

Cyclistic bike-share analysis: strategie per convertire i casual riders in membri annuali

1. Introduzione

1.1 Obiettivo del progetto

Cyclistic è un servizio di bike-sharing attivo nella città di Chicago, che offre sia piani di abbonamento annuale sia opzioni di utilizzo per singola corsa o pass giornalieri. L'obiettivo principale di questo progetto è **analizzare in che modo i “casual riders” (i clienti occasionali) utilizzano il servizio rispetto ai membri annuali**, per poi definire strategie di marketing che possano convertire un maggior numero di utenti occasionali in abbonati annuali.

Attraverso questa analisi, si mira a:

- Individuare le differenze di comportamento (durata media, orari e giorni di utilizzo, tipologia di bici, stagionalità) tra casual riders e membri.
- Fornire evidenze concrete per sviluppare campagne mirate a incrementare le sottoscrizioni annuali.
- Proporre raccomandazioni operative e ipotesi di miglioramento basate su dati oggettivi.

1.2 Domanda di business e stakeholder

La **domanda di business** a cui rispondere è:

“Come differiscono i comportamenti di utilizzo tra casual riders e abbonati annuali, e in che modo queste differenze possono essere sfruttate per convertire più casual riders in membri annuali?”

Stakeholder principali

- **Direttrice marketing, Lily Moreno:** interessata a capire quali leve promozionali utilizzare per incrementare i piani annuali.
- **Cyclistic marketing analytics team:** responsabile di raccogliere, analizzare e interpretare i dati per supportare le decisioni strategiche di marketing.
- **Cyclistic executive team:** ha il compito di approvare le raccomandazioni e allocare le risorse necessarie per implementarle.

1.3 Struttura del report

Il presente report è suddiviso nei seguenti capitoli principali:

- **Dati e metodologia**
Descrive la provenienza dei dati, le procedure di pulizia, gli strumenti adottati e i passaggi principali di preparazione del dataset.
- **Analisi esplorativa**
Approfondisce i risultati derivanti dalle query SQL e/o dalle analisi svolte, con particolare attenzione alle differenze tra casual riders e membri in termini di durata delle corse, orari di utilizzo, tipologia di bici e stagionalità.
- **Sintesi delle evidenze principali**
Raccoglie le evidenze fondamentali emerse dall'analisi, riassumendo le principali divergenze tra casual e membri.
- **Raccomandazioni e next steps**
Propone azioni pratiche per migliorare il tasso di conversione dei casual riders in membri annuali, suggerendo anche possibili approfondimenti e KPI da monitorare.
- **Limitazioni e sviluppi futuri**
Illustra i limiti riscontrati (dati anonimi, assenza di informazioni su costi/ricavi, ecc.) e ipotizza ulteriori tipologie di analisi e integrazioni di dataset per perfezionare la strategia.

Nota: Nel documento, le **Appendici** contengono principalmente le **query SQL complete**, mentre le **visualizzazioni aggiuntive** sono raccolte in una sezione separata dedicata ("Visualizzazioni"). In tal modo, il corpo del report resta focalizzato sulle evidenze e sulle raccomandazioni, mentre i dettagli tecnici e i contenuti grafici supplementari sono collocati nelle rispettive sezioni di supporto.

2. Dati e metodologia

2.1 Fonti e licenze

I dati utilizzati per l'analisi provengono dal **dataset pubblico di Cyclistic**, messo a disposizione da Motivate International Inc. sotto licenza aperta. Il dataset, anonimo e privo di informazioni personali (PII), copre un periodo di **12 mesi** relativi all'anno 2024.

- **Origine:** [<https://divvy-tripdata.s3.amazonaws.com/index.html>]
- **Formato:** 12 file **.csv**, uno per ciascun mese (da gennaio 2024 a dicembre 2024).
- **Licenza:** licenza aperta che consente l'utilizzo dei dati a fini di analisi.

Questa disponibilità pubblica e l'assenza di PII garantiscono il rispetto della privacy e la conformità alle normative vigenti in tema di protezione dei dati personali.

2.2 Descrizione dei dataset (12 file **.csv**)

Ogni file mensile contiene:

- **ride_id**: identificativo univoco della corsa
- **started_at** e **ended_at**: timestamp di inizio e fine della corsa
- **start_station_name** e **end_station_name**: nomi delle stazioni di partenza e arrivo
- **start_station_id** e **end_station_id**: ID numerici delle stazioni (in alcuni casi mancanti)
- **rideable_type**: tipologia di bici (classica, elettrica, ecc.)
- **member_casual**: indica se l'utente è un membro annuale ("member") o un utilizzatore occasionale ("casual")

Complessivamente, i 12 file includono diversi milioni di record. I **.csv** sono stati **organizzati in cartelle** separate (raw data e dati puliti) e infine caricati in un progetto su **Google BigQuery** per facilitarne la gestione e l'analisi.

2.3 Pulizia e trasformazione dei dati

2.3.1 Unione e verifica delle tabelle mensili

Per semplificare la gestione dell'intero anno, tutte le tabelle mensili caricate in BigQuery sono state **unificate** in una singola tabella finale. Al termine, sono stati eseguiti **controlli di conteggio record** e **controlli a campione** per verificare l'assenza di duplicati o anomalie macroscopiche.

(La query SQL dettagliata è fornita in Appendice A.)

2.3.2 Aggiunta di colonne calcolate

Due colonne aggiuntive sono state create per migliorare la successiva fase di analisi:

- **ride_length**: differenza tra **ended_at** e **started_at**, espressa in secondi.
- **day_of_week**: giorno della settimana (Lunedì, Martedì, ecc.) ricavato da **started_at**.

(Per i dettagli sulle query utilizzate, si rimanda alla Appendice A.)

2.3.3 Identificazione e rimozione di anomalie

Sono poi stati individuati i record con **ride_length** nullo o negativo, i quali rappresentano meno dell'1% del totale. Date le ridotte dimensioni di questa porzione, si è deciso di rimuoverli per semplificare l'analisi e prevenire distorsioni. Anche questa operazione è stata condotta tramite query SQL in BigQuery, descritte in Appendice A.

Al termine, il dataset **cleaned_data** risultante (senza valori anomali) è stato considerato pronto per l'analisi.

2.4 Strumenti utilizzati

Per svolgere le operazioni di estrazione, pulizia e analisi dei dati, si sono adottati i seguenti strumenti:

- **Google Cloud Storage:** per caricare i file di grandi dimensioni (oltre 100 MB) e renderli accessibili a BigQuery.
- **Google BigQuery:** per gestire le tabelle mensili, unificarle in un singolo dataset, ed eseguire query di pulizia e aggregazione.
- **SQL** (linguaggio di query): scelto per la sua scalabilità ed efficienza nella gestione di dataset di grandi dimensioni, oltre che per l'integrazione nativa con BigQuery.
- **Fogli Google (Google Sheets):** impiegati in una fase successiva per creare visualizzazioni rapide, a partire dai risultati estratti con BigQuery.

Con il dataset "cleaned_data" ora pronto, si può procedere all'**analisi esplorativa**, presentata nella sezione successiva, per determinare in che modo gli utenti casual differiscono dagli abbonati annuali e quali strategie possano favorire la conversione verso i piani di abbonamento.

3. Analisi esplorativa

Introduzione

In questa fase, l'obiettivo principale è rispondere alla prima domanda assegnata dal direttore marketing, Lily Moreno:

Come utilizzano le biciclette di Cyclistic i membri annuali e gli utenti occasionali in modo diverso?

Pur sapendo che l'azienda è interessata anche ai **motivi per cui i casual riders comprerebbero un abbonamento annuale** e a **come sfruttare i canali digitali per influenzarli**, in questo report ci concentreremo soprattutto sul confronto tra i due segmenti di utenza (casual vs. membri). Le evidenze ricavate serviranno in parte anche a orientare le strategie di marketing volte a incrementare le sottoscrizioni annuali e a proporre eventuali raccomandazioni operative.

