

סטטיסטיקת שילובי מקצועות מאז ועד היום

כלי לסינון והצגה נוחה של מידע מספרי אודות שילובי מקצועות לפי חתך זמן בהתבסס על מאגר Wikidata

שחר אזולאי

רקע

לפני כעשור יצא לי להיחשף לחיים שפירא, ד"ר למתמטיקה, חוקר ומרצה בנושא תורת המשחקים שגם חיבר רב מכר בתחום זה ("שיחות על תורת המשחקים") ובכך הצליח לסחוף רבים להתעניין בתחום. הרצאותיו בנושא זכו אף הן לאהדה כולל הרצאה שנתן במסגרת TED ושפירא נחשב לשם דבר בענף מתמטי זה.

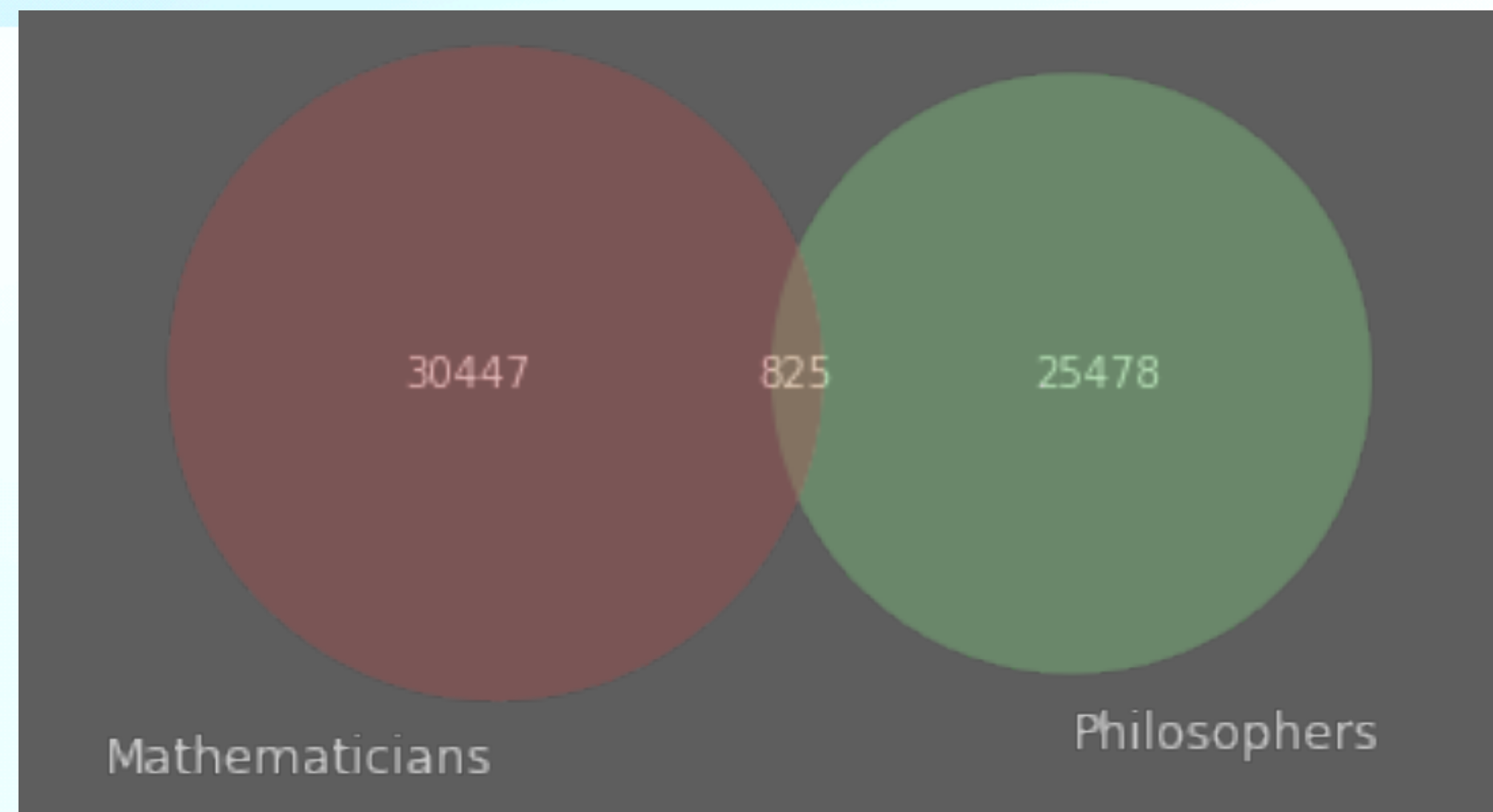
לפני כשנתיים זכיתי לקבל הזמנה פתוחה להרצאה של הדוקטור וכמובן שלא נדרשו שכנועים נוספים. רק שכותרת ההרצאה מאוד לא התאימה למה שציפיתי: "משמעות החיים על פי קהלת" היה שמה. מה למתמטיקאי שמגדיר עצמו כחילוני ולספר קהלת? התשובה היא שהקשר של שפירא לספר קהלת לא נובע מהיותו מתמטיקאי וגם לא מאמונתו הדתית, שפירא הוא גם פילוסוף והוא דואג לציין זאת גם בהרצאתו.

