

BI Developer

עם התמחות ב- Python ו-Snowflake

370 שעות

תיאור הכשרה:

בעולם טכנולוגי ועסקי, הנתונים והמידע הם הנכס האמיתי של כל ארגון. מסלול זה מכשיר את הסטודנטים לתפקיד מבוקש מפתח BI תוך כדי מתן דגש לכלים וטכנולוגיות המבוקשות כיום בשוק כגון שפת **Python ו-Snowflake**.

בינה עסקית היא תחום בטכנולוגיית המידע העוסק בבניית מערכות העוזרות לארגון להפיק מידע משמעותי מבחינה עסקית מתוך מכלול הנתונים הנאספים על ידיו. מערכות הבינה העסקית מספקות מידע היסטורי, מידע עכשווי ותחזיות בנוגע לפעילות העסקית, ובעזרתן ניתן לאתר דפוסים משמעותיים בהתנהגות והתנהלות העסקית לטובת ניהול של ארגון, וזאת ע"י בין היתר יצירת דוחות, התראות ואיתותים לצורך קבלת החלטות ניהוליות.

הצורך ב-BI קיים כמעט בכל ארגון על מנת ללמוד מהנתונים על התנהלות העסקית של הארגון ולהוביל לאופטימיזציה של התהליכים.

תוכנית הלימודים במסלול ייחודי זה הינה העדכנית ביותר וכוללת הכשרה על כלים מסורתיים בתחום BI לצד כלים וטכנולוגיות חדשות שנכנסו לארגונים רבים בתקופה אחרונה.

מבנה הקורס:

פיתוח בשפת SQL ו-TSQL - מודול זה מתמקד בכלים המאפשרים הגדרה וזיהוי של נתונים לשימוש במסגרת תהליך BI, בניית מאגר הנתונים, ניתוח הנתונים, הגדרת תפוצת הדיווח ודיווח התוצאות. המודול כולל:

- **שפת SQL** - רכישת המושגים הבסיסיים במסדי נתונים, תרגול השפה ושאלות בסיסיות.
- **שפת TSQL** - לימוד של תהליכי יצירה והגדרת טבלאות ואובייקטים שונים, עבודה עם שאלות מורכבות. T-SQL – הינה שפת הפיתוח שבאמצעותה מועברות הפקודות למערכת (על ידי כתיבת פקודות או שימוש בממשק גרפי ששולח פקודות באופן סמוי).

מחסן נתונים ותהליכי ETL – מודול המבוסס על טכנולוגיות BI של חברת מיקרוסופט, במהלכו תרכשו הבנה מעמיקה של תהליך בניית תשתית לתעבורת הנתונים העסקיים של הארגון לשימושים עתידיים. הסטודנטים ילמדו ויתרגלו על עבודה מול מחסן נתונים – **Data Warehouse (DWH) ותהליך ETL** באמצעות הכלי של מיקרוסופט **SSIS - SQL Server Integration Services** - כלי ETL המאפשר שליפת מידע ממקורות שונים, עיבודם וטעינתם ל-DWH.

בניית מודל נתונים וניתוח נתונים באמצעות שני כלים מובילים PowerBI ו-Tableau – מודולים מקנים היכרות והבנה מעשית בעבודה עם הכלים מקצה לקצה. בין הנושאים הנלמדים: עבודה מול מקורות נתונים מגוונים, בניית מודל נתונים, בניית Dashboard, חישובים בסיסיים ומתקדמים, עבודה מול שרת ועוד.

תחום נוסף שנלמד הקורס הינו אחד התחומים החמים בעולם האנליזה, אשר גרם להתפתחות וכניסה של כלים כמו Tableau ו-PowerBI – הינו תחום **Data Visualization**. במודול הזה נלמד עקרונות חשובים בויזואליזציה של הנתונים, ובניית דשבורדים תוך כדי יישום של עקרונות אלה.

