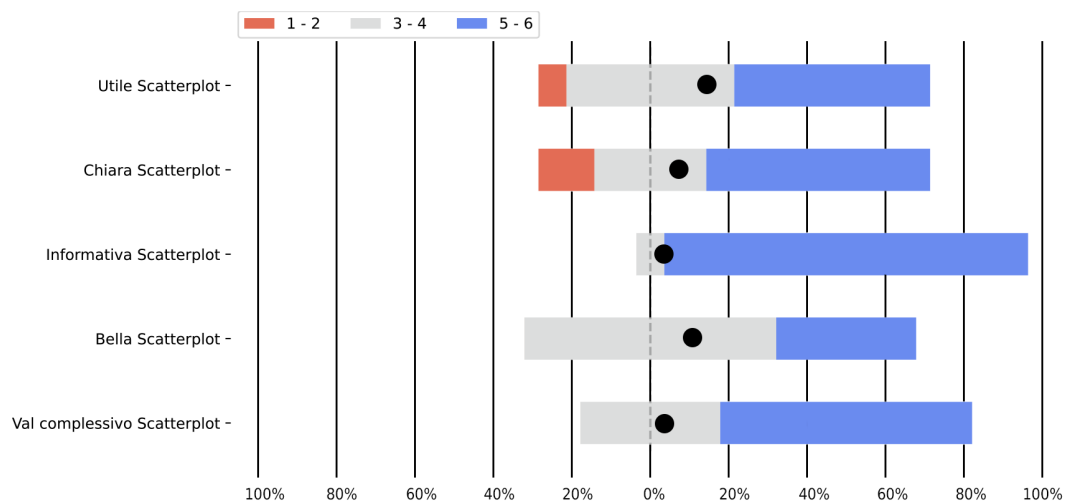
	1 - bassissimo					6 - altissimo
Valore complessivo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Sono stati analizzati i risultati ottenuti nel questionario psicometrico per ognuna delle 4 visualizzazioni. Sono stati fatti 2 tipi di valutazione: la prima mostra la correlazione tra le misure di valutazione; la seconda mostra la numerosità delle risposte per ogni categoria di ogni misura di valutazione (numero di risposte di valore 1-2 per l'utilità, numero di risposte di valore 3-4 per l'utilità, ecc.).

- **Scatterplot**



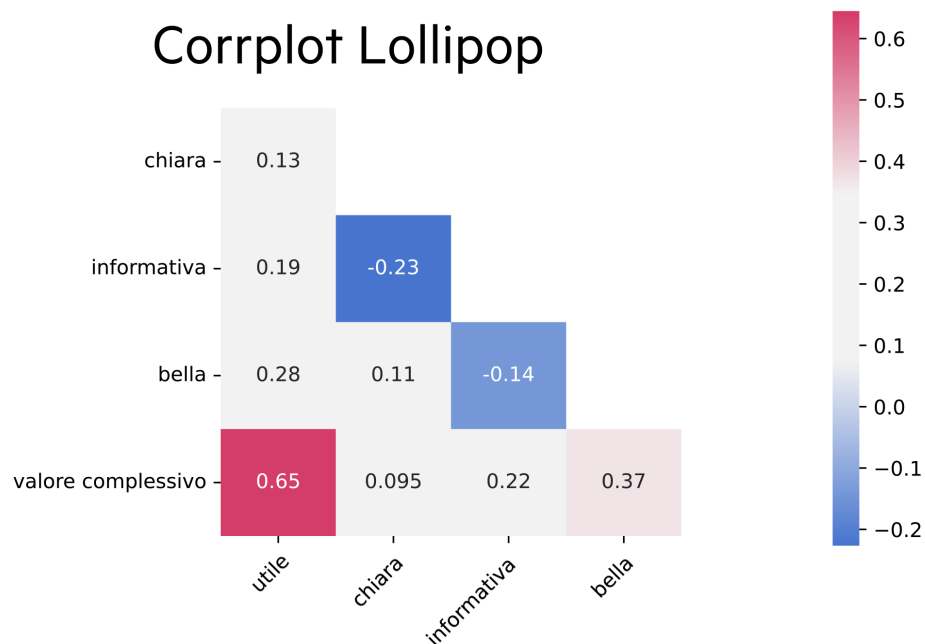
Si nota, per lo scatterplot, una buona correlazione tra chiarezza e utilità (0.7), tra chiarezza e valore complessivo (0.75) e una forte correlazione tra utilità e valore complessivo (0.87).



