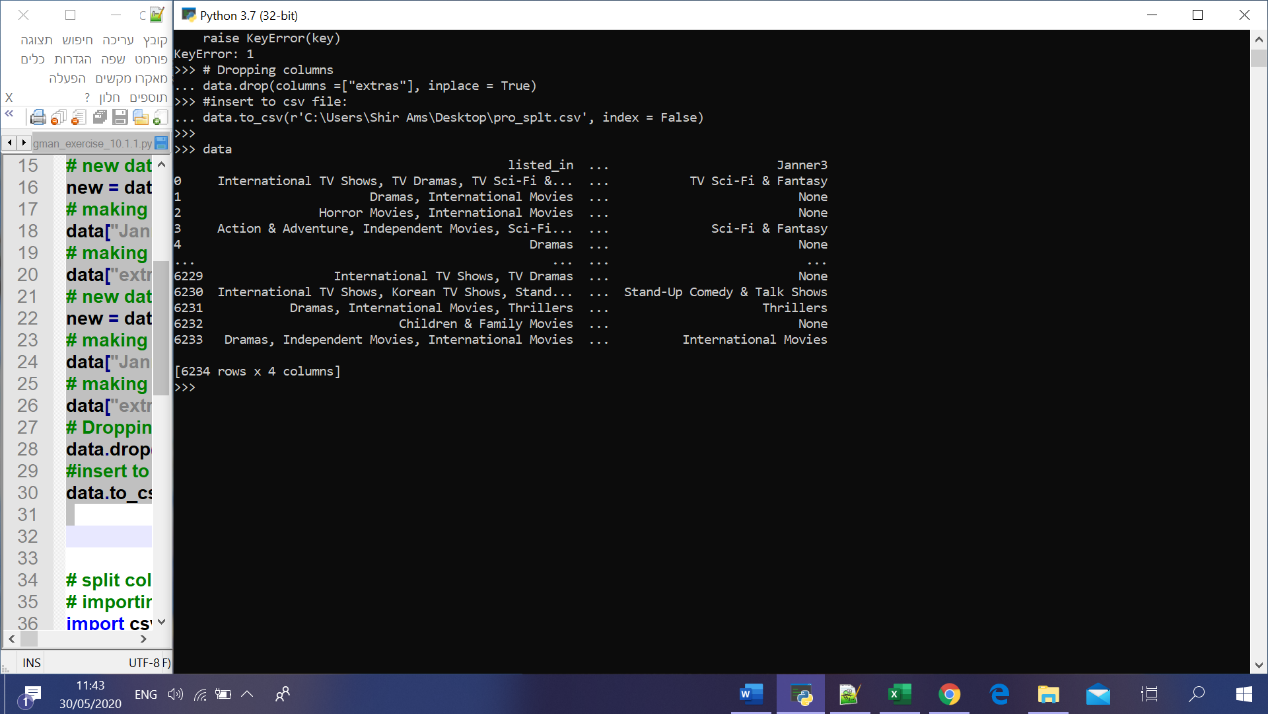
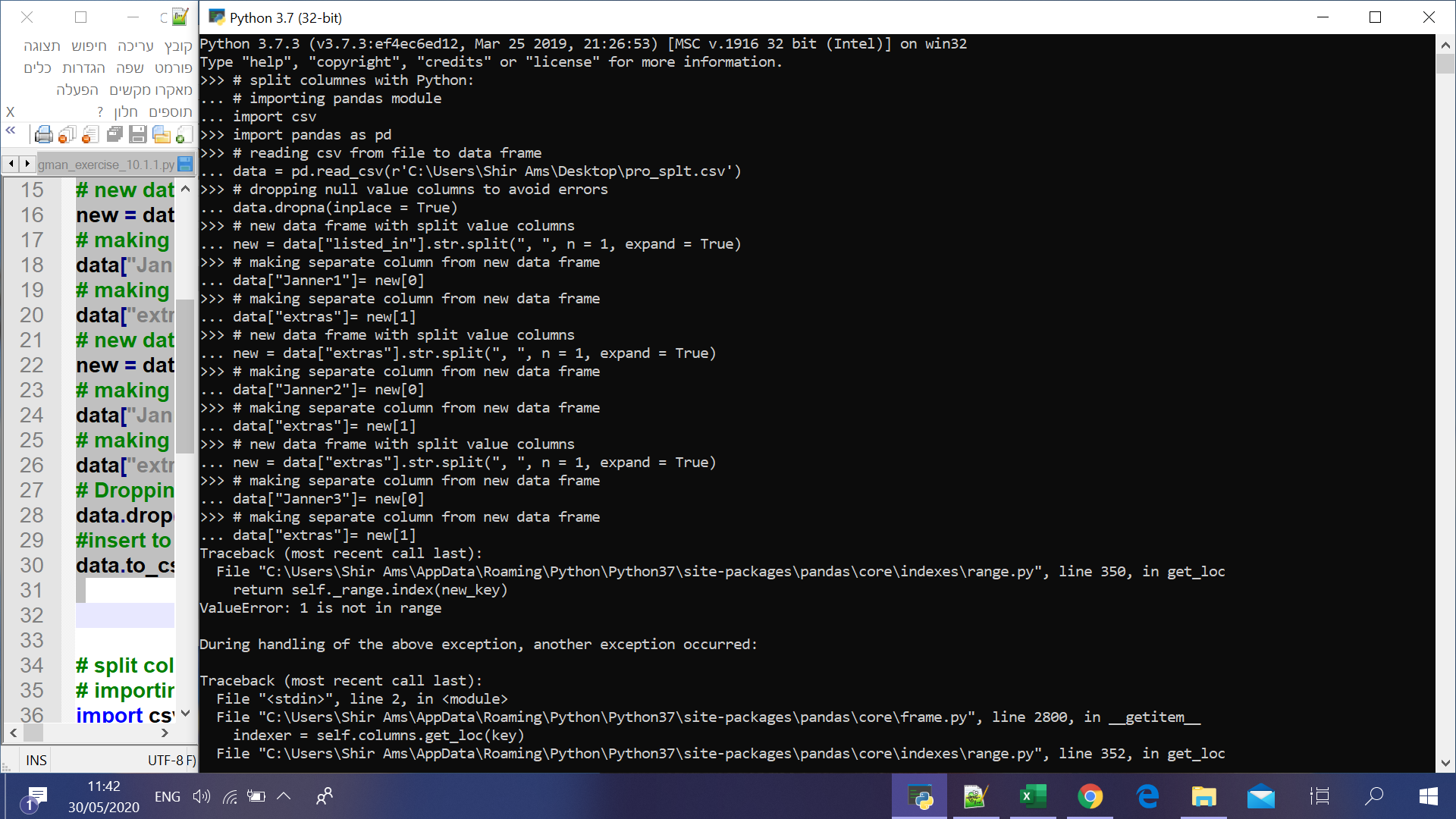
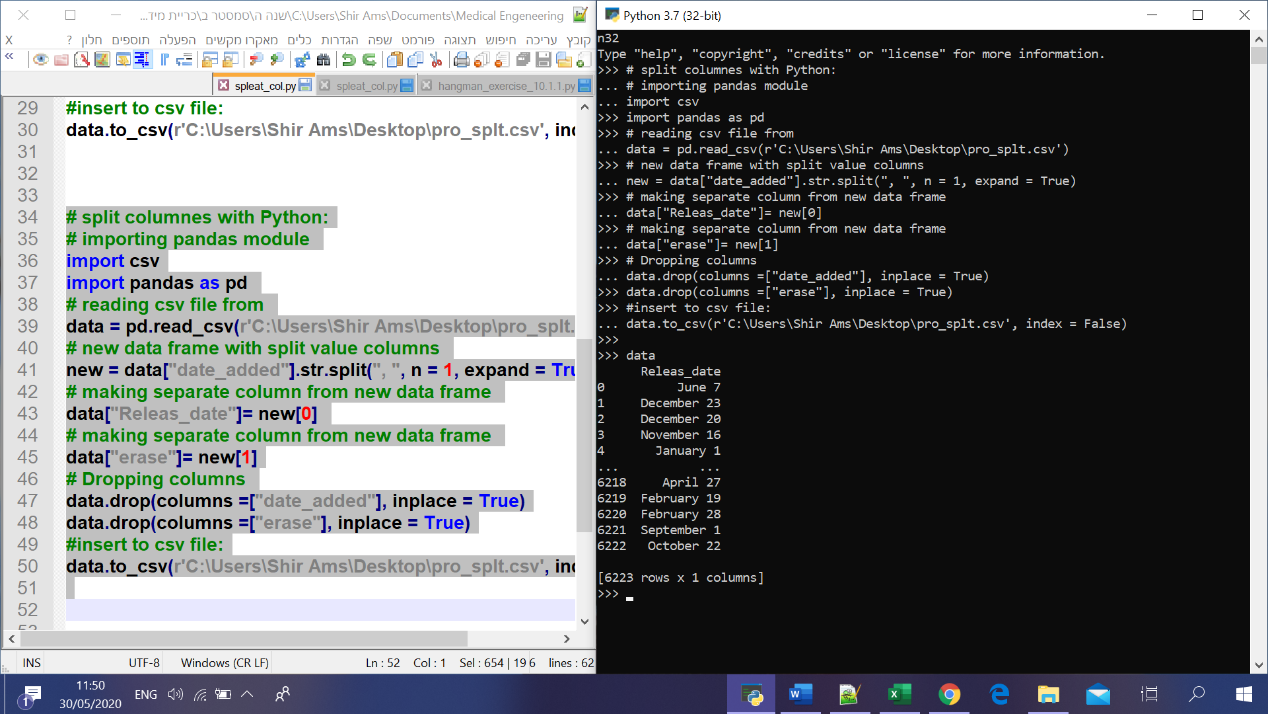
במסמך זה ישנו תיעוד של הזנת הקוד של הקובץ "spleat\_col" בפיטון גרסה 3.7, עליו הותקנה חבילת PANDAS. הטור הראשון עליו הקוד הופעל, הינו הז'אנר.  
בנוסף, מתחת לכל קוד הוצג המשתנה מטיפוס data frame אותו יצרנו, לצורך תצוגה של התוצאה של כל קוד לאחר העיבוד (את המצב הקודם ניתן לראות בקובץ המקורי).  
  
