

### מטלה 3 – בינה עסקית

המטלה נכתבה בראי תכנון דשבורד עתידי עבור הלקוח.

שאלות:

1. היחס בין מספר האשפוזים לבין מספר השורות המופיעות למטופל (1.1KPI).  
מבנה כללי: התבססות על טבלת FACT, רישום שאילתא ראשונה בעזרת window function עבור כל מספר מזהה של מטופל את היחס - מספר הפעמים שמופיע '1' בעמודת routine\_test חלקי מספר השורות הכללי שקיימות למטופל ע"י שימוש בפונקציית count. לשאילתא נבצע השמה בטבלה זמנית (שימוש ב-WITH) עליה נריץ שאילתא נוספות שתחשב ממוצע לערך עמודת היחס ובכך נקבל את המספר הסופי.
2. ערך מקדם המתאם שמתאר את עוצמת הקשר שבין הנתונים לבין משתנה המטרה (אשפוז).  
(2.1KPI) – עשרת המשתנים בעלי הערך הגבוהה ביותר ביחד למשתנה המטרה. (נדרש למצוא מבחן התאמה חד כיווני לפני כתיבת השאילתא)
- מבנה כללי: התבססות על טבלת main\_tests\_result לאחר קבלת הנוסחה למבחן ההתאמה. נזין את הנוסחה לפקודת SELECT שתבצע חישוב של ערך המתאם לכל עמודה (שימוש ב-window function לכל עמודה וקבלת הציון) \* בשלב זה נדרש לחשוב איך ואיפה הציונים יוצגו משום שאנו מבצעים את הפעולה על עמודות (הרעיון שכרגע עלה לנו הוא לבצע השמה של טבלה נוספת ובה נשמור את בעמודה הראשונה את שמות העמודות של טבלת main\_tests\_result ועמודה השניה את הציון של מתאם המדגם (שיבחר). לאחר מכן ניתן לסדר את השורות בטבלה בסדר יורד ולהשתמש בתכונת LIMIT(10) בכדי להציג רק את עשרת המשתנים בעלי ציון המתאם הגבוה ביותר.
3. תוצאת אינטרוול זמן – בדיקות שבוצעו מעבר לאינטרוול של תקופת זמן שתבחר (בחודשים/שנים) לפני האשפוז לא ישפיעו על הקשר בין המשתנה לאשפוז. (2.3KPI)
- מבנה כללי: בעת ביצוע תהליך ה-ETL אנו מתכוונים לסנן אשפוזים שהזמן שעבר בינם לבין הבדיקה האחרונה שבוצעה למטופל גדולה מערך מסוים. התבססות על טבלת FACT. מטרת השאילתא היא להראות תמיד ללקוח מבחינת הדאטה שמוצג לו לאחר הסינון מה ממוצע משך הזמן שכעת מתקיים בין אשפוז לבדיקה. בכדי לעשות זאת אנו נדרשים לכתוב שאילתת window function שתחשב עבור מטופל את משך הזמן הממוצע בין כל אשפוז שביצע לבין הבדיקה האחרונה שקדמו לו. החישוב יתבסס על רכיב התאריך של עמודת DATE ע"י חיסור התאריכים בימים. לאחר מכן נכניס את פלט השאילתא לשאילתא המרכזית שתבצע חישוב ממוצע של משך הזמן לטבלה כולה.
4. משך זמן האשפוז הממוצע.
- מבנה כללי: שימוש בטבלת additional\_hospitalizaion\_data, תחילה נדרש לחשב לכל בעמודה חדשה לכל שורת אשפוז בנתונים את משך האשפוז ע"י החסרה בין עמודת 'endshpuz' לבין עמודת 'startshpuz'. לאחר מכן נבנה שאילתה מבוססת window\_function שתאגד לכל מזהה מטופל את סכום ערכי משך האשפוז. לסיום נכניס את הפלט לשאילתא שתשלוף את הערך הממוצע לכל סכומי האפשוז וזו תהיה התוצאה – ממוצע משך האשפוזים לנתונים.
- הערה: שאילתות 5+6 יוצגו אחת ליד השניה ואופן יישומן כמעט זהה.
5. עליות חריגות של כמות אשפוזים בגיל מסוים (גרף של מספר האשפוזים כתלות בגיל המטופלים)
- מבנה כללי: התבססות על טבלת FACT. שימוש בתנאי על עמודת 'TEST\_ROUTINE' בדרישה שערכה יהיה '1' כדי לקבל את שורות האשפוזים בלבד. לאחר מכן נשתמש ב-window function על עמודת AGE ובכך נקבל את כמות האשפוזים לפי קבוצות גיל. הלקוח יוכל לזהות מגמות של כמות האשפוזים לפי גיל ולהבין האם בגיל מסוים ישנה עליה בכמות האשפוזים. זה יכול למקד אותו בתהליך קבלת החלטות לקבוצות המטרה באוכלוסייה.

6. ממוצע מספר הבדיקות כתלות בגיל המטופלים  
מבנה כללי: התבססות על טבלת FACT. שימוש בתכונה COUNT על עמודת 'TEST\_ROUTINE'  
בדרישה שערכה יהיה '0' כדי לקבל את כמות הבדיקות בלבד. לאחר מכן נשתמש ב- window  
function על עמודת AGE ובכך נקבל את כמות הבדיקות לפי קבוצות גיל נבחרת. הלקוח יוכל לזהות  
מגמות של כמות הבדיקות לפי גיל ולהבין האם בגיל מסוים ישנה עליה בכמות הבדיקות שמתבצעת.  
זה יכול למקד אותו בתהליך קבלת החלטות לקבוצות המטרה באוכלוסייה.

7. גרף שמציג את סך מקרי האשפוז לפי סיבת האשפוז  
מבנה כללי: ביצוע JOIN בין טבלת FACT ו- additional\_hospitalizaion\_data ע"י שימוש בתנאי שיוויון  
בעמודות NUMBER\_PATIENT+DATE. לאחר מכן ע"י איחוד באמצעות תכונת WINDOW FUNTION  
של סיבת האשפוז נוכל לקבל את סיבת האשפוז, מתכונת \* COUNT נקבל את מספר מקרי האשפוז  
מסיבה זו.

נתונים נוספים שיוצגו בדשבורד:

1. ערך הסף שנבחר לכמות השורות המינימלית (בדיקות ואשפוזים) של מטופל בכדי להכנס למדגם.  
(1.2KPI)
2. ערך מינימום לאשפוזים. (2.2KPI)
3. חסם לאינטרוול הזמן – בדיקות שבוצעו מעבר לאינטרוול של תקופת זמן שתבחר (בחודשים/שנים)  
לפני האשפוז לא ישפיעו על הקשר בין המשתנה לאשפוז. (2.3KPI)
4. התאריך האחרון שבו עודכנו הנתונים.
5. התאריך האחרון שבו בוצע תהליך ה-ETL.
6. המספר הכולל של מטופלים במערכת הנתונים.