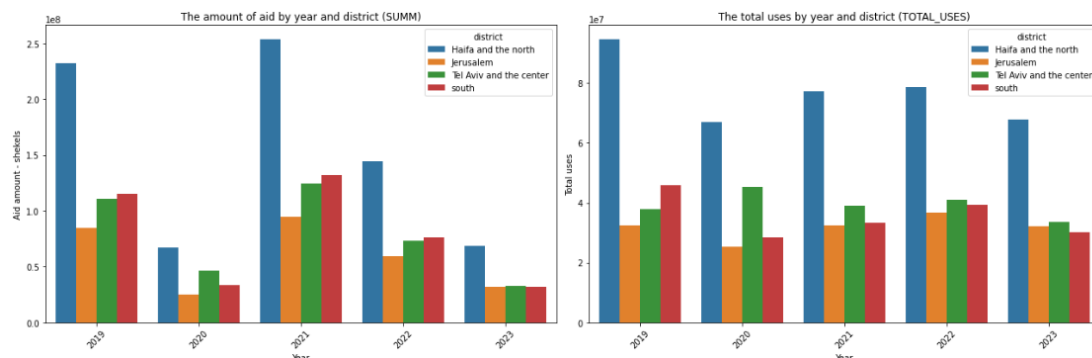
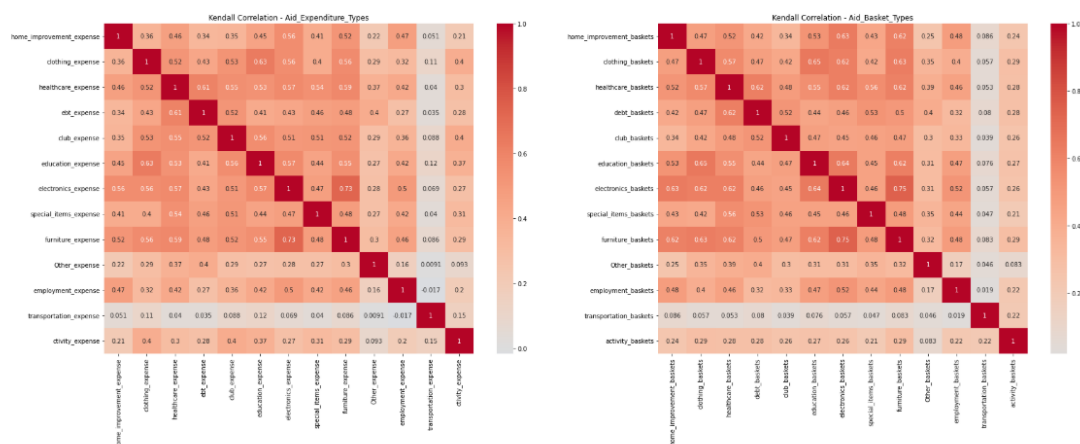


כפי שאפשר להסיק מן הנתונים החזותיים וכן מן הנתונים של הקוד, אין אפשרות להשתמש בכל השנים מהדאטה מכיוון שכל שנה מביאה נתונים שונים שמתייחסים לאותם ערים ומחוזות, ובסופו של דבר זה יוצר כפילויות שיכולות לשבש את המודל.



