

תרגיל בית 1 - EXCEL

הנחיות להגשה

1. את התרגיל יש להגיש בזוגות עד לתאריך 15/12/2019 בשעה 23:55 לתיבת ההגשה במודל.
2. יש להגיש תיקייה מקובצת כקובץ **zip** (ולא **rar**. כל עוד צורה אחרת) אחד, עם השם "EX1_ID1_ID2" בלי המרכאות וכך שבמקום ID1\ID2 יהיו מספרי ת.ז. של המגישים. בתוך התיקייה יהיו הקבצים הבאים:
 - a. קובץ אקסל עם השם "excel_ID1_ID2.xlsx" (שימו לב לסימנת **xlsx**),
 - b. קובץ וורד עם השם "EX1_ID1_ID2", בלי המרכאות וכך שבמקום ID1\ID2 יהיו ת.ז. של המגישים.
3. מספרי ת"ז של כל אחד יופיעו בגיליון בשם מגישים בתאים A1 ו-A2 בצורה קריאה וללא גלישה לתאים נוספים.
4. הגשה שלא בזוגות או שלא לפי הפורמט שהוגדר תגרור הורדה של 30 נקודות.
5. ניתן להגיש את העבודה **פעם אחת בלבד**, הגשה של יותר מהעתק אחד (הגשה כפולה או הגשה של דף תיקון) תגרור הורדה של 30 נקודות.
6. **לפני ההגשה בדקו שעברתם על הפורום ב-moodle** ושאתם מעודכנים בכל השינויים, התשובות, העדכונים וההבהרות שניתנו שם.
7. הקובץ צריך להיות ברור ונוח לשימוש, אסתטי וקל לקריאה, לדוגמא, לא יופיעו תאים עם סימן "#####" יש לעצב את התאים באופן המתאים לתוכן.
8. **שימו לב** שכאשר אתם מציירים את הגרפים ערכי ציר X חיוביים וגם ערכי y חיוביים, והגרף מצויר ברביע הראשון. יש להוסיף לכל הגרפים כותרות (צירים, כותרת ראשית, מקרא).
9. בדיקת התרגיל תהיה מדגמית – כלומר תבחרנה מספר שאלות מתוך העבודה ורק הן תבדקנה.

הנחיות כלליות

1. את התרגיל אתם יכולים לפתור במעבדות או במחשבים האישיים שלכם, אך הבדיקה תעשה לפי המחשבים במעבדות בגרסת excel 2017.
2. את כל התשובות יש לכתוב בקובץ וורד **בנוסף** לקובץ ה-excel. את התשובות של הסעיפים החשובים יש לכתוב בתאים במיועדים לכך בקובץ ה-excel (מסומן ע"י תא צהוב).
3. בכל פעם שאתם יוצרים גרף, יש לעשות זאת בלשונית המתאימה שהוכנה מראש. בנוסף יש לתת לגרף כותרת ושמות צירים בהתאם, **גם אם לא כתוב במפורש בסעיף**. יש להעתיק את הגרפים לקובץ הוורד.
4. במידה והשתמשתם בפונקציה כלשהי, יש להעתיק אותה לקובץ הוורד בסעיף המתאים.
5. עליכם לוודא שבקובץ ה word אתם מתייחסים לכל השאלות וכל הסעיפים בתרגיל (11-2).
6. **גוגל הוא החבר הכי טוב שלכם!!!** בין אם אתם לא זוכרים מה נעשה בכיתה, ובין אם יש סעיף שלא עברנו עליו במפורש, יש לבצע חיפוש באינטרנט ורק אם לא נמצאה התשובה לשאול בפורום.

תקציר

בתרגיל זה תעבדו על נתונים של חברת רכבות, אשר מציעה אפשרויות של נסיעה ברכבות לילה בין מספר תחנות. עליכם לעזור לחברה לנהל את המידע אשר היא צריכה לשמור - פרטים של תחנות מהם/אליהם היא נוסעת, פרטי נוסעים, סוגי נסיעות וכו'.

צאו מנקודות הנחה כי כל נוסע יכול להזמין רק נסיעה אחת.

קובץ הנתונים מצורף לעבודה ואתם נדרשים לעבוד עליו בלבד - Train_excel.xlsx

קובץ הנתונים:

להלן התכונות (עמודות) אשר נמצאות בקובץ הנתונים:

| שם המשתנה | הגדרה | סוג המשתנה | טווח ערכים | הערות |
|----------------------------|-------------------------------------|---------------|------------|--|
| PassengerID | מספר סידורי של כל נוסע | רציף | ? | |
| Member | האם הנוסע חבר במועדון לקוחות ? | בינארי | ? | |
| Passenger class | המחלקה של הנוסע | קטגוריאלי | ? | |
| Name | שם הנוסע | Text (string) | --- | |
| Sex | מין הנוסע | בינארי | ? | |
| Age | גיל הנוסע | רציף | ? | אם הגיל רק הוערך, אז הוא מהצורה xx.5 |
| Number of Siblings | מספר האחים/יות או בעל/אישה על הרכבת | בדיד | ? | יכול להיות גם אחים/יות חורגים. התעלמו ממאהבות וארוסות. |
| Number of Parents\Children | מספר ההורים/ילדים על הרכבת | בדיד | ? | ילדים שנסעו רק עם שמרטפית קיבלו ערך 0. |
| Code | מספר הכרטיס | Text (string) | ---- | שילוב של מספרים ואותיות |
| Price | מחיר הכרטיס | רציף | ? | |
| Departure | תחנת המוצא | קטגוריאלי | ? | |
| Baggage | משקל המזוודות של הנוסע | רציף | ? | ק"ג |