**ז'אנרים:**  
כפי שניתן לראות יש שגיאה לאחר הזנת מידע לטור קיים בפעם האחרונה. אולם למרות זאת הפיצול בוצע בצורה מיטבית.  
הטור "listed\_in" נמחק בצורה ידנית, בגלל ששכחנו להזין את מחיקת הטור שלו.

יש לשים לב כי לא בוצע פיצול לאחר השימוש בתו "&" כנהוג בדרך כלל. הסיבה לכך היא שז'אנרים אלה מקוטלגים יחד כז'אנר אחד ולא כשני ז'אנרים מפוצלים בשידורי סטרימינג של נטפליקס.



**תאריך:**

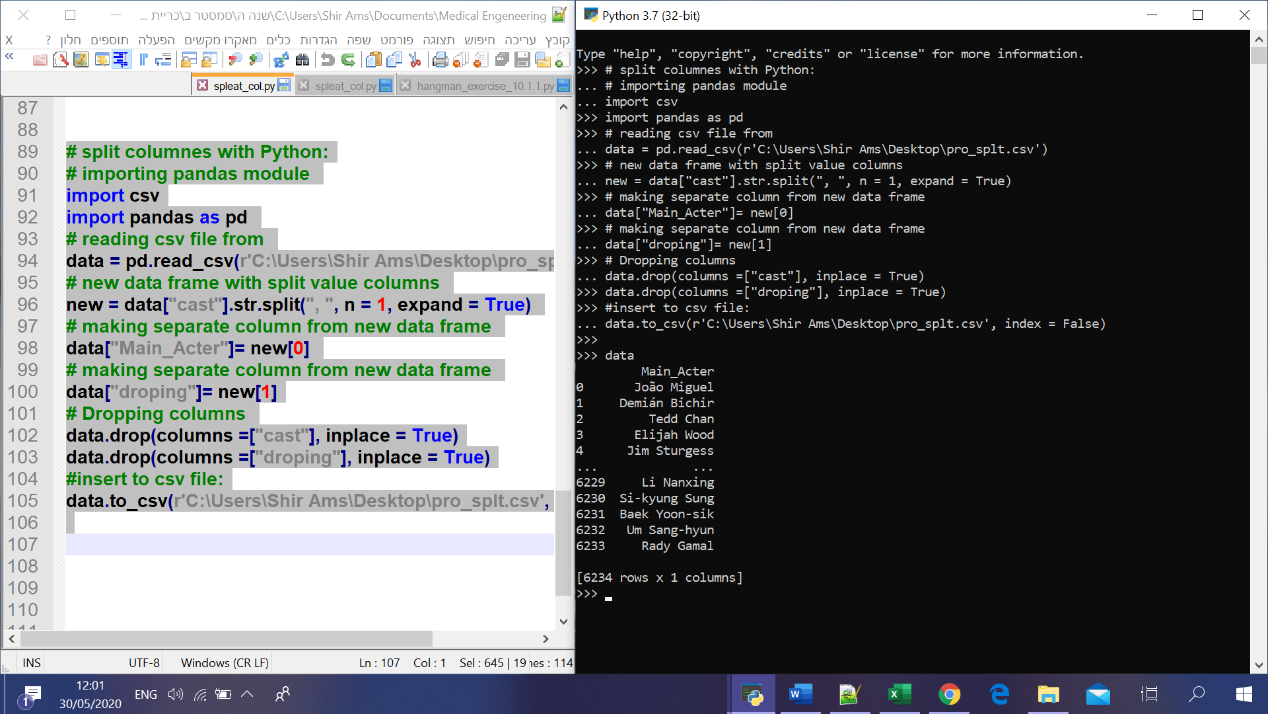
בשל כפילות במידע על שנת הוצאה, הוחלט להשאיר את הטור הקיים של שנת ההוצאה, ולהסיר את השנה מתוך התור של תאריך ההוצאה. בכך קיבלנו את התאריך של הוצאת התוכן (הסרט\ סידרה) בשני טורים: שנת הוצאה, ותאריך.

