

פרויקט בינה ואנליטיקה עסקית



מרצה: אדיר אבן

Team Members
גיל ביטון – gilbito@post.bgu.ac.il
תום דמארי – damarit@post.bgu.ac.il
שירה יעקב – shiraja@post.bgu.ac.il

תוכן עניינים

3.....	הארגון
6.....	ניתוח מקור המידע
8.....	Data Preparation
9.....	מדדי ביצוע מרכזיים - Key Performance Indicators
11.....	פלטפורמות BI&A מסורתיות
15.....	יישומי BI&A וניתוח נתונים מתקדמים
18.....	נספחים

הארגון

תיאור קצר של הארגון:

NIKE הינו תאגיד ספורט עולמי המוביל בעיצוב, פיתוח, ייצור ושיווק של נעלי ספורט, ביגוד וציוד ספורט. החברה נוסדה בשנת 1964 בארצות הברית וכיום מוכרת את מוצריה ברחבי העולם באמצעות רשת חנויות פיזיות וחנויות מקוונות, ומשרתת קהל רחב הכולל ספורטאים מקצועיים, חובבי ספורט וחובבי אופנה. היא פועלת בתעשיית ההנעלה, הביגוד וציוד ספורט.

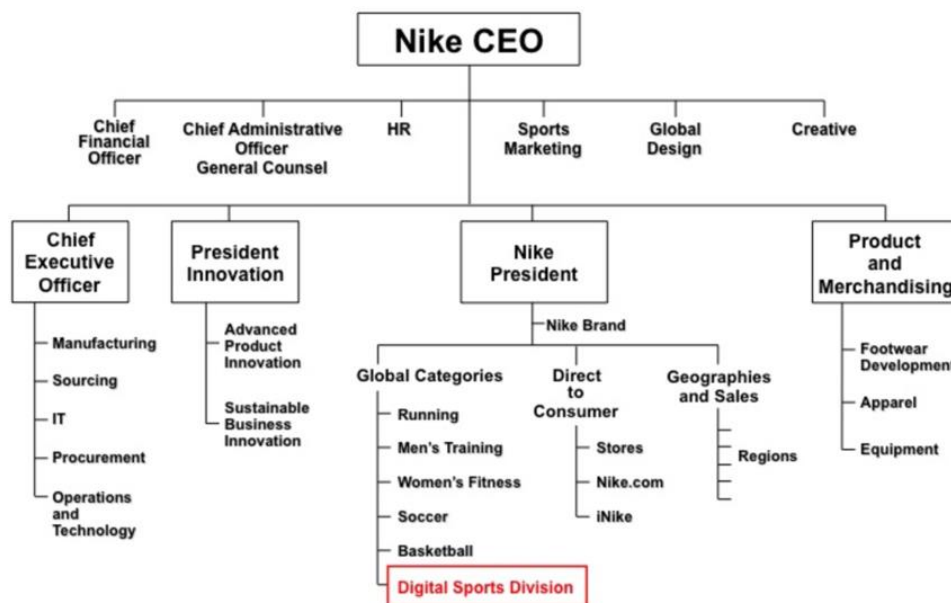
אפיון הסביבה התחרותית:

מודל 5 הכוחות של פורטר יכול לסייע בצורה משמעותית בהבנת הסביבה התחרותית של נייק. המודל מזהה חמישה כוחות עיקריים: עוצמת התחרות בתעשייה, איום כניסת מתחרים חדשים, איום של מוצרים תחליפיים, כוח המיקוח של הלקוחות וכוח המיקוח של הספקים. באמצעות ניתוח מעמיק של הכוחות הללו, ניתן להעריך את הסביבה התחרותית ולגבש אסטרטגיות בהתאם.

- **מתחרים (Competitors)** – התחרות בתעשיית הספורט היא אינטנסיבית ורבת משתתפים. ישנם שני סוגי מתחרים – מתחרים ישירים ועקיפים. מתחרים ישירים: מותגי ספורט ידועים כמו Adidas, New Balance, ASICS, Under Armour, אשר מציעים מוצרים דומים ומתחרים באופן ישיר בנייק על נתח השוק בתחום ההנעלה, הביגוד והציוד הספורטיבי. מתחרים עקיפים: מותגי אופנה ופנאי כמו Lululemon, Skechers וחנויות הנעלה המציעות מוצרים שאינם בהכרח ספורטיביים אך עדיין פונים לקהל הלקוחות של החברה. מתחרים אלו מתחרים בנייק מבחינת סגנון, נוחות ועיצוב, ולעיתים גם במחיר.
- **כניסות חדשות (New Entrants)** – תעשיית ההנעלה והביגוד היא שוק דינמי ופתוח יחסית לכניסת שחקנים חדשים. חברות חדשות עם עיצובים וטכנולוגיות חדשניות מופיעות באופן קבוע, מאיימות לקחת נתח שוק מהמומחים הוותיקים. נייק נדרשת להשקיע רבות במחקר ופיתוח כדי לשמור על יתרונה התחרותי ולספק מוצרים חדשניים ואיכותיים שמתחרים בהצלחה במוצרים של החברות החדשות.
- **לקוחות (Customers)** – ההעדפות והצרכים של הלקוחות משתנים במהירות, במיוחד בתחומים הקשורים לאופנה ולספורט. לקוחות מצפים למוצרים שמתעדכנים בהתאם לטרנדים החדשים, לשימוש בטכנולוגיות מתקדמות ולשיפור מתמיד בעמידות ובאיכות. בנוסף, קיימת דרישה גוברת למוצרים ידידותיים לסביבה ולתהליכי ייצור ברי קיימא. נייק צריכה להסתגל לשינויים אלו ולהתאים את מוצריה והאסטרטגיות שלה כדי לעמוד בצרכי הלקוחות.
- **מוצרים תחליפיים (Substitutes)** – בענף ההנעלה והביגוד ישנם מוצרים תחליפיים רבים שיכולים להשפיע על הביקוש למוצרי נייק. מוצרים תחליפיים כוללים מותגים זולים יותר או מוצרים שמציעים תכונות ייחודיות כמו נעליים אורתופדיות או פריטים מבד טכנולוגי מתקדם. על נייק לוודא שמוצריה מציעים ערך מוסף על פני התחליפים, בין אם באמצעות טכנולוגיה מתקדמת, עיצוב ייחודי או איכות גבוהה.
- **ספקים (Suppliers)** – כוח המיקוח של הספקים יכול להשפיע משמעותית על העלויות והאיכות של מוצרי נייק. ספקים המספקים חומרים או רכיבים ייחודיים יכולים להכתיב מחירים גבוהים או להקטין את זמינות המוצרים. נייק צריכה לשמור על יחסים טובים עם הספקים, למצוא ספקים חלופיים ולהשקיע בפיתוח תהליכים פנימיים כדי להקטין את התלות בספקים מסוימים ולהבטיח אספקה רציפה ואיכותית.

נוכל ללמוד על היתרון התחרותי של נייק בעזרת מודל האסטרטגיות של פורטר. המודל מתאר שלוש אסטרטגיות עיקריות שארגונים יכולים להשתמש בהן כדי להשיג יתרון תחרותי בשוק: Cost Leadership (מנהיגות עלות), Differentiation (בידול), ו-Focus (מיקוד בפלח שוק). נייק משתייכת לקטגוריית יתרון הבידול במודל, בכך שהיא פונה לקהל מאוד רחב של לקוחות ושמה דגש רב על מחקר ופיתוח (R&D) ליצירת מוצרים חדשניים בעלי טכנולוגיות מתקדמות, כגון טכנולוגיות Air Max ו-Flyknit, בכדי לספק מוצרים עם ביצועים גבוהים ואיכות מעולה. החברה מתמקדת במיתוג חזק ובקמפיינים שיווקיים חכמים עם ספורטאים מובילים, המדגישים את ערכי ההישגיות וההשראה. בנוסף, נייק מקפידה על עיצוב ייחודי ואסתטי של מוצריה, תוך מתן דגש לחוויית משתמש ושירות לקוחות מצוינים. כמו כן, היא גם מחויבת לקיימות ואחריות חברתית, מה שתורם לתדמית חיובית ולבידול מול מתחרים. כל אלו מאפשרים לנייק לשמור על יתרון תחרותי בשוק ולספק ערך מוסף ללקוחותיה.