הופתעתי מהאמירה של הדוקטור למתמטיקה לגבי היותו פילוסוף. "מה הקשר?" חשבתי, אך במחשבה שנייה נזכרתי בכל אותם מתמטיקאים גדולים שלמדתי אודות עבודתם (אפילו בתיכון) וגם הוגדרו גם כפילוסופים. פיתגורס (המאה ה-6 לפנה"ס), תלמי (המאה ה-2), תאלס (המאה ה-7), הרמב"ם (המאה ה-13) ורנה דקארט (המאה ה-17) הם רק חלק מהשמות שהופכים את השילוב הזה ליותר טבעי לאוזן ורובם אם לא כולם אף הצליחו להגיע לתובונות בתחום אחד בזכות עבודתם בתחום השני, כך שאפשר לומר שלמתמטיקה ולפילוסופיה יש נקודת השקה ואף כמה.

אך בכל זאת משהו גרם לי להסתכל על שילוב המקצועות כצירוף לא טריויאלי במחשבה ראשונה וניסיתי להבין מהו אותו הדבר. מעט מחשבה על זה הביאה אותי לשתי סברות, הראשונה היא שהשילוב הזה פשוט נשחק לאורך ההיסטוריה והשנייה יותר כוללנית והיא שפילוסופיים בני ימינו כבר לא מקבל את אור הזרקורים.

רקע - המשך

השאלה שהניעה אותי לפרוייקט הזה היא - האם אכן ניתן לאמוד ירידה בקבוצת החיתוך שבין המקצועות לאורך השנים? על השאלה הזו אני אוכל לענות בתוך כמה דקות של עבודה על מאגר Wikidata אך תהיה לי תשובה על שאלה מאוד ספציפית ונראה שניתן להכליל את השאלה כך שיתקבל כלי שיעזור לענות על שאלות כדוגמתה ועל שאלות נוספות אודות סטטיסטיקת תחומי העיסוק לאורך ההיסטוריה האנושית. מציאת חתכים בין מקצועות נוספים, מציאת 2 התחומים שמשולבים הכי הרבה וויזואליזציה של היכחדות מקצועות כמו גם בריאתם הם חלק מהדברים שהכלי יאפשר לענות עליהם בלחיצת כפתור כמו גם על השאלה "לאן הפילוסופים הולכים כשהאגם קפוא?"



תכנית עבודה

”הכל יפה על הנייר”

1. ”איסוף כלל הנתונים ממאגר Wikidata. מדובר בקובץ השוקל כ-107GB”

