מכון
טכנולוגי חולון
Holon Institute
of Technology

המחלקה למדעי הנתונים
פרויקט גמר - תואר שני

פיתוח מודל לחיזוי סיכויי העזיבה של עובדים במוקדי שירות טלפוניים בישראל

Developing a Predictive Model to Estimate the Probability of Employee Resignation in Israeli Call Centers

מגישה:
אינסה קוגן

מנחה:
ד"ר ולנטין ונצק

תאריך: 30.08.2025

תוכן העניינים

מבוא 3
Abstract 3
רשימת תרשימים 5
רשימת טבלאות 5
1. סקירת ספרות 6
1.1. סיבות וגורמים מרכזיים 6
1.2. התנהגויות נסיגה וניבוי עזיבה 7
1.3. הקשר למחקר הנוכחי 8
2. מטרות המחקר 9
3. שיטות 9
3.1. שיטת הניתוח 9
3.2. רגרסיית Cox (Cox Proportional Hazards Regression) 9
3.3. בחירת משתנים עבור Cox 12
4. הגדרת משתנים 15
4.1. אוכלוסייה ומקור הנתונים 15
4.2. הכנת הנתונים\עיבוד מקדים של הנתונים (Data Preparation) 15
5. הגדרת משתנים 19
6. ממצאים 23
6.1. תיאור המדגם והאירוע 23
6.2. הבדלים בסיסיים בין קבוצות (תיאור, לפני המודל) 24
6.3. בחירת משתנים למודל 28
6.4. תוצאות מודל Cox הסופי 29
7. דיון ומסקנות 30
7.1. דיון, מסקנות פרשנות ניהולית 30
7.2. חידוש ותרומה 31
7.3. מגבלות 32
7.4. כיווני המשך 32
8. ביבליוגרפיה 33
9. נספחים 36

מבוא

לתחלופה גבוהה של עובדים בארגון יש תג מחיר מאוד גבוה. המחיר בא לידי ביטוי בעלויות גיוס, הדרכה, עקומת למידה, הזמן הניהולי המושקע, ההשפעה של עזיבות על המורל של העובדים שנשארים ועל חלוקת עומס לא מאוזנת וגורמים עקיפים נוספים (Nagadevara, Srinivasan & Valk, 2008). עד היום נעשו הרבה מאוד מחקרים שמטרתם הייתה לבחון את הסיבות העיקריות לתחלופה גבוהה בקרב עובדים בכלל ונציגי שירות טלפוניים בפרט. מהמחקרים השונים עולה כי ישנם גורמים רבים המשפיעים על החלטת העובדים לסיים את העסקתם בארגון.

כוונת עזיבה היא רצון מודע ושקול של עובד לעזוב את הארגון, והיא יכולה לנבוע ממגוון גורמים (כגון אי שביעות רצון, שחיקה תפעולית או הזדמנויות חיצוניות). בהקשר של מוקדי שירות, גורמי שחיקה ותנאי עבודה נמצאו כקשורים לשביעות רצון, לכוונות עזיבה ולנטישה בפועל (Zito et al., 2018). חשוב להדגיש כי כוונת עזיבה עצמה איננה סופית, בעוד העזיבה בפועל היא לרוב אירוע סופי. יותר מכך, כוונת עזיבה לעיתים אף מהווה חלון הזדמנות למעסיקים להבין מאין נובע התסכול וכיצד ניתן למנוע אותו לפני שזה מגיע להחלטה הסופית של העובד לסיים את העסקה (Zhou & George, 2001).

פרויקט זה מתמקד בפיתוח כלי לניבוי סיכויי העזיבה של נציגים במוקדי שירות טלפוניים בישראל ע"י זיהוי גורמים עיקריים המשפיעים על העזיבה. הפרויקט מבוסס על נתוני עזיבה של נציגים במוקדי שירות טלפוניים בישראל ב-7 השנים הקודמות למחקר. ע"י ניתוח סטטיסטי מעמיק ופיתוח מודל המבוסס על נתוני אמת, נוכל לספק לארגונים תובנות חשובות ולאפשר שיפור בניהול המשאב האנושי.

המחקר שלנו יעסוק בניתוח וחקר נתוני העזיבות של כ-1,800 נציגי שירות לקוחות טלפוניים ממוקדים שונים בישראל. היתרון במחקר שלנו בבחירה קפדנית של המשתנים שעשויים להתגלות כמשפיעים על החלטת הנציגים לעזוב את הארגון וכמנבאים טוב את סיכויי העזיבה בוותק העובד בעת הנטישה שמוגדר מראש. כמות המשתנים שנבחרה היא גדולה מאוד, מה שמאפשר לנו לזהות קשרים חדשים שעדיין לא נבדקו במחקרים שנעשו עד כה בתחום. המטרה של המחקר זה היא להבין לעומק את הגורמים המובילים לעזיבת עובדים מצד אחד ולנבא את סיכון מצטבר לעזיבה עד זמני ותק מוגדרים מראש מצד שני ובכך לאפשר לארגונים לפתח אסטרטגיות שימור יעילות, לשפר את שביעות הרצון של העובדים ולהפחית עלויות גיוס והכשרה.

Abstract

High employee turnover carries a substantial cost for organizations. These costs are reflected in recruitment and training, the learning curve, managerial time invested, the impact of departures on the morale of employees who remain, imbalances in workload distribution, and other indirect effects (Nagadevara, Srinivasan & Valk, 2008). To date, numerous studies have examined the main reasons for high turnover among employees in general and call-center representatives in particular. Across these studies, a wide array of factors has been shown to influence employees' decisions to terminate their employment with an organization.

A turnover intention is a deliberate, conscious willingness to leave the organization and may arise from various sources (e.g., dissatisfaction, operational strain, or external opportunities). In the context of call centers, strain factors and working conditions have been linked to job satisfaction, intentions to leave, and actual attrition (Zito et al., 2018). It is important to emphasize that an intention to leave is not final, whereas actual departure is typically definitive. Moreover, an expressed intention to leave can at times provide employers with a window of opportunity to identify the sources of frustration and to intervene before this culminates in an employee's final decision to terminate the employment relationship (Zhou & George, 2001).

This project focuses on developing a model to predict the probability of resignation among call-center representatives in Israel by identifying the key factors that drive

attrition. The project is based on attrition data for call-center representatives in Israel covering the seven years preceding the study. Through in-depth statistical analysis and the development of a model grounded in real-world data, we aim to provide organizations with valuable insights and to enable improvements in human resource management.

Our study analyzes attrition data for approximately 1,800 telephone customer-service representatives from multiple centers. The strength of our study lies in the careful selection of variables that may influence representatives' decisions to leave the organization and that are likely to predict employees' tenure up to the point of separation. The relatively large set of variables allows us to identify relationships that have not yet been examined in prior research. The goal of the study is twofold: to develop a deeper understanding of the factors leading to employee attrition and to predict the time to departure. In doing so, we aim to help organizations design effective retention strategies, improve employee satisfaction, and reduce recruitment and training costs.

רשימת תרשימים

תרשים 1: משתני KPI -התפלגות לפני ואחרי השלמת ערכים

תרשים 2: התפלגות משך הזמן בחודשים עד לעזיבה/ צנזורה

תרשים 3: עקומת Kaplan-Meier לכלל העובדים

תרשים 4: סיכון מצטבר (Nelson-Aalen)

תרשים 5: עקומת Kaplan-Meier לפי גודל צוות - ללא הבדלים מובהקים (בנספח)

תרשים 6: עקומת Kaplan-Meier לפי ניצול ימי מחלה

תרשים 7: עקומת Kaplan-Meier לפי ניצול ימי חופשה (בנספח)

תרשים 8: עקומת Kaplan-Meier לפי מוקד

תרשים 9: עקומת Kaplan-Meier לפי ילדים

תרשים 10: עקומת Kaplan-Meier לפי מצב משפחתי

תרשים 11: Violin plot של גיל לפי סטטוס

תרשים 12: Violin plot של שכר יחסי לפי סטטוס

תרשים 13: Violin plot של פער ממוצע שיחות נענות לפי סטטוס

תרשים 14: מורכבות וביצועי מודל - השוואה בין 3 שיטות

תרשים 15: Forest Plot של המודל Stepwise

תרשים 16 : בדיקת פרופורציונליות (PH test) (בנספח)

תרשימים 17–21: מרטינגייל residuals (בנספח)

תרשים 22: תצפיות משפיעות (בנספח)

רשימת טבלאות

טבלה 1: טבלת משתנים מסכמת

טבלה 2: הסתברויות שרידות בנקודות זמן מרכזיות

טבלה 3: השוואה חד-משתנית של משתנים רציפים (כולל HR חד-משתני)

טבלה 4: השוואת ביצועי מודלים לפי שיטת בחירת משתנים

טבלה 5: תוצאות מודל Cox (Stepwise) – HR ורווחי סמך

טבלה 6: VIF על משתני המודל (בנספח)

טבלה 7: יחס אירועים לפרמטר EPV (בנספח)

טבלה 8: תוצאות מודל Cox Proportional Hazards עם בחירת משתנים בשיטת LASSO (בנספח)

טבלה 9: ממצאי Random Forest (בנספח)

1. סקירת ספרות

סקירת הספרות שלהלן ממפה את הידע הקיים על תחלופת עובדים בכלל ועל מוקדי שירות טלפוניים בפרט. חשוב לציין כי בספרות מבחינים בין שני מושגים שונים:

(א) כוונת עזיבה (Turnover Intention) - מבטאת את הנטייה או הרצון המודע של העובד לעזוב את הארגון ונמדד באמצעות שאלוני דיווח עצמי בסקאלת ליקרט (למשל 'אני מתכוון לעזוב את מקום העבודה בחודשים הקרובים'). מדידה זו נפוצה במחקרים בתחום התנהגות ארגונית כיוון שהיא מאפשרת איסוף רוחבי ומהיר של תצפיות במטרה לאמוד קשרים ישירים ועקיפים בין משתנים השונים ולזהות את הסיבות לכוונת עזיבה. מחקרי 'כוונת עזיבה' מדווחים בדרך כלל על מקדמי מודל β ומובהקות p.value , על שונות מוסברת R^2 , ובמחקרים שבהם נעשה שימוש ב- Structural Equation Modeling (SEM) גם על מדדי התאמת מודל (מדדים המשקפים עד כמה המודל מתאים לנתונים).

(ב) עזיבה בפועל/זמן עד עזיבה (Turnover/Attrition) - מדובר באירוע של סיום יחסי עבודה ולפעמים גם זמן עד להתרחשות האירוע. בניגוד למחקרי 'כוונת עזיבה', כאן הדגש הוא על ניבוי ואיתור מוקדם של עובדים בסיכון, על בסיס רשומות ארגוניות (תפעול, היעדרויות, דמוגרפיה ועוד) בשילוב מודלים סטטיסטיים ולמידת מכונה. לצורך הערכת הביצועים מקובל לדווח באופן כללי על מדדי דיוק של מסווגים או על יחסי סיכון (Hazard Ratios, HR) במודלי שרידות כגון- Cox בהתאם לשיטה שנעשה בה שימוש.

1.1. <u>סיבות וגורמים מרכזיים</u>

- **סביבת העבודה לעומת מאפייני התפקיד**

McKnight ועמיתיו (McKnight et al., 2009) בחנו בקרב עובדי IT האם מאפייני סביבת העבודה (כגון תרבות ארגונית ,תמיכת מנהלים וסביבת עבודה כללית) או מאפייני התפקיד (כגון אוטונומיה בעבודה , גיוון במשימות ועומס של עבודה) תורמים יותר להסבר כוונת עזיבה. מטרת המחקר הייתה לאמוד את התרומה היחסית של כל קבוצת משתנים לשונות בכוונת העזיבה ולהפיק תובנות ניהוליות לשימור עובדים. המדגם כלל 127 עובדים, והניתוח בוצע באמצעות מודל משוואות מבניות מבוססי שונות (PLS-SEM (Partial Least Squares SEM . PLS-SEM זו שיטה (סוג של SEM) המאפשרת לחשב ציון למושגים שאינם נמדדים ישירות (למשל 'שביעות רצון' או 'עומס') על בסיס כמה פריטי שאלון נצפים. כל פריט הוא אינדיקטור , והציון למושג מחושב כסכום משוקלל של הפריטים, כאשר המשקלים נקבעים מן הנתונים כך שהמודל יסביר טוב יותר את כוונת העזיבה. המודל הסביר כ-38% מן השונות בכוונת העזיבה ($R^2 \approx 0.38$) . נמצא כי מאפייני סביבת העבודה משפיעים בעיקר בעקיפין, דרך שני מתווכים מרכזיים (שביעות רצון מהעבודה ועומס) וכי תרומתם להסבר כוונת העזיבה גדולה מזו של מאפייני התפקיד. החוקרים הציגו תרשים סיבתי המבוסס על סקירת ספרות, ותוצאות המודל האמפירי נמצאו כעולות בקנה אחד עם התרשים (כלומר, המודל מתקף את התרשים). חשוב להדגיש כי כאן הדגש המחקרי הוא הסבר שונות (R^2) במסגרת מודל מסביר, שזה יעד מתודולוגי השונה מן המטרה של דיוק ניבוי המקובלת במחקרי חיזוי. (McKnight et al., 2009)

- **דיסוננס רגשי, משאבי עבודה ושביעות רצון**

במחקר שנערך במוקדי שירות לקוחות טלפוניים בטורקיה, Celik ו-Oz (2011) מצאו כי נציגי שירות חווים לעיתים קרובות דיסוננס רגשי (פער בין מה שהם מרגישים בפועל לבין מה שהם נדרשים להציג כלפי חוץ). למשל, נציג עשוי לחוש כעס או תסכול מול לקוח מתלונן, אך נדרש לשמור על אדיבות ומקצועיות. במחקר החוקרים השתמשו ברגרסיה היררכית. מדובר בשיטה שמבצעת הכנסת משתנים בבלוקים כדי לבחון תרומה של תוספת קבוצות מסבירים לכוונת עזיבה. נמצא כי דיסוננס רגשי מנבא כוונת עזיבה p<.001 , (β=0.179) עם שונות מוסברת נמוכה ($R^2 \approx 0.03$), ונמצא גם קשר מובהק בין דיסוננס להיעדרויות (Celik & Oz, 2011) .

Zito ועמיתיו (2018) בחנו 318 עובדים במוקד תקשורת באיטליה באמצעות ניתוח מסלולים (Path Analysis) במסגרת מודלי משוואות מבניות SEM קוריאנטי CB-SEM (Covariance-Based SEM). ב- CB-SEM פרמטרי המודל נאמדים כך שהקשרים בין המשתנים שהמודל מנבא (כלומר מטריצת השונויות והקוריאציות החזויה) יהיו דומים ככל האפשר לקשרים שנמדדו בפועל בנתונים (המטריצה המדגמית). התוצאות הראו שלדיסוננס רגשי קיים קשר שלילי לשביעות רצון וקשר

חיובי לכוונת עזיבה, בעוד שמשאבי עבודה מגבירים שביעות רצון ומפחיתים כוונת עזיבה .משאבי עבודה הם היבטים פיזיים, פסיכולוגיים, חברתיים או ארגוניים של התפקיד המסייעים להשיג מטרות עבודה, להפחית את עלויות הדרישות ולעודד צמיחה והתפתחות. למשל תמיכת מנהל, אוטונומיה, משוב והכשרה (Demerouti, Bakker, Nachreiner & Schaufel, 2001). שביעות הרצון תיווכה בין משאבי העבודה לכוונת העזיבה ונמצאה מנבאת שלילית חזקה של כוונת עזיבה ($\beta=-0.39$, $p<.001$) (Zito et al., 2018) .

- **מצבי לחץ ומונוטוניות, מורכבות רגשית ופיקוח הדוק**

מעבר לדיסוננס רגשי, עבודת המוקד מאופיינת במונוטוניות תפעולית לצד מורכבות רגשית גבוהה, שנובעת לרוב מעומס של כמות השיחות, עבודה מול לקוחות מאתגרים (לעיתים עוינים או מתלוננים) ופיקוח ישיר של מנהלים, הכולל הקלטת שיחות והאזנות מדגמיות לצורך בקרת איכות (Valle & Ruz, 2015; Rafaeli et al., 2004). השילוב בין עומס, שגרה ובקרה מתמדת מביאים יחד לתחושת השחיקה (Wallace, Eagleson & Waldersee, 2000; In: Valle & Ruz, 2015). חלק מהמחקרים שנעשו בתחום מצאו קשר ישיר בין הנטייה של הנציגים לעזוב לבין תחושת השחיקה. עובדים הנדרשים לעבוד במצבי לחץ גדולים, עלולים לפתח רמות גבוהות של שחיקה ובסופו של דבר לעזוב (Kumar et al., 2012) .

- **גורמי סביבת עבודה: מנהיגות, תמיכת מנהלים, ואקלים ארגוני**

גם Coomber & Barriball (2007) ציינו במאמרם שכוונת העזיבה של עובד מושפעת מגורמים רבים, בין היתר גם מהשפעת המנהלים בארגון, חוסר שביעות הרצון של העובד ומהתנהלות הארגון. חשוב לדייק שמדובר בסקירה של ספרות אמפירית ולא במחקר אמפירי. הסקירה מדגישה במיוחד את השפעת רכיבי שביעות הרצון (מנהיגות ותמיכת מנהלים, עומס בעבודה, תחושות של שחיקה, תנאים ותמורה) על כוונת עזיבה ועל עזיבה בפועל. אחד הממצאים שחוזרים על עצמם הוא שגורמי סביבת העבודה (תמיכה ניהולית, מנהיגות ואקלים) קשורים באופן עקבי לשינויים בכוונות לעזוב, בעוד שהשפעת גורמים כמו שכר או השכלה משתנה ותלויה בארגון (Coomber & Barriball, 2007).

1.2. <u>התנהגויות נסיגה וניבוי עזיבה</u>

ישנם מחקרים שמצביעים על קיום התנהגויות נסיגה, כגון איחורים והיעדרויות, שמהווים סימני אזהרה מוקדמים לעזיבה (Lee, 1997; In: Nagadevara, Srinivasan & Valk, 2008) .

החוקרים Nagadevara, Srinivasan & Valk (2008) ערכו מחקר שהתמקד בקשר בין התנהגויות נסיגה לבין נטישת עובדים. המדגם שלהם כלל 150 עובדים מחברת תוכנה גדולה בהודו (לא מוקדי שירות). הממצאים הראו שניתן לנבא עזיבה מוקדמת ברמות דיוק של 80%–90% וכן שאיחורים, היעדרויות, תוכן התפקיד, ותק ומאפיינים דמוגרפיים הם מנבאים מרכזיים של עזיבה (Nagadevara, Srinivasan & Valk, 2008).

גם Morrow ועמיתיה (1999) ערכו מחקר שהראה שניתן לנבא נטישת עובדים בין היתר באמצעות היעדרויות מרובות ורמת הביצועים של העובדים. במחקר שלהם נבחנו הקשרים בין היעדרויות וביצועי עובדים לבין נטישת עובדים באמצעות רגרסיה לוגיסטית. זהו מודל הסתברותי למשתנה דיכוטומי (עזב/לא עזב), שמחזיר יחסי סיכויים להשפעת משתנים. הביצועיים מדווחים באמצעות יחסי סיכויים ומובהקות. המדגם שלהם כלל 226 עובדים (113 עוזבים מרצון ו- 113 נשארים מתוך מדגם של 816 עובדים) נמצא קשר מובהק בין ניצול ימי מחלה לבין עזיבה מרצון וקשר שלילי בין הערכות ביצוע לעזיבה. במחקר לא דווח על דיוק המסווג אלא רק מקדמי רגרסיה ומובהקותם (Morrow et al., 1999).

אחד האתגרים הניצבים עבור מוקדי שירות טלפוניים, הוא היקף נטישת העובדים. מוקדי שירות טלפוניים הופכים במהלך השנים לנדבך מרכזי בעולם העסקים מכיוון שהם מספקים שירותים לעסקים רבים הנדרשים לתת מענה קבוע ללקוחות שלהם.

