**README – Analisi E-commerce Olist con Power BI**

**Panoramica del progetto**

**Questo progetto analizza il dataset Olist, un e-commerce brasiliano, per esplorare vendite, logistica, soddisfazione clienti e performance dei prodotti.  
Gli obiettivi principali sono:**

* **Comprendere le prestazioni complessive dell’azienda (ordini, revenue, review score)**
* **Analizzare l’efficienza operativa (tempi di consegna, tempi di approvazione, shipping share)**
* **Identificare pattern geografici nelle vendite e nelle consegne**
* **Studiare la performance per categoria di prodotto e l’efficienza di revenue**

**Il progetto utilizza Power BI per ETL, modellazione e visualizzazione, supportato da Python per la pulizia dei dati quando necessario.**

**Fonti dati**

**I dataset utilizzati sono:**

* **orders – dettagli degli ordini, date e status**
* **order\_items – prodotti, prezzi e valori di spedizione**
* **payments – informazioni sui pagamenti**
* **customers – informazioni sulla posizione dei clienti**
* **geolocation – città e stati (con problemi di duplicati e accenti)**
* **order\_reviews – punteggi delle recensioni dei clienti**

**Preparazione e trasformazione dei dati**

**1. Power Query**

* **Importati tutti i dataset in Power Query**
* **Pulite le colonne non necessarie e trasformati i tipi di dati**
* **Rinominato tabelle e colonne per maggiore chiarezza**
* **Create nuove colonne calcolate:**
  + **FULL\_PRICE = PRICE + FREIGHT\_VALUE in order\_items**
  + **DeliveryDays = order\_delivered\_customer\_date – order\_purchase\_timestamp**
  + **ApprovalDays = order\_approved\_at – order\_purchase\_timestamp**
* **Corrette le aggregazioni e categorizzazioni dei campi**
* **Pulizia della tabella geolocation in Python per rimuovere duplicati e accenti**

**2. Tabella calendario**

**Creata una tabella date per analisi temporali:**

**calendar = ADDCOLUMNS (**

**CALENDAR ( MIN ( orders[purchase\_timestamp] ), MAX ( orders[purchase\_timestamp] ) ),**

**"Anno", YEAR([Date]),**

**"Mese", MONTH([Date]),**

**"NomeMese", FORMAT([Date], "MMM"),**

**"Giorno", DAY([Date]),**

**"Settimana", WEEKNUM([Date]),**

**"GiornoSettimana", FORMAT([Date], "dddd")**

**)**

* **Permette analisi per anno, mese, giorno e giorno della settimana**
* **Collegata alla tabella orders[purchase\_timestamp]**

**Modellazione dei dati**

* **Creato schema a stella con orders come fact table e dimensioni: calendar, customers, geolocation, order\_items, payments**
* **Definite relazioni e relazioni attive per calcoli DAX corretti**
* **Impostate aggregazioni appropriate (Somma, Conteggio, Media) e categorie dati per campi geografici e numerici**

**Misure DAX principali**

* **Units Sold: Units Sold = COUNT(order\_items[product\_id])**
* **Total Revenue: somma di FULL\_PRICE**
* **Average Delivery Days e Average Approval Days basate sulle colonne calcolate**
* **Misure filtrate per status ordini (approved, delivered, shipped) per evitare valori null in review e tempi di consegna**
* **Confronti YoY e variazioni mensili di revenue e ordini**

**Pulizia dati**

* **La tabella geolocation conteneva nomi città duplicati con accenti, corretti in Python per standardizzare i nomi**
* **Controllati tutti i campi numerici e date per garantire correttezza nelle aggregazioni**
* **Aggiunte colonne differenza date in Power Query per semplificare DAX operativo**

**Visualizzazioni e pagine**

1. **Pagina Overview**
   * **KPI: Ordini, Revenue, Review Score, Avg Delivery Days**
   * **Trend mensili di ordini e revenue, con evidenza di settembre 2018 come mese critico**
2. **Pagina Operational Performance**
   * **Media tempi consegna e approvazione per mese**
   * **Trend Shipping Share**
   * **Picchi all’inizio dell’anno (Feb–Mar) e periodo festivo (Nov–Dic)**
3. **Pagina Geography**
   * **Mappa vendite per stato**
   * **Avg Delivery Days per stato (alto in RR, AP, AM)**
   * **Focus sulla regione Sud-Est come principale area di vendita**
4. **Pagina Product Insights**
   * **Revenue vs Units Sold per categoria prodotto**
   * **Evidenziate categorie ad alto valore con meno unità vendute (Watches & Gifts)**
   * **Shipping share per categoria, evidenziando costi logistici elevati**
5. **Pagine Drill-through**
   * **Analisi dettagliata per categoria prodotto, stato o singolo ordine**
   * **Pagine nascoste con pulsante “Back” per navigazione fluida**

**Insight principali**

* **Alta soddisfazione clienti: review score medio 4.1**
* **Revenue e ordini consistenti: $16M e quasi 100k ordini nel 2018**
* **Tempi medi di consegna elevati (~12.5 giorni), con picchi all’inizio e fine anno**
* **Concentrazione geografica nel Sud-Est, con criticità nel Nord (RR, AP, AM)**
* **Analisi prodotto evidenzia categorie con elevato valore di revenue ma minore numero di unità e categorie con alta incidenza di spedizione**

**Strumenti utilizzati**

* **Power BI: caricamento dati, trasformazione, modellazione e visualizzazione**
* **Python: pulizia tabella geolocation**
* **DAX: calcolo KPI, analisi temporali e filtri avanzati**

**Prossimi passi / Raccomandazioni**

* **Ottimizzare la logistica per ridurre i tempi di consegna negli stati del Nord**
* **Valutare costo di spedizione vs revenue per categorie con shipping share elevata**
* **Integrare elementi di storytelling dinamico come drill-through per stato o categoria prodotto**
* **Monitorare trend mensili per anticipare fluttuazioni stagionali**

**STAR SCHEMA:**

**Immagine che contiene testo, schermata, design, diagramma

Il contenuto generato dall'IA potrebbe non essere corretto.**

**Utilizzo Python:**

**Immagine che contiene testo, Carattere, software, schermata

Il contenuto generato dall'IA potrebbe non essere corretto.**