

# Può la paura del contagio portarci ad essere più sostenibili?

## Analisi dell'utilizzo del Bike Sharing durante il periodo del Covid-19 a Londra

Martina Roberta Cecchetto<sup>1</sup>, Giuseppe Monea<sup>2</sup>, Valentina Nelli<sup>3</sup>

### Sommario

Il fenomeno del bike sharing nasce da un nuovo modello di economia sviluppatosi nell'ultimo decennio, la cosiddetta sharing economy, grazie al quale le persone hanno iniziato ad apprezzare il potere della condivisione e dello scambio, creando un modello alternativo al consumismo. Come ogni modello economico è strettamente connesso all'aspetto sociale e in quanto tale, nell'ultimo anno, a causa del SARS-CoV-2, si è avuto modo di osservare come fosse suscettibile a paure, bisogni e abitudini della popolazione. Il progetto si propone quindi di evidenziare l'influenza che la recente pandemia ha avuto sull'utilizzo del servizio di bike-sharing nella città di Londra. Per adempiere a tale scopo sono stati raccolti e successivamente analizzati i dati dei viaggi relativi al periodo 'Gennaio 2018 - Luglio 2020'. Utilizzando la piattaforma Kafka è stato possibile simulare la velocità per ottenere le informazioni riguardanti i viaggi dell'ultima settimana di Luglio 2020. I dati così raccolti sono stati integrati a quelli riguardanti i viaggi meno recenti su MongoDB, soddisfacendo così i criteri di volume. In tal modo è stato quindi possibile realizzare delle infografiche tramite il software Tableau, grazie alle quali si è osservato che effettivamente la pandemia ha modificato il comportamento degli utenti sia per quanto riguarda il numero di viaggi effettuati sia per la loro durata.

### Keywords

Bike Sharing — Sharding — MongoDB — Kafka — Tableau

<sup>1</sup> Data Science M.Sc. matr. 852566

<sup>2</sup> Data Science M.Sc. matr. 850432

<sup>3</sup> Data Science M.Sc. matr. 860613

Indice			
1	Introduzione	2	4
2	Obiettivo e ricerca dati	2	5
3	Architettura e approccio metodologico	3	6
3.1	Volume	3	
	Sharding		
3.2	Raccolta dati	3	
3.3	Kafka	4	
4	Analisi dei risultati ottenuti		4
5	Sviluppi futuri		5
6	Fonti		5

## 1. Introduzione

Il progetto ha preso vita a partire dai seguenti quesiti:

- Quali effetti ha avuto la recente pandemia sulla mobilità urbana di Londra, in particolare sul fenomeno del bike sharing?
- La paura del contagio ha portato i cittadini ad essere più ecosostenibili?

L'emergenza sanitaria dettata dal coronavirus ha portato infatti a una revisione della mobilità urbana in tutto il mondo. Ad esempio, nel Regno Unito il governo ha annunciato un pacchetto da due miliardi di sterline volto a potenziare la mobilità pedonale e ciclabile, nonché la promozione di veicoli elettrici e sistemi di noleggio di bici ed e-scooter. D'altra parte c'è da considerare che il lockdown ha ridotto gli spostamenti e sempre più persone sono rimaste a casa. Quindi, nel periodo di pandemia, il fenomeno di bike sharing ha subito variazioni in positivo o in negativo? Al fine di trovare una risposta a questa domanda è stato necessario raccogliere i dati presso l'ente che fornisce il trasporto del bike sharing di Londra. Tali dati, tuttavia, si presentavano in maniera aggregata e non particolarmente organizzata, tramite dei csv. Per bypassare il problema della gestione dei singoli csv è stato utilizzato un algoritmo di scraping che ha permesso di semplificare l'acquisizione dei dati. Visto che il lavoro svolto si propone di creare un sistema che permetta l'acquisizione dei dati e la loro lettura in real time, la presenza dei dati solamente a consuntivo non permetteva di lavorare su strumenti che si basano sul principio di Velocity dei big data. Per questo la velocità è stata simulata tramite Kafka, una piattaforma per la gestione di dati in tempo reale, facendo sì che il producer acquisisse i dati ottenuti tramite scraping e li passasse direttamente al consumer che a sua volta sarebbe stato collegato al sistema per la memorizzazione dei dati. Questo ultimo punto è stato realizzato tramite MongoDB, un sistema di gestione basato sui documenti (Document Based Management System, DBMS), in cui i dati vengono archiviati in formato BSON (Binary JSON) e letti tramite indici. Infine, sono state realizzate delle infografiche tramite il software Tableau, adatto

