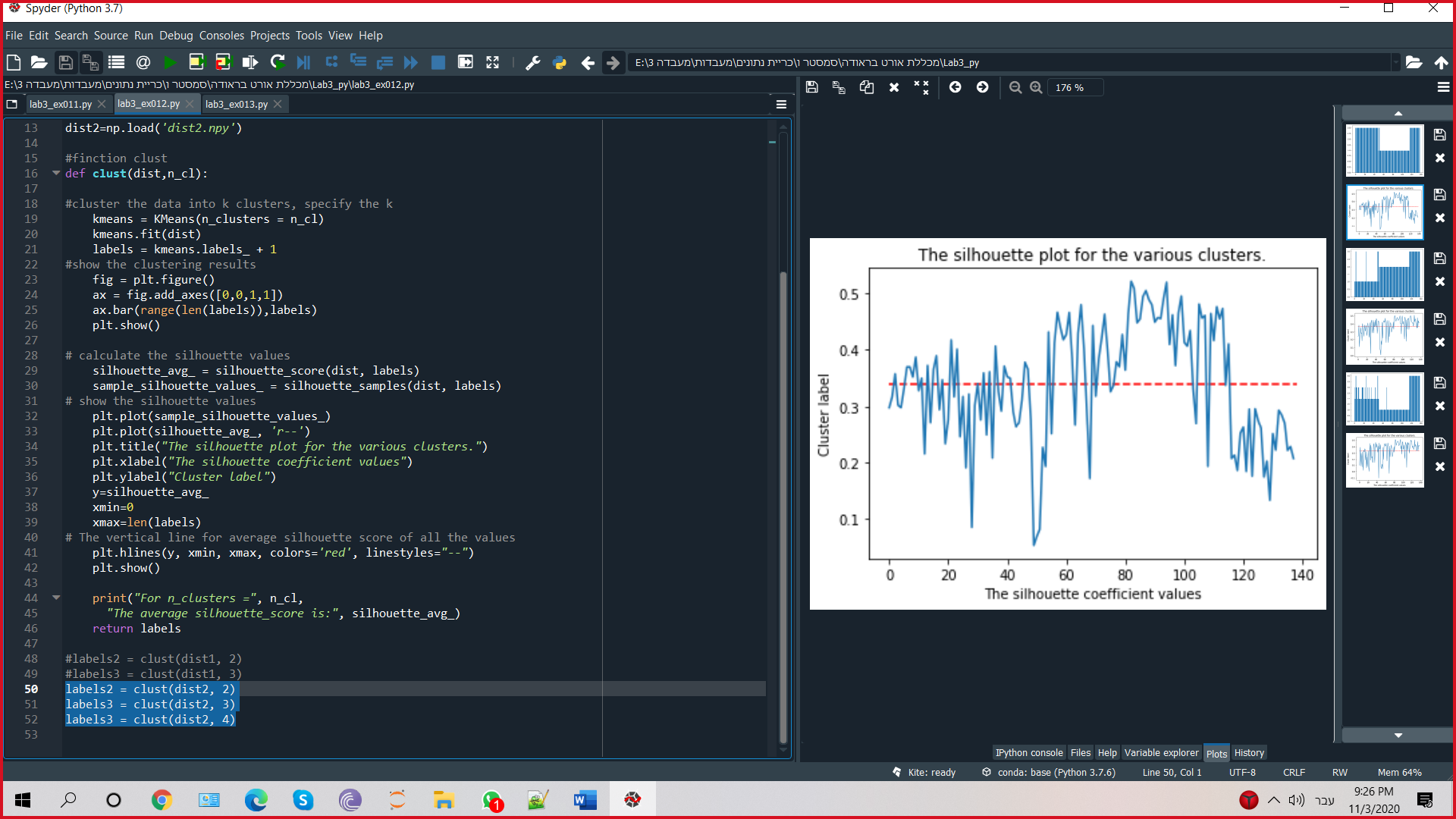
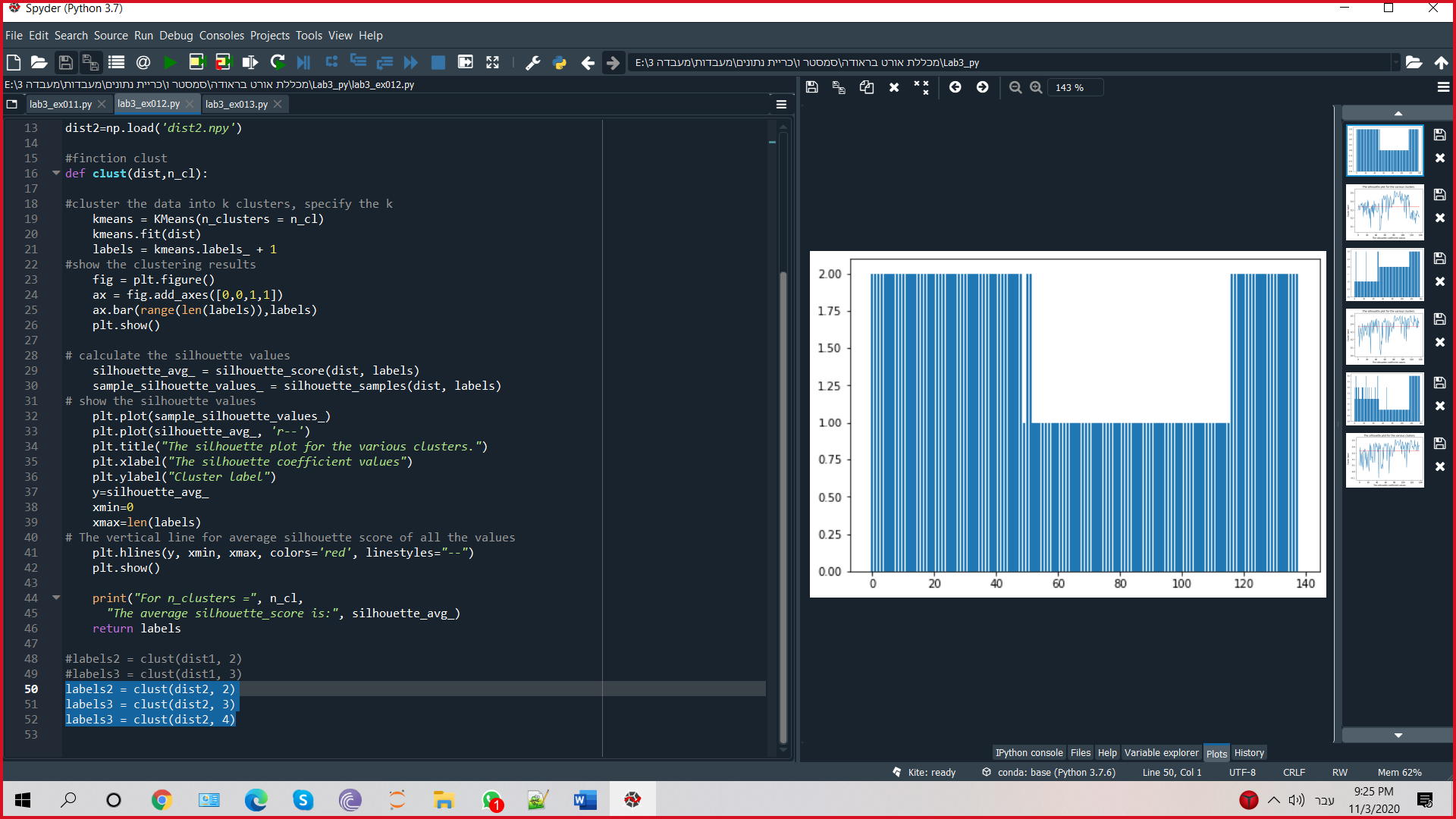
**מעבדה 3 – כריית נתונים**