מבנה ארגוני:



Galbraith, J. R. (2014). Organizational design challenges resulting from big data. *Journal of Organization Design*, 3(1), 2-13. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2458899

בראש ההיררכיה הארגונית של תאגיד נייק עומד המנכ"ל (CEO), שתחתיו פועלות יחידות אסטרטגיות מרכזיות כמו מחלקת הכספים, המחלקה המשפטית והאדמיניסטרטיבית, משאבי אנוש, שיווק ספורט, עיצוב גלובלי וקריאייטיב. כל אחת מהיחידות הללו אחראית על תחום חשוב בניהול ותפעול החברה, החל מבקרה פיננסית ועד לעיצוב ופיתוח מוצרים חדשניים. תחת המנכ"ל פועלים מנכ"ל מבצעי (Chief Executive Officer), נשיא החדשנות (President of Innovation) ונשיא נייק (President of Nike). המנכ"ל המבצעי (COO) אחראי על תחומים כגון ייצור, רכש, IT, תפעול וטכנולוגיה. נשיא החדשנות (President of Innovation) מתמקד בפיתוח מוצרים מתקדמים וחדשנות עסקית בת-קיימא. הנשיא של נייק (Nike President) מפקח על קטגוריות גלובליות, כולל ריצה, אימון גברים, כושר נשים, כדורגל וכדורסל, וכן על תחום הקשר הישיר עם הצרכנים, הכולל את החנויות הפיזיות והפלטפורמות הדיגיטליות. בנוסף, מחלקת מוצר ומסחר אחראית על פיתוח הנעלה, ביגוד וציוד, ובכך משלימה את המעטפת המלאה של פעילות התאגיד.

במסגרת הפרויקט, נתמקד ביחידות הארגוניות הבאות: מכירות (Sales), מוצר ופיתוח (Product and Development), משאבי אנוש (Human Resources), ייצור ותפעול (Manufacturing and Operations), כספים (Finance) וטכנולוגיית מידע (IT).

תהליכים עסקיים:

מכירות והפצה: במסגרת ניתוח זה, נבחן את ההבדלים בין מכירות דרך ערוצים מקוונים למכירות בחנויות פיזיות. נתמקד בפילוח הגיאוגרפי של המכירות ונבחן את פילוח הלקוחות, למשל לפי גיל ומגדר, כדי למקד מאמצי שיווק בצורה יעילה יותר. בנוסף, נבחן את שיעור המכירות האינטרנטיות כדי להעריך את הצלחת החברה במכירות דיגיטליות, לזהות מגמות בשוק ולהבין את השפעת השיווק המקוון על ההכנסות הכוללות. תהליכי המכירות וההפצה משפיעים ישירות על ההכנסות והגידול בשוק. יכולת ניהול אפקטיבית של מכירות מקוונות ומכירות בחנויות פיזיות חיונית להבטחת נתח שוק גבוה ושביעות רצון הלקוחות.

אספקת הזמנות: בתחום זה נבחן את ניהול ההזמנות והלוגיסטיקה על ידי מעקב אחר זמני האספקה וזיהוי אחוז ההזמנות שהגיעו באיחור. הניתוח יכלול גם את זיהוי הבעיות בשרשרת האספקה ותכנון אסטרטגיות לשיפור השירות והדיוק באספקת ההזמנות. מטרת הניתוח היא לשפר את הדיוק והמהירות באספקה בהתאם לנתונים המתקבלים. יעילות תהליכי האספקה והלוגיסטיקה משפיעה על נאמנות הלקוחות ושביעות רצונם. זיהוי בעיות בזמן האספקה מאפשר שיפור מתמיד והבטחת איכות השירות, מה שמגביר את שביעות רצון הלקוחות ומחזק את נאמנותם.

פיתוח מוצרים והתאמתם לשוק: נתמקד בהרכב וערך ההזמנות כדי להבין את התרומה היחסית של כל מוצר לסך ההזמנה, לזהות מוצרים מרכזיים בהזמנות ולשפר את אסטרטגיות המכירה. בנוסף, נבחן את תהליכי המחקר והפיתוח (R&D) לשיפור ביצועים ועמידות המוצרים, תוך עמידה בדרישות לקוחות משתנות ופיתוח מוצרים חדשניים. תהליכי הפיתוח והחדשנות משפיעים על היכולת של נייק להציע מוצרים חדשים ואיכותיים המותאמים לצרכי הלקוחות המשתנים ולשמור על יתרון תחרותי בשוק.

ביצועי הארגון נשענים בצורה משמעותית על תהליכים אלו, שכן הם מייצגים את מידת ההצלחה של החברה בצורה ישירה. עם זאת, ישנם תהליכים נוספים משמעותיים שלא נכללו במידול זה, כגון השפעת פרסומות ושיתופי פעולה, אשר גם להם יש חשיבות בהערכת הביצועים הארגוניים.

החלטות מפתח לתהליכים העסקיים:

החלטות ניהוליות ותפעוליות מרכזיות הקשורות לתהליכים העסקיים כוללות, בין היתר, תכנון אסטרטגיות שיווק, ניהול מלאי ושיפור שירות לקוחות עבור תחום המכירות וההפצה. בנוסף, ישנן החלטות הנוגעות לתכנון וניהול שרשרת האספקה, בחירת ספקים ומעקב אחר מדדי ביצוע (KPIs) בתחום אספקת ההזמנות. בתחום פיתוח המוצרים, יש להקנות עדיפות להשקעות במחקר ופיתוח, לתכנן לוחות זמנים ולהבטיח את איכות המוצרים. גורמים שעשויים להשפיע על החלטות אלו כוללים מגמות שוק ותחרות, התקדמות טכנולוגית, שינויים ברגולציה וחקיקה, משוב לקוחות, עלויות ייצור, ביצועי מכירות ושינויים גלובליים. לאור זאת, על המנהלים להתאים באופן מתמיד את האסטרטגיות והתהליכים שלהם לשינויים בסביבה העסקית.

ניתוח מקור המידע

אנו יצרנו את מקור הנתונים בו השתמשנו בפרויקט זה על ידי חילול הנתונים באמצעות אקסל על בסיס התפלגויות נתונים שפורסמו במקורות שונים אודות תאגיד נייק. למשל, התפלגות המכירות של נייק לפי יבשת, התפלגות גילאים ומגדר של לקוחות ועובדים ועוד. הנתונים נוצר במטרה לשקף בצורה מדויקת ככל הניתן את ההתפלגויות והטרנדים כפי שנצפו בעולם האמיתי, תוך שימוש בנוסחאות וכלים מתקדמים נוספים באקסל כפי שנלמדו בסדנה. תהליך זה תורם ליצירת מקור נתונים המייצג את ההיבטים המרכזיים של סביבת העסקים של נייק בצורה מהימנה. הנתונים משקפים תקופה שבין שנת 2006 ועד מאי 2024, ומספקים תמונה מקיפה של מגמות ודפוסים לאורך כמעט שני עשורים. פרק זמן נרחב זה מאפשר לבצע ניתוח מעמיק ולהסיק מסקנות על התפתחויות ושינויים ארוכי טווח בשוק, בהתנהגות הלקוחות ובביצועי המוצרים.

Table	Description	Attribute	Data Type	Scale	Description
Customers	List of Customers	<u>Email</u>	Varchar(30)	Nominal	Customer ID
		Password	Varchar(20)	Nominal	Password for online site
		Customer Name	Varchar(50)	Nominal	Full name of Customer
		Customer Gender	Varchar(10)	Nominal	Customer gender
		Customer Birthdate	Date	Interval	Customer birthdate
		Address-Continent	Varchar(20)	Nominal	Customer Address
		Address-Country	Varchar(20)	Nominal	
		Address-City	Varchar(20)	Nominal	
		Address-Street	Varchar(50)	Nominal	
		Has Membership	Integer	Nominal	0 - Not a member, 1 - Is a member
Products	List of Products	<u>Product ID</u>	Integer	Nominal	Product ID
		Product Name	Varchar(50)	Nominal	Name of Product
		Product Category	Varchar(20)	Nominal	Product Category (Shoe, Apparel, Sport Equipment, or Accessories)
Versions	Versions of products	<u>Product ID (Products)</u>	Integer	Nominal	Product ID
		<u>Size</u>	Real	Ordinal	Product Size (e.g., 8.5)
		<u>Color</u>	Varchar(20)	Nominal	Product Color
		Version Price	Money	Ratio	Price per version unit
Employees	List of Employees	<u>Employee ID</u>	Integer	Nominal	Employee ID
		Employee Name	Varchar(50)	Nominal	Full name of Employee
		Employee Gender	Varchar(10)	Nominal	Employee Gender
		Employee Birthdate	Date	Interval	Employee Birthdate
		Education	Varchar(20)	Ordinal	Academic education
		Role	Varchar(20)	Nominal	Role within NIKE

