**פרויקט BI**

**מגישים:**יובל מרסל ניני 207126301  
מתן רוטשטיין 312481526  
עדן טרובישי 316041169  
לישי בצר 209229061

1. **שאלה עסקית:**

מהי השפעת מגדר וניסיון של שנים על פערי שכר בין סוגי משרות שונות?

**מטרת הפרויקט:**

מטרת הפרויקט לחקור את יחסי הגומלין בין מגדר, שנות ניסיון ופערי שכר בסוגי משרות שונות. על ידי בחינת הקשר בין הגורמים הללו, ארגונים יכולים לקבל תובנות לגבי פערי שכר פוטנציאליים ולהבין אם יש מתאם ישיר בין שנות ניסיון וגידול בשכר לגברים ולנשים כאחד, על פני תפקידי עבודה שונים. הניתוח של שאלה זו יכול לעזור לארגונים לזהות אי שוויון כלשהם ולנקוט בצעדים יזומים כדי להבטיח שיטות פיצויים הוגנים, לקדם גיוון ושיווין מגדרי ובכך לספק הזדמנויות שוות לצמיחת קריירה. בנוסף, הפרויקט רלוונטי גם עבור מחפשי עבודה, לטובת השוואות שכר ובחירת עבודה בהתאם לסוג המשרה ושנות הניסיון שלהם.

**נחלק את השאלה העסקית לשתי שאלות עיקריות:**

**תת שאלה עסקית 1:** האם קיים פער בשכר בין גברים לנשים בהתאמה לסוג המשרה?

**KPI:**

1. **פערי שכר בין המינים לפי ענף:** KPI זה ימדוד את פערי השכר הממוצעים בין גברים ונשים בתעשיות שונות, כגון פיננסים, בריאות, טכנולוגיה או קמעונאות. הנתונים ייאספו ממדגם מייצג של עובדים בכל ענף וינותחו כדי לזהות פערים משמעותיים בשכר.

התאמה למודל SMART:

ספציפי: KPI זה ספציפי למדידת הפרש השכר הממוצע בין גברים ונשים בתעשיות שונות.

ניתן למדידה: ניתן לאסוף את הנתונים ממדגם מייצג של עובדים ולנתח אותם כדי לקבוע את פער השכר הממוצע לפי ענף.

בר השגה: KPI זה בר השגה מכיוון שניתן לאסוף ולנתח נתונים על פערי שכר לפי ענף.

רלוונטי: KPI זה רלוונטי לשאלת המחקר שכן הוא מתייחס ישירות האם קיים פער שכר בין גברים לנשים לפי סוג העבודה.

מוגבל בזמן: יהיה צורך לאסוף את הנתונים בתוך מסגרת זמן מסוימת כדי להבטיח דיוק ורלוונטיות.

1. **פער שכר מגדרי לפי סוג משרה:** KPI זה ישווה את השכר של גברים ונשים בקטגוריות עבודה שונות, . זה יעזור לזהות אם משרות או תפקידים מסוימים נוטים להיות מוטים יותר כלפי מגדר אחד מאשר לאחר.

התאמה למודל SMART:

ספציפי: KPI זה ספציפי להשוואת השכר של גברים ונשים בקטגוריות עבודה שונות.

ניתן למדידה: ניתן לאסוף את הנתונים ממדגם מייצג של עובדים ולנתח אותם כדי לקבוע את פער השכר הממוצע לפי סוג משרה.

בר השגה: KPI זה בר השגה מכיוון שניתן לאסוף ולנתח נתונים על פערי שכר לפי סוג עבודה.

רלוונטי: KPI זה רלוונטי לשאלת המחקר שכן הוא מתייחס לשאלה האם משרות או תפקידים מסוימים נוטים להיות מוטים יותר כלפי מגדר אחד מאשר לאחר.

מוגבל בזמן: יהיה צורך לאסוף את הנתונים בתוך מסגרת זמן מסוימת כדי להבטיח דיוק ורלוונטיות.