Nelle sottosezioni seguenti verranno analizzate diverse dimensioni d'uso:

- **Panoramica generale (volume di corse e durata totale)**
- **Analisi giorno della settimana e orario**
- **Stagionalità (mensile)**
- **Tipologia di bici (classic, electric, scooter)**
- **Distribuzione per fasce di durata**
- **Approfondimento su mediane e outlier**

Al termine, si trova una **sintesi delle evidenze principali** e successivamente alcune **raccomandazioni** che, pur derivando principalmente dall'analisi della prima domanda, potranno offrire spunti anche per le altre questioni di interesse strategico.

3.1 Panoramica generale

Questa prima panoramica si concentra su:

- **Totale corse (rides)** per tipologia di utente
- **Durata complessiva** delle corse
- **Durata media** tra casual e membri

Principali osservazioni

1. Volume di corse e durata totale

- I membri annuali effettuano **più corse complessive** rispetto ai casual (~3,7 milioni vs. ~2,1 milioni). (Figura 1)
- Tuttavia, i casual riders **totalizzano una durata aggregata maggiore** (~901.727 ore vs. ~789.149 ore), a conferma di corse mediamente più lunghe.

2. Frequenza vs. lunghezza della corsa

- I membri tendono a fare **corse brevi e frequenti** (probabilmente per pendolarismo).
- I casual riders effettuano **meno corse**, ma con tempi significativamente maggiori, indicativi di un utilizzo più ricreativo o turistico. (Figura 2)

Implicazioni di business

- **Piani con durata estesa o pacchetti ibridi** potrebbero attrarre i casual riders, che sfruttano la bici per periodi più lunghi.
- I membri pendolari, già fidelizzati con corse brevi e frequenti, beneficiano di tariffe convenienti per l'uso quotidiano.

(Vedi dettagli e query in Appendice B.1 – “Panoramica generale”)

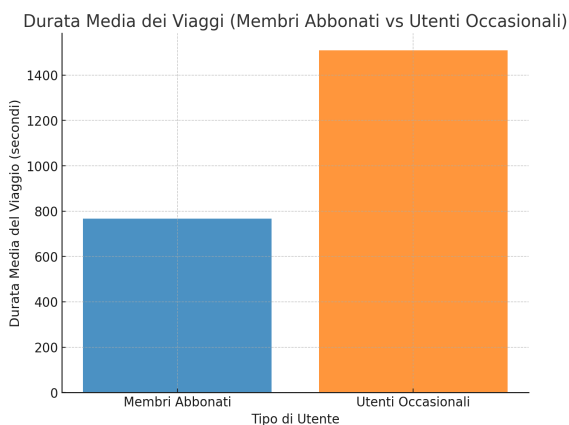


Figura 1 – Durata media delle corse (membri vs. utenti occasionali)

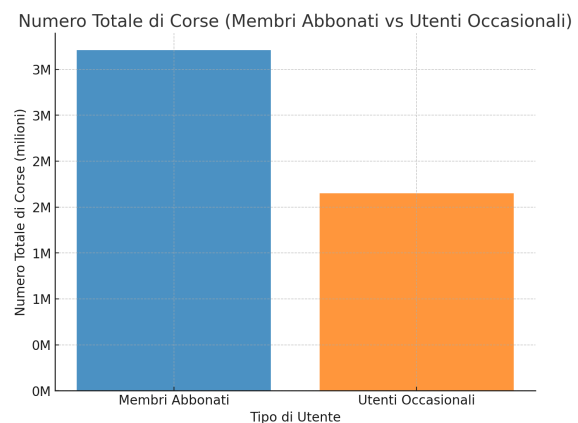


Figura 2 – Numero totale di corse (membri vs. utenti occasionali)

3.2 Analisi giornaliera e oraria

In questa sezione si esamina l'utilizzo del servizio in base ai **giorni della settimana** e alle **fasce orarie**, distinguendo i due segmenti di utenti.

3.2.1 Distribuzione per giorno della settimana

- **Membri**
 - Picco di utilizzo nei **giorni feriali** (martedì, mercoledì, giovedì), in linea con un uso principalmente lavorativo/di pendolarismo. (Figura 3)
 - **Durata media** più contenuta (~12–14 minuti).
- **Casual riders**
 - Prediligono i **weekend** (sabato e domenica), con corse notevolmente più lunghe (~28–30 minuti).
 - Carattere prevalentemente ricreativo/turistico.

Implicazioni

- **Piani weekend** o sconti speciali per convertire i casual riders più attivi nel fine settimana.
- Potenziamento della **disponibilità di bici** nelle zone turistiche nei weekend.

3.2.2 Distribuzione oraria

- **Membri**
 - Due picchi distinti: **mattina (7–9 AM)** e **tardo pomeriggio (4–6 PM)**, riconducibili a spostamenti casa-lavoro.
 - Durata media intorno a ~11–13 minuti.
- **Casual riders**
 - Aumento graduale in tarda mattinata, con **picco tra le 2 e le 5 PM**.
 - Durata spesso più che doppia rispetto ai membri.

Implicazioni

- **Promozioni pomeridiane**: sconti membership nella fascia 2–5 PM per incentivare i casual a passare a un abbonamento.
- **Ottimizzazione flotta**: concentrare bici extra nelle aree centrali/turistiche in orario pomeridiano.

(Vedi dettagli e query in Appendice B.2 – “Analisi giorno/orario”)

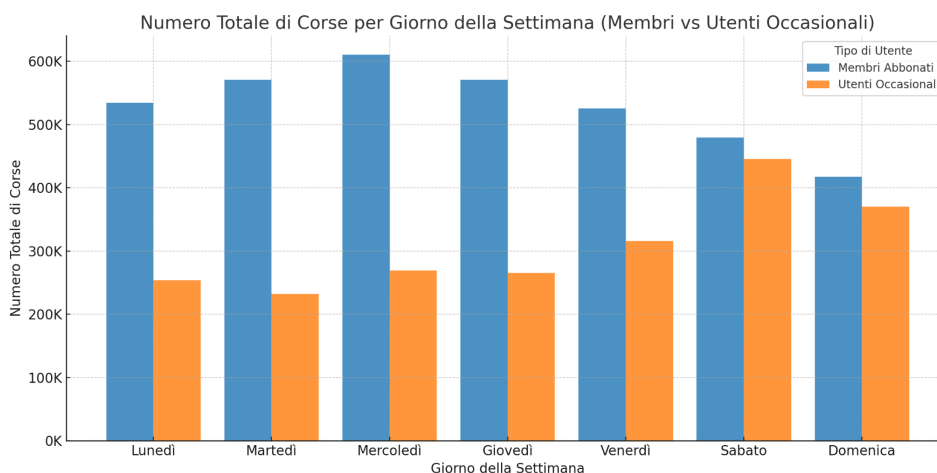


Figura 3 - Numero totale di corse per giorno della settimana (membri vs utenti occasionali)

3.3 Stagionalità (mensile)

Qui si osserva l'utilizzo del servizio su **base mensile**, per evidenziare eventuali variazioni stagionali.

1. Picchi stagionali per i casual riders

- Da maggio a settembre si registra un aumento sostanziale di corse, con un massimo a settembre (~346k corse). (Figura 4)

- La **durata media** rimane sempre più alta rispetto ai membri, confermando l'utilizzo ricreativo/turistico nella bella stagione.

2. Uso stabile dei membri

- I membri mostrano un **andamento più uniforme** durante l'anno, con un picco anch'essi a settembre (~474k corse), ma con **flessione moderata** nei mesi più freddi. (Figura 4)

Implicazioni

- **Promozioni estive:** lanciare abbonamenti stagionali o pass a lungo termine per intercettare i casual riders quando sono più propensi a utilizzare il servizio.
- **Incentivi invernali:** programmi di fidelizzazione o sconti nei mesi freddi per mantenere vivo l'interesse dei casual e incentivare i membri a non interrompere l'abbonamento.

