**Analisi Olip Store – Esame Modulo PowerBI**

1. Importati tutti i file csv con il power query, divisi in due cartelle per importanza
2. Ho iniziato a lavorare sulle dimensioni, nel caso dei prodotti ho fatto una merge con la tabella della traduzione, tenendo infine solo due colonne, ProdKey e CategoryName. Alcune prodotti non avevano una categoria, li ho dunque marchiati con la stringa “Other”.  
   Per quanto riguarda la dimensione Customer ho mantenuto solo le colonne Customer\_ID, \_city e \_state.
3. Ho riassunto le tabelle ordini in un fatto soltanto, così da avere un corretto star schema al momento della creazione del modello semantico. Grazie a una soffiata dai piani alti, ho legato tramite una serie di left join la tabella orders, collegandola con order\_items e order\_review.   
   Possiamo notare come rimangano 775 record non matchati tra le prime due tabelle, perché alcuni ordini sono stati cancellati senza avere comunque le informazioni del prodotto a loro associati nella tabella \_items. Vanno comunque mantenuti perché non sono errori, e per alcune analisi successive sono rilevanti. Prima di joinare l’ultima tabella con le review (di cui poi teniamo solo la colonna con il rating della recensione) aggrego così da avere solo una recensione per ordine singolo, tenendo solo il rating massimo di ogni recensione. Ho inoltre cambiato lo status degli ordini che risultano marchiati come “Canceled” ma comunque aventi una data di consegna al cliente, li ho chiamati “Anomaly”.
4. A questo punto possiamo caricare i dati nel modello e creare le relazioni, ovviamente nella tabella fatti sono state mantenute tutte le colonne chiave per tale operazione.
5. Sistemo le aggregazioni implicite, rinomino eventuali campi, creo le gerarchie ove possibile e aggiungo una colonna calcolata (l’avrei fatto nel power query, ma ho trovato in questo caso l’M code nettamene più ostico rispetto alla DAX) chiamata Delivery Flag, assegnando una stringa di riferimento allo stato di consegna dell’ordine; nello specifico se ancora non è stato consegnato viene marchiato come “In Process”, se è stato consegnato prima della data di consegna prevista come “On Time”, se i giorni di elaborazione dell’ordine dalla data di acquisto a quella di consegna al corriere sono di più di quelli tra il corriere e il delivery al cliente come “Delayed Fulfillment” e infine viceversa di quest’ultima, quindi “Delayed Delivery” se è stato il corriere a tardare la consegna.
6. Creo la tabella calendario usando CALENDARAUTO() con il check sulla colonna date di order\_purchase, ordino il campo month con il monthNO per la visualizzazione corretta, la lego ai fatti sullo stesso campo e creo anche la tabella “contenitore” delle mie misure.
7. A questo punto incomincio creando il report, nella prima pagina di Overview ho inserito una serie di card riepilogative, un gauge con obiettivo i ricavi dell’anno precedente e un decomposition tree che analizza i ricavi scompattandoli per anno, stato, citta e categoria, utilizzabile anche come filtro per le card e il gauge. Sulla sinistra ho inserito un menù di navigazione tra le pagine.
8. La pagina Orders Analysis risponde alle domande della prima richiesta, quindi filtri espliciti (che si ripeteranno in tutte e tre le pagine di analisi) per anno e status dell’ordine sulla sinistra, una mappa geografica che lavora con gli stati (region) impostata con la Geocoding Culture su pt-BR, al suo fianco due grafici, un barchart con il count degli ordini mese per mese, una linea che rappresenta l’andamento dell’anno precedente e come etichette la variazione percentuale tra le due misure. Come formattazione condizionale, viene evidenziato automaticamente in giallo la colonna con il count più alto, visto che molti mesi sono simili fra loro in quantità. Sotto troviamo un line chart che espone il cumulato YTD sempre degli ordini.
9. La stessa identica impostazione di pagina e grafici viene ripetuta per rispondere al quesito numero 2, oggettivamente uguale ma con in analisi i ricavi, che ottengo grazie alla sumx iterante sulla tabella sales, sommando price e freight value. La pagina in questo caso si chiama Sales Analysis.
10. In entrambe le pagine appena menzionate, se viene evidenziato un mese in particolare si può accedere alle pagine di drill-through tramite il tasto posto sopra i bar chart.  
    Quella degli ordini rimanda a una tabella, dove possiamo vedere il dettaglio di ogni ordine, comprensivo di data, status, delivery status e città di destinazione.  
    Quella dei ricavi rimanda a una matrice, che scomponendo regione per città e zip code vengono mostrati i ricavi, i ricavi dell’anno precedente, la variazione percentuale, la percentuale su tutti gli zipcode e quella su tutte le città, utilizzando così anche la funzione REMOVEFILTERS().  
    In entrambe le pagine ritroviamo il menù di navigazione incontrato nella Overview.
11. Infine, come richiesto dal terzo punto, ho creato una pagina per analizzare il rating delle recensioni, intitolata Rating Analysis. Qui ho inserito uno scatter chart per mostrare il rating medio per ogni stato, con la bolla che diventa più grande in base al count di recensioni in quel singolo stato, così da avere più chiarezza nel leggere il dato (difatti un paio di stati hanno una media molto alta, ma il numero di recensioni è bassissimo rispetto a molti altri stati); una card riepilogativa mostra il totale del conteggio al fianco dello scatter.   
    Sempre sulla parte destra troviamo anche un clustered bar chart che per ogni possibile voto (da 1 a 5) mostra il count di ognuno, quindi letteralmente scompone la card riepilogativa.  
    Sotto di loro c’è una lista per categoria di prodotto con la media delle loro recensioni mostrate in ordine decrescente.  
    NB: quei 775 record avanzati senza una categoria in questo caso vengono mostrati come (blank) perché hanno valore nullo, ne abbiamo evidenza solo dal momento che impostiamo il filtro dello status su Canceled o Unavailable.  
    Ho cercato in tutti i modi per risolvere il problema agendo sulle relazioni, ma non ho trovato nessun modo per sistemarlo se non aggiungendo una misura che funziona da tooltip, così passando il mouse sopra la eventuale categoria (blank) nel grafico viene mostrata la scritta “When selecting an Order Status marked as 'Canceled' or 'Unavailable' a Blank category may appear. Those are orders canceled before getting filled with other informations such the category of the product by the system. They are still relevant for other analysis, and they still have a rating score associated with.”.

Un altro tooltip utilizzato è implementato nello scatter chart, facendo mouseover su una bolla il tooltip mostra uno zoom sulla mappa geografica di quel particolare stato, così da avere più cognizione più chiara dello stato stesso, avendo a disposizione nemmeno il suo nome completo, ma la semplice abbreviazione con due lettere.

1. Per concludere prima di fare sharing su PowerBI Service, ho nascosto tutti i campi chiave, le pagine non necessarie all’utente finale e altri piccoli accorgimenti del caso.