כפי שניתן לראות כאשר מדובר בטור אחד לא מופיעה השגיאה מלעיל.

**שחקן ראשי:**

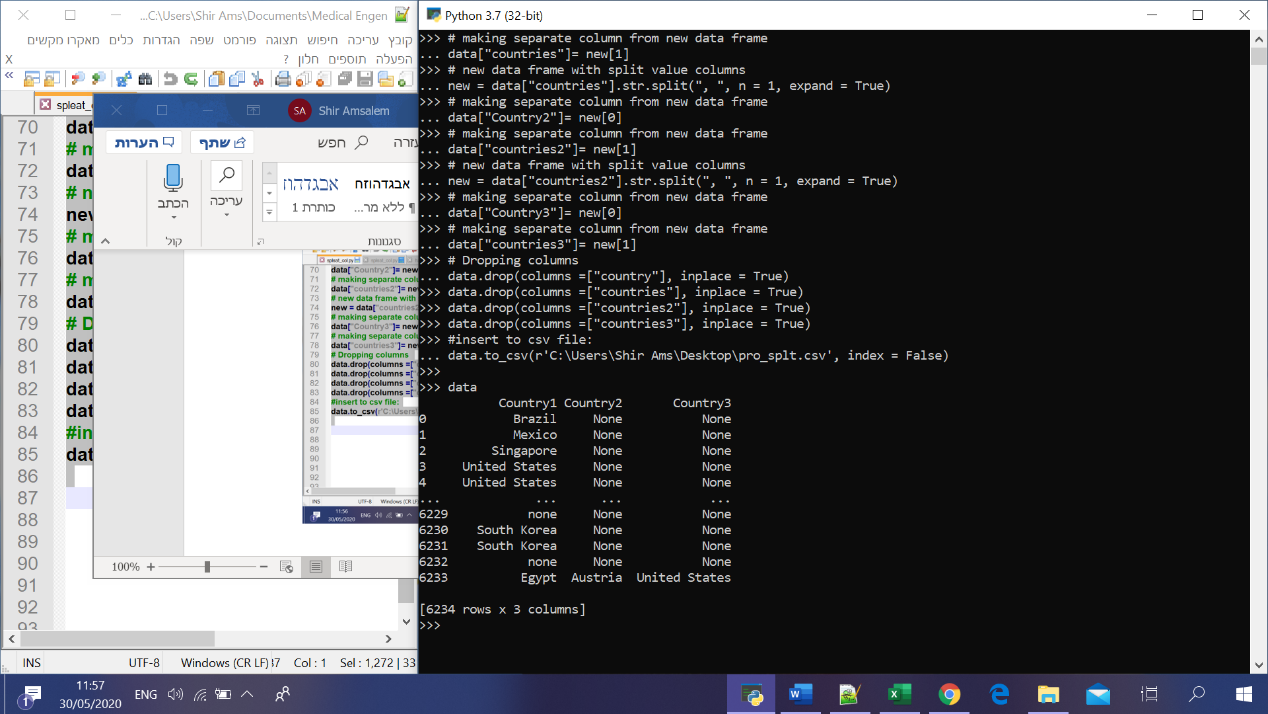
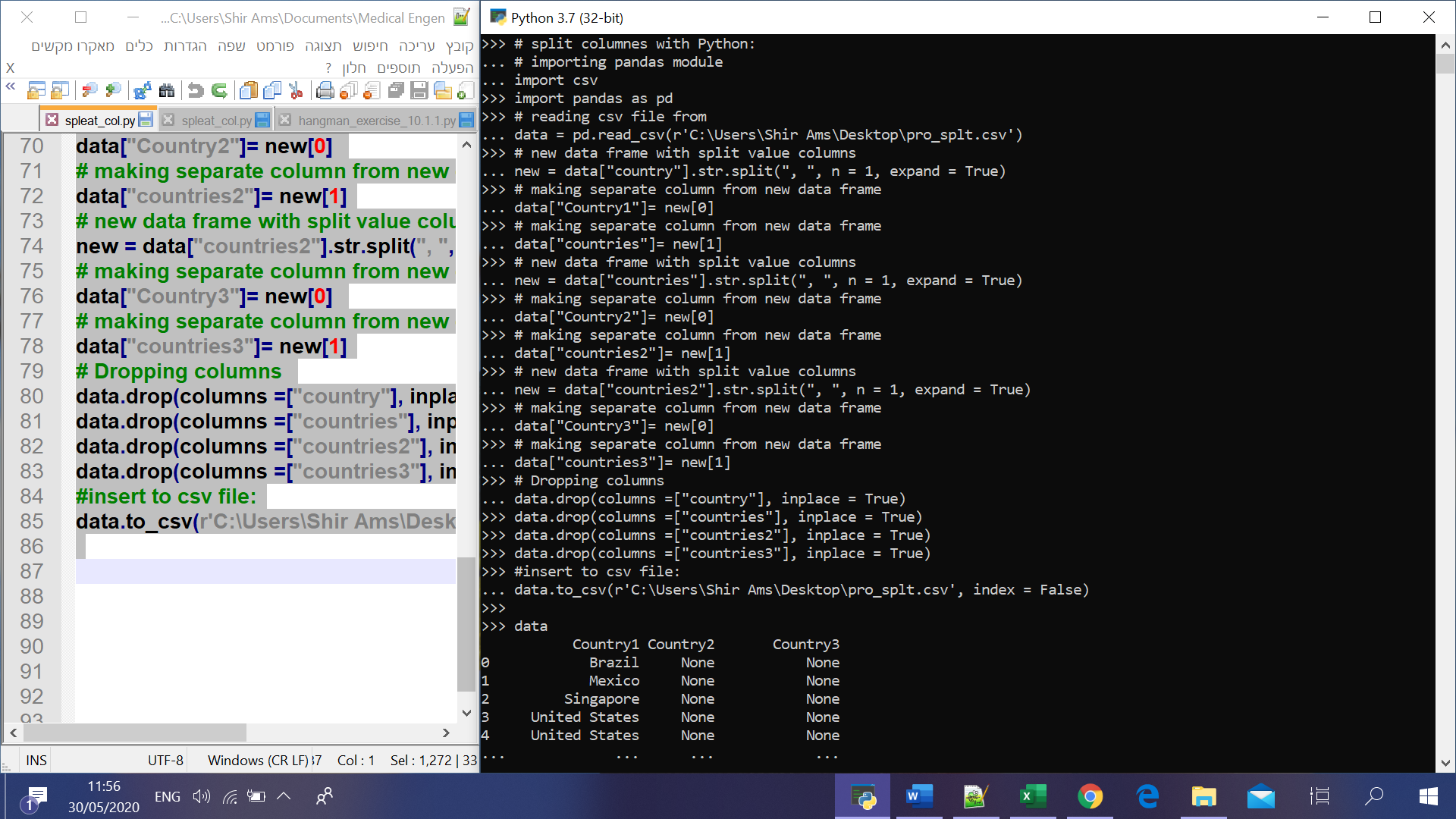
כפי שתואר במבוא- הנחות (**היפר קישור**), בטור המציין את השחקנים בתוכן (סרט\ סדרה) מופיעים כלל השחקנים. לצורת יכולת ניתוח של המידע הניתן לנו, נאלצנו להניח כי השחקן המופיע ראשון בטור "cast" הוא השחקן הראשי, ועליו נבצע את בדיקת הדירוג שלנו.

התוצאה היא טור אחד הכולל בתוכו את השחקן הראשי של כל סרט\ סדרה.