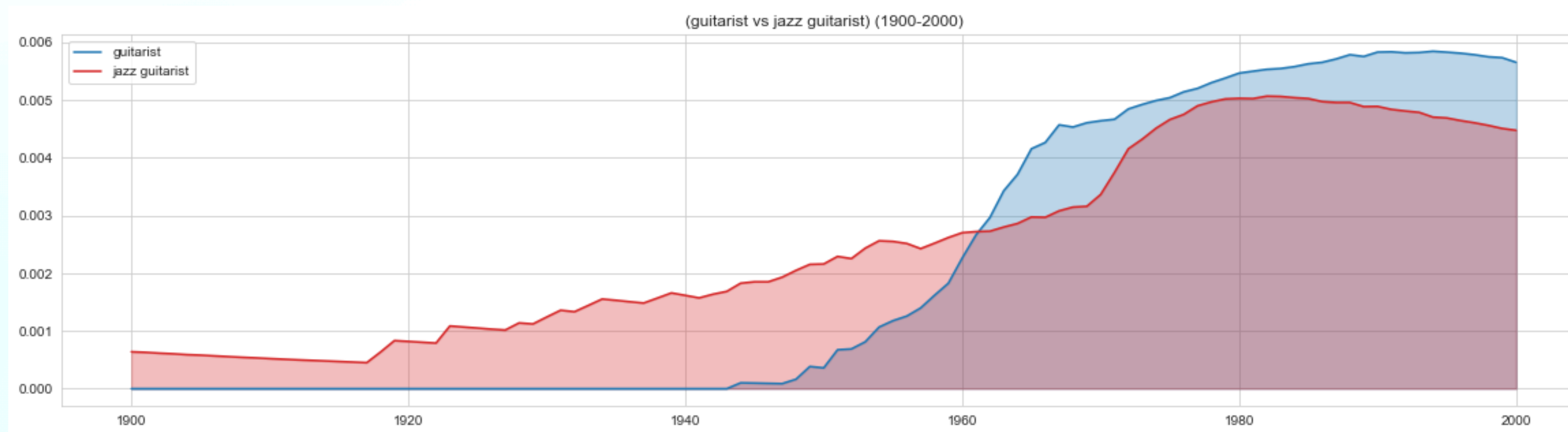
- בפועל מדובר בקובץ שמעט קשה לעבוד איתו. ה-API לא מאוד ידידותי וחסרים מקורות מידע. כמו כן כל שאלתא מצריכה מעבר על כל הקובץ כך שעל מחשב ביתי מדובר בזמן לא מבוטל.
- לבסוף החלטתי להשתמש בשאלתות SPARQL דרך חבילה בשם qwikidata בפייתון.
- גם השאלתות היו מעט מסורבלות שכן יש מגבלת זמן ריצה עבור כל שאלתא ולכן נאלצתי לחלק את השאלתות לחלקים.

2. ”חילוץ המידע הרלוונטי בלבד(קרי, כל העמודים הקשורים במקצועות/אישים)”

- בשל מגבלת הזמן על כל שאלתא יכולתי לקבל רק את רשימת כל המקצועות המתועדים ולא את כל רשימת האישים.
- לכן ביצעתי שאלתא עבור כל מקצוע ובה ביקשתי את רשימת האנשים העוסקים בכל מקצוע. מדובר בכ-3,300 שאלתות שונות וכמובן שהיה גם צורך לנקות כפילויות(אנשים העוסקים בכמה מקצועות)

3. ”ניקיון וקישור הנתונים”

- תחילה הצהרתי שארצה לנקות את כל מה שאינני מחשבי כמקצוע(כמו אנס ורוצח) אך לבסוף הבנתי שעל אף שאלו הם אינם מקצועות אולי כן ניתן להסיק מידע מעניין מהשארית המידע הזה.
- כמו כן, קיוויתי לאחד את כל אלו העוסקים במקצוע שניתן להכליל(לדוגמה כל הנגנים הם נגנים). בדוגמה מטה ניתן לראות את אוכ' הגיטריסטים וגיטריסטי הג'אז בין 1900-2000.
- כלל המקצועות עם מעט עוסקים נוקו מהמאגר ולאחר מכן בוצעו שאלתות נוספות על מנת לעדכן את האוכלוסייה הרלוונטית.



תכנית עבודה - המשך

”הכל יפה על הנייר”

4. ”בניית מבנה נתונים יעיל בשליפת המידע אותו הייתי רוצה להנגיש בעזרת הכלי”

- מבנה הנתונים העיקרי שאני משתמש בו הוא טבלה המכילה את כלל האנשים המתועדים ועבור כל אחד מהם את שנת לידתו, שנת מותו וכלל העיסוקים אליהם הוא משוייך
- מבנה נוסף הוא טבלת כל המקצועות ובעזרתה ניתן לבחור את המקצועות בהם מעוניינים וממנה גם נשלפות התוויות

