

Video Corso Data Analysis completo

Il programma dettagliato (50+ ore)

SQL Academy (9+ ore)

SQL è il linguaggio di programmazione necessario a gestire e a interrogare i **database relazionali**, uno strumento molto utilizzato dalle aziende di qualsiasi settore. La padronanza di questo linguaggio è fondamentale per chiunque voglia lavorare con i dati, tanto che una delle domande più comuni durante i colloqui di lavoro è proprio: “Sai scrivere una query?”. Il programma:

- Installazione SQL Server Express e SQL Server Management Studio
- Connessione e creazione database d'esempio
- Introduzione ai concetti di Database, Database Relazionale, RDBMS e SQL
- Selezionare e filtrare i dati
- Raggruppare i dati
- Combinare i dati di più tabelle
- Subquery: risolvere problemi complessi scomponendoli in problemi più semplici
- Utilizzare Subquery per filtrare i dati
- Approfondimento su CASE WHEN
- Utilizzare le window function
- Popolare e aggiornare un database
- Creare da zero un database
- Vincoli di chiavi primarie, esterne e check
- Progettazione concettuale, diagramma E-R e traduzione in database relazionale
- Salvare e rieseguire il codice all'interno del database
- Normalizzazione di un database
- Esercizi SQL
- Project Work

Ripasso di Excel (5+ ore)

La **conoscenza di Excel** rappresenta un'altra skill essenziale per chi vuole diventare un Data Analyst. La sua diffusione a tutti i livelli aziendali e le diverse funzionalità (formule, pivot, grafici e previsioni di serie storiche) lo rendono uno strumento di base indispensabile per i professionisti del settore. Il programma:

- Introduzione a formule e funzioni di Excel
- Funzioni Somma, Media, Dev St P, Somma Più Se e Conta Più Se
- Funzioni di tipo Testo
- Funzioni Logiche
- Cerca Verticale e Cerca X
- Nuove formule di matrice dinamica di Excel 365
- Funzioni database, sottotale, scarto e indiretto
- Come e perché creare una tabella Pivot
- Creiamo un grafico Pivot
- Filtri e aggiornamento dei dati
- Elaboriamo i dati di partenza
- Analisi su più colonne
- Layout e raggruppamenti nelle Pivot
- Approfondimento sui filtri
- Campi calcolati
- Formule su tabelle Pivot
- Creazione Modello Dati
- Aggiornamento tramite tabelle
- Calcolo degli scostamento percentuali
- Creazione di Dashboard
- Introduzione alle serie storiche
- Funzionalità Foglio previsione
- Implementazione di modelli matematici per la predizione di serie storiche
- Esercizi

Fondamenti di Python e Pandas (4+ ore)

In questo modulo impareremo da zero a programmare con il **linguaggio Python**. Secondo un approccio pratico e innovativo, finalizzeremo il nostro apprendimento di Python per imparare ad avviare dei processi di data analysis attraverso una serie di **task reali**: dall'acquisizione dei dati alla loro elaborazione con la libreria Pandas, fino all'implementazione di tecniche di preprocessing. Dando comunque il giusto peso alle istruzioni per il controllo di flusso (if, while e for) e alle tipologie più importanti di dati (liste, tuple, dizionari e set). Il programma:

- Installazione di Python, Jupyter e le principali librerie
- Import di file CSV, Excel e JSON
- Caratteristiche dei DataFrame
- Import di file strutturati
- Creare funzioni e gestire parametri di input e output
- Approfondimenti sul ciclo for e le liste
- Esempi di utilizzo dell'istruzione IF e della funzione input
- Utilizzo di variabili booleane e del costrutto while
- Approfondimento sulle istruzioni break e continue

- Utilizzo del tipo set
- Gestione degli errori con try-except
- Dizionari e import di Excel con fogli multipli
- Analogie e differenze tra liste e tuple

Nozioni di statistica (5+ ore)

In questo modulo imparerai a **descrivere le variabili**, prima singolarmente, calcolando indici di posizione, variabilità e forma, e poi combinate tra loro, per studiare il grado di associazione reciproco. Verranno introdotti anche i principali concetti di **statistica inferenziale** ed eseguiti i vari **calcoli con Excel, SQL e Python**, in modo da mettere subito in pratica quanto appreso. Il programma:

- Introduzione alla statistica
- Classificazione delle variabili statistiche
- Indici di posizione centrale
- Indici di variabilità ed eterogeneità
- Outlier e coefficiente di variazione
- Rappresentazione grafica delle variabili
- Scaling di una variabile
- Indici di forma
- Calcolo degli indici statistici con Excel, SQL e Python
- Indice V di Cramer
- Indice Eta quadro
- Covarianza tra due variabili
- Introduzione alle regressioni lineari
- Coefficiente di Pearson
- Coefficiente di Spearman
- Concetti di statistica inferenziale: campione vs popolazione
- Stime puntuali
- Test di Shapiro Wilks per la normalità di una popolazione
- Intervallo di fiducia per la media di una popolazione normale
- Test di Student sulla media di una popolazione normale

Crea algoritmi di Machine Learning con Python (4+ ore)

Questa sezione è dedicata all'implementazione di modelli **di machine learning con Python**. Inizieremo imparando a scrivere noi, da zero, il codice di importanti algoritmi come il Perceptron e il K-neighbours, utilizzando la programmazione a oggetti di Python e la libreria Numpy. In seguito

impareremo ad automatizzare le nostre attività con la famosa libreria **scikit-learn** e a visualizzare correttamente i dati con **seaborn** e **folium**.

- Implementazione del Perceptron
- Perceptron e programmazione a oggetti
- Preprocessing con Scikit-Learn
- Algoritmi di classificazione supervisionata con Scikit-Learn
- Implementazione dell'algoritmo K-neighbours
- Analisi esplorativa dei dati: calcolo dei principali indici statistici e rappresentazioni grafiche con Seaborn
- Machine Learning sul dataset Titanic: studio delle variabili, encoding e normalizzazione
- Creazione di Pipeline e convalida k-fold
- Ottimizzazione dei parametri
- Machine Learning sul dataset Diamonds e introduzione alla regressione
- Algoritmi di Regressione, clustering e analisi dell'accuratezza
- Rappresentazione di dati geografici e mappe

Analisi dati con MongoDB (5+ ore)

MongoDB è il database non relazionale più richiesto dalle aziende e conoscere questo strumento potente e flessibile è ormai essenziale per ogni Data Analyst e Data Scientist. Grazie a questo modulo potrai inserire nel tuo curriculum una competenza chiave ancora poco diffusa, che ti aiuterà a differenziarti da altri candidati durante i processi di selezione del personale. Il programma:

- Installazione di MongoDB e creazione Database
- Documenti JSON
- Collezione JSON
- Esecuzione dei primi filtri su MongoDB
- Scelta delle chiavi e altri esempi di condizioni
- Analisi dei tipi e relativo impatto sui risultati del metodo `.find()`
- Gestione dei null, espressioni regolari e filtri su array
- Aggregare i dati su MongoDB
- Eseguire filtri pre e post raggruppamento
- Eseguire raggruppamenti su chiavi di tipo array e sotto-documento
- Combinare i dati di più collection
- Creare database, collezioni e documenti
- Aggiornare il database
- Modificare chiavi di tipo array
- Funzioni `$concat`, `$ifNull` e `$cond`, `$push` e `$addToSet`
- Approfondimenti su indici, dati geospaziali e integrazione con Python
- Creazione di viste e regole di validazione
- Confronto con i Database Relazionali
- Esercizi MongoDB

- Project Work

I database Access (5+ ore)

Access è un programma utilizzato soprattutto dalle aziende che si sono affacciate da poco al mondo della Data Analysis e si sono scontrate con i limiti di Excel. È fondamentale quindi acquisire una conoscenza di base di questo software, ripassando la **struttura e le caratteristiche di un database relazionale**. Il programma:

- Database Relazionali e progetti di analisi dei dati
- Creazione delle prime tabelle su Access
- Inserimento di nuovi record nelle tabelle
- Chiavi primarie, chiavi esterne e relazioni tra tabelle
- Campi con ricerca guidata
- Popolare un Database di Access a partire da dati di Excel
- Scrivere le prime query di selezione dei dati
- Combinare i criteri con And e Or
- Join tra tabelle
- Raggruppamenti e aggregazioni con Access
- Query di aggiornamento, cancellazione e accodamento
- Campi calcolati
- Query parametriche
- Query a campi incrociati e d'unione
- Individuazione di duplicati e record mancanti
- Creazione di Maschere
- Maschere di sola lettura, inserimento di grafici, utilizzo di particolari layout e maschere di Home Page
- Maschere parametriche
- Creazione di Report semplici e con raggruppamenti
- Esercizi
- Project Work di Analisi dei dati
- Project Work sulla creazione di un applicativo con Microsoft Access

Crea Dashboard professionali con Power BI Desktop (5+ ore)