1. **פערי שכר מגדריים לפי רמת השכלה:** KPI זה ימדוד את הפרש השכר הממוצע בין גברים ונשים עם רמות השכלה דומות. לדוגמה, זה ישווה את השכר של עובדות ועובדות עם תואר ראשון או תארים מתקדמים באותו תחום כדי לזהות אם לרמת ההשכלה יש השפעה כלשהי על פערי השכר בין המינים.

התאמה למודל SMART:  
ספציפי: KPI זה ספציפי למדידת הפרש השכר הממוצע בין גברים ונשים עם רמות השכלה דומות.

ניתן למדידה: ניתן לאסוף את הנתונים ממדגם מייצג של עובדים ולנתח אותם כדי לקבוע את פער השכר הממוצע לפי רמת השכלה.

בר השגה: KPI זה בר השגה מכיוון שניתן לאסוף ולנתח נתונים על פערי שכר לפי רמת השכלה.

רלוונטי: KPI זה רלוונטי לשאלת המחקר שכן הוא מתייחס לשאלה האם לרמת ההשכלה יש השפעה כלשהי על פערי השכר בין המינים.

מוגבל בזמן: יהיה צורך לאסוף את הנתונים בתוך מסגרת זמן מסוימת כדי להבטיח דיוק ורלוונטיות.

1. **שיעור גברים ונשים בסוגי משרות**: KPI זה מודד את התפלגות גברים ונשים על פני סוגים שונים של משרות בארגון. הוא מספק תובנות לגבי הייצוג המגדרי בתפקידי עבודה שונים ועוזר לזהות כל חוסר איזון או פער מגדרי בסוגי עבודה ספציפיים. על ידי כימת שיעור הגברים והנשים בכל קטגוריית עבודה, ארגונים יכולים להעריך את הגיוון המגדרי ולקבל החלטות מושכלות כדי לקדם שוויון והכלה.

התאמה למודל SMART:

ספציפי: קובע את שיעור הגברים והנשים בסוגי עבודה שונים.

מדיד: מכמת את הייצוג של גברים ונשים בכל סוג עבודה, מה שמאפשר מדידה וניתוח אובייקטיביים.

בר השגה: ניתן להשגה על ידי ניתוח נתוני המגדר וסוגי התפקידים כדי לחשב את שיעור הגברים והנשים.

רלוונטי: רלוונטי לשאלת המחקר שכן היא חוקרת את ההתפלגות המגדרית בין סוגי משרות.

זמן מוגבל: מספק מדד מוגבל בזמן של הייצוג בנקודת זמן מסוימת.

1. **תוספת שכר לפי מין ושנות ניסיון:** KPI זה בוחן את הקשר בין עליית שכר, מגדר וניסיון של שנים. הוא מחשב את העלייה הממוצעת או האחוז בשכר עבור גברים ונשים כאחד, על סמך שנות הניסיון שלהם. על ידי ניתוח נתונים אלה, ארגונים יכולים לזהות אם יש פערים כלשהם בצמיחת השכר בין המינים ולקבוע אם לשנים של ניסיון יש השפעה ישירה על התקדמות השכר עבור גברים ונשים כאחד.

התאמה למודל SMART:

ספציפי: בוחן את העלאת השכר לגברים ולנשים על סמך שנות הניסיון שלהם.

מדיד: מכמת את העלאת השכר, ומאפשר מדידה והשוואה אובייקטיבית.

בר השגה: בר השגה על ידי ניתוח נתוני השכר לגברים ונשים ברמות ניסיון שונות וחישוב התוספת.

רלוונטי: רלוונטי לשאלת המחקר שכן היא חוקרת את הקשר בין מגדר, ניסיון ותוספת שכר.

זמן מוגבל: מספק מדד מוגבל בזמן של העלייה הקשורה לרמות ניסיון שונות עבור כל מגדר.

