

**המחלקה להנדסת תעשייה וניהול**

**התמחות:** מערכות מידע

-עבודת מחקר -



**חקר גורמי ההצלחה בתוכנית 'מעגלי תעסוקה'  
של שירות התעסוקה הישראלי.**

מגיש: חן שליו 203205984

מנחה אקדמאי: ד"ר פיני דוידוב

תוכן עניינים

[**תקציר** 3](#_Toc35436915)

[**מבוא** 3](#_Toc35436916)

[רקע על הארגון 3](#_Toc35436917)

[רקע לפרויקט 4](#_Toc35436918)

[מטרת הפרויקט 4](#_Toc35436919)

[תיאור מבנה הדוח 4](#_Toc35436920)

[**סקר ספרות** 5](#_Toc35436921)

[מונחים 5](#_Toc35436922)

[אבטלה בעולם 5](#_Toc35436923)

[תוכניות מניעה לאבטלה בעולם 6](#_Toc35436924)

[תוכניות למניעת אבטלה בישראל 7](#_Toc35436925)

[**מטרות, מדדים, תסמינים** 8](#_Toc35436926)

[תסמינים 8](#_Toc35436927)

[מטרות 8](#_Toc35436928)

[**שיטה** 9](#_Toc35436929)

[**חקר מצב קיים וניתוח נתונים** 10](#_Toc35436930)

[הוצאת הנתונים 10](#_Toc35436931)

[טיוב הנתונים 10](#_Toc35436932)

[פילוחי הנתונים 13](#_Toc35436933)

[**מענה על שאלות המחקר** 15](#_Toc35436934)

[*שאלת מחקר מס' 1: האם יש שוני בין השמות באוכלוסייה הערבית לבין השמות בשאר האוכלוסיות?* 15](#_Toc35436935)

[*שאלת מחקר מס' 2: כמה זמן דורשי עבודה נמצאים בתוכנית טרם השמתם ?* 19](#_Toc35436936)

[*שאלת מחקר מס' 3: האם יש הבדל במשך הימצאות דורשי עבודה בתוכנית לבין סוגי הצלחות (labels) ?* 20](#_Toc35436937)

[*שאלת מחקר מס' 4 : האם יש קשר בין הפעילויות בתוכנית (סדנאות ו/או מפגשים) להשמה?* 22](#_Toc35436938)

[*שאלת מחקר מס' 5: האם יש השפעה בין המאפיינים הסוציואקונומיים של דורש עבודה לבין השמה?* 34](#_Toc35436939)

[**מסקנות המחקר** 37](#_Toc35436940)

[**תרשים גאנט** 38](#_Toc35436941)

[**ביבליוגרפיה** 39](#_Toc35436942)

[קישורים לאתרים חיצוניים 39](#_Toc35436943)

[מאמרים 39](#_Toc35436944)

# **תקציר**

בפרויקט זה חקרתי את תוכנית 'מעגלי תעסוקה' של שירות התעסוקה הישראלי האמון על ידי ממשלת ישראל להוות מקור לטיפול בדורשי עבודה (דורשי עבודה) למציאת עבודה ומתן כלים ע"י תוכניות שונות). כדי לבדוק את גורמי ההצלחה של התוכנית, בוצע סקר ספרות על אבטלה, תוכניות למניעת אבטלה בארץ ובעולם ממקורות רבים ושונים וזאת כדי לקבל מידע מקיף על אבטלה. לאחר סקר הספרות והבנת מושגי האבטלה בארץ ובעולם, נשאלו שאלות המחקר ועל פיהן הוצאו הנתונים ממסד הנתונים של התוכנית ובדיקת טיב הנתונים וטיפולם בשפות ובסביבות עבודה שונות: Python וגם R וכך בוצעו המודלים השונים לחקירת שאלות המחקר והצגת המסקנות.  
בוצע כתיבת דו"ח הפרויקט ובהמשך ייכתב מאמר מחקרי אשר אנסה לפרסם בפלטפורמה הרלוונטית.

# **מבוא**

## רקע על הארגון

שירות התעסוקה הישראלי הינו תאגיד סטטוטורי( ישות משפטית אוטונומית למחצה) הנתון לפיקוחו של שר העבודה והרווחה ולמועצת שירות התעסוקה. השירות מעניק שירותי השמה ותיווך עבודה לכ-400,000 דורשי עבודה (דורשי עבודה) מידי שנה, באמצעות 60 לשכות התעסוקה הפרוסות ברחבי הארץ. כיום יש בממוצע 60,000 דורשי עבודה בשירות התעסוקה עם סוג תביעה: משותף והבטחת הכנסה.

השירות מסיע למחפשי עבודה במציאת משרה, בהפניה להכשרה מקצועית, באיתור עובדים עבור מעסיקים ועוד. בנוסף, השירות מופקד על מחקר שוק התעסוקה ופרסום הממצאים מידי תקופה. כמו כן, השירות אחראי על העברת דיווחים שוטפים למוסד לביטוח לאומי בנוגע לזכאותם של דורשי העבודה לדמי אבטלה או גמלת הבטחת הכנסה. דורשי העבודה הינם מכלל קשת האוכלוסייה הרחבה של מדינת ישראל המקבלים בין היתר את סדנאות הארגון, פעילויות תומכות השמה, ייעוץ פסיכולוגי, ועדות רפואיות ותיווך משרות להשמה.

שירות התעסוקה כולל את כל הלשכות והמרכזים של מעגלי תעסוקה ובפרט את המטה הראשי. תפקיד המטה הינו לבקר, לעקוב ולמתן פתרונות מעשיים וניהוליים ללשכות בנושאים השונים בעבודה משותפת במחלקות השונות במטה, המחלקות הינם: המחלקה המשפטית, אגף משאבי אנוש, מחלקת המחקר ותכנון, מחלקת המעסיקים ושירות, לשכת מנכ"ל, מחלקת השמה, מחלקת תפעול ותכנון, מחלקת הדרכה ופיתוח ארגוני, מחלקת רווחה, מחלקת ביטחון, אגף בכיר טכנולוגיות דיגיטליות ומידע, אגף חשבות, אגף א.רכשים ולוגיסטיקה.

## רקע לפרויקט

כעובד לשעבר בשירות התעסוקה במטה הראשי בתפקיד: אנליסט בקרה לתוכנית 'מעגלי תעסוקה', שמתי לב לאופן ביצוע הבקרות הכוללות על דורשי עבודה בתוכנית. רוב מדדי הבקרות מקבלים 'הערכת מומחים' (מנהלים) ולא נעשים בכלים סטטיסטים או בכלי בקרה ייעודים מלבד כלי Excel.  
בנוסף כאשר רציתי לאמוד את מצב התוכנית בהשוואה למחקרים שנעשו בשנים קודמות כדי להגדיל את ההסתברות להצלחת התוכנית ע"י בחינת חלופות או בחינת גורמי ההצלחה של דורשי עבודה בתוכנית, גיליתי שאין מחקרים כאלו או דומים לה שנעשו בעבר וזאת כי התוכנית קיימת רק משנת 2014 ופעלה מ2016. בנוסף, ניגשתי לסמנכ"ל תכנון ולצוות מערכות מידע לקבלת אישור להכנסת כלי BI לארגון או כלי סטטיסטי כדי לבצע בקרות נוספות, אך נתקלתי בדחייה חוזרת של הבקשות (נכון לשנת 2020 שירות התעסוקה יכניסו SAS).

המטרה העיקרית של התוכנית היא מתן כלים יעודיים והכשרה מקצועית לדורשי עבודה והקטנת מספר דורשי עבודה ע"י הסבתם חזרה לשוק העבודה (השמה), הגדלת מספר המושמים והקטנת מספר מקרי דלת מסתובבת ו/או נגרעים והשהיות של דורשי עבודה.

## מטרת הפרויקט

בפרויקט הגמר ארצה לחקור את המאפיינים של דורשי עבודה שהיו בתוכנית טרם השמתם, למצוא תוצאות ומסקנות בעזרת ביצוע ניתוח הנתונים בצורה סטטיסטית ומחקרית של תוכנית מעגלי תעסוקה בשירות התעסוקה הישראלי בעזרת כלים סטטיסטים ייעודים לביצוע פרויקט מחקרי מסוג זה. כיום בתוכנית מעגלי תעסוקה ניתנים מגוון רחב של כלים וסדנאות המשפיעים ועוזרים לדורשי עבודה למצוא עבודה בשוק העבודה, אך אין מסלול אחיד/מוגדר מראש לדורש עבודה בו הוא עובר, לכן ישנם מסלולים שונים וסיבות שונות לכך שדורשי עבודה הצליח למצוא ולהיכנס לעבודה (השמה עצמית או השמה בתמיכת לשכה). על כל אלו ארצה לכתוב מאמר אקדמאי בנושא ולנסות לפרסמו.

## תיאור מבנה הדוח

דוח הפרויקט כולל בתוכו שישה פרקים. ראשית יוצג פרק המבוא בו תהיה הקדמה, מידע כללי ומונחים להמשך. בפרק השני ארחיב בנושא סקר הספרות שמהווה בסיס איתן ובו מסקנות ממאמרים שונים, מונחים מקצועיים. פרק זה יתמוך וייתמך בכל חלק בפרויקט.  
הפרק השלישי הוא המדדים והיעדים לפרויקט ובו יפורטו מטרות הפרויקט, שאלות המחקר, המדדים והמוטיבציה לביצוע הפרויקט. הפרק הרביעי, פרק השיטה יוצגו המתודולוגיה והכלים שייעשה בהם שימוש לאורך שלבי הפרויקט השונים. בפרק החמישי, פרק חקר המצב הקיים הכולל איסוף נתונים וניתוחם בצורה מעמיקה ומענה על שאלות המחקר ותוצאות המודלים.  
הפרק השישי הוא פרק המסקנות אשר יכלול את מסקנות המחקר לאחר ניתוח המודלים ובדיקת התוצאות. יש לציין כי סקר הספרות נעשית לכל אורך הפרויקט.

# **סקר ספרות**

## מונחים

**אבטלה**: מצב שבו אדם המסוגל לעבוד ורוצה בכך אינו מצליח למצוא עבודה תמורת שכר ‎[10].  
נהוג להבדיל בין מובטל לבין בלתי מועסק:  
**מובטל**: בעל כושר עבודה המעוניין לעבוד ואינו מצליח למצוא עבודה תמורת שכר. **בלתי מועסק**: בעל כושר עבודה שאינו עובד ואינו מעוניין בעבודה, ומבקש עבודה, המחפש עבודה באופן פעיל.

## אבטלה בעולם

כיום האבטלה הינה בעיה כלל עולמית ובאיחוד האירופי בפרט. במאמרו של Danuše נאמר: "אין שום ספק שאבטלה מציגה את הבעיה החברתית הגדולה ביותר באיחוד האירופי" ‎[1]. מתוך בדיקת המצב העולמי (נכון לתאריך 27/11/2019) מדינת ישראל ממוקמת במקום ה10 מתוך 36 מדינות הOECD עם 3.7% אבטלה במדד הOECD לאבטלה ‎(1) , בנוסף ישראל נמצאת מתחת לממוצע של מדד זה (5.2%). כפי שנאמר במאמר של Sol:"מסתבר שבמדינות האיחוד האירופי באופן כללי רבע מהמובטלים הסובלים מבעיות כלכליות סובלים גם מבעיות בריאות, התמכרות לאלכוהול ו/או אפליה" ‎[2]. Zwinkels מגלה שהסיכוי של מובטלים לחזור לעבודה פוחת משמעותית ברגע שהוחרמה הקצבה שלהם. עבור מובטלים הסיכוי לחזור לעבודה פוחת ב35% ולמוגבלים ב12%. Zwinkels מסיק כי מובטלים עם חובות בעייתיים אכן מתקשים לחזור לעבודה יותר מאשר מובטלים ללא חובות (Zwinkels 2015; Zwinkels&Guiaux 2015) ‎[3], ‎[4]. מחקרים בקרב אנשים באמסטרדם בנושא סיוע סוציאלי מראים כי מסלולי הקלה בחובות בהקשר של שירותי תעסוקה אכן מביאים ליותר השמות, וייתכן שההשמה ללא מסלולי ההקלה מוגבלת יחסית (Koning 2014) ‎[5]. האבטלה נחשבת לאחד מגורמי הסיכון העיקריים להדרה חברתית. הרחקה מהעבודה עשויה להביא להתנכרות מהחברה, ובתורו להגדיל את הסיכונים לתלות ארוכת טווח ברווחה חברתית, התאבדות, הפיכה לפושעת או קורבן לפשע ותמיכה פוליטית במפלגות קיצוניות. יתרה מזאת, הדרה חברתית עשויה לא רק להשפיע על הדור הנוכחי, אלא שהיא תועבר לדור הבא (Polhan 2019) ‎[18].פרויקט מחקר בקרב ספקי שירותי תעסוקה באירופה מצביע על כך שברגע שחובו של מובטל משתפר, הסיכוי שלו לעבוד נוטה לגדול, זאת אומרת שליטה בחובות עבור מובטלים הנזקקים לעזרה סוציאלית-מדינית יובילו להגדלת הסיכוי להחזיר את המובטל לשוק העבודה (Sol & Kok 2014, p.144.p.168) ‎[6]. במחקרו של Pohlan הראה שישנו הקשר גם בין מאפיינים סוציואקונומיים לזמן האבטלה והשפעתה לאורך זמן ‎[16]. במאמרם של Artazcoz, Benach, Borrell and Cortès נמצא שיש קשר הדוק בין אבטלה לבין מצב הבריאותי והנפשי של האדם וישנם הבדלים בין המינים בהשפעות הקשורות לאחריות משפחתית ולמעמד חברתי ‎[17].