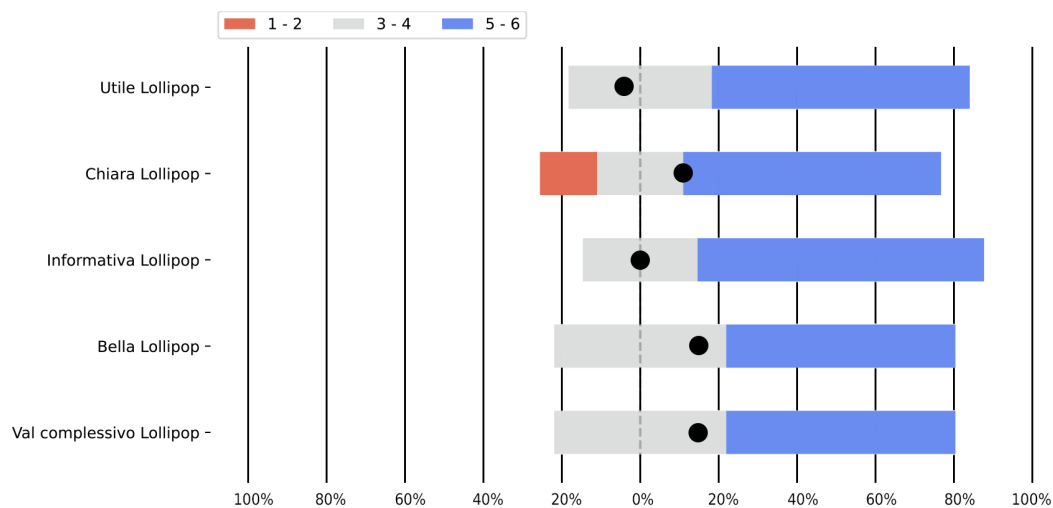
Per lo scatterplot sono stati ottenuti per la maggior parte risultati positivi per tutte le misure di valutazione.

Si nota che, per quanto riguarda i valori centrali (3-4), per tutte le misure di valutazione i risultati tendono ad assumere valore 4 piuttosto che valore 3; questo è visibile dalla posizione del pallino nero (che ha la dimensione del diametro pari all'intervallo di confidenza) all'interno dell'area grigia.

- Lollipop



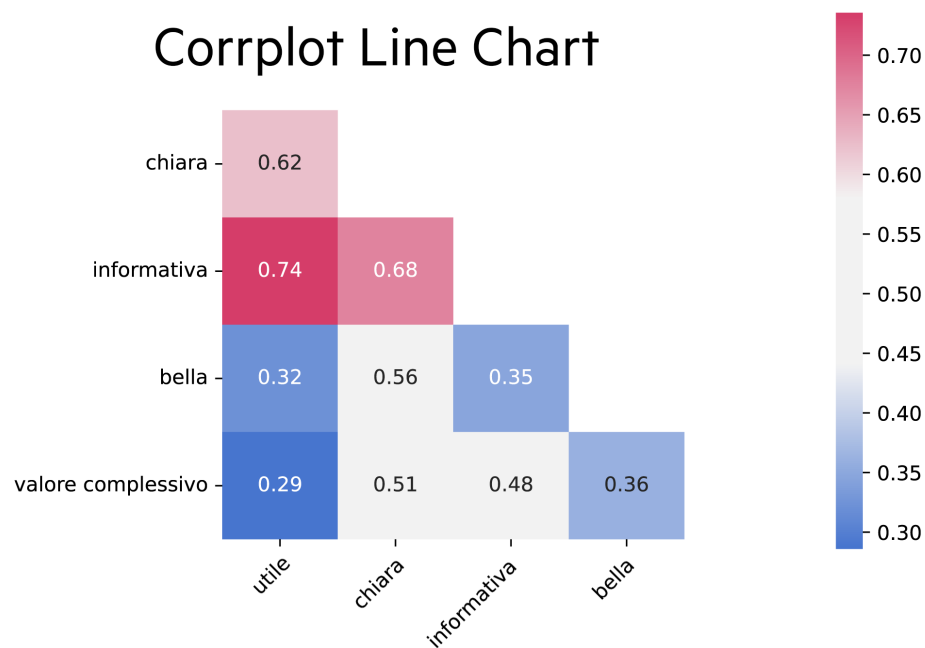
Si nota, per il lollipop, una discreta correlazione tra utilità e valore complessivo (0.65); le altre misure di valutazione non sono particolarmente fortemente correlate.