שם מגיש: אילון בן סימון ת"ז: 312162951

שם מגיש: סער ויקטור ת"ז: 312392822

**תחילה בדקנו עבור wind =20000:**

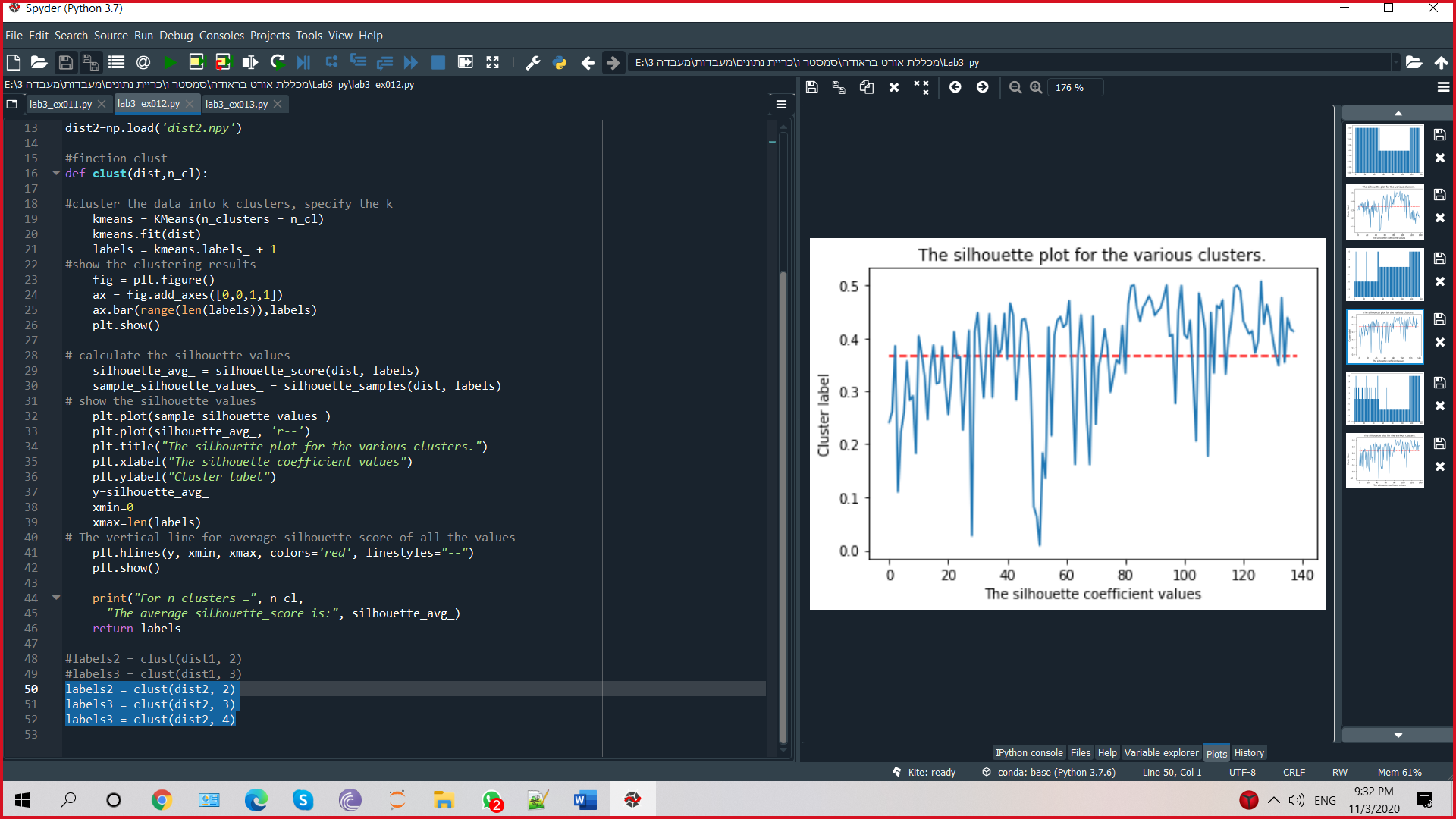
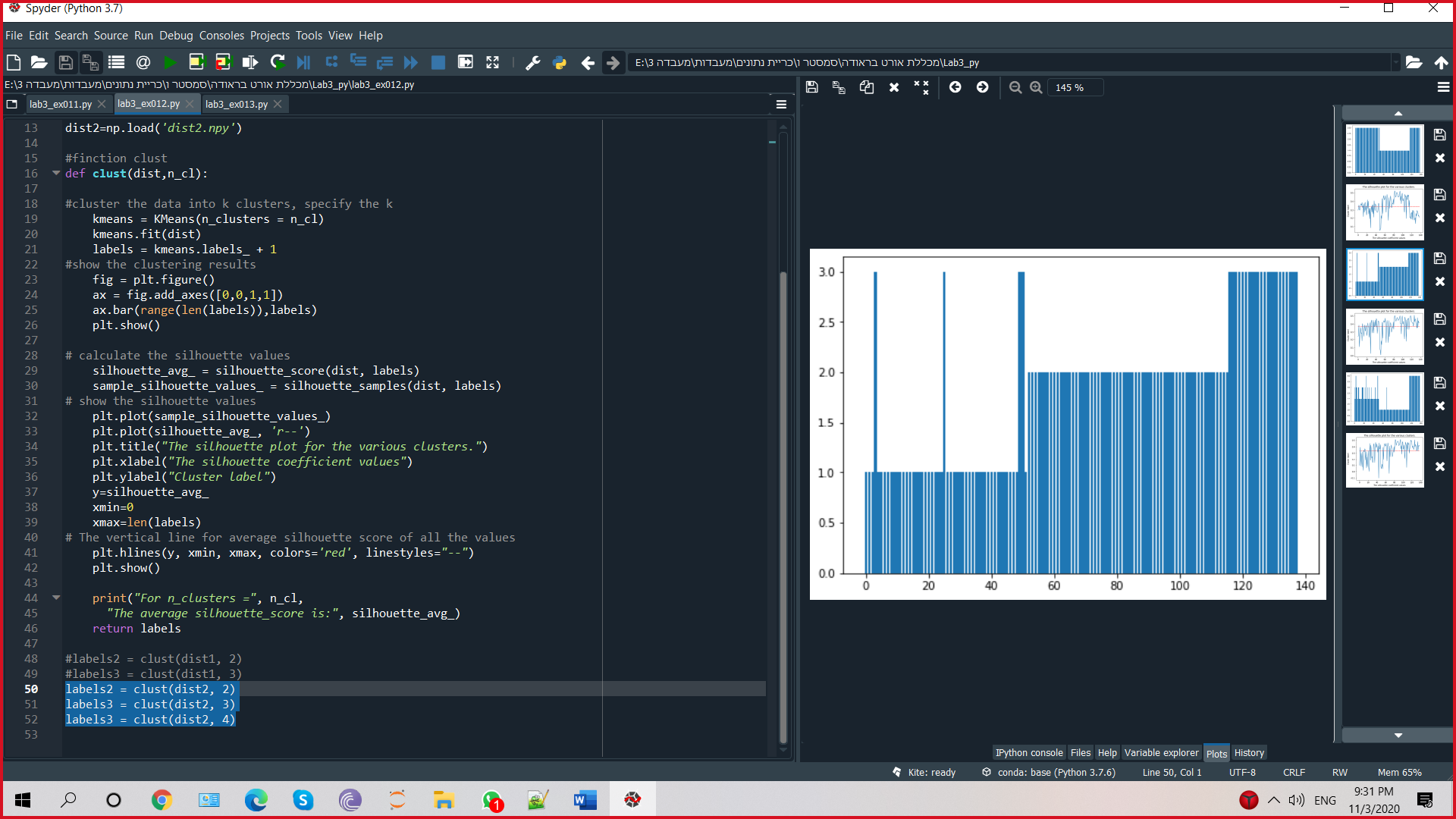
עבור חלוקה ל- cluster=2:



כאשר מחלקים ל-2 clusters ערך הסילואט הממוצע הינו:

.0.340305959162652

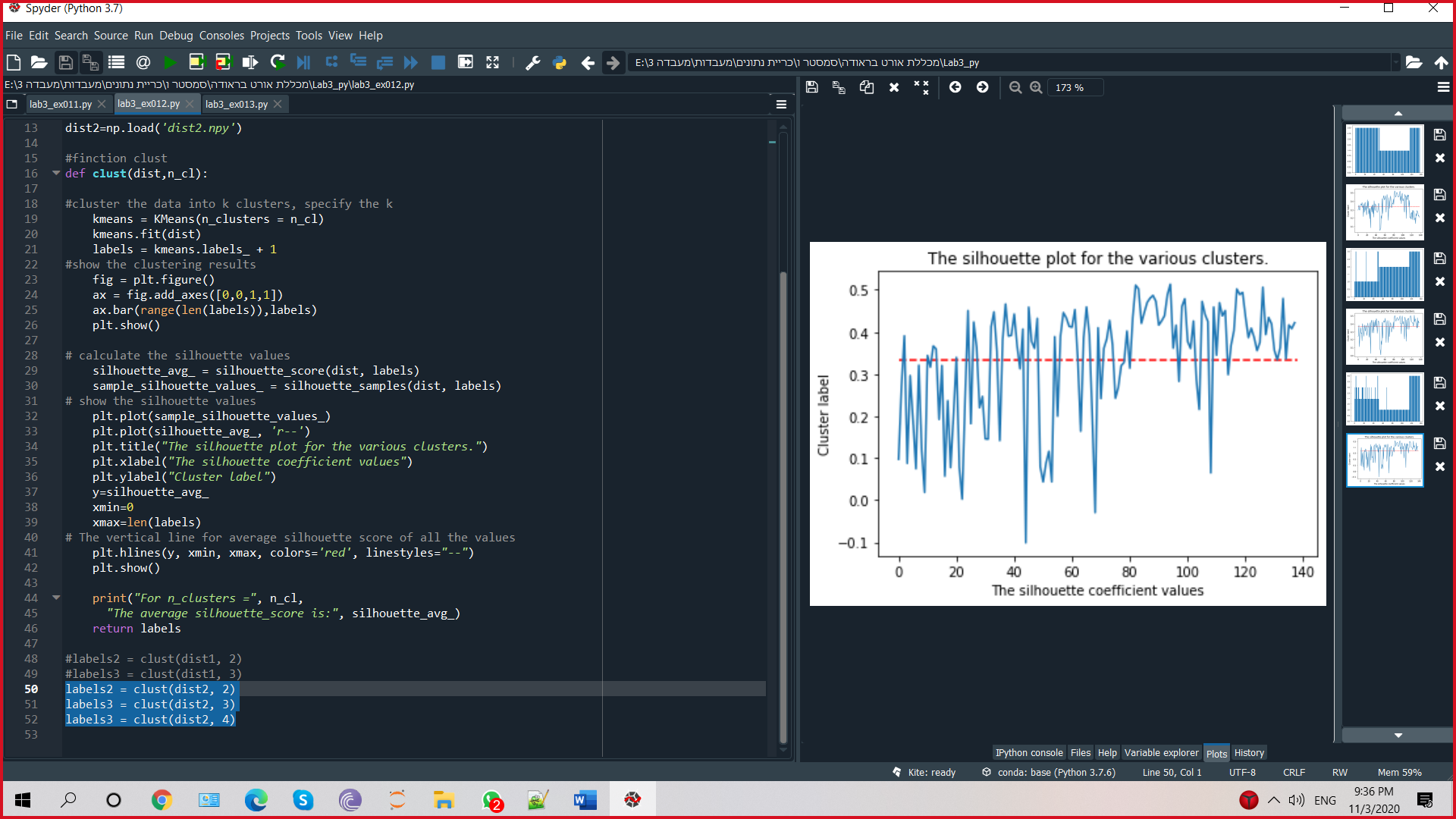
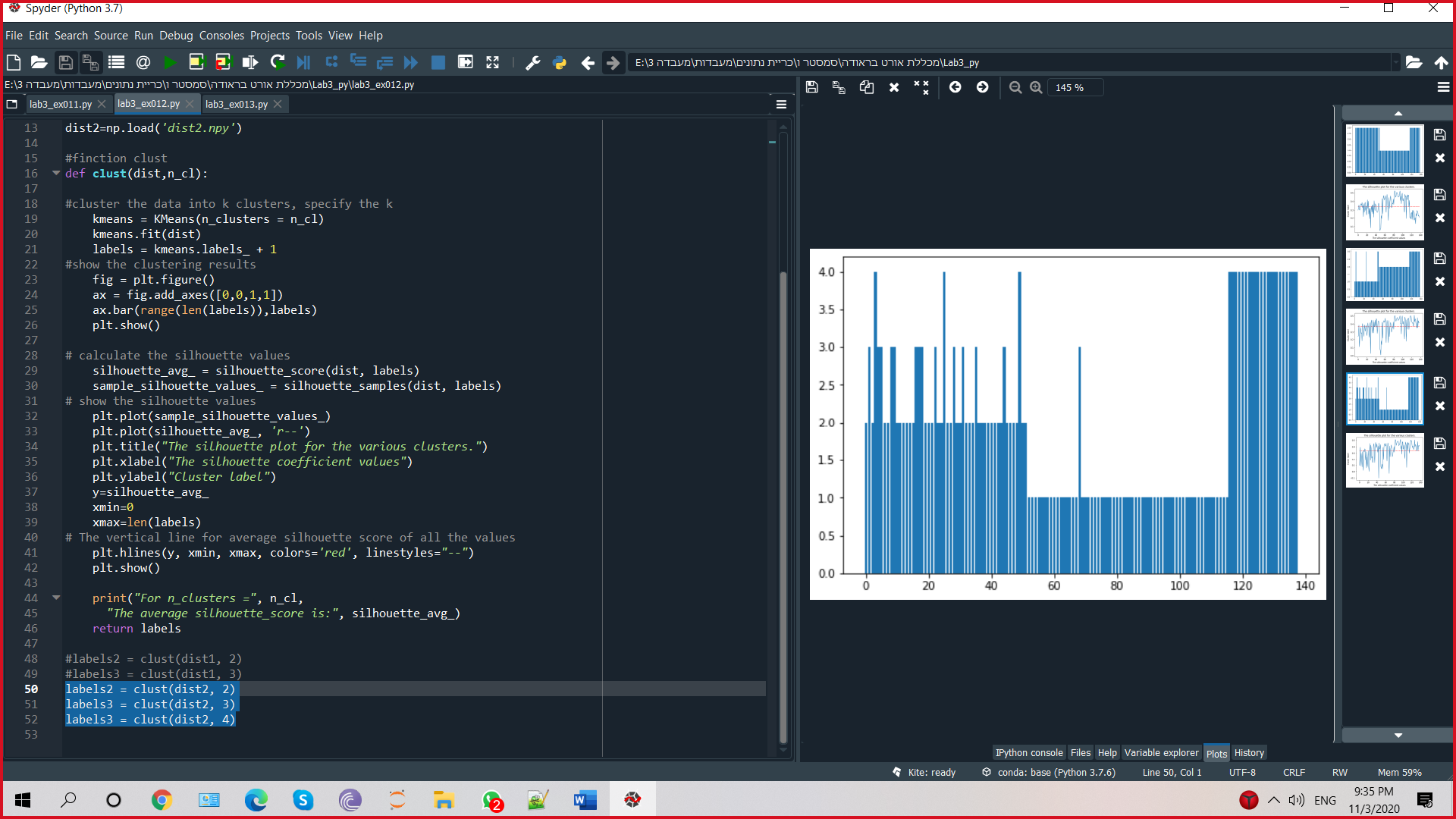
עבור חלוקה ל- cluster=3:



כאשר מחלקים ל-3 clusters ערך הסילואט הממוצע הינו:

. 0.3677785492843207

עבור חלוקה ל- cluster=4:



כאשר מחלקים ל-4 clusters ערך הסילואט הממוצע הינו:

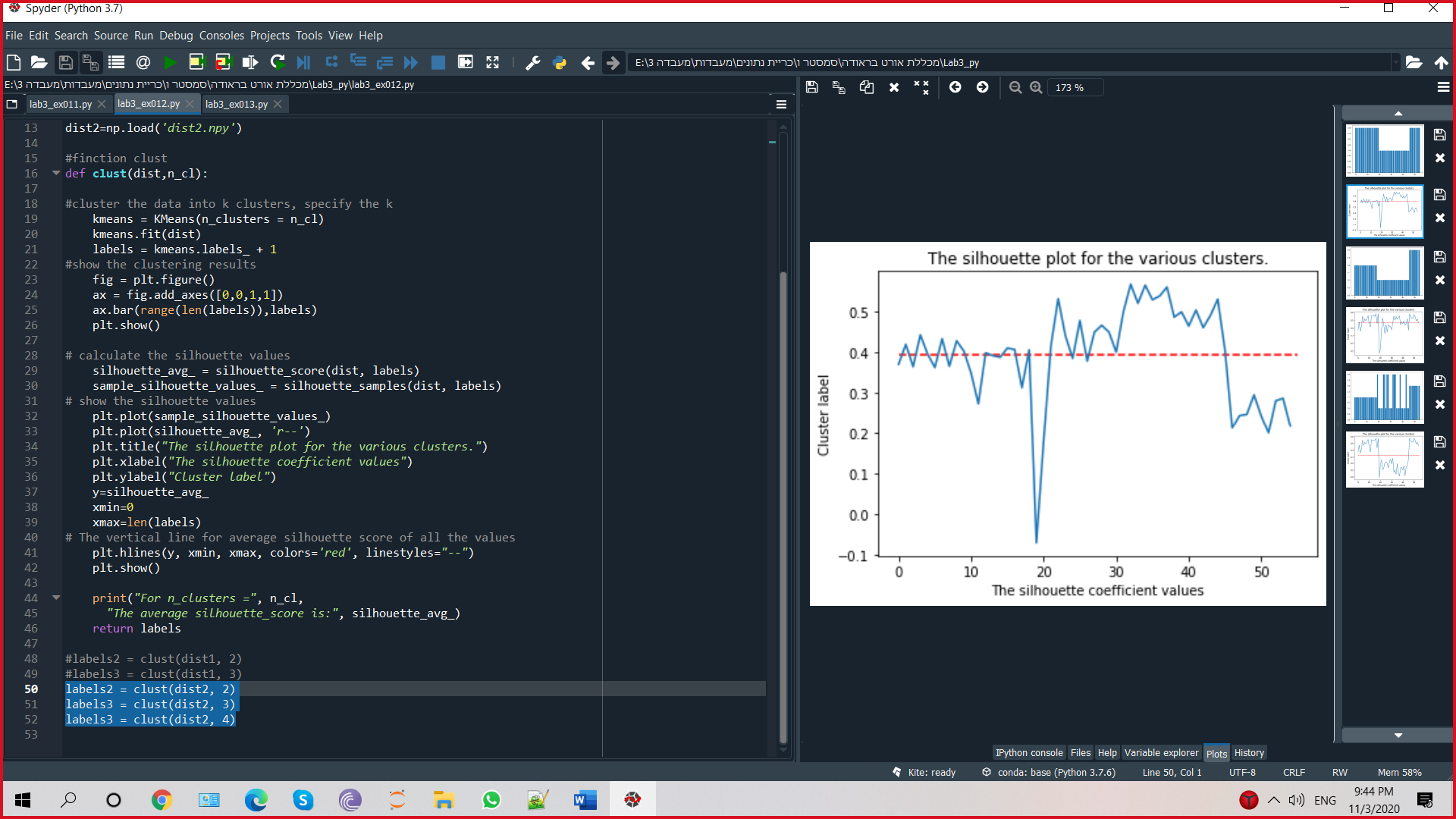
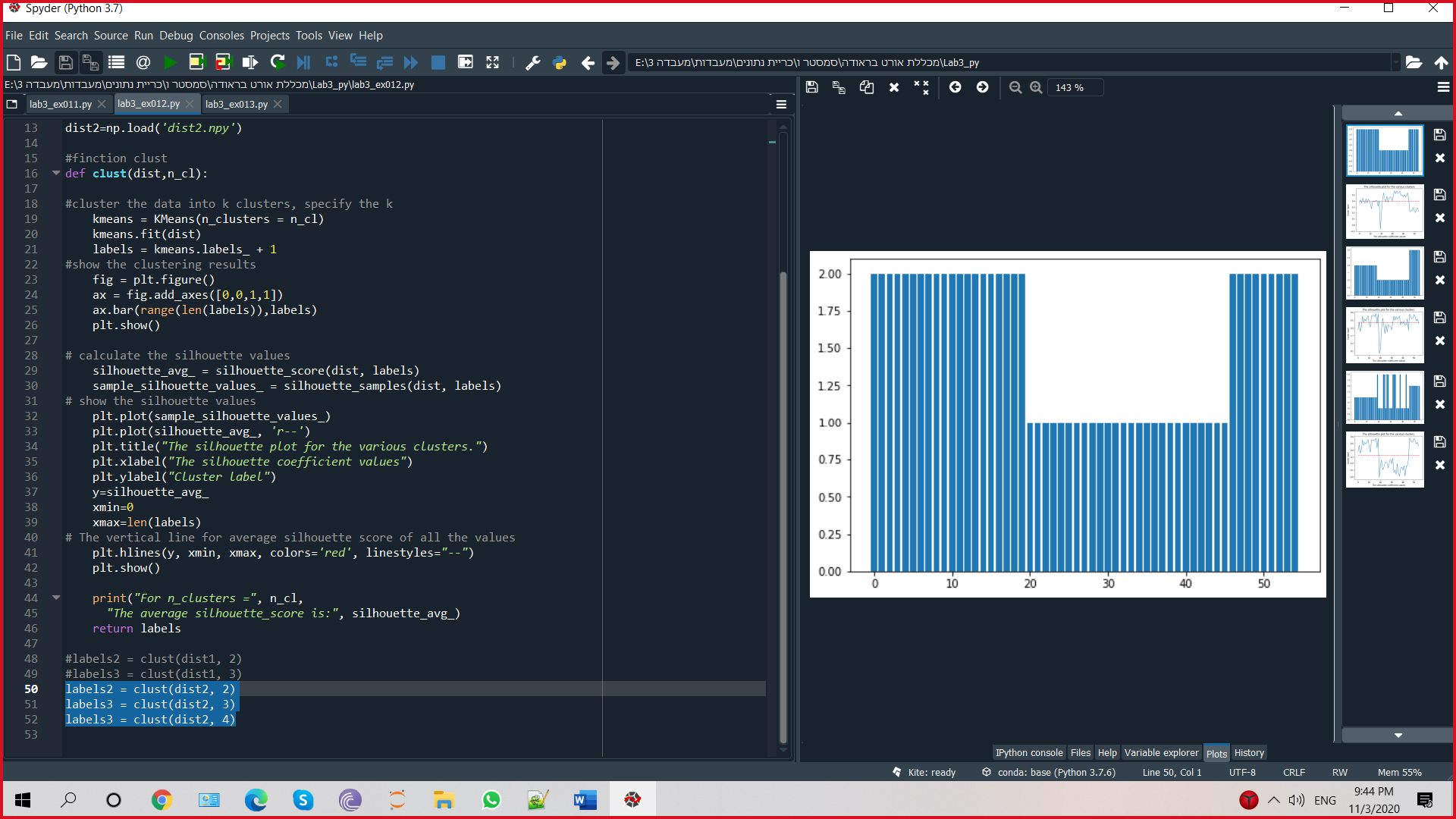
. 0.33501500535969125

מסקנות ניסוי wind = 20000:

ניתן לראות שכאשר החלוקה היא לשלושה clusters אנחנו מקבלים את ערך הסילואט הממוצע הגבוה ביותר (0.36) כלומר החלוקה הטובה ביותר היא כאשר מחלקים את ה-clusters ככמות הספרים השונים.

**לאחר מכן בדקנו עבור wind =50000:**

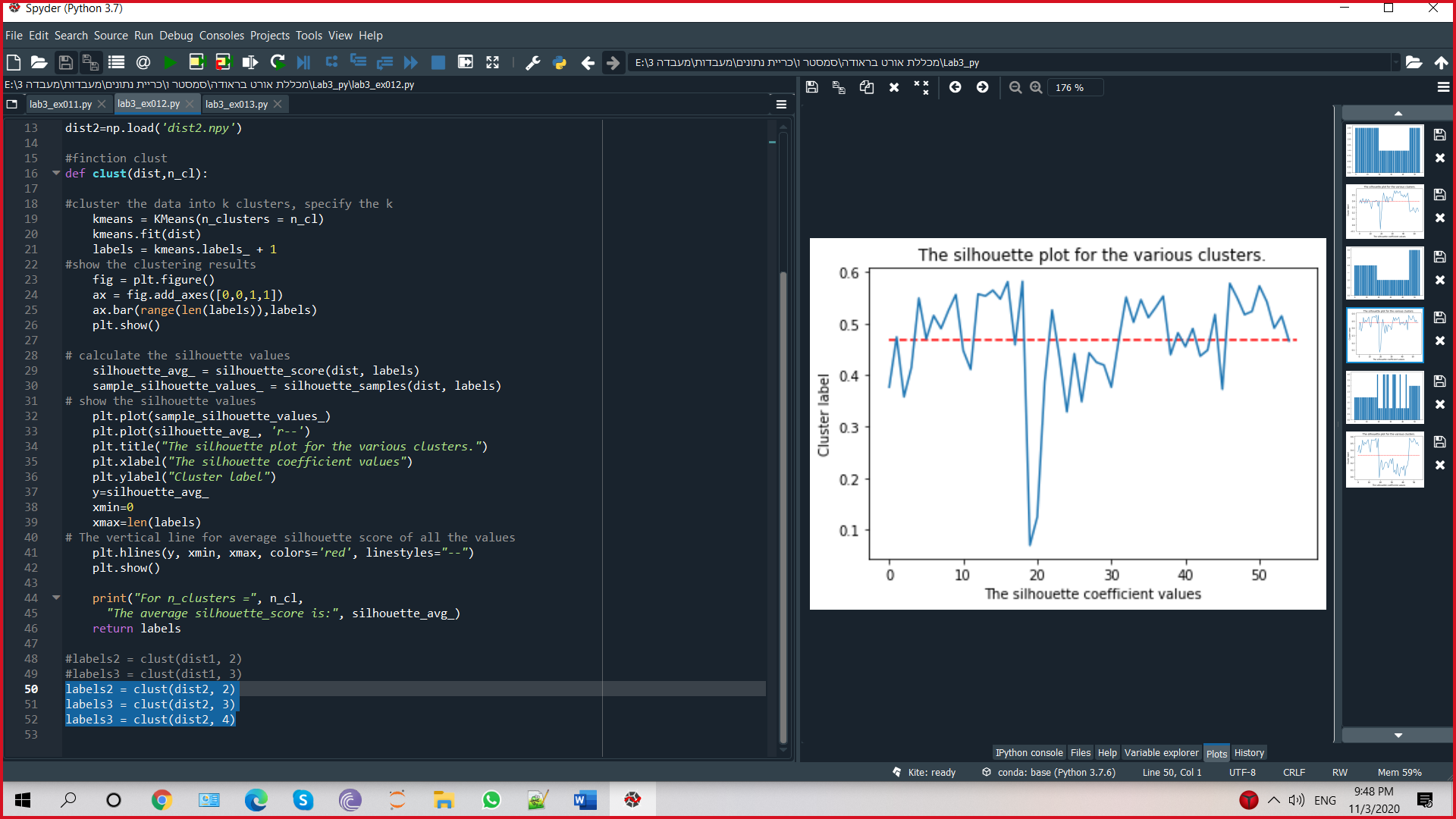
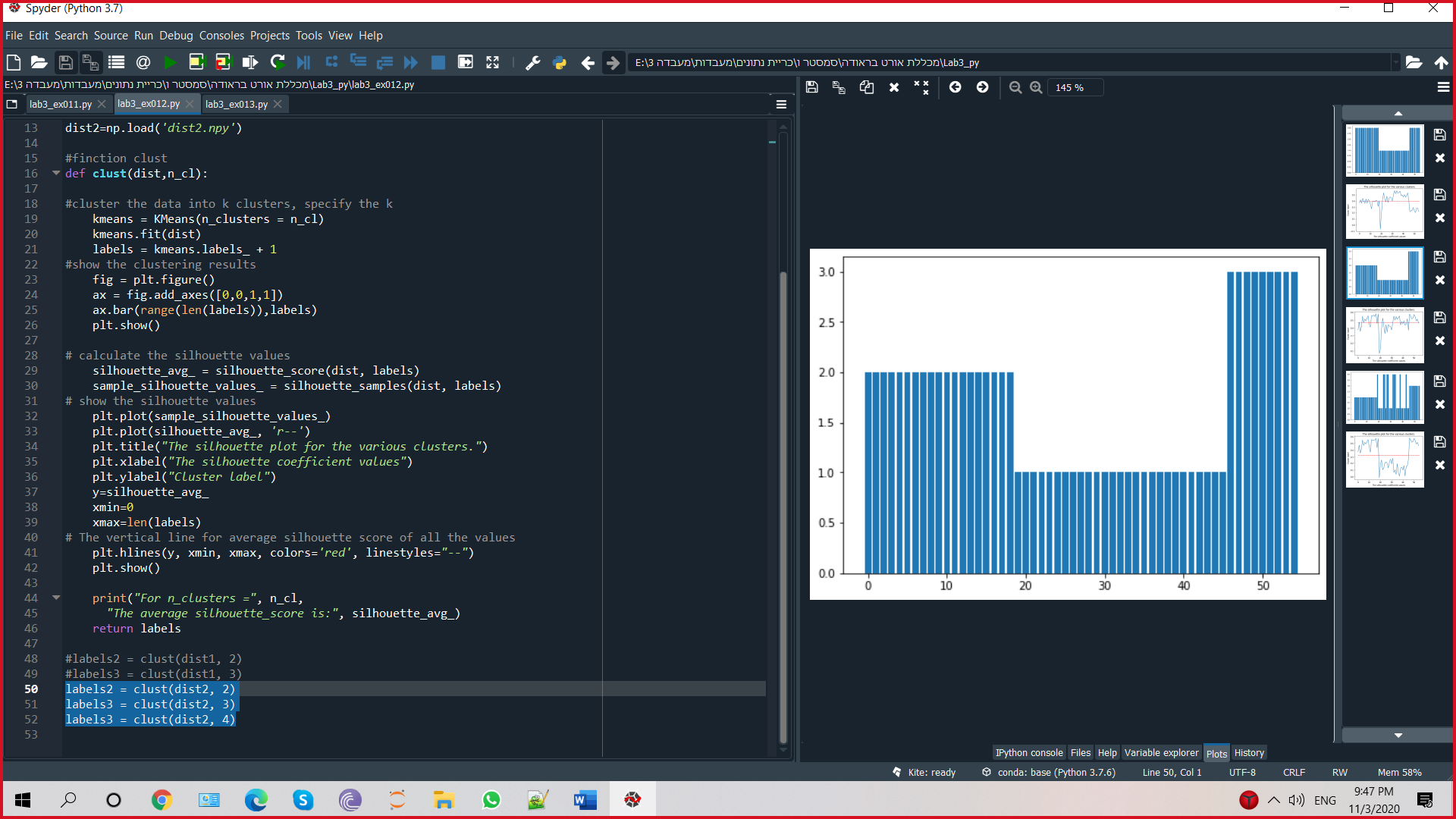
עבור חלוקה ל- cluster=2:



כאשר מחלקים ל-2 clusters ערך הסילואט הממוצע הינו:

0.3930581688497115.

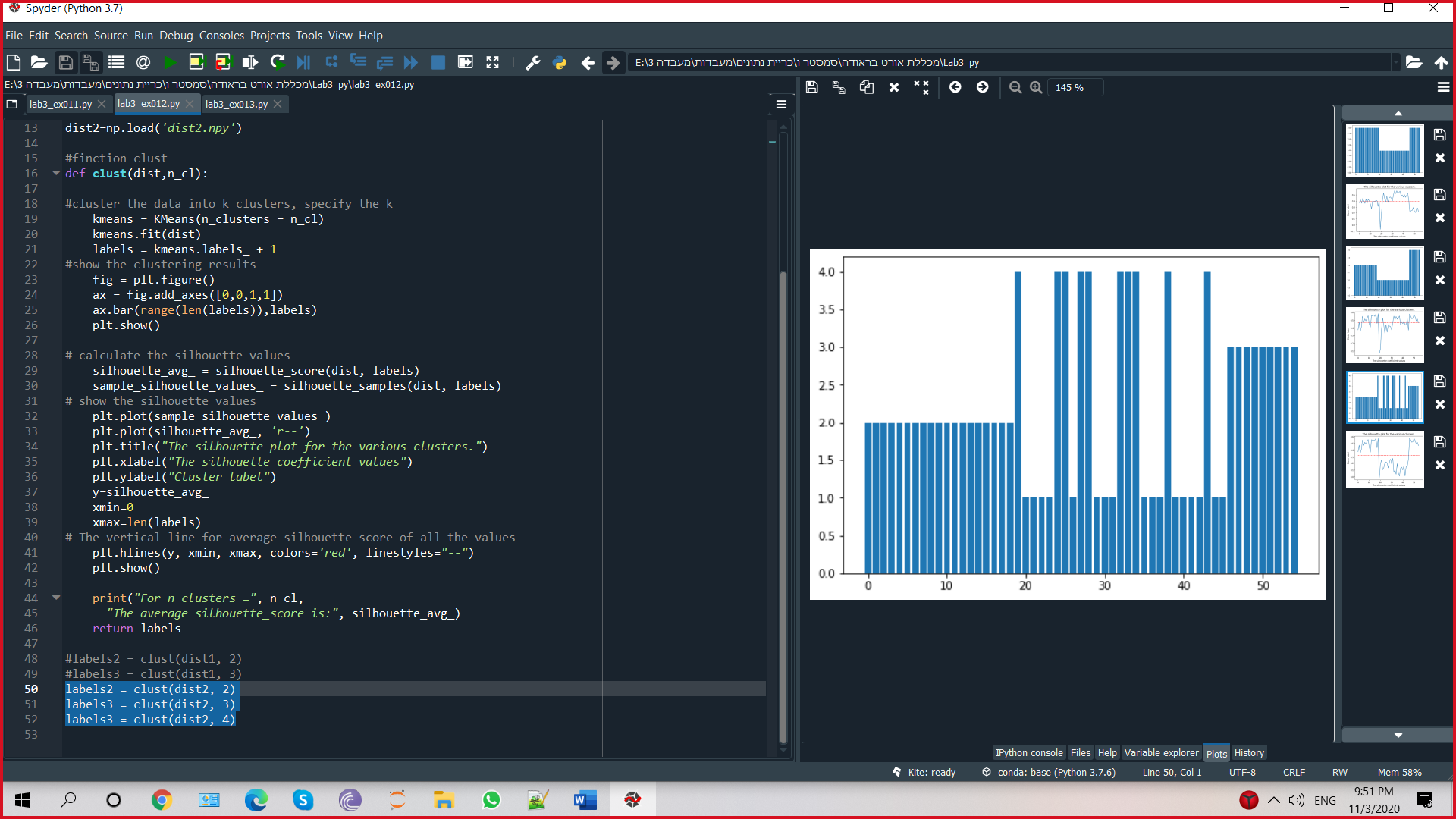
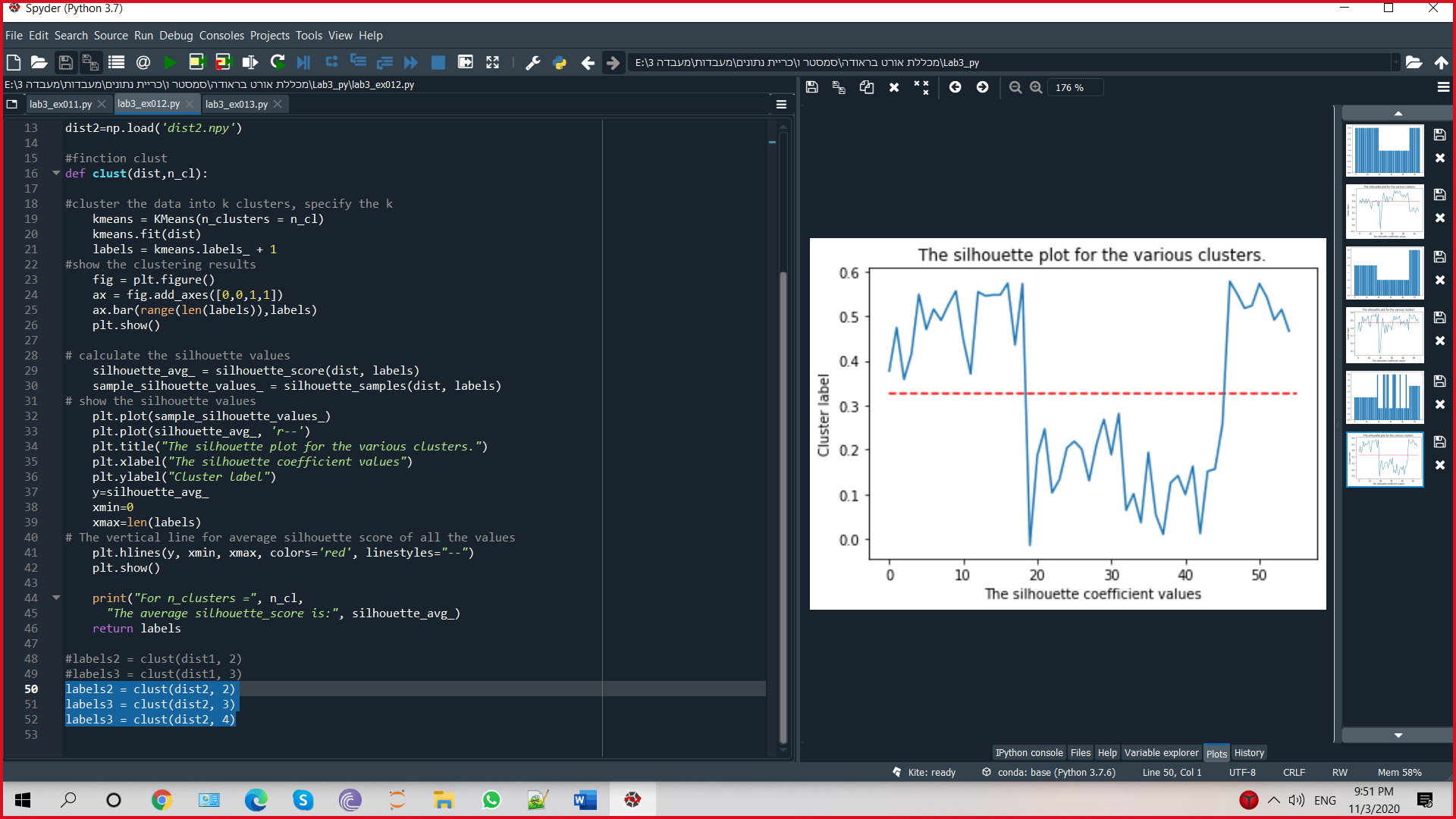
עבור חלוקה ל- cluster=3:



כאשר מחלקים ל-3 clusters ערך הסילואט הממוצע הינו:

0.4695826724282637.

עבור חלוקה ל- cluster=4:

כאשר מחלקים ל-4 clusters ערך הסילואט הממוצע הינו:

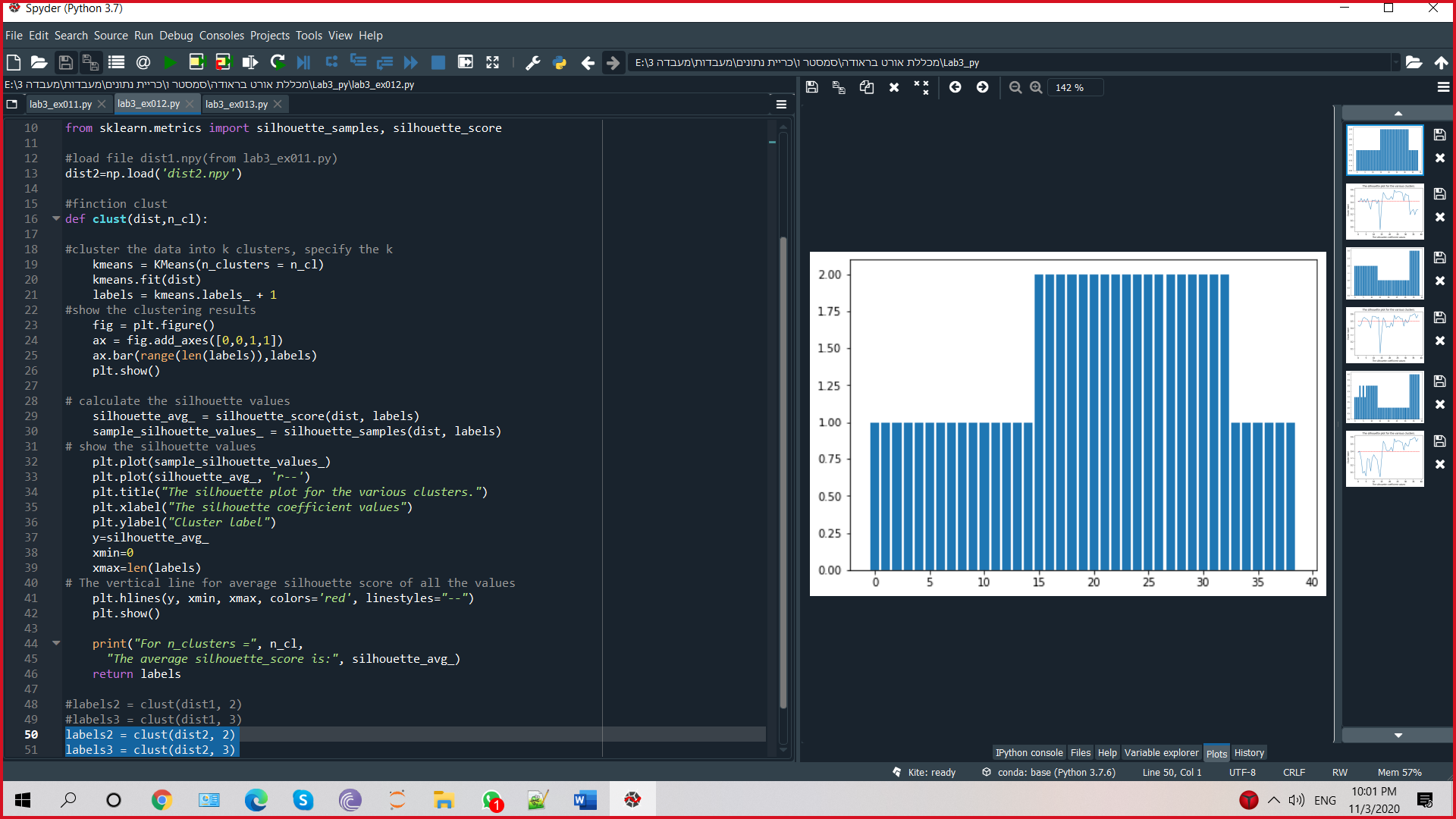
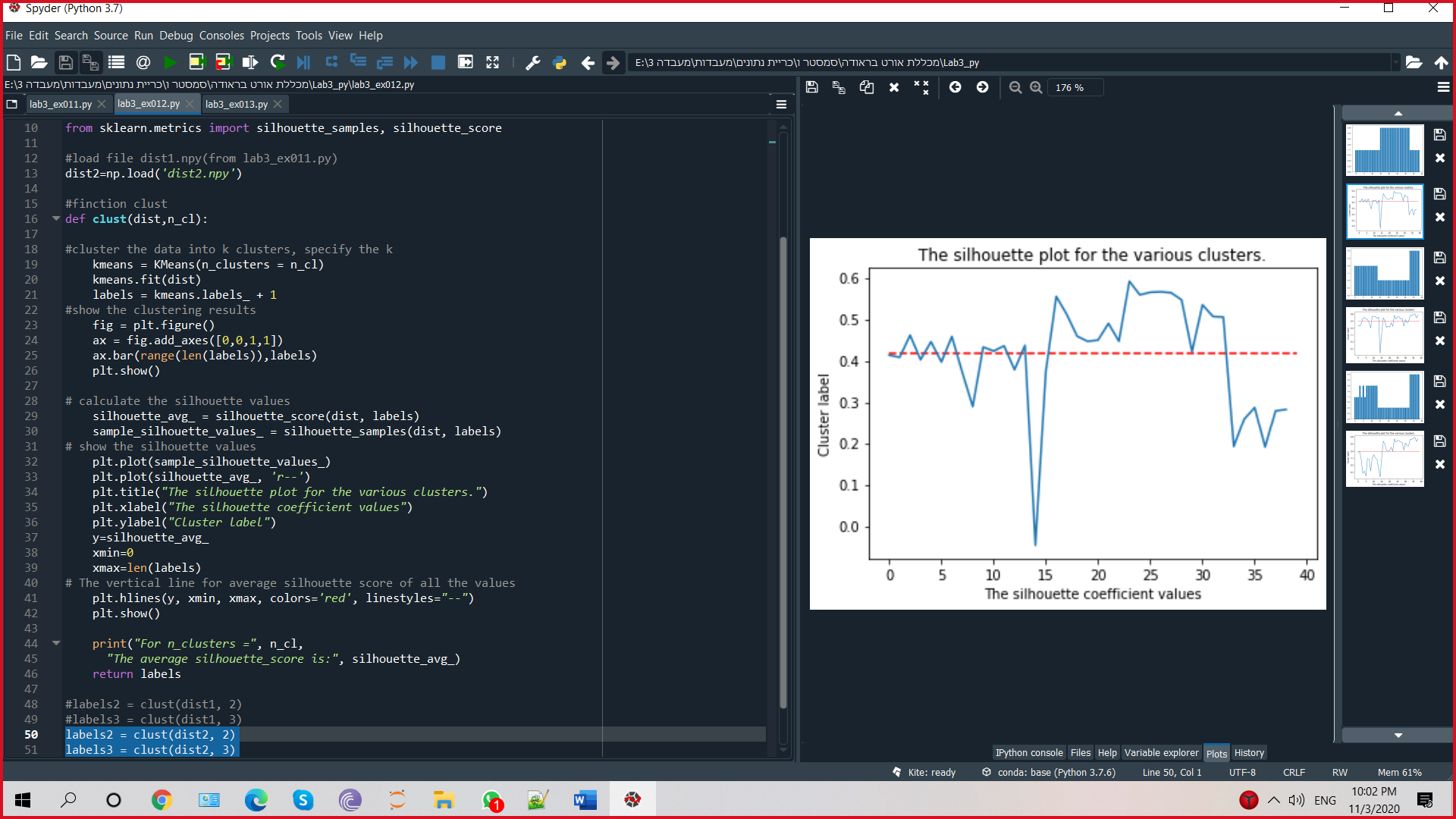
0.3274184686457054.