Data Manipulation with Python - מודול זה מעניק ידע מעשי בפיתוח בשפת **Python**, שהיא השפה המובילה כיום בעבודה עם נתונים. במהלך המודול נלמד כלים לעבודה עם נתונים ממקורות שונים ולהצגתם. נפתח מאפס קוד, נלמד את הנושאים הנדרשים בסביבה מונחית-עצמים (Object-Oriented), שהיא המתודה הסטנדרטית כיום בפיתוח תוכנה. בנוסף, נכיר את הספריות שמציעה השפה לעבודה עם נתונים, החל מהספריה העשירה של השפה Python Standard Library ונדע כיצד להיעזר בה. חבילות יעודיות נוספות של פייתון, **Pandas, Numpy, Matplotlib** יעניקו לסטודנטים ידע במניפולציה על הנתונים, כולל תחקור הנתונים, ביצוע חישובים מתמטיים ויצירת גרפים לצורך הצגת ממצאי התחקור. את השלב הזה נסכם בפרויקט מעשי במסגרתו הסטודנטים יציגו תהליך בניית Data Pipeline עם אפשרויות ניתוח והפקת תובנות מהנתונים.

בלעדי בנאייה קולג'! מודול Snowflake - הכלי החדשני והמוביל בתחום BI.

ריבוי מקורות המידע, גידול בכמות הנתונים והשימושים השונים והמגוונים שנדרש לבצע על הנתונים תוך כדי צורך גובר ביעילות, פשטות ומהירות הביאו לכך שמחשני הנתונים המסורתיים כבר לא יכולים להוות מענה הולם. פלטפורמת Snowflake נועדה לתת מענה בדיוק לדרישות בשטח ע"י יצירת בסיס נתונים אנליטי המנוהל בענן ומאפשר לייבא נתונים לתוך מערכת מרכזית אחת המסוגלת לתמוך בנפחי מידע גבוהים ועם יכולות שליפה מהירות מאד.

Snowflake מצמצמת באופן משמעותי את העלויות, המורכבות של המערכת וחוסר הגמישות של מערכות מסורתיות, ומאפשרת אחסון ועיבוד של הנתונים תוך כדי ניצול היתרונות העצומים של סביבת הענן לטיפול גם בנתונים ביג דאטה.

מסיבות אלו נחשבת פלטפורמה זו לאחת הטכנולוגיות המובילות כיום את תחום הדאטה בענן.

במסגרת מודול ייחודי ובלעדי זה, נלמד את הסטודנטים לטעון נתונים ל-Snowflake, לשלב קוד SQL עם הפלטפורמה, נלמד על היכולות הניתוח והאנליטיקה של הפלטפורמה וכיצד היא ממנפת את היכולות של SQL וכלי BI ארגוניים.

פרויקט התמחות – פרויקט מעשי המסכם את הידע שנרכש במהלך ההכשרה. תהליך ההתמחות מאפשר לסטודנטים לבצע אינטגרציה בין התכנים השונים שנלמדו במסלול, לצבור ניסיון מעשי ולרכוש דרכי עבודה נכונות כדי לבסס את דרככם כעובדים בתעשיית ההייטק בישראל. תהליך ההתמחות יתבצע במסגרת משולבת של ליווי מנטור (במכללה) ועבודה אישית בבית. מטרת תהליך ההתמחות, לאפשר לסטודנטים לצבור ניסיון מעשי ולבצע פרויקט מסדי נתונים ובינה עסקית.

קהל היעד:

מסלול הכשרה זה מיועד למועמדים בעלי רקע טכני בסיסי במערכות מידע, אנשי תשתיות, מפתחים, בודקי תוכנה. כמו כן המסלול מיועד לבעלי תארים אקדמיים בתחומים מערכות מידע, תעשייה וניהול, מנהל עסקים, כלכלה, ניהול עם התמחות במערכות מידע, בעלי אוריינטציה טכנולוגית המעוניינים להשתלב בעבודה כמפתחי BI ומומחי Data.

דרישות הקורס:

- על מנת לתרגל בשעות הפנאי במחשבים האישיים, לא במסגרת השיעורים בכיתה, יש להצטייד במחשב חזק שיעמוד בדרישות הטכנולוגיות.
- נוכחות 80% מינימום
- הצלחה בבחינות ופרויקטים

תוכנית לימודים:

Part 1: SQL Basic

שיעורי מכינה לשפת SQL ולמושגים בסיסיים ביותר במסדי נתונים הכוללים תרגול מעשי.