ישנם תופעות של אבטלה שונות: אבטלה מחזורית, אבטלה סמויה, אבטלה טבעית, אבטלה מבנית, אבטלה טכנולוגית, אבטלה עונתית וכו'. האבטלה המחזורית הזכורה ביותר הייתה במשבר הכלכלי העולמי בשנת 2008 אשר זכור בשמו "המיתון הגדול" אשר החל ביולי 2007, והתפשט אחר כך לכל רחבי העולם. מצב זה גרם לאבטלה מחזורית ענקית אשר השפיע על כלל האנשים ברחבי העולם מבחינת עסקים ואבטלה בפרט. בארצות הברית גם במהלך השנים 1921–1929 נקלעה למצב "השפל הגדול". בימים אלו ניצבים אנשים אל מול סוג נוסף של אבטלה: אבטלה טכנולוגית. בשנים האחרונות יותר ויותר משרות נתפסות על ידי רובוטים, מכונות אשר עושות את מלאכת היד בצורה טובה יותר ומדויקת יותר במגוון מקצועות ומקומות עבודה.

## תוכניות מניעה לאבטלה בעולם

יותר ויותר מדינות פועלות בתוכניות שונות כדי להילחם בתופעה וביניהם הקמת מקומות תעסוקה, תוכניות העשרה וקידום תעסוקתי, תוכניות הסבת מקצוע, ארגונים העוזרים בתחום, דמי אבטלה ועוד. מתוך מאמר של [Winkelmann](https://wol.iza.org/authors/rainer-winkelmann) ‎[7], הוכח כי אבטלה מקומית גבוהה יותר מחלישה את נורמת (מוסר) העבודה, לכן באזורים בהן יש פשיעה / אבטלה / חוסר רצון תעסוקתי גבוה ביחס למקומות אחרים תהיה השפעה גדולה יותר לאבטלה.

בפינלנד חישבו את ההסתברויות לאבטלה ארוכת טווח (אבטלה > 12 חודשים) לבני 25-28 וגילו כי 4.5% מהנשים ו -7.1% מהגברים חוו או יחוו אבטלה ארוכת טווח (מתוך 46,521) ‎[15].

באירופה הוקם ארגון שירות התעסוקה האירופאי (EURES- Europe Employment Service) ‎(2) בשנת 1994 והארגון פועל בשיתוף פעולה עם כלל מדינות אירופה במציאת עבודה לתושבי אירופה בגילאים 18-35, בה דורשי עבודה יכולים לחפש עבודה גם במדינתם וגם במדינה אחרת באירופה כולה וכל זאת ללא תשלום. השירות מאפשר ראיונות עבודה דרך שיחת וידאו, עזרה במעבר למדינה שאליה התקבל דורש העבודה (מגורים וטיסה),לימודי שפה, המרת תעודות רלוונטיות ועוד.  
לפי האתר של EURES (נכון לתאריך 12.12.2019) ישנם 2,191,759 משרות וכ3,658,553 דורשי עבודה אשר מגישים משרות, לפי מספרים אלה אפשר להשים ~60% כדרושי עבודה אך עדין ישאיר ~40% ללא עבודה גם בארץ אחרת. כמובן שלא כל דורשי העבודה יושמו בפועל וגם מספרים אלו מבוססים רק על דורשי העבודה הרשומים למערכת, למשרות הרשומות בה. לכן אתייחס לנתון של מדד הOECD העומד על 6.3% באיחוד האירופי אבטלה.

בשוודיה מדד האבטלה עומד על 6.5% וביטוח האבטלה מורכב משני מרכיבים: ביטוח בסיס כללי (סכום הבסיס) ואובדן ביטוח הכנסה (גמלה הקשורה בהכנסה). אף אחד משני המרכיבים בביטוח האבטלה אינו נבחן באמצעים. סכום הבסיס זמין לאדם שאינו חבר בקופת האבטלה או קשור אליו, או שלא היה חבר במשך מספיק זמן. משמע רק אדם אשר משלם חודש בחודשו למען ביטוח האבטלה יהיה זכאי לקבל קצבה תקופה ארוכה לעמוד בתנאי החברות, שכרוך בכך שהיה חבר בקרן או היה קשור לפחות שנה ללא הפרעה. עם זאת, המבקש נדרש למלא את התנאי הבסיסי ואת תנאי העבודה לקבלת גמלה בהתאם לביטוח הבסיס (Gabriella) ‎[8].

בסין מדד האבטלה עומד על 3.61% (2019) וכבר בשנת 1990 ממשלת סין החלה את מאמציה בהכנסת מדיניות אכיפה וסטנדרטיזציה בשוק העבודה אשר מחזקת ועוזרת לעסקים בחיפוש  
עובדים, בהשמתם ובהתאמתם. סין פיתחה תשתית חזקה של מערכת מידע (The Hùkǒu System שיטת מרשם אוכלוסין המכילה כלל המידע על האדם) ובה דורשי עבודה יכולים להיעזר ולהעריך את התאמתם לתפקידים אליהם רוצים להתמיין. בשנת 1999 ממשלת סין קראה לכל המגזרים החברתיים לאמץ את המערכת ולשים לב לתעודות בית הספר ולתעודות ההסמכה הנמצאות בה וזאת כדי להגדיל יכולת וסיכויי ההעסקה.

נקודת המוצא של מובטל ביפן (מדד האבטלה לפי OECD הוא 2.3%) היא שאין לו זכאות להטבת אבטלה או שמיצה את זכאותו, יש תמריץ חזק מאוד למצוא עבודה. במצב ספציפי זה, שירות התעסוקה היפני אינו צריך להקדיש מאמץ גדול להפעלת תוכניות השמה מכיוון שלמובטלים יש תמריצי עבודה חזקים. כתוצאה מכך מאמץ תוכניות אלו ביפן, כפי שנמדד במאמץ ההוצאה הציבורית על מדיניות השמה, הוא קטן יחסית. בשנת 2011, יפן הוציאה מתחת ל0.3% מGDP (תוצר מקומי גולמי) על מדיניות השמה, פחות ממחצית מממוצע ה- OECD (Martin) ‎[9].

## תוכניות למניעת אבטלה בישראל

בישראל, כמו בשאר העולם, הקימו ארגון ממשלתי אשר ישרת את הציבור בשם "שירות התעסוקה" הפועלת בממשק ישיר עם מול ביטוח לאומי. בעקבות הקמת השירות הממשלתי, החלו בחיפושים אחר תוכנית אב שתהווה אבן דרך התחלתית, אכן כך קרה. בשנת 2004 הקימו תוכנית בשם מהל"ב (מהבטחת הכנסה לתעסוקה בטוחה) אשר פעלה עד שנת 2007 כפי שקמה בארה"ב (שמה היה "חוק האחריות האישית והזדמנויות תעסוקה" או בשמה המוכר "תוכנית וויסקונסין"). מטרת התוכנית הייתה להעביר חלק מהאנשים החיים בעוני מהישענות על קצבאות הביטוח הלאומי אל שוק העבודה. הייתה תוכנית וויסקונסין לאחת מתוכניות הדגל של ממשלות ישראל בכל הקשור להתמודדות עם עוני ומצוקה. יחד עם זאת, עוד לפני שהופעלה, בדיונים הארוכים אודות מאפייניה, היא הפכה יעד לביקורת חריפה מצד גורמים פוליטיים, ארגונים חברתיים וחלק מאוכלוסיית היעד. בעקבות הביקורות הרבות כנגד התוכנית, ישראל כיום פועלת בכמה רבדים על מנת למגר את תופעת האבטלה וביניהם בשירות התעסוקה קמו מספר תוכניות כגון: מעגלי תעסוקה, Age And Work, מפתח (חרדים), תנופה, אשת חיל ועוד. תוכניות אלו נותנות כלים להשמת דורשי עבודה בשוק העבודה. בתוכנית מעגלי תעסוקה תחת שירות התעסוקה אנו מקבלים דורשי עבודה (דורשי עבודה) אשר מגיעים מביטוח לאומי (עם סוג תביעה: משותף והשלמת הכנסה). כחלק מהתוכנית דורשי עבודה עוברים סדנאות, אינטייקים (פגישות אישיות) , קורסים ועוד כלים רבים. לשנת 2019 מדד ההשמה העצמית במעגלי תעסוקה (עדכני לתאריך 12.12.2019) עומד על 7.49% מכלל ההשמות בתוכנית (63.94% תמיכת לשכה, 28% לשכתית, מספר כולל של 12,223 השמות). אין הבדל בין הגדרת המדד תמיכת לשכה ולבין השמה עצמית, לכן במעגלי תעסוקה ניתן לומר כי אחוז ההשמה העצמית הינו 71.43%. אין לומר רק מהנתון לעצמו האם התוכנית אפקטיבית, יש לחקור כמובן את הסיבות למדד זה ולמי שנספר בו. תחילה אבדוק האם יש קורלציה ישירה בין מאפייני התוכנית להשמות או שלא.

בהחלטת הממשלה מס' 3790, 13.05.2018 הוחלט על "צמצום פערים חברתיים כלכליים ופיתוח כלכלי במזרח ירושלים" בהמשך להחלטה מס' 1775 והחלטה מס' 2684, זאת לאור הצורך בחיזור יכולת ההשתלבות של תושבי מזרח ירושלים בחברה ובכלכלה הישראלית וכן מתוך המטרה לחזק את חוסנה הכלכלי והחברתי של הבירה כולה ‎[13].

שיעור המועסקים באוכלוסייה הערבית הוא 50.5% לעומת 78% באוכלוסייה היהודית, וגם הכנסותיה של האוכלוסייה הערבית מעבודה קטנות במידה ניכרת מהכנסות האוכלוסייה היהודית. אחת התוצאות של מצב זה היא שהאוכלוסייה הערבית ענייה בהרבה מהאוכלוסייה היהודית ‎[14].

# **מטרות, מדדים, תסמינים**

## תסמינים

תוכנית מעגלי תעסוקה בשירות התעסוקה מכילה מגוון רחב מקשת האוכלוסייה. כיום מדידת ההצלחה של התוכנית נמדדת במספר פרמטרים מידי חודש: מספר מושמים, מספר פעילים בתוכנית, מספר מקרי דלת מסתובבת, מספר נגרעים, מספר השהיות, מספר בעלי פעילות בתקופה הנוכחית, מספר בעלי רצף פעילות של חצי שנה ויותר ועוד. כל המדדים הנ"ל אינם בעלי גבולות בקרה קבועים מראש, אלא מוערכים ע"י התבוננות כללית של מספר דורשי עבודה פעילים בחודש לעומת כמות המושמים לתקופה הנבדקת.

המדדים הקיימים כיום לא מהווים אינדיקציה קונקרטית לגבולות בקרה או הגדרה מדויקת לכמות גבוהה או נמוכה של דורשי עבודה בתוכנית, למספר השמות חודשי, למספר פעילויות ועוד. גם אם יצטרפו לתוכנית מספר גבוה של דורשי עבודה לעומת חודשים קודמים אין לכך התייחסות מיוחדת ולהפך.  
כעובד השירות שמתי לב לניהול וניתוח דורשי עבודה בתוכנית ולכן החלטתי לפתח מחקר בו אנתח את הנתונים בתוכנית ואגלה את הגורמים להצלחה באילו שכן הושמו דרך התוכנית, בין אם בעזרת השירות ובין אם בהשמה עצמאית. זהו אתגר גדול ומשמעותי בפן האישי (להצלחת הפרויקט) וגם בפן הלאומי-מדיני (חסכון כספי ומציאת גורמי שורש).