		Employee Continent	Varchar(20)	Nominal	Continent of Employment
		Employee Country	Varchar(20)	Nominal	Country of Employment
		Annual Salary	Money	Ratio	Employee's Annual Salary
Interests	Version's interests saved by Customers	<u>Customer Email (Customers)</u>	Varchar(30)	Nominal	Customer Identifier
		<u>Product ID (Versions)</u>	Integer	Nominal	Version's Product ID
		<u>Size (Versions)</u>	Real	Ordinal	Size of the Product Version
		<u>Color (Versions)</u>	Varchar(20)	Nominal	Color of the Product Version
		Interest Date	DateTime	Interval	Date when the Interest was recorded
Orders	Product Orders made by Customers	<u>Order ID</u>	Integer	Nominal	Order Number
		Order Date	DateTime	Interval	Date when the Order was placed
		Order Arrival Date	DateTime	Interval	Date when the Order arrived
		Order Type	Varchar(10)	Nominal	1 - In Store, 2 – Online
		<u>Customer Email (Customers)</u>	Varchar(30)	Nominal	Customer Identifier
		<u>Employee ID (Employees)</u>	Integer	Nominal	Employee Identifier
		Order Discount	Decimal	Ratio	Order discount percentage (e.g., 0.05)
Includes	Products in each Order	<u>Order ID (Orders)</u>	Integer	Nominal	Order Identifier
		<u>Product ID (Versions)</u>	Integer	Nominal	Version's Product ID
		<u>Size (Versions)</u>	Real	Ordinal	Size of purchased Version
		<u>Color (Versions)</u>	Varchar(20)	Nominal	Color of purchased Version
		Quantity	Integer	Ratio	Number of Units Sold

Data Preparation

בחרנו להשתמש במפתח עזר עבור כל טבלה שבה המפתח הראשי מורכב יותר משדה אחד.
 בנוסף, על מנת להימנע ממבנה פתית השלג בין VERSIONS לבין PRODUCTS בחרנו לבצע שיטות
 ולהוסיף את השדות של PRODUCTS לתוך טבלת VERSIONS. בהמשך בחרנו להכליל טבלאות
 מגרעיניות גבוה ולכן החלטנו לבצע השטחה של טבלת VERSIONS על טבלאות העובדה
 INTERESTS ו INCLUDES.

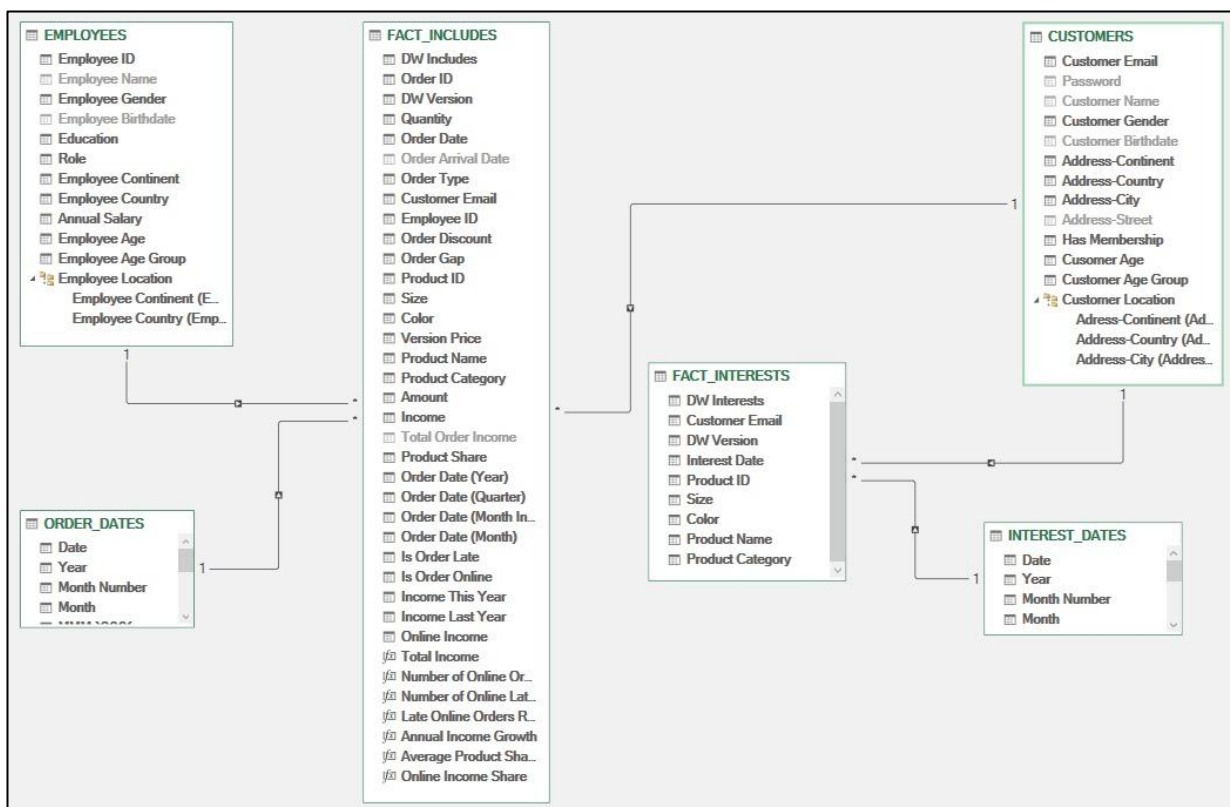
טבלאות מימד:

- **CUSTOMERS** (Customer Email, Password, Customer Name, Customer Gender, Customer Birthdate, Address-Country, Address-City, Has Membership, Customer Age, Customer Age Group)
- **EMPLOYEES** (Employee ID, Employee Name, Employee Gender, Employee Birthdate, Education, Role, Employee Country, Annual Salary, Employee Age, Employee Age Group)

טבלאות עובדה:

- **FACT - INTERESTS** (DW Interest, {Customer Email} (CUSTOMERS), {DW Version, Product ID, Size, Color} (VERSIONS), Interest Date)
- **FACT - INCLUDES** (Order ID (ORDERS), {Product ID, Size, Color} (VERSIONS), Quantity)

סכמת כוכב:



מדדי ביצוע מרכזיים - Key Performance Indicators

מדדי ביצוע מרכזיים (KPIs) הם מדדים חיוניים המשמשים להערכת הצלחה וביצועים של הארגון בהתקדמות לעבר המטרות והשגת יעדים אסטרטגיים. המדדים בהם בחרנו לניתוח הנתונים:

1. צמיחה שנתית בהכנסות

שיעור השינוי בהכנסות משנה קודמת לשנה הנוכחית -

$$\text{Annual Income Growth} = \text{SUM}(\text{Income This Year}) / \text{SUM}(\text{Income Last Year}) - 1$$

לצורך החישוב יצרנו שני שדות מחושבים Income This Year וגם Income Last Year בטבלת FACT_INCLUDES אשר מציגים את ההכנסות השנתיות של החברה מההזמנות בשנה הנוכחית ובשנה שעברה בהתאמה, תוך שימוש בהעמסה של מספר הרשומות המתועדות באותה שנה.

מדד הצמיחה השנתית בהכנסות הוא קריטי להערכת היעילות של האסטרטגיות העסקיות של החברה לאורך זמן ורווחיות החברה. מדד זה משקף את היכולת של החברה לייצר הכנסות נוספות באמצעות שיפור במכירות, הרחבת קווי מוצרים, כניסה לשווקים חדשים והגדלת נתח השוק. יתרה מכך, המדד מסייע לזיהוי מגמות וצמיחה פוטנציאלית, ולתכנון עתידי של יעדים עסקיים. מדד זה שייך לפרספקטיבת ה-BSC הפיננסית, שכן הוא מתמקד בהערכת הביצועים הכספיים של החברה ומשקף את הצלחתה הכלכלית. המדד משקף הערכת אפקטיביות, כיוון שהוא מציג נקודת מבט חיצונית על הצלחת החברה בהשגת מטרותיה העסקיות והפיננסיות (ראה נספח 1.1). דוגמא לשאלה עסקית: בכמה הכנסות החברה גדלו השנה ביחס לשנה שעברה?

2. נתח ההכנסות ממכירות מקוונות

אחוז ההכנסות ממכירות שמתבצעות דרך ערוצים מקוונים ביחס לסך ההכנסות הכולל השנה -

$$\text{Online Income Share} = \text{SUM}(\text{Online Income}) / \text{Total Income}$$

יצרנו משתנה מחושב Online Income בטבלת INCLUDES_FACT אשר מציג את ההכנסות מביצוע הזמנות מקוונות בלבד:

$$\text{Online Income} = \text{IF}(\text{Order Type} = \text{"Online"} \text{ THEN } 1 \text{ ELSE } 0) * \text{Income}$$

בנוסף, הגדרנו מטריקה Total Income אשר סוכמת את ההכנסה מכל פריט בהזמנה:

$$\text{Total Income} = \text{SUM}(\text{Income})$$

מדד שיעור ההכנסות ממכירות המקוונות מתמקד בניתוח הצלחת החברה במכירות דרך הערוצים המקוונים. המדד מסייע להבין את ההצלחה האסטרטגית של החברה במגזר הדיגיטלי וכיצד היא מצליחה לעמוד בצרכים ובציפיות הלקוחות בתחום זה. כך ניתן לזהות מגמות בצרכנות הדיגיטלית, להתאים את הממשק ללקוחות ולהשקיע יותר במאמצי שיווק מקוונים, ובכך לשפר את חוויית הלקוח הדיגיטלית. המדד מהווה חלק מפרספקטיבת ה-BSC של הלקוחות, שכן הוא מתמקד בניתוח חוויית הלקוח והצלחת החברה במכירות דרך ערוצים מקוונים. המדד משקף הערכת אפקטיביות מכיוון שהוא מודד את יכולת החברה להגיע ללקוחות דרך הערוצים המקוונים ולהצליח למכור בערוצים אלו (ראה נספח 1.2).