alla visualizzazione interattiva delle infografiche realizzate con i dati precedentemente raccolti.

## 2. Obiettivo e ricerca dati

Il servizio di bike sharing di Londra offre la possibilità di noleggiare una bici da una flotta di circa 11500 presso una delle oltre 750 docking stations situate in particolare nella parte centrale della capitale inglese. Il progetto è iniziato osservando l'utilizzo del servizio a seconda delle fasce orarie e delle stazioni, e cercando la sua crescita potenziale nel tempo, per poi focalizzarsi sull'andamento durante il 2020: lo scopo finale era, infatti, quello di analizzare l'impatto della pandemia di SARS-CoV-2 sui noleggi. La domanda che ci si è posti è, quindi, come il Covid-19 abbia influito sulle abitudini di utilizzo del servizio di bike sharing nella città di Londra. Per poter svolgere le analisi sono stati individuati i dati relativi ai noleggi delle bici a partire da gennaio 2018. I dati relativi ai viaggi sono stati raccolti dal sito di "Transport for London" (TfL). Ogni viaggio costituisce una singola riga del file e si presenta con i seguenti attributi:

- Rental id: codice univoco del viaggio;
- Duration: durata del viaggio espressa in secondi;
- Bike id: codice univoco della bici noleggiata;
- End Date: data di fine noleggio;
- EndStation id: docking station di fine noleggio;
- EndStation Name: nome del docking station di fine noleggio;
- Start Date: data di inizio noleggio;
- StartStation id: docking station di inizio noleggio;
- StartStation Name: nome del docking station di inizio noleggio;

Per poter analizzare i dati è stato necessario utilizzare anche un file che contenesse le informazioni relative alle singole docking stations.

Per ogni stazione è presente:

- Id: codice univoco;
- Common name: nome completo;

- Lat: latitudine;
- Lon: longitudine;
- Installed: True/False a seconda della presenza effettiva della docking station;
- Locked: True/False a seconda dello stato della docking station (True corrisponde a una stazione bloccata);
- Instal Date: data di installazione;
- Removal ID: codice presente nel caso di rimozione della stazione;
- Temporary: True/False a seconda dell'utilizzo solamente temporaneo della stazione;
- NbDocks: numero di postazioni (docks) per ogni stazione.

Ottenuti anche questi dati si è potuto procedere con la fase di preprocessing. I dati ottenuti dal sito del TfL non hanno richiesto particolari interventi. Nel file sulla posizione delle docking stations, invece, è stato necessario correggere la longitudine di alcune stazioni. Un numero notevole di esse infatti risultava locato nella campagna canadese invece che nel centro di Londra. È stato così rilevato che la longitudine di alcune di esse era “106” invece che “0,106”. In questo modo è stato possibile localizzarle geograficamente e poi rappresentarle tramite una mappa con il software Tableau.

### 3. Architettura e approccio metodologico

Per rispondere alle domande che ci si è posti, è stato necessario acquisire una grande quantità di dati e pensare ad un metodo per analizzarli in tempo reale o, quanto meno, in un lasso temporale adeguato che, ai fini dell'analisi, si è ritenuto essere un numero randomico tra 1 e 10 secondi. Per i suddetti motivi, si è deciso di affrontare il progetto con particolare attenzione a due delle V caratteristiche dei BIG Data. La prima V è quella del Volume, poiché nonostante i dati fossero solamente in formato testuale la loro quantità era più che notevole e abbondantemente in grado di superare i 2GB richiesti. La seconda V è quella di Velocità poiché i dati, pur se attraverso una simulazione, sono stati trattati come se incorporati con cadenza oraria.