## מטרות

המטרות המרכזיות הן חקר גורמי ההצלחה של תוכנית 'מעגלי תעסוקה' וכתיבת מאמר אקדמאי. המטרות המשניות/נלוות הן: למידה מעמיקה על אבטלה ותוכניות למניעתה בעולם ובישראל בפרט.

שאלות המחקר:

1. האם יש שוני בין השמות באוכלוסייה הערבית לבין השמות בשאר האוכלוסיות ?
2. כמה זמן דורשי עבודה נמצאים בתוכנית טרם השמתם ?
3. האם יש הבדל במשך שהייתם דורשי עבודה בתוכנית לבין סוגי הצלחות (labels) ?
4. האם יש קשר בין הפעילויות בתוכנית (סדנאות ו/או מפגשים) להשמה ?
5. האם יש השפעה בין המאפיינים הסוציואקונומיים של דורש עבודה לבין השמה?

בכדי לנתח את המצב הקיים של התוכנית, איעזר בניתוחים שנעשו בעבר על דורשי עבודה בתוכנית, במחקרים שפורסמו בארץ ובעולם על אבטלה, רכישת קורסים לניתוח מידע בעזרת R ובעזרת Python.

# **שיטה**

בפרק זה אפרט על כלים שונים אותם למדתי במהלך התואר ובכלים לניתוח מידע בפרט.  
במהלך ביצוע חקר מצב קיים הכלים הנתונים לשימוש כיום בשירות, הבחנתי כי יש חוסר בכלים בהם ניתן לנתח את המידע הקיים בארגון ובתוכנית בפרט.

בתחילת ביצוע הפרויקט, חשבתי על שאלות המחקר ולפיהם הוצאתי שלד גולמי של המידע הקיים בתוכנית 'מעגלי תעסוקה', טייבתי את הנתונים, חקר השאלות ופיתוח המודלים נעשו בשפת R בסביבת העבודה R-studio וגם בשפת Python בעזרת Jupyter Notebook. כל אלו יובילו לעשות הערכה והשוואה בין הממצאים לבין המציאות ולראות האם התוכנית אכן מביאה לתוצאות הצפויות, הסקת מסקנות ומתן המלצות להמשך.

הקורסים שעברתי היוו בסיס איתן להתחלת הפרויקט, אך בכדי לבצע את כלל הניתוחים הסטטיסטים, הייתי צריך להרחיב את הידע שלי בכתיבת קוד בשפת R ובשפת Python או השימוש בSPSS מעבר למה שלמדתי בקורסים: "כריית ידע ולמידת מכונה", "טכנולוגיות מתקדמות באינטרנט" בהנחיית ד"ר רוני הורוביץ או בקורסים "סטטיסטיקה יישומית" , "אופטימיזציה הנדסית", "הסתברות" שהיו בהנחיית ד"ר בנימין ברסקי. הרחבתי את הידע שלי בכתיבת קוד בשפת R ובשפת Python ע"י קניית קורסים בUdemy וצפיה בסרטונים ומדריכים ברשת לפיתוח אלגוריתמים וכיצד לבצע ניתוחי מידע שונים.

שאלות מחקר

איסוף מידע

הכנה וטיוב נתונים

ניתוח המודל

הערכת תוצאות והצגת מסקנות

# **חקר מצב קיים וניתוח נתונים**

## הוצאת הנתונים

כיום לארגון מסד נתונים עם המון טבלאות המנוהל דרך Microsoft SQL Server ולו המון טבלאות לא משומשות וספר מפתחות ישנן ולא מעודכן. מי שמתעסק ומכיר את מסד הנתונים בצורה הכי טובה הם נציגה מטעם Malam אשר נמצאת במשרד.

כדי למשוך את כלל הנתונים בשאילתת SQL זה די מסובך, לכן העדפתי להוציא זאת דרך מחולל הדוחות ולאחר מכן לחבר את כלל הקבצים בעזרת VLOOKUP לתאים ולעמודות בקובץ Excel אחד המכיל כ80 עמודות וכ56,000 שורות (כל שורה היא דורשי עבודה וכל עמודה היא תכונה של הדורשי עבודה).

## טיוב הנתונים

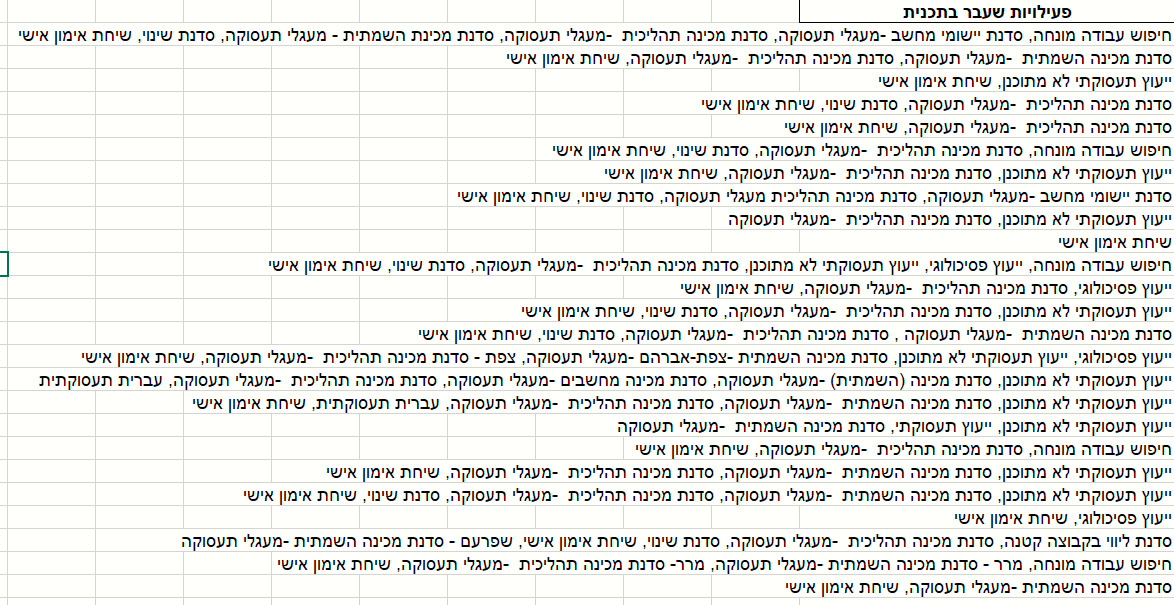
לאחר יצירת עותק לקובץ התחלתי לבצע בדיקת טיב הנתונים שהוצאו בעזרת Python ובשפת R, יצרתי גרף בשפת R המראה את טיב הנתונים וקיום הנתונים בתאים.

**איור א.1 : בדיקת טיב וקיום הנתונים**

(א.1 באיור זה ניתן לראות את הdataframe צבוע בשחור וצהוב, צהוב מהווה 14% מהנתונים ומבשר על תאים ריקים לאותם תאים, שחור מהווה 86% מהנתונים ומראה שקיים נתון באותם תאים).

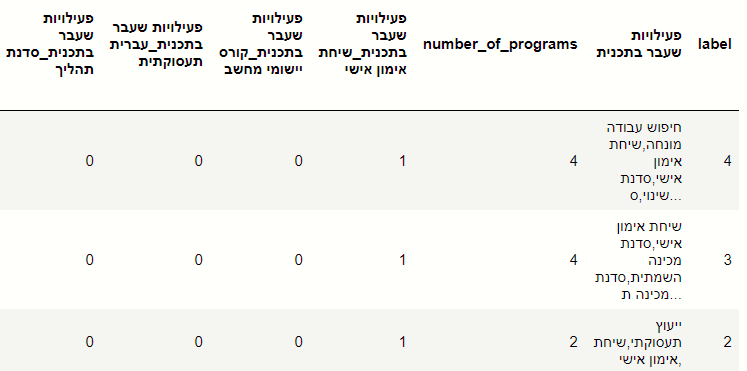
להפתעתי גיליתי כי 14% מהם פסולים (ריקים, חסרים או לא נכונים) ורק 86% מהם תקינים.

**איור א.2 : עמודת פעילויות שעבר בתוכנית**



בנוסף היה עלי לשנות את עמודת "פעילויות שעבר בתוכנית" עקב כך כשפותחים פעילות חדשה ניתן לתת מלל חופשי לשם התוכנית ולא לבחור את שמה מרשימה מוגדרת מראש.  
לכן נוצרו כ1860 ערכים ייחודים והיה צורך לסננן, לקטלג אותן וליצור רשימה של פעילויות קטנה יותר ומוסכמת מראש. לאחר מניפולציית הנתונים הצלחתי להגיע ל30 ייחודיים בלבד.  
לאחר פעולות אלה, היה צורך לפרוס את כלל הפעילויות כדי לסמן האם הדורשי עבודה אכן היה באותה פעילות, במידה ודורש העבודה היה בפעילות יקבל את הערך 1 ובמידה ולא קיבל 0.

**איור 1.ג : סימון פעילות לכל דורש עבודה**



היה צורך בהגדרת משתנה המטרה שבעזרתו אוכל לחקור ולנתח את שאלות המחקר בעזרת הנתונים ומעצם העובדה שהמושג "הצלחה" הינו נתון לפירוש שונה בכל סיטואציה, החלטתי להגדיר אותה בעזרת 4 קטגוריות שחולקו ע"י תנאים מתוך הכרת המערכת, היגיון בריא והכרת הנתונים:  
1) הצלחה מסחררת- ללא מקרי דלת מסתובבת, השמה מכניסה לתוכנית=1, ללא תאריך חידוש פעילות.  
2) הצלחה בינונית- ללא מקרי דלת מסתובבת, השמה מכניסה לתוכנית > 1, ללא תאריך חידוש פעילות או ללא תאריך חידוש רישום.  
3) הצלחה חלשה- כל דורשי עבודה שאינם נמצאים בקטגוריות 1,2,4.  
4) אי הצלחה / כישלון- ללא תאריך גריעה, ללא השמות מכניסה לתוכנית או מספר מקרי דלת מסתובבת > 0.

כל קטגוריה נקראת label ולאחר החלוקה של דורשי עבודה בדקתי שסה"כ כמות מכל label אכן שווה לסה"כ דורשי עבודה מהdataframe.

'עומק אבטלה בחודשים' שכל ערכיה הם integers מ0-100 חולקו ב10 עם עיגול התוצאה כדי לקבל עמודה "מקוטלגת".

עמודת 'גיל' קוטלגה לפי מה שנהוג בשירות התעסוקה, קבוצות של: 18-29, 30-39, 40-49, 50-54, 55+.

עמודת 'השכלה' גם לה יש המון ערכים כי ניתן להכניס בה טקסט חופשי, לכן בוצעה חלוקה לקטגוריות: יסודי, תיכון, תואר, תעודת מקצוע וללא השכלה.

עמודת 'דת' גם היא חולקה לדתות העיקריות: יהודי, נוצרי, דרוזי, מוסלמי, אחר. כמובן שכל אחת מהם מכילה כמה סוגים ולכן הוחלט לקבצם. מקרי קצה לא ישפיעו על הנתונים לכן קובצו ל'אחר'.

בעמודת 'ארץ לידה' ישנם ערכים מוזרים השייכים לישראל אך מופיעים בתור ערך שונה ממנה: יהודה ושומרון, רמת הגולן, עזה וסיני. לפי היגיון בריא החלטתי לשייכם לערך 'ישראל'.

עמודת 'אחוזי נכות' חולקו לקטגוריות: [0,1) , [1,20) , [20, 40) , [40,60) ,[60,100].

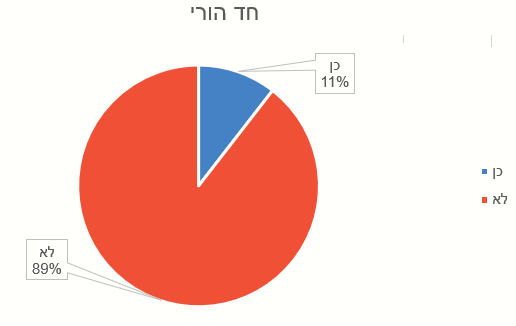
בעמודת 'שפה' יש אפשרויות רבות כי לכל שפה יש את הרמה שלה ורמתן מתחילה מ'בסיסית'. לכן החלטתי שאפשר להוריד את כל התוספות ולהשאיר רק את שמה של השפה שאותה מדבר הדורשי עבודה.

