



## FINAL PROJECT

### 1. שאלה עסקית + KPI's

- מטרת פרויקט זה היא לקבוע את הענף בו מערכת Mac זוכה לדירוג גבוה יותר. על ידי מינוף העקרונות של מחסני נתונים ושימוש בתהליכי ETL, ינותח מערך נתונים נרחב. באמצעות ניתוח זה, מטרת הפרויקט היא לזהות את פלח התעשייה שבו מערכת Mac השיגה את הדירוגים הגבוהים ביותר. ממצאים אלה יספקו תובנות חשובות לקבלת החלטות ותכנון אסטרטגי, ויאפשרו לבצע פעולות מושכלות בתעשייה שזוהתה כדי לשפר את הביצועים והתחרותיות של מערכת Mac.
- באיזה תעשייה מערכת Mac מקבלת דירוג גבוה יותר?

#### KPI'S:

- א. דירוג ממוצע של מערכות Mac בכל התעשיות ביחד/בנפרד - בעזרת דירוג נוכל לראות האם המערכת רלוונטית לכל התעשיות ביחד או האם רלוונטית רק לחלק מהתעשיות (ספציפית).
- ב. מדידת נתח השוק של מערכת Mac בכל תעשייה - נבצע בדיקה על ידי סכימה של השימוש במערכת Mac ובכך נוכל להסיק על נתח השוק של המערכת ואף להציע תוכנית לייעול על מנת שהמערכת תהיה בעלת שוק נרחב.
- ג. אחוז מערכות Mac המדורגות גבוה - בעזרת הנתונים ודירוג המערכת נוכל לראות היכן מערכות Mac מדורגות הכי גבוה, היכן קיים לה יתרון בשוק ולהבין איך ניתן לקדם אותה במקומות אחרים על מנת לקבל גם שם אחוז גבוה.

### 2. הגדרת DATA WAREHOUSE

- א. מתוך הסכמות הקיימות ובהתבסס על טבלת הנתונים שבחרנו במטלה הקודמת סיכמת כובב תהיה המתאימה ביותר, מכיוון שהיא מספקת דרך פשוטה ואינטואיטיבית לארגן ולנתח את הנתונים שלנו. על ידי ארגון הנתונים שלנו בצורה זו, נוכל לבצע בקלות שאילתות כדי לנתח את הקשרים בין הממדים והמדדים. לדוגמה, נוכל לנתח בקלות כיצד התמחור משפיע על דירוגי משתמשים עבור תעשייה או קטגוריה ספציפית, או כיצד מספר המשתמשים משתנה לפי קנה המידה העסקי וסוג המשתמש.
- ב. להלן פירוט של האופן שבו ניתן לארגן את טבלת הנתונים לפי סכמת כובבים:

#### 1. Fact Table:

- OS key
- Industry key
- Product id (measure)
- Mobile apps (measure)
- Rating (measure)

## 2. Dimension Tables:

- OS Dimension:
  - OS Key
  - Operating System Name
- Industry Dimension:
  - Industry Key
  - Industry Category

ג. דוגמא ל use case בשאלת המחקר שלנו – "מדידת נתח השוק של מערכת Mac בכל תעשייה- בבצע בדיקה על ידי סכימה של השימוש במערכת Mac ובכך נוכל להסיק על נתח השוק של המערכת ואף להציע תוכנית לייעול על מנת שהמערכת תהיה בעלת שוק נרחב."

על מנת שנוכל לבצע ניתוח כזה על נתח השוק של המערכת, נצטרך לעשות איחוד בין טבלאות שונות (OS - Rating). בעזרת שימוש ב"Over" נבצע הפחתת שימוש בJoin מאחר והפונקציה מאפשרת שליפה מהירה של מידע וקבלת פלט, חוסך זיכרון (שמירת המידע המקורי ללא שינוי) ומקצר את זמן ההמתנה של המשתמש. דבר זה יכול לעזור לחברת התוכנה לקבל החלטות מושכלות לגבי פיתוח מוצר, אסטרטגיות תמחור ומיקוד לקוחות, מה שמוביל לתוצאות עסקיות משופרות.

