

תרגיל בית 1 - EXCEL

הנחיות להגשה

1. את התרגיל יש להגיש בזוגות 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 "Answers_ID1_ID2",בלי המרכאות וכך שבמקום ID1\ID2 יהיו ת.ז. של המגישים.
3. מספרי ת"ז של כל אחד יופיעו בגיליון בשם מגישים בתא A1 ו A2 בצורה קריאה וללא גלישה לתאים נוספים.
4. הגשה שלא בזוגות או שלא לפי הפורמט שהוגדר תגרור הורדה של 30 מקודות.
5. ניתן להגיש את העבודה **פעם אחת בלבד**, הגשה של יותר מהעתק אחד (הגשה כפולה או הגשה של דף תיקון) תגרור הורדה של 30 מקודות.
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 חייביים וגם ערכי y חייביים, 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.
2. את כל התשובות יש לכתוב בקובץ וורד **בנוסף** לקובץ ה-excel. את התשובות של הסעיפים החישוביים 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

תרגיל זה כולל נתונים שנאספו ממטוס נוסעים.

קובץ הנתונים המעודכן מצורף לעבודה, ואתם נדרשים לעבוד עליו בלבד - Excel_ID1_ID2.xlsx

תזכורת - קובץ הנתונים:

להלן חזרה על התכונות (עמודות) אשר נמצאות בקובץ הנתונים (שימו לב כי ישנם שינויים קלים):

שם המשתנה	הגדרה	סוג המשתנה	טווח ערכים	הערות
PassengerID	מספר סידורי של כל אדם.	רציף	?	
Member	האם הנוסע חבר במועדון לקוחות ?	בינארי	?	
Passenger class	המחלקה של הנוסע.	קטגוריאלי	?	
Name	שם הנוסע	Text (string)	---	
Gender	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 המטוס	בדיד	?	
Number of Parents\Children	מספר ההורים/ילדים על המטוס	בדיד	?	
Code	מספר הכרטיס	Text (string)	----	שילוב של מספרים ואותיות
Price	מחיר הכרטיס.	רציף	?	
Departure	נמל המוצא	קטגוריאלי	?	
Baggage	משקל המזוודות של הנוסע	רציף	?	ק"ג

תיאור התרגיל

1. טענו את קובץ הנתונים (Excel_ID1_ID2.xlsx) לתוך EXCEL.
2. (6 מקודות) 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 המקסימאלי. לדוגמא: '5-65'.
 - c. עבור משתנים קטגוריאליים כתבו את הערכים של כל הרמות שהמשתנה יכול לקבל, לדוגמא: 'Young, old'.
3. (6 מקודות) עבור המשתנים age, price, baggage חשבו את הממוצע, החציון והשכיח (יש מקום מיועד בתחתית הטבלה). במידה ויש הבדלים בין המדדים, הסבירו ממה הם יכולים לנבוע.
4. (6 מקודות) עבור המשתנים Member, Passenger class, Gender חשבו את השכיח (Mode). שימו לב כי מדובר במשתנה מסוג text, חשבו כיצד לפתור את הבעיה.
5. (10 מקודות) 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 ברבוע. את מי משני מהמשתנים ניתן לדעתכם לנבא מדויק יותר לפי הגיל? הסבירו מדוע.
6. (10 מקודות) 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 נמל המוצא (Departure).
7. (10 מקודות) מהי הקורלציה ע"פ Pearson בין (יש מקום מיועד 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. (10 מקודות) 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. (10 מקודות) חשבו כמה (יש מקום מיועד מימין לטבלה המקורית):
 - a. גברים יש במחלקה העליונה.
 - b. אנשים שעלו על המטוס ב-Southampton הינם חברי מועדון.
10. (16 מקודות) 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 המקודה, אין צורך להעתיק את התוצאה לקובץ הוורד. מה הפונקציה מחשבת? כיצד היא מתחברת לתוצאות שקיבלתם לסעיף 3?
 - b. הוסיפו כעת עוד עמודה, ובה סווג 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. (16 מקודות) חשבו באמצעות טבלת ציר בלשונית "Pivot table":
 - a. עבור גבר ואישה, את מחיר הכרטיס הממוצע, בכל מחלקה.
 - b. עבור כל נמל מוצא, כמה נוסעים הינם חברי מועדון בכל מחלקה.
 - c. עבור גבר ואישה, את סטיית התקן (standard deviation) של הגיל, בכל נמל מוצא.
 - d. עבור כל מחלקה, מהו מחיר הכרטיס המינימאלי והמקסימאלי.