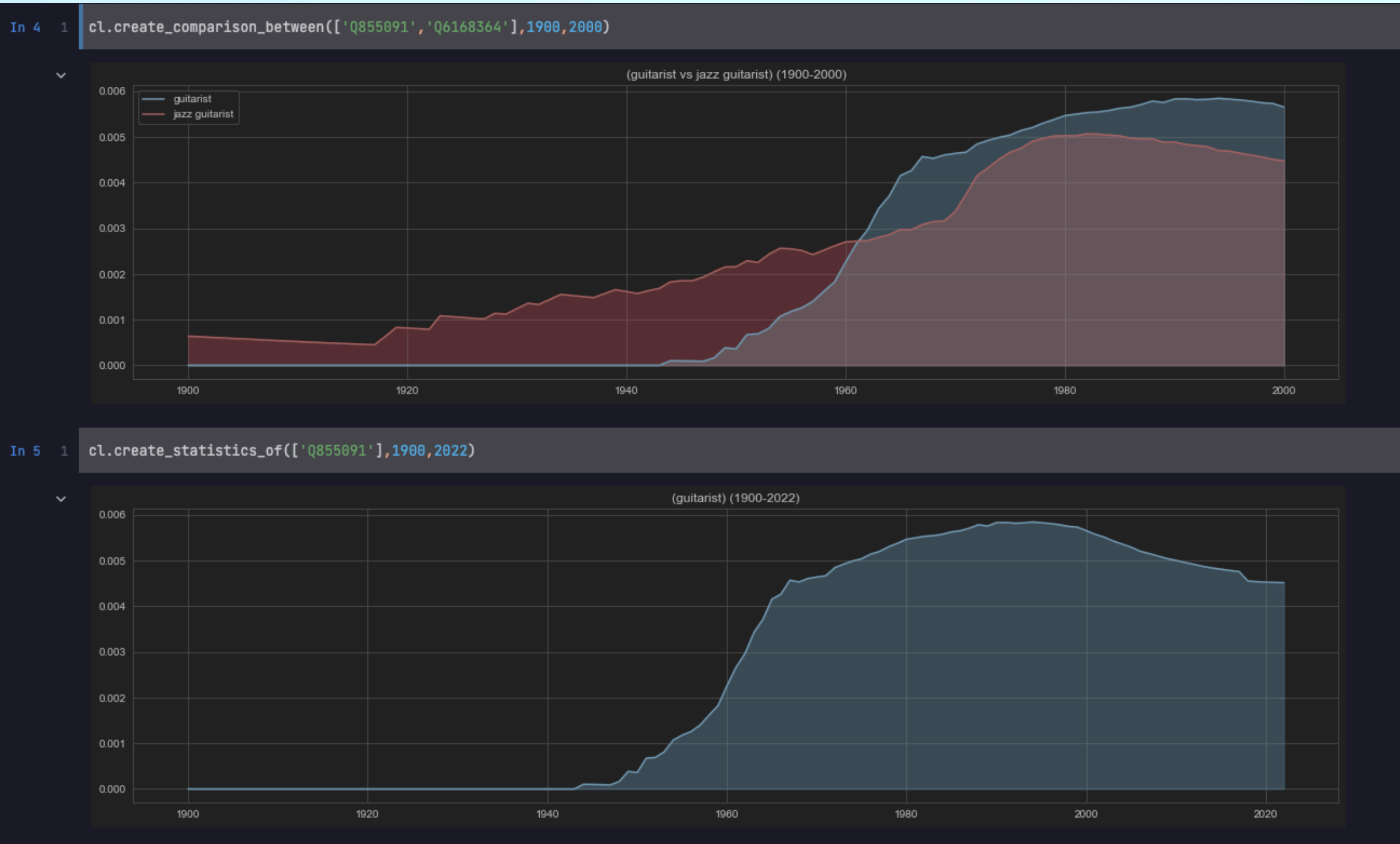
5. ”יצירת API נוח בצורת פונקציות לתשאול מאגר הנתונים שבניתי”

- גם כאן הצלחתי לעמוד ביעד ויצרתי מחלקה יחסית קומפקטית שעושה שימוש יעיל בקוד על מנת לשלוף ולייצא את המידע הרלוונטי

6. ”בניית UI לכלי והוספת ויזואליזציות מתאימות למידע שחוזר מהשאלות השונות”

- כאן הייתה הנפילה הקשה ביותר ולאחר מלחמות מרובות הבנתי שפחות נועדתי ליישומי צד-קדמי(front-end בשפה מדיצינה)

