
מבוא לניתוח נתונים 4212410

סמסטר קיץ תשפ"ב

מועד א' 20.10.2022

ד"ר ליהי דראי

משך הבחינה: 2 שעות

המבחן בכיתת מחשב ללא אינטרנט

חומר עזר מותר: כל חומר מודפס. חומר דיגיטלי על דיסק-און-קי

חומר עזר אסור: מכשירים דיגיטליים (לפטופ, טאבלט, שעון, טלפון, מחשב)

אין להשתמש או להגיש מחברות

הנחיות לסטודנטים:

- במודל תמצאו: קובץ פייתון (test_a.ipynb) וקובץ דאטה (data_file.csv)
- העתיקו את שני הקבצים למחשב, ושימו אותם באותה התיקייה.
- כתבו את התשובות בתוך קובץ הפייתון (test_a.ipynb).
- יש להגיש את הקובץ test_a.ipynb בלבד. טופס הבחינה ומחברת הבחינה לא יבדקו.
- הגשת המבחן – דרך מתזמן הבחינות בלבד.
- אין להשתמש בלולאות for או בספריות שלא נלמדו בקורס.
- ניתן להשתמש בכל הפקודות מהספריות: pandas, numpy, matplotlib, seaborn, sklearn

טור א

בהצלחה!!!

עליכם לנתח את קובץ הנתונים data_file.csv. הקובץ נמצא במתזמן המבחנים. יש להעתיק אותו למחשב ולקרוא אותו משם. הקובץ מכיל נתוני דירוגים של מסעדות.

Area code	האזור בו נמצאת המסעדה
Location	מיקום המסעדה
Gender	מגדר (של האורח)
YOB	שנת לידה (של האורח)
Marital Status	מצב משפחתי (של האורח)
Activity	תעסוקה (של האורח)
Budget	דירוג יוקרה (של המסעדה)
Cuisines	סוג אוכל
Alcohol	האם מגישים אלכוהול
Smoker	האם האורח מעשן
Food Rating	דירוג האוכל שנתן האורח
Service Rating	דירוג השירות שנתן האורח
Overall Rating	דירוג כללי שנתן האורח
Often A S	האם האורח מבקר קבוע במסעדה זו

שאלה 1 (25%)

- א. צרו טבלה המראה עבור כל סוג אוכל (Cuisines) בכל דירוג יוקרה (Budget) את הדירוג הכללי (Overall Rating) הממוצע, הנמוך ביותר, הגבוה ביותר וכן את מספר הדירוגים שניתנו.
- ב. מהו סוג האוכל (Cuisines) עם הממוצע הכי גבוה בדירוג יוקרה (Budget) מספר 4?

שאלה 2 (25%)

הציגו שני איורים אחד ליד השני:

- א. השוואה בין הדירוג (Overall Rating) החציוני של סוגי אוכל (Cuisines), עם הבחנה בין גברים ונשים. יש להציג את כל סוגי האוכל (Cuisines) פרט לאוכל צרפתי (French).
- ב. התפלגות הדירוג (Overall Rating) באחוזים של גברים לעומת נשים, עבור כל הדאטה פרט לאוכל צרפתי.

שאלה 3 (25%)

- א. איזו סוג מסעדה (Cuisines) קיבלה את הדירוג הממוצע (Overall Rating) הגבוה ביותר מגברים שתחום עיסוקם (Activity) הוא להיות סטודנטים והם מעשנים לעיתים קרובות (Often)?
- ב. הציגו באיור את הדאטה מסעיף א.

שאלה 4 (25%)

בחרו 6 פיצ'רים (עמודות נתונים). צרו שתי מפות חום מהפיצ'רים שבחרתם והציגו אותן אחת ליד השני. דאגו שהמפות יכילו גם את ערכי הקורלציה במספרים ולא רק צבעים.

הערה: זה לא משנה באיזה סוג קורלציה אתם בוחרים להשתמש.

- א. במפה אחת השתמשו בכל הנתונים
- ב. במפה השנייה השתמשו רק בנתונים של האזור (Area code) שקיבל את כמות הדירוגים הגדולה ביותר (אם הצגתם את Area code בסעיף א, אז עכשיו יהיו לכם רק 5 פיצ'רים, זה בסדר).
- ג. עבור הנתונים מסעיף ב, צרו איור המראה את הקשר בין שני הפיצ'רים עם הקורלציה השלישית הכי גדולה.