**תת שאלה עסקית 2:** האם קיים קשר ישיר בין כמות שנות הניסיון לבין עלייה בשכר?

**KPI:**

1. **עליית שכר לפי שנות ניסיון:** KPI זה ימדוד את עליית השכר הממוצעת עבור אנשים עם רמות ניסיון שונות, כגון 1-5 שנים. הנתונים ייאספו ממדגם מייצג של עובדים בענף מסוים או בקטגוריית עבודה מסוימת וינותחו כדי לקבוע אם יש מתאם ישיר בין שנות ניסיון לשכר.

**התאמה למודל SMART:**

ספציפי: KPI זה ספציפי למדידת הקשר בין שנות ניסיון לעלייה בשכר.

ניתן למדידה: ניתן לאסוף את הנתונים ממדגם מייצג של עובדים ולנתח אותם כדי לקבוע את עליית השכר הממוצעת עבור רמות ניסיון שונות.

בר השגה: KPI זה בר השגה מכיוון שניתן לאסוף ולנתח נתונים על עליות שכר לפי שנות ניסיון.

רלוונטי: KPI זה רלוונטי לשאלת המחקר שכן הוא מתייחס ישירות לקשר בין ניסיון ושכר.

מוגבל בזמן: יהיה צורך לאסוף את הנתונים בתוך מסגרת זמן מסוימת כדי להבטיח דיוק ורלוונטיות.

1. **עליית שכר לפי כותרת משרה:** KPI זה ישווה את עליית השכר הממוצעת עבור אנשים בעלי אותו תפקיד אך רמות ניסיון שונות. לדוגמה, זה ישווה את השכר של מהנדסי תוכנה ברמת התחלה עם ניסיון של 1-2 שנים לאלה עם ניסיון של 3-5 שנים. KPI זה יעזור לזהות האם יש דפוס עקבי של עליות שכר על פני כותרות עבודה שונות או אם הוא משתנה בהתאם לענף או לחברה.

התאמה למודל SMART:

ספציפי: KPI זה ספציפי להשוואת עליות שכר עבור אנשים בעלי אותו תפקיד אך רמות שונות של ניסיון.

ניתן למדידה: ניתן לאסוף את הנתונים ממדגם מייצג של עובדים ולנתח אותם כדי לקבוע את עליית השכר הממוצעת עבור כל תפקיד.

בר השגה: KPI זה בר השגה מכיוון שניתן לאסוף ולנתח נתונים על עליות שכר לפי כותרת תפקיד.

רלוונטי: KPI זה רלוונטי לשאלת המחקר שכן הוא מתייחס לשאלה האם הקשר בין ניסיון ושכר משתנה בין כותרות עבודה שונות.

מוגבל בזמן: יהיה צורך לאסוף את הנתונים בתוך מסגרת זמן מסוימת כדי להבטיח דיוק ורלוונטיות.

1. **עליית שכר לפי מגדר ושנות ניסיון:** KPI זה ימדוד את עליית השכר הממוצעת לעובדות ועובדות עם רמות ניסיון שונות, כגון 1-2 שנים, 3-5 שנים או 5+ שנים. הנתונים ייאספו ממדגם מייצג של עובדים בענף מסוים או בקטגוריית עבודה מסוימת וינותחו כדי לקבוע אם יש הבדלים בעליות שכר בין עובדות ועובדות ברמות ניסיון שונות. KPI זה יעזור לזהות האם מגדר משחק תפקיד ביחס בין ניסיון ושכר.

התאמה למודל SMART:

ספציפי: KPI זה ספציפי למדידת עליית השכר הממוצעת לעובדות ועובדות עם רמות שונות של ניסיון.

ניתן למדידה: ניתן לאסוף את הנתונים ממדגם מייצג של עובדים ולנתח אותם כדי לקבוע את עליית השכר הממוצעת לכל מגדר לפי שנות ניסיון.

בר השגה: KPI זה בר השגה מכיוון שניתן לאסוף ולנתח נתונים על עליות שכר לפי מגדר ושנות ניסיון.