(Vedi dettagli e query in Appendice B.3 – “Analisi mensile”)

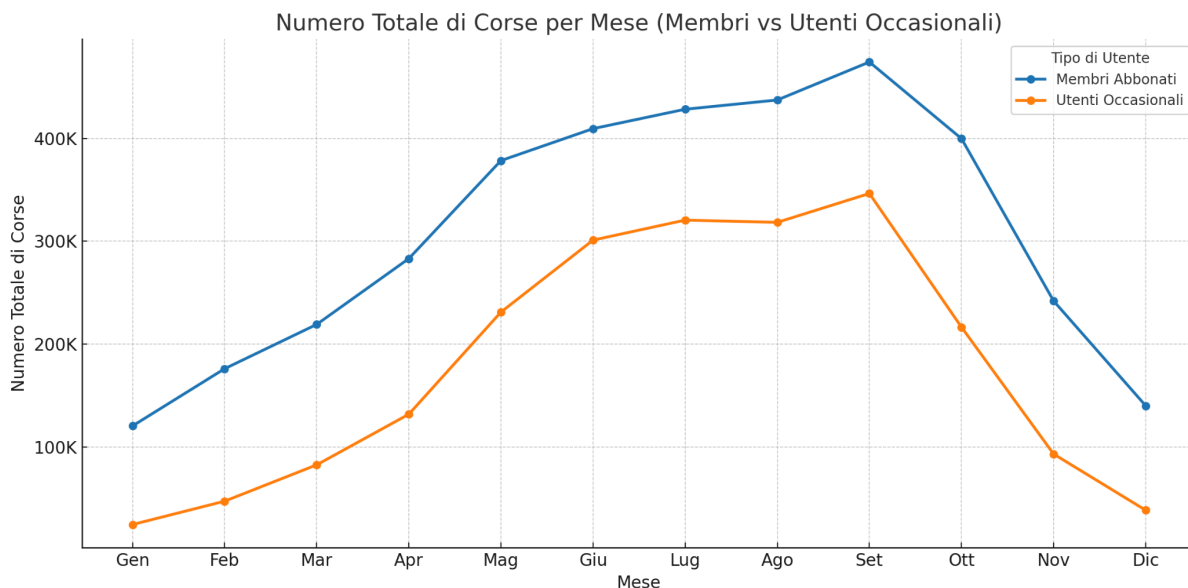


Figura 4 - Numero totale di corse per mese (membri vs utenti occasionali)

3.4 Tipo di mezzo (classic vs. electric vs. scooter)

In questa sezione si esamina la **preferenza** dei due segmenti di utenti (membri e casual riders) per i diversi tipi di bici: **classica (classic_bike)**, **elettrica (electric_bike)** e **scooter (electric_scooter)**.

- **Bici elettriche**
 - La più popolare per entrambi i segmenti. (Figura 5)

- I membri effettuano corse **frequenti ma più brevi** (~11 min di media).
- I casual riders registrano corse leggermente più lunghe (~14–15 min), specialmente nei weekend e nelle ore pomeridiane.
- **Bici classiche**
 - I casual riders fanno corse **molto più lunghe** (fino a ~38 min di media), tipicamente usate per svago/turismo. Si concentrano nei weekend e nelle ore centrali della giornata.
 - I membri mantengono un uso diffuso (14–15 min di media), soprattutto negli spostamenti feriali di pendolarismo (picchi 7–9 AM e 4–6 PM).
- **Scooter**
 - Rappresentano la **tipologia meno diffusa** in assoluto, soprattutto per i membri.
 - I casual riders li utilizzano un po' di più, con medie intorno agli 11–12 min; la preferenza maggiore si riscontra comunque nel weekend e nelle fasce orarie di svago.

Implicazioni di business

- **Offerte speciali su e-bike:** vista la popolarità per tutti, potrebbe essere vantaggioso proporre un piano di membership dedicato alle bici elettriche, specie per i casual rider nel weekend.
- **Pacchetti turistici con bici classiche:** i casual che prediligono corse lunghe (fino a ~38 min) potrebbero essere attratti da **pacchetti** in collaborazione con hotel/agenzie turistiche.
- **Valutazione flotta scooter:** l'uso marginale tra i membri suggerisce di **indagare se mantenere o ridurre** la flotta, oppure se proporre iniziative promozionali mirate a incrementarne l'utilizzo (magari con test gratuiti o sconti specifici).

(Vedi dettagli e query in Appendice B.4 – “Analisi tipo di mezzo”).

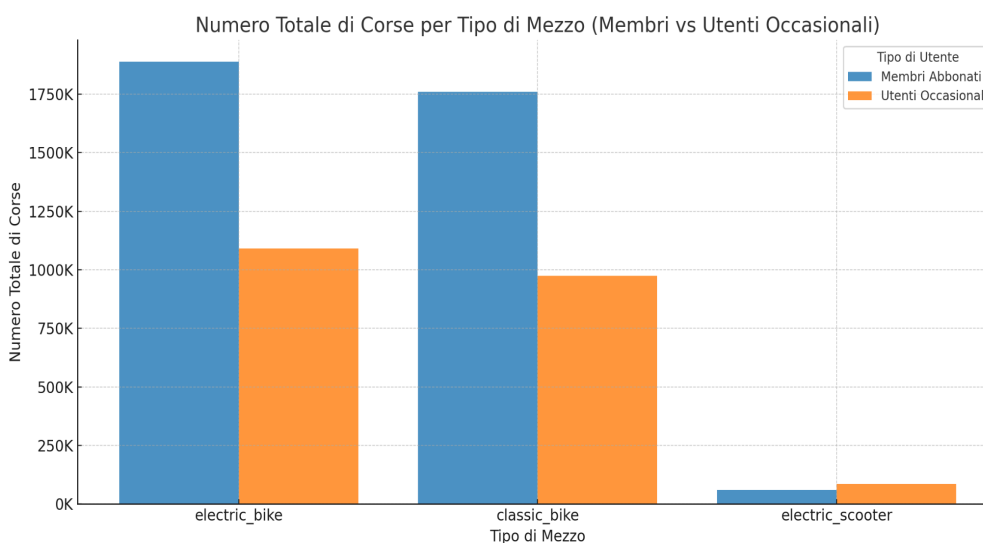


Figura 5 -
Numero totale di corse per mezzo (membri vs utenti occasionali)

3.5 Distribuzione durata corse (fasce di durata)

Questa parte dell'analisi suddivide le corse in fasce di durata prestabilite (<5 min, 5–15 min, 15–30 min, 30–60 min, >60 min) per evidenziare come i due segmenti di utenti sfruttino il servizio in termini di tempo di percorrenza.

Principali osservazioni

- **Membri**
 - Oltre il 75% delle corse dura **meno di 15 minuti**, confermando una forte propensione a spostamenti brevi (pendolarismo).
 - Solo una piccola frazione (>60 min) rappresenta corse molto lunghe (circa 25.969 corse, meno dell'1% del totale), spesso legate a situazioni eccezionali o a un utilizzo sporadico. (Figura 6)
- **Casual riders**
 - Più del **37%** delle corse eccede i 15 minuti, e quasi **130mila** viaggi superano l'ora di durata.
 - Tale distribuzione segnala un uso ricreativo o turistico, con meno vincoli di tempo rispetto ai pendolari.

Implicazioni di business

1. **Piani con ride time esteso**
 - Abbonamenti che coprano corse fino a 45–60 minuti senza sovrapprezzo potrebbero essere accattivanti per i casual riders, soliti fare viaggi più lunghi.
2. **Pricing progressivo**
 - Prevedere una tariffa più elevata oltre una certa soglia (ad es. 30 o 45 min) per i non abbonati. Questo incentiverebbe la sottoscrizione di membership, visto che i casual superano spesso i 30 min.