- Setting up development environment - Introducing SQL Server Management Studio
- Querying Data Using T-SQL Language
 - Basic Concepts of Databases
 - SQL Server Data Types
 - Retrieving Data: SELECT Statement
 - Filtering Data: WHERE, Logical Operators and Expressions
 - Organizing Retrieved Data: ORDER BY, GROUP BY, TOP N
 - Using System Supplied Scalar Functions
 - Using Aggregate (group) Functions
 - Multiple Table Access: JOINS
 - Using Simple Sub-Queries
 - Updating Data Using DML Statements: UPDATE, INSERT, DELETE

Part 2: Database Design

מודול זה מקנה לתלמידים הבנה בסיסית וידע בהקמת מסדי נתונים ובהגדרות התצורה הפיזית שלהם, וכן בשימוש בשפת SQL ליצירה של אובייקטים ולהגדרת יחסי גומלין ו-Data Integrity.

-
- Database Design
 - Working with Data Types
 - Designing and Implementing Tables
 - Ensuring Data Integrity through Constraints
 - Table Structures and Indexing
 - Indexes and SQL statements performance
 - Advanced Table Structures: Partitioning, Columnstore Indexes

Part 3: Advanced SQL techniques and T-SQL Programming

מודול זה מקנה לתלמידים את עקרונות התכנות בשפת T-SQL לצורך כתיבת של תוכניות לעיבוד נתונים (תנאים, לולאות, פרוצדורות, פונקציות וטריגרים) ולניסוח מתקדם של שאילתות SQL.

- Advanced SQL tools and techniques
 - Working with Advanced Subqueries (Derived Tables, Correlated subqueries, EXISTS)
 - Using Ranking and Analytic functions
 - The MERGE Statement
 - Understanding Transactions and Locks
- Programming with Transact-SQL
 - Using variables
 - Displaying data to the Query and Results panes
 - Control flow – Loops and Conditions
 - Working with result sets – Temporary tables and Table variables
- Database Code Objects
 - Designing and Implementing Views
 - Designing and Implementing Stored Procedures
 - Designing and Implementing User-Defined Functions
 - Additional code objects topics – Overview
 - Working with triggers
 - .NET languages (CLR) integration
 - R language integration
 - XML and JSON support – Working with semi-structured data

Part 4: Data Warehouse Fundamentals

מבוא לבינה עסקית - מודול המיועד להציג את העקרונות הבסיסיים בתכנון פרויקט BI ובתכנון ועיצוב מחסן נתונים (DWH).

- Business Intelligence and Data Warehousing
 - Defining Data Warehouse Concepts and Terminology
 - Planning and Managing the Data Warehouse Project
- Designing a Dimensional Model
 - Normalization
 - Star & Snowflake Schemas
- Advanced Data Modelling
 - Surrogate keys
 - Snapshots and Logs
 - Slowly Changing Dimensions

Part 5: Integration Services 2016 (SSIS)

מודול זה מתמקד בפיתוח של תהליכי ETL מבוססים SSIS.

- Introduction to SQL Server 2016 Integration Services

- Introduction to ETL architecture
- Introduction to Microsoft SQL Server Integration Services.
- Connection Manager Levels.
- Basic Data Flow Task Transformations.
- Control Flow
 - Using Variables & Parameters
 - Control Flow Tasks
 - Control Flow Precedent Constraints
 - Control Flow Containers
- Debugging and Troubleshooting
 - Debugging
 - Logging & Event Handlers
- Deploying and Configuring SSIS Packages
 - Deploying Packages

Part 6: ETL Concepts and Techniques

מודול זה עוסק בשיטות המקובלות ובתכנון תהליכי טעינת הנתונים למחסן הנתונים (ETL).

- ETL Concepts and Techniques
 - Designing the ETL process
 - The Staging area
 - Full and Incremental Loading Techniques
 - Retaining Data Integrity
 - QA and Validation
 - ETL Administration and Monitoring

Part 7: Power BI

מודול זה מתמקד בהקמת מודל נתונים ובניית תשתית לניתוח נתונים באמצעות הכלי. כולל שימוש בשפת DAX והקמה של דוחות אינטראקטיביים ודשבורדים ניהוליים בפלטפורמה החדשנית של Power BI.