רלוונטי: KPI זה רלוונטי לשאלת המחקר שכן הוא מתייחס האם ישנם הבדלים כלשהם בעליות השכר בין עובדות ועובדות ברמות שונות של ניסיון.

מוגבל בזמן: יהיה צורך לאסוף את הנתונים בתוך מסגרת זמן מסוימת כדי להבטיח דיוק ורלוונטיות.

1. **תוספת שכר על סמך הערכת ביצועים**:

KPI זה מעריך את עליית השכר על סמך דירוגי הערכת ביצועים. הוא מודד את העלייה הממוצעת או האחוזים בשכר לעובדים עם רמות הערכת ביצועים שונות (למשל, נמוכה, בינונית, גבוהה). על ידי ניתוח נתונים אלה, ארגונים יכולים להעריך האם יש מתאם בין דירוגי הערכת ביצועים וגידול בשכר. KPI זה עוזר לארגונים לזהות אם תהליך הערכת הביצועים מותאם לתוספות השכר ומספק תובנות לגבי הטיות או חוסר עקביות פוטנציאליות.

התאמה למודל SMART:

ספציפי: מנתח את עליית השכר בהתבסס על דירוגי הערכת ביצועים (למשל, נמוך, בינוני, גבוה).

מדיד: מכמת את עליית השכר עבור כל דירוג הערכת ביצועים, מה שמאפשר מדידה והשוואה אובייקטיבית.

בר השגה: ניתן להשגה על ידי ניתוח נתוני השכר על פני דירוגי הערכת ביצועים שונים וחישוב העלייה.

רלוונטי: רלוונטי לשאלת המחקר שכן היא בוחנת את הקשר בין הערכת ביצועים לעלייה בשכר.

זמן מוגבל: מספק מדד מוגבל בזמן של העלייה הקשורה לכל דירוג הערכת ביצועים.

1. **גידול בשכר לפי מחלקה:**

KPI זה בוחן את העלאת השכר לפי מחלקות בתוך הארגון. הוא מודד את הגידול הממוצע או האחוז בשכר לעובדים במחלקות שונות. על ידי השוואת שיעורי הגידול בשכר בין מחלקות, ארגונים יכולים לזהות שינויים בתוספות השכר ובפערים פוטנציאליים. KPI זה עוזר לארגונים להבין אם למחלקות מסוימות יש גידול שכר גבוה יותר מאחרות ומאפשר להם לנקוט באמצעי תיקון במידת הצורך כדי להבטיח העלאת שכר הוגנת ועקבית בכל המחלקות.

התאמה למודל SMART:

ספציפי: חוקר את הגידול במשכורות במחלקות שונות.

מדיד: מכמת את הגידול בשכר עבור כל מחלקה, ומאפשר מדידה וניתוח אובייקטיביים.

בר השגה: ניתן להשגה על ידי ניתוח נתוני השכר בין מחלקות שונות וחישוב שיעורי הצמיחה.

