

Percorso completo di Data Analysis

Programma completo (40+ ore)

Comincia dall'SQL (7+ ore)

Il primo video-corso che ti consiglio di seguire è l'**SQL Academy**. L'SQL è il linguaggio di programmazione per gestire e interrogare i Database Relazionali. Per quanto non sia il linguaggio più "trendy" e in "hype" nelle varie pubblicità dei master o nei post su LinkedIn, posso assicurarti che è la skill fondamentale per chiunque voglia approcciarsi da professionista al lavoro dell'analisi dei dati. I Database Relazionali sono infatti diffusissimi in tutti i contesti aziendali e se "sai scrivere una query" è una domanda che ti faranno in praticamente ogni colloquio relativo ad una posizione nel mondo Tech. 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: parte 1
- 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
- Creare un database
- Vincoli di chiavi primarie, esterne e check
- Popolare e aggiornare un database
- 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

Fai un ripasso di Excel (3,5+ ore)

Per quanto abbia dei limiti per i quali in alcuni casi è necessario passare ad uno strumento più potente come un Database, **essere fluenti in Excel** è comunque una skill fondamentale per chiunque

voglia lavorare con i dati. La sua iper-diffusione in tutti gli ambiti e uffici aziendali e le sue infinite possibilità di utilizzo tra formule, pivot e grafici lo rendono uno strumento indispensabile. Non ti servirà essere un "mega-esperto" perché vedrai che, *arrivati ad un certo punto*, le aziende passano ad utilizzare altri strumenti. Ma può essere possibile saper creare una rete neurale su Python senza saper fare un grafico su Excel? Hai a disposizione due mini-corsi per imparare a **scrivere formule e funzioni** (questo corso è gratuito per tutti!) e **creare tabelle pivot**. 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
- 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
- Esercizi

Un po' di statistica è essenziale (4+ ore)

Prima di partire con il Machine Learning, sarà essenziale fare un ripasso dei principali concetti di statistica. Impareremo 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. Introduciamo anche i principali concetti di statistica inferenziale, come le stime e gli intervalli di fiducia. Eseguiremo i vari calcoli con Excel, SQL e Python in modo da mettere subito in pratica i concetti teorici studiati. 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 (6+ ore)

A questo punto potrai iniziare a "*fare sul serio*" imparando le tecniche di programmazione per l'**Analisi dei dati e il Machine Learning con il linguaggio Python**. Il corso è strutturato per partire da zero con Python, seguendo un approccio innovativo e concreto: non seguiremo il classico canovaccio teorico *variabili, liste, tuple, dizionari, eccetera* ma impareremo a programmare tramite una serie di task reali dei processi di Analisi dei dati: imparerai ad acquisire i dati, lavorarli con la libreria pandas, implementare tecniche di PreProcessing e Machine Learning tramite la programmazione a oggetti e la libreria scikit-learn. Il programma

- Installazione di Python, Jupyter e le principali librerie
- Import di file CSV, Excel e JSON
- Caratteristiche dei DataFrame
- Import di file strutturati
- Organizzare il codice in funzioni
- Filtrare i dati
- Aggregare i dati - metodo groupby
- Combinare i dati - metodo merge
- Combinare i dati - metodo concat
- Ordinare i dati e modificare la struttura di un DataFrame
- Modificare e trasformare i dati in un DataFrame
- Esercitazione sul pre-processing dei dati: import del file Iris, filtrare e ordinare i dati, gestire i null, gestione variabile target e suddivisione in training e test, valorizzazione null, normalizzazione dati, aggiunta colonna pesi e conversione dati in array di numpy
- Implementazione del Perceptron
- Perceptron e programmazione a oggetti
- Machine Learning con Scikit-Learn - parte 1
- Machine Learning con Scikit-Learn - parte 2

- 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

Amplia le tue skill con MongoDB (5+ ore)

MongoDB è il **Database Non relazionale** con maggiore richiesta nel mercato del lavoro. Il formato JSON è ormai uno standard nella strutturazione dei dati; conoscere uno strumento potente e flessibile come MongoDB per la gestione e analisi dei dati in questo formato è un must-have per ogni Data Analyst e Data Scientist. Con il **modulo di MongoDB** potrai inserire una competenza chiave nel tuo curriculum ma ancora non molto diffusa e che quindi ti distinguerà da molti altri professionisti. 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

Ripassiamo la struttura dei DB con Access (5+ ore)

Il **corso di Access** ti permetterà di fare un utile ripasso sulle peculiarità e le strutture di un Database relazionale. Spesso Access è il primo passo fatto dalle aziende che hanno iniziato a lavorare in modo strutturato sui dati e si sono scontrate con i primi limiti di Excel. Anche in questo caso non ci servirà diventare "mega-esperti" di Access in quanto a questo punto avremo già le basi per utilizzare strumenti più avanzati come SQL Server, MongoDB o Python. Tuttavia non possiamo pensare di presentarci come esperti di Database senza avere una dimestichezza di base con il famoso prodotto del pacchetto Office. 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

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

Il modulo di Power BI Desktop ti insegnerà le basi di uno dei software più utilizzati per la Business Intelligence. Imparerai ad importare, trasformare e presentare i dati tramite Dashboard professionali, interattive e dall'alto impatto visivo. Vedremo come utilizzare il potente strumento Power Query per controllare vari step nelle procedure di import, ponendo particolare attenzione alla gestione degli errori e dei formati. Dopo aver creato le opportune relazioni tra le tabelle e arricchito il nostro modello con il linguaggio DAX, saremo pronti a creare e personalizzare con pochi click dei grafici dall'alto potere informativo. 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

Diventa un esperto di SQL Server (5+ ore)

A questo punto avremo delle solide basi di molti strumenti di analisi dei dati, l'ultimo passo è quello di specializzarsi in qualcosa per potenziare le skill su un particolare strumento. Quello che troverai nella mia libreria è un **video-corso con contenuti pressoché unici** in italiano sul Database SQL Server, analizzando le tecniche più moderne ed efficaci per migliorare e monitorare le prestazioni di uno dei database più utilizzati dalle principali aziende mondiali. 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

Torna alla pagina del corso <https://www.yimp.it/corso-completo-di-analisi-dei-dati/>