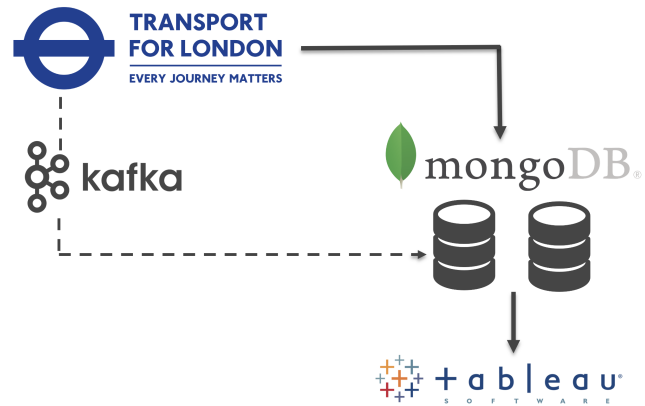


Figura 1. Architettura

#### 3.1 Volume

##### 3.1.1 Sharding

A causa dell'ingente volume di dati (3 gb) che avremmo dovuto immagazzinare si è ritenuto necessario effettuare l'operazione di sharding. Per effettuare tale operazione è stato utilizzato il database NoSQL MongoDB configurato in modalità sharding. Grazie a questa configurazione il carico di lavoro è stato così distribuito da MongoDB su più server.

La simulazione dell'architettura è stata sviluppata in locale tramite dei container docker, in questo modo ogni singola macchina è rappresentata dai diversi container su cui è attiva una istanza di MongoDB. Abbiamo perciò deciso di creare un cluster di due shard, ognuno contenente tre nodi. In questo modo l'architettura risulta facilmente scalabile e permette di integrare anche informazioni successive alla data del 4 agosto 2020.

#### 3.2 Raccolta dati

Una volta configurato il nostro database in modalità sharding il primo passo per ottenere i dati necessari alle analisi preposte è stato quello di acquisirli dal sito del gestore TfL (Transport for London).

Per ottenere il più rapidamente possibile tutti i file necessari si è realizzato un algoritmo di scraping che raccoglieva solamente i file relativi ai journeys con date successive al 1 gennaio 2018 fino al 28 Luglio 2020, escludendo i file contenenti informazioni non inerenti. Lo scraping è stato realizzato utilizzando sia la libreria selenium (utilizzando il webdriver chromedriver per la sua versatilità con

il browser Chrome) sia beautiful soup (in quanto concilia semplicità d'uso a dei risultati efficienti). Durante questo passaggio ci si è confrontati con problemi nel riconoscere tutti i file utili, poiché alcuni erano salvati in maniera arbitraria rispetto allo standard utilizzato per la maggior parte.

Nello specifico, lo standard era il seguente:

- 216JourneyDataExtract27May2020-02Jun2020.csv

In alcuni casi, invece, il file si presentava con una numerazione differente all'inizio del nome oppure con un uso di simboli di spaziatura differente rispetto al '-', come ad esempio:

- 02bJourneyDataExtract21Feb16-05Mar2016.csv

Una volta ottenuti i file, si è costruito un unico dataframe e si è effettuato l'inserimento sul database configurato in modalità sharding su MongoDB.

### 3.3 Kafka

Per i dati del periodo 29 Luglio 2020 – 4 Agosto 2020 si è deciso di simulare la velocità tramite il software Kafka. È stato quindi creato il topic "bike journeys" e un producer che mandasse a periodi randomici, di valore compreso tra 1 e 10 secondi, i dati dei viaggi ad un consumer che successivamente li caricava su MongoDB.