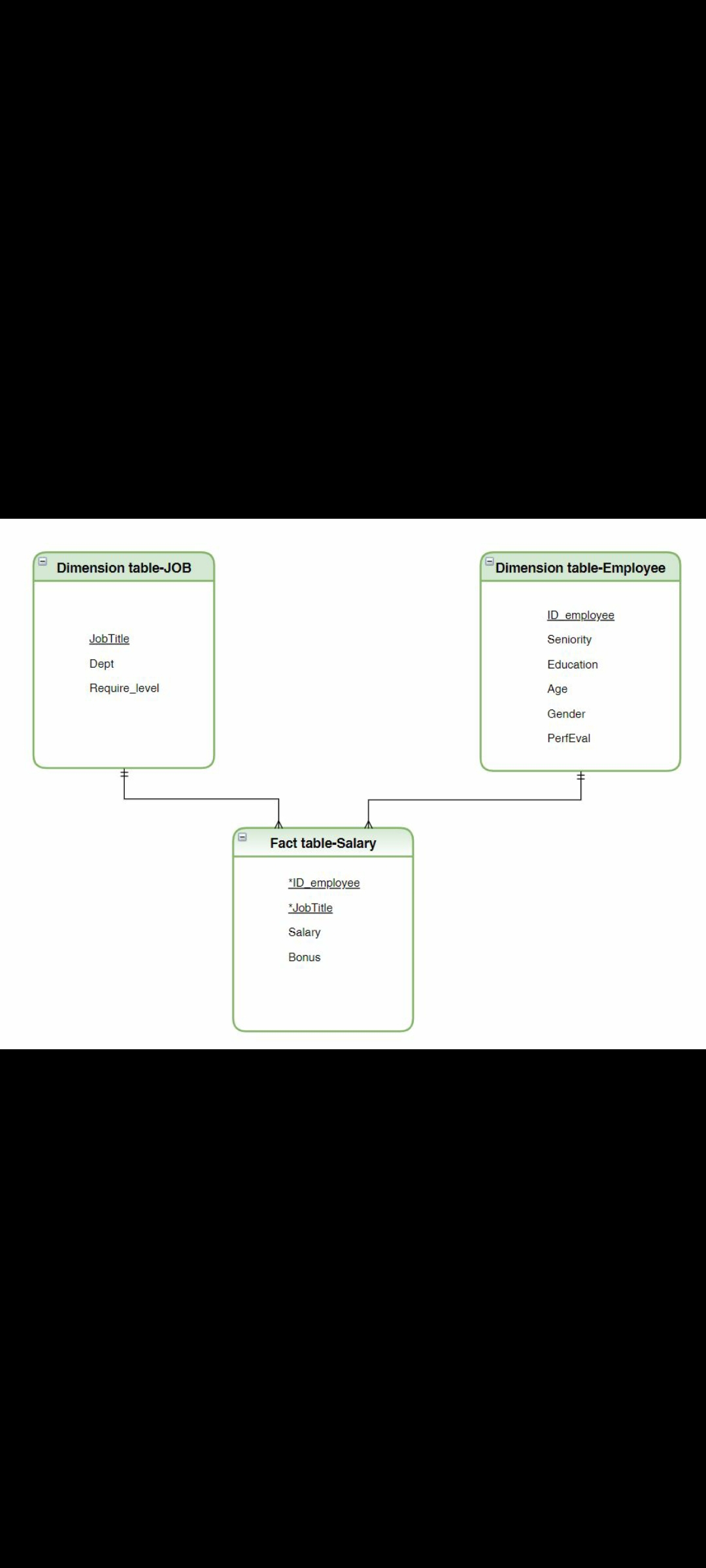
רלוונטי: רלוונטי לשאלת המחקר שכן היא חוקרת את הקשר בין המחלקה לגידול בשכר.

זמן מוגבל: מספק מדד מוגבל בזמן של קצב הצמיחה בתוך כל מחלקה.

1. **הגדרת Data Warehouse:**
   1. **בחירת סכמה:**

הסכמה שאנו בוחרים לתהליך הETL הינה: Star. יש לנו טבלת נתונים יחידה ופשוטה ולכן לא נדרשת בחירת סכמה מורכבת. סכמת star נבחרה בשל הפשטות, יכולות הפרדת הנתונים והגמישות שלה, התורמים לעיבוד נתונים יעיל, ניהול ותחזוקת המערכת.

