**ניב קוטק – 208236315 קובץ פלט לתרגיל מספר 3**

|  |  |  |  |
| --- | --- | --- | --- |
| שאלה | מטלה | פתרון | |
| 2 | Age | **נומינלי** – כי יש להתייחס לכל המשתנים המיוצגים באמצעות מספרים/ טקסטואליים שמשקפים לצורך כוונה קטגוריאליים . (בעל אופי קטגוריאלי) בנוסף הוא משתנה Numerical עם מספר קטן של ערכים וכן הקשר שלו הוא לא לינארי עם המשתנה התלוי. | |
| Children | **נומינלי** – כי הוא משתנה Numerical ואין לו קשר לינארי עם המשתנה התלוי וגם יש לו מספר קטן של ערכים [0,5] | |
| Catalogs | **כמותי** – כי יש לו קשר לינארי | |
| 3 | AmountSpent | **זנב ימין** – לפי בחינה ויזואלית (בחינה מספיקה) וגם לפי מבחן NCV אשר דוחה את השערת האפס יש לנו שונות משתנה ולכן המשתנה התלוי אינה מתפלג נורמלי. בנוסף ניתן לראות לפי היסטוגרמה שיש הרבה בין 0 ל 25,000 של הוצאות בחנות וככל שהמספר עולה יש פחת . | |
|  | Salary | **זנב ימין** – לפי בחינה ויזואלית (בחינה מספיקה) וגם לפי מבחן NCV אשר דוחה את השערת האפס יש לנו שונות משתנה ולכן המשתנה התלוי אינה מתפלג נורמלי. בנוסף ניתן לראות לפי היסטוגרמה שיש לי הרבה בשכורות בין 0e לבין 1e+5 וככל שהמספר עולה יש פחת . | |
| 4 | טבלת הרגרסיה (תמונת מסך) |  | |
| מדדי ביצועים על training | MAPE **6.156297** | RMSE **1792.276** |
| מדדי ביצועים על validation | MAPE **6.491986** | RMSE **1564.478** |
| איזו תוצאה טובה יותר? | התוצאה הנמוכה יותר היא ב-training (6.156297) לעומת validation (6.491986). לכן, ה-MAPE הטוב יותר נמצא ב-**training**. | התוצאה הנמוכה יותר היא ב-validation (1564.478) לעומת training (1792.276). לכן, ה-RMSE הטוב יותר נמצא ב-**validation**. |
| משמעות המדדים | **מסתכל על השגיאה ביחס לגודל האמיתי שלה** **באחוזים** כלומר הוא מספק תמונה של השגיאות היחסיות ויכול להיות יותר שימושי כאשר רוצים לדעת באיזו מידה המודל נוטה לטעות באחוזים מהערכים האמיתיים, מה שמאפשר השוואה בין בעיות שונות עם סדרי גודל שונים. | **נותן עונש גדול יותר לטעות גדולה  ועונש עבור שגיאה קטנה זה לא ממש ישנה לה** כלומר הוא מספק תמונה של השגיאות האבסולוטיות  ונותן יותר משקל לשגיאות גדולות. |
| 5 | ChatGPT | **ChatGPT -4 Input:**  I have an R code for a multiple regression model. Write me code to calculate RMSE and MAPE to evaluate the model's performance.  **Output**:  # Assuming you have the actual and predicted values  # Replace these with your actual and predicted values  actual <- c(Your\_Actual\_Values) # Replace with your actual values  predicted <- c(Your\_Predicted\_Values) # Replace with your predicted values from the model  # Calculate RMSE  rmse <- sqrt(mean((actual - predicted)^2))  cat("RMSE:", rmse, "\n")  # Calculate MAPE  mape <- mean(abs((actual - predicted) / actual)) \* 100  cat("MAPE:", mape, "%", "\n")  **results**: # The results of both RMSE and MAPE were the same. # He used the Metrics library that was not taught in class | |