מסקנות ניסוי wind = 50000:

ניתן לראות שכאשר החלוקה היא לשלושה clusters אנחנו מקבלים את ערך הסילואט הממוצע הגבוה ביותר (0.46) כלומר החלוקה הטובה ביותר היא כאשר מחלקים את ה-clusters ככמות הספרים השונים.

**לבסוף בדקנו עבור wind =70000:**

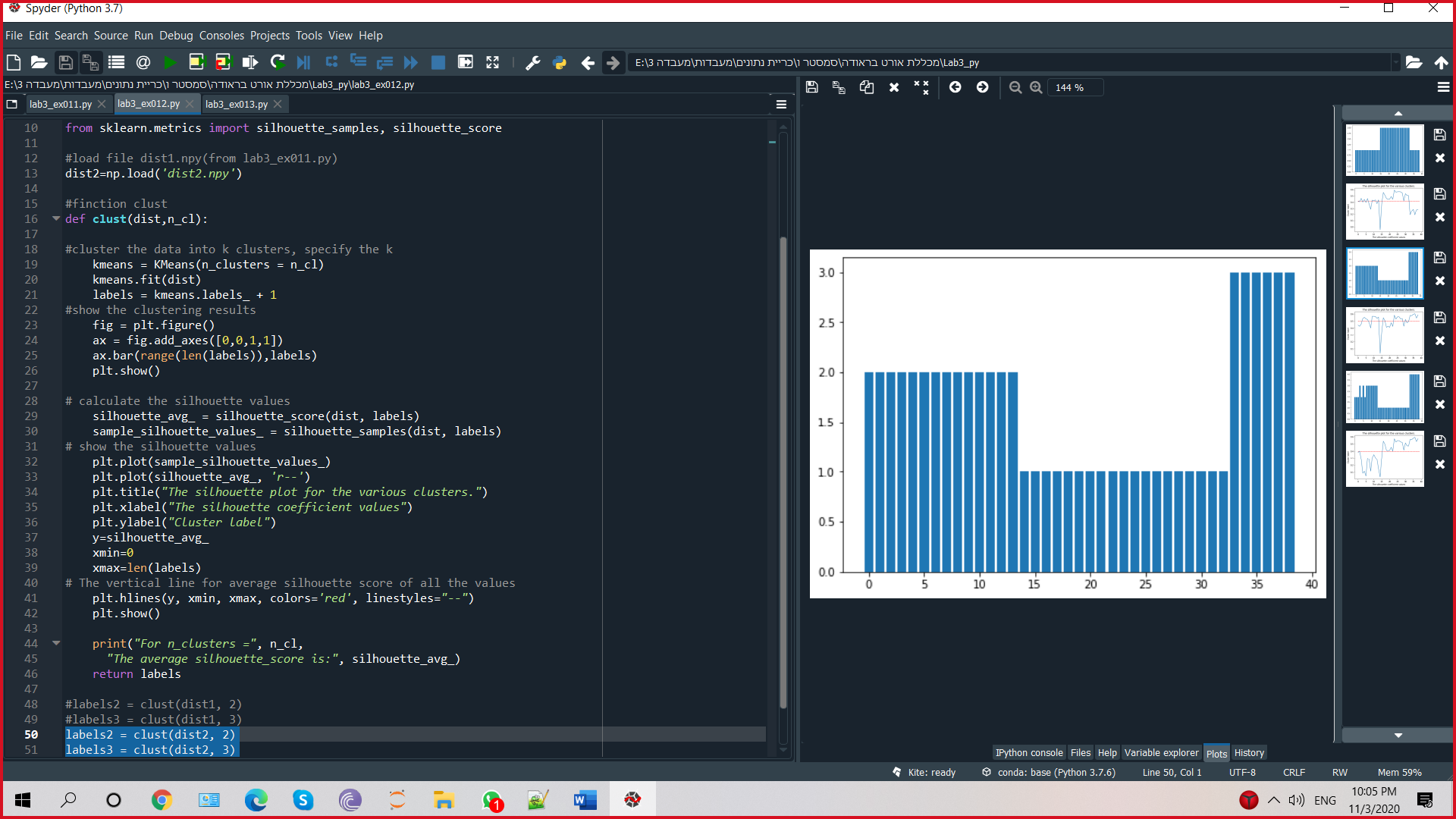
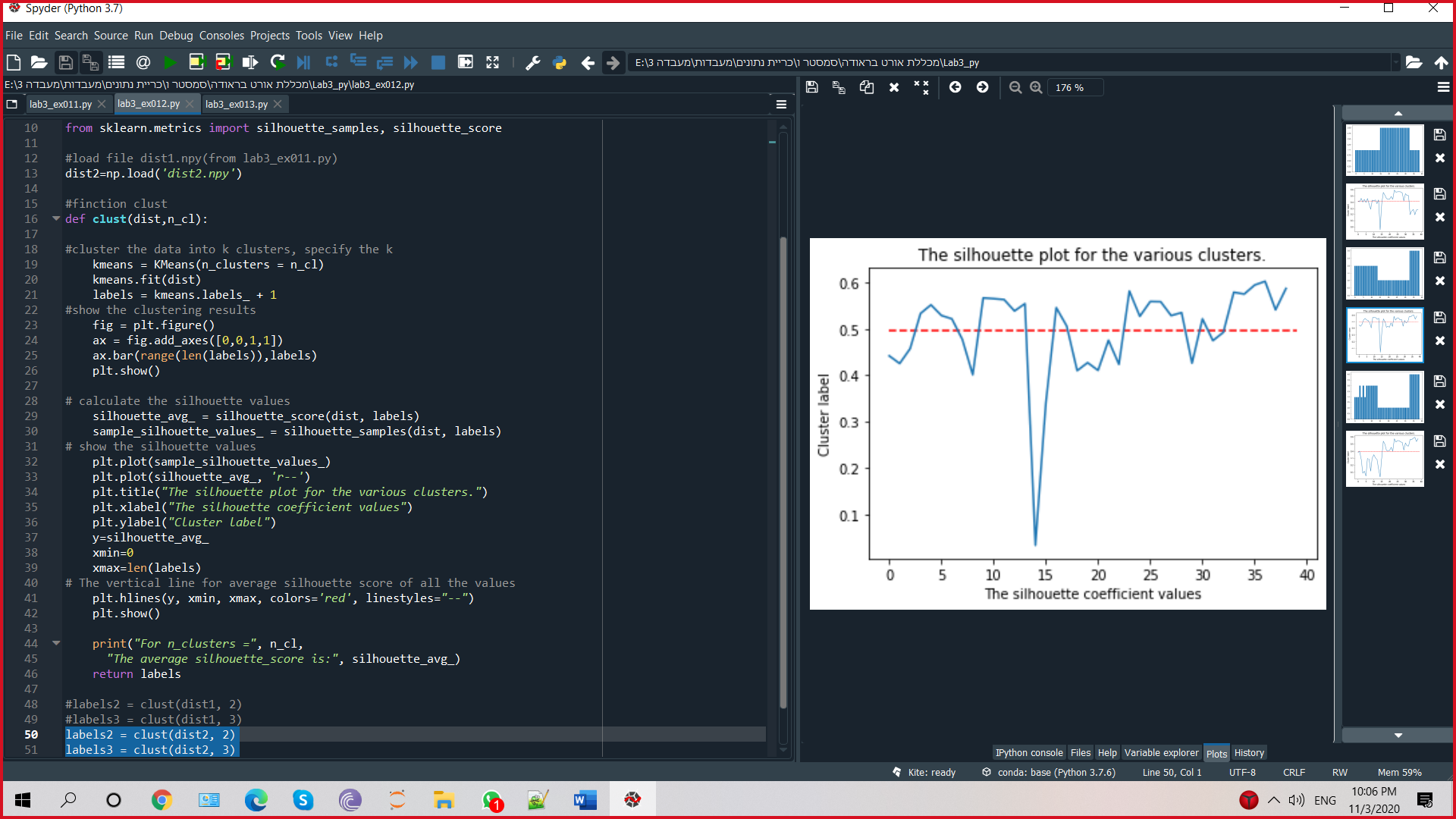
עבור חלוקה ל- cluster=2:

כאשר מחלקים ל-2 clusters ערך הסילואט הממוצע הינו:

0.4178817500517704.

