

בית הספר למוסמכים במינהל עסקים ע"ש ליאון רקנאטי

1242.3271 – טיפול יישומי בנתוני אנליטיקה עסקית Data Handling for Business Data Analytics

סמסטר א' – תשפ"ג

קבוצה	יום בשבוע	שעה	מטלה	מרצה	דואר אלקטרוני
01	א'	21:30-18:45 (מחצית שניה)	בחינה	גב נועה אקר	noaeker@mail.tau.ac.il

שעת קבלה – בתיאום מראש במייל

עוזר הוראה – מר בן זרח bzerah@gmail.com

היקף הלימודים

היקף הי"ס לקורס : 1

1 י"ס = 4 ECTS – European Credit Transfer and Accumulation System (ECTS), ערך הניקוד של הקורס במוסדות להשכלה גבוהה בעולם שהינם חלק מ"תהליך בולוניה".

תיאור הקורס

תחום Data Science והשימושים שלו בתחום העסקי (Business Data Analytics) כוללים טכנולוגיות ושיטות המאפשרות איסוף ועיבוד כמויות עצומות של נתונים ממגוון רחב של מקורות שאינם בהכרח מאורגנים. נתונים אלו, המכונים באופן פופולרי Big Data מאפשרים לגופים עסקיים קבלת תובנות עסקיות חשובות ושיפור משמעותי בביצועים העסקיים.

שלב הכרחי בתהליכי Data Science – הינו שלב Data Processing שכולל השגת נתונים גולמיים ממקורות פנימיים וחיצוניים, עיבוד וארגון הנתונים, המרתם לפורמטים המתאימים לניתוח, וניהול הנתונים. נתונים גולמיים אלו יכולים להימצא במסגרת הארגון, באינטרנט, במאגרי מידע, או בידי גופים אחרים. הנתונים יכולים להופיע בצורה מאורגנת או בלתי מאורגנת (unstructured data), בפורמטים שונים, חלקם סטנדרטיים וחלקם ייחודיים.

בלימוד טכניקות Data Science ההנחה המובנית ברוב המקרים היא שכל הנתונים זמינים ומוכנים לניתוח. יחד עם זאת, בעבודה מעשית עם נתונים, נדירים מאוד המצבים בהם כל הנתונים זמינים ומאורגנים. מטרת קורס זה הינה להקנות לסטודנטים יכולות תכנות בסיסיות ושיטות עבודה מעשית עם נתונים: השגת נתונים ממקורות שונים, הכנתם לפורמט המתאים לניתוח נתונים, ניהולם, וארגונם לצורך הפקת תובנות מתוך הנתונים (כפי שנלמד בקורסים אחרים בפקולטה).

קורס זה יתמקד בראש ובראשונה ביסודות התכנות. יילמדו כלים ותכנות בשפת Python. באמצעות הכלים שיירכשו, יובנו תהליכים להשגת נתונים ממקורות שונים. תבוצע הכרות עם פורמטים API מקובלים באינטרנט לצורך "שאיבת נתונים" והכרות עם בסיסי נתונים רלציונים ושפת SQL לצורך ניהול נתונים.

קורס זה מיועד בעיקר לסטודנטים ללא רקע בתכנות (או רקע בסיסי בשפת תכנות שאינה Python) המעוניינים:

- לרכוש ידע בסיסי לעבודה מעשית עם נתונים בתחום Data Science.
- לרכוש כלים מעשיים להשגת נתונים ועיבודם.
- לנהל פרויקטים בתחום הדאטה ו/או לעבוד במקביל לאנשי דאטה, וכן לסטודנטים המעוניינים לבחון באופן ביקורתי דרכי השגת נתונים.

קורס זה מהווה קדם מומלץ (לאנשים ללא רקע תכנותי) לקורסים נוספים בהתמחות הדאטה.

תפוקות למידה

- עם סיום הקורס בהצלחה יוכל הסטודנט:
1. לתכנת תכנות בסיסי בשפת Python.
 2. להבין תהליכים השגת הנתונים ממקורות שונים.
 3. "לשאוב נתונים" מהאינטרנט ולעבדם תוך שימוש בפורמטים ו-APIs מקובלים.
 4. להבין את היסודות לבסיסי נתונים רלציונים ושפת SQL לניהול נתונים.
 5. להביא נתונים לתצורה המתאימה לניתוח מתקדם בהמשך (כפי שנלמד בקורסים אחרים בפקולטה).

הערכת הסטודנט בקורס והרכב הציון

אחוז	מטלה	תאריך	הערות
20%	4-5 תרגילי בית		
80%	מבחן	כמפורט בלוח הבחינות	

לקבלת ציון עובר בקורס יש לקבל ציון עובר בכל אחד ממרכיבי הציון הסופי. הערות לגבי ציונים:

יתכן בונוס אישי של עד 5% מהציון על השתתפות ותרומה משמעותית לדיון בהרצאות.

