

בינה עסקית**מבחן סופי - מועד ב'****סמסטר א' תשע"ח**

מרצה: פרופסור אילן שמשוני

מתרגל: מר אבנר אינוז

משך הבחינה שלוש שעות.

ניתן להשתמש בכל חומר עזר כתוב ומחשב כיס.

בבחינה ארבע שאלות.

יש לענות על השאלות במחברת המצורפת.

שים לב כי בידך 5 דפים כולל דף זה.

הבחינה מיועדת לגברים ונשים כאחד ומנוסחת בלשון זכר מטעמי נוחות בלבד.

שאלה	ניקוד מקסימלי	ציון
1	40	
2	25	
3	25	
4	20	
סה"כ	110	

בהצלחה!

(1) Data Warehouse, MDX, Data Warehouse, MDX, ETL (40 נקודות)

חברת התעופה הבינלאומית "התאם-לי-מטוס" מעוניינת לעשות מעקב שוטף על ההוצאות בבניית המטוסים השונים שמייצרת ברחבי העולם לאנשי הצווארון הלבן. לשם כך אתם מתבקשים לתכנן עבורה מחסן נתונים. הנתונים הנחוצים יישלפו ממערכת ה OLTP המשמשת לניהול החברה.

עבור כל מטוס שומרים את קוד המטוס, שמו ואת הארץ והעיר שבה המטוס נבנה. בבניית מטוס נעשה שימוש במספר קבלני משנה הקיימים ברחבי העולם ולכל בניית מטוס יש בנוסף מהנדס ראשי ומעצב פנים. כל הגורמים הנ"ל נחשבים כעצמאים ויכולים לעבוד גם בבניית מטוסים אחרים בחברה.

בכל בנייה של מטוס נעשה שימוש בחומרים שונים מתוך רשימה המתעדכנת על ידי המהנדס הראשי ומעצב הפנים, לכל מטוס מזמינים מכל חומר כמות מסוימת במחיר מסוים (הניחו שהמטבע עימו החברה עושה שימוש הינו דולר והכל כבר מומר למטבע הדולר).

להלן מספר דוחות שהנהלת "התאם-לי-מטוס" מעוניינת לקבל מהמערכת:

- כמות חומרים שהוזמנו מכל סוג במהלך חודש אפריל 2017. לכל חומר יש לציין את הקוד שלו ואת שמו.
- בכל שבוע במהלך שנת 2017 ציין/י מה היתה עלות ההוצאה ($\text{sum}(\text{amount} * \text{price})$) אך ורק מהזמנת חומרים.
- עבור כל קבלן משנה, הצג את הקוד שלו, שמו וכמות הפועלים הממוצעת שהעסיק לחודש ועלותם.
- עבור כל מהנדס ראשי שעלותו גבוהה יותר מסך העלות בגין של 3 מעצבי הפנים המרוויחים הכי הרבה בשנת 2017, יש להציג את קוד של המהנדס הראשי, שמו ואת סך העלות של מהנדס הפנים שהרוויח הכי פחות, על התוצאות להראות באופן הבא:

העלות של מעצב הפנים הנמוכה ביותר	
שם מהנדס ראשי 1	קוד מהנדס ראשי 1
...
שם מהנדס ראשי N	קוד מהנדס ראשי N

- לכל מטוס יש להציג את עלותו הממוצעת בשנת 2017. עלות מטוס היא סכימה של המחיר שמשולם למהנדס ראשי, למעצב פנים, לקבלני המשנה ולחומרים שהשתמשו בהם.
- עבור כל ארץ ועיר במערכת יש להציג את מספר הפועלים החודשית הממוצעת בבניית מטוסים בשנת 2017 לפי חודש. יש להציג רק ערים בהן מספר הפועלים השנתית גדולה מ-30 פועלים.

		צרפת			...	יוון			ישראל	
		...	פריס		רודוס	...	אתונה	תל אביב	הרצליה	...
2017	ינואר	...	12.0	...	35.9	...	20.2	42.3	28.5	...
	פברואר

	דצמבר

ז. החברה חוששת כי בשנה האחרונה ישנה מעילה כספית בבניית מטוסים בערים מסוימות, ולכן מעוניינת לבדוק את עלות החומרים הכוללת שהיה בה שימוש בכל עיר בשנת 2017, שנת 2016 וההפרש ביניהם. מנכ"ל החברה מעוניין להציג רק את הערים שבהם הגידול באחוז היה גבוה מ 10%.