Questo procedimento è stato realizzato in quanto, in condizioni di lavoro ideali, i dati sarebbero stati caricati man mano che venivano prodotti, rendendo quindi il tempo di caricamento nettamente inferiore rispetto a quello che sarebbe stato necessario per caricarli in maniera simulata tutti in una volta.

I dati ottenuti si presentavano già adatti alle analisi che sarebbero state svolte ed erano raccolti come dataframe le cui singole righe si riferivano al singolo viaggio.

2. La durata media di un viaggio è rimasta costante?
3. Se c'è stata una variazione, quali sono le docking station che hanno subito una maggiore perdita o un maggiore incremento?

Analizzando i dati e le visualizzazioni si è notato come il bike sharing sia un fenomeno che presenta una forte stagionalità, dovuta chiaramente ai fenomeni atmosferici come ad esempio la temperatura fredda dei periodi invernali. Se ci si concentra sui primi sette mesi dell'anno appare chiaro che il diffondersi della pandemia ha frenato drasticamente l'utilizzo della bici a partire da febbraio fino ad aprile. Da maggio, dall'allentamento del lockdown, le persone hanno ripreso a utilizzare il servizio e persino più all'anno precedente senza però tornare sui livelli del 2018. Ciò è dovuto, probabilmente, alla voglia di tornare a uscire e svagarsi all'aperto dopo le settimane in cui tendenzialmente si era rimasti in casa a causa dello spargersi del virus.

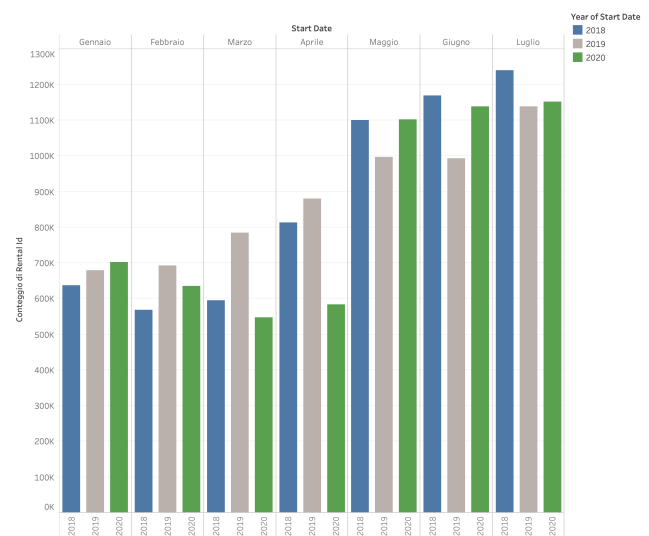


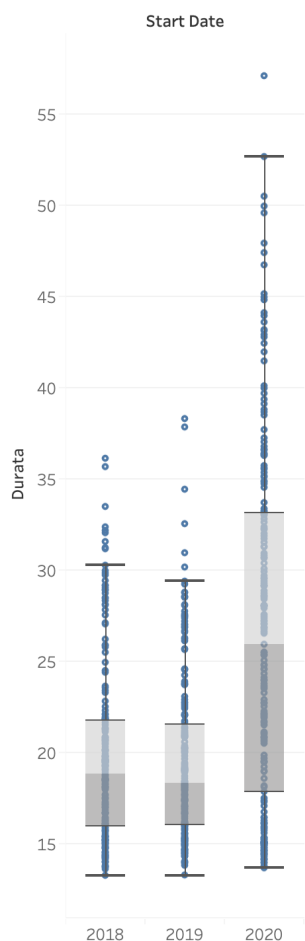
Figura 2. Andamento totale viaggi gennaio-luglio

## 4. Analisi dei risultati ottenuti

Per rispondere alle domande poste all'inizio le analisi si sono concentrate principalmente su tre aspetti:

1. Il numero di viaggi in totale del periodo Gennaio-Agosto è variato?

Quello del numero di viaggi, tuttavia, non è l'unico aspetto interessante da analizzare. Un'altra variabile particolarmente utile si è rivelata essere la durata dei viaggi. Per studiarla si è calcolata per ogni giorno la durata media di un singolo viaggio.