* עפ"י תקנון האוניברסיטה תלמיד חייב להיות נוכח בכל השיעורים (סעיף 5).

* מועד הבחינה יפורסם באתר הפקולטה - לוח בחינות.

* תלמיד, הנעדר משיעור המחייב השתתפות פעילה או שלא השתתף באורח פעיל, רשאי המורה להודיע למזכירות כי יש למחוק את שמו מרשימת המשתתפים. (התלמיד יחויב בתשלום בגין קורס זה)

פירוט המטלות בקורס

בקורס יינתנו כארבעה עד חמישה תרגילי בית אותם יש להגיש במהלך הסמסטר לפי המפורט בהמשך.

כל אי עמידה במי ממטלות הקורס מחיבת הודעה מראש (במייל) למרצה

מדיניות שמירה על טווח ציונים

החל משנה"ל תשס"ט מונהגת בפקולטה מדיניות שמירה על טווח ציונים בקורסי התואר השני.

עקרונות השיטה חלים על כל קורסי התואר השני, ומדיניות השמירה על טווח הציונים תיושם לגבי הציון הסופי בקורס זה. מידע נוסף בנושא זה מתפרסם בהרחבה באתר הפקולטה.

[לתקנוני מדיניות שמירת טווח ציונים](#)

הערכת הקורס ע"י הסטודנטים

בסיומו של הקורס הסטודנטים ישתתפו בסקר הוראה על מנת להסיק מסקנות לטובת צרכי הסטודנטים והאוניברסיטה.

אתר הקורס

אתר הקורס יהווה המקום המרכזי בו ימסרו הודעות לסטודנטים, לפיכך מומלץ להתעדכן בו מדי שבוע, לפני השיעור, ובכלל – גם בתום הסמסטר. (לצורך תיאום ענייני הבחינה למשל).

שקפי הקורס יהיו באתר הקורס באתר.

לתשומת לבכם - בכיתה ידונו גם נושאים (ובפרט דוגמאות) שאינם מופיעים בשקפים או מופיעים בכותרת בלבד. כל אלו הינם חלק בלתי נפרד מחומר הקורס.

תכנית הקורס *

הערות	קריאת חובה	נושאים	תאריך	שבוע
	- Data Science for business chapters 1,2. - Chapman et al. 2000 Learning Python chapters 1-3	Data Processing as a Part of the Business Data Analytics Process Introduction to Data Science. The benefits of using Big Data for business insights. The data science process. Working methodologies in Data Science, CRISP-DM cycle. Python – most commonly used recent technology tool for writing data related applications: Introduction, writing a first program and script. Programming entities: variables, data types, dynamic typing, casting, expressions.	11.12.22	1
	Learning Python chapters 4, 6, 12-14	Python. Flow Control: conditional statements, iterations over data, loops.	18.12.22	2
	Learning Python chapters 9, 16-18	Python. Working methodologies in writing code. Memory model. Functions.	01.01.23	3
	Learning Python chapters 4, 7	Python. Advanced data types and structures: tuples, lists, dictionaries. Strings fundamentals for text processing. Sorting.	08.01.23	4
	Learning Python chapter 7	Processing Textual-based Data Sources Using structured and unstructured textual sources for obtaining business data. Regular expressions as a tool for retrieval data from textual sources. Dictionaries and dictionary-based sentiment analysis example.	15.01.23	5
		Retrieving Big Data from External Sources Basic web scraper – for scraping online product information, reviews, and competitive information.	22.01.23	6

	Database Systems chapters 2, 6	Managing and Storing Data Introduction to Relational Databases and Structured Query Language (via Python): Tables, Keys, Queries, Joins, Aggregation.	29.01.23	7
--	-----------------------------------	--	----------	---

*התכנית הינה בסיס לשינויים.

הערות:

חלק משמעותי בקורס הינו **תרגול ולימוד עצמי**. משך לימוד חלק מהנושאים אינו בהכרח הרצאה אחת. יתכן שחלק מההרצאות יוחלפו בהרצאות אורח או בנושאים עדכניים.

קריאת רשות

Chapman, Pete, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rudiger Wirth. "CRISP-DM 1.0 Step-by-step data mining guide." (2000). <http://the-modeling-agency.com/crisp-dm.pdf>

Garcia-Molina, H., Ullman J. D., Widom J., "Database Systems: The Complete Book", Pearson Prentice Hall, 2009, 2002.

Lutz, M., "Learning Python", 5th edition, O'Reilly Publisher.

Provost, F., Fawcett, T., "Data Science for Business", O'Reilly Publisher, 2013