ד. מקרה שימוש נוסף שבו הגישה שלנו יעילה ומסייעת למשתמש לקבל החלטות הוא בזיהוי השפעת התמחור על הדירוג של מערכת Mac. על ידי איחוד הנתונים מטבלאות "דירוג" ו"תמחור", נוכל לנתח את הקשר בין התמחור של מערכת Mac והדירוג המקביל לה. ניתוח זה יכול לספק תובנות לגבי אסטרטגיות התמחור שמניבות דירוגים גבוהים יותר ושביעות רצון הלקוחות. על ידי שימוש בפונקציות צבירה כגון "group by" ו"sum", נוכל לחשב את הדירוג הממוצע עבור טווחי מחירים שונים ולזהות כל מתאם בין תמחור לשביעות רצון הלקוחות. מידע זה יכול לסייע לחברת התוכנה בקביעת אסטרטגיות תמחור אופטימליות המתאימות להעדפות הלקוח ולמקסם את קבלת המערכת בשוק. גישת זו מסייעת להפחית את הצורך בהצטרפות מרובות, מאפשרת שליפה מהירה יותר של מידע ומאפשרת לחברה לקבל החלטות מונעות נתונים באופן מיידי.

## 3. תהליך ETL

1. תהליך ETL (Extract, Transform, Load) משמש לאיסוף נתונים ממקורות שונים, להפוך אותם לפורמט רצוי ולטעון אותם למחסן נתונים או למערכת יעד אחרת. הנה דוגמה לתהליך ETL לאיסוף הנתונים עם עמודות OS, Industry, Product id, Mobile apps, Rating:

### הוצאה:

השלב הראשון בתהליך ה-ETL הוא לחלץ את הנתונים מהמקור שלו. במקרה זה, הנתונים עשויים להיות מאוחסנים במקורות שונים, כגון גיליונות אלקטרוניים, מסדי

נתונים או קבצי CSV. הנתונים נשלפים ומועברים לאזור היערכות בו ניתן לנקות ולהפוך אותם.

#### טרנספורמציה:

בשלב הטרנספורמציה הנתונים עוברים ניקוי, סטנדרטיזציה והפיכה לפורמט המתאים לניתוח.

ניתן לבצע את ההמרות הבאות על הנתונים:

- סטנדרטיזציה של מזהה המוצר על ידי הסרת תווים או רווחים מיוחדים.
- סיווג התעשייה בהתבסס על קטגוריה.
- סטנדרטיזציה של סוג הפריסה ל-on-premise, ענן או היברידי.
- סיווג התמחור בהתבסס על קטגוריה.
- סטנדרטיזציה של סולם הדירוג מ-1-3 ו מ-3.5-5.

#### טוען:

ברגע שהנתונים עוברים טרנספורמציה, הם נטענים למערכת היעד, כגון מחסן נתונים, לצורך ניתוח.

יש לבצע אופטימיזציה של מערכת היעד לסוגי השאילתות והניתוח שיבוצעו על הנתונים. במהלך שלב הטעינה, ניתן לנקות, לסנן או לצבור את הנתונים כדי לייעל את השימושיות שלהם. לדוגמה, ניתן להסיר רשומות כפולות, לזקוף ערכים חסרים או לצבור נתונים ברמה גבוהה יותר, כגון חודשי או רבעוני.

לסיכום, תהליך ה-ETL לאיסוף הנתונים עם לאיסוף הנתונים עם עמודות OS, Industry, Product id, Mobile apps, Rating כרוך בחילוץ הנתונים מהמקור שלהם, הפיכתם לפורמט סטנדרטי המתאים עבור ניתוח, וטעינתו למערכת יעד המותאמת לשאילתות וניתוח.

## 4. ניתוח ה-DATA WAREHOUSE

### 1. In which industry is the Mac operating system rated higher?

```
SELECT industry, AVG(rating) AS average_rating
```

```
FROM (
```

```
SELECT industry, rating, ROW_NUMBER() OVER (PARTITION BY industry ORDER BY  
rating DESC) AS rn
```

```
FROM FactTable
```

```
WHERE system = 'Mac'
```

```
) AS ranked
```

```
WHERE rn = 1
```

GROUP BY industry;

**2. What are the average ratings for operating systems with mobile apps?**

SELECT AVG(rating) AS AverageRating

FROM FactTable

WHERE MobileApps = 'Y';

**3. Is an operating system that has a mobile app rated higher on average than operating systems without a mobile app, within each industry?**

