

Report stage: Analisi qualitativa del mercato e valutazione della strategia di marketing da applicare ad un potenziale e-commerce, attraverso la costruzione di KPIs, nel settore dello sport rotellistico.

1.Obiettivi

Lo scopo di questo lavoro, svolto nel periodo tra Febbraio e Maggio 2025, è stato quello di effettuare una prima analisi qualitativa di mercato, attraverso l'osservazione di competitor, e di sviluppare metriche di valutazione coerenti che supportino una strategia di marketing, in ottica di una realizzazione, potenziale, di un'attività di e-commerce nel settore dello sport rotellistico.

Il punto di riferimento per la definizione e le categorie degli sport su rotelle, nonché l'equipaggiamento specifico, è stato il sito ufficiale della Federazione Italiana Sport Rotellistici (fisir.it).

Le metriche di valutazione per questa analisi qualitativa del mercato relativo allo sport su rotelle sono costituite da KPI (Key Performance Indicator) anch'esse di carattere qualitativo/descrittivo, in quanto le informazioni di riferimento per la realizzazione di queste metriche si basano su siti di vendita online già esistenti, i cui dati sono stati estratti tramite web scraping.

La scelta della fonte, quindi, è stata modulata sulla base delle KPI da realizzare e, dal momento che i dati disponibili riguardavano quelli pubblicati nei rispettivi siti web, e di conseguenza nessun tipo di informazione sui volumi di vendita risulta accessibile per creare KPI di carattere predittivo, le metriche generiche di partenza vertevano principalmente sul catturare variazioni di prezzo, variazioni di sconto, valutazione dei prodotti (in formato di punteggio), numero recensioni e brand più ambiti.

Sulla base di queste premesse le fonti reputate idonee, ovvero che avessero almeno i dati di prezzo, score, recensioni e nome del brand relativi ai prodotti, oltre che una struttura non troppo complessa o pesantemente protetta da misure anti-scraping, e un catalogo di prodotti sufficientemente ampio, sia in termini di numero complessivo sia di categorie di prodotto, si sono rivelate due, una italiana e una inglese: Decathlon Italia e RollerSnakes UK. Chiaramente entrambe le fonti presentavano le medesime informazioni chiave al fine di garantire continuità e coerenza tra i dati raccolti.

La scelta di Decathlon, inoltre, è stata supportata dalla natura accessibile della sua brand reputation, dal fatto che si è sempre presentato come "Il negozio degli sportivi", ovvero un punto di riferimento per tutti coloro che desiderano avvicinarsi ad uno sport specifico, grazie anche alla vastissima gamma di categorie sportive che copre, dai principianti ai più esperti e appassionati. RollerSnakes UK, invece, è stata scelta come riferimento estero, dal momento che, tra i principali siti di roller esteri, questo risultava il più adatto date le premesse elencate in precedenza e le risorse disponibili in termini computazionali.

2.Svolgimento

Il primo passo di questo lavoro è stato quello di analizzare la struttura dei siti web più idonei allo scopo del progetto, attraverso l'impiego di strumenti per sviluppatori, alcuni già integrati nel browser, come per Safari e Chrome.

Questi strumenti permettono di ispezionare gli elementi di un sito nella loro forma HTML, oltre che osservare le chiamate di rete che il sito fa, che possono rivelarsi utili per inferire possibili chiamate API e, quindi, ricavare in maniera più diretta le informazioni. In questo modo sono stati recuperati i tag HTML necessari per sviluppare i rispetti script per Decathlon e RollerSnakes, come, ad esempio, “div.page-container” e “div.collection__main” che indicano i contenitori degli elementi-prodotto per ciascuna pagina del sito, rispettivamente per l’una e l’altra pagina web.

Anzitutto il metodo impiegato per entrambi è stata una combinazione di Selenium, che ha permesso l’esecuzione di un webdriver che naviga il sito, unito a BeautifulSoup per il parsing HTML delle pagine e quindi l’effettivo recupero dei dati, in ambiente di lavoro locale, nello specifico VS Code.

Entrambi gli script seguono la medesima logica: come prima cosa viene effettuato l’accesso alla pagina di partenza, dopodiché vengono recuperati in una lista i link dei prodotti da tutte le pagine di ciascuna categoria di prodotto, come nel caso di Decathlon, che ne presentava 5, o direttamente dalla pagina di partenza, come per RollerSnakes, e una volta esaurite le pagine e raccolti i link dei prodotti inizia l’estrazione effettiva dei dati, secondo i tag HTML dei rispettivi siti web; infine i dati estratti vengono salvati in un file JSON, dentro una cartella salvata in locale direttamente dallo script.