דוגמא לשאלה עסקית: בכמה אחוזים גדלו המכירות המקוונות בשנה האחרונה?

3. אחוז ההזמנות שהגיעו באיחור

היחס בין מספר ההזמנות שהגיעו באיחור (מעל 10 ימים) השנה לבין מספר ההזמנות המקוונות הכולל השנה (הזמנות שאינן מקוונות מסופקות בעת הרכישה) -

$$\text{Late Online Orders Rate} = \text{Number of Online Late Orders} / \text{Number of Online Orders}$$

לצורך חישוב המדד יצרנו שדה מחושב Is Order Late בטבלת FACT_INCLUDES אשר בודקת האם הפריט בהזמנה הגיע אחרי 10 ימים ומעלה: $\text{Is Order Late} = \text{IF}(\text{Order Gap} > 10 \text{ THEN } 1 \text{ ELSE } 0)$, ולאחר מכן ביצענו העמסה לפי מספר הפריטים בהזמנה. כך ניתן בעזרת סכימה פשוטה לחשב את שתי המטריקות $\text{Number of Online Late Orders}$ ו- $\text{Number of Online Orders}$:

$$\text{Number of Online Orders} = \text{SUM}(\text{Is Order Online})$$

$$\text{Number of Online Late Orders} = \text{SUM}(\text{Is Order Late})$$

מדד אחוז ההזמנות המאוחרות חיוני להערכת איכות תהליך האספקה והשירות של החברה. המדד מאפשר לזהות בעיות בשרשרת האספקה ובתהליך ההזמנה, ולנקוט צעדים לשיפור הדיוק והמהירות באספקת ההזמנות. המדד שייך לפרספקטיבת התהליכים הפנימיים של ה-BSC, שכן הוא מתמקד באיכות השירות והיעילות של תהליכי העבודה הפנימיים. בנוסף, מדד זה משקף הערכת יעילות, מכיוון שהוא מעריך את יכולת החברה לעמוד בזמני אספקה ולספק שירות איכותי ללקוחותיה (ראה נספח 1.3).

דוגמה לשאלה עסקית: מהו אחוז ההזמנות שהגיעו באיחור השנה בהשוואה לשנה שעברה?

4. נתח ממוצע של מוצר בהזמנה

ממוצע של יחס עלות המוצר מסך עלות ההזמנה הכוללת -

$$\text{Average Product Share in Order} = \text{AVERAGE}(\text{Product Share})$$

לצורך חישוב המדד יצרנו ראשית שדה מחושב Total Order Income בטבלת FACT_INCLUDES אשר מייצג את סך ההכנסה מההזמנה. לאחר מכן, השתמשנו בשדה זה כדי לחשב שדה מחושב Product Share אשר מחשב את יחס ההכנסה מרכישת מוצר בהזמנה מסוימת (שדה Income) מתוך סך ההכנסה מאותה הזמנה (שדה מחושב Total Order Income):

$$\text{Product Share} = \text{Income} / \text{Total Order Income}$$

מדד זה חיוני להערכת ההרכב והערך של ההזמנות הכוללות, ומאפשר להבין את התרומה היחסית של כל מוצר לסך ההזמנה. המדד מסייע לזהות אילו מוצרים הם המרכזיים בהזמנות הלקוחות, ומשקף את יכולת החברה לתמחר מוצרים בצורה אופטימלית ולהבין את העדפות הלקוחות. תרומתו העיקרית היא בהבנה מעמיקה של התפלגות המוצרים בהזמנות, זיהוי מוצרים מובילים ושיפור אסטרטגיות המכירה והתכנון הלוגיסטי. מדד זה שייך לפרספקטיבת התהליכים הפנימיים של ה-BSC, שכן הוא מסייע בניהול יעיל של תהליכי המוצרים וההזמנות ומשקף את האיכות והאפקטיביות של תהליכי העבודה הפנימיים. המדד מעריך יעילות בכך שהוא מודד את יכולת החברה לנהל את המוצרים וההזמנות שלה בצורה מיטבית ולשפר את תהליכי העבודה הפנימיים בהתאם להעדפות הלקוחות (ראה נספח 1.4).

דוגמה לשאלה עסקית: מהם 5 המוצרים בעלי הנתח הממוצע הגבוה ביותר?

פלטפורמות BI&A מסורתיות

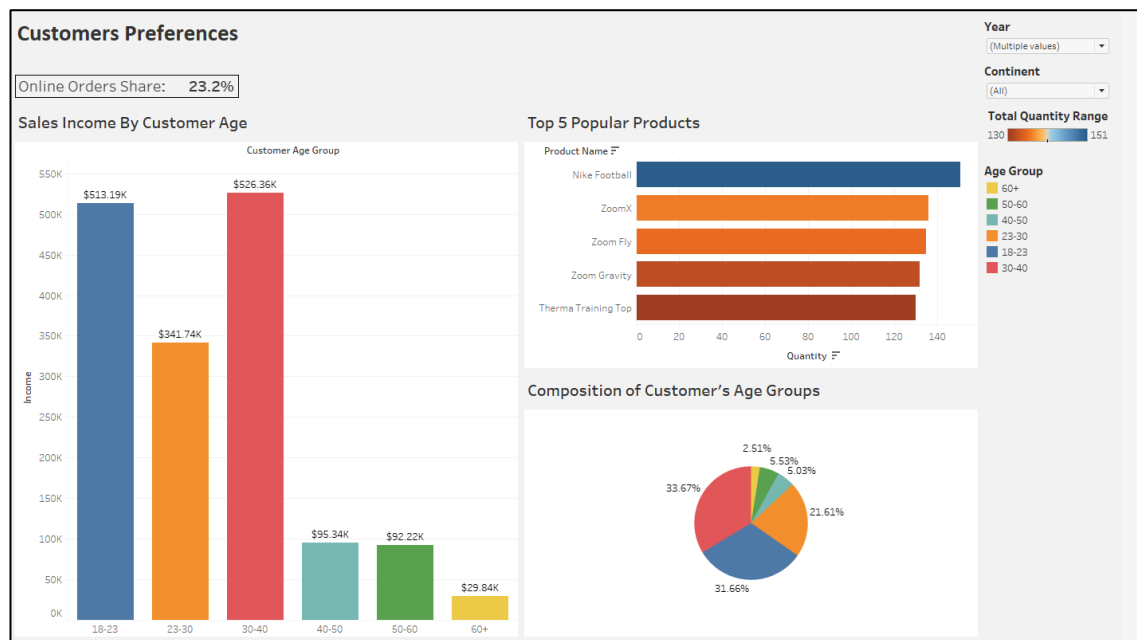
1. העדפת לקוחות:

דוח העדפת לקוחות נועד לספק תובנות מעמיקות על דפוסי הקנייה והעדפות הלקוחות. הדוח כולל תרשימים כמו אחוז ההזמנות אונליין, חמשת המוצרים הפופולריים ביותר, הכנסות ממכירות לפי קבוצות גיל, וחלוקת קבוצות הגיל של הלקוחות. באמצעות מדדי מפתח (KPI) כמו נתח ממוצע של מוצר בהזמנה (Product Share in Order) ונתח המכירות המקוונות (Online Sales Share), ניתן להבין את הרגלי הקנייה של הלקוחות וניתוח על פי קבוצות גיל.

תובנות אלו מאפשרות למנהלים בדרג הביניים לקבל החלטות ניהוליות טקטיות מבוססות נתונים לצורך תכנון לטווח הקצר, הקצאת משאבים וניטור ביצועים. הדוח מציע אפשרויות סינון לפי שנה, יבשת, מגדר וקבוצת גיל, מה שמאפשר ניתוח דינמי וממוקד של הנתונים.

השאלות העסקיות שהדוח עונה עליהן כוללות: "אילו מוצרים הם הפופולריים ביותר בקרב קבוצות גיל שונות?" ו"כיצד מתחלקות ההכנסות בין ערוצי המכירה השונים?", תובנות אלו מסייעות למנהלים להתאים את טקטיקות השיווק בהתאם לצרכים והעדפות של קבוצות הלקוחות השונות, לשפר את ניהול המלאי בהתאמה ולהתמקד בערוצים הרווחיים ביותר.