(Vedi dettagli e query in Appendice B.5 – “Distribuzione durata corse”)

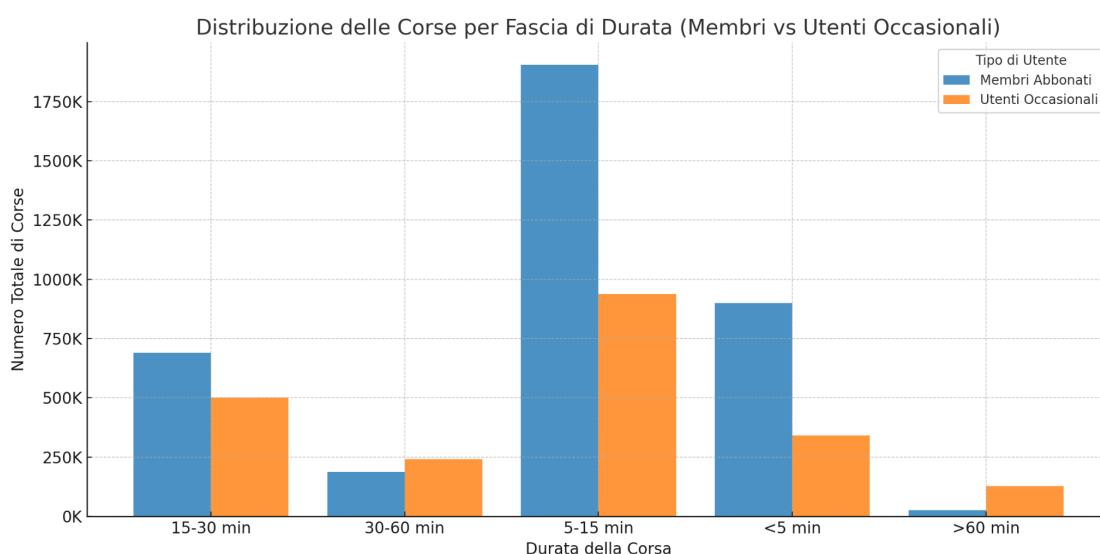


Figura 6 - Distribuzione delle corse per fascia di durata (membri vs utenti occasionali)

3.6 Approfondimento: mediana, percentili e outlier

Per completare l'analisi sulla durata delle corse, si è valutato l'impatto di eventuali **outlier**, ovvero corse eccezionalmente brevi (<1 minuto) o eccezionalmente lunghe (>18 ore). Tali valori possono *distorcere* la media, rendendola meno rappresentativa. In particolare:

- **Corse >18 ore:** Rappresentano circa **0,17%** del totale. Non possiamo escludere a priori che siano tutte dovute a errori tecnici (alcuni pass giornalieri potrebbero teoricamente coprire molte ore), ma il loro numero è esiguo.
- **Corse <1 minuto:** Costituiscono circa **2,2%** del totale. Potrebbero essere sblocchi accidentali o ripensamenti immediati. In assenza di una policy che le indichi come errori, vengono mantenute nel dataset, ma se ne segnala la presenza.

3.6.1 Calcolo della mediana e dei percentili

Oltre alla **media**, si è calcolata la **mediana** (50° percentile) e alcuni **quantili** significativi (25°, 75°, 90°, 95° percentile) tramite la funzione `APPROX_QUANTILES()`. Per semplicità, la query completa è riportata in Appendice B.6 (SQL principali per la mediana e i percentili).

- **Mediana (50° percentile)**
 - *Member:* 521 secondi (~8,7 min)
 - *Casual:* 723 secondi (~12,0 min)
- I casual riders conservano corse più lunghe anche a metà distribuzione.
- **25° e 75° percentile**
 - *Member:* 305 sec (~5 min) e 887 sec (~14,8 min)
 - *Casual:* 401 sec (~6,7 min) e 1364 sec (~22,7 min)
- Il 75% dei casual riders effettua corse fino a ~23 min, contro ~15 min per i membri.
- **Percentili alti (90°, 95°)**
 - *Member:* fino a ~1900 sec (~31,7 min al 95° percentile)
 - *Casual:* ~3960 sec (~66 min al 95° percentile)
- Conferma che esiste una quota di corse molto lunghe in mano ai casual.

Distribuzione della durata delle corse per percentili (Membri Abbonati vs Utenti Occasionali)

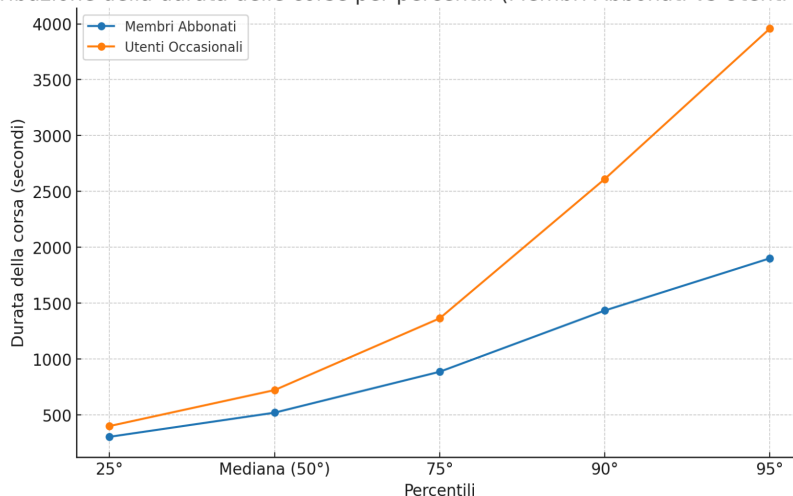


Figura 7 -
Distribuzione
delle corse per
percentili
(abbonati vs
utenti
occasionali)

3.6.2 Sensitivity analysis e impatto degli Outlier

Per valutare l'effetto di valori estremi sulla media delle durate, è stata condotta una breve **sensitivity analysis**:

- **Corse >18h** (circa 0,17% dei record):
Rimuovendole, la media dei casual riders passa da ~1509 secondi (~25 min) a ~1225 secondi (~20 min), mentre quella dei membri subisce variazioni trascurabili.
- **Corse <1 min** (circa 2,2% dei record):
Filtrandole, la media dei casual riders aumenta di poco, ma non capovolge le conclusioni generali sulle differenze tra i due segmenti.

Conclusione:

Anche eliminando gli outlier più estremi (oltre 18h o sotto 1 minuto), la media dei casual riders resta comunque **sensibilmente più elevata** rispetto ai membri, confermando un utilizzo tendenzialmente più prolungato e una varianza maggiore nelle corse dei casual.

4. Sintesi delle evidenze principali

- **Durata delle corse**
 - I **casual riders** effettuano corse mediamente più lunghe in ogni fascia temporale, con una quota significativa (>37%) che supera i 15 minuti e un 5–6% che va oltre i 60 minuti.
 - I **membri** prediligono spostamenti brevi (oltre il 75% delle corse <15 minuti), evidenziando un uso frequente e pendolare.
- **Giorni e orari**
 - I **membri** mostrano picchi nei giorni feriali (martedì, mercoledì, giovedì) e nelle fasce pendolari (7–9 AM, 4–6 PM).
 - I **casual** prediligono i weekend (sabato e domenica), con corse più lunghe e un picco orario tra le 2 e le 5 PM.
- **Stagionalità**
 - Entrambi i segmenti raggiungono il picco di corse a settembre (circa 474k per i membri e 346k per i casual), mostrando un andamento mensile sostanzialmente simile sul fronte dei volumi. Tuttavia, le **durate medie** si distinguono nettamente: i membri mantengono valori piuttosto stabili (intorno a 8–14 min), mentre i casual toccano picchi di oltre 20 min nei mesi estivi, confermando un uso più ricreativo/turistico legato alla bella stagione.
- **Tipologia di mezzo**
 - Le **bici elettriche** sono la scelta più diffusa per entrambi i segmenti, ma i membri le usano in modo più “funzionale” (corse brevi e frequenti).
 - Le **bici classiche** attirano corse molto lunghe dei casual (fino a 38 min di media), specialmente nel weekend.
 - Gli **scooter** restano poco rilevanti per i membri, mentre i casual li impiegano leggermente di più, anche se con durata media inferiore rispetto alle classiche.