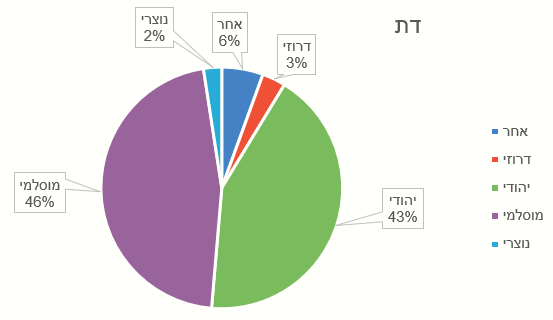
עמודת 'רישיונות' עברה המרה מתת חלוקה של הקטגוריות לקטגוריות עצמן.

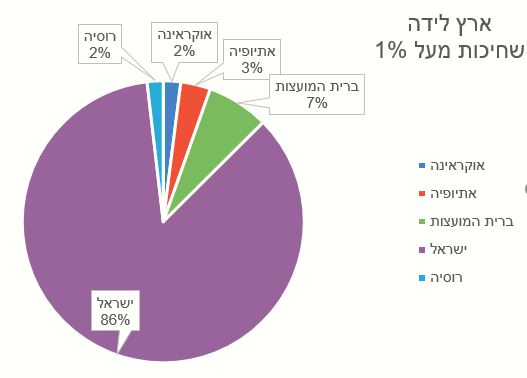
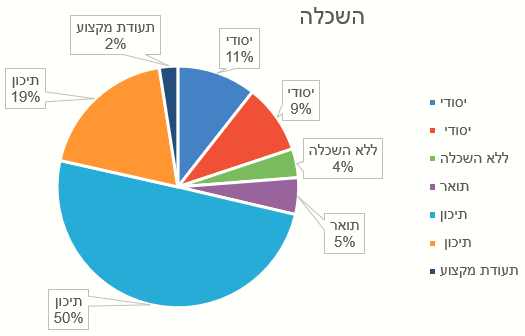
ערכים בעמודת 'מצב משפחתי' היו שנויים במחלוקת (ערכים כמו: פנוי\ה, פוליגם, בברית, מרשם) לכן הומרו לעיקריות: רווק, נשוי, לא ידוע. אליהם כמובן שיש גם את: אלמן, גרוש.

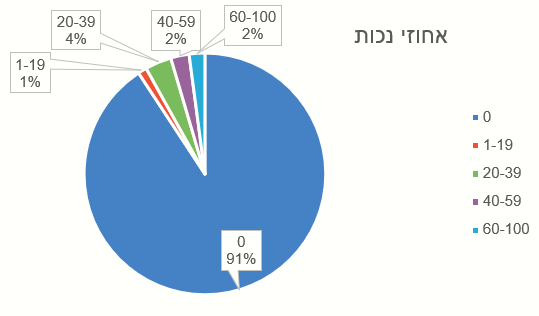
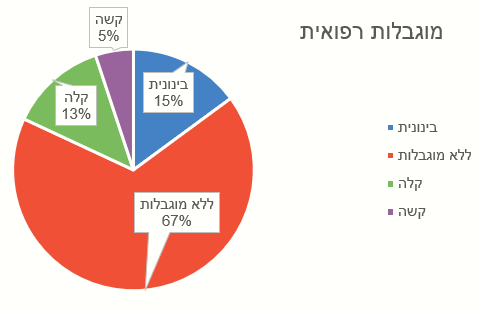
'מספר ילדים עד גיל 18' עברו קטלוג של: 0-7 , 8+.

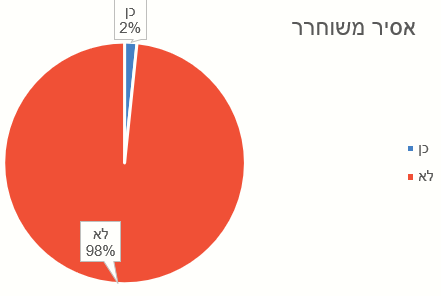
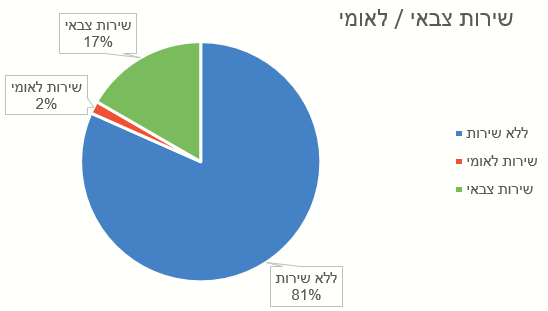
## פילוחי הנתונים

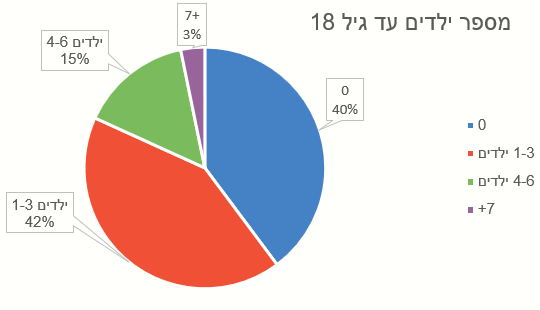


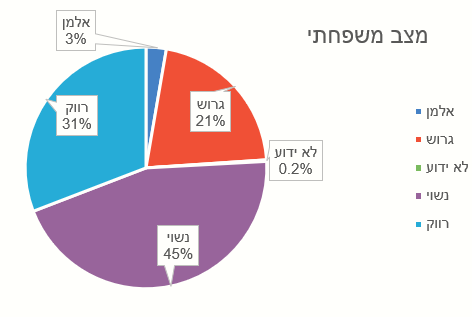












# **מענה על שאלות המחקר**

### שאלת מחקר מס' 1: *האם יש שוני בין השמות באוכלוסייה הערבית לבין השמות בשאר האוכלוסיות?*

שאלה זו נבעה גם בשירות התעסוקה ובממשלת ישראל עקב רצון להפעלת שוויון זכויות.  
לכן הוטל על שירות התעסוקה להקים סדנאות ומערכים ייעודיים לאוכלוסייה הערבית ודוברי ערבית מטעם הממשלה עם תכנים המותאמים להם, פיתוח מערכים חדשים (תפוח) ועוד.  
השירות מסתכל על המספרים של דורשי עבודה בחברה הערבית - מספר גבוה של דורשי עבודה ערבים שאינם מושמים דרך התוכנית. לדבר זה סיבות רבות כמו: הבדלי תרבויות, מסורת, סביבה גיאוגרפית, מעמד חברתי ועוד. הייני סבור שאין הבדל בהשמתם.

(ההשערה שתוחלת ההפרש=0).

(ההשערה שתוחלת ההפרש ≠ 0)*.*

הוספתי לdataframe עמודה חדשה בשם is\_arab שבה מי שערבי יסומן 1 ומי שלא 0. הסימון לערבי יתקבל אם הדורשי עבודה דובר שפה הערבית וגם שדתו לא יהודית. (\*הערה - ישנם לא ערבים (שאינם יהודים) שדוברים ערבית אבל שיעורם באוכלוסייה הנו נמוך ולכן לא ישפיע רבות על התוצאה).

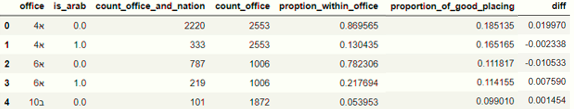
בנוסף ניתן גם סימון למי שהוא בהשמה טובה מאוד או השמה בינונית (label = 1 or label =2) בעמודה נוספת בשם is\_good\_placing. מי שעונה על קריטריון זה יקבל is\_good\_placing =1 אחרת 0.

**איור 3.א : סימון 'השמות טובות' וסימון 'האם הערבי' לכל דורש עבודה**

לאחר זה בוצעה חלוקה כדי לדעת את הפרופורציה של דורשי עבודה ערבים ולא ערבים מכל לשכה ומהחלוקה is\_arab. ההבדלים מיוצגים בעמודת diff כדי לאמוד את ההבדלים בהשמות באותה לשכה.

diff = [(is\_arab=0). propotion\_of\_good\_placing] – [(is\_arab=1). propotion\_of\_good\_placing]

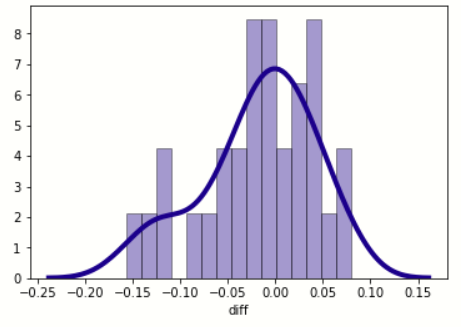
ארצה לראות האם הנתונים מתפלגים נורמלית לכן אני רוצה להגיע למצב בו יהיו לפחות 30 לשכות בהן מספר הערבים יהווה ביסוס סטטיסטי להשערתי. לכן לאחר ביצוע ניסוי וטעיה, ראיתי שצריך לפחות 14 דורשי עבודה ערבים בלשכה כדי להגיע למצב בו יש 30 לשכות יחודיות.  
לאחר הגעה לאותם 30 לשכות בעצם ערכתי השוואה בין הפרשי הזמן בין הערבים בלשכה ללא ערבים ואת שיעורי הפרופורציה בהשמה הטובה ופרופורציית דורשי עבודה ערבים בלשכה.

**איור 3.ב : חלוקת פרופורציית 'השמות טובות' והפרשן באוכלוסייה הנדגמת**

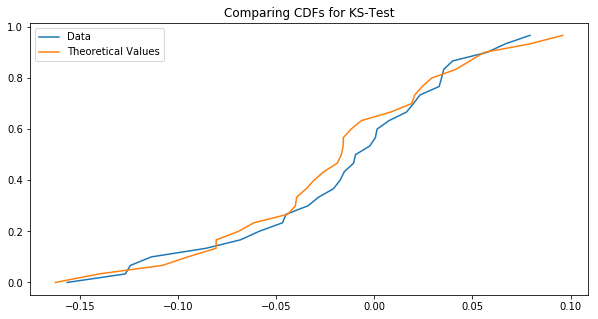
לדוגמה: לשכה א4 הערבים מהווים 333 אנשים שהם 13.04% מתוך דורשי עבודה השייכים ללשכה והשמתם היא 16.51%. לשכה א4 הלא ערבים מהווים 86.95% ולהם יש 18.51% השמה. ההפרשים בהשמות הוא 1.99%.

כל מה שנותר הוא לבדוק האם ההפרשים מתפלגים בצורה נורמלית, בשביל כך אשתמש בהיסטוגרמה אשר תראה את כמות הדגימות ביחס להפרש ונעביר פונקציית צפיפות.

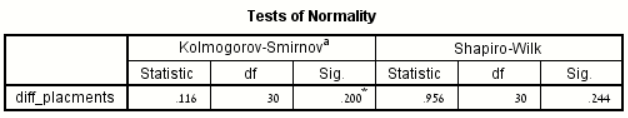
**איור 3.ג : היסטוגרמה של עמודת 'diff' ופונקציית צפיפות**

(בציר X מוצג ההפרשים ובציר הY מוצג כמות הדגימות) אומד לשונות הוא 0.0035 .  
כפי שניתן לראות פונקציית הצפיפות של הנתונים (קו כחול) אכן מרמזת שהנתונים מתפלגים נורמלית (כמובן שלא די בכך ויש לבצע מבחן סטטיסטי שיבוצע בהמשך).

צירפתי את עמודת 'diff' לגרף פונקציית התפלגות מצטברת לראות את הדמיון ביניהן:

**איור 3.ד : פונקצית ערכי 'diff' והערכים התיאורתיים**

כפי שניתן לראות יש דמיון מאוד גדול, לכן כדי לוודא האם הפרשי ההשמות הטובות מתפלגים בצורה נורמלית, ביצעתי מבחן Kolmogorov Smirnov דרך SPSS לאשש זאת.

**איור 3.ה :** **מבחן קולמוגורוב סמירנוב**

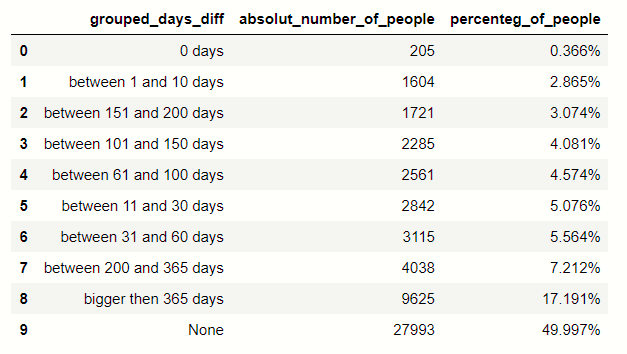
ניתן לראות בבירור כי הערך הקריטי  *, ולכן נאמר כי הנתונים מתפלגים בצורה נורמלית.*בגלל שהנתונים מתפלגים נורמלית השתמשתי במבחן T כי יש לי 30 לשכות מתוך 60 שנדגמו עם מעל ל14 דורשי עבודה ערבים כפי שהוגדר מראש כאשר:

ותוצאת היא -0.0182 . כאשר האומד הסטטיסטי הוא: שווה   
ל. -1.686 מפה לפי מבחן T: t (0.95,29) = 1.699,  **,**. לכן מסיקים כי (ערך הקריטי) גדולה מ ולכן נסיק כי נקבל את ונאמר כי אין הבדל מובהק בין השמת הערבים ובין השמת הלא ערבים. הרי שטענתי אינה מוכחת מהבדיקה ובכך עניתי על שאלת המחקר.