מבוא לניתוח נתונים 4212410

סמסטר קיץ תשפ"ב

מועד א' 20.10.2022

ד"ר ליהי דראי

משך הבחינה: 2 שעות

המבחן בכיתת מחשב ללא אינטרנט

חומר עזר מותר: כל חומר מודפס. חומר דיגיטלי על דיסק-און-קי

חומר עזר אסור: מכשירים דיגיטליים (לפטופ, טאבלט, שעון, טלפון, מחשב)

אין להשתמש או להגיש מחברות

הנחיות לסטודנטים:

- במודל תמצאו: קובץ פייתון (test_a.ipynb) וקובץ דאטה (data_file.csv)
- העתיקו את שני הקבצים למחשב, ושימו אותם באותה התיקייה.
- כתבו את התשובות בתוך קובץ הפייתון (test_a.ipynb).
- יש להגיש את הקובץ test_a.ipynb בלבד. טופס הבחינה ומחברת הבחינה לא יבדקו.
- הגשת המבחן – דרך מתזמן הבחינות בלבד.
- אין להשתמש בלולאות for או בספריות שלא נלמדו בקורס.
- ניתן להשתמש בכל הפקודות מהספריות: pandas, numpy, matplotlib, seaborn, sklearn

טור ב

בהצלחה!!!

עליכם לנתח את קובץ הנתונים data_file.csv. הקובץ נמצא במתזמן המבחנים. יש להעתיק אותו למחשב ולקרוא אותו משם. הקובץ מכיל נתוני דירוגים של מסעדות.

Area code	האזור בו נמצאת המסעדה
Location	מיקום המסעדה
Gender	מגדר (של האורח)
YOB	שנת לידה (של האורח)
Marital Status	מצב משפחתי (של האורח)
Activity	תעסוקה (של האורח)
Budget	דירוג יוקרה (של המסעדה)
Cuisines	סוג אוכל
Alcohol	האם מגישים אלכוהול
Smoker	האם האורח מעשן
Food Rating	דירוג האוכל שנתן האורח
Service Rating	דירוג השירות שנתן האורח
Overall Rating	דירוג כללי שנתן האורח
Often A S	האם האורח מבקר קבוע במסעדה זו

שאלה 1 (25%)

- א. צרו טבלה המראה עבור כל סוג אוכל (Cuisines) בכל דירוג יוקרה (Budget) את הדירוג הכללי (Overall Rating) החציוני, הנמוך ביותר, הגבוה ביותר וכן את מספר הדירוגים שניתנו.
- ב. מהו סוג האוכל (Cuisines) עם החציון הכי גבוה בדירוג יוקרה (Budget) מספר 4?

שאלה 2 (25%)

הציגו שני איורים אחד ליד השני:

- א. השוואה בין הדירוג הממוצע (Overall Rating) של סוגי אוכל (Cuisines), עם הבחנה בין סוגי מעשנים (Smoker). יש להציג את כל סוגי האוכל פרט לאוכל איטלקי (Italian).
- ב. התפלגות הדירוג (Overall Rating) באחוזים של אנשים שאינם מעשנים, מעשנים לעיתים קרובות ומעשנים חברתית (Never, Often, Socially), עבור כל הדאטה פרט לאוכל איטלקי.

שאלה 3 (25%)

- א. איזו סוג מסעדה (Cuisines) באזור (Area code) 123 קיבלה הכי הרבה ביקורות (Overall Rating) מנשים שתחום עיסוקן (Activity) הוא מקצועי (Professional)?
- ב. הציגו באיור את הדאטה מסעיף א.

שאלה 4 (25%)

בחרו 6 פיצ'רים (עמודות נתונים). צרו שתי מפות חום מהפיצ'רים שבחרתם והציגו אותן אחת ליד השני. דאגו שהמפות יכילו גם את ערכי הקורלציה במספרים ולא רק צבעים.

הערה: זה לא משנה באיזה סוג קורלציה אתם בוחרים להשתמש.

- א. במפה אחת השתמשו בכל הנתונים.
- ב. במפה השנייה השתמשו רק בנתונים של אנשים שהמצב המשפחתי שלהם (Marital Status) הוא רווקים (Single).
- ג. עבור הנתונים מסעיף ב, צרו איור המראה את הקשר בין שני הפיצ'רים עם הקורלציה השלישית הכי גדולה.