במחקרי מוקדים נמצא כי שיטות שונות של כריית נתונים (שיטות סטטיסטיות שונות ומודלים של למידת מכונה) מאפשרות לזהות תבניות המסייעות לניבוי עזיבה ולקבלת החלטות שימור. כבר במאמר מוקדם הראו Valle, Varas & Ruz (2012) כי נתוני התפעול העשירים הנאספים במוקדי שירות ומכירות טלפוניים מהווים תשתית נוחה ליישומי מודלים שונים של כריית נתונים (המאמר לא עסק בחיזוי עזיבות אלא בביצועי נציגים). בעבודת ההמשך Valle & Ruz (2015), פיתחו מודל

לחיזוי עזיבה בפועל על בסיס רשומות ביצוע שוטפות של נציגים - מדדי ביצוע מרכזיים KPI (Key Performance Indicators) על 2,407 רשומות של נציגי מכירות. נמצא כי שילוב משתני שינוי לאורך זמן (דלתא/מגמות ביחס לעבר ולממוצע המוקד) משפר את יכולת הניבוי , וכי מגמות ירידה עקביות במדדים תפעוליים עשויות לשמש אות אזהרה מוקדם לסיכון עזיבה ולהכוונת התערבויות שימור (Valle, Varas & Ruz, 2012; Valle & Ruz, 2015) .

לסיכום, החוקרים מצביעים על כך, כי ניתן לנבא את סיכון העזיבה. השימוש בשיטות סטטיסטיות שונות ולמידת מכונה מאפשר לנבא תופעות של נטישה וכך מסייע לקבל החלטות ניהוליות שונות לטובת הארגון והעובדים (Valle & Ruz, 2015) . יכולת ניבוי של עזיבות, יחד עם ידע על הסיבות המרכזיות לעזיבה, צפויות להוביל לחיסכון בהוצאות גיוס והדרכה, להפחתת העלויות הנובעות מעקומת הלמידה של עובדים חדשים, לשיפור הרמה המקצועית ורמת השירות ללקוחות. ארגון יוכל למקד ולהתאים את מאמצי שימור לרמת סיכון העזיבה, להיערך טוב יותר מבחינת כוח אדם, לשפר את הפריון ואת רמת המוטיבציה ולהביא לשיפור האווירה במוקד באמצעות ניטור מהיר של עובדים מתוסכלים. כמו כן, הנציגים יזכו לתשומת לב של המנהלים, הערכה טובה יותר וחלוקה מאוזנת יותר של העומס, פעולות שימור ייעודיות ושיפור תנאי העסקה.

1.3. <u>הקשר למחקר הנוכחי</u>

המחקר שלנו מתמקד בעזיבות של נציגים ממוקדי שירות טלפוניים בישראל. רוב הארגונים שיש בהם מוקד שירות טלפוני מתמודדים עם שיעורי תחלופה גבוהים מאוד אשר עומדים על מעל 30% בשנה במוקדים רבים (ב־51% מהמוקדים), כאשר התחלופה הגבוהה ביותר שנמדדה הייתה 46% בממוצע במגזר התחבורה (ובענף האאוטסורסינג מיקור חוץ נמדד ממוצע של כ־35%). כך עלה ממחקר שנערך בטכניון על ידי פרופסור ענת רפאלי ותלמידיה מהפקולטה להנדסת תעשייה וניהול. המחקר התבסס על ראיונות עם מנהלי 80 מוקדים טלפוניים בישראל והחוקרים מצאו כי בין הסיבות העיקריות לתחלופה הגבוהה נמצאים גורמים כגון אופי העבודה ושחיקה רגשית (Rafaeli et al., 2004).

חשוב להבין כי עבודה במוקדי שירות טלפוניים כוללת אינטראקציה קבועה ואינטנסיבית עם אנשים. במקרים רבים הם נדרשים לתת מענה לתגובות עוינות של לקוחות ולתחושת דחייה. מסיבה זו, עבודה בשירות לקוחות במוקדי שירות טלפוניים דורשת מהנציגים מבנה אישיותי ופרופיל פסיכולוגי מתאים (Sawyerr et al., 2009; as cited in Valle & Ruz, 2015) .

ישנם מחקרים רבים שבחנו את המשמעות של נטישה מוקדמת של עובדים והצביעו על כך, כי נטישה משפיעה במידה רבה על ההתנהלות של החברה, הביצועים שלה וגם על מצבה הפיננסי (Nagadevara, Srinivasan & Valk, 2008; Kumar et al., 2012) . מוקדי שירות טלפוניים מספקים שירות לחברות וארגונים ונדרשים לעבוד מול לקוחות הקצה. עבור העובדים והארגון יש השפעות רבות כאמור כאשר העובדים עוזבים את העבודה במפתיע. לעזיבתם של העובדים ישנה חשיבות רבה על התוצרים, הפרודוקטיביות וההכנסות של הארגון. הכשרה וגיוס של עובדים עולים הרבה מאוד כסף לארגון ולכן ארגונים רבים מעוניינים להבין את המנגנונים השונים שגורמים לעובדים לקבל החלטה לעזוב. כך סבר Kumar ועמיתיו, שבחנו את הקשרים בין אוריינטציה הומנית בארגון, יחסי חליפין חברתיים והזדהות ארגונית לבין שחיקה (Kumar et al., 2012) .

ברמת הארגון, תחלופה של עובדים גורמת לירידה בביצועים של הארגון וביעילות של הארגון. למרות קיומה של ספרות מחקרית בינלאומית ענפה על עזיבת עובדים במוקדי שירות (למשל (Morrow et al., 1999; Nagadevara, Srinivasan & Valk, 2008; Zito et al., 2018 , בישראל קיים מחסור יחסית גדול במחקרים אמפיריים כמותיים בתחום זה. מעבר לדו״ח הסקר הרחב של רפאלי ועמיתיה על 80 מוקדים (Rafaeli et al., 2004) כמעט שלא נבחנו לאורך זמן נתוני כח אדם בארגונים בישראל. ומכאן גדל הצורך במחקר הנוכחי .

פרויקט זה מתמקד בפיתוח כלי לניבוי סיכויי העזיבה של נציגים במוקדי שירות טלפוניים בישראל ע״י זיהוי גורמים עיקריים המשפיעים על העזיבה.

מטרת מחקר זה היא לפתח מודל חיזוי של הסיכוי לעזיבת נציגי שירות לקוחות במוקדי שירות טלפוניים בישראל, תוך בחינת הגורמים המרכזיים המשפיעים על עזיבתם, באמצעות ניתוח נתוני אמת שנאספו ממוקדי שירות בישראל במשך שבע שנים.

2. מטרות המחקר

- לזהות את גורמי הסיכון המרכזיים המשפיעים על עזיבה.
- לאמוד את הסיכוי לעזיבה עד זמני ותק מוגדרים מראש (t=6,12,18,24 חודשים) לכל פרופיל עובד.
- לספק לארגונים כלים פרקטיים להתערבות מוקדמת, לצורך הפחתת תחלופת עובדים.

3. שיטות

3.1. שיטת הניתוח

המודל שנבחר במסגרת מחקר זה הוא מודל Cox, שמאפשר לבחון את ההשפעה של משתנים שונים על סיכון מצטבר לעזיבה עד זמני ותק מוגדרים מראש. בחירת המשתנים למודל נעשתה בעזרת שלוש שיטות: RandomForest*, LASSO* (Least Absolute Shrinkage and Selection Operator), Stepwise. כל אחת מהשיטות הפיקה מודל סופי. לבסוף, נערכה השוואה בין שלושת המודלים הסופיים לפי ביצוע ניבוי ופשטות המודל (מספר משתנים מינימלי), ע״מ לבחור את המודל האופטימלי. מדד ההשוואה העיקרי היה C-index (Concordance Index), המודד את אחוז הזוגות שהמודל מדרג נכון מבחינת יחס הסיכון לעזיבה. הבחירה במדד C-index כקריטריון ההשוואה נובעת מכך שהוא מותאם במיוחד לנתוני הישרדות שבהם קיימת צנזורה. בניגוד למדדים כמו Accuracy או MSE, שאינם מתחשבים בממד הזמן ובצנזורה, ה־C-index בוחן את יכולת המודל לדרג נכון זוגות של נבדקים לפי סדר ההתרחשות של האירוע. בכך הוא נותן הערכה אמינה ליכולת הניבוי של המודל בהקשר של זמני עזיבה. בנוסף, ערכי C-index ניתנים לפרשנות אינטואיטיבית (0.5=אקראי, 1.0=מושלם), מה שמקל על השוואה בין מודלים שונים. לבסוף שיטת Stepwise הניבה את המודל הכי טוב מבין השיטות.

3.2. רגרסיית Cox (Cox Proportional Hazards Regression)

ניתוח הישרדות היא שיטה לניתוח נתונים שבהם מודדים את הזמן עד שמתרחש אירוע (למשל עזיבת עובד). בלב הגישה עומדת פונקציית השרידות S(t), שמבטאת את ההסתברות ׳לשרוד׳ (כלומר, להישאר בלי שיתרחש האירוע) מעבר לזמן t. כשמשלבים את S(t) בתוך מודלי סיכונים פרופורציונליים (ובעיקר מודל Cox) היא הופכת לכלי שימושי להערכת סיכון מדויקת ולתחזית לאורך זמן.

הגדרות:

- **אירוע:** התופעה שאותה מנתחים כנקודת סיום. במחקר הנוכחי מדובר בעזיבת עובד.
- **צנזורה:** תצפיות שלא חוו את האירוע עד סוף תקופת המעקב מסומנים כמצונזרים.
- **תחילת מעקב T(0) :** תחילת העסקה/כניסה לתפקיד.
- **משך הזמן עד להתרחשות האירוע T:** המשתנה שמודד את הזמן (למשל ימים/חודשים) מרגע תחילת המעקב ועד התרחשות האירוע הנחקר. בדוגמה שלנו – הזמן מתחילת העסקה ועד עזיבת העובד.

פונקציית השרידות:

$$S(t) = P(T > t)$$

כלומר ההסתברות שזמן האירוע האקראי T יעלה על t (למשל, שעובד עדיין מועסק אחרי t חודשים).

תכונות מרכזיות של $S(t)$:

- $S(0) = 1$ - בתחילת המעקב כולם ׳שורדים׳ (אף אירוע עוד לא התרחש).
- $\lim_{t\to\infty} S(t) = 0$ - אם מניחים שכל אחד בסוף יחווה את האירוע, אחרת אם קיימת תת אוכלוסייה שלעולם לא תחווה אירוע , הגבול יהיה חיובי.

- S(t) מונוטונית לא עולה ולרוב רציפה מימין. ככל שהזמן חולף מצטברים אירועים, ולכן ההסתברות 'עדיין לא אירע' לא יכולה לעלות.

קשר לפונקציית ההתפלגות המצטברת :

$$S(t) = 1 - F(t) \Longrightarrow F(t) = P(T \leq t)$$

בהקשר למחקר הנוכחי $S(6)$ היא ההסתברות שעובד עדיין מועסק אחרי 6 חודשים מתחילת עבודתו, $1 - S(6)$ היא ההסתברות שעזב עד אז.

פונקציית הסיכון (hazard):

קצב ההתרחשות המיידי של אירוע בזמן t, בתנאי שלא אירע עד t. הגדרה מתמטית :

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}((t \leq T < t + \Delta t | T \geq t)}{\Delta t} = -\frac{d}{dt} \ln S(t)$$

הקשר לשרידות דרך הסיכון המצטבר :

$\Lambda(t) = \int_0^t \lambda(u) du$ מכאן מתקבל : $\mathrm{S(t)} = \exp\{-\Lambda(t)\}$

פרשנות בתוך מודל הסיכונים הפרופורציונליים (Cox PH)

במודל סיכונים הפרופורציונליים של Cox (PH), פונקציית הסיכון המותנית בוקטור משתנים x נתונה על ידי :

$$\lambda(t|x) = \lambda_0(t) \exp(\beta^T x)$$

כאשר $\lambda_0(t)$ היא פונקציית הסיכון הבסיסית (הלא-פרמטרית) ו- β הם מקדמי הרגרסיה.

שקול לכך, הלוג-סיכון (Log-Hazard) ויחס הסיכונים (HR) :

$$log\left(\frac{\lambda(t|x)}{\lambda_0(t)}\right) = \beta^T x$$

כאשר :

- $X_1, \dots, X_p$ הם משתנים מסבירים
- $\lambda(t|X_1, \dots, X_p)$ הוא הסיכון בנקודת הזמן t בהינתן $X_1, \dots, X_p$
- $\lambda_0(t)$ הוא הסיכון הבסיסי (הבלתי מותנה) בנקודת הזמן t
- $\beta_1, \dots, \beta_p$ הם מקדמי הרגרסיה.

כלומר בלוג סיכון הקשר ל- x_i ליניארי ,ו- $\exp(\beta_i)$ = הוא יחס הסיכונים לשינוי יחידה ב- x_i בהנחת PH, ה- HR קבוע בזמן.

פונקציית הסיכון מודדת את ההסתברות המיידית להתרחשות המאורע בזמן t, בהנחה שהנבדק שרד עד זמן זה. המודל מבטא את פונקציית הסיכון כתוצאה מהכפלת פונקציית הסיכון הבסיסית $\lambda_0(t)$ (baseline hazard) באקספוננט של צירוף לינארי של המשתנים המסבירים :

$$\lambda(t|x_i) = \lambda_0(t) \exp\left(\beta_1 x_1 + \cdots + \beta_p x_p\right)$$

מכאן נוכל לקבל את פונקציית השרידות המותנית :

$$S(t|x) = exp\{-\Lambda_0(t) exp\{\beta^T x\}\} = [S_0(t)]^{exp(\beta^T x)}$$

כאשר. $\Lambda(t) = \int_0^t \lambda_0(u)du$ היא הסיכון המצטבר הבסיסי, ו- $S(t) = \exp\{-\Lambda(t)\}$ היא השרידות הבסיסית. לכן, ההסתברות לאירוע עד t

$$P(T \leq t|x) = 1 - S(t|x)$$

(לפרוט נוסף, ראו Vancak ועמיתים, 2022).

המודל לא מניח מראש שום הנחה על התפלגות הזמן T. יחד עם זאת, המודל מניח שהמשתנים המסבירים לא משתנים עם הזמן. המודל אינו אומד ישירות את פונקציית הסיכון אלא את לוג יחסי הסיכונים.

מכאן e^{β_i} הוא יחס הסיכונים (hazard ratio) של המשתנה המסביר X_i בהינתן שאר המשתנים המסבירים במודל. אמידת הפרמטרים $\beta_1, \dots, \beta_p$ נעשית בשיטת הנראות המקסימלית החלקית. במילים אחרות, מקדמי הרגרסיה (β) מתורגמים ליחס הסיכונים e^{β_i}, כך שניתן לפרש בקלות אם משתנה מגדיל או מפחית את הסיכון לאירוע.

ייחודו של המודל בכך שהוא סמיפרמטרי, כלומר מבנה שבו חלק אחד לא פרמטרי (כאן: פונקציית הסיכון הבסיסית $\lambda_0(t)$ וחלק שני פרמטרי ($\exp(\beta^T x)$ הקשר בין משתנים לסיכון). הוא אינו מניח צורת פונקציה מוגדרת מראש לפונקציית הסיכון הבסיסית (Baseline Hazard), אלא מתמקד באמידת מקדמי ההשפעה של המשתנים על יחס הסיכונים (Hazard Ratio) לאורך זמן. המשמעות היא שהמודל מאפשר גמישות רבה ומותאם למצבים שבהם קשה או בלתי אפשרי להניח התפלגות מוגדרת מראש לזמני ההישרדות (Therneau & Grambsch, 2000).

אמידת מקדמי המודל מתבצעת באמצעות שיטת ה־Partial Likelihood, שמאפשרת לאמוד את β מבלי לאמוד במפורש את פונקציית הסיכון הבסיסית. בשיטה זו, עבור כל מועד שבו התרחש מאורע, מחשבים את ההסתברות של אותו נבדק לחוות את המאורע מתוך קבוצת הסיכון, כלל הנבדקים ששרדו עד אותו מועד. הכפלת הסתברויות אלה לאורך כל האירועים יוצרת את פונקציית הנראות החלקית, שממקסום שלה מתקבלים אומדני המקדמים.

למודל יתרונות בולטים: הוא משלב באופן טבעי משתנים כמותיים וקטגוריאליים, מתמודד היטב עם נתונים מצונזורים, ואינו מוגבל על ידי הנחות התפלגות נוקשות. עם זאת, הוא רגיש להפרת הנחת הפרופורציונליות, ופרשנותו עשויה להסתבך במקרים של אינטראקציות מורכבות או משתנים שתלויים בזמן לאירוע .(Box-Steffensmeier & Zorn, 2001) .

ישנה דרישה טכנית למודל - היעדר מולטיקוליניאריות בין המשתנים (No Multicollinearity) - כלומר, היעדר מתאם גבוה מאוד בין משתנים מסבירים , כדי לאפשר אמידת β יציבה. למשל, מדדי ביצוע שונים (KPI) שנבדקים במחקר לא זהים זה לזה (במידה ויש מתאם גבוה, ניתן להסיר או לאחד בין המשתנים). בדיקה נפוצה באמצעות חישוב ערכי VIF (Variance Inflation Factor) (מקובל כי VIF<5 או לכל היותר VIF<10).

הנחות עיקריות של המודל

- הנחת יחס סיכונים קבוע (Proportional Hazards) - אחת ההנחות המרכזיות של המודל היא פרופורציונליות, כלומר, ההשפעה של משתנה מסביר (כגון מגדר או שכר) על הסיכון נשארת קבועה לאורך כל תקופת המעקב. לדוגמה: אם הסיכון של עובד גבר לעזוב את מקום העבודה כפול מהסיכון של אישה בחודש הראשון, אז גם בחודש השישי יחס הסיכונים נשאר כפול. בדיקות נפוצות להנחה זו כוללות ניתוח שאריות מסוג Schoenfeld ובדיקת Therneau-Grambsch, לצד המחשה גרפית.

- צנזורה בלתי תלויה (Independent Censoring) - כלומר המשתתפים שלא חוו את האירוע (כלומר, צונזרו) דומים מבחינת הסיכון לעזיבה לאלו שכן חוו את האירוע - אילו היינו ממשיכים לעקוב אחריהם. בהתייחס לפרויקט הנוכחי, העובדים שעדיין מועסקים בסוף תקופת המעקב לא שונים מהותית (למשל בגיל, שכר, הערכה מקצועית) מאלו שעזבו את

החברה. זוהי הנחה שקשה לבדוק ישירות, אך אפשר להצדיק אותה על בסיס מבנה המחקר.

- הנחת ליניאריות של המשתנים הרציפים (Linearity in the Log-Hazard) - הנחה נוספת כוללת קשר לינארי בין משתנים רציפים ללוג־סיכון. כלומר, המשתנים הרציפים (כמו גיל, משכורת) משפיעים על פונקציית הסיכון באופן ליניארי בלוג של הסיכון. אם הקשר לא ליניארי - תיתכן סטייה מהנחות המודל. בדיקות נפוצות להנחה זו מתבצעות באמצעות שאריות מרטינגלים (Martingale Residuals), או על ידי הוספת פולינומים או ספליינים למודל ובחינת ההתאמה.

- תצפיות בלתי תלויות (Independent Observations) - הנחה נוספת היא שהאירועים של כל נבדק בלתי תלויים זה בזה. כלומר, ההסתברות שעובד אחד יעזוב אינה תלויה בהחלטה של עובד אחר. בהקשר של המחקר הנוכחי, חשוב לציין כי הנתונים מתפרסים על פני שבע שנות מעקב, ובמהלכן נציגים שונים החלו את עבודתם במועדים שונים ובמוקדים שונים. עובדה זו מחזקת את ההנחה, שכן זמני ההתחלה השונים מבטיחים שהתצפיות אינן משקפות תהליך קולקטיבי אחיד אלא מהוות מסלולי העסקה נפרדים ובלתי תלויים. לפיכך, אין השפעה ישירה של מועד ההתחלה של עובד אחד על משך ההעסקה של עובד אחר.

לסיכום, רגרסיית Cox מהווה כלי עוצמתי ומתאים ביותר לניתוח הישרדות של עובדים בחברה, ועל כן ניתוח הנתונים בפרויקט זה יתבסס על המודל זה.

3.3. בחירת משתנים עבור Cox

<u>בחירת משתנים בשיטת LASSO עבור Cox</u>

כבסיס, נעשה שימוש במודל Cox הרגיל שהוצג לעיל (ר' בפרק שיטות 3.2), שבו פרמטרי β נאמדים באמצעות מזעור פונקציית הלוג-נראות החלקית. הרחבת LASSO שהוצעה על ידי Tibshirani (1997) מוסיפה אילוץ מסוג L_1 על המקדמים, כך שסכום הערכים המוחלטים של β מוגבל:

$$\hat{\beta} = \text{argmax}\ell_p(\beta), subject\ to \sum|\beta_j| \leq S$$

כאן S הוא גודל האילוץ - סף שמגביל את סכום הערכים המוחלטים של המקדמים.