מבוא

- Self Service BI
- מבנה וארכיטקטורה של הכלי
- תהליך פיתוח עם הכלי (הבנת ה-Flow)

ייבוא נתונים ועריכתם באמצעות Power BI Query Editor

- היכרות עם Power BI Query Editor
- Extract - יבוא נתונים ממקורות שונים :
 - קבצים (אקסל, טקסט)
 - SQL
 - תיקיות
- Transform - עריכת הנתונים ואוטומציה לתהליך
- יצירת מימד תאריכים (יאפשר חיתוך והצגת המידע לפי יום בשבוע, חודש, רבעון, שנה)
- פיתוח אוטומציה לעריכת נתונים :
 - חיבור בין טבלאות
 - יצירת עמודות מחושבות
 - סינון נתונים

- פיצול וחיבור עמודות
- טיוב נתונים

- היכרות עם שפת "M"

מידול - Modeling ויצירת מדדים אשר יאפשר לנו בהמשך חקר נתונים אינטרקטיביים

- כללים ביצירת מודל נתונים
- מתודולוגיות בפיתוח מודל נתונים
- היכרות עם Star Schema ו-Snowflake Schema
- התווית קשרים בין טבלאות

מדדים

- סקירה ואפיון של מדדים ו-KPI's
- הכרות עם שפת Dax והשפעת מודל הנתונים על כתיבת המדדים
- פיתוח מדדים בשפת Dax
- פיתוח מדדים מבוססי זמן לשוואה בין תקופות ונתונים מצטברים

דוחות

- עיצוב ופיתוח של דוחות
- טיפים ושיטות עבודה לעיצוב דוחות
- שימוש בפילטרים וויזואליזציות שונות
- הוספת ויזואליזציות ומניפות צבעים מחנות מיקרוסופט
- שימוש ב Bookmarks, tooltips ו-Drillthrough
- שימוש בכפתורים
- עיצוב מותאם למובייל

Power BI Service – סביבת הענן של Power BI

- היכרות עם סביבת הענן
- העלאת המודל מכלי הפיתוח לבסיסת הענן
- התקנה וחיבור של רכיב ה-Gateway
- יצירת דשבורדים
- שיתוף דוחות ודשבורדים
- ניהול הרשאות למשתמשים ובסיסות עבודה
- קביעת זמני רענון אוטומטיים והגדרת מקורות הנתונים ב-Gateway

Part 8: Tableau

הכלי המוביל לניתוח נתונים עסקיים, שהופך במהירות טבלאות ומספרים מסורבלים להמחשות גרפיות ברורות. כלי לבנייה והצגת דוחות דינאמיים בתצורה גרפית מתקדמת המאפשר יכולת ניתוח סטטיסטיות ייחודיות.

- Introduction
- Working with single and multiple data sources
- Connecting to Data
 - Creating a Data Source
 - Manipulate Data
 - Advanced Data Sources
- Desktop Basics
 - Tabs
 - Status bar
 - Panes

-
- Cards
 - Toolbar/Menus
 - Show Me
 - Full Screen
 - Creating a Report
 - Understanding Fields
 - Basic Worksheet Options
 - Filtering & Sorting
 - Viewing Data
 - Parameters
 - Organizing Data
 - Advanced Reports
 - Calculations
 - Advanced Quick Table Calculations
 - Advanced Filtering & Sorting
 - Advanced table calculations
 - LOD calculations
 - Data Blending
 - Data Source Performance Considerations
 - Advanced Data Blending
 - Advanced Charts
 - Table options
 - Analytics
 - Maps
 - Pages
 - Highlighter
 - Formatting
 - Captions
 - Annotations
 - Text manipulation
 - Format menus
 - Creating and using parameters
 - Defining subsets of your data
 - Viewing distributions
 - Statistics and forecasting
 - Dashboards
 - Creating a new dashboard
 - Sizing and Layout
 - Adding objects
 - Filters
 - Dashboard actions
 - A Touch on Tableau Server
 - Roles and permissions
 - Data security