### שאלת מחקר מס' 2: *כמה זמן דורשי עבודה נמצאים בתוכנית טרם השמתם ?*

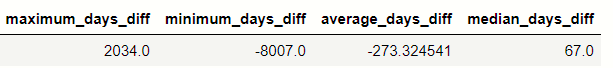
בניתי dataFrame נוסף (df\_entry\_dates) עם העמודות של "תאריך השמה אחרונה", "תאריך דיווח השמה אחרונה", "תאריך כניסה ראשונית לתוכנית". בחלק מנהתונים יש תאריכים ריקים ("NAT"). רציתי לראות את ההפרשים בין "תאריך השמה אחרונה" לבין "תאריך כניסה ראשונית לתוכנית".

df\_entry\_dates.loc[:,'days\_diff'] = df['תאריך השמה אחרונה'] - df\_entry\_dates['תאריך כניסה ראשונית לתוכנית']

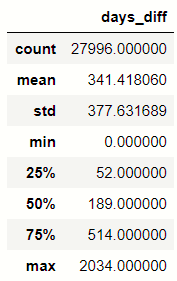
**איור 3.ו : חלוקת 'days\_diff'** **ואחוז של כל קטגוריה מהכלל**

ניתן לראות בבירור כי 49.99% מדורשי עבודה בתוכנית אינן בעלי "תאריך השמה" וכן יותר מ25% יש בעיתיות עם הנתונים שלהם ("תאריך כניסה ראשונה לתוכנית" > "תאריך השמה אחרונה").

**איור 3.ז : מקסימום, מינימום, ממוצע וחציון זמן שהייה בתוכנית לפני ואחרי ניקוי**

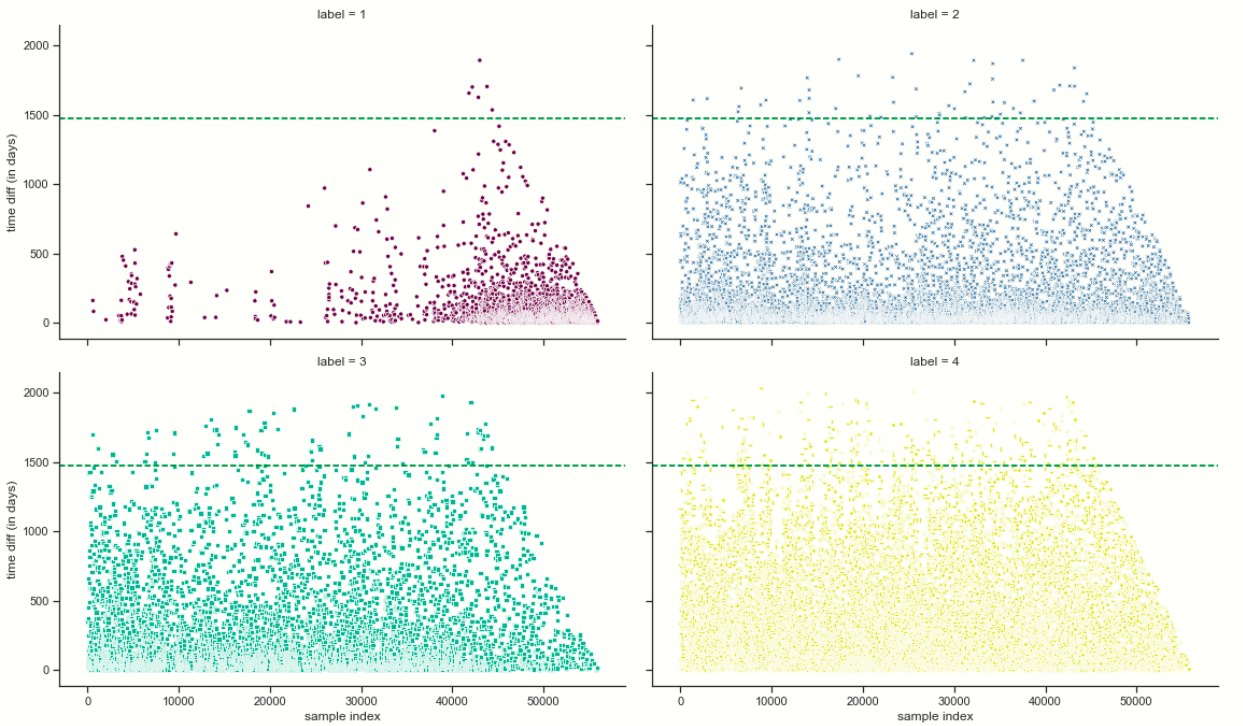


לאחר ניקוי הרשומות שהבן הdiff שלילי, בוצע סיכום של הנתונים:



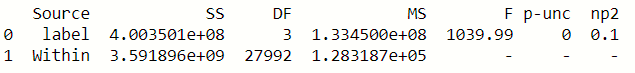
### שאלת מחקר מס' 3: האם יש הבדל במשך שהייתם דורשי עבודה בתוכנית לבין סוגי הצלחות (labels) ?

בהמשך לשאלת מחקר מס' 2, רציתי להבין האם יש באמת הבדל בין הבדלי זמן בין הצלחות (labels) שונים?



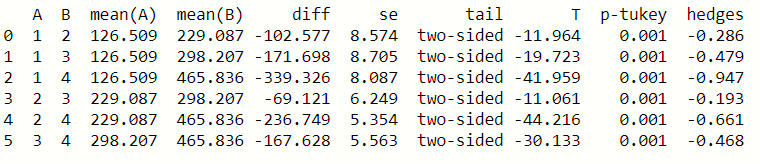
**איור 3.ח : חלוקת דורשי עבודה לפי label ופרק זמן הימצאותו בתוכנית**

מהגרפים נראה שככל שדורש העבודה נמצא בסיווג טוב יותר (label), השהות שלו בתוכנית מתקצר.  
אך לא די רק בגרף של הנתונים להסיק זאת, יש לבצע מבחן סטטיסטי ANOVA שיוכיח זאת:

**איור 3.ט : ANOVA בין label** 

ניתן לראות ש קטן מ0.05 לכן נאמר כי יש הבדל מובהק בין סיווג הצלחת דורשי עבודה בתוכנית (label) לבין זמן השהות שלו בתוכנית.

מבחן ניתוח השונות מצביע על כך שהאוכלוסיות שונות זו מזו, עולה השאלה בין אילו זוגות של אוכלוסיות יש הבדל מובהק. ביצעתי מבחן Tukey post hoc שיבחן הבדלי זמנים בין כל זוג label בנפרד.

**איור 3.י: Tukey Post Hoc Test בין כל זוג label שונים**

במבחן Post Hoc יש מספר k קונטרסים ידוע מראש והם מסוג ספציפי וכל קונטרס מסתכל על הפרש התוחלות בין זוגות של קבוצות, בשיטה זו מחשבים סטטיסטי ורואים האם ההפרש גדול/קטן מהסטטיסטי. בשיטה זו לא מנפחים את שהוא ההסתברות לטעות. השימוש במבחן זה נכון יותר לעומת מבחן T רגיל אשר יגדיל את k והסיכוי לא לדחות את קטן יותר מ , בנוסף גם הסיכוי לדחות גדול מ*.*

באופן ברור ומובהק, ישנם הבדלי זמנים בין label שונים וזאת על פי שתי תוצאות המבחנים שבוצעו להשערות.

### שאלת מחקר מס' 4 : האם יש קשר בין הפעילויות בתוכנית (סדנאות ו/או מפגשים) להשמה?

כפי שהוסבר בטיוב הנתונים, הנתונים בעמודת "פעילויות קודמות שעבר בתוכנית" הם לא מתוך רשימה מוגדרת מראש, אלא נתונה למלל חופשי על ידי מתאמי השמה שונים, נוצר תוכניות ופעילויות רבות. כאן עלה הצורך לטייב את הנתונים לתוכניות ופעילויות מוגדרות מראש (רשימה).  
לאחר סינון רב מ1860 פעילויות שהיו בעמודה, נותרנו עם 30 פעילויות ייחודיות שהוכנסו בחזרה לdataframe.

programs\_list = [ 'חיפוש עבודה מונחה'

,'סדנת מכינה השמתית מעגלי תעסוקה'

,'ייעוץ תעסוקתי'

,'סדנת מכינה תהליכית מעגלי תעסוקה'

,'סדנת יישומי מחשב מעגלי תעסוקה'

,'שיחת אימון אישי'

,'ייעוץ פסיכולוגי'

,'סדנת ליווי בקבוצה קטנה'

,'עברית תעסוקתית'

,'זימון למנהל'

,'סדנת מכינה השמתית'

,'סדנת שינוי'

,'סדנת מכינה מחשבים מעגלי תעסוקה'

,'סדנת מכינה מעגלי תעסוקה'

,'יישומי מחשב מעגלי תעסוקה'

,'סדנת השמה פלוס מעגלי תעסוקה'

,'סדנת מכינה תהליכית'

,'סדנת מכינת השמתית מעגלי תעסוקה'

,'סדנת השמה'

,'מכינה תהליכית מעגלי תעסוקה'

,'סדנת תהליך'

,'סדנת מכינה מעורב מעגלי תעסוקה'

,'סדנת השמה מעגלי תעסוקה'

,'סדנת רכבת מעגלי תעסוקה'

,'סדנת תהליך מעגלי תעסוקה'

,'קורס יישומי מחשב'

,'יישומי מחשב מעגלי תעסוקה'

,'קורס יישומי מחשב'

,'סדנת שינוי מעגלי תעסוקה'

,'מכינה השמתית',]

programs = df\_programs\_labels['פעילויות שעבר בתוכנית']

l =[]

**for** i **in** programs:

p = ''

**for** x **in** programs\_list:

**if** x **in** str(i):

p = p + x + ','

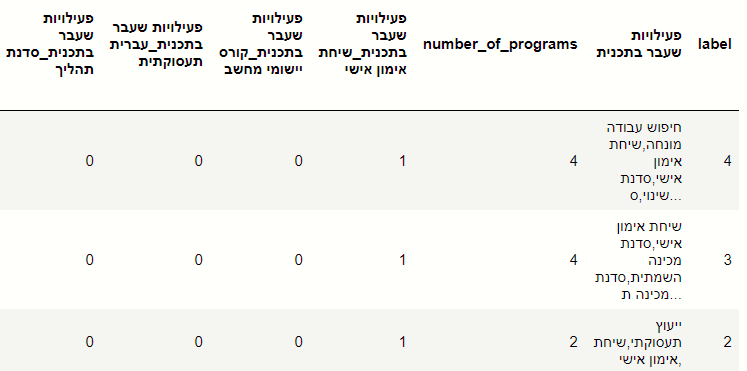
**if** p == '':

p= **None**

l.append(p)

df\_programs\_labels.loc[:,'פעילויות שעבר בתוכנית'] = l

פה היה צורך לשנות את עמודת "פעילויות שעבר בתוכנית" למשתנה קטגוריאלי ופריסת כלל הפעילות בעמודה חדשה ומתן סיווג בינארי האם העבר את התוכנית = 1 אם לא 0.

