

Analisi dei requisiti per il progetto di analisi dei dati relativi alla vendita di una categoria di articoli tramite e-commerce

Caso di studio

Si vogliono analizzare i dati relativi alle vendite di scarpette da arrampicata nell'arco di un mese, per estrapolare le preferenze dei clienti in termini di articoli, prezzi, metodo di pagamento scelto e verificare se il modo di utilizzo da parte del cliente coincida con quello consigliato dal produttore.

Dataset utilizzati

Vengono utilizzati 3 set di dati; gli attributi che permettono di rapportare i diversi dataset sono sottolineati.

Lista degli articoli in vendita (shoes_list.xml)

Il file contiene la lista degli articoli in vendita che coincide con la lista degli articoli presenti sul sito. Ogni articolo è identificato tramite l'attributo *model* del tag *shoe*.

Tag	Attributi	Descrizione	Tipo di dato	Requisiti
<i>shoes_list</i> (root element)				
<i>shoe</i>	<u><i>model</i></u> (nome dell'articolo fornito dal produttore)	Articolo in vendita	Stringa	
<i>type</i>		Tipo di chiusura (<i>velcro</i> , <i>shoelace</i> -stringhe o <i>elastic</i> -elastico)	Stringa	Valori validi: - <i>velcro</i> - <i>shoelace</i> - <i>elastic</i>
<i>price</i>		Prezzo in euro	Decimale	
<i>weight</i>		Peso in grammi di una scarpa n.39	Intero	
<i>sole</i>	<u><i>stype</i></u> (tipo di suola fornito dal)	Spessore della suola in mm	Decimale	

	produttore)			
<i>suggested_use</i>		Utilizzo consigliato dal produttore	Stringa	Valori validi: -indoor -outdoor -indoor_outdoor

Lista delle transazioni (transactions_aug2022.json)

Il file contiene la lista delle transazioni effettuate durante il mese tramite il sito. Per poter effettuare un acquisto è necessario essere iscritti al sito. È possibile inserire un codice sconto che dà diritto alla detrazione di una percentuale (definita dal codice stesso) dal prezzo di listino. Ogni transazione è identificata tramite il numero progressivo *order_number*.

Chiave	Descrizione	Tipo di dato	Requisiti
<i>order_number</i>	Numero d'ordine	Stringa	Progressivo
<u><i>username</i></u>	Nome utente	Stringa	Valore univoco
<i>n_articles</i>	Numero di articoli acquistati	Intero	
<u><i>articles</i></u>	Nome degli articoli acquistati	Stringa se <i>n_articles</i> =1 Lista di stringhe se <i>n_articles</i> >1	Valori presenti nella lista degli attributi <u><i>model</i></u> degli oggetti <i>shoe</i> nel file <i>shoes_list.xml</i>
<i>payment_method</i>	Metodo di pagamento utilizzato	Stringa	Valori validi: <i>PayPal</i> , <i>MasterCard</i> , <i>Visa</i>
<i>order_date</i>	Data dell'ordine	Oggetto json con chiavi <i>day</i> , <i>month</i> , <i>year</i>	Valori compatibili con il formato data
<i>order_hour</i>	Ora dell'ordine	Oggetto json con chiavi <i>hour</i> , <i>minute</i> , <i>second</i>	Valori compatibili con il formato orario
<i>discount_code</i>	Specifica se sia stato utilizzato un codice sconto	Booleano	<i>true</i> se è stato utilizzato un codice sconto <i>false</i> se non è stato utilizzato un codice sconto
<i>discount</i>	Percentuale di sconto sul totale	Intero	Valore compreso tra 5 e 70

Lista delle recensioni (reviews.csv)

Il file contiene la lista delle recensioni lasciate dagli utenti. Qualora l'utente abbia acquistato più di una tipologia di articolo, può lasciare una recensione per ogni articolo acquistato. I campi *gender*, *experience*, *shoe_model* e *real_use* sono compilati dall'utente tramite un menù a tendina. Ogni recensione è identificata tramite la coppia di valori *username* e *shoe_model*.

Campo	Descrizione	Tipo di dato	Requisiti
<u><i>username</i></u>	Nome utente	Stringa	Valore univoco
<i>gender</i>	Genere dell'utente, non è obbligatorio compilare questo campo	Stringa	Valori validi: - <i>m</i> - <i>f</i> - <i>null</i>
<i>age</i>	Età dell'utente	Intero	Compreso tra 18 e 109
<i>experience</i>	Grado di esperienza dell'utente	Stringa	Valori validi: - <i>beginner</i> - <i>intermediate</i> - <i>advanced</i>
<u><i>model</i></u>	Modello di scarpa recensito	Stringa	Valore presente nella lista degli attributi dell'oggetto " <i>shoe</i> " nel file <i>shoes_list.xml</i>
<i>real_use</i>	Modo di utilizzo da parte dell'utente	Stringa	Valori validi: - <i>indoor</i> - <i>outdoor</i> - <i>indoor_outdoor</i>
<i>review</i>	Recensione	Stringa	Dato non strutturato