**מדינה:**

חלק מהתכנים צולמו במדינה אחת, חלק במספר מדינות.  
כרגע לא נסתכל על כמות השורות הריקות של הטורים במדינות השניות והשלישיות, אלא נפצל אותן ונשקול להשאירן כפי שהשארנו את הז'אנרים. זאת לצורך ספירה שלהן בבדיקה שלנו אילו מדינות מניבות את הדירוג הגבוהה ביותר.

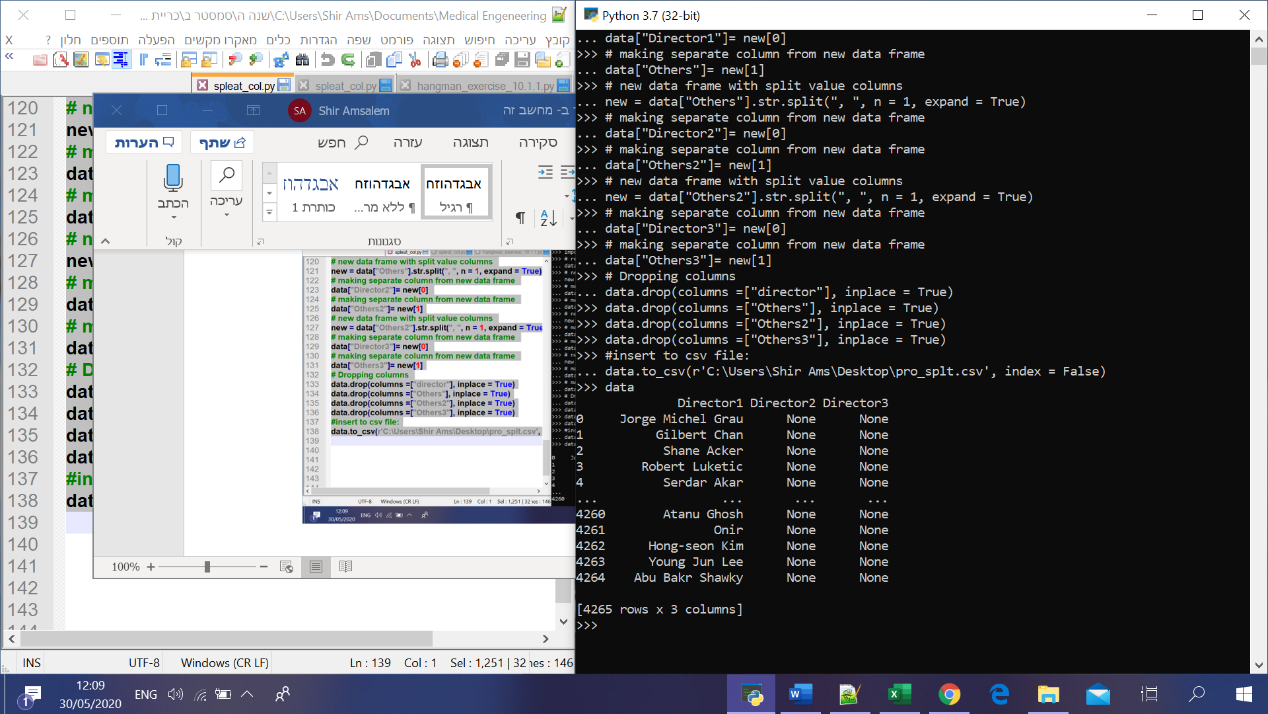
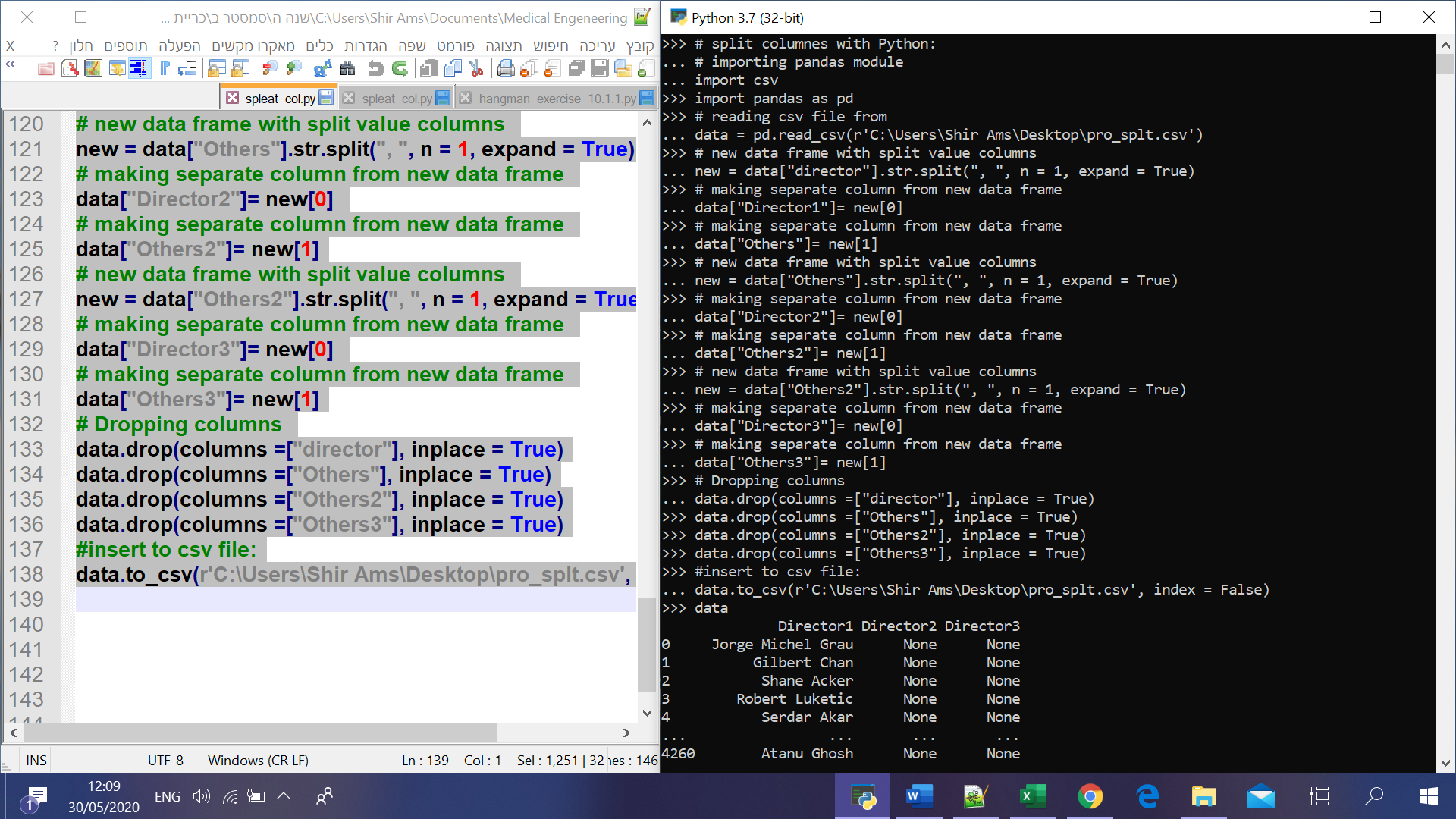