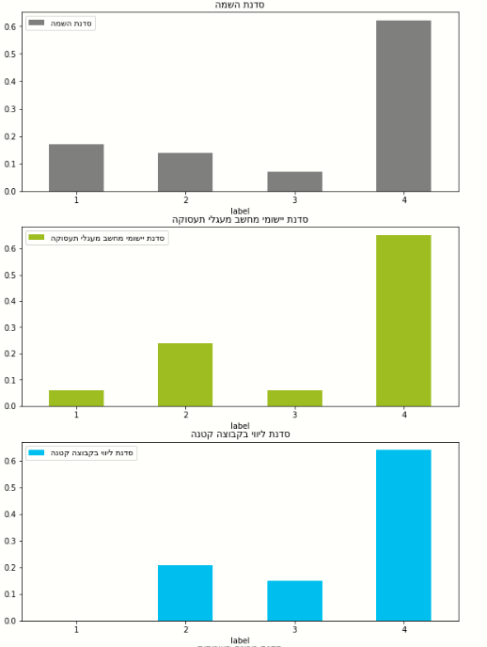
**איור 3.כ : פריסת הפעילויות של דורש העבודה וסכימתן** 

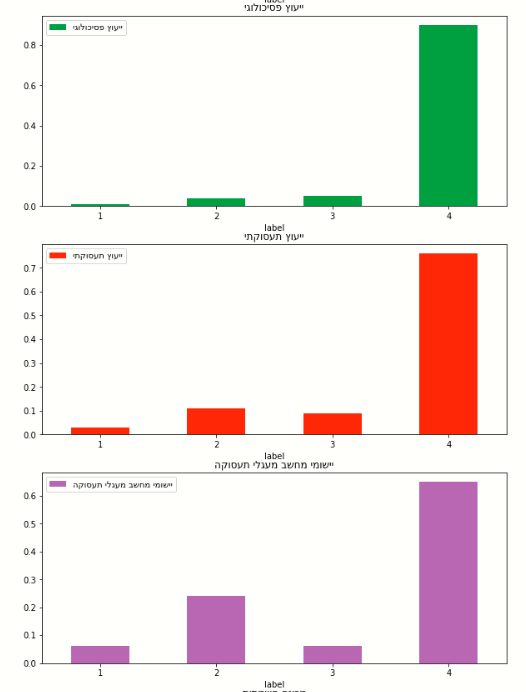
לאחר פעולות אלו, היה לסכום את כלל הפעילויות בהן label הוקצה ולראות את האחוז מכלל הפעילויות.

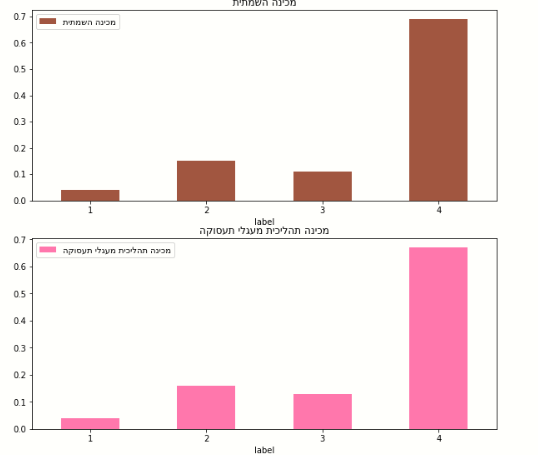
**איור 3.ל : סכום המשתתפים מכל label ואחוז מכלל הפעילויות**

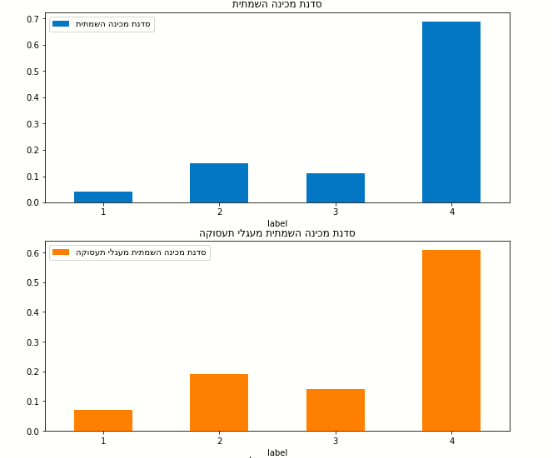
לפי הטבלה לעיל, נראה כי מי שתחת label = 1 ממעיט מאוד את מספר הפעילויות שלו באופן כללי ובכל פעילות בפרט. (ישנם גרפים המראים את כמות המשתתפים מכל label מכל פעילות באופן נפרד(.

**איורים 3.מ : הפעילויות ואחוז מכל label**

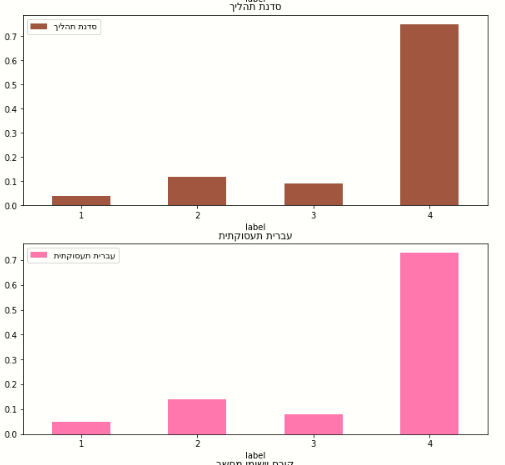


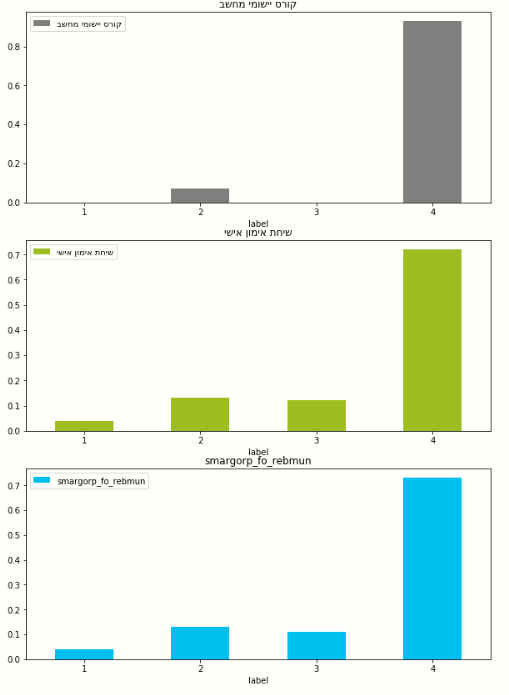


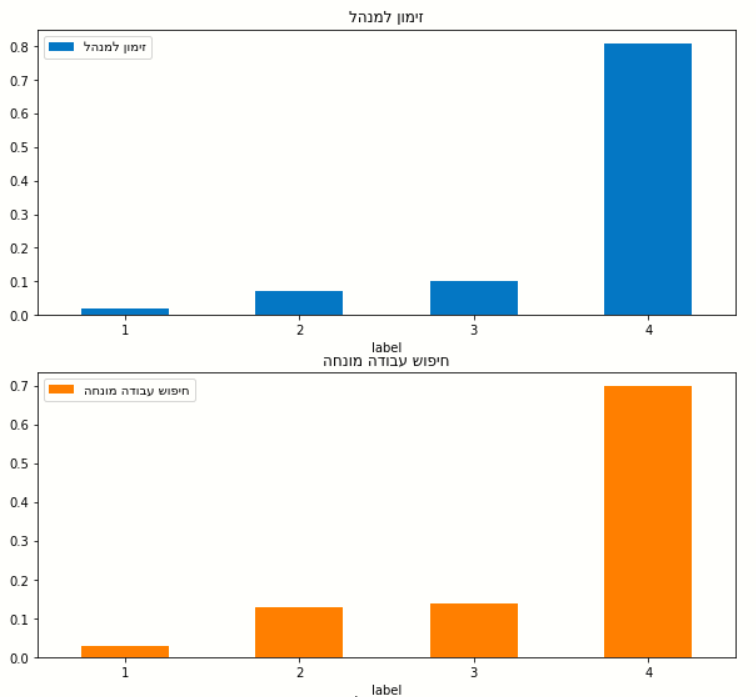




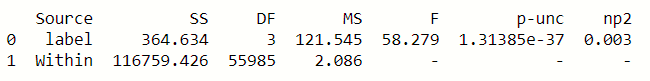






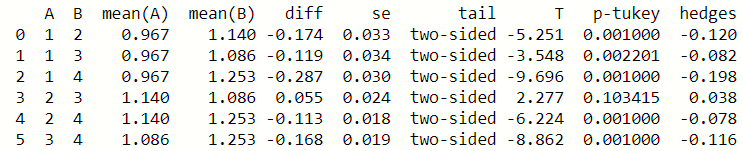


בנוסף היה צורך לראות האם יש הבדל מובהק בשונות בין הפעילויות בתוכנית, לכן בוצע מבחן ANOVA.

**איור 3.נ : מבחן ANOVA לשוני בין הפעילויות**

הערך הקריטי קטן מ ולכן נאמר כי יש הבדל מובהק בין הפעילויות השונות לבין הזמן בו נמצא דורשי עבודה בתוכנית.

היה צורך להמשיך את חקירת ההבדל בין הפעילויות בעזרת מבחן Tukey Post Hoc:

**איור 3.ס : מבחן Tukey Post Hoc לשוני בין זוגות label**

ניתן לראות בבירור כי ישנו הבדל בין מספר הפעילויות בין סוגי ההצלחות השונות (label), חוץ מקבוצות 2 ו3.

לאחר חקירת כל הפעילויות השונות כבודדים, רציתי ליצור מודל עץ החלטות אשר יקבל דגימה של דורשי עבודה ויתאים את המאפיינים הסוציואקונומיים ואת הפעילויות שלו אל מול dataframe שממנו נדגם.

t = **function**(df, sample){

parameters = c()

programs = c()

**for** (i **in** colnames(df)){

**if** (!(grepl('programs', i)) & ('label' != i )){

parameters= rbind(parameters,i)

}

**else**{

programs= rbind(programs,i)

}

}

sample\_parameters = c()

**for** (i **in** parameters){

**if** (!is.na(sample[,c(i)])){

sample\_parameters = rbind(sample\_parameters,i)

}

}

a = TRUE

b = FALSE

**while** ((a ==TRUE) | (b == FALSE)) {

sample\_parameters = sample\_parameters[-length(sample\_parameters)]

filtered\_df = semi\_join(df, sample, by = sample\_parameters, copy = FALSE )

a = any(table(filtered\_df$label)<40)

b = dim(filtered\_df)[1] > 160

}

*## create tree*

df\_prepered\_for\_tree = filtered\_df[,programs]

*# finde the nin label in order to reate equal samplqs from al labels*

min\_label = min(summary(factor(df\_prepered\_for\_tree$label)))

*# pulling sampels with higher programs number of each label.*

v = df\_prepered\_for\_tree[0:0,]

**for** (i **in** c(1,2,3,4)){

s = which(df\_prepered\_for\_tree$label==i)

d = df\_prepered\_for\_tree[s,]

d = d[order(-d$number\_of\_programs),]

a = d[c(1:min\_label-1),]

v = rbind(v,a)

}

df\_prepered\_for\_tree = v

*# factorization of columns:*

**for**(i **in** 1:ncol(df\_prepered\_for\_tree)){

df\_prepered\_for\_tree[,i] <- as.factor(df\_prepered\_for\_tree[,i])}

df\_prepered\_for\_tree = df\_prepered\_for\_tree[,-c(20)] *# drop column "number of programs"*

*# function which revers strings oreder in vector*

strReverse <- **function**(x)

sapply(lapply(strsplit(x, NULL), rev), paste, collapse="")

*# apply function and remove unnecessary strings*

colnames( df\_prepered\_for\_tree) = gsub("programs\_", "", colnames( df\_prepered\_for\_tree))

colnames( df\_prepered\_for\_tree) = strReverse(colnames( df\_prepered\_for\_tree))

fit <- rpart(

lebal~.,

data = df\_prepered\_for\_tree,

method = 'class',

control = list( minsplit = 5,maxdepth = 10)

)

df\_summary = sqldf('select label,

avg(days\_diff) days\_avg,

median(days\_diff) days\_median,

min(days\_diff) min\_diff,

max(days\_diff) max\_diff

from filtered\_df

where days\_diff >= 0

group by 1')

writeLines(paste('\'nthe results based on' ,min\_label\*4,'sampels\n'))

writeLines('\n---- time differences summary ----\n')

print(df\_summary)

rpart.plot(fit)

**return** (fit)

}

הסבר לקוד:

T פונקציה שמקבלת df וsample, יוצרת 2 וקטורים ריקים של פרמטרים (מאפיינים סוציואקונומיים) ופעילויות (programs) ומוציאה כפלט עץ ההחלטות.

יש לולאה שמצרפת כל התוכניות של הדגימה ואת המאפיינים שלו לשימוש בהמשך.  
וקטור נוסף sample\_parameters ריק ובו הפרמטרים של האדם, לאחר מכן נבצע פילטור של האנשים הדומים ביותר שמינימום הדגימות יהיו 160 ומכל label לפחות 40 דגימות.