SELECT Industry, MobileApps, AVG(Rating) AS AverageRating

FROM FactTable

GROUP BY Industry, MobileApps

HAVING COUNT(\*) > 1

ORDER BY Industry, MobileApps;

**4. What are the top 3 industries with the highest average rating for each operating system?**

SELECT OS, Industry, AverageRating

FROM (

SELECT OS, Industry, AVG(rating) AS AverageRating, RANK() OVER (PARTITION BY OS ORDER BY AVG(rating) DESC) AS Ranking

FROM FactTable

GROUP BY OS, Industry

) AS Subquery

WHERE Ranking <= 3

ORDER BY OS, Ranking;

**5. Is there a difference in the average rating of OS between industries with a high number of Mac users and industries with a low number of Mac users?**

SELECT industry,

NTILE(4) OVER (PARTITION BY industry ORDER BY rating DESC) AS rating\_bucket,



AVG(rating) AS avg\_rating

FROM FactTable

WHERE OS = 'Mac'

GROUP BY industry

ORDER BY avg\_rating DESC;

**6. Which Operating system has the highest rating within each industry?**

SELECT industry, OS, rating

FROM (

SELECT industry, OS, rating, ROW\_NUMBER() OVER (PARTITION BY industry ORDER  
BY rating DESC) AS rn

FROM FactTable

) AS ranked

WHERE rn = 1;

**7. Is there a correlation between the number of users and the average rating of mobile apps across different categories?**

SELECT category, AVG(rating) AS average\_rating, SUM(no\_of\_users) AS  
total\_users

FROM FactTable

WHERE mobile\_apps = 'Y'

GROUP BY category

ORDER BY total\_users DESC;

**8. How does the pricing of products vary across different user types and industries?**

SELECT industry, user\_type, AVG(pricing) AS average\_pricing

FROM FactTable

GROUP BY industry, user\_type

ORDER BY average\_pricing DESC;

**9. Are there any significant differences in ratings between different business scales for a specific operating system?**

```
SELECT OS, business_scale, AVG(rating) AS average_rating  
FROM FactTable  
WHERE OS = 'Mac'  
GROUP BY OS, business_scale  
ORDER BY average_rating DESC;
```

**10. How does the average rating of Mac systems compare between industries with and without mobile apps?**

```
SELECT industry, AVG(rating) AS average_rating  
FROM FactTable  
WHERE OS = 'Mac'  
GROUP BY industry  
ORDER BY average_rating DESC;
```

## 5. מסקנות

1. למערכות הפעלה עם אפליקציות MOBILE יש דירוג ממוצע, המעיד על הדירוג הכולל של מערכות הפעלה אלו. מסקנה זו מספקת תובנות לגבי הדירוגים הממוצעים במיוחד עבור מערכות הפעלה המציעות יישומים ניידים.
2. מערכות הפעלה עם אפליקציה סלולרית, בממוצע, זוכות לדירוג גבוה יותר בהשוואה למערכות הפעלה ללא אפליקציה מובייל, בתוך כל ענף. מסקנה זו מבוססת על השוואת דירוגים ממוצעים בין סוגים שונים של מערכות הפעלה (עם ובלי אפליקציות סלולריות) בכל ענף.
3. ייתכן הבדל בדירוג הממוצע של מערכות הפעלה בין תעשיות עם מספר גבוה של משתמשי Mac לבין תעשיות עם מספר נמוך של משתמשי Mac. מסקנה זו מאפשרת השוואה של דירוגים ממוצעים בין שני סוגי התעשיות הללו, מה שיכול להצביע על שינויים בדירוגים המבוססים על שכיחות משתמשי Mac.
4. ייתכן שיש מתאם בין מספר המשתמשים והדירוג הממוצע של אפליקציות לנייד בקטגוריות שונות. מסקנה זו מציעה שיכול להיות קשר בין מספר המשתמשים לבין הדירוג הממוצע של אפליקציות סלולריות בקטגוריות שונות.



5. התמחור של המוצרים משתנה בין סוגי משתמשים ותעשיות שונות. מסקנה זו מצביעה על כך שיש הבדלים בתמחור המבוססים על סוגי משתמשים ותעשיות, כפי שמעידים ערכי התמחור הממוצעים הקשורים לכל שילוב.