צילום דוח העדפת לקוחות



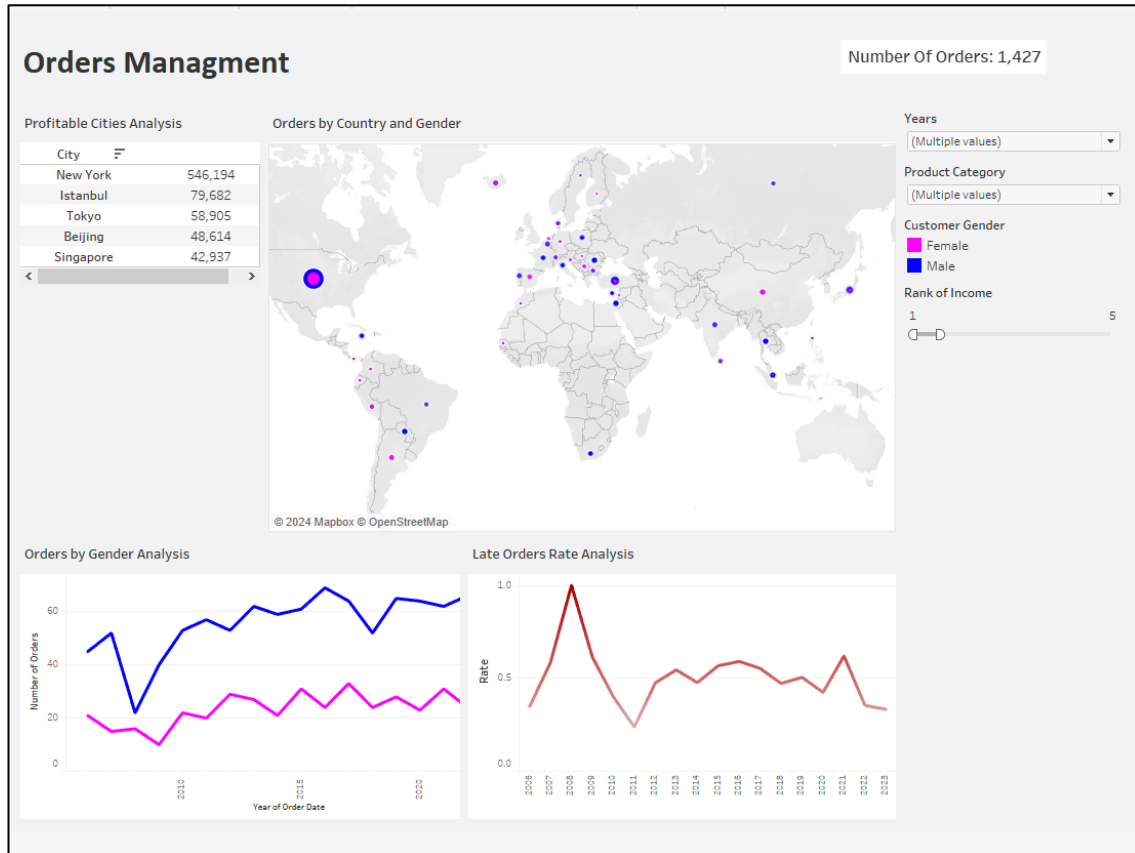
2. ניהול הזמנות

דוח ניהול הזמנות מתמקד בזיהוי מגמות וביצועים לפי אזור ומגדר, במטרה לשפר את תהליכי הלוגיסטיקה והאספקה, לזהות בעיות פוטנציאליות ולהנחות את הקצאת המשאבים האזוריים באופן יעיל. הדוח כולל טבלה המציגה את ההכנסות בערים השונות, עם אפשרות לסינון לפי דירוג כמו חמש הערים הרווחיות ביותר או עשר הערים הכי פחות רווחיות. בנוסף, מוצגת מפה אינטראקטיבית המדגימה את כמות ההזמנות בכל מדינה בפילוח לפי מגדרי (גברים ונשים), וגרף המראה את התפלגות כמות ההזמנות לאורך השנים עם פילוח מגדרי. כמו כן, ישנו גרף המציג את שיעור ההזמנות שהגיעו באיחור לאורך השנים (שימוש ב-KPI של Late Order Rate), וכן מוצגת

כמות ההזמנות הכוללת. הדוח מציע אפשרויות סינון אינטראקטיביות לפי מגדר הלקוח, קטגוריית מוצר, שנה וחודש ההזמנה, מה שמאפשר ניתוח מדויק וממוקד של הנתונים.

תובנות מהדוח תומכות בקבלת החלטות מושכלת מצד המנהלים הבכירים לשיפור תהליכי האספקה, תכנון אסטרטגי והתרחבות לשווקים פוטנציאליים ורווחיים.

צילום דוח ניהול הזמנות



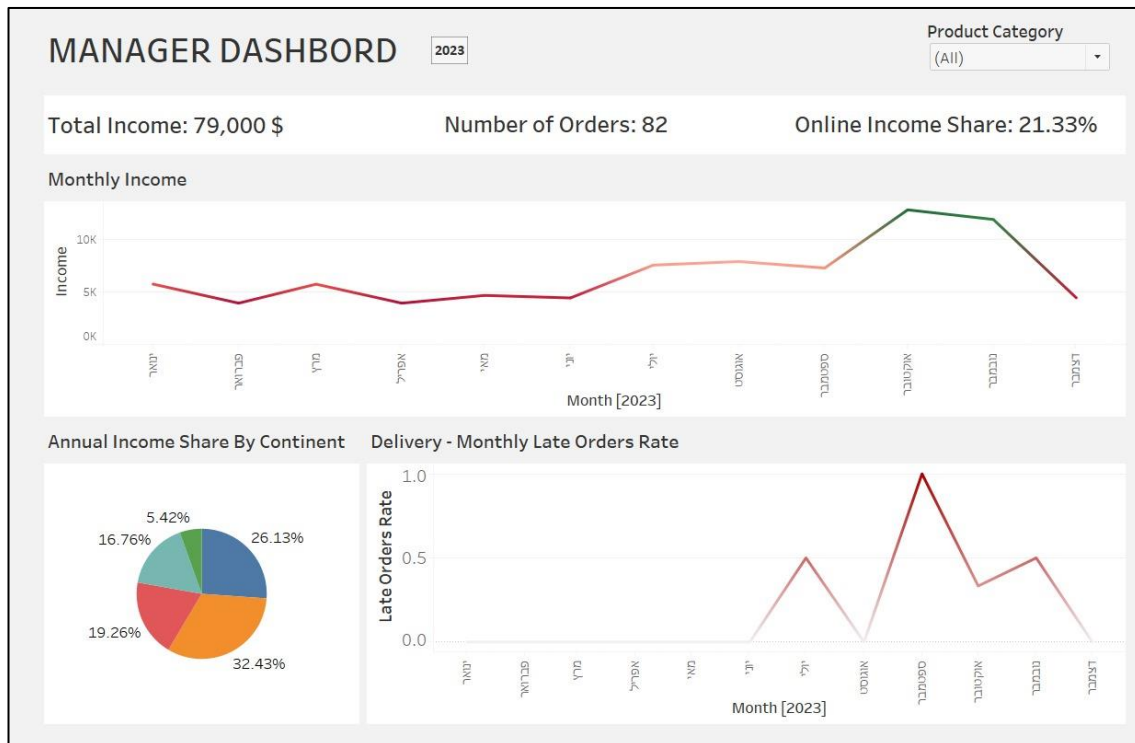
3. לוח מחוונים (DASHBOARD)

לוח המחוונים שלנו מיועד להנהלה הבכירה בארגון ומספק תמונה מקיפה ומדויקת של ביצועי הארגון בתחומים כמו מכירות, הפצה ורווחים. הדשבורד מציג נתונים עדכניים מהשנה הקודמת, עם אפשרויות סינון אינטראקטיביות (drill down) המאפשרות פילוח לפי קטגוריית מוצר, ובאמצעות לחיצה על נתונים בגרפים, כך שהלוח כולו מתעדכן אוטומטית בהתאם לבחירות המשתמש. כך, המנהלים יכולים להתמקד בניתוחים מעמיקים לפי הצורך, ולקבל החלטות מושכלות לשיפור תהליכים והתמקדות בשווקים בעלי פוטנציאל רווח גבוה.