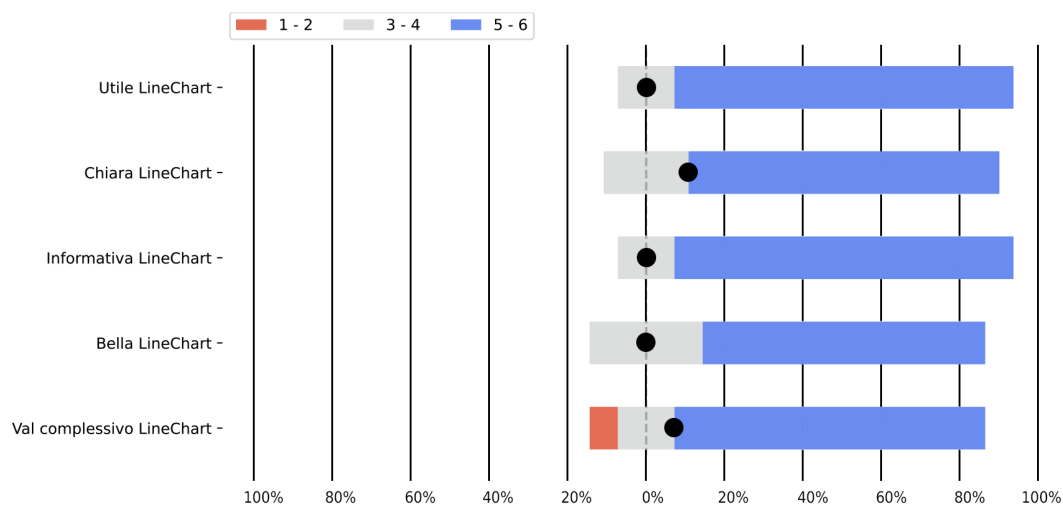
Per il lollipop sono stati ottenuti per la maggior parte risultati positivi per tutte le misure di valutazione.

Si nota che, per quanto riguarda i valori centrali (3-4), per le misure di valutazione chiarezza, bellezza e valore complessivo i risultati tendono ad assumere valore 4 piuttosto che valore 3. Per l'utilità assumono più valori pari a 3 che a 4; per l'informatività hanno pari voti.

- Line Chart



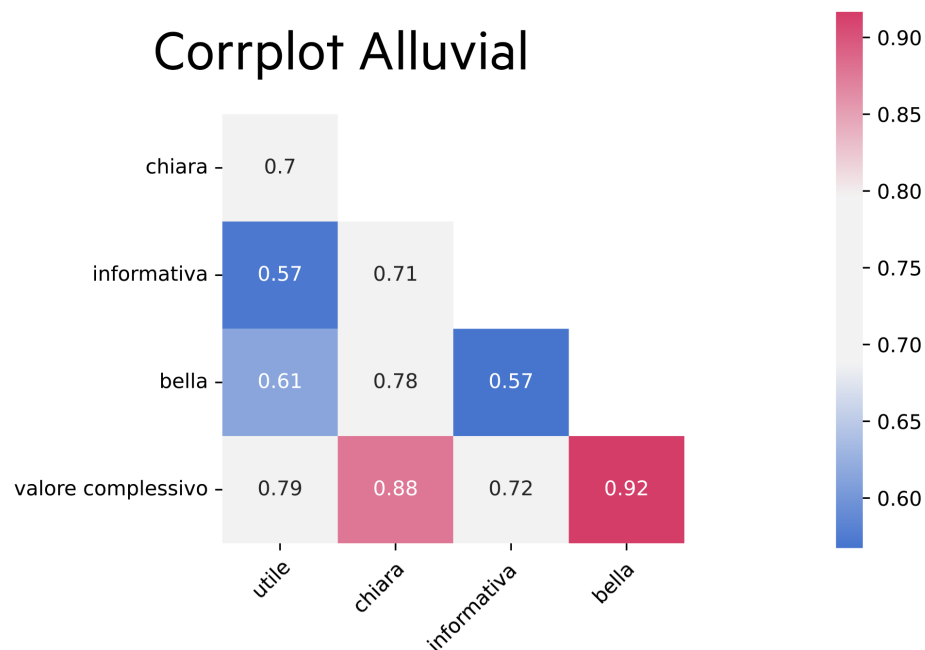
Si nota, per il line chart, una buona correlazione tra chiarezza e utilità (0.62), tra utilità e informatività (0.74) e tra chiarezza e informatività (0.68).