אפשר לנסח את אותה בעיה בדיוק גם בצורה שקולה גם באמצעות כופל לגראנז' λ:

$$\hat{\beta} = \text{argmax}\left(\ell_p(\beta) - \lambda\sum|\beta_j|\right)$$

כאן λ הוא כופל לגראנז' שמאזן בין איכות ההתאמה (הלוג-נראות) לבין רמת הענישה על גודל המקדמים. במילים אחרות - S מגדיר גבול עליון לסכום המקדמים, ו- λ מגדיר את עוצמת הענישה. שני הניסוחים שקולים מתמטית.

בניסוח הראשון, פרמטר הענישה S נבחר באמצעות קרוס ולידציה. אילוץ זה מביא לכך שמקדמים מסוימים נדחקים בדיוק לאפס, ולכן נוצרת בחירה אוטומטית של תת קבוצה של משתנים. זה עובד על עיקרון מסננת שלא רק מתאימה מודל Cox לנתונים, אלא גם מחפשת תוך כדי מי מהמשתנים באמת תורם לניבוי עזיבה. המשתנים שמוסיפים הרבה מידע נשמרים, בעוד משתנים חלשים או מיותרים נדחקים בהדרגה החוצה.

שלבי עבודה של האלגוריתם:

1. סטנדרטיזציה של משתנים - כדי שהענישה תחול באופן אחיד, לפני שמתחילים, מבצעים 'נרמול' של כל המשתנים (למשל גיל, שכר, מספר שיחות). LASSO מעניש את המקדמים

לפי הגודל המספרי שלהם. אם נשאיר משתנים בסקאלה שונה, משתנה עם מספרים גדולים יותר ייענש יותר חזק.

2. בחירת פרמטר רגולריזציה (λ) - הפרמטר λ קובע כמה חזקה הענישה של LASSO . כדי לבחור את הערך הנכון, משתמשים בקרוס ולידציה (CV) . כלומר, מחלקים את הדאטה לכמה קבוצות, מאמנים על חלק ובודקים על החלק האחר. חוזרים שוב ושוב עד שמוצאים את ערך λ שמאזן בין דיוק לפשטות.

3. אמידה - התאמת מודל Cox עם אילוץ L_1 . בשלב זה מריצים את מודל Cox, אבל במקום להתאים אותו חופשי לכל המשתנים, מוסיפים את האילוץ של LASSO (סכום הערכים המוחלטים של המקדמים מוגבל).

4. בחירת משתנים - משתנים עם מקדמים שלא שווים ל-0 נשמרים במודל, אחרים מתאפסים.

5. מודל סופי –המודל עם λ האופטימלי המאזן בין מורכבות לדיוק. בסוף התהליך מתקבל מודל Cox , שמכיל רק את המשתנים החשובים ביותר לפי הנתונים.

(לפרוט נוסף, ראו Tibshirani, 1997).

<u>בחירת משתנים בשיטת Random Forest</u>

שיטת Random Forest, שהוצגה על ידי Breiman (2001), מהווה הרחבה של שיטות מבוססות עצי החלטה ומתבססת על הרעיון של 'למידה באנסמבל' (Ensemble Learning). מדובר במושג שמתאר גישה בלמידת מכונה שבה לא מסתמכים על מודל יחיד, אלא משלבים כמה מודלים (בדרך כלל פשוטים) כדי לקבל תחזית אחת חזקה, יציבה ומדויקת יותר. בניגוד לעץ החלטה יחיד, שהוא לעיתים לא יציב ורגיש לשינויים בנתונים, Random Forest בונה יער של עצים רבים, שכל אחד מהם מאומן על דגימה אקראית שונה של הנתונים (Bootstrap) ובכל צומת נבחרת תת קבוצה אקראית של משתנים מועמדים לפיצול.

הגדרה פורמלית לפי ב-Breiman (2001):

"יער אקראי הוא אוסף של מסווגים בצורת עצים $\{h(x,\Theta k), k=1, \ldots\}$, כאשר כל עץ נבנה על בסיס וקטור אקראי שונה $\{\Theta k\}$ וכל עץ 'מצביע' על קטגוריה.

במאמרו Breiman פיתח הגדרות מתמטיות מתקדמות כגון margin function, שגיאת הכללה (generalization error) וגבולות הסתברותיים. כאן אספק יותר הסבר אינטואיטיבי.

בצורה אינטואיטיבית ופשוטה יותר, כפי שמופיע באחד מספרי היסוד בלמידה סטטיסטית - 'The Elements of Statistical Learning' (Hastie, Tibshirani & Friedman, 2009), ניתן לחשוב על Random Forest כעל ממוצע של עצי החלטה:

$$f(\hat{x}) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

כאשר $f(\hat{x})$ היא התחזית של היער עבור תצפית חדשה x, $T_b(x)$ היא התחזית של העץ ה-b, ו-B הוא מספר העצים ביער.

- ברגרסיה: התחזית היא ממוצע התחזיות מכל העצים.
- בסיווג: התחזית היא הקטגוריה שנבחרה על ידי רוב העצים.

עץ החלטה בודד הוא מודל חיזוי פשוט שמבוסס על סדרת שאלות ותשובות (פיצולים), עד שמתקבלת החלטה סופית. היתרון שלו בפשטות, אך הוא לא יציב . RF פותר את זה על ידי שילוב של הרבה עצים. השילוב של מספר רב של עצים מקטין שונות ומונע למידת יתר (overfitting) ולכן יוצר תחזיות יציבות יותר מעץ יחיד.

מעבר לדיוק החיזוי Random Forest מאפשר להעריך את חשיבות המשתנים באמצעות מדדים כגון:

- Mean Decrease in Impurity (MDI) - עד כמה כל משתנה תרם להפחתת שונות/אנטרופיה לאורך כל העצים.
- Mean Decrease in Accuracy (MDA) - ירידה בדיוק החיזוי כאשר מערבבים ערכים של משתנה מסוים (Permutation Importance).

מדדים אלו מספקים כלי לסינון משתנים חשובים, שניתן להשתמש בהם לצורך בחירת משתנים גם במודל Cox.

<u>בחירת משתנים בשיטת Stepwise עבור Cox</u>

מדובר בתהליך אוטומטי שבו בכל צעד בודקים גם הוספה של משתנה חדש למודל וגם הסרה של אחד המשתנים שכבר נכללים בו. ההחלטה אם להרחיב או לצמצם את המודל נקבעת לפי קריטריון איכות שמוגדר מראש. התהליך נעצר כשכבר אין שיפור במדד. כך מתקבל המודל האופטימלי לפי שיטה זו (1976) Hocking,.

ב- Cox אנחנו מסתמכים על לוג-נראות חלקית (Therneau & Grambsch ,2000):

$$\ell_{\mathrm{p}}(\beta) = \sum_{i:\delta_i=1}(\beta^T x_i - \log \sum_{i\in R_i} e^{\beta^T x_i})$$

במחקר זה הקריטריון העיקרי לבחירה היה מזעור AIC על בסיס הלוג-נראות החלקית של Cox. AIC (מינימיזציה), אומד את איבוד המידע ומאזן בין התאמה לנתונים לבין מורכבות. ב- Cox מחושבת על בסיס לוג-נראות חלקית (ערך נמוך יותר של AIC מעיד על מודל טוב יותר (Akaike,1974).

$$\mathrm{AIC} = -2\ell_{\mathrm{p}}(\hat{\beta}) + 2\mathrm{k}$$

כאשר k זה מספר הפרמטרים במודל. לחלופין, ניתן לשקול BIC (ענישה חזקה יותר על מורכבות) (Schwarz ,1978) או AICc (תיקון למדגמים קטנים) (Hurvich & Tsai 1,989).

כלל עצירה Stepwise (ΔAIC):

$\Delta\mathrm{AIC} = \mathrm{AIC}_{current} - \mathrm{AIC}_{candidate} \geq \varepsilon$, לעיתים $\varepsilon = 2$ (Burnham & Anderson, 2002).

כלומר, אם ה- AIC יורד לפחות ב-2 ($\Delta\mathrm{AIC} \geq 2$ לטובת המועמד), מקבלים את המהלך. כשאין שום הוספה/הסרה שמורידה את ה- AIC לפחות ב-2 - עוצרים, וזה המודל הסופי.

שלבי עבודה של אלגוריתם:

1. נקודת פתיחה - מתחילים או ממודל ריק, או ממודל בסיס כגון משתני בקרה שקובעים מראש.
2. צעד 'קדימה' - (Forward) מחשבים לכל משתנה שעדיין לא במודל מה יקרה אם נוסיף אותו. בוחרים את המשתנה שמביא את הירידה הכי גדולה ב-AIC ובתנאים שהירידה עולה על הסף שקבענו מראש, ומכניסים אותו .
3. צעד 'אחורה' (Backward) - בודקים את כל המשתנים שכבר במודל: אם הסרה של אחד מהם משפרת את AIC או שהוא אינו מובהק לפי סף שנקבע , מסירים אותו.
4. חזרה עד עצירה: חוזרים על שלבים 2-3 עד שאין הוספה/הסרה שמביאה לשיפור בקריטריון. מקובל להשתמש בכלל עצירה של $\Delta\mathrm{AIC} \geq 2$.
5. מודל סופי: המודל האחרון הוא הפתרון האופטימלי שנבחר לפי קריטריון.

4. הגדרת משתנים

4.1. אוכלוסייה ומקור הנתונים

הנתונים למחקר זה נאספו ממספר מוקדי שירות טלפוניים בישראל, וכוללים מידע אודות כ-1,800 נציגי שירות. המאגר כולל נציגים אשר עזבו את הארגון לצד נציגים שעדיין פעילים, ונאסף לאורך תקופה של שבע שנים. המשתנים שנבחנו במסגרת המחקר נבחרו בקפידה על בסיס סקירת ספרות מקיפה וידע מקצועי, במטרה לבחון את הגורמים המשפיעים על עזיבת עובדים במוקדי שירות טלפוניים בישראל, ולבנות מודל לניבוי סיכויי העזיבה עד לתקופה המוגדרת מראש (6 , 12, 18, 24 חודשים). משתני המחקר שנבחרו משקפים את מגוון ההיבטים התפקודיים, האישיים והארגוניים של העובד: נתוני העסקה, נתוני נוכחות, נתוני שכר ומדדי תפקוד (KPI) משיחות במוקד.

שלב של איסוף הנתונים היה מלווה באתגרים רבים. לרוב, האתגרים נבעו ממספר סיבות שהיו משותפות לארגונים שהיו מעוניינים להשתתף במחקר. הסיבה הנפוצה והעיקרית היא חוסר ניהול מאגר הנתונים אחיד עבור כלל התחומים (מחלקת כספים, מחלקת גיוס, הדרכה משאבי אנוש, רווחה ועוד). נוצר רושם שבגין כל עובד מנהלים הרבה מאוד מידע חשוב שניתן להשתמש בו לטובת העובד והארגון, אבל עקב חוסר סדר בניהול של המידע שנאסף, והיעדר תפקיד ייעודי בארגון שיהיה אחראי לניהול מסד נתונים אחיד בצורה מסודרת ונגישה, בפועל השימוש בו מוגבל מאוד. למשל, הנתונים מנוהלים במערכות שונות שלא מתממשקות אחת עם השנייה או מתממשקות באופן חלקי בלבד. בנוסף, ישנם נתונים רבים שמנוהלים באופן ידני באמצעות קבצי אקסל. כאמור, בחלק ניכר של הארגונים אין מסד נתונים אחד והמידע מפוזר בין הגורמים השונים בארגון. היו ארגונים שהחליפו לאחרונה מערכות וחלק ניכר מהמידע קשה לשחזור, מה שלא איפשר השתתפות שלהם במחקר זה. כל אלו הפכו את התהליך של איסוף הנתונים לקשה במיוחד. בכל שלב של הנפקת הנתונים התגלו ערכים חסרים רבים, חוסר התאמה בין נתונים ממערכות השונות, חלק ניכר של הנתונים לא היו אמינים (בגלל שמערכות שונות מדברות 'בשפות שונות' ובעת ההתממשקות הרבה מידע נאבד). מסיבות אלו, לאורך כל הדרך של איסוף נתונים היה צורך בהרבה פשרות בבחירת המשתנים למחקר. בשלב של איסוף הנתונים רוכזו תצפיות של כ-70 משתנים שונים, שמתוכם, עקב האתגרים שפורטו מעלה, בוצע עיבוד נתונים מקדים שהביא ל- 21 משתנים בלבד.

4.2. הכנת הנתונים\עיבוד מקדים של הנתונים (Data Preparation)

על מנת להפיק מהנתונים שנאספו מקסימום תועלת למחקר הנוכחי, בוצע עיבוד מקדים:

- <u>סינון עובדים בוותק מעל 3 חודשים</u>

  קיימים לא מעט מחקרים אקדמיים שמראים שהחודשים הראשונים (לעיתים מוגדרים במפורש כ-90 הימים הראשונים) הם שלב ייחודי של 'סוציאליזציה', שבו תהליכים כגון חפיפה, הכשרה, הבהרת תפקיד, תמיכה חברתית והתאמת ציפיות הם המנבאים המרכזיים של פרישה מוקדמת, בניגוד לגורמים שמשפעים על עזיבה בוותק גדול יותר בארגון (כמו שחיקה מתמשכת). לכן בוצע סינון רשומות והסרת עובדים עם ותק של פחות משלושה חודשים (לצורך נטרול השפעת תקופת הסתגלות בעת קליטה) (Saks, Uggerslev, & Fassina, 2007; Kammeyer-Mueller, Wanberg, Rubenstein, & Song, 2013).

- <u>השמטה של משתנים עם מעל 70% ערכים חסרים:</u>

  בין הנתונים שנאספו, היו נתונים רבים שנבדקו כמשתנים בעלי עניין, בהתבסס על ספרות קודמת שמצאה קשר בין הערכות ביצוע לבין עזיבה (Morrow et al., 1999) , ובין התנהלות מנהלים בארגון ומרכיבי שכר שונים לבין עזיבה (Coomber & Barriball ,2007). עם זאת, בכמות גדולה של משתנים אלו במאגר הנוכחי, בעקבות האתגרים שפורטו מעלה, נמצאו ערכים חסרים/ לא תקינים עבור כ־70% מהנציגים (החוסרים הם בטווח בין 1100 ל- 1249 נציגים מתוך 1818). דוגמא למשתנים: ציון הערכה – מטרתו הייתה להסיק על שביעות רצון של מנהל הישיר מהתפקוד/ התנהלות/ עבודת צוות ותחומים נוספים של העובד. מדובר בשאלון דיגיטאלי שמנהל ממלא עבור כל עובד שלו. השאלון כולל כ 30 שאלות על תפקודו של העובד בתחומים שונים. בסיום מילוי השאלון, המערכת מנפיקה ציון משוכלל בסולם מ 6 עד 10 בהתאם לתשובות שניתנו על ידי המנהל; שיוך למנהל צוות – מטרתו הייתה לבחון הבדלים בסיכויי עזיבה בין המנהלים השונים כדי להסיק על השפעת התנהלות מנהלים על האירוע). הכנסת משתנים אלו למודלי רגרסיה הייתה גורמת לאיבוד ניכר של נתונים (השמטת מעל שני

שליש מהמדגם) ולכן הוחלט שלא לכלול אותם במודל הסופי. הגישה תואמת את ההמלצות הסטטיסטיות בתחום, לפיהן משתנים עם שיעורי חוסרים גבוהים מאוד עלולים להטות את תוצאות הניתוח (Little & Rubin, 2019) .

- <u>השלמת ערכים חסרים</u>:

לאחר השמטה של משתנים עם מעל 70% ערכים חסרים, היה צורך לטפל בערכים חסרים במשתני KPI. כאמור, בחלק מהעובדים נמצאו ערכים חסרים רבים, שנבעו ממגבלות טכניות של מערכות המידע ומאי רציפות בתיעוד (23%-26% חוסרים). כדי למנוע איבוד משמעותי של נתונים ולשמור על כוח סטטיסטי, נדרש לבצע השלמה שתאפשר לנצל את המידע הקיים במלואו. הוחלט להשלים את הערכים החסרים באמצעות שיטת MICE (Multivariate Imputation by Chained Equations) בשילוב מודל רגרסיה מסוג Bayesian Ridge .
ההתבססות על שיטת MICE נובעת מתפיסת החסרים על פי המודל של Rubin (1987), המבחין בין שלושה סוגים של חסרים:

1. MCAR (Missing Completely At Random) - חסרים שמופיעים באקראי, ללא קשר לנתונים אחרים. הנחה זו נחשבת לא ריאליסטית במחקרי שדה.
2. MNAR (Missing Not At Random) - חסרים שתלויים בערכים שאינם נצפים. מצב זה נחשב בעייתי במיוחד, שכן לא ניתן להשלים בצורה אמינה את הערכים החסרים.
3. MAR (Missing At Random) - חסרים שתלויים רק בנתונים הנצפים. זהו מצב סביר יותר במחקרי HR ובו השתמשנו כאן. בהנחה זו ההשלמות מניבות אומדים חסרי הטיה. על בסיס הנחת MAR יושמה שיטת MICE, שכן היא מתחשבת בקשרים בין המשתנים ובכך שומרת על מבנה הנתונים המקורי.

בגלל שמשתני KPI נמדדים ביחידות שונות (למשל, משך שיחה נמדד בשניות ,שיחות לשעה במספרים עם שתי ספרות, והפסקה באחוזים), היה חשוב לבצע סטנדרטיזציה (z-score) לכל משתני KPI כדי למנוע הטיות $z = \frac{x-\mu}{\sigma}$.

לאחר מכן, מתבצעת השלמה לכל משתנה חסר בתורו באמצעות בניית מודל רגרסיה המנבא את הערך החסר על סמך יתר משתנים. אחרי שמשלימים את כל הערכים החסרים, מחזירים את הנתונים לסקאלה המקורית ,כדי לשמור על פרשנות ביחידות המקוריות (שניות, שיחות לשעה, אחוזים). במחקר הנוכחי, השיטה יושמה בנפרד עבור כל מוקד שירות. כלומר, קודם הנתונים פוצלו לפי מזהה המוקד ואז כל קבוצה הושלמה בפני עצמה, כדי להתאים למבנה הייחודי ודפוסי הביצועים של כל מוקד.

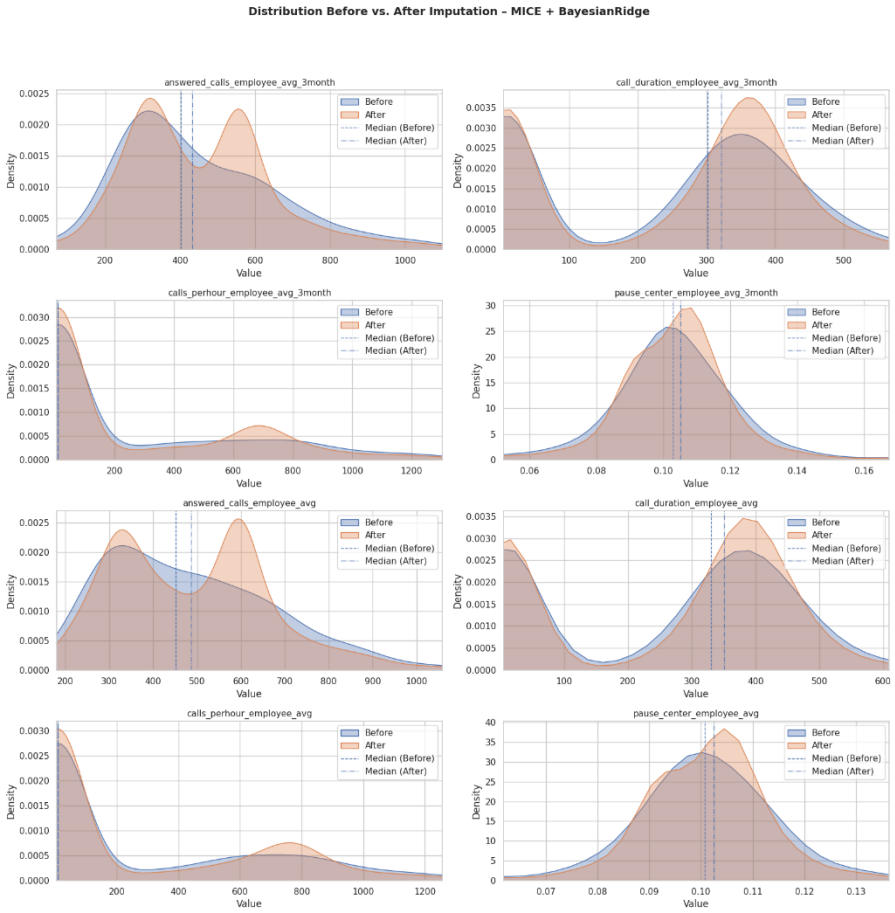
חשוב להדגיש שוב כי במחקר זה נעשה שימוש בשיטת MICE. ההנחה המרכזית העומדת בבסיס השיטה היא כי הנתונים החסרים מקיימים את הנחת - MAR. כלומר, ההסתברות לכך שערך מסוים חסר תלויה רק בערכים הנצפים (שאינם חסרים), ולא בנתונים שאינם נצפים. תחת הנחה זו, ההשלמות משקפות את המבנה הקיים בנתונים ואינן מייצרות הטיה שיטתית.