הדשבורד כולל מגוון נתונים מרכזיים כגון סך ההכנסות, מספר ההזמנות, ונתח המכירות המקוונות, אשר יחד מספקים תמונה כוללת על ביצועי הארגון. בנוסף, מוצגים בו גרפים ותרשימים חיוניים כולל גרף של הכנסות בכל חודש, גרף של אחוז ההזמנות שהגיעו באיחור בחלוקה לחודשים (KPI): אחוז ההזמנות שהגיעו באיחור, ותרשים פאי של נתח ההכנסות של כל יבשת. גרף ההכנסות החודשי מציג את השינויים בהכנסות לאורך השנה, ומאפשר לזהות מגמות עונתיות או אירועים חריגים. גרף אחוז ההזמנות שהגיעו באיחור מציג את שיעור ההזמנות שהגיעו באיחור בכל חודש, ובכך מסייע בזהוי בעיות בתהליכי האספקה ובשיפור היעילות הלוגיסטית. תרשים פאי של נתח ההכנסות לפי יבשות מאפשר לזהות שווקים רווחיים ולנתח את פיזור ההכנסות הגלובלי.

לוח המחוננים, עם כלל המדדים והכלים האינטראקטיביים שבו, מעניק להנהלה תמונה כוללת ומפורטת על ביצועי הארגון, ומאפשר קבלת החלטות אסטרטגיות שמטרתן לשפר את ביצועי החברה ולהגדיל את הרווחיות בשווקים השונים.

צילום לוח מחוננים



4. OLAP

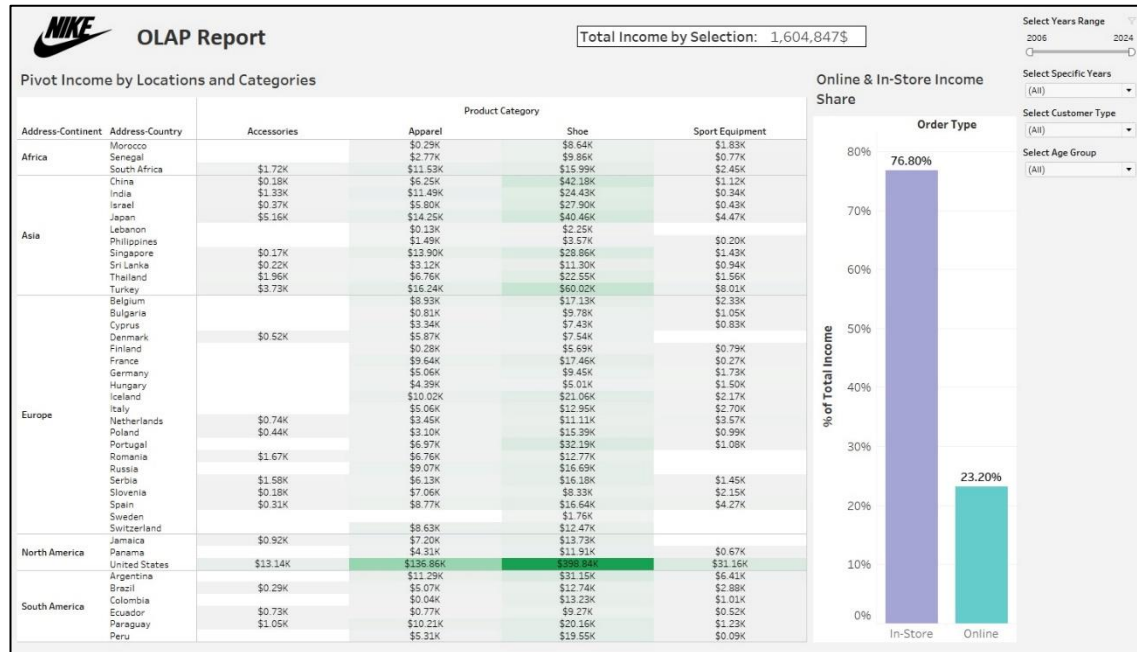
דוח ה-OLAP שפיתחנו מאפשר חקירה אינטראקטיבית וגמישה של הנתונים, תוך שימוש במגוון משתני מימד ועובדה. בטבלת ה-Pivot מוצגים סך הרווחים מכל קטגוריית מוצר בחלוקה ליבשות, עם אפשרות להעמקת הרזולוציה למדינות, ערים ולמוצרים ספציפיים (Drill Down).

בנוסף, הדוח מציג את סך ההכנסות הכולל בהתאם לבחירות שבוצעו על ידי המשתמש, לצד תרשים עמודות המשווה את נתח ההכנסות מהאינטרנט להכנסות מהחנויות הפיזיות. לחיצה על כל נתון בטבלה או בתרשימים מפעילה סינון (filter) המעדכן את כל מרכיבי הדוח באופן אוטומטי, ומאפשרת למנהלים לבצע ניתוח מעמיק ומקיף של הנתונים בצורה אינטראקטיבית ודינמית. הדוח כולל גם אפשרויות סינון מתקדמות לפי טווח שנים, בחירה של שנים ספציפיות, קבוצות גיל, וסוג הלקוח (חבר מועדון או לא). אפשרויות אלה מסייעות לנו לבצע ניתוח ממוקד ומדויק בהתאם לצרכים והפרמטרים החשובים לארגון.

כלי זה תומך בהחלטות ניהוליות קריטיות, כגון זיהוי קטגוריות מוצרים רווחיות לפי אזורים גיאוגרפיים, הבנת חלוקת ההכנסות בין מכירות בחנות למכירות מקוונות, וניתוח מגמות ארוכות טווח בנתח המוצרים הממוצע בהזמנות. באמצעות תובנות אלו, ניתן לבצע אופטימיזציה של המלאי והצעות המוצרים בהתאם לפופולריות האזורית, ולזהות הזדמנויות לשיפור תהליכי המכירה והלוגיסטיקה.

בפיתוח הדוח שולבו עקרונות ה-Data Storytelling כדי לסייע בהעברת מסרים משמעותיים מתוך הנתונים. עקרונות אלו מאפשרים להציג את התובנות המתקבלות בצורה ברורה ומובנת, שתומכת בתהליך קבלת החלטות המבוססות על נתונים. פירוט המלא וצילומי מסך של ה-Data Storytelling מופיעים בנספח 2.

צילום OLAP:



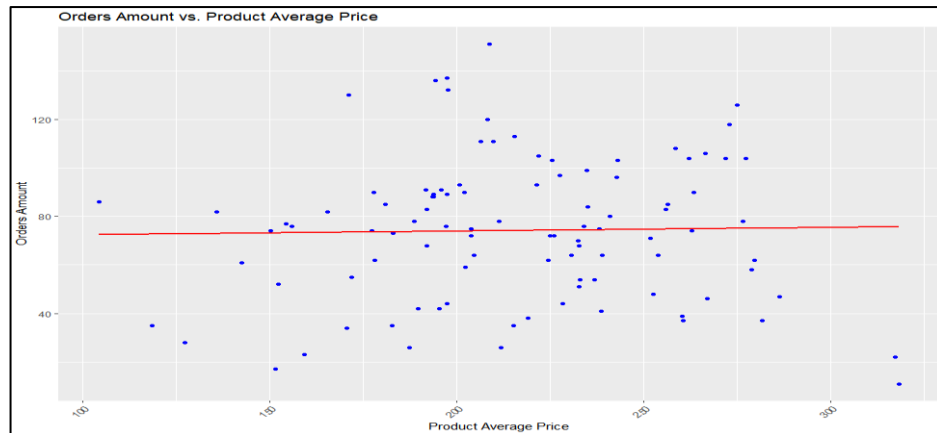
יישומי BI&A וניתוח נתונים מתקדמים

יישום א': תחקור אנליטי מאשש מלווה בניתוח השערות סטטיסטי

1. מבחני קורלציה - בחינת הקשר בין מחיר מוצר לכמות הרכישות שלו בפילוח מגדרי

נרצה לבחון את השאלה העסקית: **"האם קיים קשר בין עלות המוצר לכמות הרכישות שבוצעו ממנו?"**. לצורך בחינת הקשר בין עלות המוצר לכמות הרכישות שבוצעו ממנו, בחרנו לבצע מבחן קורלציה ולהיעזר בתרשים הפיזור. מבחן זה מאפשר לבחון האם קיים קשר בין המשתנים וכיוון הקשר – חיובי או שלילי. באופן תאורטי, היינו מצפים שככל שמחיר המוצר גבוה יותר, כמות הרכישות תקטן. כלומר, אם המוצר יקר במיוחד, סביר להניח שהלקוחות ירכשו ממנו פחות יחידות. על כן, ציפינו לקורלציה שלילית. במידה והייתה מתקבלת קורלציה שלילית משמעותית, ייתכן והיינו שוקלים להוזיל מחירים עבור מוצרים יקרים במיוחד. כדי לבחון את הקשר, חישבנו את המחיר הממוצע של כל מוצר ואת סך הכמות שנרכשה ממנו (נספח 3.1.1).

תוצאות הקורלציה שהתקבלו מצביעות על ערך נמוך מאוד של 0.0218, אשר מצביע על כך שלא קיים קשר מובהק בין מחיר המוצר לכמות הרכישות ממנו (פירוט החישוב בנספח 3.1.2).