Questi risultati, confermati anche dall'analisi dei **quantili** (mediana e percentile), evidenziano due segmenti di clientela con pattern d'uso distinti: da un lato, i casual riders con un forte orientamento leisure/turistico (weekend, estate, corse lunghe), dall'altro i membri annuali più regolari e pendolari (giorni feriali, corse brevi). Tali differenze offrono **spunti decisivi** per la definizione di strategie di marketing, tariffe e promozioni mirate a incrementare le sottoscrizioni annuali.

5. Raccomandazioni e prossimi passi

Alla luce delle differenze emerse tra casual riders e membri annuali, si propongono le seguenti azioni per **massimizzare la conversione** dei casual in abbonati annuali:

1. **Implementare piani weekend e pass estesi**
 - **Motivazione:** I casual riders concentrano l'uso nel fine settimana, spesso con corse più lunghe.
 - **Azione:** Offrire abbonamenti "Weekend Plus" che comprendano corse fino a 60 minuti senza sovrapprezzo, rendendo l'abbonamento più conveniente di un pay-per-ride prolungato.
2. **Promozioni mirate nelle fasce orarie pomeridiane**
 - **Motivazione:** I casual mostrano un picco tra le 2 e le 5 PM, con corse molto lunghe.
 - **Azione:** Lanciare sconti per l'upgrade a un piano annuale se il rider supera regolarmente 30 min di corsa in quella fascia. Un'email automatica o una notifica push potrebbe ricordare il risparmio potenziale di un membership.
3. **Bundle turistici con le bici classiche**
 - **Motivazione:** Le corse su bici classiche sono molto più lunghe per i casual, specialmente nei mesi estivi.
 - **Azione:** Partnership con hotel, o pacchetti "City Tour" integrati con l'abbonamento annuale (o stagionale) per spingere i turisti a optare per un piano più lungo.
4. **Focus su e-bike e commuting**
 - **Motivazione:** Entrambi i segmenti amano le e-bike, ma i membri le sfruttano per brevi tragitti pendolari.
 - **Azione:** Creare un "Commuter Membership" dedicato alle e-bike, con vantaggi specifici (es. priorità nelle ore di punta, tariffe scontate su corse brevi).
5. **Pricing progressivo per le corse più lunghe**
 - **Motivazione:** I casual superano spesso i 15–30 minuti.
 - **Azione:** Applicare una tariffa incrementale oltre i 30 min per i non abbonati, incentivando la sottoscrizione di un abbonamento che copra corse lunghe.

Prossimi passi operativi:

- **Test A/B** su diverse proposte tariffarie (weekend pass, pass estivo, commuter pass).
- **Monitoraggio conversion rate:** confrontare il tasso di upgrade da casual a member nei prossimi 3–6 mesi.
- **Analisi di customer feedback:** raccogliere recensioni, sondaggi e suggerimenti dei clienti sulle nuove offerte.

6. Limitazioni e Sviluppi Futuri

Pur avendo fornito una panoramica esaustiva sul comportamento di casual riders e membri, esistono alcune **limitazioni** da considerare:

1. **Assenza di dati sui costi/ricavi**
 - Non disponiamo di informazioni finanziarie dettagliate per valutare la redditività di ciascun piano. Le raccomandazioni si basano quindi sul volume e sulla durata delle corse, piuttosto che sui margini di profitto.
2. **Dataset anonimo**
 - Non avendo ID utente univoci, non possiamo misurare la fidelizzazione individuale né identificare casual riders che effettuano molte corse singole (potenziali candidati all'abbonamento).
3. **Corse outlier**
 - Le corse oltre 18h e quelle sotto 1 min sono state incluse in parte nell'analisi, a meno di sensibility check. Se fossero errori di registrazione, potrebbero alterare leggermente le medie.
4. **Mancanza di analisi geospaziale**
 - I dati sulle stazioni di partenza/arrivo risultano in parte mancanti; un'analisi più dettagliata delle aree turistiche o residenziali migliorerebbe la comprensione dei flussi di traffico e dell'allocazione delle bici.

Sviluppi futuri:

- **Integrazione con dati di localizzazione e meteo:** per comprendere ancor meglio le dinamiche di utilizzo e la correlazione con le condizioni climatiche.
- **Analisi di churn:** se fosse possibile tracciare gli ID, si potrebbe studiare la percentuale di abbandono (churn) fra i membri e gli occasionali che ripetono corse.
- **Approfondimenti demografici:** eventuali dati su età, genere o reddito potrebbero aiutare a segmentare ulteriormente il pubblico e sviluppare campagne ancora più mirate.

Appendice

Appendice A – Preparazione e pulizia dei dati

Questa sezione descrive i passaggi seguiti per l'integrazione, la pulizia e la preparazione del dataset per l'analisi.

A.1 Creazione della tabella unificata

Per aggregare i 12 dataset mensili in un'unica tabella, è stata utilizzata la seguente query:

```
CREATE OR REPLACE TABLE
`fluted-bot-433911-f3.cyclistic_data_2024.all_trips` AS
SELECT * FROM
`fluted-bot-433911-f3.cyclistic_data_2024.january_2024`
UNION ALL
SELECT * FROM
`fluted-bot-433911-f3.cyclistic_data_2024.february_2024`
UNION ALL
...
UNION ALL
SELECT * FROM
`fluted-bot-433911-f3.cyclistic_data_2024.december_2024`;
```

A.2 Verifica dei record nella tabella unificata

```
SELECT COUNT(*) AS total_rows
FROM `fluted-bot-433911-f3.cyclistic_data_2024.all_trips`;

SELECT *
FROM `fluted-bot-433911-f3.cyclistic_data_2024.all_trips`
LIMIT 10;
```


A.3 Creazione di ride_length e day_of_week

```
CREATE OR REPLACE TABLE
`fluted-bot-433911-f3.cyclistic_data_2024.all_trips_cleaned` AS
SELECT *,
    TIMESTAMP_DIFF(ended_at, started_at, SECOND) AS ride_length,
    FORMAT_TIMESTAMP('%A', started_at) AS day_of_week
FROM `fluted-bot-433911-f3.cyclistic_data_2024.all_trips`;
```

A.4 Rilevazione di valori nulli o negativi

```
SELECT COUNT(*) AS null_or_negative_count
FROM `fluted-bot-433911-f3.cyclistic_data_2024.all_trips_cleaned`
WHERE ride_length <= 0 OR ride_length IS NULL;
```

A.5 Creazione della tabella finale pulita

```
CREATE OR REPLACE TABLE
`fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data` AS
SELECT *
FROM `fluted-bot-433911-f3.cyclistic_data_2024.all_trips_cleaned`
WHERE ride_length > 0;
```

Appendice B

Appendice B.1 - Query per totale corse e durata totale

B.1.1 - Query: statistiche di base (totale corse e durata media per categoria di utilizzatore)

```
SELECT
    member_casual,
    COUNT(*) AS total_rides,
    ROUND(AVG(ride_length), 2) AS avg_ride_length_seconds,
    ROUND(MAX(ride_length), 2) AS max_ride_length_seconds,
    ROUND(MIN(ride_length), 2) AS min_ride_length_seconds
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
```

```
GROUP BY member_casual;
```

B.1.2 - Durata totale dei viaggi per categoria di utilizzatore

```
SELECT
    member_casual,
    COUNT(*) AS total_rides,
    SUM(ride_length) AS total_ride_duration_seconds,
    ROUND(SUM(ride_length) / 3600, 2) AS total_ride_duration_hours
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY member_casual;
```

Appendice B.2 – Analisi giornaliera e oraria

Questa appendice raccoglie le query SQL che hanno permesso di analizzare la distribuzione delle corse per giorno della settimana e per ora del giorno, sia per gli utenti occasionali che per i membri annuali.

B.2.1 – Distribuzione per giorno della settimana

```
SELECT
    member_casual,
    day_of_week,
    COUNT(*) AS total_rides,
    ROUND(AVG(ride_length), 2) AS avg_ride_length_seconds
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY member_casual, day_of_week
ORDER BY member_casual, day_of_week;
```

B.2.2 – Distribuzione oraria (totale corse)

```
SELECT
    EXTRACT(HOUR FROM started_at) AS hour_of_day,
    member_casual,
    COUNT(*) AS total_rides
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY hour_of_day, member_casual
ORDER BY hour_of_day, member_casual;
```

B.2.3 – Distribuzione oraria (durata media)

```
SELECT
    EXTRACT(HOUR FROM started_at) AS hour_of_day,
    member_casual,
    ROUND(AVG(ride_length), 2) AS avg_ride_length_seconds
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY hour_of_day, member_casual
ORDER BY hour_of_day, member_casual;
```

B.2.4 – Weekday vs weekend

```
SELECT
    CASE
        WHEN day_of_week IN ('Saturday', 'Sunday') THEN 'Weekend'
        ELSE 'Weekday'
    END AS day_type,
    EXTRACT(HOUR FROM started_at) AS hour_of_day,
    member_casual,
    COUNT(*) AS total_rides,
    ROUND(AVG(ride_length), 2) AS avg_ride_length_seconds
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY day_type, hour_of_day, member_casual
ORDER BY day_type, hour_of_day, member_casual;
```

Appendice B.3 – Analisi mensile

Questa appendice riporta la query SQL utilizzata per analizzare come il numero di corse e la durata media variano di mese in mese per i due segmenti di utenti.

```
SELECT
    EXTRACT(MONTH FROM started_at) AS month,
    member_casual,
    COUNT(*) AS total_rides,
    ROUND(AVG(ride_length), 2) AS avg_ride_length_seconds
FROM
    `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY
    month, member_casual
```

```
ORDER BY
    month, member_casual;
```

Appendice B.4 – Analisi tipo di mezzo

B.4.1 – Utilizzo delle diverse tipologie di bici

```
SELECT
    rideable_type,
    member_casual,
    COUNT(*) AS total_rides,
    ROUND(AVG(ride_length), 2) AS avg_ride_length_seconds
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY rideable_type, member_casual
ORDER BY total_rides DESC;
```

B.4.2 – Differenze orarie e giornaliere per tipologia di bici

```
SELECT
    rideable_type,
    member_casual,
    CASE
        WHEN day_of_week IN ('Saturday', 'Sunday') THEN 'Weekend'
        ELSE 'Weekday'
    END AS day_type,
    EXTRACT(HOUR FROM started_at) AS hour_of_day,
    COUNT(*) AS total_rides,
    ROUND(AVG(ride_length), 2) AS avg_ride_length_seconds
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY rideable_type, member_casual, day_type, hour_of_day
ORDER BY rideable_type, member_casual, day_type, hour_of_day;
```

Appendice B.5 – Distribuzione durata corse

```
SELECT
  member_casual,
  CASE
    WHEN ride_length < 300 THEN '<5 min'
    WHEN ride_length BETWEEN 300 AND 900 THEN '5-15 min'
    WHEN ride_length BETWEEN 900 AND 1800 THEN '15-30 min'
    WHEN ride_length BETWEEN 1800 AND 3600 THEN '30-60 min'
    ELSE '>60 min'
  END AS duration_category,
  COUNT(*) AS total_rides
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY member_casual, duration_category
ORDER BY member_casual, duration_category;
```

Appendice B.6 – SQL principali per la mediana e i percentili

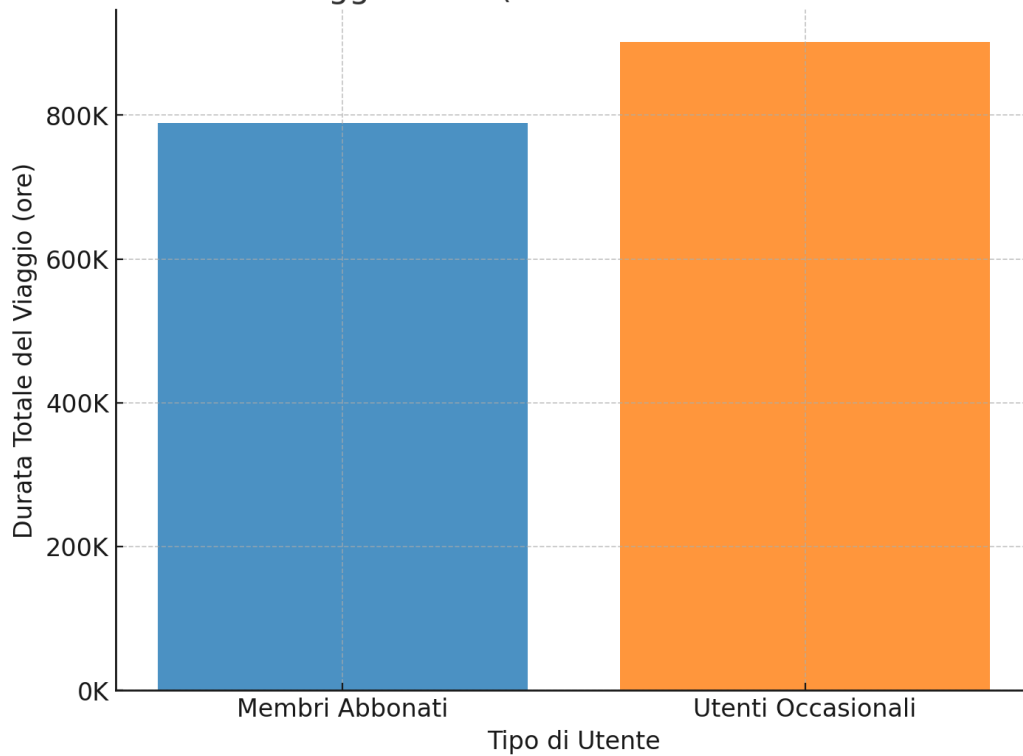
```
SELECT
  member_casual,
  APPROX_QUANTILES(ride_length, 100)[OFFSET(50)] AS
median_ride_length,
  APPROX_QUANTILES(ride_length, 100)[OFFSET(25)] AS percentile_25,
  APPROX_QUANTILES(ride_length, 100)[OFFSET(75)] AS percentile_75,
  APPROX_QUANTILES(ride_length, 100)[OFFSET(90)] AS percentile_90,
  APPROX_QUANTILES(ride_length, 100)[OFFSET(95)] AS percentile_95
FROM `fluted-bot-433911-f3.cyclistic_data_2024.cleaned_data`
GROUP BY member_casual;
```