השיטה נבחרה על פני גישות פשוטות יותר(כגון KNN או מילוי בחציון), שכן היא מאפשרת לנצל מידע ממספר רב של משתנים ולשמר את הקשרים ביניהם. במודל האמידה נעשה שימוש ב- Bayesian Ridge, המתמודד היטב עם בעיות של קולינאריות ומייצב את אמדני המקדמים. בהקשר זה חשוב להדגיש כי המודל מניח התפלגות אפריורית רגולרית למקדמים (Normal prior) עם ענישה ריבועית (כלומר האמידה משלבת את הנתונים בפועל עם הנחת 'ריכוך' מראש, המקטינה שונות ומייצבת את האומדנים (Van Buuren & Groothuis-Oudshoorn, 2011).

תהליך ההשלמה בוצע עבור ארבעה מדדי KPI כמותיים מרכזיים (מספר השיחות שנענו, משך השיחה הממוצע, מספר השיחות לשעה ואחוז הפסקה מתוך שעות עבודה), הן ברמת העובד (employee_avg) והן ברמת המוקד (center_avg) . בכך שוחזרו הערכים החסרים באופן עקבי עם דפוסי הנתונים הקיימים, ולא נותרו כמעט ערכים חסרים. במקרים נדירים בהם מדד מסוים היה חסר לחלוטין בכל הרשומות של מוקד מסוים, ההשלמה נעשתה באמצעות חציון כללי כפתרון גיבוי.

יש להדגיש כי שום שיטת השלמה (imputation) אינה משחזרת מידע אבוד אלא רק נתונים חסרים , ולכן עצם ההשלמה אינה יוצרת מידע חדש. התהליך עשוי להפחית באופן מסוים את הפיזור ואת ערכי הקיצון, ועל כן כלול לגרום לאמידת חסר של סטיות התקן המדגמית. עם זאת, יתרונו המרכזי בכך שהוא מאפשר ניתוח מלא של כלל המשתנים תוך שמירה על דפוסי הקשרים המקוריים.

לאחר ההשלמה של ערכים החסרים בוצעה השוואת התפלגויות (לפני ואחרי ההשלמה) בכל אחד משש המדדים. ההשוואה בוצעה באמצעות מבחן קולמוגורוב–סמירנוב (KS). מוצגים מטה כמה תרשימים לצורך הדגמה:



תרשים 1: **משתני KPI - התפלגות לפני ואחרי השלמת ערכים חסרים**

הגרפים מציגים צפיפות מדגמית (Kernel Density) של כל מדד KPI לפני ההשלמה (כחול) ואחריה (כתום). הקווים המקווקווים מסמנים את החציון לפני ואחרי ההשלמה.
ניתן לראות בגרפים כי ברוב המדדים הקווים הללו מתלכדים או קרובים מאוד זה לזה, כלומר החציון נשמר. בנוסף, קיימת חפיפה גבוהה בין העקומות, וצורת ההתפלגות הכללית נשמרת, למעט סטיות קטנות בזנבות ובערכי קיצון.
כלומר, השלמה בשיטת MICE - לפי שיוך למוקד שמרה על המיקום והצורה הכללית של ההתפלגויות והפחיתה במתינות את הפיזור וערכי קיצון - ללא הסטה של המרכז כולו. ממצאים אלו תומכים בכך שההשלמה לא עיוותה את מדדי ה- KPI, אלא שיחזרה ערכים חסרים באופן עקבי עם המבנה הקיים בנתונים.
עם זאת, יש לציין כי גם אם האומדים המרכזיים (ממוצע, חציון) חסרי הטיה, ההשלמה עשויה לגרום להערכת חסר של סטיית התקן, שכן היא מפחיתה את השונות הנצפית. המשמעות היא שהפיזור עשוי להיראות קטן מעט מכפי שהיה אילו כל הנתונים היו נצפים בפועל.

- <u>יצירת משתנים חדשים</u>

נתוני KPI - לאחר השלמת ערכים חסרים ל 12 משתני הביצוע (KPI) שהיוו רכיב קריטי במודל המחקר, כגון ממוצע משך השיחות, מספר שיחות לשעה, מספר שיחות נענות, כפי שהם נרשמו במערכות טלפוניה במוקדים, הוחלט ליצור מהם משתנים חדשים. בעיקרון, מדובר ב- 4 משתני ביצועים תפעוליים שונים (רציפים), שכל אחד מהם נאסף שלוש פעמים (כל פעם בהסתכלות שונה): נתונים בגין ממוצע שנתי עבור העובד, נתונים בגין ממוצע 3 חודשים לפני האירוע/ צנזורה, ממוצע שנתי בגין המוקד שאליו העובד משויך. מטרת האיסוף הייתה ליצור מכל השדות שנאספו משתנים חדשים, שמהם ניתן יהיה להסיק על מגמת שינוי בתפקוד העובד/ בהתנהלות משמעתית של העובד/ בעמידה ביעדי המוקד ועוד. לצורך כך יצרנו משתנים חדשים על ידי חישוב פער/יחס בין המשתנים השונים על פי פירוט מטה:

אינדיקציה לירידה/ עליה בתפקוד העובד ב- 3 חודשים אחרונים לפני האירוע/ צנזורה = ממוצע של ביצוע מדד על ידי העובד ב 3 חודשים אחרונים לפני האירוע/ צנזורה - ממוצע שנתי של ביצוע מדד על ידי העובד

אינדיקציה לעמידה/ אי עמידה בממוצע של המוקד ב- 3 חודשים אחרונים לפני האירוע/ צנזורה = ממוצע של ביצוע מדד על ידי העובד ב 3 חודשים אחרונים לפני האירוע/ צנזורה - ממוצע שנתי של ביצוע מדד על ידי המוקד שאליו העובד משויך

אינדיקציה לשינוי באופן התנהלות משמעתית של העובד ב- 3 חודשים אחרונים לפני האירוע/ צנזורה = ממוצע % הפסקה מתוך סך * שעות לוגין ב 3 חודשים אחרונים / ממוצע % הפסקה מתוך סך שעות לוגין בשנה אחרונה

* שעות לוגין - כל זמן שהנציג מחובר למערכת המוקד במהלך המשמרת (או בתקופת מדידה) נספר כלוגין ,בלי תלות אם הוא פנוי או מטפל בשיחה . בזמן זה הנציג עשוי להיות במצבים שונים – מוכן לשיחה, בשיחה, בתדריך של הבוקר ,ממתין לשיחה, או "לא מוכן" מסיבה מוגדרת כגון,הפסקה, הדרכה וכו'(Gans, Koole & Mandelbaum, 2003).

חישובים של פער ויחס במדדים תפעוליים (KPI) מקבלים תמיכה אמפירית במחקרים רבים בתחום ניבוי עזיבה וזיהוי דפוסי התנהגות טרם עזיבה (Enders & Tofighi, 2007; Singer & Willett, 2003 ; Valle & Ruz, 2015; Valle, Varas & Ruz, 2012; Morrow et al., 1999).

נתוני שכר - מאגר הנתונים כולל נתון של שכר שעתי. שימוש בנתון בפני עצמו עלול לייצר הטיה, מכיוון שמדובר בנציגי שירות שלרוב מקבלים שכר מינימום הקבוע בחוק. כידוע, ב־2017 שכר המינימום היה נמוך בהרבה מאשר ב־2024. לכן, עובד 'זול' יחסית ב־2024 יכול להיראות 'עשיר' לעומת עובד מ־2017, למרות שבפועל כוח הקנייה/רמת השוק שונה. זה מייצר אפקט מדומה כאילו שכר קשור לעזיבה, כשבפועל זה יכול להיות פשוט אפקט הזמן. מסיבה זו וכדי עדיין לנצל את הנתון ולבדוק השפעה של שכר על סיכויי העזיבה, נוצר משתנה חדש תוך כדי שימוש בנתונים נוספים ('פרמיה ממוצעת חודשית' ושכר מינים בשוק כפי שפורסם מעת לעת על ידי משרד העבודה). באמצעות שלושת הנתונים בוצע חישוב לצורך יצירת משתנה 'שכר יחסי', ובכך לבצע נורמליזציה שמאפשרת השוואה נוחה בין עובדים בכל השנים.
להלן שלבי החישוב שבוצע:

- חישוב **שכר חודשי כולל** (למשרה מלאה) = (שכר לשעה × 1.78*) + פרמיה ממוצעת חודשית.

**1.78 – ממוצע של שעות תקן חודשי לצורך חישוב שכר חודשי לעובדים בשכר שעתי. שיטה שמקובלת על ידי חשבי שכר בישראל.*

- **נרמול לפי שכר מינימום באותה שנה כפי שפורסם על ידי משרד העבודה (שנת עזיבה/שנת הנפקת דוחות בגין פעילים)** כלומר:

שכר יחסי = שכר חודשי למשרה מלאה / שכר מינימום באותה שנה

קיבוץ קטגוריות וקידוד ב- Excel לפני טעינה ל-Python

כל פעולות הקיבוץ, האיחוד והקידוד של המשתנים הקטגוריאליים בוצעו מראש בחוברת Excel , עוד לפני טעינת הנתונים לסביבת ה- Python. הבחירה לבצע את השלבים הללו באקסל נועדה להבטיח עקביות בין מקורות המידע השונים ולשפר שקיפות וביקורתיות (ניתן לראות ולערוך את טבלאות המיפוי ישירות בחוברת). לצורך שחזור מלא של התהליך נשמרה טבלת מיפוי בחוברת הנתונים, הכוללת: שמות המשתנים, רמות המקור, הקידוד לאחר האיחוד.

עם זאת, יש לציין כי נתוני העובדים עצמם חסויים ואינם נגישים לציבור. מסירתם תתבצע רק אם תעלה דרישה מפורשת ורק באישור מוקדי השירות, בהתאם לכללי הגנת הפרטיות והמדיניות הארגונית.

- <u>קיבוץ קטגוריות דלילות במשתנים קטגוריאליים</u>:

בשלב עיבוד הנתונים אוחדו קטגוריות בעלות שכיחות זניחה במשתנים קטגוריאליים, על-מנת לייצב את האמידה במודל Cox, למנוע סטיות תקן מנופחות ולצמצם סיכון לבעיות 'היפרדות' (למשל קטגוריה ללא אירועים) וליחס אירוע. לדוגמא, במשתנה מצב משפחתי, שכלל במקור 6 קטגוריות (גרוש, אלמן, רווק ועוד), שחלקן הופיעו במקרים בודדים בלבד, בוצע קיבוץ לשתי תתי משפחות תוכן רחבות יותר (נשוי/ לא נשוי) ובמקביל נשמרה קטגוריית "אחר/לא ידוע" כדי לא לאבד מידע. דוגמא נוספת, שיוך למנהל צוות. החלוקה למנהלי צוותים התגלתה כלא אמינה עקב בעיות טכנולוגיות בשליפת הנתונים ולכן לא ניתן היה להשתמש בנתון זה. כדי לא לאבד מידע נבדקה כמות של נציגים בכל אחד מהצוותים ובוצעה חלוקה לקטגוריות על פי גודל הצוות ולא על פי שיוך למנהל (למשל 10-15 נציגים, 15-20 נציגים ועוד). הקיבוץ בוצע לפי מה שמקובל במוקדי שירות טלפוניים, לצורך פשטות הפרשנות והפחתת דימוי ממדים מיותרים לפני קידוד. בכל האיחודים נבדקה קירבה מושגית בין רמות ונלקח בחשבון שיקול סטטיסטי (הבטחת מספר תצפיות/אירועים מספק לכל רמה), תוך שמירת טבלת מיפוי מלאה והרצה עד שלא נותרו רמות עם אפס אירועים. הפרקטיקה תואמת הנחיות מתודולוגיות בניתוח רגרסיות הישרדות/כלליות הממליצות לאחד קטגוריות נדירות כדי לשפר יציבות, הפיכות ועוצמה סטטיסטית של המודל (Harrell, 2015; Vittinghoff & McCulloch, 2007).

- <u>קידוד משתנים קטגוריאליים לערכים מספריים</u>

בשלב עיבוד הנתונים בוצעה המרה של המשתנים הקטגוריאליים לערכים מספריים, תוך ביצוע הבחנה בין משתנים סדרתיים (כגון חודש/ שנה) לבין משתנים נומינליים. במשתנים סדרתיים נשמר סדר טבעי באמצעות קידוד בסדר עולה , ומשתנים נומינליים קודדו על פי גודל מנמוך לגבוה (ר' פירוט בפרק 4.3). במשתנה הבינרי, כגון מצב משפחתי, נעשה קידוד 0/1: 0=לא נשוי ו-1=נשוי . טבלת מיפוי מלאה נשמרה לכל משתנה, והקידוד שימר את הסדר כאשר הוא בעל משמעות, בעוד שבמודל עצמו (Cox) הוסבו משתנים נומינליים ל- one-hot.

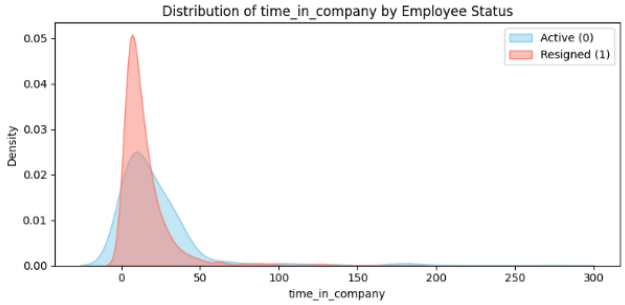
one-hot. הוא אופן קידוד למשתנה נומינלי (ללא סדר), שבו כל קטגוריה הופכת לעמודת 'דגל' נפרדת עם ערכים 0/1. בשורה של נבדק יש 1 רק בעמודה של הקטגוריה שלו ו-0 בשאר. במודל Cox כל עמודה מקבלת מקדם נפרד, וכשמוציאים אחת כקטגוריית ייחוס (k-1), כל מקדם מתפרש כ- HR של אותה קטגוריה ביחס לקטגוריית הייחוס.

חשוב לציין כי משתני resignation_reason / resignation_month / resignation_year סומנו כ- post-event ולא הוכנסו למודל הניבוי כדי למנוע דליפת מידע. הקידוד שלהם שימש לתיאור ולבקרת איכות בלבד.

5. הגדרת משתנים

משתה תלוי

- **משך הזמן בחודשים עד לעזיבה/ צנזורה (time_in_company):** משתנה רציף, המודד את מספר החודשים מתחילת ההעסקה ועד לעזיבה (עבור עובדים שסיימו את המעקב כשהם עדיין פעילים, מדובר בזמן מתחילת ההעסקה ועד למועד של סיום המעקב). המחקר כלל עובדים בעלי ותק של מעל 3 חודשים בלבד, במטרה לנטרל השפעת עזיבות בשלבי קליטה והדרכה.



תרשים 2 : **התפלגות משך הזמן בחודשים עד לעזיבה/ צנזורה**

כלל המשתנים שנבדקו במחקר

- **סטטוס העובד (status):** משתנה דיכוטומי, המהווה אינדיקטור למאורע (הבחנה בין מאורע לצנזורה). המשתנה קודד (0=פעיל, 1=עזב).

<u>נתוני העסקה</u>

- **מוקד שאליו משויך העובד (call_center)** : מאגר נתונים כולל תצפיות של נציגים מ 7 מוקדים שונים. אף שמדובר בשבעה מוקדים ספציפיים, במחקר זה המשתנה הומר למדד סדרתי של גודל מוקד :כל מוקד שובץ לאחת משבע דרגות גודל מקובלות בשיח של מוקדי שירות (1=קטן מאוד, 7=גדול מאוד), כך שהקידוד משקף סדר עולה בגודל ולא זהות נומינלית. הבחירה לייצג את המוקד באמצעות דרגת גודל נועדה ללכוד השפעה מערכתית של קנה-מידה (Scale) על סיכון העזיבה, ולהימנע מהטלת פרמטר נפרד לכל מוקד (one-hot)בשעה שמטרתנו העיקרית היא פרשנות ותכלול. כדי להבטיח יציבות אמידה, הוגדר סף ייצוג מספיק של אירועים בכל דרגת גודל (5% מהמדגם) ,ונערכה בדיקת מגמה-קוקס חד-משתני עם קידוד סדרתי. נמצא יתרון מובהק למודל הקטגורי, נשמר הייצוג הסדרתי (חסכוני בפרמטרים וקל לפרשנות).
- **גודל הצוות שאליו משויך העובד (Team_size):** במאגר הנתונים ישנה חלוקה ליותר מ-60 צוותים על פי שיוך למנהל צוות, שבדיעבד התגלה שרוב המוחלט של צוותים לא קיימים כבר בפועל וחלק ניכר ממנהלי הצוותים כבר לא עובדים במוקד. מי שמשויך לצוותים הנ"ל אלה עובדים שכבר עזבו בשנים הקודמות בלבד. על מנת להפוך את הנתון למשתנה שניתן יהיה להשתמש בו, הוגדר לכל צוות גודל ממוצע של נציגים (לצוותים שנסגרו, הוגדר הגודל בעת פעילותם, לפי דיווח מנהל המוקד), ולאחר מכן בוצעה חלוקה לחמש דרגות סדרתיות המקובלות במוקדי שירות. (10-15 נציגים, 15-20 נציגים וכו'). כאמור, הקידוד בוצע על פי גודל הקבוצה כפי שנמסר על ידי מנהל המוקד שאליו הצוות היה משויך, מנמוך לגבוה, סה"כ 5 קטגוריות מקודדות 1-5. כדי להבטיח יציבות אמידה הוגדר מראש סף ייצוג מספק של אירועים (עזיבות) לכל קטגוריה (לפחות 5% מן המדגם), בפועל כל הקטגוריות עמדו בדרישות הסף. לצורך ולידציה של החלוקה ,הורץ מודל קוקס חד משתני, נמצאו הבדלים מובהקים סטטיסטית בין הדרגות ,אך גודלי האפקט קטנים מאוד . לכן, הוחלט לשמור על חלוקה לקטגוריות כפי שנעשתה לצורך פרשנות קלה יותר למנהלים.
- **סיבת עזיבה (resignation_reason):** ערך קיים עבור עובדים שעזבו בלבד. משתנה ששימש לתיאור ולא הוכנס למודל ההישרדות. במאגר הנתונים קיימים 6 סיבות עזיבה. למשתנה בוצע קידוד על פי כמות עובדים שמשויך לכל אחד מהקטגוריות (מנמוך לגבוה). טווח המספרים של הקידודים מופיע בטבלה מטה (0-5).
- **שנת עזיבה (resignation_year)** : ערך קיים עבור עובדים שעזבו בלבד. משתנה ששימש לתיאור ולא הוכנס למודל ההישרדות. משתנה סדרתי שנשמר עבורו סדר טבעי באמצעות

קידוד על פי סדר עולה. טווח מספרים של הקידודים מופיע בטבלה מטה (0-7).

- **חודש עזיבה (resignation_month)**: ערך קיים עבור עובדים שעזבו בלבד. משתנה ששימש לתיאור ולא הוכנס למודל ההישרדות. משתנה סדרתי שנשמר עבורו סדר טבעי באמצעות קידוד על פי סדר עולה. טווח מספרים של הקידודים מופיע בטבלה מטה (0-11).

נתוני שכר

- **שכר יחסי (relative_salary)**: משתנה רציף, שנוצר באמצעות חישוב של יחס בין שכר עובד למשרה מלאה (גם אם בפועל עובד במשרה חלקית) לבין שכר מינימום בשוק בתקופה הרלוונטית. נמדד בשקלים. הנתון מהווה אינדיקציה לשיעור שכר העובד מתוך שכר מינימום (מעל 1 משמעות שעובד מרוויח מעל המינימום ומתחת ל-1 שהוא מרוויח פחות ממינימום)
- **ממוצע חודשי של פרמיה (bonus_avg):** בכל מוקד קיים מודל בונוס וכל נציג מקבל את הפרמיה בהתאם לתפקודו ועמידה ביעדי המודל. הנתון משקף ממוצע חודשי של גובה הפרמיה שחושב לשנה אחרונה. הנתון רציף ונמדד בשקלים.
- **אחוז משרה (part_time_percent)**: משתנה רציף

נתוני נוכחות

- **האם עלה ניצול ימי חופשה בשלושת החודשים שלפני עזיבה/מועד סיום מעקב (avg_annual_vacation_increase):** ההשוואה בוצעה אל מול ממוצע ניצול חופשה השנתי. משתנה קטגוריאלי שקודד (0=כן, 1=לא).
- **האם עלה ניצול ימי מחלה בשלושת החודשים שלפני עזיבה/מועד סיום מעקב (avg_annual_sick_increase):** ההשוואה בוצעה אל מול ממוצע ניצול מחלה השנתי. משתנה קטגוריאלי שקודד (0=כן, 1=לא).