	חיפה	מדריד	לונדון	ווינגטון
2017	50,000,000	121,512,000	111,500,000			42,421,000
2016	40,000,500	100,000,000	101,000,000			35,000,500
Difference	9,999,500	21,512,000	10,500,000			7,420,500

(א) **(10 נקודות)** הגדר/י star schema כך שניתן יהיה לבצע את כל הדוחות הנ"ל. יש להסביר את כל השדות המחושבים.

(ב) **(2 נקודות)** בחר/י באחד המימדים אשר יצרת בסעיף א' אשר ניתן לפרק אותו והצג/י אותו ביחד עם טבלת העובדות בצורת סכמת פתית שלג.

(ג) ממש/י ב MDX את הדוחות הבאים :

i. **(8 נקודות)** דו"ח ד'.

ii. **(8 נקודות)** דו"ח ו'.

iii. **(8 נקודות)** דו"ח ז'.

(ד) **(4 נקודות)** תאר/י במילים כיצד יתבצע תהליך ה ETL ממסד הנתונים התפעולי לסכמה אותה יצרת בסעיף א'. בתיאורך יש להתייחס לפעולות אגרגציה (ממוצע, סכימה, מינימום וכו') פעולות מתמטיות בסיסיות, נתונים לא רלוונטיים ונתונים שאין באפשרותך לדעת על בסיס טבלאות אלו.

(2) אשכול (25 נקודות)

לפני שנתיים נערכו בישראל בחירות. ועדת הבחירות הכללית פרסמה את התוצאות באתר שלה בקובץ excel. הקובץ מייצג טבלה. בכל שורה נתונים מספר הקולות שכל מפלגה קיבלה באותו יישוב.

ש"ס	מ"ר"צ	המשותפת	עבודה	ליכוד	
					עכו
					חיפה
					כפר יאסיף
					יגור

התבקשתם להפעיל מאשכלים על הנתונים. ניתן להגדיר כל נקודה כשורה בטבלה או עמודה.

- (א) מה המשמעות של תוצאות אשכול לפי שורות ושל אשכול לפי עמודות.
- (ב) איזה פונקציות מרחק הייתם מגדירים בין שתי שורות ובין שתי עמודות. ולמה בחרתם אותן.
- (ג) תנו דוגמא אחת לנקודות שנראה לכם שהיו מאושכלות ביחד לכל אחת מבעיות האשכול שהוגדרו לעיל. תנו דוגמאות ספציפיות ונמקו (סיבות פוליטיות דמוגרפיות).
- (ד) באיזה אלגוריתם אשכול שלמדנו הייתם משתמשים ולמה.
- (ה) לכל יישוב נתונים גם מאפיינים דמוגרפיים כמו רמה כלכלית, סוג יישוב: עיר, מושב, קיבוץ, וכו'. בהינתן שתי תוצאות אשכול של שני אלגוריתמים שונים איך הייתם בוחרים ביניהם תוך שימוש בנתונים הדמוגרפיים.
- (ו) איך ניתן לנצל את תוצאות האשכול שקיבלתם לקראת הבחירות הבאות שייערכו עוד שנתיים.

(3) Adaboost (25 נקודות)

נתון dataset מסווג שכל נקודה שייכת לאחת משתי מחלקות. היית רוצה ללמוד מסווג adaboost כשכל מסווג בסיסי הוא עץ החלטה בעומק 1 (אי שוויון אחד).

- (א) איך תבחר את המסווג הבסיסי הראשון.
- (ב) איך תבחר את המסווג הבסיסי השני. האם הוא זהה לראשון?
- (ג) האם ניתן להפעיל את ה adaboost כשיש יותר משתי מחלקות אפשריות? אם כן אז איך תשנה את האלגוריתם כדי שזה יהיה אפשרי?
- (ד) אם נתונים לך כמסווגים פשוטים קבוצה סופית ידועה של עצים בעומק מסוים איך היית מממש את אלגוריתם ה adaboost.

(4) SOM (20 נקודות)

- (א) הסבר מהו SOM חד-מימדי ודו-מימדי (צורת התוצאה).
- (ב) בהינתן נקודת קלט $x(t)$ תאר איך ה SOM החד-מימדי מטפל בנקודה זו.
- (ג) האם האלגוריתם יטפל בנקודה זו שוב במהלך הריצה שלו?
- (ד) נתון הקלט הדו-מימדי הבא (כל צורה היא קבוצה גדולה של נקודות):
- צייר את הפלט של SOM חד-מימדי, SOM דו-מימדי ו Kmeans כש $K=5$.