Power BI Desktop è uno dei software più utilizzati per la Business Intelligence e la Data Visualization. In questo modulo vedremo come importare, trasformare e presentare i dati creando **dashboard interattive e di grande impatto visivo**. Inoltre, impareremo a utilizzare il potente strumento Power Query per controllare le procedure di import e arricchire il modello con il linguaggio DAX. Il programma:

- Installazione di Power BI Desktop
- Operazioni di import dei dati con Power Query
- Creazione delle prime Dashboard
- Formattazione e interazione con Dashboard
- Gestione degli errori e dei formati con Power Query
- Trasformazione e pulizia dei dati
- Combinare più fonti in un'unica query
- Raggruppare e filtrare i dati
- Creazione di relazioni tra tabelle
- Dashboard che combinano i dati di più tabelle
- Utilizzo di alcune funzioni DAX: RELATED, BLANK, DATEDIFF, IF
- Gestione di stringhe e date con il DAX
- Creazione e utilizzo di misure con il DAX
- Approfondimenti sulle rappresentazioni grafiche e matriciali
- Analisi grafiche per confrontare colonne numeriche
- Misure con DAX
- Funzioni FILTER e CALCULATE
- Grafici combinati, a nastro, a dispersione e alberi di scomposizione
- Navigare la dashboard con il Drill-through
- Project Work

Usa Tableau per la Business Intelligence (2+ ore)

Approfondiamo la conoscenza degli strumenti per la Business Intelligence con un'introduzione a **Tableau**, nella sua versione didattica Public Edition. Vedremo come utilizzare questo software per **realizzare dashboard**, ma anche per acquisire dati in vari formati, arricchire le informazioni con le colonne calcolate, gestire i dati mancanti e creare relazioni tra diverse fonti di dati. Il programma:

- Installazione e casi d'uso di Tableau Public
- Import e arricchimento dei dati
- Grafici per combinare dimensioni e misure
- Creiamo e interagiamo con la nostra prima Dashboard
- Funzioni per la gestione dei dati mancanti
- Applicare filtri sui dati
- Rappresentazioni tabellari
- Grafici per analizzare contemporaneamente più misure e più dimensioni
- Import dei dati da file di testo con delimitatore
- Import di file JSON
- Import da Access
- Creare relazioni tra più fonti dati

Diventa un esperto di SQL Server (6+ ore)

Nell'ultima parte del corso ci concentreremo su **SQL Server**, uno dei database più diffusi nelle grandi aziende. Analizzeremo le tecniche più efficaci per migliorare e monitorare le prestazioni di un database SQL server e scrivere query più efficienti. Il programma:

- Creazione database di test
- Variabili, viste parametriche, Cross Apply e Outer Apply
- Window function
- Pivot e Unpivot dei dati
- Grouping sets - rollup - cube
- Transazioni e gestione degli errori
- Chiavi primarie, indici clustered e non clustered
- Fattori che influenzano l'utilizzo di un indice non clustered: operatori key lookup e scan
- Operazioni di ordinamento e rimozioni duplicati
- Algoritmi di join
- Covered index, Filtered Index e indici multi-colonna
- Importanza dell'ordine delle colonne in un clustered index
- Tuning di query: sostituire quando possibile UNION con UNION ALL, Condizioni Sargable
- Tuning di query: utilizzare la clausola GROUPING SETS, confronto tra LEFT JOIN e NOT EXISTS
- Tuning di query: utilizzo di tabelle temporanee e analisi delle performance delle Window function
- Tuning di query: pre-raggruppamenti, SQL dinamico, riflessioni su viste, join e clausola IN
- Cercare proattivamente le query più problematiche
- Warning nei piani d'esecuzione e approfondimento sulle statistiche

Progetti finali

In questa sezione ti presenterò la traccia di nuovi progetti di analisi dei dati da svolgere in autonomia, pensati per applicare quanto imparato durante i corsi. Ti basterà inviarmi i tuoi lavori via mail per ricevere un feedback dettagliato su quanto svolto.

- Riflessioni su licenze e citazioni
- Import dei dati e progetto con SQL
- Import dei dati e progetto con Pandas
- Import dei dati e progetto di Business Intelligence
- Import dei dati e progetto di Machine Learning
- Progetto di analisi dei dati: confronto tra due fonti dati
- Progetto Fanta Formula 1 con T-SQL avanzato
- Progetto Fanta Formula 1 con strumenti di Business Intelligence

Torna alla pagina del corso <https://www.yimp.it/corso-data-analysis/>