**W15 D1 – Introduzione a Power BI**

**Business Intelligence**

La Business Intelligence (BI) è un insieme di processi, metodologie e strumenti software che supportano il processo decisionale aziendale. L’obiettivo principale è quello di trasformare i dati disponibili in informazioni, e successivamente in conoscenza, così da fornire un supporto concreto alle decisioni dei manager.

Un dato diventa realmente un’informazione quando suggerisce o permette di intraprendere un’azione. Le due principali sfide della BI sono:

* **la raccolta e la ristrutturazione dei dati**, che devono essere corretti, affidabili e ben organizzati;
* **lo storytelling**, cioè la capacità di comunicare i risultati in maniera chiara ed efficace, così che chi prende le decisioni possa comprenderli e agire di conseguenza.

**I task di un Data Analyst**

Il lavoro di un Data Analyst non si limita a “leggere i dati”, ma comprende una serie di attività strutturate:

1. **Comprensione dello scenario**  
   È la fase iniziale, in cui si raccolgono i requisiti, si individuano le sorgenti dati e si esplorano i dataset per acquisire familiarità con il contesto.
2. **Preparazione del dato**  
   Consiste nell’analizzare la qualità e l’integrità dei dati, correggere eventuali errori, eliminare ridondanze e ristrutturare i dataset in modo che siano leggibili e facilmente utilizzabili per aggregazioni e calcoli.
3. **Costruzione del modello logico**  
   Implica la definizione delle relazioni tra le tabelle, necessarie per collegare informazioni provenienti da fonti diverse e permettere analisi coerenti.
4. **Visualizzazione e analisi**  
   La creazione di report interattivi e dinamici consente un’esplorazione efficace dello scenario, aiutando a individuare pattern, anomalie o tendenze.
5. **Gestione della soluzione**  
   Una volta implementata, la soluzione di BI deve essere gestita: occorre occuparsi degli aggiornamenti dei dati, della distribuzione dei report, della gestione degli accessi e del monitoraggio continuo.

**Che cos’è Power BI**

Power BI è una raccolta di servizi software, applicazioni e connettori che operano insieme per trasformare dati provenienti da origini diverse in informazioni dettagliate, coerenti e visivamente coinvolgenti.

Gli elementi principali di Power BI sono:

* **Power BI Desktop**: l’applicazione per Windows che consente di sviluppare report e modelli.
* **Power BI Service**: un servizio SaaS (Software as a Service) che permette di pubblicare, condividere e collaborare sui contenuti.
* **Power BI Mobile**: l’app per dispositivi mobili che consente di consultare report e dashboard ovunque ci si trovi.

**Definizioni fondamentali**

* **Visualization**: un singolo oggetto grafico (es. grafico a linee, a barre, a torta).
* **Report**: insieme di più visualizzazioni organizzate in una o più pagine; è dinamico, interattivo e filtrabile.
* **Dashboard**: insieme selezionato di visualizzazioni provenienti da diversi report riunite in un’unica pagina; è statica.
* **Tile**: la singola visualizzazione all’interno di una dashboard.

**Flusso tipico di Power BI**

Il processo standard in Power BI si sviluppa in quattro fasi:

1. **Acquisizione dei dati**, da file CSV, Excel, database locali (on-premise), database in cloud, SharePoint.
2. **Creazione del report**, con modellazione e visualizzazione dei dati.
3. **Pubblicazione del report** sul servizio Power BI.
4. **Condivisione**, per permettere ad altri utenti o gruppi di accedere e analizzare le informazioni.

**Modalità di storage dei dati**

Power BI offre due modalità principali per gestire i dati:

* **Import**  
  I dati vengono importati e copiati localmente all’interno del file Power BI (.pbix). Questo garantisce tutte le funzionalità della piattaforma e un’esperienza utente ottimale, ma ha dei limiti: il volume massimo dipende dalla licenza, e gli aggiornamenti devono essere manuali o pianificati.
* **DirectQuery**  
  I dati non vengono copiati, ma interrogati direttamente nella sorgente. È utile per dataset molto grandi e per scenari real-time, ma può risultare meno performante e presenta limitazioni su alcune trasformazioni e funzioni DAX.

**Power Query Editor**

In Power BI Desktop, il primo passo è l’acquisizione e preparazione dei dati. Questo compito viene svolto attraverso **Power Query Editor**, uno strumento che consente di collegarsi a sorgenti dati locali (Excel, CSV, TXT, database on-premise) o cloud (SharePoint, database in cloud).

Dopo la pulizia e ristrutturazione, i dati vengono usati per la creazione dei report. Power Query Editor è un vero e proprio strumento ETL (Extract, Transform, Load) ed è presente anche in altri prodotti Microsoft come Power Platform, Azure Data Factory e Fabric.

**Flat Files ed Excel**

Molte aziende archiviano i dati in file CSV o TXT, oppure in cartelle di lavoro Excel. Le criticità principali di queste fonti dati sono:

* il percorso del file: se cambia, il report può rompersi;
* modifiche strutturali al file sorgente: l’eliminazione o rinomina di colonne può generare errori nei report.

Quando ci si collega a un database relazionale, bisogna scegliere la modalità di storage dei dati e selezionare le tabelle da importare. In alternativa, è possibile scrivere query personalizzate specificando il database di riferimento.

Un’altra opzione utile è la connessione a una cartella (**Folder**): in questo modo Power BI esegue un **append** automatico dei file contenuti al suo interno, utile ad esempio per gestire estrazioni periodiche di file con la stessa struttura.

**Profilazione campi**

Power Query consente di analizzare la qualità e la distribuzione dei dati già in fase di caricamento.

* La sezione **Column Quality** mostra la percentuale di valori validi, errori e valori vuoti.
* La sezione **Column Distribution** mostra la frequenza dei valori, distinguendo tra valori distinti e valori univoci.

È fondamentale assegnare sempre il tipo di dato corretto a ciascuna colonna, in coerenza con l’informazione che essa rappresenta.  
Nota: la vista di anteprima mostra solo le prime 1000 righe; se il dataset è molto grande, occorre abilitare la profilazione sull’intero set di dati.

**Append e Merge**

In Power Query esistono due operazioni fondamentali per combinare i dati:

* **Append**: unisce i record di due o più tabelle con la stessa struttura (accodamento).
* **Merge**: combina i record di due tabelle in base a un campo chiave comune.

Le tipologie di join disponibili in un’operazione di Merge sono:

* **Left Outer**: restituisce tutte le righe della prima tabella con le corrispondenze della seconda.
* **Right Outer**: restituisce tutte le righe della seconda tabella con le corrispondenze della prima.
* **Full Outer**: restituisce tutte le righe di entrambe le tabelle.
* **Inner**: restituisce solo le righe con corrispondenze in entrambe le tabelle.
* **Left Anti**: restituisce le righe della prima tabella che non hanno corrispondenze nella seconda.
* **Right Anti**: restituisce le righe della seconda tabella che non hanno corrispondenze nella prima.