Statistiche proposte

Si intendono proporre le seguenti statistiche:

S1) Per quanto riguarda le **preferenze** dei clienti relative agli **articoli** scelti, si calcola il numero di pezzi venduti per ogni articolo, visualizzandoli tramite un istogramma; questo valore verrà confrontato con il prezzo di ogni articolo e visualizzato tramite uno *scatter plot*. Si confronteranno inoltre il numero di pezzi venduti con la tipologia (*type*), peso e spessore della suola, evidenziando per ognuno di questi parametri l'influenza di età, genere ed esperienza del cliente.

S2) Per quanto riguarda le **preferenze** dei clienti relative all'**importo speso**, si calcola questo valore e lo si confronta con il genere, l'età e l'esperienza del cliente, visualizzando la relazione che intercorre tra l'importo e ognuno di questi valori tramite *scatter plot*. Nella visualizzazione, si evidenziano i clienti che hanno utilizzato un codice sconto per l'acquisto.

S3) Per quanto riguarda le **preferenze** dei clienti relative al **metodo di pagamento** si calcola il numero di transazioni con ogni metodo di pagamento; questo valore viene confrontato con il genere del cliente e con l'età del cliente, visualizzando la relazione che intercorre tra il metodo di pagamento e ognuno di questi due valori tramite *scatter plot*.

S4) Per quanto riguarda il **modo di utilizzo**, per ogni articolo si ottiene il modo di utilizzo reale da parte dei clienti che hanno lasciato una recensione e lo si confronta con il modo di utilizzo suggerito dal produttore. Per ogni articolo si otterrà il numero di clienti che utilizzano l'articolo secondo le indicazioni (*suggested_use = real_use*) e il numero di clienti che non utilizzano l'articolo secondo le indicazioni (*suggested_use ≠ real_use*). In questo secondo caso si evidenzierà anche la differenza tra i modi d'uso e il grado di esperienza del cliente.

Outcome atteso dal sistema

Sarà possibile per il Data Analyst effettuare le seguenti **considerazioni** relative alla vendita di una categoria di articoli tramite e-commerce in un mese.

La statistica S1 dà informazioni sulle **preferenze** dei clienti relative agli **articoli scelti**, evidenziando gli articoli più e meno venduti e l'influenza del prezzo sulla scelta dell'articolo da acquistare. Si ottengono informazioni relative all'influenza di genere ed età del cliente sulla scelta di una determinata tipologia di prodotto (caratterizzata dal tipo di chiusura e dal peso).

La statistica S2 dà informazioni sulle **preferenze** dei clienti relative all'**importo speso**, evidenziando l'influenza del genere e dell'età sull'importo speso. Permette di valutare strategie di marketing mirate ad una specifica classe di clienti e la scelta di fornire codici sconto specifici.

La statistica S3 dà informazioni sulle **preferenze** dei clienti relative al **metodo di pagamento** utilizzato. Permette di valutare quali metodi di pagamento mantenere attivi sul sito e, eventualmente, quali nuovi metodi di pagamento simili o complementari potrebbero essere aggiunti per migliorare l'esperienza del cliente.

La statistica S4 dà informazioni sul **modo di utilizzo** reale da parte dei clienti; questa informazione è utile per valutare la comunicazione dei suggerimenti per l'utilizzo. Potrebbe portare alla scelta dell'apertura di una sezione specifica del sito, una descrizione più mirata, la produzione di contenuti esplicativi da diffondere anche tramite i social.

Applicando la stessa analisi ad ulteriori dataset relativi ad altri mesi e/o anni, qualora siano disponibili, sarà possibile per il Data Scientist valutare l'andamento dei parametri in un periodo temporale più lungo per ottenere, grazie a **tecniche predittive**, una stima dell'andamento delle vendite, su cui basare le decisioni.

Le informazioni ricavate potranno essere la base per un **racconto** tramite il sito o le pagine social dell'e-commerce, in cui diverse categorie di clienti espongono la propria esperienza con un articolo adatto al loro grado di abilità, per guidare i nuovi clienti nella scelta dell'articolo più adatto a loro.

Glossario

Termine	Descrizione	Sinonimo	Collegamenti
e-commerce	Piattaforma tramite cui vengono acquistati gli articoli	sito	transazione
cliente	Utente del sito che compie un acquisto	utente	Recensione, transazione
modello	Nome del modello dell'articolo	Nome dell'articolo	Recensione, transazione
Modo di utilizzo	Può essere outdoor, indoor oppure entrambi. Può essere suggerito dal produttore oppure il modo d'uso effettivo da parte dell'utente	Modo d'uso	Recensione, articolo