תיאור התרגיל

1. טענו את קובץ הנתונים (Train_excel.xlsx) לתוך EXCEL.
2. העתיקו את הטבלה מהעמוד הקודם לקובץ התשובות, והשלימו את העמודה 'טווח ערכים' באופן הבא:
 - a. בתא שמצויין בו '---', אין צורך להשלים.
 - b. עבור משתנים רציפים ובדידים כתבו את הערך המינימאלי והערך המקסימלי. לדוגמא- '5-65'.
 - c. עבור משתנים קטגוריאליים כתבו את הערכים של כל הרמות שהמשתנה יכול לקבל, לדוגמא- 'Young, old'.
3. עבור המשתנים age, price, baggage חשבו את הממוצע, החציון והשכיח (יש מקום מיוחד בתחתית הטבלה). במידה ויש הבדלים בין המדדים, הסבירו ממה הם יכולים לנבוע.
4. עבור המשתנים Member, Passenger class, Sex חשבו את השכיח (Mode). שימו לב כי מדובר במשתנה מסוג text, חשבו כיצד לפתור את הבעיה.
5. התבוננו במשתנים מחיר הכרטיס, משקל המזוודות וגיל, ועברו ללשונית "age graphs":
 - a. צרו שני גרפי פיזור: אחד עבור השינוי במחיר הכרטיס כפונקציה של גיל והשני עבור השינוי במשקל המזוודות כפונקציה של גיל. את מי משני מהמשתנים ניתן לדעתכם לנבא מדויק יותר לפי הגיל? הסבירו מדוע.
 - b. בצעו חישוב רגרסיה לינארית לניבוי המשתנה מחיר הכרטיס כפונקציה של משתנה הגיל וכן לניבוי המשתנה משקל המזוודות כפונקציה של משתנה הגיל. כתבו בקובץ הוורד את הנוסחא (משוואה) אליה הגעתם, ואת ערך ה-R בריבוע. את מי משני מהמשתנים ניתן לדעתכם לנבא מדויק יותר לפי הגיל? הסבירו מדוע.
6. צרו גרף מסוג Pie של המשתנה Passenger Class בלשונית "PClass Pie chart" (היעזרו בפונקציה 'countif'). שימו לב, כי עליכם לעשות זאת על נתוני הטקסט הגולמיים (רמז- הוסיפו עמודה שמכילה את הערכים האפשריים). בנוסף, עליכם להציג על הגרף את האחוז שכל רמה מהווה, ולא את מספר המאורעות. האם התוצאות מתיישבות עם ההיגיון? בצעו זאת שוב עבור המשתנה תחנת המוצא (Departure).
7. מהי הקורלציה ע"פ Pearson בין (יש מקום מיוחד מימין בלשונית המקורית):
 - a. מחיר הכרטיס לבין מספר המחלקה של הנוסע? הסבירו את התוצאה.
 - b. האם הנוסע חבר מועדון לבין מספר ההורים/ ילדים שיש לו? הסבירו את התוצאה.
8. בצעו צביעה אוטומטית כך:
 - a. משתנה הגיל בצורה הדרגתית מהנמוך לגבוה.
 - b. משתנה מספר האחים/יות כך שמי שיש לו 0 התא ייצבע בכחול, בין 1-3 בירוק, ומי שמעל בצהוב.
9. חשבו כמה (יש מקום מיוחד מימין לטבלה המקורית):
 - a. גברים יש במחלקה העליונה.
 - b. אנשים שעלו על הרכבת ב-Southampton הינם חברי מועדון.
10. בלשונית 'Price- PERCENTRANK' בצעו את הפעולות הבאות:
 - a. הוסיפו עמודה אשר מחשבת את דירוג מחיר הכרטיס לרשומה ביחס לשאר הרשומות (השתמשו בפונקציה 'PERCENTRANK'). הציגו 2 ספרות אחרי הנקודה, אין צורך להעתיק את התוצאה לקובץ הוורד. מה הפונקציה מחשבת? כיצד היא מתחברת לתוצאות שקיבלתם לסעיף 3?
 - b. הוסיפו כעת עוד עמודה, ובה סווגו כל כרטיס ל"cheap" אם נמצא בשליש התחתון (קטן או שווה 0.33), "moderate" (בין 0.33 ל-0.67) אם נמצא בשליש האמצעי ו-"expensive" אם נמצא בשליש העליון (גדול או שווה ל-0.67). לאיזה עמודה אחרת בנתונים המקוריים הגיוני להשוות את התוצאה שקיבלתם ומדוע? בצעו השוואה זו, והביאו את יחס הערכים התואמים מתוך כלל הרשומות כנתון מספרי שייתמוך במסקנה שלכם (רמז- הוסיפו עמודה שממירה את העמודה שהוספתם לצורה בה תוכלו להשוות אוטומטית).
11. חשבו באמצעות טבלת ציר בלשונית "Pivot table":
 - a. עבור גבר ואישה, את מחיר הכרטיס הממוצע, בכל מחלקה.
 - b. עבור כל תחנת מוצא, כמה נוסעים הינם חברי מועדון בכל מחלקה.
 - c. עבור גבר ואישה, את סטיית התקן (standard deviation) של הגיל, בכל תחנת מוצא.
 - d. עבור כל מחלקה, מהו מחיר הכרטיס המינימאלי והמקסימאלי.