דוגמה להרצה בקובץ המחברת



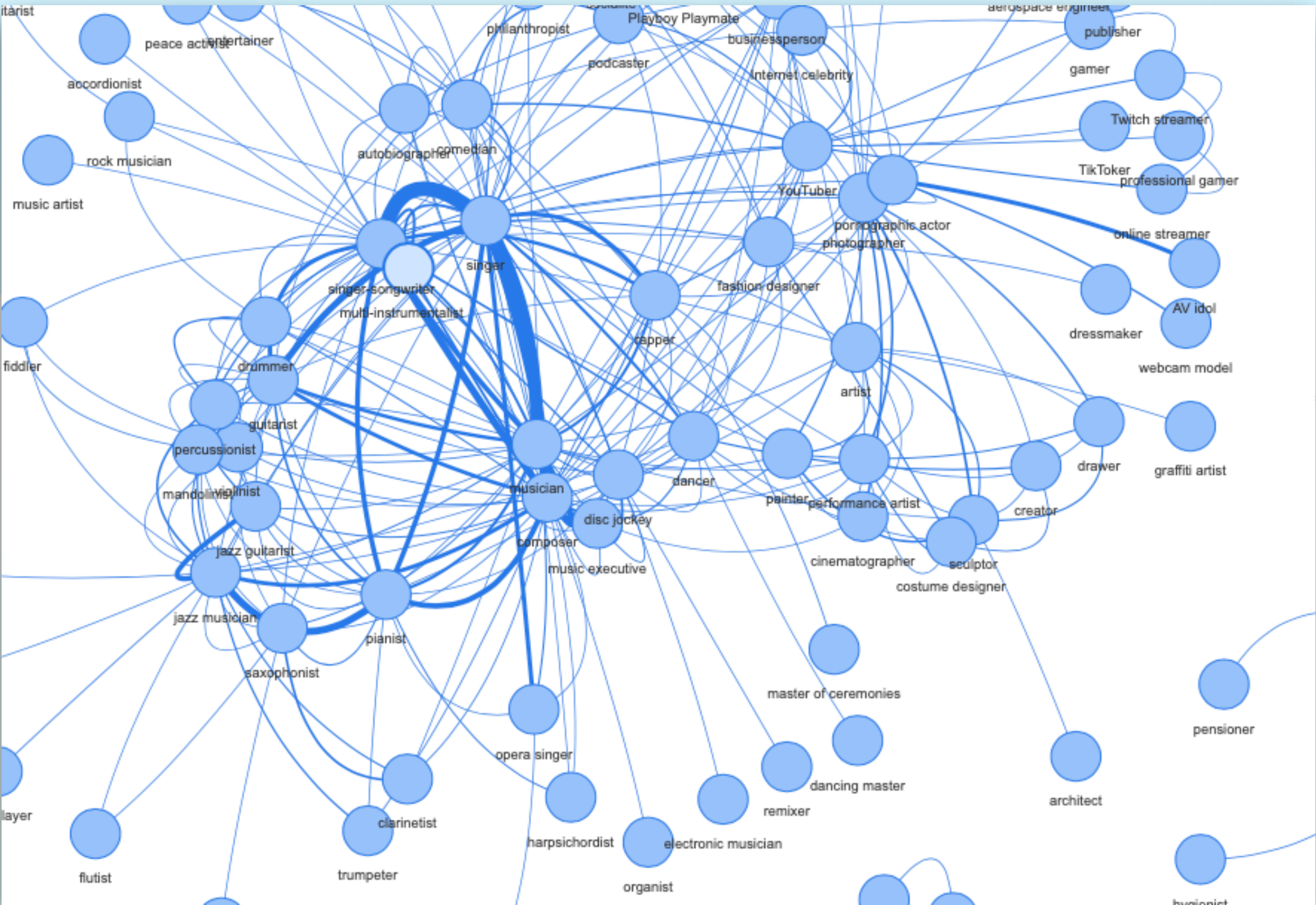
7. ”שיעל תביא לי 100”

- אמן

היעד

בסך הכל די עמדתי ביעדים. אמנם הוט זה משהו שמאוד רציתי אך בפועל זהו הדבר הכי פחות חשוב. כמו כן החלטתי להוסיף לא רק את היוזאליזציה של אוכלוסיית עוסקים בשילוב מסויים אלא גם השוואה בין מקצועות שונים.

בנוסף, ניתן לייצר גרף קודקודים בין מקצועות שונים(דוגמה למטה) כאשר כל קשת מחברת בין שני מקצועות שיש מישהו שעסק בשניהם. עובי הקשת נקבע ע"י כמות העוסקים העונים להגדרה.



#	תווית באנגלית	תווית בעברית	מס' עוסקים מתועדים
1	researcher	חוקר	1766428
2	politician	פוליטיקאי	738496
3	writer	סופר	316449
4	actor	שחקן	304050
5	painter	צייר	188848
6	journalist	עיתונאי	143731
7	singer	זמר	107108
8	composer	מלחין	96375
9	musician	מוזיקאי	86657
10	military	איש צבא	83831
11	physician	רופא	76816
12	businesspers	איש עסקים	74739
13	artist	אמן	71662
14	teacher	מורה	67993
15	photographer	צלם	64401

15 תחומי העיסוק המתועדים ביותר

גרף(חלקי) של יחסי מקצועות בין השנים 1900-2022 (מינימום 10 עוסקים בכל שילוב)