**Figura 3.** Boxplot durata media viaggio

Come emerge dal boxplot nel 2020 la durata media si sposta nettamente verso l'alto. Si suppone che questa variazione sia conseguenza di un utilizzo diverso del bike sharing rispetto agli anni precedenti, magari in sostituzione a forme di trasporto più veloci come metro o bus, per effettuare viaggi più lunghi. I dati sono anche ben distribuiti.

Da questi risultati abbiamo ricavato tre infografiche (Figura 4, Figura 5, Figura 6).

Si può trovare la versione interattiva al link:

- [https://public.tableau.com/profile/martina.roberta.cecchetto#!/vizhome/Finale\\_15996730766570/Story1](https://public.tableau.com/profile/martina.roberta.cecchetto#!/vizhome/Finale_15996730766570/Story1)

Dalle mappe in particolare si nota che i viaggi hanno

subito un calo nelle zone centrali, dove in condizioni standard il grosso degli spostamenti è dovuto a motivi turistici o per recarsi al lavoro. In queste zone si concentrano infatti la maggior parte degli uffici, delle stazioni e dei punti di interesse turistico. Con la presenza del Covid-19, invece, i maggiori spostamenti sono stati effettuati nelle zone più esterne del servizio ricche di zone residenziali e di punti verdi.

## 5. Sviluppi futuri

I risultati ottenuti mostrano chiaramente come il Covid-19 abbia alterato alla radice le abitudini degli abitanti e dei turisti, cambiando l'uso che veniva fatto del servizio di bike sharing. Sarebbe interessante, pertanto, continuare ad analizzare questo fenomeno per tutto il 2020. Si potrebbe così capire se effettivamente questi cambiamenti siano solamente momentanei.

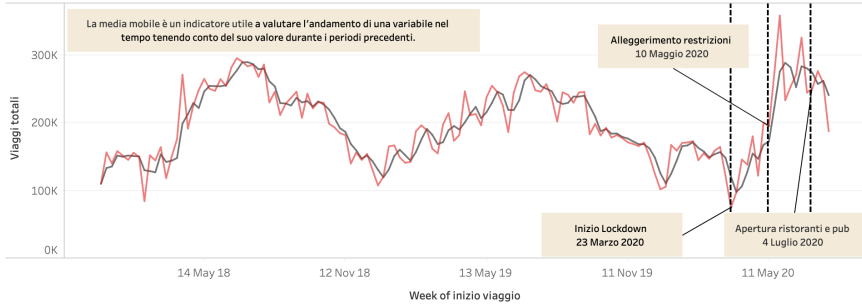
Un ulteriore spunto di analisi sarebbe fornito inoltre dall'integrazione con i dati di utilizzo degli altri mezzi pubblici per poter stabilire meglio quanto, oltre alle misure restrittive sulla circolazione dovute alla pandemia, si debba considerare il fattore psicologico; questo potrebbe aver portato gli utenti a preferire il bike sharing a mezzi più affollati. Continuando l'analisi quando la paura del virus andrà scemando, si capirà quale può essere il futuro della mobilità urbana.

Parallelamente potrebbe essere interessante studiare con lo stesso metodo le abitudini negli spostamenti per quanto riguarda le altre grandi città dove mobilità, inquinamento e ripresa post Covid-19 sono temi attuali e di grande importanza, come nella città di Londra.