יש לציין כי ניתן לראות שאכן הוספת טורים חדשים במקום הזנה לטור קיים אכן פתר את השגיאה שהוצגה בקוד הראשון.  
  
התוצאה היא שלושה טורים שמתעדים את המדינות בהן התוכן צולם.

**במאים:**

בדומה למדינות, טור זה פוצל לשלושה טורים רלוונטיים על מנת לבדוק בנפרד את השפעתם של כל אחד מהבמאים על "הצלחת" התוכן, כלומר הדירוג גבוהה שניתן לו.

גם כאן ריבוי הטורים מנע שגיאת קוד, ובוצע בצורה מושלמת.



**אורך התוכן:**

שלב זה בוצע כתוצאה מצורך לאחר דיסקרטיזציה. עבור אורך התוכן נבטל את השימוש במילים "עונה" ו"דקות" בהן אין צורך בשל השימוש בטור "type" הקובע אם מדובר בסידרה- כאשר מחולקת לעונות, או מדובר בסרט- שאורכו הוא בדקות.

המטרה של זה לבדוק את השארתינו כי ככל שהזמן צפיה ערוך יותר, כך הדירוג נמוך יותר- כלומר ישנו קשר ישיר הפוך (קורולציה) בין נתונים אלה לבין הדירוג, מה שיסייע לנו להסיר את טור זה מעיבוד המידע.

להלן הפעולות שבוצעו:

