

הנחיות לפרויקט הסיום – סטטיסטיקה למדע הנתונים (90% מהציון הסופי)

מטרה כללית

הפרויקט נועד לבדוק את היכולת שלכם ליישם באופן עצמאי ומקצועי את רוב הכלים והחשיבה הסטטיסטית שנלמדו בקורס על מערך נתונים אמיתי – מהשאלה ועד למסקנות בעלות משמעות.

נושא ונתונים

- כל סטודנט מקבל נושא מוגדר מראש + קובץ נתונים אחד
- אין אפשרות לבחור נושא אחר או להחליף / להוסיף קובץ נתונים

מה נדרש להגיש

- מסמך / קובץ / מצגת / דוח – בכל פורמט שנראה לכם מתאים
- המטרה היא להראות הבנה עמוקה, חשיבה סטטיסטית נכונה, קבלת החלטות מושכלות ויכולת להעביר מסר ברור
- אין מבנה חובה – אתם מחליטים איך לארגן את החומר (רצף לוגי, סיפור, דוח קלאסי, מצגת מורחבת, נוטבוק עם הסברים וכו')

אלמנטים שכדאי שיופיעו (בצורה כלשהי)

- הצגת השאלה / המטרה העסקית-מחקרית
- תיאור ראשוני של הנתונים (גודל, סוגי משתנים, בעיות בולטות)
- מה עשיתם עם הנתונים לפני הניתוח (חסרים? חריגים? טרנספורמציות? משתנים חדשים?)
- אילו שיטות ניתוח בחרתם – ולמה דווקא אותן
- תוצאות עיקריות (טבלאות / גרפים / מדדים)
- אבחון ואימות (שאריות? הנחות? השוואת מודלים? over-fitting?)
- מסקנות ישירות לשאלה + פרשנות בעלת משמעות
- מגבלות + מה לא הצלחתם לעשות / מה היה יכול לשפר את העבודה

דרישות טכניות מינימליות

- קוד נקי ומתועד (Python או R או שילוב) – חייב להיות ניתן להריץ אותו
- גרפים קריאים עם כותרות ותוויות
- טבלאות מסודרות (לא צילומי מסך מבולגנים)

קריטריוני הערכה עיקריים

- טיפול מעמיק ומנומק בנתונים + EDA איכותי
- בחירה, יישום ואבחון מודלים
- השוואה בין מודלים והסקת מסקנות מבוססות
- פרשנות עסקית/מדעית + מודעות למגבלות
- איכות הכתיבה, הויזואליזציה והארגון

הפרויקט הזה הוא המרכיב המרכזי של הציון בקורס.
הוא נועד לדמות עבודה אמיתית של מדען/ת נתונים – מהשאלה ועד ההמלצה המעשית.

בהצלחה רבה!