## 6. Fonti

- <https://github.com/kayne87/mongodb-sharding-docker>
- <https://tfl.gov.uk/modes/cycling/santander-cycles>
- <https://docs.mongodb.com/manual/sharding/>

## Andamento dei viaggi totali in bici



Docking station di partenza  
(All)

Docking station di arrivo  
(All)

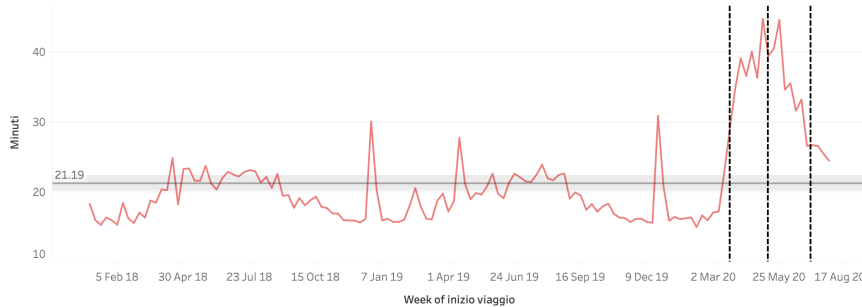
I grafici presenti mostrano l'andamento dell'utilizzo totale delle bici e della durata media di un singolo viaggio da gennaio 2018 a fine luglio 2020.

I primi due anni mostrano un andamento simile con una certa stagionalità e dei picchi simili, corrispondenti al periodo natalizio (in cui i mezzi pubblici effettuano meno corse o sono assenti). Nel 2020 l'utilizzo è nettamente inferiore rispetto agli anni precedenti.

Tuttavia questo andamento appare differente se viene analizzata la durata media del singolo viaggio.

Nella storia successiva sarà possibile comprendere quanto effettivamente tale aumento, in termini percentuali, sia importante rispetto all'andamento dell'anno precedente.

## Andamento della durata media di un viaggio singolo



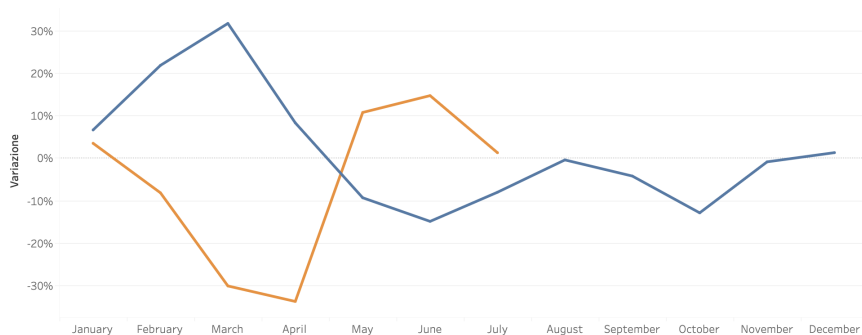
**Legenda**

Grafico 1  
 ■ Viaggi effettuati  
 ■ Media mobile viaggi effettuati

Grafico 2  
 ■ Durata media

Figura 4. Prima infografica

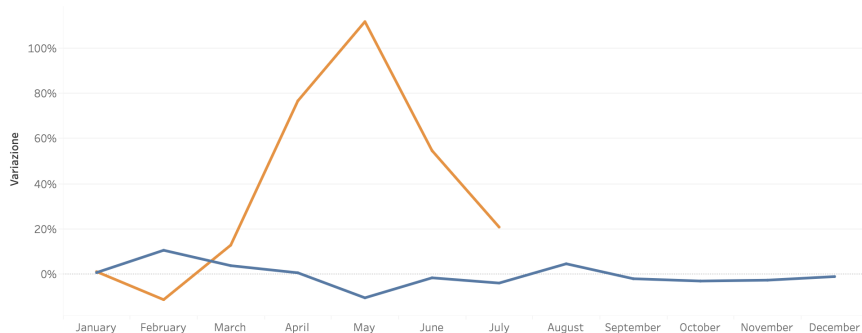
## Variazione dell'utilizzo totale in % rispetto agli anni precedenti



Su questa infografica è possibile visualizzare l'andamento della variazione in termini percentuali dell'utilizzo totale del bike sharing e la durata media di un singolo viaggio rispetto all'anno precedente.

