**דו"ח תרגיל 4**

**מגישים:**

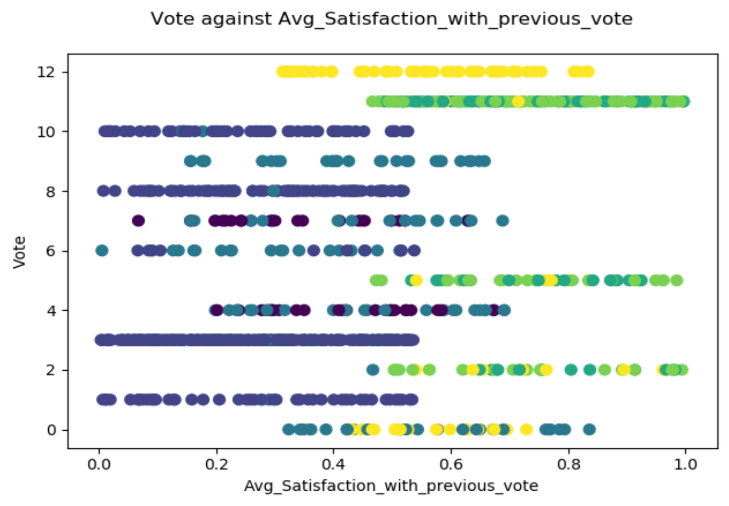
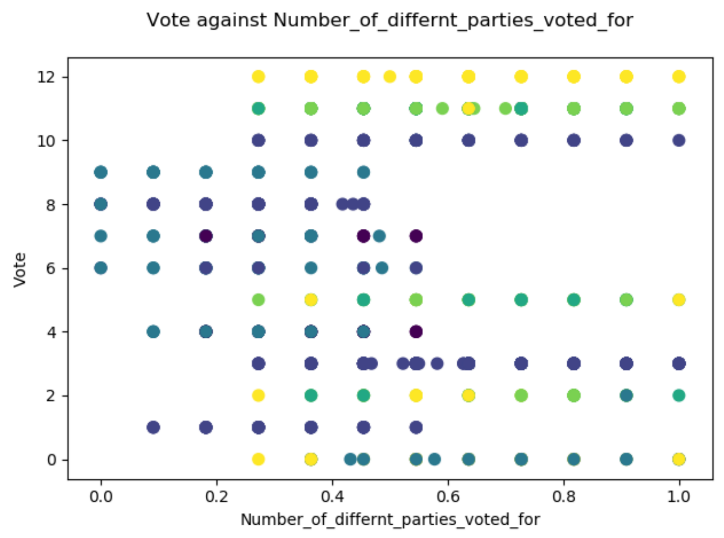
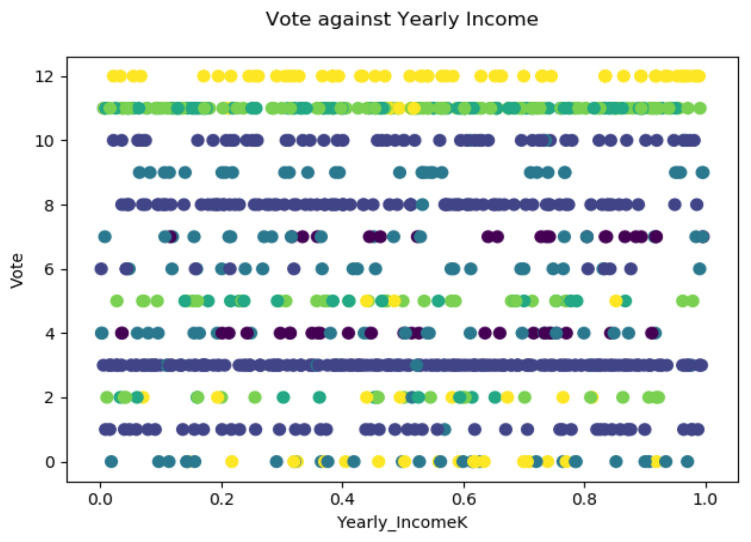
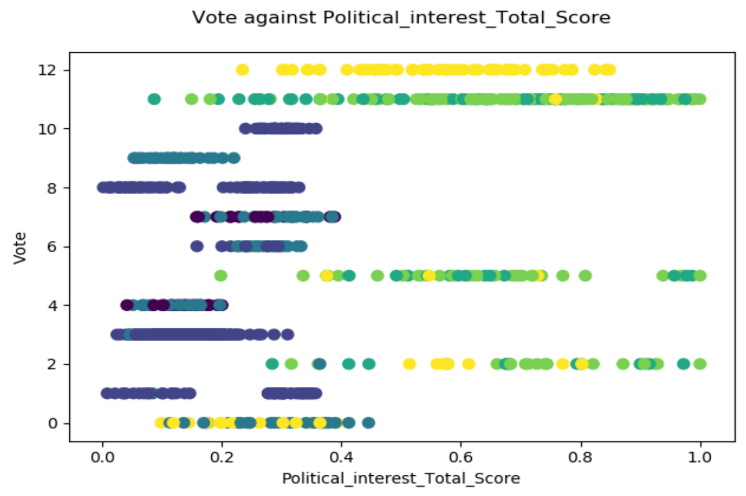
**עומר עצמון – 208669457**