נתונים אישיים

- **מצב משפחתי (family_status)**: משתנה שבמקור היה עם כמה קטגוריות ומספר הקטגוריות צומצם ל – 2. משתנה קטגוריאלי שקודד (0=נשוי, 1=לא נשוי).
- **מגדר (gender)**: משתנה קטגוריאלי שקודד (0=זכר, 1=נקבה).
- **גיל בשנים (age):** משתנה רציף .
- **האם לעובד יש ילדים מתחת לגיל 18 או לא (children)**. משתנה קטגוראלי שקודד (0=אין ילדים, 1=יש ילדים)

נתוני ביצועים

- **הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות במוקד כולו (answered_calls_vs_center_avg)**: משתנה רציף (מספר שלילי מהווה אינדיקציה לאי עמידה של העובד בממוצע של המוקד).
- **הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות של העובד בשנה אחרונה (answered_calls_vs employee_avg)**: משתנה רציף (מספר שלילי מהווה אינדיקציה לירידה בתפקוד של העובד בכמות שיחות נענות).
- **הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה במוקד כולו (call_duration_vs_center_avg):** משתנה רציף ונמדד בשניות (מספר שלילי מהווה אינדיקציה לאי עמידה של העובד בממוצע של המוקד).
- **הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה של העובד בשנה האחרונה (call_duration_vs_employee_avg):** משתנה רציף ונמדד בשניות (מספר שלילי מהווה אינדיקציה לירידה בתפקוד במשך שיחה של העובד).

- **הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה במוקד כולו (calls_perhour vs_center_avg):** משתנה רציף (מספר שלילי מהווה אינדיקציה לאי עמידה של העובד בממוצע של המוקד).
- **הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה של העובד בשנה אחרונה (calls_perhour_vs_employee_avg)** : משתנה רציף (מספר שלילי מהווה אינדיקציה לירידה בתפקוד של העובד).
- **יחס בין ממוצע אחוז זמן הפסקה מתוך שעות לוגין של העובד ב-3 חודשים אחרונים לפני סיום מעקב לבין ממוצע של העובד בשנה אחרונה (pause_self3m_vs_selfannual):** משתנה רציף (מספר גדול מ-1 מהווה אינדיקציה להתדרדרות בהתנהגות משמעתית של העובד, נוטה לצאת ליותר הפסקות במהלך שעות העבודה).

<u>טבלה משתנים מסכמת</u>

משתנה	קידוד/טווח ערכים	כמות ערכים חסרים
משך הזמן בחודשים עד לעזיבה או לצנזורה (time_in_company)	מ 3 ועד 274.13	0
סטטוס העובד (status)	0 = עובד פעיל 1= עזב	0
מוקד שאליו משויך העובד (call_center)	1 =מוקד קטן מאוד 2 = קטן 3 =בינוני קטן 4 =בינוני 5 =בינוני גדול 6 =גדול 7 =גדול מאוד	0
גודל הצוות שאליו משויך העובד (Team_size)	1 =עד 10 נציגים 2 =בין 10 ל 15 נציגים 3=בין 15 ל 20 נציגים 4 =בין 20 ל 25 נציגים 5 =מעל 25 נציגים	0
שכר יחסי (relative_salary)	מ 0.94 ועד 2.22	0
ממוצע חודשי של פרמיה (bonus_avg)	מ 0 ועד 5,780.63	0
אחוז משרה (part_time_percent)	מ 0.03 ועד 1.77	0
האם עלה ניצול ימי חופשה ב-3 חודשים לפני סיום מעקב (avg_annual_vacation_increase)	0 = כן 1= לא	0
האם עלה ניצול ימי מחלה ב-3 חודשים לפני סיום מעקב (avg_annual_sick_increase)	0 = כן 1= לא	0
מצב משפחתי (family_status)	0 = לא נשוי 1 =נשוי	0
מגדר (gender)	0 = זכר 1 = נקבה	0
גיל בשנים (age)	מ 17 ועד 61	0
האם לעובד יש ילדים או אין ילדים (children)	0 = ללא ילדים 1 =עם ילדים	0
הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות במוקד כולו (answered_calls_vs_ center _avg)	מ 647.67- ועד 948.557	462
הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות של העובד בשנה אחרונה (answered_calls_vs_employee_avg)	מ 745.53- ועד 636.17	597
הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע	מ 277.19- ועד 1,488.26	462

משך השיחה במוקד כולו (call_duration_vs center_avg)		
הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה של העובד בשנה האחרונה (call_duration_vs_employee avg)	מ -273.66 ועד 99.93	597
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה במוקד כולו (calls_perhour_vs_center_avg)	מ -759.67 ועד 1,437.67	462
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה של העובד בשנה אחרונה (calls_perhour_vs employee_avg)	מ -848.90 ועד 494.67	597
יחס בין ממוצע אחוז זמן הפסקה מתוך שעות לוגין של העובד ב-3 חודשים אחרונים לפני סיום מעקב לבין ממוצע של העובד בשנה אחרונה (pause_self3m_vs_selfannual)	מ 0 ועד 2.32	463

טבלה 1 : **טבלת משתנים מסכמת**

6. ממצאים

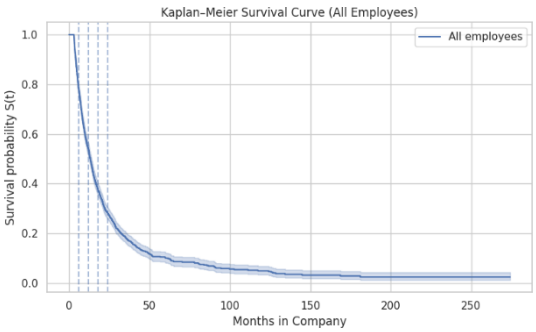
6.1. תיאור המדגם והאירוע

המדגם כולל N=1,751 נציגים (לאחר ניקוי). כפי שנשלפו ממערכות ה־HR בסוף שנת 2024. המדגם הורכב משתי קבוצות : (1) כלל העובדים שעזבו את החברה במהלך 7 שנים שקדמו למועד הנפקת הדוחות, ו־(2) כלל העובדים שהיו פעילים במערכת נכון למועד ההנפקה.

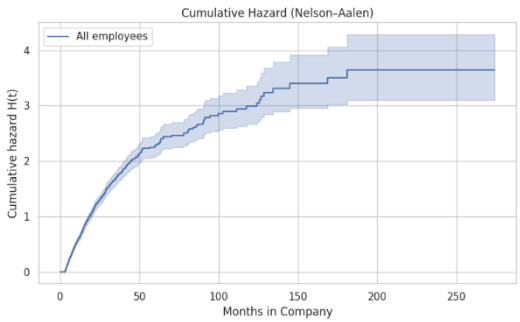
עבור כל עובד היה זמין המשתנה **ותק במוקד** (time_in_company), אשר שיקף את משך הזמן המצטבר מתחילת ההעסקה ועד לנקודת הסיום (עזיבה או צנזורה). מידע זה אפשר לחשב רטרוספקטיבית את משך הזמן עד לעזיבה (עבור העובדים שעזבו במהלך שבע השנים האחרונות) או עד לנקודת הצנזורה (עבור העובדים שהיו פעילים בסוף התקופה).

בכך, אף שהמחקר מבוסס על צילום מצב יחיד של כלל העובדים בסוף 2024, השימוש במשתנה הוותק סיפק את האינפורמציה הדרושה לביצוע ניתוח הישרדות מלא במסגרת מודל Cox. בסך הכול תועדו 1,339 אירועי עזיבה ו־ 412 עובדים מצונזרים (נותרו פעילים במועד ההפקה).

<u>הישרדות כללית (Kaplan–Meier & Nelson–Aalen)</u>



תרשים 3 : **עקומת Kaplan-Meier (KM) לכלל העובדים, עם רצועות סמך 95%.**



תרשים 4 : **סיכון מצטבר (Nelson-Aalen) - ממחיש את "קצב" ההתרחשות לאורך זמן.**

עקומת קפלן- מאייר (איור 3) מתארת את הסתברות השרידות לכלל העובדים לאורך תקופת המעקב. כבר בתחילת המעקב נצפית ירידה חדה בשרידות. לאחר חצי שנה של מעקב ההסתברות המותנית להישארות היא כ- 78% ,(S(6)=0.78) . ולאחר שנה כ- 54% בלבד(S(12)=0.54) .

חשוב להדגיש כי ערכי ההישרדות המוצגים אינם מתייחסים בהכרח לתקופת העבודה הראשונה של כל עובד, שכן חלק מהעובדים נכנסו למעקב עם ותק שונה. במודל Coxהוותק (time_in_company) שימש כמשתנה התלוי, ולכן לא ניתן היה לכלול אותו גם כמשתנה מסביר. יחד עם זאת, ניתן היה לשקול בעתיד לבחון משתנה נוסף של 'ותק בתחילת המעקב' כמשתנה מסביר, על מנת לבדוק האם לעובדים שנכנסו עם ניסיון מצטבר ארוך יותר יש פרופיל הישרדות שונה. (רווחי סמך בטבלה 2).

טבלה 2 : הסתברויות שרידות בנקודות זמן מרכזיות

חודשים	S(t)	95% CI
6	0.780	[0.759, 0.799]
12	0.539	[0.515, 0.563]
18	0.374	[0.350, 0.398]
24	0.282	[0.259, 0.305]

עקומת הסיכון המצטבר (Nelson-Aalen , איור 4) מדגישה שהסיכון לעזיבה מצטבר בקצב מהיר במיוחד בשנה הראשונה, ולאחר מכן ממשיך לעלות, אך בשיפוע מתון יותר. ככל שהזמן חולף, רווחי הסמך מתרחבים. זה מהווה עדות לאי ודאות גבוהה יותר לגבי השרידות בקרב עובדים המעטים ששורדים לאורך שנים רבות.

לסיכום, החודשים הראשונים של המעקב, ובעיקר השנה הראשונה, מהווים חלון סיכון קריטי. יש להדגיש כי מדובר בפרק הזמן שנמדד במסגרת המעקב (2017–2024), ועובדים שונים נכנסו אליו עם רמות ותק שונות. אף על פי כן, מרבית העזיבות התרחשו בשלב זה, דבר שמצביע על החשיבות של אסטרטגיות קליטה, חפיפה, ליווי ותמריצים כבר בתקופה זו.

6.2. הבדלים בסיסיים בין קבוצות (תיאור, לפני המודל)

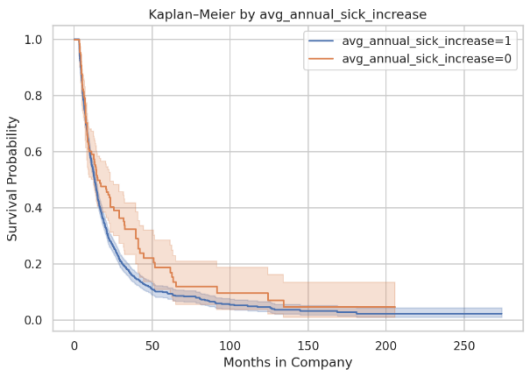
גודל צוות (Team_size)

נבדקו עקומות הישרדות לפי חמש קטגוריות גודל צוות (תרשים 5 בנספח). התפלגות העקומות כמעט זהה, ולא נמצא הבדל מובהק סטטיסטית במבחן(Log-rank (p≈0.42 . כלומר, גודל הצוות כשלעצמו אינו משפיע על סיכון העזיבה, ולכן לא מהווה משתנה קריטי למדיניות שימור.

עליה בניצול ימי מחלה ב-3 חודשים לפני סיום המעקב (avg_annual_sick_increase)

נבדקו עקומות הישרדות לפי עלייה או אי עלייה בניצול ימי מחלה בשלושת החודשים שקדמו לסיום המעקב (תרשים 6). נמצא כי לעובדים עם עלייה בשימוש בימי מחלה סיכון גבוה יותר לעזוב מוקדם, בהשוואה לעובדים שלא הגדילו שימוש (Log-rank p≈0.04) .

עלייה חריגה בניצול ימי מחלה בחודשים האחרונים עלולה לשמש אינדיקציה מוקדמת לשחיקה או חוסר מחויבות .איתור תבנית כזו בזמן אמת יכול לאפשר התערבות ניהולית - שיחה אישית, בחינת עומס העבודה או הצעת תמיכה נוספת - לפני שהעובד אכן מחליט לעזוב.



תרשים 6 : גרף Kaplan - Meier לפי המשתנה avg_annual_sick_increase

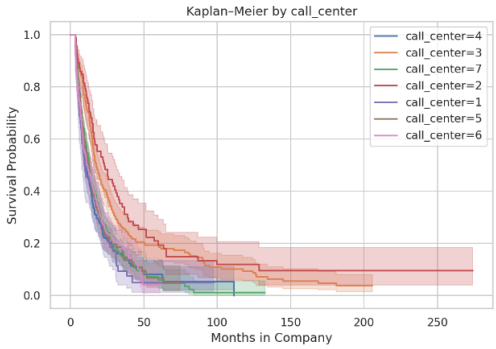
עליה בניצול ימי חופשה ב-3 חודשים לפני סיום המעקב (avg_annual_vacation_increase):

נבדקו עקומות הישרדות לפי עלייה או אי-עלייה בניצול ימי חופשה בשלושת החודשים שקדמו לסיום המעקב (תרשים 7 בנספח). נמצא הבדל גבולי : עובדים שלא הגדילו את ניצול ימי החופשה שרדו מעט יותר לאורך זמן בהשוואה לעובדים שהגדילו שימוש (Log-rank p≈0.05).

משמעות ניהולית : אף כי מבחינה סטטיסטית ההבדל מובהק, מבחינה מעשית מדובר בהשפעה קטנה בלבד. ייתכן כי עלייה בשימוש בחופשה משקפת נסיבות חיצוניות (חופשה משפחתית, אירוע

אישי) ולא בהכרח שחיקה בעבודה. לפיכך, אינדיקטור זה פחות חזק לשימור עובדים בהשוואה לניצול ימי מחלה.

מוקד (call_center)



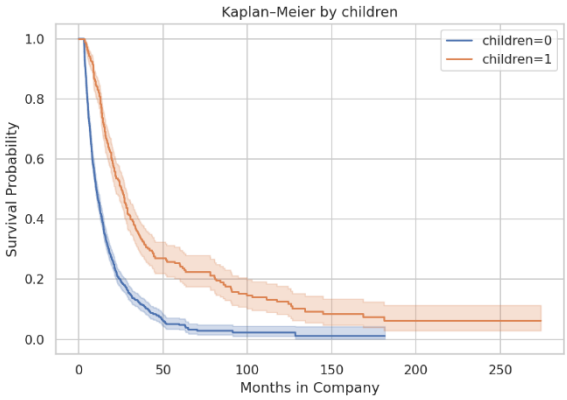
תרשים 8 : **גרף Kaplan-Meier לפי מוקד (call_center)**

להבדיל מהמשתנים הקודמים, כאן רואים פערים ברורים בין המוקדים. מוקד 2 (אדום) מציג שרידות גבוהה בהרבה מהשאר, גם אחרי 50+ חודשים. מוקדים אחרים (1, 3, 4, 5, 6, 7) עקומות יורדות מהר יותר, עם דפוסי הישרדות שונים. הפערים לא רק ויזואליים: מבחן Log-rank בין הקבוצות יוצא מובהק מאוד ($p<0.001$).

עקומות הישרדות לפי מוקד מצביעות על שונות בולטת (תרשים 8). למשל, במוקד 2 החציון עד לעזיבה גבוה בהרבה (≈16 חודשים) לעומת מוקדים אחרים שבהם החציון נע סביב 7-9 חודשים. מבחן Log-rank מחזק הבדל מובהק סטטיסטית בין המוקדים ($p<0.001$).

ההבדלים בין מוקדים מעידים כי לא רק מאפיינים אישיים או תפעוליים משפיעים על הישרדות העובדים, אלא גם גורמים ארגוניים ברמת המוקד (תרבות הניהול, עומס, תהליכי חפיפה). ממצא זה פותח פתח ללמידת שיטות ממוקדים שמצליחים לשמר עובדים לאורך זמן, והטמעתם במוקדים אחרים.

ילדים (עם ילדים / בלי ילדים) (Children)

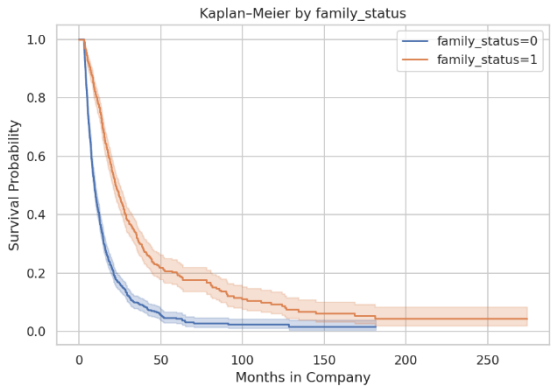


תרשים 9 : **גרף Kaplan-Meier לפי נוכחות ילדים**

עקומות הישרדות לפי נוכחות ילדים (תרשים 9) מראות הבדל ברור. לעובדים עם ילדים שיעור שרידות גבוה יותר באופן מובהק לאורך כל תקופת המעקב. מבחן Log-rank מצביע על הבדל מובהק סטטיסטית ($p<0.001$).

כלומר, עובדים עם ילדים הם קבוצה יציבה יחסית, עם סיכוי נמוך יותר לעזיבה בכל שלבי המעקב. מבחינה ניהולית ניתן לפרש ממצא זה בשתי דרכים משלימות: מצד אחד, ניתן להשקיע במדיניות תומכת-משפחה (כגון גמישות בשעות עבודה, אפשרות עבודה חלקית מותאמת, סיוע ברווחה) כדי להמשיך לשמר קבוצה זו לאורך זמן, כחלק מהאסטרטגיה לחיזוק ההון האנושי היציב. מצד שני, ניתן לטעון שהאתגר האמיתי נמצא דווקא בקרב עובדים ללא ילדים, אשר נוטים לעזוב מוקדם יותר ולכן מהווים אוכלוסייה פגיעה יותר. עבור קבוצה זו יש חשיבות לפתח תמריצים אחרים - למשל בונוסים כספיים, מסלולי קידום מהירים או פעולות שמחזקות תחושת שייכות ומחויבות לארגון.

מצב משפחתי (נשוי / לא נשוי) (Family Status)



תרשים 10 : **גרף Kaplan-Meier לפי מצב משפחתי**

עקומות ההישרדות לפי מצב משפחתי (תרשים 10) מראות כי עובדים נשואים (כתום) נשארים בארגון לאורך זמן יותר מעובדים שאינם נשואים (כחול). ההבדל בולט בעיקר בשנתיים הראשונות. ההבדל מובהק סטטיסטית (Log-rank $p\approx0.03$) גם אם ההשפעה נראית מתונה יחסית. חשוב להדגיש כי מדובר בניתוח חד-משתני שאינו מבקר על משתנים נוספים (כגון גיל, ילדים או שכר יחסי), ולכן יש להתייחס לממצא זה ככיוון כללי בלבד ולא כמסקנה סופית. המשמעות היא שבשלב זה אין

להסיק מכך המלצות אופרטיביות, אלא לראות בו אינדיקציה ראשונית שתיבחן בהמשך בניתוח הרב-משתני.

מגדר (gender)

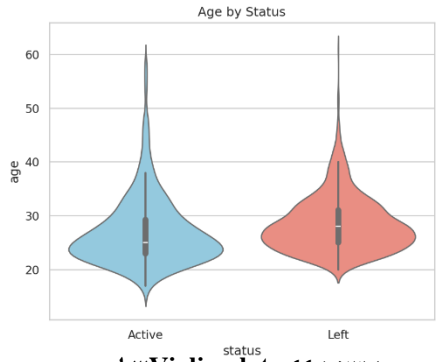
עקומות הישרדות לפי מגדר (תרשים 11 בנספח) מראות דמיון כמעט מלא בין נשים וגברים. ההבדל בשרידות אינו מובהק סטטיסטית (Log-rank p>0.2). כלומר, בניגוד למשתנים אחרים (כגון ילדים או מצב משפחתי), מגדר אינו נמצא כגורם מבדיל בהישארות במוקדי שירות, ולכן אין הצדקה למדיניות שימור מגדרית ייחודית.