Visualizzazioni aggiuntive

Sezione B1

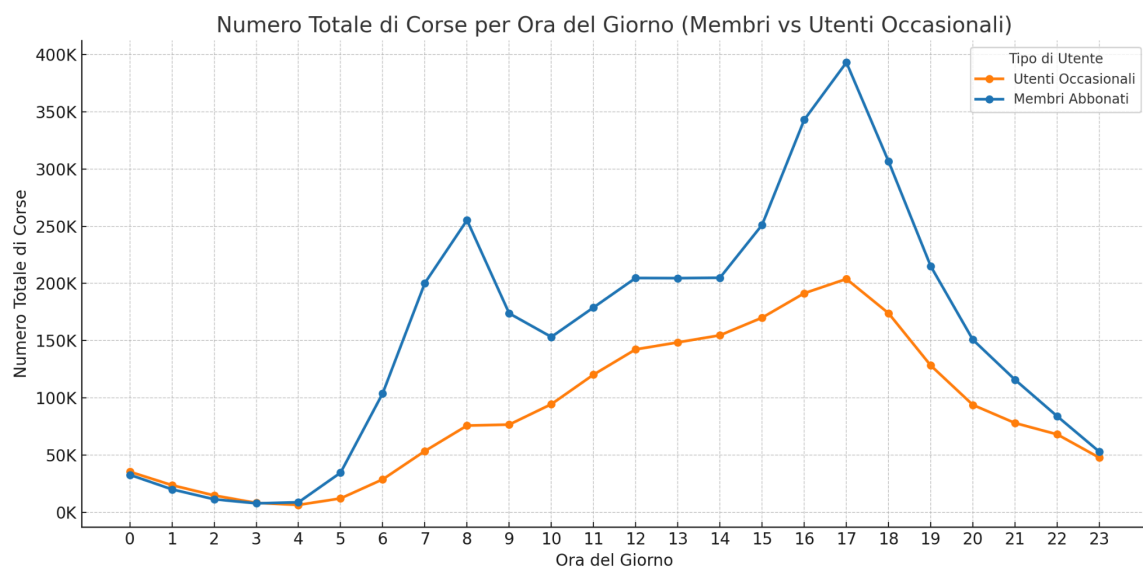
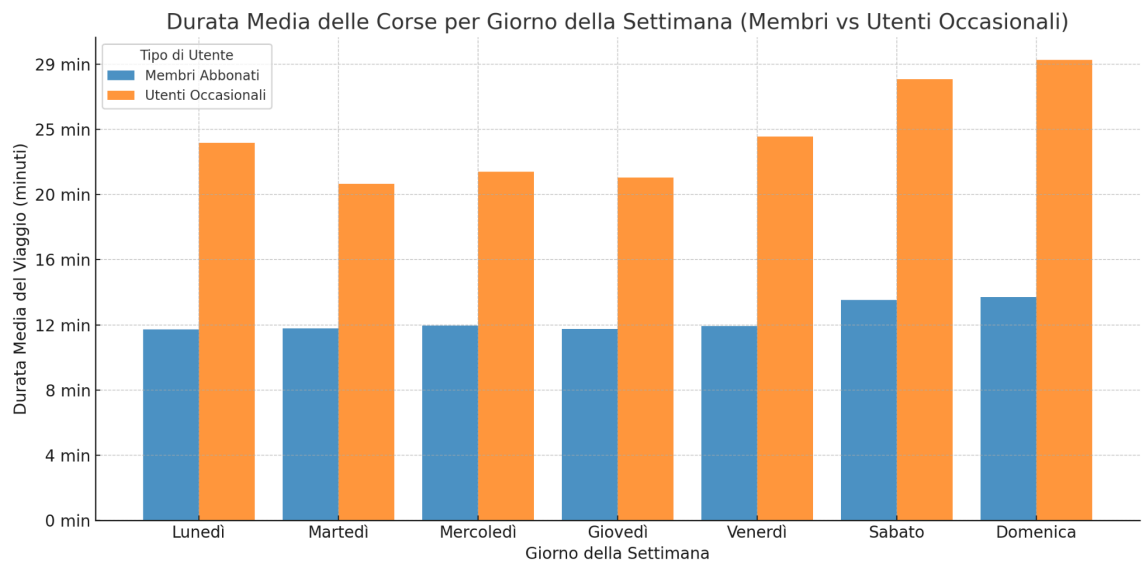
Il grafico mostra la durata totale delle corse per membri e casual riders. Nonostante i membri effettuino più corse, i casual riders accumulano una durata complessiva maggiore, suggerendo un uso più ricreativo/turistico.

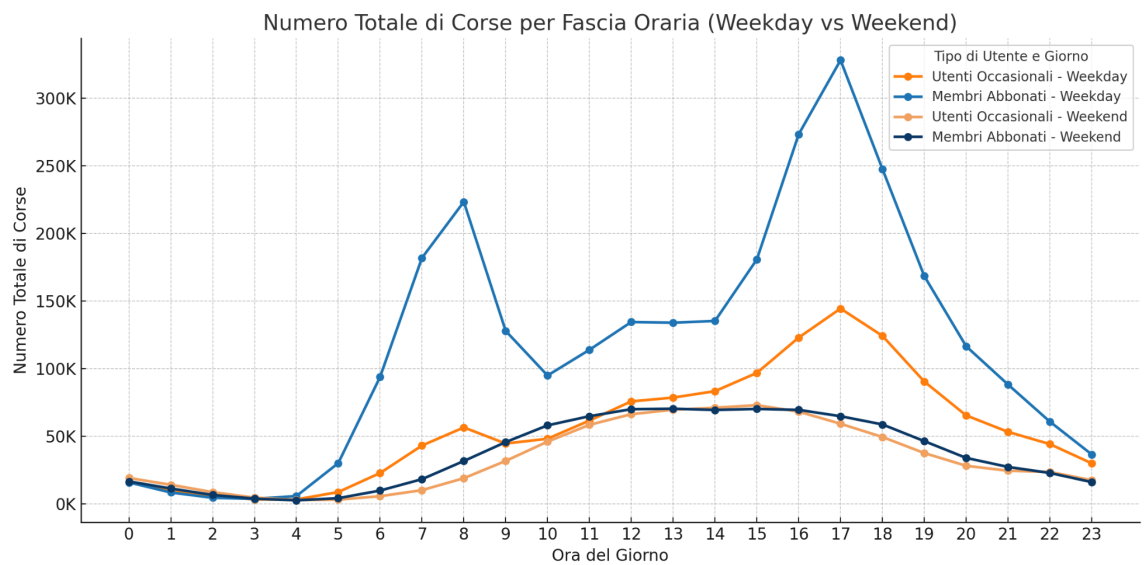
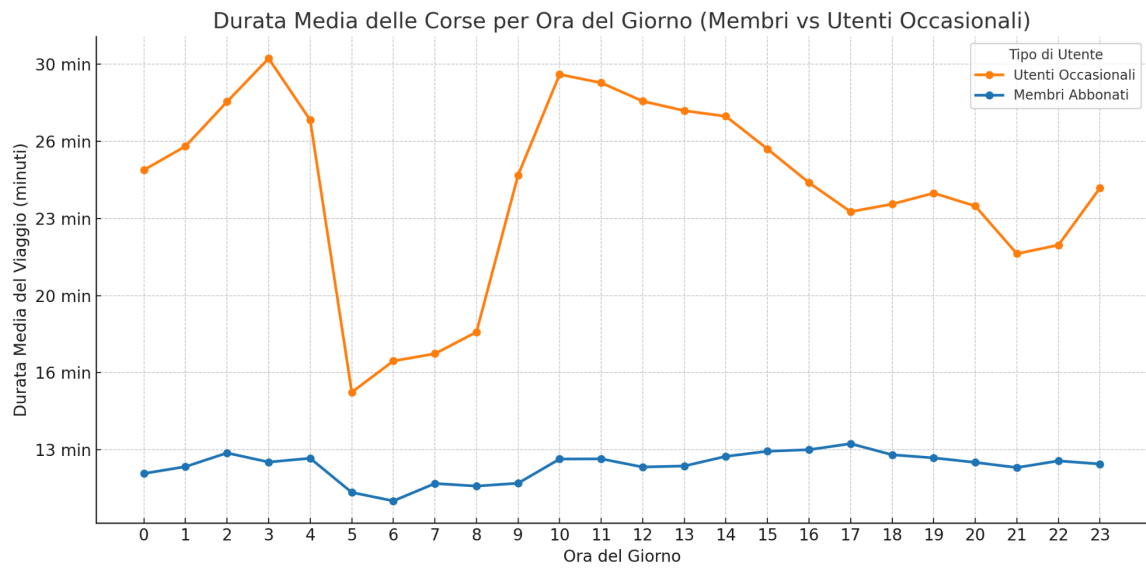
Durata Totale dei Viaggi in Ore (Membri Abbonati vs Utenti Occasionali)



