UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Python and Databases - 2

2627-2-E1807M015-T2

Obiettivi formativi

Il corso si propone di fornire agli studenti una comprensione generale dei temi dell'innovazione tecnologica nelle imprese e nella società, della gestione dei dati e delle informazioni, dei sistemi informativi aziendali e delle loro implicazioni organizzative e gestionali, con particolare riguardo al settore dei servizi finanziari e dell'amministrazione delle imprese.
Al termine del corso, lo studente sarà in grado di:

1. Comprendere le logiche della programmazione in Python.
2. Interrogare ed estrarre dati da database relazionali utilizzando il linguaggio SQL.
3. Utilizzare le librerie di Python per pulire, manipolare e analizzare dati economici e aziendali (es. vendite, bilanci, anagrafiche clienti).
4. Integrare i due mondi: estrarre dati da un database tramite codice Python e automatizzare la reportistica di base.
   Con l'analisi dati con Python e SQL lo studente potrà trasformare dati grezzi in insight strategici, automatizzando il processo ETL (estrazione, trasformazione, caricamento), pulendo dataset complessi, eseguendo analisi esplorative (EDA) e creando modelli predittivi o report visuali per supportare le decisioni, aziendali basate su fatti concreti, coprendo l'intero ciclo di vita del dato, dalla raccolta alla modellazione.

Contenuti sintetici

Il corso permette agli studenti di usare gli strumenti Python e database SQL in modo integrato e consapevole, ai fini dell'analisi e della visualizzazione dei dati in tutti gli ambiti aziendali, con particolare riferimento al mondo economicio e dell'amministrazione delle imprese.
Dopo lo studio dei fondamenti di Python e dei database SQL per l'analisi dei dati e previsioni economico-aziendali, si procederà all'integrazione dei due strumenti al fine di supportare decisioni aziendali basate su fatti concreti.
Nell'ultima parte del corso si analizzerà un caso di studio, ad esempio analizzando le performance di vendita di un'azienda.

Programma esteso

Fondamenti di Python per il business
L'obiettivo è fornire le basi della programmazione, superando lo scoglio iniziale della sintassi.
Introduzione: Python per l'economia. L'ambiente di lavoro Anaconda e Jupyter Notebook.
Basi del linguaggio: Variabili, tipi di dato (numerici, stringhe, booleani), liste e dizionari. Rappresentare informazioni aziendali semplici.
Controllo di flusso: Cicli (for, while) e costrutti condizionali (if, elif, else). Creazione di funzioni di base per calcoli economici (es. calcolo IVA, sconti, VAN semplice).
Analisi dei dati con Python.
L'obiettivo é passare da dati grezzi a informazioni utili.
Introduzione a Numpy e Pandas: L'ecosistema dei dati in Python. Importazione di file CSV ed Excel aziendali.
Gestione dei DataFrame: Esplorazione dei dati, selezione di righe e colonne, gestione dei dati mancanti.
Manipolazione avanzata: Raggruppamenti (groupby), tabelle pivot in Python, unione di dataset diversi.
Basi di dati e SQL.
Come vengono archiviati i dati nelle grandi aziende e come estrarli.
Modellazione dei dati: Cos'è un database relazionale. Entità, attributi, chiavi primarie e chiavi esterne. Lettura di un diagramma E-R (Entity-Relationship) applicato a un caso aziendale (es. Clienti - Ordini - Prodotti).
Estrazione dati con SQL - Le Basi: Il linguaggio SQL. Le istruzioni SELECT, FROM, WHERE, ORDER BY. Operatori logici (dopo Python, in SQL).
Aggregazioni e Relazioni: Funzioni di aggregazione (SUM, AVG, COUNT), GROUP BY. Come incrociare tabelle diverse utilizzando le JOIN.
Integrazione e Caso di studio aziendale
L'integrazione dei due strumenti e la messa in pratica.
Python + SQL: Connettere un database relazionale tramite Python usando le librerie apposite. Eseguire query SQL direttamente all'interno di un Jupyter Notebook e salvare i risultati in un DataFrame Pandas.
Visualizzazione di base: Cenni a Matplotlib per creare grafici essenziali, quali barre o linee di tendenza, da inserirein report aziendali.
Progetto finale / Caso di studio: Esercitazione guidata end-to-end, ad esempio sull'analisi delle performance di vendita di un'azienda. Dalla query sul database per estrarre lo storico vendite, passando per la pulizia dei dati in Pandas, fino alla generazione di un grafico sulle marginalità di prodotto.

Prerequisiti

Conoscenze matematiche e logiche di base. Conoscenza base della lingua inglese.
Utile la conoscenza dell'organizzazione tabellare dei dati.

Metodi didattici

Il corso è tenuto in lingua italiana. Alcune slide/testi e parti di approfondimento potranno essere in lingua inglese.
Lezioni, esercitazioni e sessioni laboratoriali in aula/laboratorio. Alcuni esercizi, approfondimenti, project work potranno essere assegnati in laboratorio/aula da terminare a casa.
Le lezioni saranno erogate prevalentemente in presenza.

Modalità di verifica dell'apprendimento

La verifica dell'apprendimento consiste in:

1. Una **prova scritta** esercizi sul programma svolto quali - titolo esemplificativo - modellazione basi di dati, query SQL, script Python che prevedono l'uso di DataFrame Pandas;
2. Integrazione eventuale - a discrezione del docente per specifici casi - attraverso una **prova orale** di discussione della prova scritta ed eventuali integrazioni attraverso questiti specifici.

La prova sarà valutata in trentesimi.

Testi di riferimento

Downey, A. B. (2015). Think Python: How to Think Like a Computer Scientist, Version 2.4. Green Tea Press. (disponibile online con Creative Commons licence, https://greenteapress.com/thinkpython2/thinkpython2.pdf).
W3C Python tutorial (https://www.w3schools.com/python/).
Dispense, materiale del corso, eventuali testi addizionali pubblicati sulla pagina elearning.

Periodo di erogazione dell'insegnamento

Secondo semestre

Lingua di insegnamento

Italiano

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ | PARITÁ DI GENERE | LAVORO DIGNITOSO E CRESCITA ECONOMICA | IMPRESE, INNOVAZIONE E INFRASTRUTTURE