משתנים רציפים

Variable	Median_Active	Median_Left	Mean_Active	Mean_Left	MannWhitney_p	HR	HR_CI_lower	HR_CI_upper
גיל בשנים	25.00	28.00	27.06	28.43	0.00	0.96	-0.05	-0.03
שכר יחסי	1.14	1.09	1.15	1.10	0.00	0.03	-3.93	-2.88
הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות במוקד כולו	-38.71	-39.83	-36.48	-51.86	0.06	1.00	0.00	0.00
הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות של העובד בשנה אחרונה	-18.43	-20.25	-25.11	-31.61	0.19	1.00	0.00	0.00
הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה במוקד כולו	-0.46	-0.42	-8.63	-5.84	0.55	1.00	0.00	0.00
הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה של העובד בשנה האחרונה	-2.43	-11.75	-13.55	-16.41	0.03	1.00	0.00	0.00
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה במוקד כולו	-0.10	-0.19	63.31	24.09	0.00	1.00	0.00	0.00
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה של העובד בשנה אחרונה	0.33	0.29	-11.21	-18.83	0.64	1.00	0.00	0.00
יחס בין ממוצע אחוז זמן הפסקה מתוך שעות לוגין של העובד ב-3 חודשים אחרונים לפני סיום מעקב לבין ממוצע של העובד בשנה אחרונה	1.02	1.02	1.02	1.02	0.97	0.69	-0.78	0.05

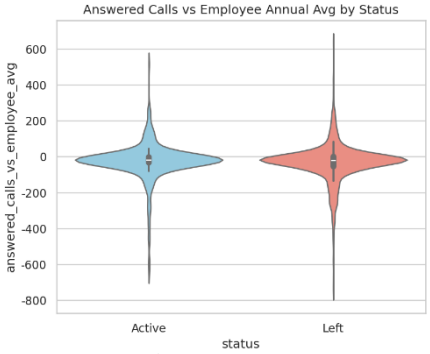
טבלה 3 : **השוואה חד-משתנית של משתנים רציפים בין עובדים פעילים לנושרים (כולל HR חד-משתני)**

טבלה 3 מציגה את ההשוואה החד משתנית במשתנים רציפים בין עובדים פעילים לעוזבים. עבור כל משתנה מוצגים החציון והממוצע לכל קבוצה, ערך ה- p.value של מבחן Mann-Whitney (השוואה בין שתי קבוצות בלתי תלויות), וכן יחס הסיכונים (HR) חד משתני עם רווחי סמך 95%. הסיבה לשימוש ב- Mann-Whitney היא שמדובר במבחן לא פרמטרי ,שאינו מניח התפלגות נורמלית של המשתנים או שוויון שונויות בין הקבוצות. מאחר ובנתונים נמצאו משתנים רציפים עם התפלגויות לא סימטריות ולעיתים גם חריגות, השימוש במבחן Mann-Whitney עדיף על מבחן t הקלאסי, שהוא מבחן פרמטרי שמניח נורמליות ושוויון שונויות.



תרשים 11 : **Violin plot של גיל לפי סטטוס**

תרשים 12 : **Violin plot של שכר יחסי לפי סטטוס**



תרשים 13 : **Violin plot של הפער בממוצע שיחות נענות (3 חודשים אחרונים מול שנתי) לפי סטטוס**

תרשים 11 מציג את התפלגות הגיל לפי סטטוס. למרות המובהקות הסטטיסטית, ניכר חיתוך גבוה בין הקבוצות, מה שמעיד על אפקט קטן בלבד. עוזבים הם בממוצע מעט מבוגרים יותר מפעילים. האפקט קטן ולכן לא ניתן להסיק כאן משמעות ניהולית חזקה.

תרשים 12 מציג את התפלגות השכר היחסי לפי סטטוס. נמצא הבדל מובהק בין עובדים פעילים לעוזבים. עובדים פעילים נוטים להרוויח שכר גבוה יותר ביחס לשכר המינימום, בעוד שעובדים שעזבו נטו לקבל שכר שקרוב יותר לרמת המינימום. ככל שהשכר היחסי גבוה יותר, כך הסיכון לעזיבה יורד משמעותית (HR≈0.68 לכל עלייה של 0.1 ביחס לשכר המינימום, כלומר עלייה של 10% ביחס לשכר המינימום).

יש לזכור כי ההבדל בין עובדים פעילים לעוזבים משקף גם את אורך תקופת ההעסקה. עובדים שנשארו עד סוף המעקב בהכרח ותיקים יותר, ולכן סביר שנהנו מעליות שכר במהלך הזמן. אמנם המדד שבשימוש כאן הוא שכר יחסי (שכר חודשי ל-100% משרה ביחס לשכר המינימום באותה תקופה), אך עדיין ייתכן שחלק מהפער בין הקבוצות מוסבר בהשפעת הוותק והעלאות מצטברות לאורך זמן. עם זאת, גם לאחר התחשבות במגבלה זו, התוצאה מצביעה על מגמה עקבית וברורה, ששכר גבוה יותר ביחס למינימום קשור ליציבות תעסוקתית גבוהה יותר ולסיכון מופחת לנטישה.

ממצא זה מדגיש את תפקידו המרכזי של גובה השכר ביחס למינימום כגורם שעלול למנוע את העזיבה. תוספות שכר מתונות מעל למינימום עשויות לשפר באופן ניכר את יציבות העובדים ולהקטין נטישה מוקדמת.

תרשים 13 מציג את ההתפלגות של הפער בממוצע מספר השיחות הנענות בשלושת החודשים האחרונים לעומת הממוצע השנתי של אותו עובד, לפי סטטוס (פעילים מול עוזבים).

נראה כי בקרב העובדים שעזבו (ימין, אדום) קיימת נטייה חזקה יותר לערכים שליליים - כלומר, הם ענו על פחות שיחות בחודשים האחרונים ביחס לשיעור המענה שלהם לאורך השנה. אצל העובדים הפעילים (שמאל, כחול) הפיזור קרוב יותר לאפס, מה שמעיד על יציבות גבוהה יותר בביצועים.

מבחן Mann-Whitney מצביע על הבדל מובהק סטטיסטית בין הקבוצות ($p<0.001$) . בנוסף, במודל Cox חד משתני התקבל HR≈1.10 לכל ירידה של 100 שיחות ביחס לממוצע השנתי של העובד, כלומר ירידה מתמשכת במספר השיחות הנענות מהווה אינדיקטור משמעותי לסיכון עזיבה.

תבנית זו יכולה לשמש כאות אזהרה מוקדם. עובדים שמתחילים להראות ירידה עקבית במספר השיחות הנענות לעומת הביצועים ההיסטוריים שלהם נמצאים בסיכון מוגבר לעזוב. ניטור של מדד זה בזמן אמת יכול לאפשר למנהלים לזהות מצבי שחיקה, לפתוח שיחה עם העובד ולהציע תמיכה או התאמות בתקופות העומס, ובכך להפחית סיכון לעזיבה.

מלבד המשתנים שתוארו לעיל, נבחנו גם מספר משתנים רציפים נוספים (המוצגים בטבלה 1). שני המשתנים מתוכם (call_duration_vs_center_avg ו-call_duration_vs_employee_avg) מבטאים את הפער בין משך השיחה הממוצע של העובד (בשניות) לבין הממוצעים (של המוקד ושל העובד עצמו בהתבוננות שנתית). נמצא כי הפערים אינם גדולים בממוצע בין עוזבים לפעילים, ולא נצפו הבדלים מובהקים במיוחד, אם כי קיימת מגמה של ירידה קלה ביכולת לעמוד בממוצעים לפני עזיבה.

בבדיקה של ממוצע שיחות לשעה (calls_perhour_vs_center_avg לעומת calls_perhour_vs employee_avg) ניתן לראות שבעוד שהשוואה לממוצע של מוקד לא הראתה קשר מובהק לעזיבה, הפער מול ביצועי העובד עצמו נמצא בעל משמעות גבוהה יותר, ירידה בממוצע שיחות לשעה ביחס לשנה האחרונה הייתה קשורה לעלייה בסיכון עזיבה. הדבר מחזק את הממצא שכבר הוצג (בגרף על שיחות נענות), ולפיו השוואה של העובד לעצמו היא אות אזהרה מדויק יותר מאשר השוואה לאחרים.

משתנה pause_self3m_vs_selfannual משקף את שינוי דפוסי ההפסקות של העובד בשלושת החודשים האחרונים ביחס לשנה השוטפת. נמצא כי קיימים עובדים שנטו להגדיל את משך ההפסקות לפני עזיבה, אך האפקט קטן ואינו מובהק סטטיסטית ברמה שמאפשרת להסיק מסקנה ברורה.

בסך הכול, משתנים אלו מצביעים על מגמות עקביות אך פחות חדות בהשוואה למשתנים שתוארו בהרחבה. משמעות הדבר היא שהתרומה שלהם להבנת סיכוני העזיבה קיימת, אך היא משנית ביחס למשתנים הכלכליים (כמו שכר יחסי) ולמדדים התפעוליים המרכזיים (כמו ירידה בממוצע שיחות שנענות ביחס להיסטורית העובד).

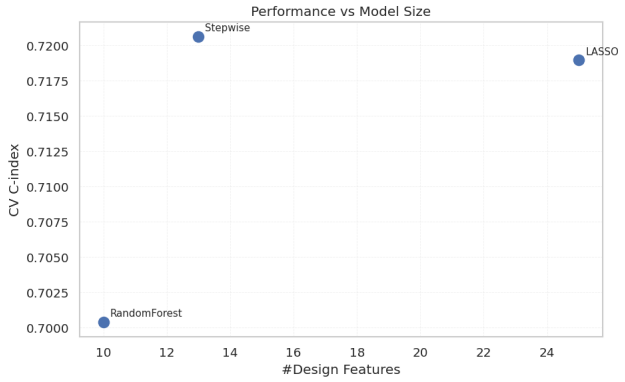
6.3. בחירת משתנים למודל

Base design built: X_base shape=(1751, 26)

	Method	Num_features_design	C-index_CV_mean	C-index_CV_sd
0	RandomForest (proxy → Top-K → Cox)	10	0.700392	0.011699
1	LASSO (screen+z → Cox L1)	25	0.718929	0.013837
2	Stepwise (ΔC-index add/drop)	13	0.720615	0.014998

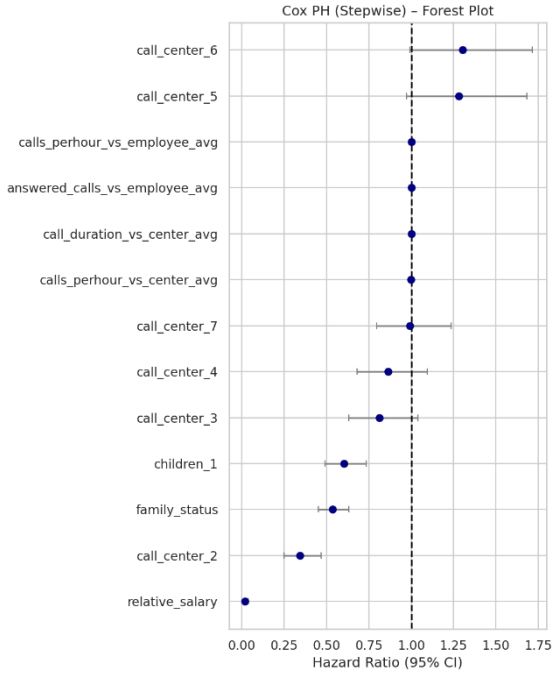
טבלה 4 : **השוואת ביצועי מודלים לפי שיטת בחירת משתנים**

לצורך בחירת משתנים למודל Cox נבחנו שלוש גישות שונות: LASSO (טבלה 8 בנספח) Random Forest (טבלה 9 בנספח) ו-Stepwise. כל אחת מהשיטות הפיקה סט משתנים מעט שונה, אך כל המודלים הציגו ביצועים דומים מאוד מבחינת יכולת ניבוי (C-index סביב 0.70-0.72 : טבלה 4).



תרשים 14 : **מורכבות וביצועי המודל - השוואה בין 3 שיטות**

תרשים 14 מציג את הקשר בין מורכבות המודל לבין ביצועיו. ניתן לראות כי מעבר ל-10-12 משתנים לא מושג שיפור מהותי בדיוק. הבחירה הסופית נעשתה בשיטת Stepwise, מתוך שיקול של פשטות ויכולת פרשנות ניהולית, שכן היא אפשרה לשמור על מספר מצומצם של משתנים תוך שמירה על ביצועי ניבוי גבוהים.



תרשים 15 : **Forest Plot של המודל Stepwise**

תרשים 15 מציג Forest Plot של המודל שנבחר בשיטת Stepwise . בתרשים מתוארים יחס הסיכונים (HR) ורווחי הסמך של המשתנים שנותרו במודל לאחר סינון. ניתן לראות כי מספר משתנים כלכליים, דמוגרפיים ותפעוליים ממשיכים להיות משמעותיים גם לאחר בדיקה תוך כדי התחשבות במשתנים אחרים :

- **שכר יחסי:** בעל ההשפעה החזקה ביותר. עובדים שמקבלים שכר יחסי (למינימום) גבוה יותר נמצאים בסיכון משמעותית נמוך יותר לעזוב.
- **משתנים דמוגרפיים:** גם מצב משפחתי ונוכחות ילדים נמצאו כגורמים שמורידים את הסיכון לעזיבה(HR נמוך מ-1) , המצביעים על נטייה לשרידות גבוהה יותר.
- **מאפייני מוקד** : מספר מוקדים שונים הציגו יחס סיכונים שונה באופן מובהק מהמוקד הייחוס , זה מהווה אינדיקציה לכך שהבדלים בין מוקדים (כגון גודל המוקד, תרבות ניהולית, עומס או תהליכים פנימיים) משפיעים על שרידות העובדים.

- **מדדי KPI תפעוליים**: משתנים כמו ירידה בקצב השיחות לשעה או במספר השיחות הנענות ביחס לשנה הקודמת נמצאו קשורים לעלייה בסיכון לעזיבה. לעומת זאת, השוואה לביצועי המוקד (center_avg) תרמה פחות להסבר סיכויי העזיבה, אך נשארה כחלק מהמעטפת התפעולית שנבחרה במודל.

תרשים 15 מדגים בבירור כי המודל שנבחר מצליח לשלב משתנים מתחומים שונים (כלכלה, דמוגרפיה ותפעול) ולשמור על מודל יחסית פשוט, תוך שימור יכולת הניבוי הגבוהה שנמצאה גם בשיטות בחירה אחרות.

6.4. תוצאות מודל Cox הסופי

covariate	HR	HR_lower	HR_upper	p	z	coef	se(coef)
שכר יחסי	0.02	0.01	0.04	0.00	-12.17	-3.91	0.32
מצב משפחתי	0.53	0.45	0.63	0.00	-7.22	-0.63	0.09
מוקד קטן	0.34	0.25	0.47	0.00	-6.67	-1.07	0.16
הורים עם ילדים	0.60	0.49	0.74	0.00	-4.94	-0.51	0.10
הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות של העובד בשנה אחרונה	1.00	1.00	1.00	0.00	3.53	0.00	0.00
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה של העובד בשנה אחרונה	1.00	1.00	1.00	0.00	3.12	0.00	0.00
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה במוקד כולו	1.00	1.00	1.00	0.05	-1.98	0.00	0.00
מוקד גדול	1.30	0.99	1.71	0.06	1.90	0.26	0.14
מוקדבינוני עד גדול	1.28	0.97	1.68	0.08	1.77	0.25	0.14
מוקד בינוני עד קטן	0.81	0.63	1.04	0.10	-1.66	-0.21	0.13
הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה במוקד כולו	1.00	1.00	1.00	0.15	1.45	0.00	0.00
מוקד בינווני	0.86	0.68	1.10	0.23	-1.21	-0.15	0.12
מוקד גדול מאוד	0.99	0.80	1.24	0.94	-0.08	-0.01	0.11

טבלה 5: **תוצאות מודל- Cox Proportional Hazards (Stepwise) יחסי סיכון (HR) ורווחי סמך 95%**

לאחר תהליך בחירת המשתנים בשיטת Stepwise , התקבל מודל סופי הכולל שילוב של משתנים דמוגרפיים, כלכליים ותפעוליים .טבלה 5 מציגה את יחסי הסיכונים (Hazard Ratios - HR) , רווחי הסמך (95% CI) וערכי המובהקות (p-value) עבור כל אחד מהמשתנים שנכללו במודל.

- **שכר יחסי :** נמצא כמשתנה בעל ההשפעה החזקה ביותר. ככל שהשכר היחסי גבוה יותר (כלומר העובד משתכר מעל שכר המינימום בתקופה הרלוונטית), כך קטן הסיכון לעזיבה בצורה משמעותית. מדובר בגורם המפחית את הסיכון לעזיבה בצורה מובהקת (HR <1, p<0.001).

- **משתנים דמוגרפיים :** נוכחות ילדים ומצב משפחתי נמצאו אף הם כגורמים שמפחיתים את הסיכון לעזיבה (HR<1) . המשמעות היא שעובדים בעלי מחויבויות משפחתיות נוטים להישאר לאורך זמן רב יותר.

- **מאפייני מוקד:** נצפו הבדלים מובהקים בין מוקדים שונים. חלקם מציגים HR נמוך מ-1 (סיכון מופחת יחסית למוקד הייחוס), בעוד אחרים קרובים או גבוהים מ-1 (סיכון מוגבר). הדבר מעיד כי קיימת השפעה ניכרת לגורמים מערכתיים ותרבותיים ברמת המוקד.

- **מדדי KPI תפעוליים :** משתנים המתארים ירידה בביצועי העובד ביחס לעצמו(כגון answered_calls_vs_employee_avg או calls_perhour_vs_employee_avg) נמצאו כמעלים סיכון לעזיבה (HR>1), בעוד שהשוואה לביצועי המוקד (center_avg) הציגה השפעה קטנה יותר. בכך המודל מאשר כי השוואה של ביצועים אישיים (אל מול עובד עצמו) היא אינדיקציה חזקה יותר לסיכון עזיבה מאשר השוואה חיצונית (אל מול המוקד).

ממצאים אלו מראים כי גם גורמים כלכליים (שכר יחסי), גם מאפיינים דמוגרפיים (מצב משפחתי וילדים), וגם מדדי ביצוע תפעוליים (ביצועים אחרונים מול ביצועים היסטוריים) משפיעים באופן מובהק על הסיכון לעזיבה. הדבר מחזק את הצורך בניהול רב ממדי: מדיניות שכר מותאמת, זיהוי וניטור מדדי ביצוע בזמן אמת, ומתן אפשרות ללמידה בין מוקדים.

7. דיון ומסקנות

7.1. דיון, מסקנות פרשנות ניהולית

ממצאי המחקר הנוכחי מאשרים כי נטישת עובדים במוקדי שירות טלפוניים בישראל היא תופעה מורכבת המושפעת משילוב של גורמים אישיים, כלכליים ותפעוליים. באמצעות מודל Cox נמצא כי השכר היחסי הוא הגורם המגן החזק ביותר: עובדים ששכרם גבוה יותר ביחס לשכר המינימום היו בעלי סיכוי משמעותית נמוך יותר לעזוב. במילים אחרות, כל עלייה בשכר מעבר למינימום מהווה תמריץ חשוב ומקטינה בצורה מובהקת את סיכון הנטישה. ממצא זה אינו רק תוצאה סטטיסטית אלא נושא עמו משמעות ניהולית ברורה - מדיניות תגמול שממקמת את העובדים מעל קו המינימום, אפילו בהפרש קטן, עשויה להוות כלי אפקטיבי לצמצום עזיבות מוקדמות.

יש להדגיש כי רבים מנציגי השירות אינם מועסקים במשרה מלאה ולעיתים עובדים בהיקף של 60%–80%. המשמעות היא שהכנסתם החודשית בפועל נמוכה עוד יותר, גם כאשר שכרם השעתי קרוב לשכר המינימום. במציאות זו, כל תוספת קטנה בשכר השעתי מתורגמת לשינוי משמעותי בתחושת היציבות הכלכלית וביכולת העובד לפרנס את עצמו או את משפחתו. לפיכך, עצם העובדה שנמצא קשר כה חזק בין שכר יחסי לבין עזיבה, מצביע על כך שכנראה הסיכון גבוה במיוחד בקבוצת עובדים במשרה חלקית.