Part 9: Data Visualization

עקרונות בהצגה ויזואלית של נתונים - Data Visualization

- מה זה DataViz ולמה צריך את זה
- סוגי ויזואליזציות בעולם העסקי

- Story Telling - איך לספר את הסיפור של המספרים
- סוגי גרפים, מפות וטבלאות
- הטיות קוגניטיביות בהצגת נתונים + חוקי הגשטלש
- צבעוניות ומשמעותה
- כיצד להשתמש בטקסטים וטיפוגרפיה בצורה נכונה
- אינטראקטיביות ויתרונותיה בהצגת הדאטה

בניית דשבורדים שמתאימים למשתמשים

- מהו תהליך העבודה הנכון על דשבורד
- צרכים של המשתמשים
- סוגי דשבורדים
- בניית היררכיה ולייאוט - הגדרה, מטרה ופעולה
- בחירת סוג גרף, מפה או טבלה
- שימוש בצבעים כדי להדגיש את הבעיות
- שימוש בטקסטים וטיפוגרפיה בצורה נכונה
- אינטראקטיביות בדשבורדים

Part 10: Internship Project

פרויקט התמחות הינו משימה מובנית עם הנחיה ושלבי ביצוע הכוללת דרישה לשימוש נכון ויעיל בכלים שנלמדו עד לשלב הזה. התהליך הינו החל מעבודה מול מקור נתונים, כולל שליפה ועיבוד נתונים רלוונטיים לשאלה עסקית, בניית מחסן נתונים ותהליך ETL, בניית מודל נתונים ודשבורד ב-PowerBI תוך כדי שימוש בטכניקות ויזואליזציה מתאימות ומתקדמות. התוצר של הפרויקט הינו דשבורד שעונה לצרכים שהועלו בהצגת הצורך העסקי.

Part 11: Data Manipulation with Python

בפרק זה נלמד לתכנת בשפת Python ונתוודע אל סביבות העבודה Jupyter (IDE) ו-IPython. מעבר למושגי היסוד נכיר את הספריות השימושיות לעבודה על הנתונים.

Basic Python

- Fundamentals
 - Intro
 - Python essentials
 - The working environment
- Data types
 - Numbers
 - Strings
 - Booleans
 - None
- Collections
 - Lists
 - Tuples
 - Dictionaries
 - Sets
- Control flow
 - if...else
 - for...in
 - list comprehension
 - while
- Textual interface

-
- input
 - format

Intermediate Python

- Functions
 - User-defined functions
 - *args and **kwargs
 - Built-in functions
 - Lambda expressions
- Text files
- The standard library
 - import
 - datetime

Python Tools for working with Data

- The *NumPy* library
 - Array
 - Broadcasting
- The matplotlib library
 - matplotlib objects
 - Plotting
 - Seaborn
- The *pandas* library
 - Series and Index
 - DataFrame
 - GroupBy
 - Visualizations
- Practice: Use cases and EDA
 - Diamonds
 - IMDB
 - San Francisco Crime
 - Adult US Census
- General tools
 - Intro to regular expressions (re)
 - API's and Connecting with Data Resources (JSON, SQLAlchemy)

Part 12: Snowflake

- Introduction to Snowflake
 - Overview of Snowflake's architecture and key features
 - Comparison with other cloud data warehousing solutions
 - Getting started with Snowflake
 - Understanding the Snowflake computing and storage layers
 - Navigating the Snowflake web interface and using the worksheet
- Data loading and transformation
 - Loading data into Snowflake using various methods (e.g., COPY command, data connectors, etc.)
 - Understanding Snowflake's support for semi-structured data and JSON

-
- Transforming data using Snowflake's SQL
 - Data querying and analysis
 - Understanding Snowflake's query optimizer
 - Using Snowflake's SQL to query and analyze data
 - Optimizing query performance using features such as clustering, materialized views, and Result Cache
 - Using Snowflake's built-in analytics functions
 - Advanced data querying and analysis using Snowflake's Virtual Warehouses
 - Data integration and management
 - Integrating Snowflake with other tools and systems (e.g., Tableau, Power BI, etc.)
 - Managing and monitoring Snowflake using the web console and Snowflake's Data Management feature
 - Best Practices and real-world use cases
 - Best practices for data loading, transformation, and querying in Snowflake
 - Real-world use cases and examples of Snowflake in action
 - Case studies using Snowflake
 - Hands-on exercises and labs throughout the course to reinforce the concepts covered