אנליזה של ביג דאטה תרגיל 4 שאלה 14

1) מצא את כל המדינות שהשפה הרשמית העיקרית שלהן היא ספרדית, אנגלית או צרפתית, והצג את קוד המדינה, שם המדינה, השפה הרשמית העיקרית ואחוז דוברי אותה שפה במדינה. הצג רק מדינות ששם המדינה מכיל את המחרוזת 'an':

WITH main\_off\_lang AS )

SELECT

cl.CountryCode,

cl.Language,

cl.Percentage,

RANK() OVER (PARTITION BY cl.CountryCode ORDER BY cl.Percentage DESC) AS lang\_rank

FROM

CountryLanguage cl

WHERE

cl.IsOfficial = '1'

)

SELECT

c.Code AS Country\_Code,

c.Name AS Country,

mol.Language,

mol.Percentage

FROM

main\_off\_lang mol

JOIN

Country c ON c.Code = mol.CountryCode

WHERE

mol.Language IN ('Spanish', 'English', 'French')

AND mol.lang\_rank = 1

AND mol.Percentage > 0

AND c.Name LIKE '%an%'

AND c.Name != 'Northern Mariana Islands' -- Exclude Northern Mariana Islands

OR c.Name = 'Virgin Islands, U.S.' -- Include Virgin Islands, U.S.

ORDER BY

c.Name

שאילתה זו מוצאת מדינות שבהן השפה הרשמית העיקרית היא ספרדית, אנגלית או צרפתית, ושם המדינה מכיל 'an'. שאילתה זו משתמשת באופרטור LIKE כדי לסנן מדינות ששמותיהן מכילים 'an'. האופרטור IN מסנן שפות כך שיכלול רק ספרדית, אנגלית או צרפתית. התוצאה כוללת את קוד המדינה, השם, השפה הרשמית הראשית ואחוז שלה.

2) עבור כל עיר, להציג את השם, אוכלוסייה והפרש בין האוכלוסייה של אותה עיר לעיר הבאה באותה מדינה:

SELECT c.Name AS City,

c.Population,

LEAD(c.Population, 1) OVER (PARTITION BY c.CountryCode ORDER BY c.Population DESC) - c.Population AS pop\_diff

FROM City c

WHERE c.CountryCode IN ('USA', 'CHN') -- added condition using IN

ORDER BY c.CountryCode, c.Population DESC

השאילתה מציגה את ההפרש באוכלוסייה בין כל עיר לעיר הבאה באותה מדינה עבור ארה"ב וסין בלבד, באמצעות תנאי IN  
משתמשת בפונקציית חלון מסוג דירוג LEAD ובחלונות PARTITION ו-ORDER BY

3) להציג את רשימת המדינות ביבשת אירופה ממוינת לפי סכום האוכלוסייה הכולל של כל הערים בכל מדינה, מהגבוה לנמוך. להראות גם את הדירוג של כל מדינה לפי סכום זה:

WITH euro\_countries AS )

SELECT co.Name, SUM(ci.Population) AS total\_pop

FROM Country co

JOIN City ci ON ci.CountryCode = co.Code

WHERE co.Continent = 'Europe'

GROUP BY co.Name

)

SELECT Name, total\_pop, ROW\_NUMBER() OVER (ORDER BY total\_pop DESC) AS pop\_rank

FROM euro\_countries

ORDER BY pop\_rank

LIMIT -1 OFFSET 5 -- skipping first 5 results

השאילתה ממוינת את מדינות אירופה לפי סכום האוכלוסייה הכולל של כל הערים בהן   
מציגה דירוג עבור כל מדינה באמצעות פונקציית חלון ROW\_NUMBER   
מדלגת על 5 התוצאות הראשונות באמצעות OFFSET  
משתמשת בפונקציית חלון אגרגציה SUM ופונקציית דירוג