ממצא זה תואם את הספרות המחקרית המצביעה על הקשר הישיר בין תגמול הולם לבין נטייה להישאר בארגון (Coomber & Barriball, 2007; Kumar et al., 2012) . חוקרים אלו הצביעו על כך שהשכר הוא חלק מרכזי בציפיות של העובד מהארגון: הוא מסמל הערכה, ביטחון כלכלי ותחושת שייכות. כאשר העובדים חשים כי הם מתוגמלים באופן נמוך מדי, במיוחד בתפקידים עתירי עומס רגשי ותפעולי כמו מוקדי שירות, הם נוטים להרגיש חוסר איזון בין השקעתם לבין התמורה שהם מקבלים - מה שמגביר את הסיכון לנטישה.

במובן זה, המחקר הנוכחי מחזק את ההבנה שתוספות קטנות בשכר, מודלים של בונוסים או התאמות יוקר מחיה אינן 'תוספת שולית', אלא עשויות להוות תנאי בסיסי להישאר בארגון. לארגונים יש כאן כלי פרקטי, השקעה נקודתית בשכר או בתמריצים מותאמים יכולה לייצר השפעה ישירה על ירידת שיעורי התחלופה, במיוחד בקרב עובדים במשרות חלקיות.

גורמים דמוגרפיים כמו מצב משפחתי והורות לילדים נמצאו אף הם כבעלי השפעה על סיכון העזיבה. תוצאה זו משתלבת היטב עם הספרות המחקרית, שמצאה כי עובדים עם מחויבות משפחתית, בין אם מדובר בנישואים ובין אם מדובר בגידול ילדים, נוטים לגלות יציבות תעסוקתית גבוהה יותר ולהפגין נאמנות ארגונית מוגברת (Morrow et al., 1999; Zito et al., 2018) . המחקר הנוכחי מצא באופן מובהק כי עובדים נשואים ועובדים עם ילדים נמצאים בסיכון נמוך יותר לעזיבה, השפעתם נותרה מובהקת גם כאשר נבדקה בתוך המודל הרב משתני, הכולל במקביל את משתני השכר והביצועים התפעוליים.

הסבר אפשרי לכך הוא כי מחויבות משפחתית מגבירה את הצורך של העובד בביטחון כלכלי ותעסוקתי, מצמצמת את הנכונות לקחת סיכונים בקריירה ומעודדת הישארות במקום עבודה יציב. יתר על כן, עובדים עם ילדים עשויים לייחס חשיבות גדולה יותר להטבות עקיפות כמו גמישות בשעות, סביבת עבודה יציבה והבנת מנהלים את צורכי המשפחה, גורמים אשר לרוב אינם נמדדים ישירות במדדים כלכליים, אך יש להם תרומה מכרעת על חוויית העובד.

ממצא זה מחזק את ההבנה ששימור עובדים אינו נשען רק על תגמול כספי אלא גם על היכולת של הארגון להציע איזון בין עבודה לחיים פרטיים . השקעת הארגון במדיניות שתומכת בתא משפחתי בדרכים שונות כגון, אפשרות לעבודה חלקית מותאמת, אפשרות עבודה מהבית, גמישות בשעות עבודה, או סיוע בהוצאות רווחה – עשויה לתרום באופן משמעותי להפחתת שיעורי הנטישה, במיוחד בקרב אוכלוסיות המוגדרות כ'פגיעות' יותר מבחינת גורמים חיצוניים.

בכך, המחקר מעיד על כך כי התמיכה באיזון בין עבודה ומשפחה אינה רק 'ערך מוסף' אלא רכיב אסטרטגי במדיניות שימור. ארגון המזהה את המשקל של גורמים דמוגרפיים יכול לפתח תוכניות

מותאמות אישית לשימור עובדים בהתאם למאפיינים המשפחתיים שלהם, ובכך לא רק להפחית נטישה אלא גם לחזק את תחושת המחויבות והזיקה של העובדים לארגון.

בתחום התפעולי, הממצאים חידדו כי דווקא מדדים המשווים את העובד לעצמו לאורך זמן (כגון ממוצע כמות שיחות נענות וממוצע כמות שיחות לשעה) הם מנבאים חזקים יותר לעזיבה מאשר השוואה לביצועי המוקד כולו. משמעות הדבר היא כי ירידה בביצועיו האישיים של עובד ביחס להיסטוריה שלו מהווה אינדיקציה מוקדמת וברורה יותר לסיכון עזיבה מאשר ההשוואה החיצונית לקולגות במוקד. ממצא זה מתיישב היטב עם מחקרים קודמים שנעשו על חשיבות מגמות ביצוע אישיות כמנבאות עיקריות לנטישה (Valle & Ruz, 2015) , ומציע כלי ישים במיוחד לניהול. ניטור שוטף של ביצועי כל עובד ביחס לביצועיו ההיסטוריים מאפשר זיהוי מוקדם של עובדים המצויים במגמת ירידה , גם אם עדיין עומדים ברמת הביצועים הממוצעת של המוקד, ובכך מאפשר הפעלת התערבות ממוקדת לשימור. מבחינה ניהולית, מדובר בפרקטיקה פרואקטיבית- לאתר את ה'שחיקה שקוראת בשקט עוד לפני שהיא מתבטאת בהישגים נמוכים בהשוואה לאחרים.

הממצאים לגבי השונות בין מוקדים שונים מוסיפים רובד נוסף להבנת התופעה, בכך שהם מצביעים על כך שגורמים ארגוניים רחבים יותר (כגון תרבות ניהולית, תהליכי חפיפה, מודלים של תמריצים ועומס עבודה) עשויים להסביר את ההבדלים. יחד עם זאת, יש לזכור כי ייתכן גם שמדובר בהבדלים הקשורים לאופי האוכלוסייה שמגויסת בכל מוקד. מיקום גיאוגרפי, למשל, עשוי למשוך אוכלוסיות שונות (מוקד בפריפריה עשוי למשוך יותר עובדים בעלי משפחות וילדים, בעוד מוקד במרכז כגון תל אביב, עשוי למשוך אוכלוסייה צעירה וניידת יותר). כלומר, ייתכן שחלק מהשונות נובעת לא רק מתרבות הניהול אלא גם מהרכב חברתי דמוגרפי שונה של העובדים.

יש לציין כי במסגרת מחקר זה לא נותחה השפעת אירועים חיצוניים חריגים כגון מלחמות. מאחר שתקופת המעקב כללה את פרוץ מלחמת "חרבות ברזל" (7.10.2023) הנמשכת עד היום, ייתכן שלמלחמה היו השלכות משמעותיות על דפוסי ההעסקה והעזיבה (למשל, עלייה בנטישת עובדים עקב גיוס מילואים ממושך, קשיים בתחבורה ציבורית, מעבר אזורי מגורים או לחץ כלכלי). בהיעדר משתנה ישיר המייצג את השפעת המלחמה, ממצאי המודל הנוכחי אינם מתייחסים לגורם זה. זוהי מגבלה חשובה שיש להביאה בחשבון, והיא מדגישה את הצורך במחקרי המשך שיבחנו במפורש את השפעתם של אירועים לאומיים ביטחוניים על תחלופת עובדים.

7.2. חידוש ותרומה

המחקר הנוכחי מוסיף נדבך חשוב לספרות, הן מהבחינה המתודולוגית והן מהבחינה האמפירית. ראשית, הוא נשען על נתוני אמת רחבי היקף שנאספו עבור יותר מ-1,700 נציגי שירות לאורך תקופה של שבע שנים, דבר המאפשר לבחון תהליכי עזיבה על פני זמן ולא רק בתמונת מצב רגעית. בשונה מרוב המחקרים שנעשו עד כה שהתמקדו במשתנים סוציולוגיים או פסיכולוגיים (למשל שאלוני שביעות רצון וכוונת עזיבה), כאן נעשה שימוש בנתוני ביצוע ותפעול בפועל, כפי שהם נרשמים במערכות הארגוניות. יתרון זה הופך את הממצאים לברי יישום מיידי במוקדי שירות.

שנית, המחקר מתמקד במוקדי שירות בישראל - תחום שבו הספרות האקדמית דלה יחסית. אף שתחלופת עובדים במוקדים בישראל גבוהה מאוד בהשוואה בינלאומית, כמעט ולא נערכו מחקרים אמפיריים ארוכי טווח שהתבססו על נתוני אמת. בכך העבודה הנוכחית נותנת מענה לפער מחקרי חשוב ומאפשרת לארגונים מקומיים לקבל תובנות המותאמות להקשר התרבותי והעסקי הייחודי בישראל.

שלישית, השימוש במודל הישרדות (Cox) מאפשר לא רק לחזות אם עובד יעזוב או יישאר, אלא גם להעריך את ההסתברות לעזיבה בכל נקודת זמן לאורך מסלול ההעסקה. זהו יתרון משמעותי לעומת מודלים בינאריים, משום שהוא מאפשר לזהות "חלונות סיכון" קריטיים (כגון השנה הראשונה לעבודה), ולבנות סביבם אסטרטגיות ניהול ממוקדות.

תרומה ייחודית נוספת של המחקר היא יצירת משתני KPI חדשים שהוגדרו כהפרשים לאורך זמן (למשל בין ביצועי עובד בשלושת החודשים האחרונים לבין ביצועיו הממוצעים השנתיים). גישה זו מאפשרת לזהות מגמות פנימיות ושחיקה מתמשכת עוד לפני שהיא ניכרת בהשוואה לממוצעי המוקד. כך מתקבלים מדדי סיכון פרקטיים ואותות אזהרה שניתן להפעיל באמצעות מערכות

תומכות החלטה בזמן אמת. במילים אחרות, תרומת המחקר אינה רק תאורטית אלא גם יישומית: המחקר מספק כלי ישים לזיהוי עובדים בסיכון ולפיתוח מנגנוני שימור מותאמים.

7.3. מגבלות

עם זאת, למחקר הנוכחי יש מספר מגבלות שיש להביא בחשבון.

ראשית ,הנתונים נאספו ממספר מוקדים מסוימים בישראל, ולכן ייתכן שהכללה לענפים אחרים מוגבלת. כל מוקד פועל בהקשר ניהולי ותרבותי שונה, ולפיכך ייתכן שהקשרים שנמצאו לא יתבטאו באותה עוצמה בענפים אחרים במשק או במדינות אחרות.

שנית ,חלק מהמשתנים נבנו על בסיס חישובים ועיבוד משני (למשל מדד שכר יחסי),מה שעשוי להכניס הטיות מסוימות. אף שהשיטה נדרשה כדי לאפשר השוואות הוגנות בין תקופות ושנים שונות, תמיד קיים סיכון שחישובים נגזרים משקפים גם רעש שנובע מתנודות חיצוניות ולא רק מתהליכים פנים ארגוניים.

שלישית ,למרות השימוש המתקדם בשיטת להשלמת ערכים חסרים בשיטת ,MICE תמיד קיים חשש שהחוסרים עצמם אינם MAR. לדוגמה, ייתכן שעובדים מסוימים או מוקדים מסוימים נוטים יותר להיעדר באופן שיטתי (MNAR), והדבר עשוי להשפיע על תוצאות המודל.

רביעית ,המודל מתבסס על הנחת הפרופורציונליות של יחסי הסיכונים .(Proportional Hazards) למרות שהבדיקות הסטטיסטיות העידו כי ההנחה סבירה ברמה הגלובלית, נמצאו חריגות מתונות במשתני KPI דינמיים - מה שמעיד כי ייתכן שההשפעות משתנות לאורך זמן. מצב זה מדגיש את הצורך במחקרי המשך שיבחנו מודלים מתקדמים יותר, כגון.time-varying coefficients

לבסוף, אחת המגבלות המרכזיות והמשמעותיות ביותר היא איכות תשתיות הנתונים הארגוניות . כבר בשלב איסוף הנתונים התברר כי לא קיים מסד נתונים אחיד שבו מרוכזים כל נתוני העובד. המידע נשמר במערכות שונות - מערכות נוכחות, מערכות טלפוניה, מערכות שכר, מחלקות משאבי אנוש וכספים, שלעתים אינן מתממשקות ביניהן כלל או מתממשקות באופן חלקי בלבד. חלק מהמידע נוהל ידנית באמצעות קבצי ,Excel דבר שהגביר את הסיכוי לטעויות ולחוסרים. כתוצאה מכך, נדרש מאמץ רב של איחוד, ניקוי והשלמת נתונים כדי לבנות קובץ מחקרי שניתן לעבוד איתו. תהליך זה אומנם בוצע בהצלחה, אך הוא חושף מגבלה מערכתית: ללא תשתית נתונים אינטגרטיבית, ארגונים מתקשים לנצל את המידע העשיר שברשותם לטובת קבלת החלטות מושכלות.

כאן חשוב להדגיש: יצירת מסד נתונים אחיד לא מחייבת מעבר לכלי תוכנה חדש או לניהול אחיד של כלל המידע במערכת אחת. הפתרון הפשוט יותר הוא שמירה של כלל הנתונים הקיימים בטבלאות מסודרות במקום נגיש אחד ,(Data Warehouse) והקצאת גורם בארגון שמכיר היטב את הנתונים ויודע לנהל אותם. השקעה נקודתית כזו תאפשר בעתיד לארגון לבצע ניתוחים מתקדמים בצורה יעילה, מהירה ואמינה יותר, תוך חיסכון במשאבים והגדלת הערך מהמידע הקיים. יתר על כן, מהלך כזה יתרום לא רק למחקר אקדמי אלא גם ליישומים פרקטיים כגון מערכות תומכות החלטה לצורך התרעה מוקדמות לעובדים בסיכון, כלי BI לשימור עובדים, ותכנון כוח אדם מבוסס נתונים.

7.4. כיווני המשך

המשך מחקר בתחום יכול להעמיק ולהתרחב במספר כיוונים משמעותיים:

ראשית ,יש מקום לפתח מודלים מתקדמים המאפשרים אפקטים משתנים בזמן. מודלים כאלה יתמודדו טוב יותר עם משתני KPI דינמיים, שהשפעתם אינה קבועה אלא משתנה לאורך חיי ההעסקה. כך, למשל, ייתכן שירידה בביצועי השיחות היא מנבא חזק במיוחד בששת החודשים הראשונים, אך פחות קריטית לאחר מכן - תובנה שניתן להגיע אליה רק באמצעות מודלים גמישים יותר.

שנית ,כדאי להשוות בין מודל Cox לבין אלגוריתמים מתקדמים של למידת מכונה, כגון XGBoost Survival Random Forest , או DeepSurv כדי לבחון אם ניתן לשפר את דיוק החיזוי ואת יכולת

הכללת המודל. שילוב בין יתרונות Cox (פשטות ופרשנות ניהולית) לבין אלגוריתמים חזקים יותר לניבוי עשוי להוות כיוון מחקר פורה.

שלישית ,הרחבת המדגם לענפים שונים היא חיונית. המחקר הנוכחי התמקד בכמות מצומצמת של מוקדי שירות בישראל, וכדי להגיע למסקנות כלליות יותר יש צורך במחקרים חוזרים בסקטורים מגוונים (כגון בריאות, חינוך או היי-טק) ובתרבויות ארגוניות שונות. רק באמצעות חזרות כאלה ניתן יהיה לבחון את עמידות המודל ואת הכללתו מעבר להקשר המקומי.

רביעית ,יש מקום לשלב נתוני שביעות רצון סובייקטיביים (למשל, שאלוני עובדים) עם נתוני תפעול אובייקטיביים .שילוב שכזה יכול לספק תמונה הוליסטית - מצד אחד תחושות ועמדות העובדים, ומצד שני ביצועים בפועל - ולשפר את היכולת לזהות עובדים בסיכון גבוה לעזיבה.

חמישית, מחקרי המשך יכולים לבחון את תרומת הוותק ההתחלתי של העובד בעת הכניסה למעקב. במחקר הנוכחי הוותק שימש כמשתנה תלוי (משך הזמן עד לעזיבה או לצנזורה), אך לא נכלל במפורש כמשתנה מסביר. עם זאת, ייתכן שעובדים שנכנסו למעקב עם ותק מצטבר ארוך יותר מציגים פרופיל הישרדות שונה מזה של עובדים חדשים יחסית. שילוב של ותק התחלתי כמשתנה מסביר במחקרי המשך עשוי להעמיק את ההבנה לגבי דפוסי העזיבה ולחדד את אסטרטגיות השימור עבור קבוצות עובדים שונות.

לבסוף, אחד הכיוונים המרכזיים ביותר להמשך מחקר ולפיתוח יישומים הוא בחינה של מודלים על בסיס נתוני אמת מתוך מסד נתונים אחיד בארגון. כפי שצוין במגבלות, חוסר באינטגרציה בין מערכות שונות אילץ שימוש בהשלמות ובאיחודי נתונים מורכבים. מחקר המשך שיוכל להתבסס על מאגר נתונים מסודר, שבו כלל המידע נשמר בטבלאות במקום נגיש אחד ומנוהל על ידי גורם מקצועי שמכיר את הנתונים, יספק בסיס יציב יותר לבניית מודלים. מצב כזה יפחית את הצורך בהשלמות מלאכותיות, ישפר את איכות התובנות, ויאפשר יישום ישיר של הממצאים לצרכים ניהוליים ולמערכות התרעה בזמן אמת. מעבר לכך, יצירת מחקרים חוזרים על ארגונים שבהם כבר הוסדר ניהול הנתונים תהווה הוכחת היתכנות לכך שההשקעה בתשתיות מידע משתלמת גם מבחינה מחקרית וגם מבחינה יישומית.

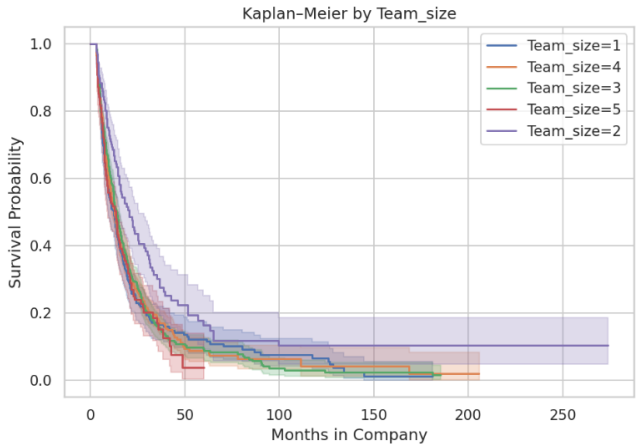
8. ביבליוגרפיה

1. **Box-Steffensmeier, J. M.,** & **Zorn, C.** (2001). Duration models and proportional hazards in political science. *American Journal of Political Science, 45*(4), 951–967.
2. **Celik, A. D.,** & **Oz, U. E.** (2011). The effects of emotional dissonance and quality of work life perceptions on absenteeism and turnover intentions among Turkish call center employees. *Procedia – Social and Behavioral Sciences, 30*, 2515–2519.
3. **Coomber, B.,** & **Barriball, K. L.** (2007). Impact of job satisfaction components on intent to leave and turnover for hospital-based nurses: A review of the research literature. *International Journal of Nursing Studies, 44*, 297–314.
4. **Demerouti, E., Bakker, A. B., Nachreiner, F.,** & **Schaufeli, W. B.** (2001). The job demands–resources model of burnout. *Journal of Applied Psychology, 86*(3), **499–512**.
5. **Kammeyer-Mueller, J. D., Wanberg, C. R., Rubenstein, A. L.,** & **Song, Z.** (2013). *Support, undermining, and newcomer socialization: Fitting in during the first 90 days. Academy of Management Journal, 56*(4), 1104–1124.
6. **Kumar, M., Singh, S., Rai, H.,** & **Bhattacharya, A.** (2012). Measuring humane orientation of organizations through social exchange and organizational identification: Facilitation and control of burnout and intent to quit. *International Journal of Organization Theory & Behavior, 15*(4), 520–547.
7. **Lee, T. W.,** & **Maurer, S. D.** (1997). The retention of knowledge workers with the unfolding model of voluntary turnover. *Human Resource Management Review, 7*(3), 247–275.*[בסקירה "Lee, 1997" מכסה את האזכור]*
8. **McKnight, D. H., Phillips, B.,** & **Hardgrave, B. C.** (2009). Which reduces IT turnover intention the most: Workplace characteristics or job characteristics? *Information & Management, 46*(3), 167–174.