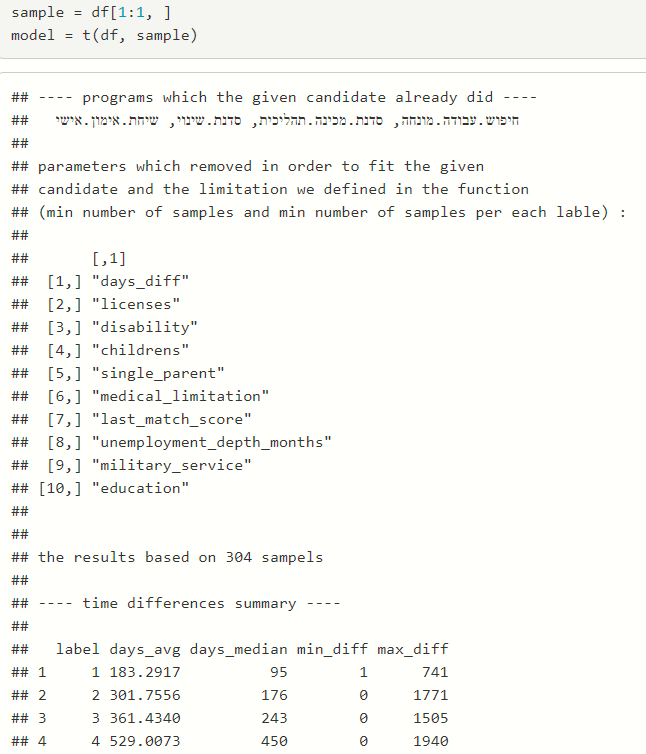
דגימה מתוך שאר df ביחס לlabel המינימלי כדי לא ליצור מדגם מוטעה, לכן נשאל מאיזה label יש את מס' הדגימות הקטן ביותר, ואז נשווה את מספר הדגימות משאר הlabels לmin\_label.

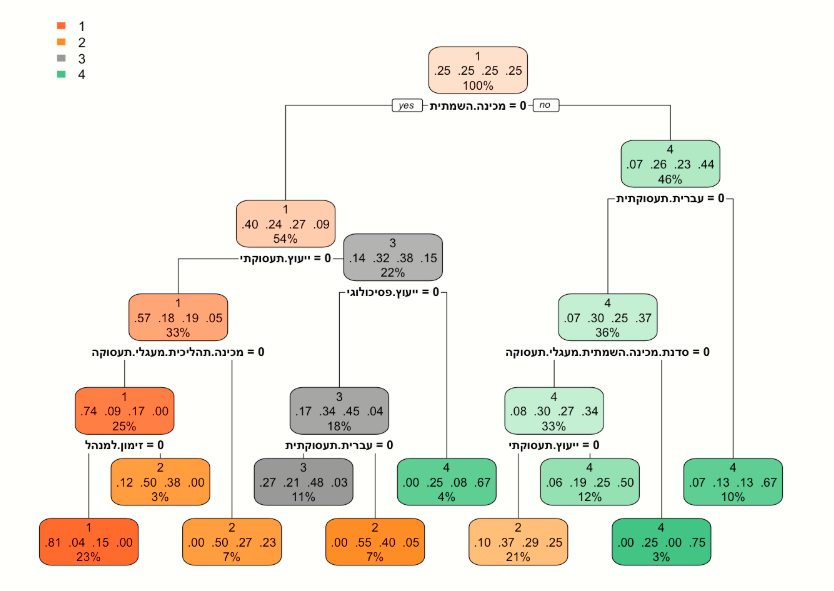
לאחר ההשוואה נדגום את האנשים שביצעו את מספר התוכניות הגדול ביותר כדי להגדיל את הסיכוי להשוואה מקסימלית אל מול הדגימה (הדורשי עבודה הנבחר sample). ביצוע פקטוריזציה לעמודות ונהפוך את שם העמודות (reverse) בגלל העברית והורדת צירוף programs\_"".

לאחר מכן fit מבצעת את בניית עץ ההחלטות לפי מספר דגימות מינימלי לבניית העץ הוא 5 והמקסימום הוא 10 (הדגימות פה מדברות על הפעילויות), ההחלטה על מספרים אלו נבעה מהכירות עם התוכנית.

Df\_summary מבצעת שאילתת sql על הdataframe כדי לקבל סטטיסטיקות של זמנים

כעת, מוציאים פלט של העץ ואיתו הרישום של סטטיסטיקות הזמנים, מספר הדגימות שנעשו אליהם השוואה.

דוגמה:

**איור 3.ע :** **פלט מודל עץ החלטות**

המודל דוגם דורשי עבודה ומציג את המסלול המובהק ביותר בשבילו על פי האנשים הדומים לו, הן מבחינת התוכניות שביצעו והן מבחינת מאפיינים סוציואקונומיים.

לדוגמה המוצגת לעיל, נראה עץ ההחלטות עבור דורשי עבודה שנדגם (sample[1:1]) מתוך dataframe (df).  
הדורשי עבודה הושווה במאפייניו הסוציואקונומיים ובתוכניות שעבר אל מול 304 דורשי עבודה מתוך df.  
מסלול העץ מצביע שלדורשי עבודה מומלץ לא לעבור אף תוכנית מהתוכניות הבאות כפי שהיו אצל הדורשי עבודה שאליהם השווה וזאת כדי שיגיע לlabel הטוב ביותר (1).

ההחלטה להשתמש בעץ ההחלטות נובעת מעצם שהוא מראה את המסלול (סדר הפעולות) המומלץ ביותר. כמובן שישנם עוד מודלים רבים להחלטות מסוג זה, אך לאחר בחינת אלטרנטיבות הוחלט שמודל זה יהיה הטוב ביותר למצב הנתון.

המודל מראה את המסלול המובהק ביותר, לכן יש להתייחס לעץ כאל סדר הפעולות המומלץ ביותר לאותו דורשי עבודה נדגם. כמובן שאין בעץ הבטחה חד משמעית שלפי המסלול המוצע, דורשי עבודה יושם בצורה זו.

## שאלת מחקר מס' 5: האם יש השפעה בין המאפיינים הסוציואקונומיים של דורש עבודה לבין השמה?

שאלה זו רוצה לבדוק האם קיים מאפיין סוציואקונומי אשר משפיע על השמה יותר מאשר מאפיינים אחרים. לכן הגדרתי מהם המאפיינים הסוציואקונומיים של דורשי עבודה: דת, גיל, חד הורי, מגדר, רמת השכלה, ישוב, שפה, ארץ לידה, מצב משפחתי, ילדים עד גיל 18, סיווג דוע, אחוזי נכות, מגבלה רפואית, רישיונות, שירות צבאי, אסיר משוחרר וחודש השמה ('תאריך השמה אחרונה').

בניתי Dataframe חדש אשר מורכב מעמודות אלו וטייבתי את הנתונים בהם.  
איחדתי בין הדתות לדתות מרכזיות: יהודי, נוצרי, מוסלמי ודרוזי, אחרת יקבל ערך: אחר.  
רמת השכלה סווגה לעיקריות: יסודי, תיכון, תואר ותעודת מקצוע. גילאים קובצו לפי הנהוג בשירות התעסוקה: 15-29, 30-39, 40-49, 50-54, 55+. בעמודת 'ארץ הלידה' ישנם כמה מקומות בישראל שהתחלקו: ישראל, יהודה ושומרון, רמת הגולן. כל ארץ לידה שהיה בה פחות מ1% תדירות, ירדה מהנתונים כי הם מקרי קצה שלא ישפיעו על התוצאות בהמשך.

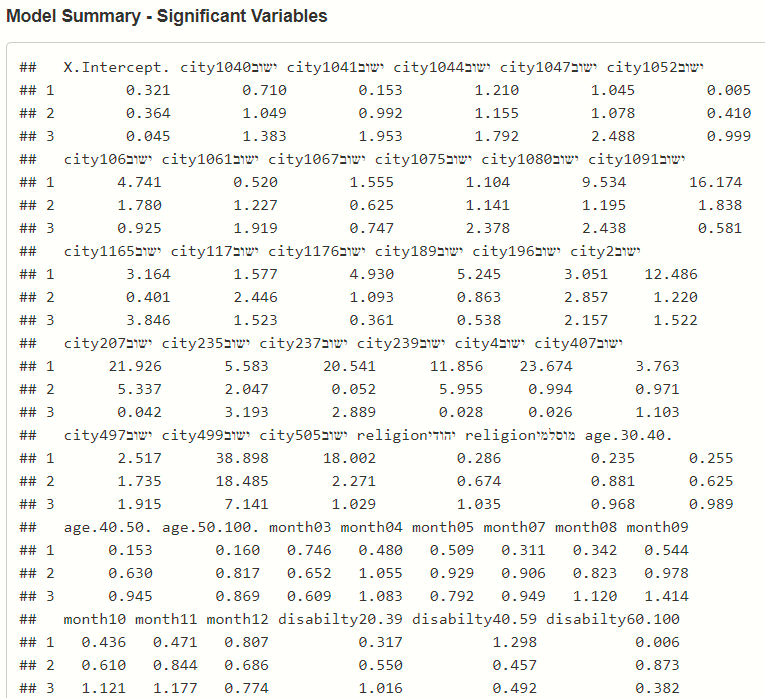
אשתמש במודל רגרסיה מולטינומית המשווה את המודל עם כלל המשתנים, למודל עם חותך בלבד.  
מודל זה מוכר בשיטות מחקר כמו במחקר של Xiaoye Zhang, Yiping Tian, Mindong Bai, Zhitao Zhang ‎[11].

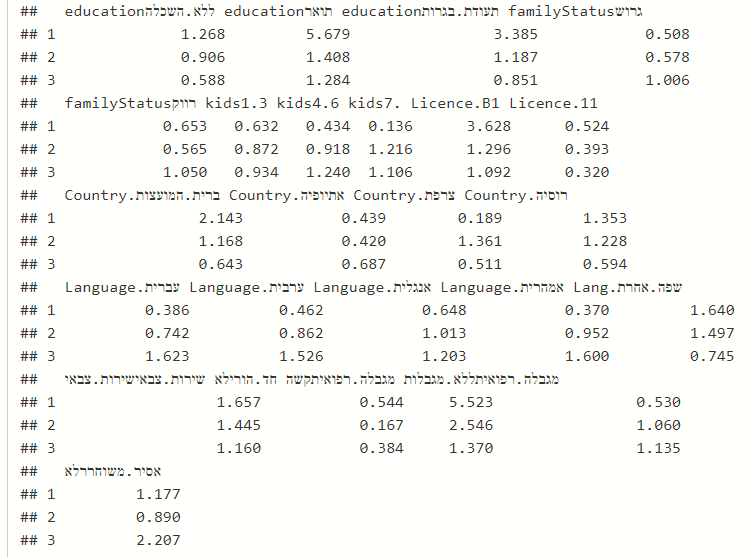
הנוסחה של McFadden, מחסרת מ-1 את החלוקה של לוג האומדים ולכן, ככל שהערך המתקבל (בין 0 ל-1) גבוה יותר כך המודל מובהק יותר.

חשוב להבין שברגרסיה מולטינומית מרובת משתנים קשה להגיע לתוצאה שקרובה ל1 כיוון שצריך פרמטרים מסבירים חזקים מאוד על מנת להעלות את הערך של האומד של McFadden.

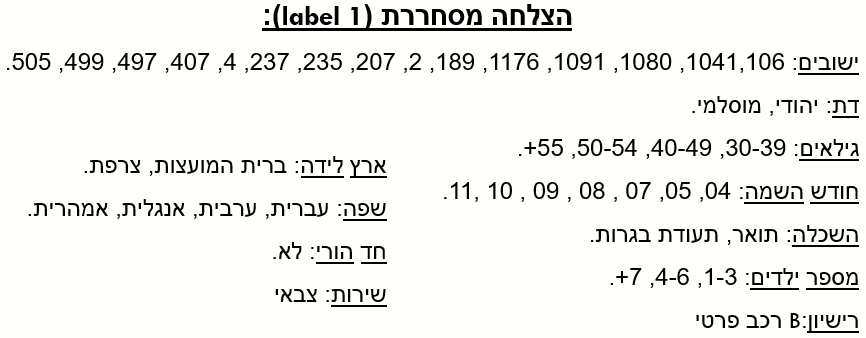
בניסיון לשחק עם המודל ולהכניס אינטראקציות בין משתנים ולהסיר משתנים לא רלוונטיים, לא הושגה תוצאה סטטיסטית יותר מובהקת מהמודל המלא. במקרה של הנתונים שלי, אומד McFadden הוא 0.2368146.  
לפי דבריו של McFadden ‎[12] האומד במודל להתאמה מצוינת תעמוד בין 0.2 לבין 0.4, לכן ניתן לומר שזה מובהק סטטיסטית וכי יש התאמה בין המשתנים הסוציואקונומיים לבין סוג ההשמה. המשתנים שמציג המודל הם בעלי סיכוי גדול יותר להשפעה על תוצאת ההשמה (מאיזה סוג label יהיה דורשי עבודה).