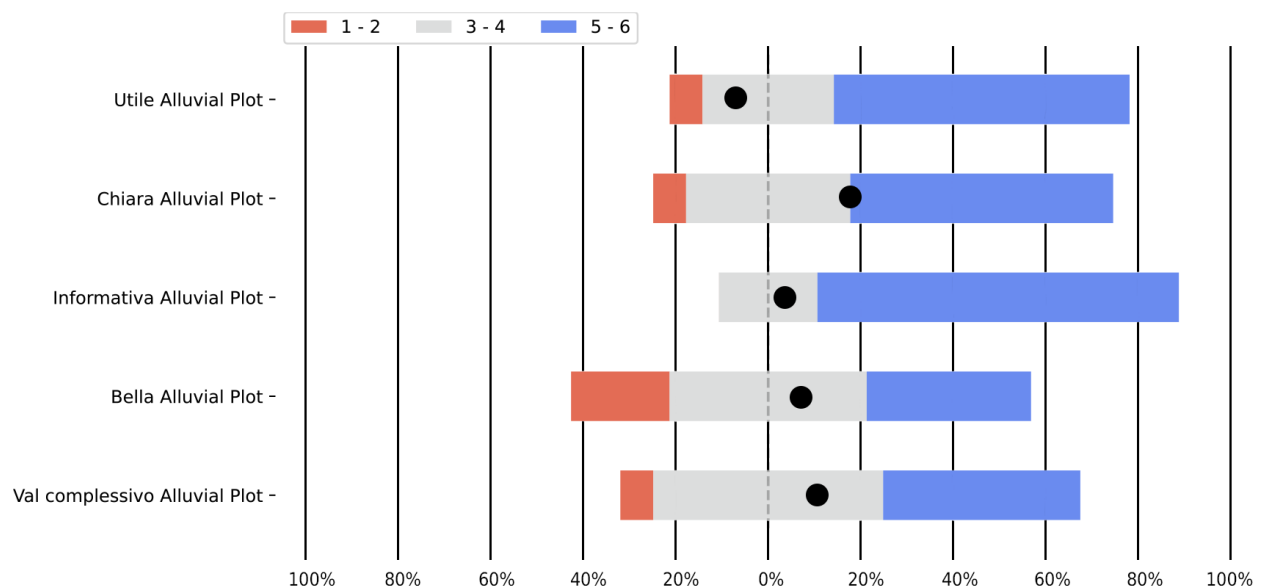
Anche per il line chart sono stati ottenuti per la maggior parte risultati positivi per tutte le misure di valutazione.

Si nota che, per quanto riguarda i valori centrali (3-4), per la chiarezza e il valore complessivo i risultati tendono ad assumere valore 4 piuttosto che valore 3. Per utilità, informatività e bellezza hanno lo stesso numero di valori pari a 3 e pari a 4.

- **Alluvial Plot**



Si nota, per l'alluvial plot, una buona correlazione tra praticamente tutte le misure di valutazione; da sottolineare la forte correlazione tra chiarezza e valore complessivo (0.88) e tra bellezza e valore complessivo (0.92).



Infine, per l'alluvial plot sono stati ottenuti per la maggior parte risultati positivi per le misure di valutazione. Questo grafico ha ottenuto risultati un po' altalenanti per quanto riguarda la bellezza, anche se, anche per questa misura di valutazione, sono stati registrati più valori positivi (4-5-6) che negativi (1-2-3).

Si nota che, per quanto riguarda i valori centrali (3-4), per tutte le misure di valutazione i risultati tendono ad assumere valore 4 piuttosto che valore 3. Fa eccezione solo l'utilità che ha registrato una maggioranza di valori pari a 3 piuttosto che pari a 4.

Conclusioni e sviluppi futuri

In conclusione, l'obiettivo del progetto era quello di dimostrare, tramite analisi e visualizzazioni, chi fosse il miglior pilota esistente nella storia della Formula 1.

Dai dati è emerso che:

1. Tra tutti i campioni del mondo quelli maggiormente titolati risultano essere Hamilton e Schumacher (7 titoli a testa);
2. Tra tutti i campioni del mondo quelli che hanno realizzato il maggior numero di punti in carriera risultano nuovamente Hamilton (al primo posto, con più di 4000 punti) e Schumacher (al secondo posto);
3. Tra i due piloti, quello più costante, con la vettura più affidabile e meno coinvolto in incidenti è risultato essere Hamilton.

Sicuramente il dominio della Mercedes (team di appartenenza di Hamilton) negli ultimi 8 anni e la maggiore affidabilità delle nuove vetture hanno influito sui risultati ottenuti dai due campioni.

Nonostante ciò, alla luce di tutti i risultati ottenuti nelle varie infografiche, ci sentiamo di sbilanciarci dando una risposta alla domanda iniziale: secondo noi (e secondo quanto analizzato) Lewis Hamilton risulta il miglior pilota di sempre.

Possibili sviluppi futuri a questo progetto potrebbero essere apportati realizzando ulteriori visualizzazioni su temi non analizzati come, ad esempio, realizzando un Line Chart con il numero di vittorie e la stagione in cui sono state realizzate; si potrebbe altrimenti provare ad analizzare la costanza dei piloti considerando la percentuale di vittorie e il numero di gare disputate; o ancora, considerando i migliori 5 piloti, si potrebbe cercare quello che ha ottenuto un maggior incremento prestazionale durante la carriera.