Sezione B2 - Analisi giornaliera e oraria

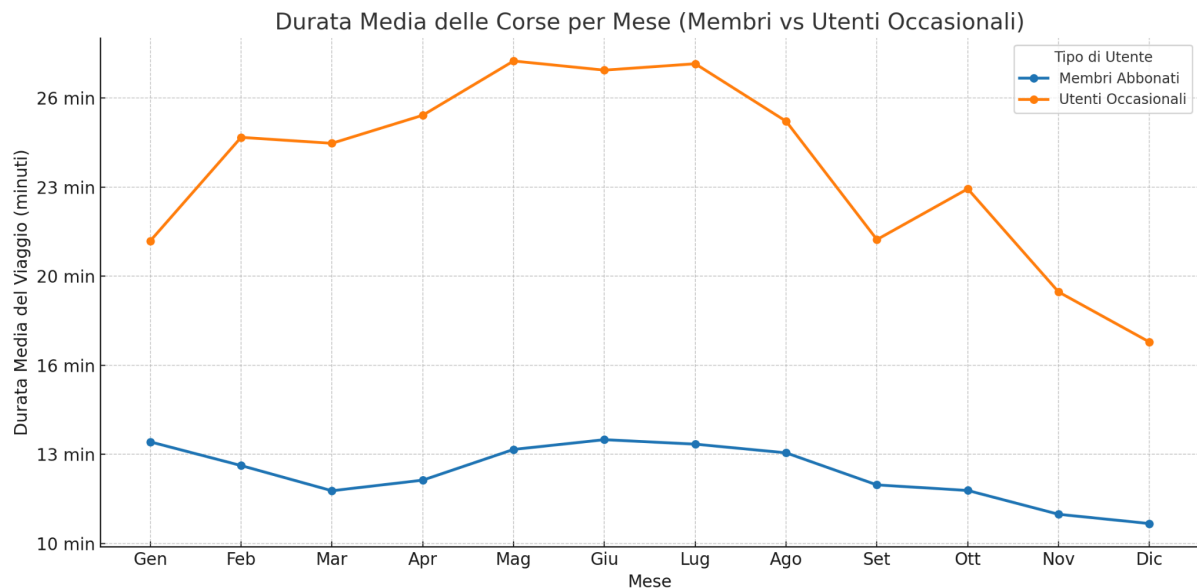
Questa sezione presenta una serie di visualizzazioni che illustrano le differenze nell'utilizzo del servizio tra membri e casual riders in base ai giorni della settimana e alle fasce orarie.





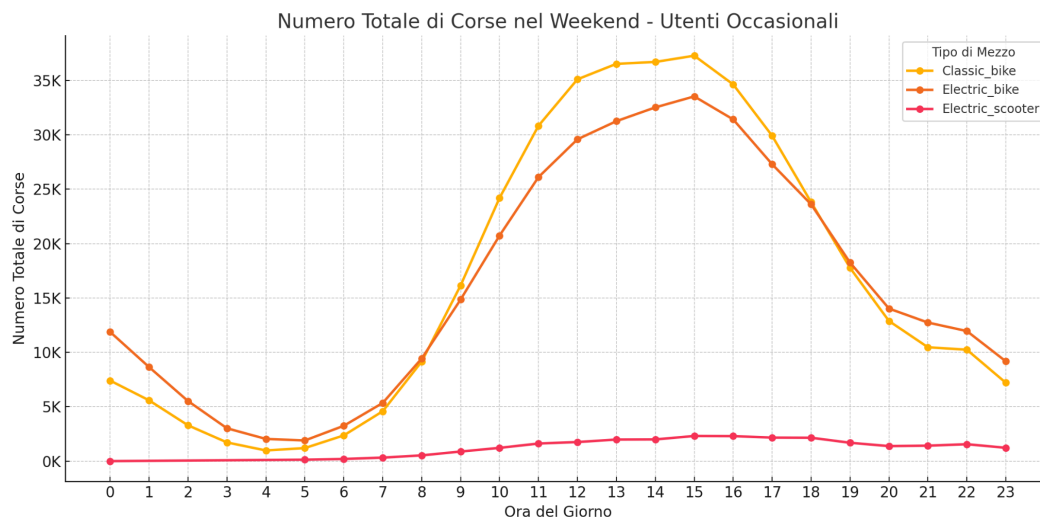
Sezione B3 – Analisi mensile (stagionalità)

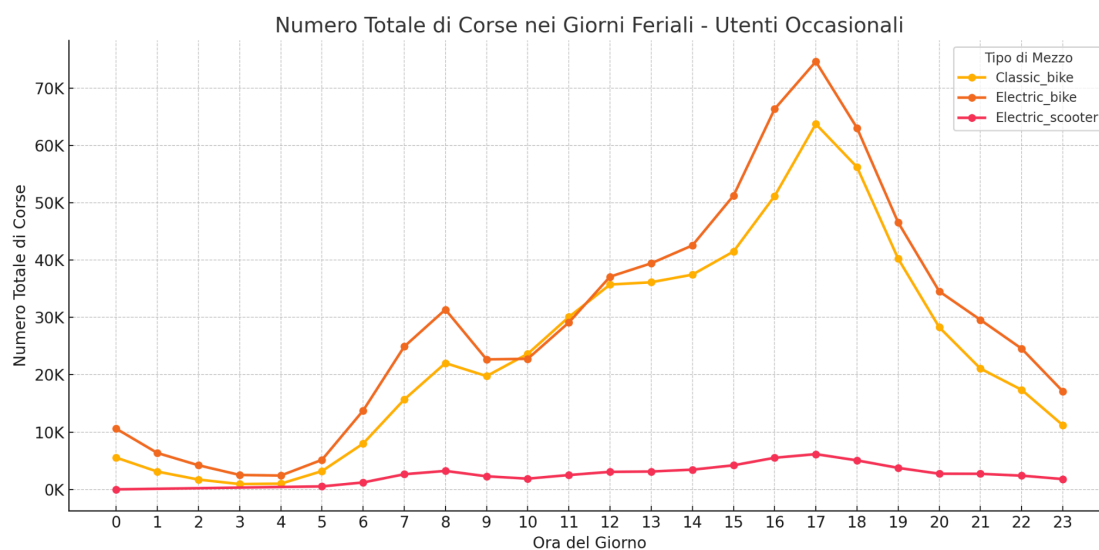
Questa visualizzazione mostra la variazione della durata media delle corse nel corso dell'anno, evidenziando un aumento nei mesi più caldi, soprattutto tra i casual riders. I membri mantengono una durata più stabile.



Sezione B4 - Analisi tipologia di bici

Questa sezione presenta visualizzazioni sull'uso dei diversi tipi di bici (classiche, elettriche, scooter) da parte di membri e casual riders, distinguendo tra giorni feriali e weekend. Inoltre, evidenzia le differenze nella durata media delle corse per ciascuna tipologia di bici.





Sezione B5 - Distribuzione durata corse

Visualizzazione della durata delle corse per fasce di tempo. I membri effettuano per lo più corse brevi (5–15 min), mentre i casual riders registrano viaggi più lunghi, con molti oltre i 30 minuti.