התוצאות הבאות הן המאפיינים הסוציואקונומיים המשפיעים ביותר והמובהקים ביותר סטטיסטית ביחס למשתנים הנוספים על השמתו או הצבתו של דורשי עבודה מכל אחד מהlabels:

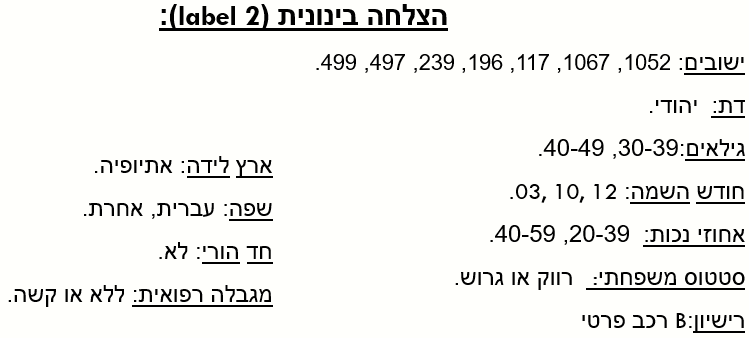




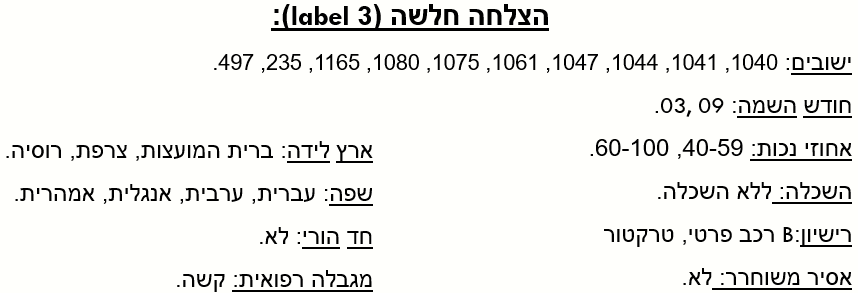
**סיכום תוצאות המודל:**



לפי היגיון בריא ו/או נורמות חברתיות, למי שיש תואר אמור להצליח יותר מאשר למי שיש רק תעודת בגרות או ללא השכלה. ישנם 10 ישובים ערביים לעומת 5 יהודיים. יש קשר ישיר בין מספר ילדים לחד הורי. אפשר לנחש או לחקור כיוון של חודש ההשמה שאולי מצביע על אבטלה תקופתית.



היגיון בריא אומר שיהיה קשה יותר להצליח למי שבעל אחוזי נכות גבוהים יותר, או מצב משפחתי גרוש (מזונות). ישנם 3 ישובים ערביים לעומת 4 יהודיים. אם מוצאך אתיופי, ישנו סיכוי גדול יותר להצלחה נמוכה מאלו של ברית המועצות או צרפת.



לפי היגיון בריא ו/או נורמות חברתיות, למי שאין השכלה כלשהי, אמור להצליח פחות / לא להצליח מאילו שכן בעלי השכלה. אם המוצא של דורשי עבודה רוסי, אז יש פה מובהקות שתהיה מlabel 3. בסוג הצלחה זו ישנם 7 ישובים יהודיים לעומת 3 ערביים. גם לגבי מגבלה רפואית קשה, ניתן להבין שככל שהמגבלה הרפואית קשה יותר, סיכוי ההשמה פחות איכותיים כי ישנם מגבלות לאדם בעבודתו.

# **מסקנות המחקר**

לאחר ביצוע מבחנים שונים, ניסוי כמה כיווני חשיבה וניתוחים שונים, ישנם מספר מסקנות מניתוח הנתונים ומענה על שאלות המחקר על תוכנית 'מעגלי תעסוקה' של שירות התעסוקה הישראלי:

* אין הבדל בין השמות (labels 1 and 2) מהאוכלוסייה הערבית לשאר האוכלוסייה.
* מודל עץ ההחלטות ומבחנים סטטיסטים הראו כי ככל שדורש העבודה יבצע פחות פעילויות בתוכנית, כך הסיכוי להצלחה איכותית יותר תגדל.
* ישנה השפעה בין זמן הימצאותו בתוכנית לבין סוג הצלחתו (השמה איכותית).
* ישנה השפעה בין מאפיינים סוציואקונומיים של דורש עבודה על סוג ההצלחה בתוכנית.

# **תרשים גאנט**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| שם פעילות | משך | התחלה | סיום | בוצע או לא |
| **קביעת פרויקט גמר** | **18 days** | **חמש' 06/06/19** | **שני 01/07/19** | **V** |
| קביעת פגישה עם מנהלת אישית בארגון | 1 day | חמש' 06/06/19 | חמש' 06/06/19 | V |
| העברת החלטה למחלקה משפטית | 9 days | ראש' 09/06/19 | רבי' 19/06/19 | V |
| קביעת פגישה עם סמנכ"ל הארגון | 1 day | חמש' 20/06/19 | חמש' 20/06/19 | V |
| קביעת פגישה עם פיני | 1 day | ראש' 23/06/19 | ראש' 23/06/19 | V |
| החלטה על שאלת המחקר | 1 day | שני 24/06/19 | שני 24/06/19 | V |
| הפקת מסמך דוח יזום | 4 days | שלש' 25/06/19 | ראש' 30/06/19 | V |
| קביעת פגישה עם ד"ר בנימין ברסקי וקבלת אישור לפרויקט | 1 day | שני 01/07/19 | שני 01/07/19 | V |
| **סקר ספרות** | **123 days** | **שלש' 02/07/19** | **שישי 20/12/19** | **V** |
| איסוף מאמרים | 52 days | שלש' 02/07/19 | רבי' 11/09/19 | V |
| קריאת מידע על אבטלה | 52 days | ראש' 01/09/19 | שני 11/11/19 | V |
| שליחת מאמרים לקניה | 7 days | שלש' 12/11/19 | רבי' 20/11/19 | V |
| סינון מאמרים | 24 days | שלש' 12/11/19 | ראש' 15/12/19 | V |
| לקיחת מסקנות ונתונים | 6 days | ראש' 15/12/19 | שישי 20/12/19 | V |
| **חקר מצב קיים** | **129 days** | **שני 12/08/19** | **חמש' 06/02/20** | **V** |
| למידת הארגון | 80 days | שני 12/08/19 | ראש' 01/12/19 | V |
| דוח תכנון מפורט | 8 days | שני 25/11/19 | רבי' 04/12/19 | V |
| איסוף נתונים | 5 days | חמש' 12/09/19 | רבי' 18/09/19 | V |
| עיבוד נתונים | 97 days | חמש' 19/09/19 | שבת 01/02/20 | V |
| בדיקת טיב הממצאים | 4 days | ראש' 02/02/20 | רבי' 05/02/20 | V |
| הצגת המידע והמסקנות | 1 day | חמש' 06/02/20 | חמש' 06/02/20 | V |
| **דוח ומצגת אמצע** | **17 days** | **ראש' 09/02/20** | **שני 02/03/20** | **V** |
| הכנת דוח אמצע | 4 days | ראש' 09/02/20 | רבי' 12/02/20 | V |
| הכנת המצגת | 4 days | ראש' 09/02/20 | רבי' 12/02/20 | V |
| אישור מנחה | 1 day | חמש' 13/02/20 | חמש' 13/02/20 | V |
| תיקונים | 3 days | ראש' 16/02/20 | שלש' 18/02/20 | V |
| חזרה על ההצגה | 13 days | רבי' 12/02/20 | שישי 28/02/20 | V |
| הצגת המצגת אמצע | 2 days | ראש' 01/03/20 | שני 02/03/20 | V |
| **הצגת פרויקט גמר** | **66 days** | **שלש' 03/03/20** | **שלש' 02/06/20** |  |
| הכנת המצגת | 20 days | שלש' 03/03/20 | שני 30/03/20 | V |
| הכנת דוח סופי | 20 days | שלש' 31/03/20 | שני 27/04/20 | V |
| תיאום מול בית הדפוס | 1 day | שלש' 28/04/20 | שלש' 28/04/20 |  |
| בדיקת המצגת | 1 day | שלש' 31/03/20 | שלש' 31/03/20 |  |
| אישור מנחה | 2 days | רבי' 01/04/20 | חמש' 02/04/20 |  |
| שליחת המצגת לבדיקה | 7 days | רבי' 01/04/20 | חמש' 09/04/20 |  |
| תיקונים | 17 days | ראש' 12/04/20 | שני 04/05/20 |  |
| חזרות על הצגת הפרויקט | 6 days | שלש' 05/05/20 | שלש' 12/05/20 |  |
| הצגה סופית | 2 days | שני 01/06/20 | שלש' 02/06/20 | שלש' 02/06/20 |

# **ביבליוגרפיה**

## קישורים לאתרים חיצוניים

1. <https://data.oecd.org/unemp/unemployment-rate.htm-> The Unemployment Rate by the OECD

1. <https://ec.europa.eu/eures/public/language-selection> - The EURES

## מאמרים

1. Danuše Nerudová "Alternative approaches for assessing the European countries economic and social results". *17th International Conference Enterprise and Competitive Environment 2014* Volume 12, Pages 1-804 (2014).
2. Els Sol C.C.A.M. (2016). Getting unemployed with problematic debt problems to work. A quick scan of debt help strategies by European Public Employment Services. Universiteit van Amsterdam, AIAS Working Paper 167.
3. Zwinkels, W. en Guiaux, M. (2015) "Schulden belemmeren terugkeer naar werk" ESB, 1000 (4722),690-92.
4. Zwinkels, W. (2015), "Wie heeft schuld? Een kwantitatieve analyse van schulden bij uitkeringsgerechtigden". Amsterdam:UWV.
5. Koning, P. (2014) "Door schuldhulpverlening uit de bijstand". *ESB* jrg 99 (4677), 38-41.
6. Sol, C.C.A.M. & K. Kok (2014) *Fit or Unfit: Theorie en praktijk van Re-integratie*. Amsterdam: RVO.

1. [Rainer Winkelmann](https://wol.iza.org/authors/rainer-winkelmann) "Unemployment and happiness". *doi: 10.15185/izawol.94* October 2014.
2. Gabriella Sjögren Lindquist, "Unemployment insurance, social assistance and activation policy in Sweden", *Implementierung the news basic allowance for job seekers, Berlin.* 17-18 April 2007 [https://www.researchgate.net/profile/Gabriella\_Lindquist/publications 13 April 2015](https://www.researchgate.net/profile/Gabriella_Lindquist/publications%2013%20April%202015).
3. John P. Martin " Activation and Active Labour Market Policiesin OECD Countries- Stylized Facts andEvidence on their Effectiveness", *Geary Institute, University College Dublin and IZA*, 2014.
4. International Labour Organization (ILO) Resolutions Concerning Economically Active Population, Employment, Unemployment and Underemployment Adopted by the 13th International Conference of Labour Statisticians, October 1982, para. 10
5. Xiaoye Zhang, Yiping Tian, Xiaofang Zhang, Mindong Bai, Zhitao Zhang, "Use of multiple regression models for predicting the formation of bromoform and dibromochloromethane during ballast water treatment based on an advanced oxidation process", 4 August 2019.
6. Daniel McFadden, "Quantitavie methods for analyzing travel behavior of individuals: some recent developments", November 2, 1977.
7. צמצום פערים חברתיים כלכליים ופיתוח כלכלי במזרח ירושלים | מספר החלטה 3790 | משרד ראש הממשלה.
8. פעולות המדינה לעידוד שילובה של האוכלוסייה הערבית בתעסוקה, דוח שנתי 66ג, התשע"ו 2016.
9. Tea Lallukka, Martta Kerkelä, Tiina Ristikari, Marko Merikukkac, Heikki Hiilamo, Marianna Virtanen, Simon Øverland, Mika Gissler, Jaana I. Halonen, "Determinants of long-term unemployment in early adulthood: A Finnish birth cohort study", Volume 8, August 2019, 100410.
10. Pohlan, “Unemployment and social exclusion”, Volume 164, August 2019, Pages 273-299.
11. Lucía Artazcoz, MPH, Joan Benach, PhD, Carme Borrell, PhD, and Immaculada Cortès, MPH, "Unemployment and Mental Health: Understanding the Interactions Among Gender, Family Roles, and Social Class", *American Journal of Public Health* January 2004, Vol 94, No. 1
12. Laura Pohlan , *Journal of Economic Behavior and Organization 164 (2019)*, "Unemployment and social exclusion" , 273–299.