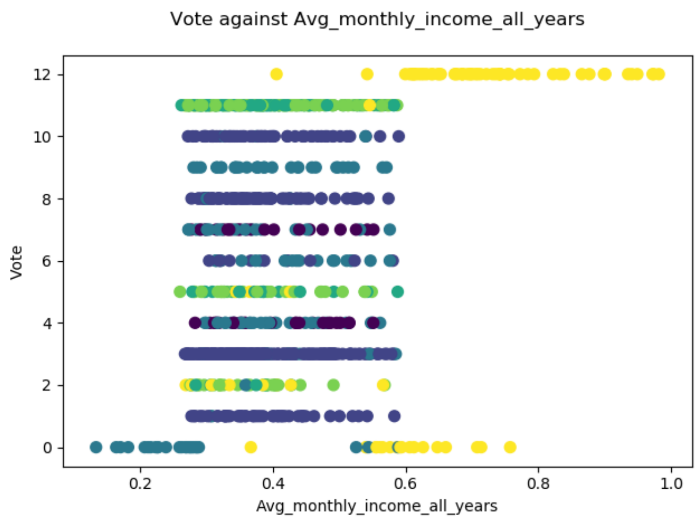
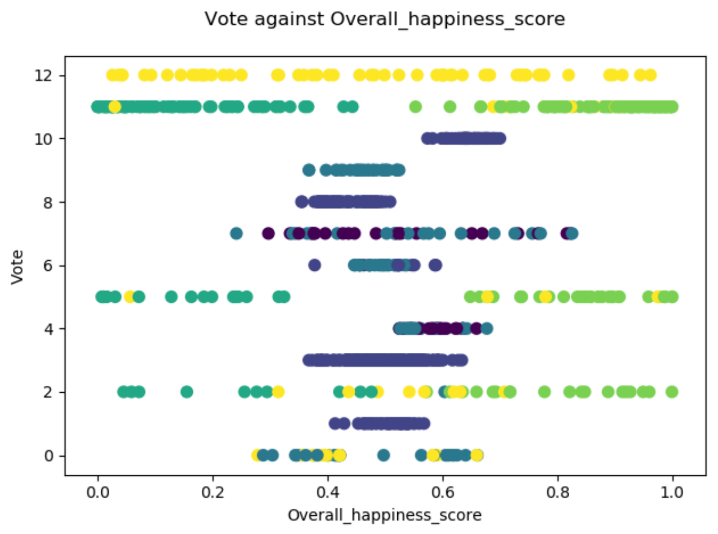
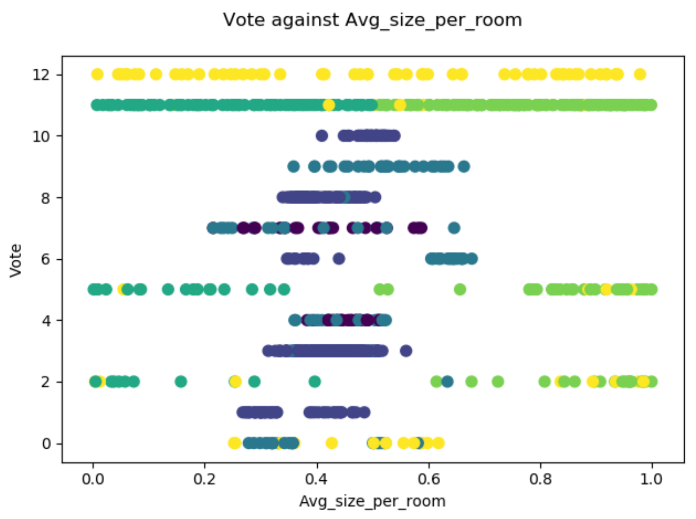
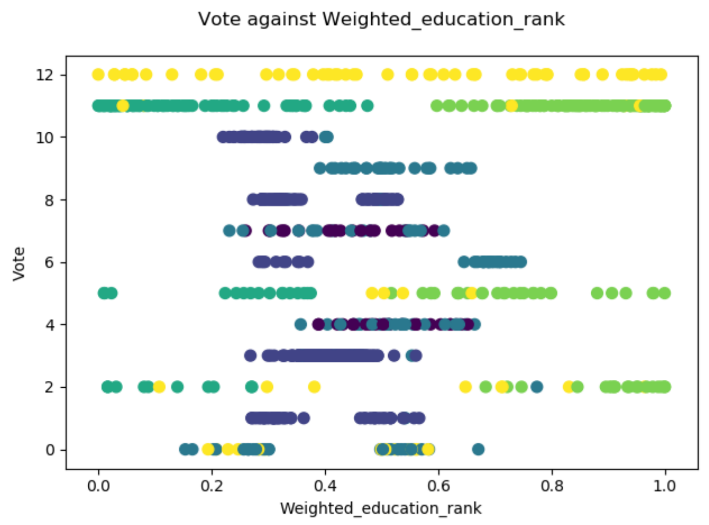
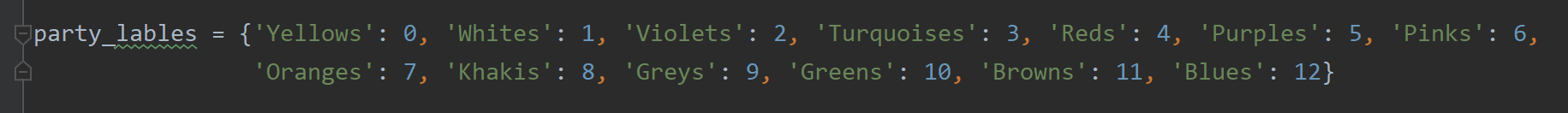
**ולאד קיל – 312866775**