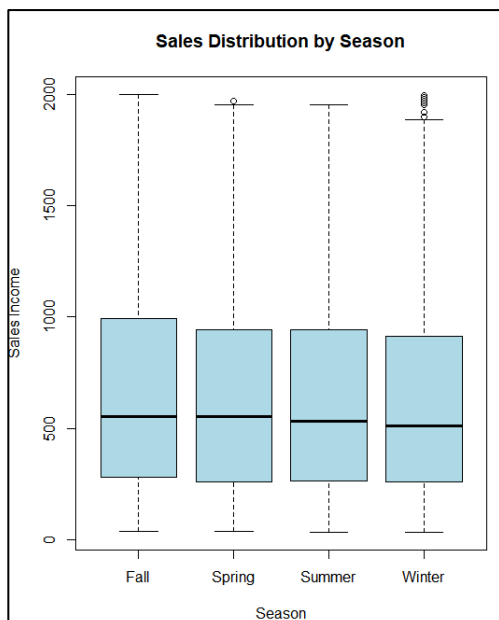


2. השוואת מכירות עונתיות - ANOVA

בהמשך למסקנות מלוח המחוונים, נרצה לבחון את השאלה העסקית: **"האם קיים הבדל בהתפלגות המכירות בעונות השונות?"**. על מנת לעשות זאת, נשתמש במבחן ANOVA מכיוון ומדובר בהשוואה של מעל ל-2 קבוצות.

$$\begin{cases} H_0: \text{Same average income} \\ H_1: \text{else} \end{cases} \quad \text{השערות המבחן:}$$

על מנת לחשב את הרווחים הממוצעים לכל עונה, הוספנו שדה קטגוריאל "Season" שמציין את העונה שבה נעשתה ההזמנה (נספח 3.2.1). חישבנו את ממוצעי המכירות בכל עונות השנה על פי החישוב בנספח 3.2.2. את הנתונים הצגנו בתרשים Boxplot בכדי לאפשר השוואה וויזואלית תוך התחשבות בסטיות התקן ותצפיות חריגות. על פי תוצאות ANOVA, התקבל $Pvalue > 0.5$ (0.591) ועל כן לא נדחה את השערת האפס ונקבע שבר"מ 5% ממוצעי המכירות בעונות השנה אינם שונים זה מזה (פירוט התוצאות בנספח 3.2.3).

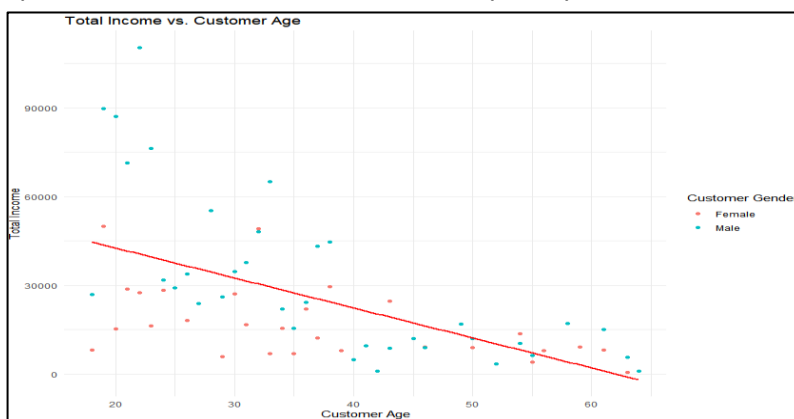


3. בחינת הקשר בין גיל ומגדר הלקוח לסך עלויות הרכישה שלו – מבחן t ומבחני קורלציה

נרצה לבחון את השאלה העסקית: "**האם קיים קשר בין גיל ומגדר הלקוח לבין סך עלויות הרכישה שלו?**". כדי לבחון את הקשר בין גיל הלקוח, מגדר הלקוח וסך עלויות הרכישה שלו, ביצענו מספר ניתוחים סטטיסטיים. תחילה, בחנו את ההבדלים בממוצעי הרכישות בין גברים לנשים באמצעות

$$\begin{cases} H_0: \text{Same average} \\ H_1: \text{else} \end{cases} \quad \text{מבחן t בלתי תלוי, השערות המבחן:}$$

תוצאות מבחן ה-t הראו כי ערך ה-t הוא 0.519 וערך p-value של 0.6039. מאחר והתקבל $Pvalue > 0.5$, נאמר שבר"מ 5% לא קיים הבדל מובהק בממוצע הרכישות בין גברים לנשים (פירוט על מבחן t בנספח 3.3.3). את הקשר בין גיל הלקוח לסך הרכישות בחנו באמצעות ניתוח קורלציה (נספח 3.3.2). עבור מבחן הקורלציה בצענו חישוב עבור סך הרכישות בפילוח לפי גיל (נספח 3.3.1). תוצאות הקורלציה לבחינת הקשר בין גיל הלקוח לסך הרכישות הראו ערך של (-0.5817). מה שמצביע על קשר בינוני-חזק בכיוון שלילי בין גיל הלקוח לסך ההכנסות, כלומר ככל שהלקוח מבוגר יותר ההכנסות שלו לחברה נמוכות יותר. את התוצאות בחרנו להציג באמצעות גרף הפיזור של גיל הלקוח וסך הרכישות שלו, בפילוח לפי מין הלקוח ע"י חלוקה לצבעים. הגרף מאפשר תצוגה חזותית של הקשר בין המשתנים וזיהוי מגמות נוספות על פי מין הלקוח.



יישום ב': חיזוי מספר היחידות שימכרו בשנה עבור כל מוצר בכל מדינה

הערכת הביקוש למוצר מסוים במדינה מאפשר ייעול של תהליכי הייצור והמכירה. לכל מדינה נייצר בהתאם לביקוש השנתי החזוי מהמוצר, נבצע אספקה של כמות היחידות הצפויה ונמנע מעלויות אחסנה מיותרות. זה יאפשר שינוע סחורה באופן יעיל בין המדינות השונות, חיסכון בעלויות משלוח ומיקסום הרווחים. כיוון שמדובר בחיזוי של כמות יחידות שהינו מספר חיובי ושלם, החלטנו לעגל כלפי מעלה את ערכי החיזוי מתוך ההנחה שנשאף לעמוד בביקוש ולכן נעדיף לייצר יותר יחידות. **בחירת המשתנים** - בחרנו לבחון את השפעת המשתנים הקטגוריאליים הבאים על חיזוי כמות היחידות שנרכשו ממוצר ספציפי במדינה: קבוצת גיל הלקוח, מגדר הלקוח, חברות במועדון, וסוג ההזמנה. בנוסף, בחנו את השפעת המשתנים הכמותיים הבאים: מחיר המוצר וגיל הלקוח. **הכנת הנתונים** - מיזוג על ידי פעולת INNER JOIN של טבלאות הנתונים "FACT – INCLUDES" ו-"CUSTOMERS". לאחר המיזוג, המרנו משתנים קטגוריאליים לפקטורים וביצעו השטחה של הנתונים על פני הזמן כדי להסיר את מימד הזמן. ההשטחה נעשתה על ידי אגרגציה שנתית תוך

שימוש בסכום לכמות, בממוצע למחיר ולגיל, ובחירה של הערך השכיח עבור המשתנים הקטגוריאליים. פירוט המשתנים (מומנטים) שהופקו מופיע בנספח 4.1.

אימון מודל רגרסיה לינארית - הנתונים חולקו למערכי אימון (80%) ומבחן (20%).

תוצאות המודל - הפלט המלא מופיע בנספח 4.2.

```
Residual standard error: 0.1993 on 1035 degrees of freedom
Multiple R-squared: 0.9794, Adjusted R-squared: 0.979
F-statistic: 3068 on 16 and 1035 DF, p-value: < 2.2e-16
> cat("Root Mean Squared Error (RMSE):", rmse, "\n")
Root Mean Squared Error (RMSE): 0.8708854
```

ניתוח התוצאות -

- מתוצאות המודל ניתן לראות כי התקבלו ערכי Adjusted R-squared (0.9794) ו-R-squared (0.979) המעידים על כך שהמודל מסביר כ-97.9% מהשונות בסך היחידות שנמכרו בשנה. כלומר, ישנה התאמה טובה של המודל הלינארי לחיזוי.
- שגיאת ה-RMSE (Root Mean Square Error) מודדת את השורש הריבועי של ממוצע השגיאות בריבוע במודל, כלומר את הסטייה הממוצעת של התחזיות מהערכים בפועל. ככל שערך שגיאה זו קטן יותר, כך המודל נחשב למדויק יותר. במודל שלנו התקבל ערך RMSE של 0.8708854, שמעיד על כך שהסטייה הממוצעת של התחזיות מהערכים האמיתיים היא כ-0.87 יחידות מוצר, כלומר הבדל ממוצע של פחות מיחידה אחת בין הכמויות החזויות לאמיתיות, מה שמעיד על דיוק יחסי גבוה של המודל.
- ערך ה-p-value הנמוך שהתקבל עבור מבחן ה-F מעיד על מובהקות גבוהה של המודל, במילים אחרות לפחות משתנה מסביר אחד במודל הוא מובהק ותורם בצורה משמעותית להסברת כמות היחידות השנתית שנמכרות ממוצר מסוים במדינה.