* 1. **תיאור הData Warehouse:**



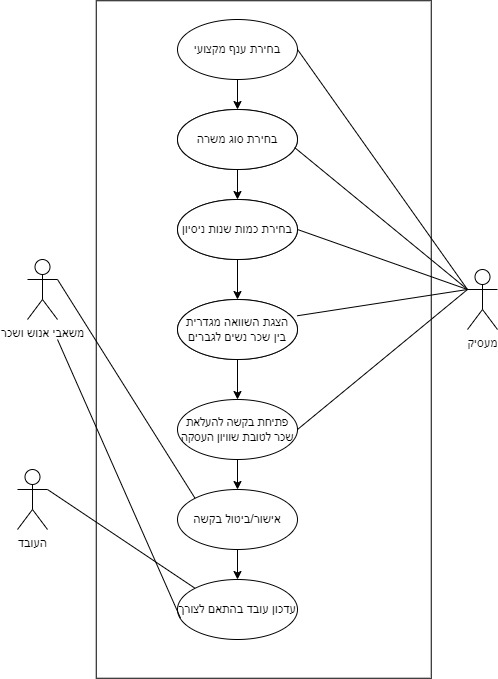
* 1. **Use-caseהמתאר מדוע בחירתנו נכונה לשם קבלת החלטות למשתמש:**
     1. **Use-case ראשון:** תהליך חיפוש משרה ובקשה לראיון

A picture containing diagram, sketch, drawing, line

Description automatically generated

תרשים הUSE CASE שהצגנו יעיל ועוזר בקבלת החלטות המשתמש מהסיבות הבאות:

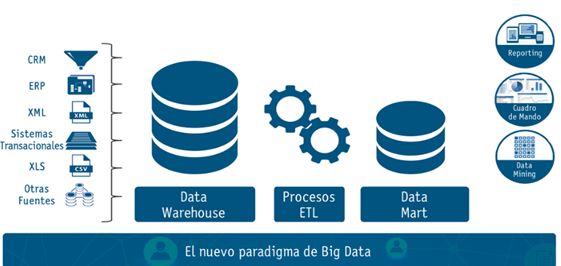
* התרשים מתאר בבירור את השלבים העוקבים בתהליך חיפוש העבודה ומסייע למשתמש לנווט בתהליך ללא בלבול.
* התרשים מאפשר למשתמש לסנן ולהציג הזדמנויות עבודה ספציפיות לתחומי העניין והכישורים שלו, מקל על קבלת החלטות על ידי הצגת אפשרויות עבודה רלוונטיות.
* התרשים מספק למשתמש מידע מקיף על משרה מסוימת, עוזר לקבל החלטות מושכלות בנוגע לקבלת ראיון עבודה ולתאימות לדרישות המשרה.
* התרשים מאפשר למשתמש להביע את התעניינותו בעבודה מסוימת ולהתחיל את שלב הגיוס, מסמיך אותו לקראת הבטחת הזדמנות עבודה ומעניק לו תחושת שליטה ומעורבות בתהליך.
* התרשים מזהה את בעלי העניין השונים בתהליך חיפוש העבודה כמו מחפשי העבודה, חברות ההשמה והמעסיקים, משפר את השקיפות ומאפשר תקשורת יעילה.
  + 1. **Use-case שני:** תהליך השוואת שכר ובקשה להעלאה (תוך ארגוני)



תרשים הUSE CASE שהצגנו יעיל ועוזר בקבלת החלטות המשתמש המעסיק מהסיבות הבאות:

* התרשים מתאר בבירור את השלבים העוקבים בתהליך תוך ארגוני של המעסיקים בכדי להבין האם לעובדים שלהם מוענק שכר שקיים בו אי שיווין מגדרי-תעסוקתי בהתאם לסוג המשרה ומין העובד.
* התרשים מאפשר למשתמש המעסיק לסנן ולהציג את שכר העובד בהתאם ל-ענף המקצועי שלו, שנות הניסיון שלו והמין שלו.
* למשתמש המעסיק תוצג השוואת שכר מגדרית בהתאם לסינון ולקריטריונים שבחר, ולאחר מכן יוכל להחליט אם קיים צורך להעלאת שכר לטובת שוויון העסקה. הבקשה תעלה למשאבי אנוש ואחראי השכר בארגון. באם הבקשה תאושר, העובד יעודכן בהחלטת העלאת שכרו.
* התהליך מאפשר לחברות לעשות תהליך של בחינה והשוואת עצמית ותוך ארגונית בכדי לממש העסקה שוויונית הן מבחינה מגדרית והן מבחינת ניסיון העובד.

