

L'esercizio è mirato a prendere confidenza con la produzione di dati nei formati XML, CSV e JSON.

Si esegua la progettazione di un sistema per l'analisi di dati. La progettazione deve prevedere:

1. L'individuazione di un argomento di proprio interesse multi-dominio
2. La produzione di almeno tre dataset rispettivamente nei formati XML, CSV, e JSON.
3. Richiesta 1: Il dataset XML dovrà contenere almeno 10 figli del nodo padre e almeno 2 nidificazioni per ogni figlio, dovranno inoltre essere presenti tag con e senza attributi.
4. Richiesta 2: il dataset JSON dovrà contenere almeno 10 oggetti JSON contenenti diversi tipi di dato (tra cui array e altri oggetti JSON)
5. Richiesta 3: il dataset CSV dovrà essere composto da un header e almeno 10 righe. Inoltre per campi a propria scelta sarà necessario inserire almeno una virgola, un carattere di nuova linea, e i doppi apici.

Procedere con la redazione di un documento che dovrà includere:

- La descrizione del caso di studio scelto
- Analisi e raccolta dei requisiti (generici e poi più specifici)
- Strutturazione dei requisiti in gruppi di frasi omogenee
- Glossario dei termini
- Descrivere nel dettaglio quale sarebbe l'outcome atteso dal sistema (analisi di business / analisi predittive), come potrebbe essere di supporto nelle decisioni ed eventualmente quali 'storie' è possibile ricavare

2

PRODUZIONE DATI OPEN.

DESCRIZIONE DEL CASO DI STUDIO SCELTO

JustDrink.it è un servizio in rete per ordinazione e consegna. La piattaforma online permette agli utenti di ordinare il proprio beverage delivery a domicilio e agisce come intermediario tra il negozio prescelto e il cliente. Con questo progetto si vuole realizzare l'analisi delle informazioni relative al customer satisfaction in particolare si vuole stabilire quale fattore influisce maggiormente sul livello di soddisfazione dei clienti.

TIPOLOGIE DI DATI

Abbiamo a disposizione tre dataset Customer Reviews in formato .csv, Order Processing Time in formato .xml e Order Details in formato .json. Tutti i tre dataset sono ad utilizzo interno e non possono essere classificati secondo il sistema di classificazione degli open data, non hanno nessun tipo di licenza o metadati.

CUSTOMER REVIEWS

order_id	numeric
date	data
rating	numeric
comment	text

ORDER PROCESSING TIME

order_id	numeric
processing_time	numeric

ORDER DETAILS

order_id	numeric
items	numeric
missing_items	boolean
price	numeric
driver_id	numeric
driver_rating	numeric

ANALISI E RACCOLTA DEI REQUISITI

Si vuole realizzare l'analisi delle informazioni relative al customer satisfaction in particolare si vuole stabilire quale fattore influisce maggiormente sul livello di soddisfazione dei clienti. Abbiamo a disposizione i seguenti dataset:

- dati sulla valutazione degli utenti per ogni singolo ordine (10 in totale) e i relativi commenti,
- i tempi di gestione di ogni singolo ordine,
- le informazioni dettagliate che riguardano ogni singolo ordine come: il numero degli articoli, la completezza dell'ordine, l'importo, il codice indicativo e il voto del driver.

STRUTTURAZIONE DEI REQUISITI IN GRUPPI DI FRASI OMOGENEE

Per gli ordini si hanno a disposizione: un codice identificativo univoco, il tempo di gestione dell'ordine, il contenuto del carrello, i dati sulla completezza dell'ordine, l'importo totale e le informazioni sul driver incaricato (codice univoco e voto personale).

Per la soddisfazione dei clienti per ogni singolo ordine (identificato da un codice univoco), si hanno a disposizione la data, il voto assegnato dall'utente in una scala da 1 a 10 e il relativo commento.

Per l'analisi del livello di soddisfazione dei clienti si intendono proporre le seguenti statistiche:

- la media, la deviazione standard e la mediana dei voti in base al tempo di gestione dell'ordine ("basso" 1-20 min, "medio" 21-30 min, "alto" sopra 31min).
- la media, la deviazione standard e la mediana dei voti divisi in base alla categoria che descrive la completezza dell'ordine ("incompleto", "completo").
- la media, la deviazione standard e la mediana dei voti per categoria che identifica il voto del driver ("basso" 1-4, "medio" 5-8, "alto" 9-10).
- la media, la deviazione standard e la mediana dei voti per fascia di prezzo del ordine ("basso" 1-20 EUR, "medio" 21-40 EUR, "alto" sopra 40 EUR).

GLOSSARIO DEI TERMINI

Termine	Descrizione	Sinonimi	Collegamenti
Cliente	Persona che effettua un acquisto	Utente	Ordine
Ordine	Transazione confermata		Cliente

OUTCOME ATTESO

Grazie ai dati raccolti sarà possibile ottenere le informazioni relative al livello di soddisfazione dei clienti elencati precedentemente e individuare i fattori che hanno maggior peso sul livello di soddisfazione dei clienti.

Calcolando la media, la deviazione standard e la mediana dei voti per categorie di tempi di gestione dell'ordine possiamo verificare se con la diminuzione del tempo di attesa cresce la soddisfazione dei clienti.

Calcolando la media, la deviazione standard e la mediana dei voti per categorie di completezza dell'ordine possiamo verificare se, in caso di ordini incompleti, il cliente rimane meno soddisfatto.

Calcolando la media, la deviazione standard e la mediana dei voti per categorie del voto del driver possiamo verificare se la professionalità del corriere ha impatto sul livello di soddisfazione del cliente.

Infine, calcolando la media, la deviazione standard e la mediana dei voti per fasce di prezzo possiamo osservare l'impatto di tali importi sulla soddisfazione degli utenti.

Comparando i risultati ottenuti evidenziamo i fattori che possono influenzare il livello di soddisfazione del cliente e a cui il business dovrebbe prestare maggiori attenzioni.

ANALISI DI BUSINESS

- Definition of targets; la definizione dei processi che maggiormente impattano sulla soddisfazione dei clienti aiuta il business a individuare i target che devono essere raggiunti dai dipendenti.
- Definition of customer preferences; la definizione degli aspetti del servizio che impattano maggiormente la soddisfazione dei clienti.
- Definition of marketing strategies; la conoscenza delle preferenze degli utenti può essere utilizzata nelle campagne di marketing per sottolineare i punti forti del servizio.

ANALISI PREDITTIVA

- I rapporti tra i singoli fattori e il livello di soddisfazione del cliente determinati possono essere utilizzati per prevedere come le variazioni di un parametro impatta il livello della soddisfazione dei clienti. Per esempio, come la diminuzione del tempo di gestione dell'ordine può aumentare la customer satisfaction e i relativi valori.
- I rapporti tra i singoli fattori e il livello di soddisfazione possono anche essere utilizzati per proporre cambiamenti funzionali, per esempio dando priorità ai driver con un voto più alto con lo scopo di aumentare la soddisfazione dei clienti.

ESEMPIO