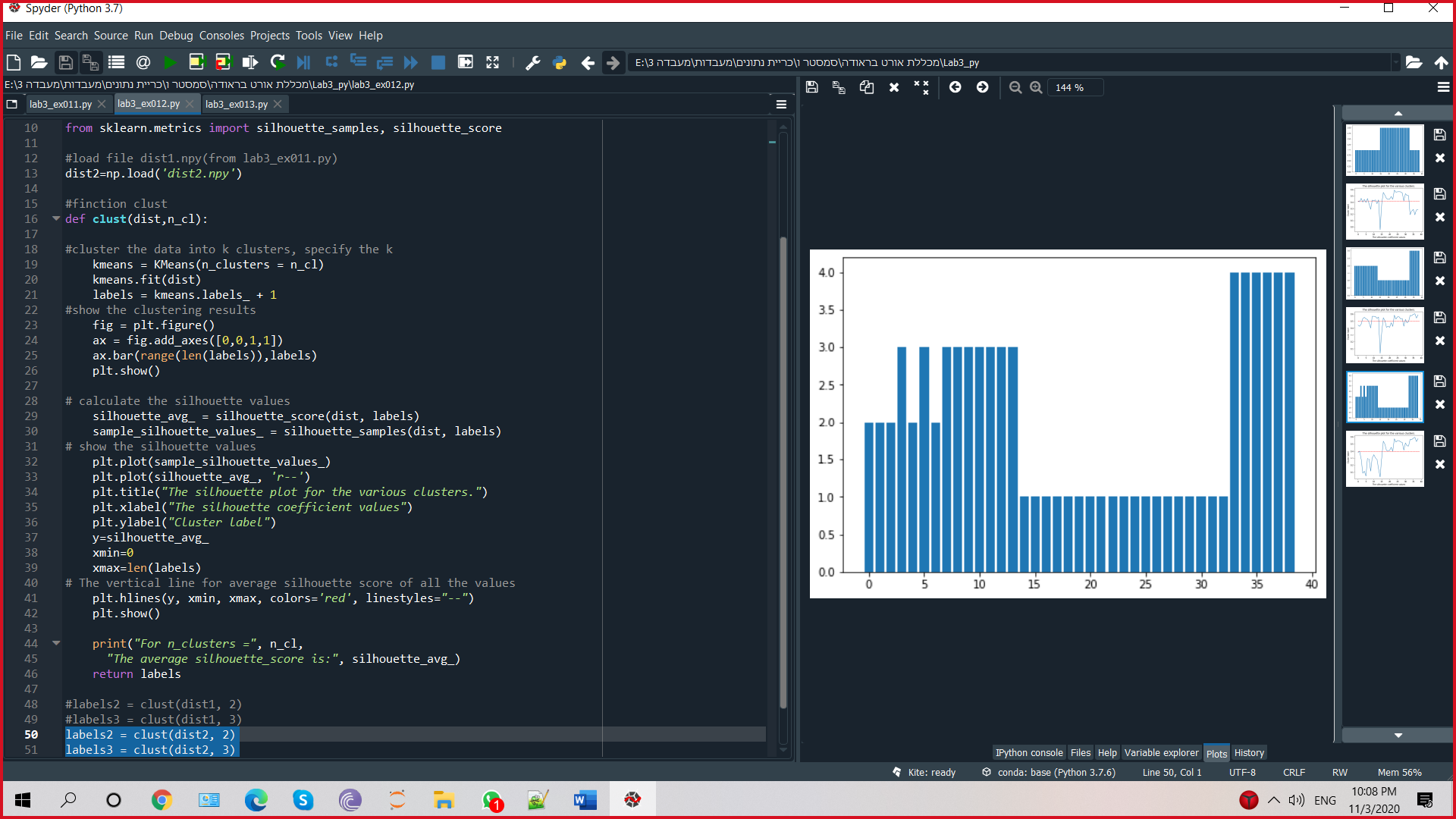
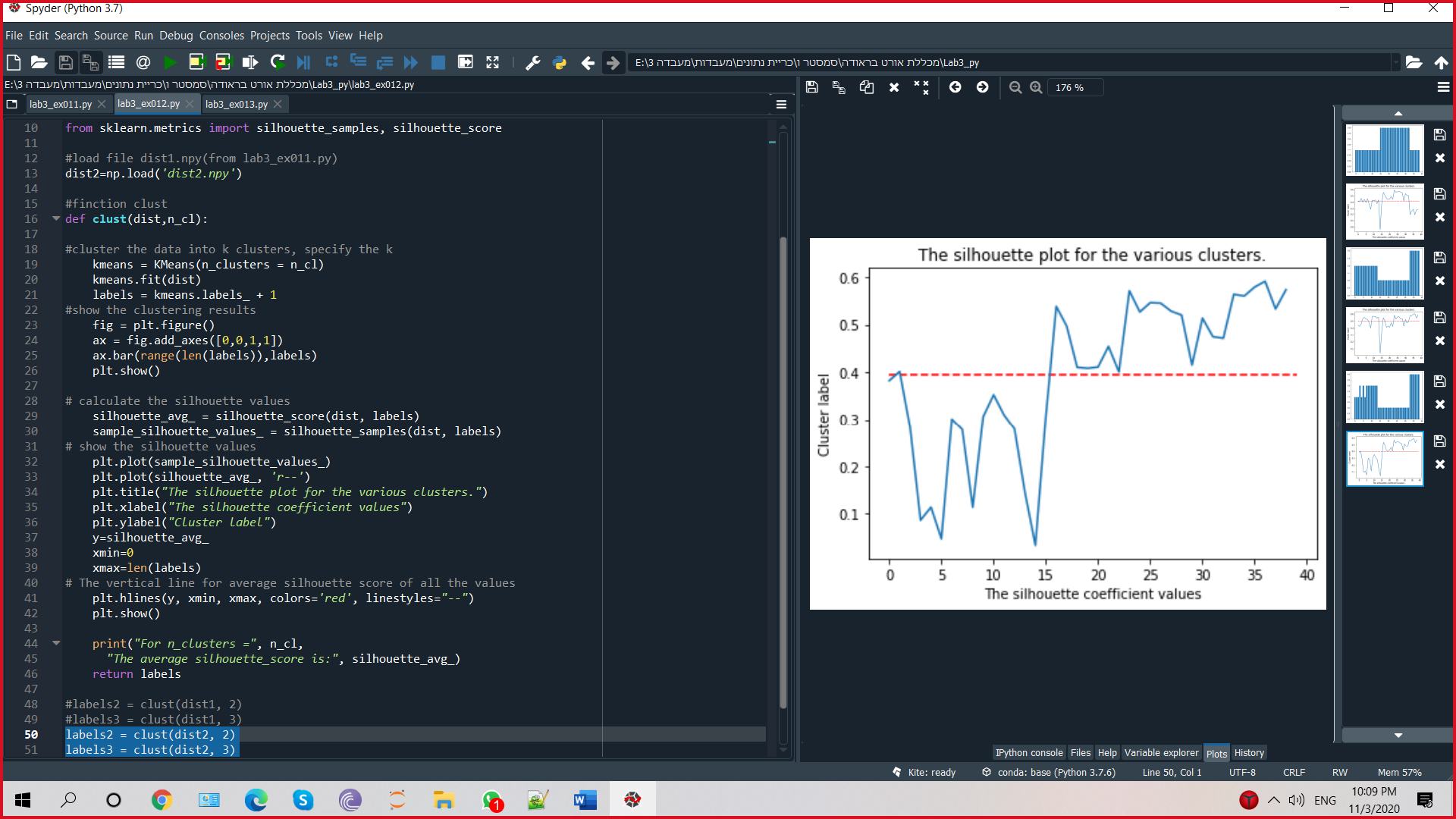
עבור חלוקה ל- cluster=3:

כאשר מחלקים ל-3 clusters ערך הסילואט הממוצע הינו:

0.49681302254325144.

עבור חלוקה ל- cluster=4:



כאשר מחלקים ל-4 clusters ערך הסילואט הממוצע הינו:

0.3941500423717535.

מסקנות ניסוי wind = 70000:

ניתן לראות שכאשר החלוקה היא לשלושה clusters אנחנו מקבלים את ערך הסילואט הממוצע הגבוה ביותר (0.49) כלומר החלוקה הטובה ביותר היא כאשר מחלקים את ה-clusters ככמות הספרים השונים.

מסקנות כלליות מהניסוי:

ניתן להסיק שאם משתמשים ב-clusters ככמות הספרים השונים שבידינו נקבל ערך סילואט גבוה יותר מאשר אם נשתמש בפחות או יותר clusters. בנוסף ניתן לקבוע כי ככל שנגדיל את ערך ה-wind כך נקבל ערך טוב יותר אבל עבור ערכים גדולים, הגדילה בערך הסילואט כמעט ולא תורגש.