**Clustering**

את הקלאסטרינג ביצענו על ידי מודל GMM, כיוון שיש לנו מספר רב של נקודות וניתן להניח כי אנשים מתפלגים באופן גאוסי על פני מספר רב. בדקנו את ההיפר-פרמטרים שלנו בעזרת cross validation על מנת לקבל את הפרמטרים הטובים ביותר וראינו כי המודל מחלק הכי טוב ל-6 קלאסטרים כאשר ה-type הינו Full, פונקציית ההערכה בה השתמשנו לבדיקת איכות הפרמטרים הינה BIC.

**ניתוח תוצאות**

לאחר האימון הפעלנו את המודל על ה-test set והוצאנו גרפים שמראים את ההתפלגות של ההצבעה כתלות בכל אחד מהפיצ'רים שלנו:

המפלגות ממוספרות כך:  


ניתן לראות ערבוב רב בין המפלגות השונות והקלאסטרים השונים. לכן על מנת להגיע לקואליציה התחלנו לאחד בין קלאסטרים שונים על מנת להגדיל את המרחק הממוצע בין הקלאסטרים השונים עד שנגיע לקלאסטרים שמופרדים כמה שיותר. את זה נעשה על ידי איטרציות כאשר בכל איטרציה נחבר את שני הקלאסטרים שחיבורם יעלה בצורה מקסימלית את ממוצע הניקוד לפי silhouette. ניקוד זה בודק את המרחק בין קלאסטרים והוא נע בין -1 ל- 1 כאשר ניקוד 1 מייצג שהנקודות רחוקות אחד מהקלאסטרים האחרים ולכן נשאף להביא את הניקוד כמה שיותר קרוב לאחד. תוצאה:



ניתן לראות כי על ידי חיבור הקלאסטר הכחול עם הקלאסטר הטורקיז נקבל את הקואליציה שלנו בעלת המפלגות: (לפי המעבר שהראנו קודם)  
whites, Turquoises, reds, pinks, oranges, Khakis, greys, greens