Si osserva come il Covid-19 abbia impattato sull'utilizzo di tale servizio.

## Variazione durata media del singolo viaggio in % rispetto agli anni precedenti

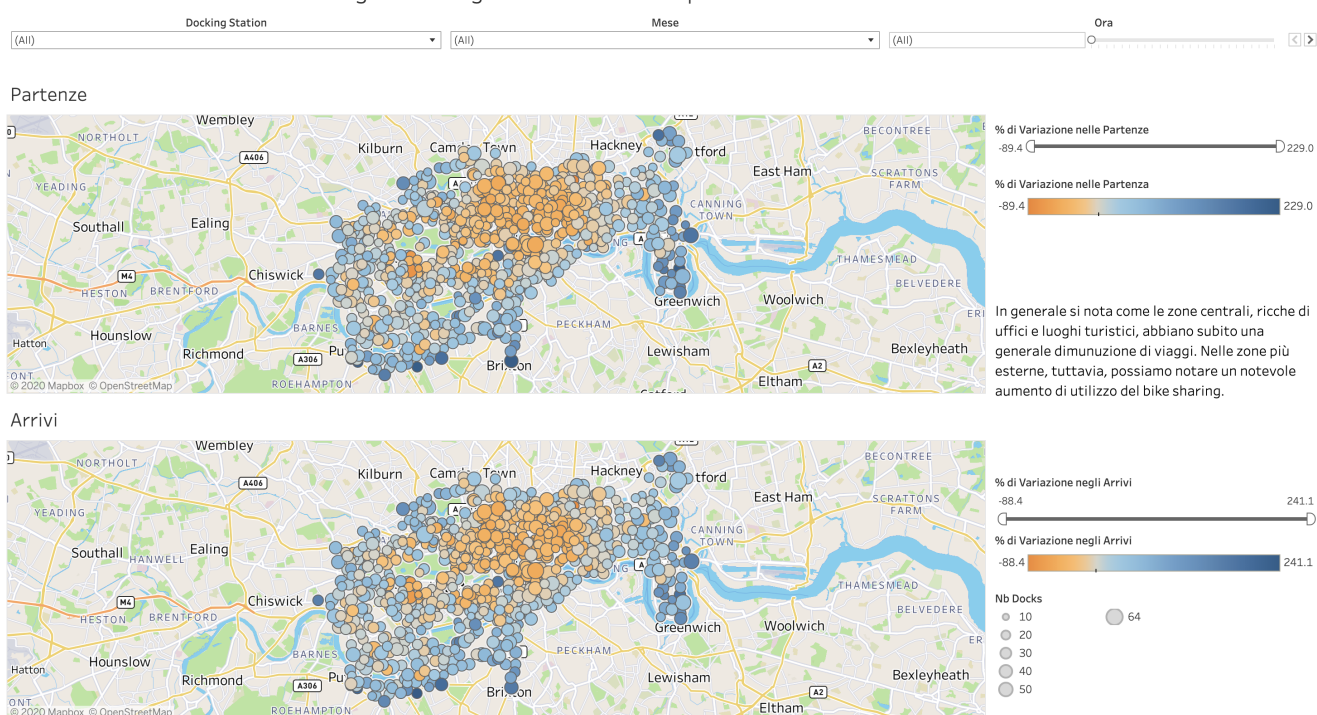


**Scopri di più:**  
 Se si desidera è possibile analizzare i cambiamenti selezionando in particolare una stazione di partenza e una di arrivo per valutare l'utilizzo del servizio tra due punti della città.  
 Ad esempio selezionando come docking station di partenza "Waterloo Station 1, Waterloo" si osserva un decremento rispetto al 2019 nell'utilizzo ma un aumento nella durata media dei viaggi. Ma cosa succede nelle altre stazioni? Scopri nella prossima storia.

Figura 5. Seconda infografica



Come sono variati i flussi nelle singole docking station nel 2020 rispetto al 2019?



**Figura 6.** Terza infografica