1. **תהליך הETL:**



* **חילוץ:** ניגשנו לאתר Kaggle וחיפשנו סט נתונים בהתאם לנושא שבחרנו לחקור. לאחר מציאת סט נתונים עשיר ומגוון בסוגי השדות המוכלים בו, הורדנו את סט הנתונים קובץ מסוג CSV ופתחנו אותו בעזרת תוכנת Excel.
* **טרנספורמציה:**
* **הסרת כפילויות**
* **טיפול בערכים חסרים וחריגים**
* **הבטחת שלמות הנתונים-**

המרת ערכי מגדר לערכים סטנדרטיים – זכר, נקבה.

הבטחת ערכים עקביים לגיל, שכר, ותק.

* **ביצוע חישוב נתונים-**

על סמך וותק עובד הגדרנו את רמת העובד: וותק של 1-2 שנים: junior, 3-4 medium , 5 שנים ומעלה: senior.

* **הכללת פעולת הTransform ומימוש תהליך על ידי פונקציות קיימות עבור כלל העמודות:**

עמודת JobTitle : שימוש בפונקציית CLEAN\_STRING   
עמודת Salary : שימוש בפונקציית CLEAN\_NUM

עמודת Bonus : שימוש בפונקציית CLEAN\_ NUM

עמודת ID\_employee : שימוש בפונקציית CLEAN\_ NUM

עמודת Gender : שימוש בפונקציית CAPITAL\_LETTER

עמודת Age : שימוש בפונקציית CLEAN\_ NUM

עמודת PerfEval : שימוש בפונקציית CLEAN\_ NUM

עמודת Education : שימוש בפונקציית CAPITAL\_LETTER

עמודת Dept : שימוש בפונקציית CAPITAL\_LETTER

עמודת Require\_level : שימוש בפונקציית CAPITAL\_LETTER

עמודת Seniority : שימוש בפונקציית CLEAN\_ NUM

* **טעינה:**
* **קביעת סכימת היעד- סכימת כוכב.**

הגדרנו את הטבלאות, העמודות והקשרים על סמך הדרישות העסקיות שהגדרנו.

יצרנו תכונות מתאימות בטבלאות העובדות והממדים כדי לייעל את ביצועי השאילתות. בין התכונות הגדרנו מפתחות ראשיים, מפתחות זרים ותכונות אחרות המבוססות על דפוסי השימוש.

* **מיפוי וטעינת הנתונים-**

מיפינו את הנתונים שעברו טרנספורמציה לסכימת הכוכב וביצענו טעינה לטבלאות המתאימות.

* **בדיקת איכות הנתונים-**

אמתנו את איכות הנתונים, בדקנו האם יש חוסר עקביות, ערכים חסרים או חריגים.

1. **ניתוח הData Warehouse:**

הוגדרו 10 שאילתות SQL העוזרות לענות על השאלה העסקית.

1. **מסקנות:**

מסקנות וממצאים מתוך השאילתות:

* עבור ענפים שונים ואף סוגים שונים של משרות, מצאנו כי קיים פער משמעותי בין גברים לנשים. השכר של הגברים היה גבוה משל הנשים.
* קיימים פערים בין גברים לנשים במשרות זהות ובהתאמה לכמות שנות ניסיון. נשים קיבלו משכורות נמוכות יותר מגברים גם כאשר כמות שנות הניסיון שלהם והמשרה הייתה זהה.
* קיימים פערי שכר בין גברים לנשים גם כאשר רמת ההשכלה שלהם זהה. יש לציין כי ההפרש בין השכר של גברים ונשים דומה בכלל רמות ההשכלה.
* קיימת עלייה בשכר ככל שמספר שנות הניסיון עולה, הן בהשוואה בין הענפים השונים והן בהשוואה בין המשרות השונות.