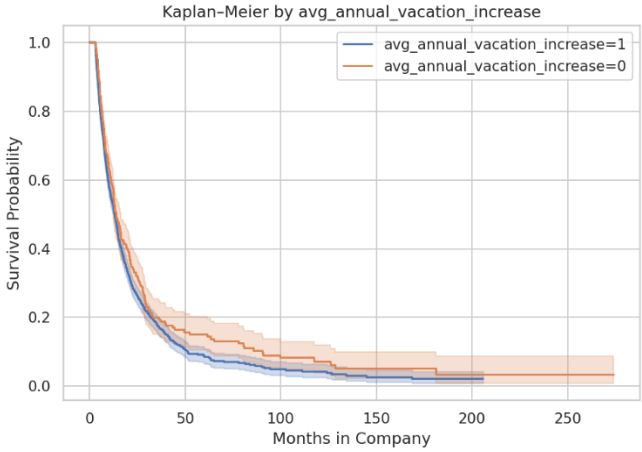
9. **Morrow, C. P., McElroy, C. J., Laczniak, S. K.,** & **Fenton, B. J.** (1999). Using absenteeism and performance to predict employee turnover: Early detection through company records. *Journal of Vocational Behavior, 55*, 358–374.
10. **Nagadevara, V., Srinivasan, V.,** & **Valk, R.** (2008). Establishing a link between employee turnover and withdrawal behaviours: Application of data mining techniques. *Research and Practice in Human Resource Management, 16*(2), 81–99.
11. **Rafaeli, A. et al** (2004). *Call center industry: Operations and human resource management*. Technion – Israel Institute of Technology. **Technical report.** *[אם יש לך גרסת דו"ח עם רשימת מחברים מפורטת – אפשר להשאיר כמותאם]*
12. **Saks, A. M., Uggerslev, K. L.,** & **Fassina, N. E.** (2007). Socialization tactics and newcomer adjustment: A meta-analytic review and test of a model. *Journal of Vocational Behavior, 70*(3), 413–446.
13. **Sawyerr, O. O., Srinivas, S.,** & **Wang, S.** (2009). Call center employee personality factors and service performance. *Journal of Services Marketing, 23*(5), 301–317.*[מכסה את "Sawyerr et al., 2009; as cited in …"]*
14. **Therneau, T. M.,** & **Grambsch, P. M.** (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.
15. **Valle, M. A., Varas, S.,** & **Ruz, G. A.** (2012). Job performance prediction in a call center using a naive Bayes classifier. *Expert Systems with Applications, 39*(11), 9939–9945.
16. **Valle, A. M.,** & **Ruz, A. G.** (2015). Turnover prediction in a call center: Behavioral evidence of loss aversion using Random Forest and Naïve Bayes algorithms. *Applied Artificial Intelligence, 29*, 923–942.
17. **van Buuren, S.,** & **Groothuis-Oudshoorn, K.** (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.
18. **Wallace, C. M., Eagleson, G.,** & **Waldersee, R.** (2000). The sacrificial HR strategy in call centres. *International Journal of Service Industry Management, 11*(2), 174–184.
19. **Zhou, J.,** & **George, J. M.** (2001). When job dissatisfaction leads to creativity: Encouraging the expression of voice. *Academy of Management Journal, 44*(4), 682–696.
20. **Zito, M., Emanuel, F., Molino, M., Cortese, C. G., Ghislieri, C.,** & **Colombo, L.** (2018). Turnover intentions in a call center: The role of emotional dissonance, job resources, and job satisfaction. *PLOS ONE, 13*(2), e0192126.
21. **Hocking, R. R. (1976).** The analysis and selection of variables in linear regression. *Biometrics, 32*(1), 1–49.
22. **Akaike, H. (1974).** A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.
23. **Schwarz, G. (1978).** Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.
24. **Hurvich, C. M.,** & **Tsai, C.-L.** (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297–307.
25. **Burnham, K. P.,** & **Anderson, D. R.** (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). Springer.
26. **Vancak, V., Goldberg, Y.,** & **Levine, S. Z.** (2022). The number needed to treat adjusted for explanatory variables in regression and survival analysis: Theory and application. *Statistics in Medicine, 41*(17), 3299–3320. https://doi.org/10.1002/sim.9418
27. **Little, R. J. A.,** & **Rubin, D. B.** (2019). Statistical Analysis with Missing Data (3rd ed.). Wiley
28. **Enders, C. K.,** & **Tofighi, D.** (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. Psychological Methods, 12(2), 121–138.
29. **Singer, J. D.,** & **Willett, J. B.** (2003). Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press
30. **Gans, N., Koole, G.,** & **Mandelbaum, A.** (2003). Telephone call centers: Tutorial, review, and research prospects. Manufacturing & Service Operations Management, 5(2), 79–141

31. **Harrell, F. E., Jr.** (2015). Regression Modeling Strategies (2nd ed.). Springer
32. **Vittinghoff, E.,** & **McCulloch, C. E**. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. American Journal of Epidemiology, 165(6), 710–718.
33. **Tibshirani, R.** (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine, 16*(4), 385–395. https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
34. **Breiman, L.** (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
35. **Hastie, T., Tibshirani, R.,** & **Friedman, J.** (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
36. **Kogan, I.** (2025). Employee-Attrition-Cox-Model [Computer software]. GitHub. https://github.com/Kogaines/Employee-Attrition-Cox-Model

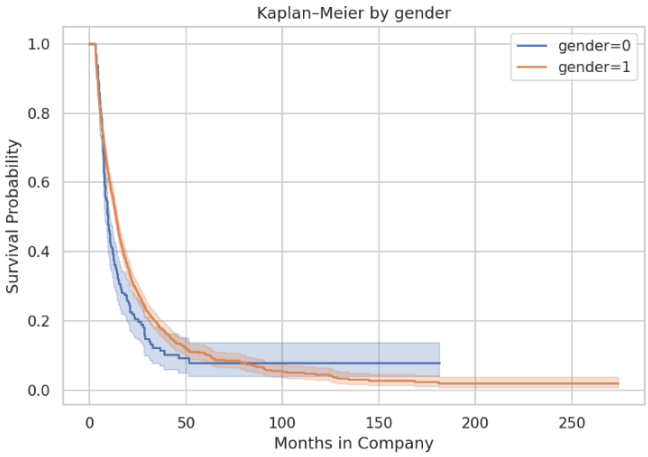
9. נספחים



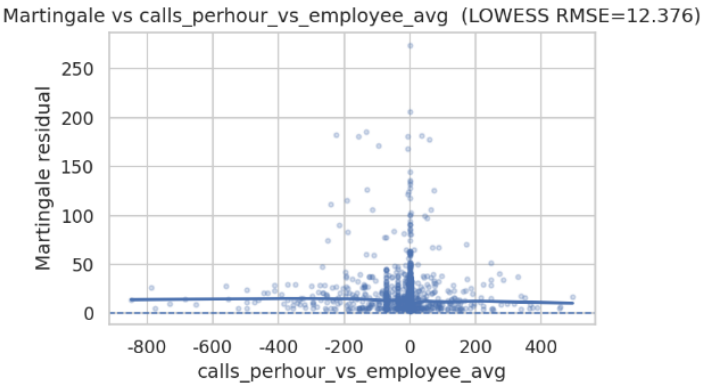
תרשים 5 : **עקומות Kaplan–Meier לפי גודל צוות**



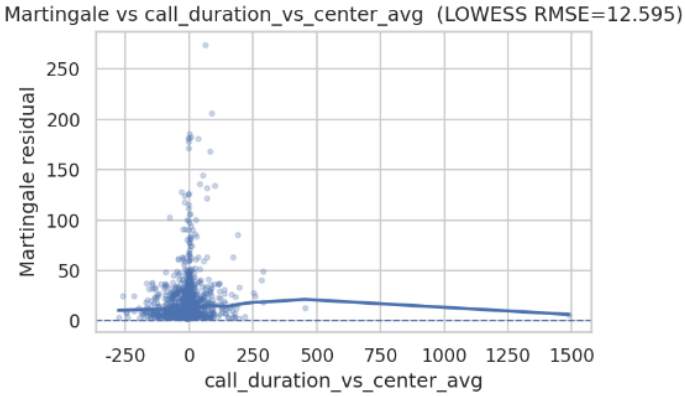
תרשים 7 : **עקומות Kaplan–Meier לפי המשתנה avg_annual_vacation_increase**



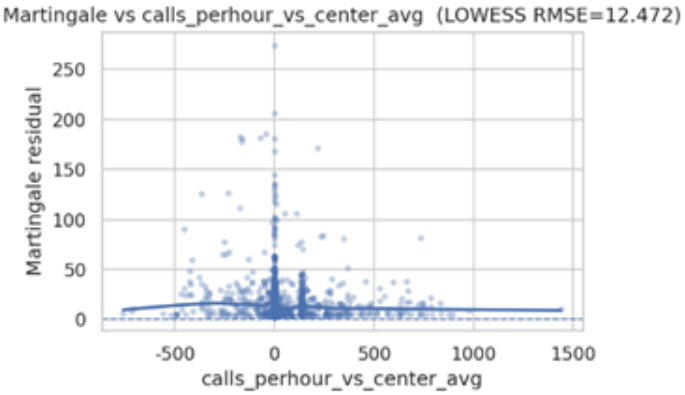
תרשים 11 : **עקומות Kaplan–Meier לפי מגדר (gender)**



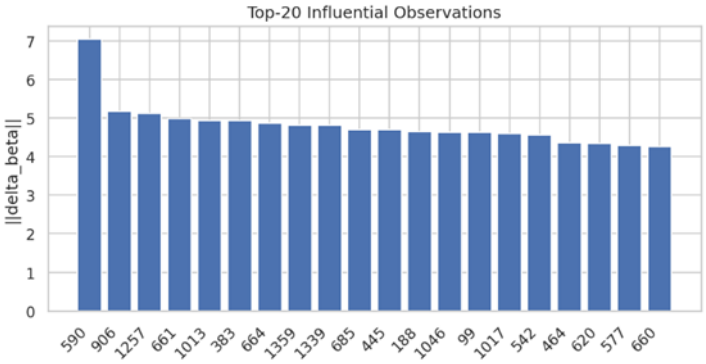
תרשים 19 : **בדיקת ליניאריות למשתנה calls_perhour_vs employee_avg**



תרשים 20 : **בדיקת ליניאריות למשתנה call_duration_vs_center_avg**



תרשים 21 : **בדיקת ליניאריות למשתנה calls_perhour_vs_center_avg**



תרשים 22 : **תצפיות משפיעות**

covariate	HR	HR_lower	HR_upper	p	z	coef	se(coef)
שכר יחסי	0.01	0.01	0.03	0.00	-12.74	-4.29	0.34
מצב משפחתי	0.58	0.48	0.69	0.00	-6.15	-0.55	0.09
מוקד בינוני	0.31	0.21	0.47	0.00	-5.54	-1.16	0.21
גיל בשנים	0.97	0.96	0.98	0.00	-5.33	-0.03	0.01
אחוז משרה	1.81	1.40	2.34	0.00	4.52	0.59	0.13
הורים עם ילדים	0.66	0.54	0.81	0.00	-3.89	-0.41	0.11
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה של העובד בשנה אחרונה	1.00	1.00	1.00	0.00	3.56	0.00	0.00
אין עליה בניצול ימי מחלה ב-3 חודשים לפני סיום מעקב	1.36	1.09	1.69	0.01	2.77	0.31	0.11
הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות של העובד בשנה אחרונה	1.00	1.00	1.00	0.01	2.67	0.00	0.00
מוקד בינוני עד קטן	0.78	0.60	1.02	0.07	-1.80	-0.25	0.14
הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה במוקד כולו	1.00	1.00	1.00	0.09	-1.69	0.00	0.00
הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה במוקד כולו	1.00	1.00	1.00	0.09	1.68	0.00	0.00
מגדר	0.86	0.73	1.03	0.10	-1.66	-0.15	0.09
מוקד גדול	1.31	0.95	1.80	0.10	1.66	0.27	0.16
צוות בין 15 ל 20 נציגים	1.13	0.95	1.35	0.16	1.40	0.13	0.09
מוקד ביינוני עד גדול	1.23	0.90	1.67	0.19	1.32	0.21	0.16
הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה של העובד בשנה האחרונה	1.00	1.00	1.00	0.28	1.07	0.00	0.00
אין עליה בניצול ימי חופשה ב-3 חודשים לפני סיום מעקב	0.92	0.79	1.07	0.29	-1.05	-0.08	0.08
צוות בין 10 ל 15 נציגים	1.18	0.82	1.70	0.38	0.89	0.17	0.19
צוות מעל 25 נציגים	1.11	0.87	1.41	0.41	0.82	0.10	0.12
מוקד ביינוני	0.93	0.69	1.24	0.61	-0.51	-0.08	0.15
הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות במוקד כולו	1.00	1.00	1.00	0.70	0.38	0.00	0.00
מוקד גדול מאוד	0.97	0.76	1.24	0.81	-0.24	-0.03	0.12
יחס בין ממוצע אחוז זמן הפסקה מתוך שעות לוגין של העובד ב-3 חודשים אחרונים לפני סיום מעקב לבין ממוצע של העובד בשנה אחרונה	1.04	0.65	1.66	0.87	0.16	0.04	0.24
בין 20 ל 25 נציגים	1.00	0.83	1.22	0.99	0.01	0.00	0.10

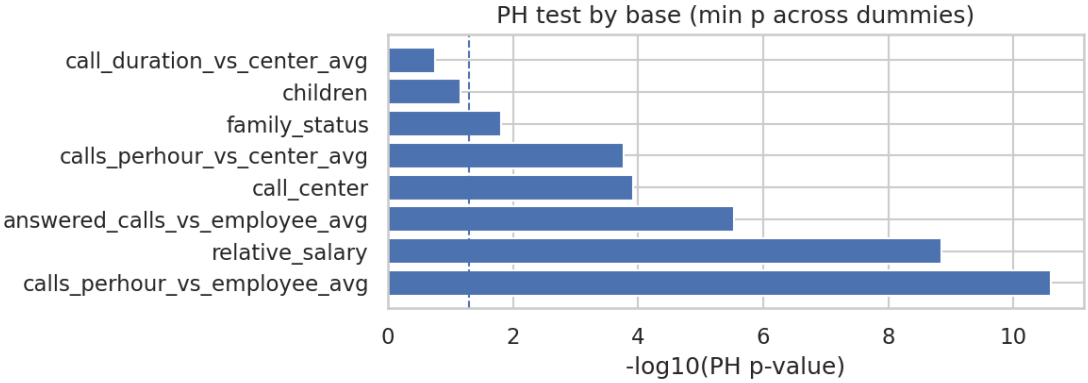
טבלה 8 : **תוצאות מודל** Cox Proportional Hazards **עם בחירת משתנים בשיטת**– LASSO
יחסי סיכון ,(HR) **רווחי סמך** 95% **וערכי מובהקות**

מדד	תוצאה
RF AUC (5-fold CV)	0.008 ± 0.839
RF in-sample AUC (sanity only)	0.955
משתנים שנבחרו (≤60% או טופ 10):	אחוז משרה, שכר יחסי, גיל בשנים, הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה במוקד כולו, הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה של העובד בשנה האחרונה, הפרש בין ממוצע משך השיחה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע משך השיחה במוקד כולו, הפרש בין ממוצע שיחות לשעה של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות לשעה של העובד בשנה אחרונה, הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות במוקד כולו, הפרש בין ממוצע שיחות נענות של העובד ב-3 חודשים לפני סיום מעקב לבין ממוצע שיחות נענות של העובד בשנה אחרונה, מצב משפחתי
Cox on RF-features: CV C-index	0.012 ± 0.700
PH global p (median of mins)	1.82e-07

טבלה 9 : **ממצאי** Random Forest

בדיקות ההנחות של מודל Cox.

בדיקת פרופורציונליות (Proportional Hazards – PH)

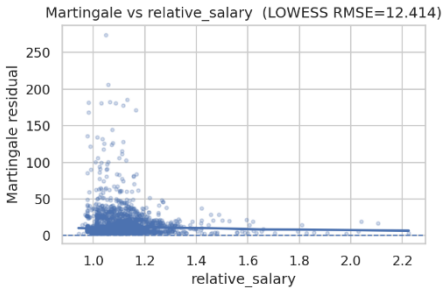


תרשים 16 : **בדיקת פרופורציונליות (Proportional Hazards – PH)**

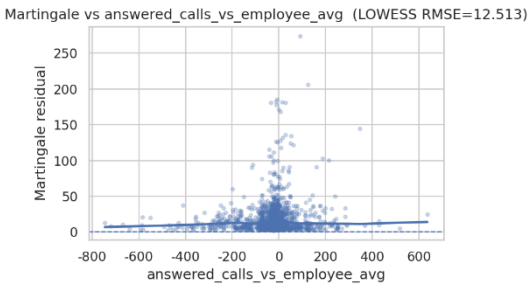
באמצעות מבחן Schoenfeld נבדקה הנחת הפרופורציונליות של יחסי הסיכונים. תרשים 16 מציג לכל משתנה את ערך ה־ pמהמבחן, לאחר המרה ל- $log_{10}(p)$. ככל שהעמודה גבוהה יותר ה- p קטן יותר. הקו המקווקו מסמן את סף המובהקות המקובל $p = 0.05$ (כלומר, $log_{10}(0.05) \approx$ 1.3). עמודות שמשמאל לקו (מתחת לסף) : ערכי p גבוהים, ולכן אין עדות להפרת ההנחה (כלומר, ההנחה מתקיימת). עמודות שחורגות ימינה מהקו (מעל הסף) : ערכי p נמוכים, מה שמעיד על אפשרות של הפרת הנחת הפרופורציונליות.

ברמה הגלובלית ההנחה נמצאה סבירה, אולם במספר משתנים (כגון relative_salary, answered_calls_vs_employee_avg, calls_perhour_vs_employee_avg), נמצאו ערכי p.value נמוכים המעידים על סטיות מסוימות (תרשים 16). עם זאת, ההפרות לא היו דרמטיות ולכן המשתנים נשמרו במודל, תוך פרשנות זהירה לכך שייתכן והשפעתם משתנה לאורך זמן. משתנים דמוגרפיים (כגון ילדים ומצב משפחתי) לא הראו עדות להפרה.

בדיקת ליניאריות



תרשים 17 : **בדיקת ליניאריות למשתנה שכר יחסי**



תרשים 18 : **בדיקת ליניאריות למשתנה** answered_calls_vs_employee_avg

הנחת הליניאריות בין משתנים רציפים לבין log-hazard נבדקה באמצעות שאריות מרטינגלים (תרשימים 17-21). עבור relative_salary התקבלה התאמה טובה (קו LOWESS סביב אפס, ללא עיוות משמעותי). לעומת זאת, עבור משתני KPI (כמו answered_calls_vs_employee_avg ו־calls_perhour_vs_employee_avg) נצפו סטיות קלות המעידות על כך שייתכן קשר לא ליניארי מלא. מאחר והסטיות אינן חריפות, המשתנים נותרו במודל תוך ציון מגבלה זו.

בדיקת מולטיקולינאריות (VIF)

התקבלה בדיקת VIF עבור משתני המודל בלבד (טבלה 6). כל הערכים היו מתחת לסף המקובל של 10, ולפיכך לא נמצאה בעיית קולינאריות חמורה. משתנה relative_salary הראה ערך גבוה יחסית אך עדיין בגבולות הסביר.

	feature	VIF
2	children	1.951738
1	family_status	1.913371
7	calls_perhour_vs_center_avg	1.253857
5	calls_perhour_vs_employee_avg	1.210550
0	relative_salary	1.141486
3	call_center	1.047939
6	call_duration_vs_center_avg	1.044843
4	answered_calls_vs_employee_avg	1.039110

טבלה 6 : VIF **רק על משתני המודל**

אירועים לכל משתנה (EPV)

במודל התקבל יחס של כ־103 אירועים לכל פרמטר (1,339 עזיבות, 13 פרמטרים), נתון העולה בהרבה על כלל האצבע המקובל (≥10). לכן אין חשש לחוסר יציבות סטטיסטית .

(EPV) יחס אירועים לפרמטר	מספר פרמטרים במודל	מספר אירועים
103.0	13	1339

טבלה 7 : **יחס אירועים לפרמטר EPV**

תצפיות משפיעות

נבדקו שאריות dfbeta (תרשים 22 בנספח). נמצאו מספר תצפיות בודדות עם השפעה יחסית גבוהה (כגון תצפית 590), אך הפער לא דרמטי ולא מצדיק הסרה. המסקנה היא שהמודל אינו מונע על ידי תצפיות חריגות בודדות.

קוד המקור של הפרוייקט נמצא ב-*repository* של מחברת העבודה בקישור הבא:

https://github.com/Kogaines/Employee-Attrition-Cox-Model