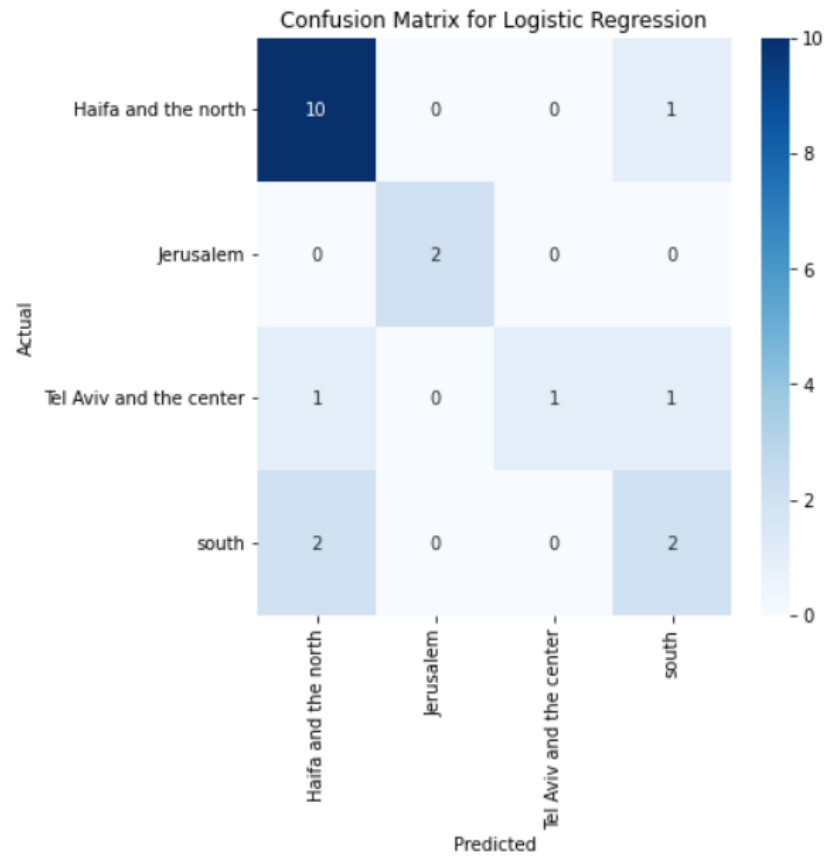
השוואת העמודות של סכום הכסף מול העמודות של הסלים.

כפי שאפשר לראות עומדו בפנינו האופציה או לקחת את העמודות החדשות שנצרו משני עמודות פר סוג סיוע, או לבחור בעמודה אחת - מספר סלים \ סך הכל סכום פר סיוע. מכיוון שגם מטריצת הקורלציה דומה וגם הממוצעים וסטיות התקן. נשתמש רק בעמודה אחת - מספר סלים.(יותר קל להתמודד איתם ולא עם העמודות של סך הכל בשקלים פר סל)



מטריצת בילבול – המודל המוצלח ביותר

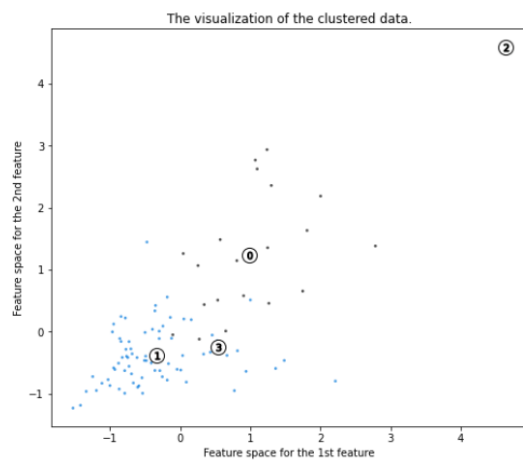
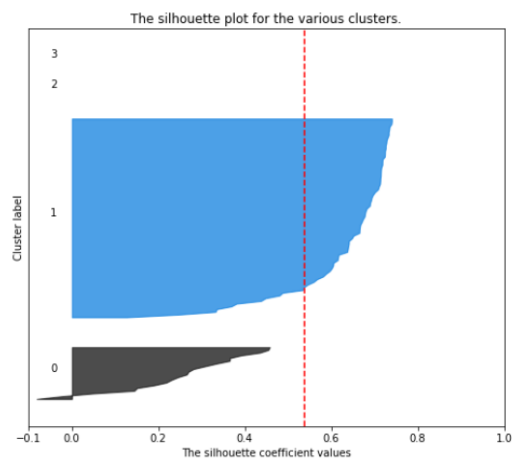
Model: Logistic Regression
Confusion Matrix:



Model: Random Forest
Confusion Matrix:

קלאסטינג – means-K המודל המוצלח ביותר

Silhouette analysis for KMeans clustering on data with n_clusters = 4



קלאסטינג – Agglomerative Clustering המודל המוצלח ביותר

Testing for 3 clusters:

Silhouette Score for 3 clusters: 0.5062058218928224