Dopo il periodo di scraping, che si è verificato settimanalmente (al fine di scongiurare possibili blocchi dell’IP da parte dei siti coinvolti) tra l’inizio di aprile e metà maggio 2025, i file JSON sono stati caricati su MongoDB, che si prestava molto bene per lo scopo del progetto e il successivo sviluppo delle query per la costruzione dei KPI. Inoltre, come interfaccia per il database è stato utilizzato MongoDB Compass, che ha permesso di visualizzare in maniera aggregata i dati, poiché localmente erano file separati, e di verificare le query che sono state realizzate in un python notebook su VS Code, attraverso l’importazione dei file, già aggregati su MongoDB Compass, con la libreria di PyMongo.

Per la realizzazione dei KPI, tuttavia, i dati di Decathlon si sono rilevati, dopo un primo parsing dei dati estratti, più significativi sia da un punto di vista qualitativo che quantitativo, motivo per cui il notebook finale realizzato per concludere lo stage contiene osservazioni basati su quest’ultimi; la scelta è stata anche supportata dal fatto che i KPI avrebbero mantenuto un carattere generale e flessibile, così da poter essere applicati in futuro ad altri dati, e, quindi, da un punto di vista puramente concettuale, non avrebbe impedito la buona riuscita del lavoro.

Quindi la pipeline finale del progetto è stata:

- **Primo step:** analisi della struttura dei siti web adatti per lo scopo e per risorse disponibili
- **Secondo step:** sviluppo degli script e fine tuning degli scraper
- **Terzo step:** estrazione settimanale dati e salvataggio in formato di file JSON
- **Quarto step:** storage dei file su MongoDB e aggregazione attraverso MongoDB Compass
- **Quinto step:** creazione delle query per KPI significativi e rappresentazioni grafiche

3. Risultati/Conclusioni

Prima di enunciare i KPI più importanti sviluppati per questo progetto di stage, bisogna tener conto che il periodo di raccolta dei dati, cioè circa due mesi (terminato una settimana prima della fine dello stage allo scopo di sviluppare i KPI), non si è rilevato abbastanza lungo per poter catturare variazioni più significative e apprezzabili legate ai prezzi, agli sconti e altri indicatori.

A questo proposito si è effettuato, nel periodo di agosto, un'ulteriore estrazione di dati per cercare di osservare una variazione tra le principali variabili considerate, ovvero prezzo, score e recensioni. Questa volta, però, non si è potuto raccogliere dati da Decathlon a causa di una complicazione tecnica dovuta ad un cambiamento nella struttura del sito e ad un conseguente tentativo senza successo di recupero dei dati, probabilmente rallentato dalle troppe richieste verso il sito in poche ore.

Per questo motivo si sono impiegati i dati di RollerSnakes UK, il cui scraper funziona ancora perfettamente, che si sono rilevati utili allo sviluppo della strategia di marketing come da obiettivi del progetto.

I KPI più importanti sviluppati sono i seguenti:

- **Prezzo Medio per Mese e Categoria**
- **Trend Prezzo Medio per Mese e Categoria**
- **Trend Sconto Medio Settimanale** (mostrato anche per categoria)
- **Sconto Medio (%) per Categoria (media periodo)**
- **Top Brand per Recensioni**
- **Recensioni Medie per Intervallo di Sconto**
- **Prodotti più stabili nel punteggio**
- **Disponibilità dei prodotti per brand**
- **Correlazione tra sconto e numero Recensioni**

Sebbene i dati di RollerSnakes non avessero informazioni disponibili sullo sconto, a differenza di Decathlon, hanno comunque conferito, grazie all'estrazione di Agosto, degli insight utili per supportare una strategia di vendita, reso evidente dal KPI del "Prezzo Medio per Mese e Categoria" che, rappresentato graficamente, mostra una evidente stagionalità dei prezzi, i quali, per alcune categorie di prodotto, si dimostrano più bassi in media.

Nonostante la scarsità di osservazioni per ottenere degli insight più apprezzabili e utili in un contesto strategico, si ritiene che, almeno da un punto di vista concettuale e di significato, i KPI sviluppati possano comunque essere utili e potenzialmente impiegati per una prima valutazione qualitativa del mercato di questo settore sportivo e, nel contesto di questo progetto, possano diventare più significativi in un periodo di osservazione maggiore e, quindi, con un volume di dati superiore a quello raccolto.