משמעות המשתנים המובהקים שהתקבלו -

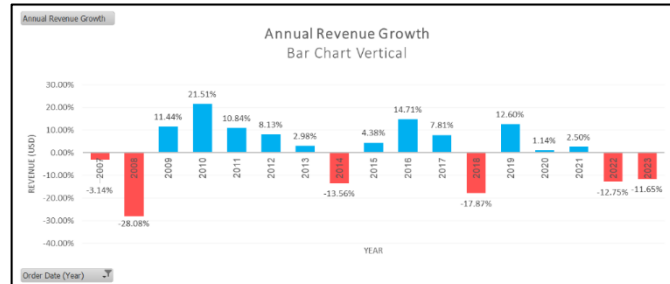
- Max_Annual_Quantity ו-Min_Annual_Quantity: המשתנים הללו מראים את הכמות השנתית המינימלית והמקסימלית שנמכרה לאורך השנים. שניהם נמצאו מובהקים מאוד ומשפיעים באופן חיובי על הכמות השנתית הממוצעת החזויה. ככל שהכמות השנתית המקסימלית והמינימלית גבוהים יותר, כך הכמות השנתית הממוצעת תהיה גבוהה יותר.
- Avg_Age: הגיל הממוצע נמצא מובהק ומשפיע באופן שלילי על הכמות השנתית הממוצעת. ככל שגיל הלקוחות הממוצע עולה, הכמות השנתית הממוצעת של היחידות הנמכרות יורדת.
- Customer_Age_Group: קבוצת הגיל השכיחה של הלקוחות משפיעה באופן חיובי ומובהק על הכמות השנתית הממוצעת.
- Most_Frequent_Gender: משפיע באופן חיובי ומובהק על הכמות השנתית הממוצעת.
- Avg_Male_Count: משפיע באופן שלילי ומובהק על הכמות השנתית הממוצעת.
- Avg_Online_Count: משפיע באופן חיובי ומובהק על הכמות השנתית הממוצעת.

תוצאות מדד VIF לבחינת מולטי-קולינאריות של המודל מצורפות בנספח 4.3.

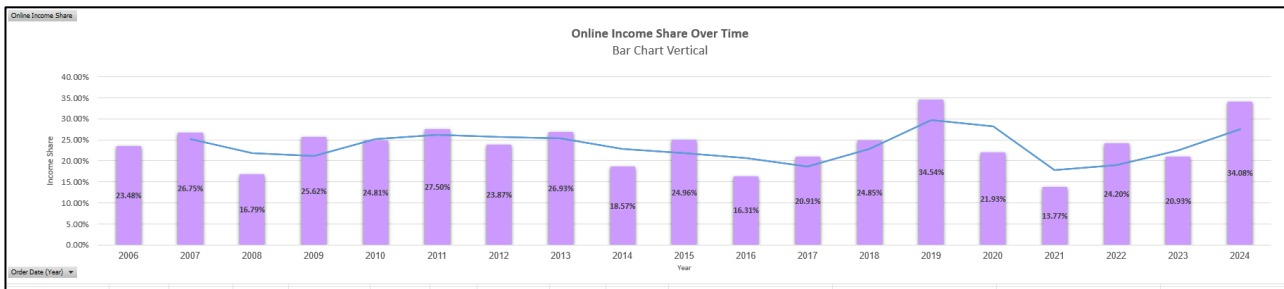
הנחות מודל רגרסיה לינארית נבחנו בפירוט בנספח 4.2. לאור ההפרות בהנחות הרגרסיה הלינארית, נמליץ להמשיך ולבחון מודלים נוספים כדי לשפר את יכולת החיזוי. למשל, ניתן לנסות לשפר את המודל באמצעות טרנספורמציות על המשתנים או שימוש במודלים לא לינאריים.

נספח 1: KPI

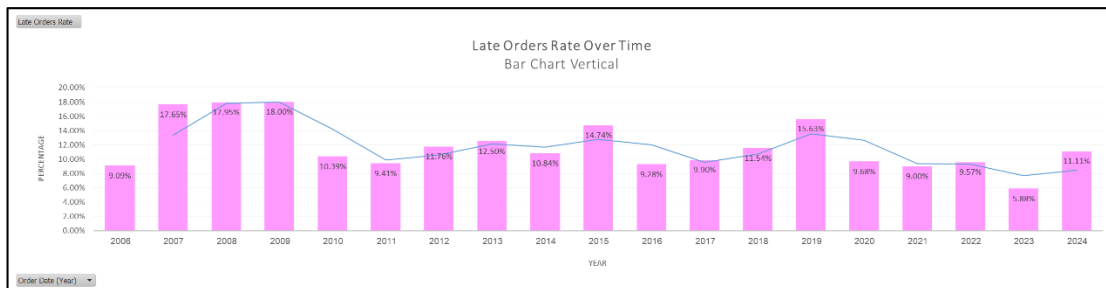
נספח 1.1: KPI Annual Income Growth



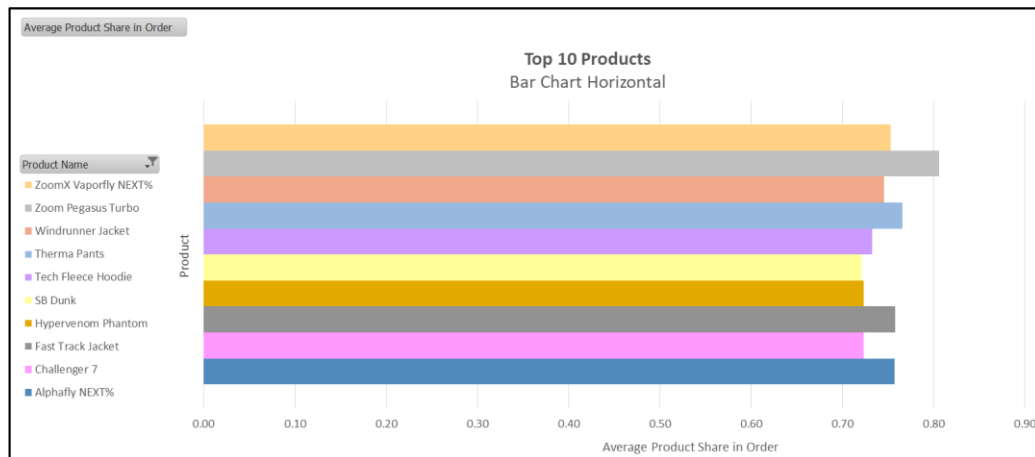
נספח 1.2: KPI Online Income Share

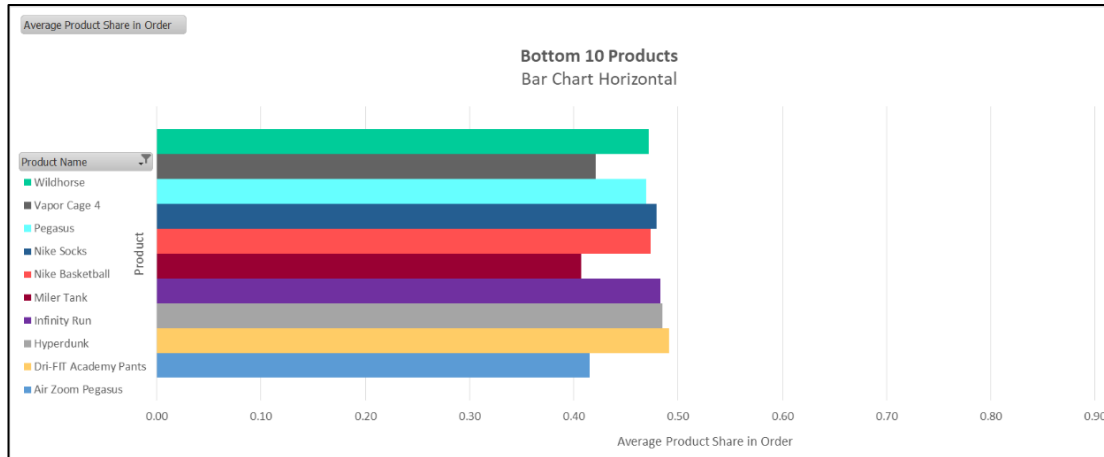


נספח 1.3: KPI Late Order Rate



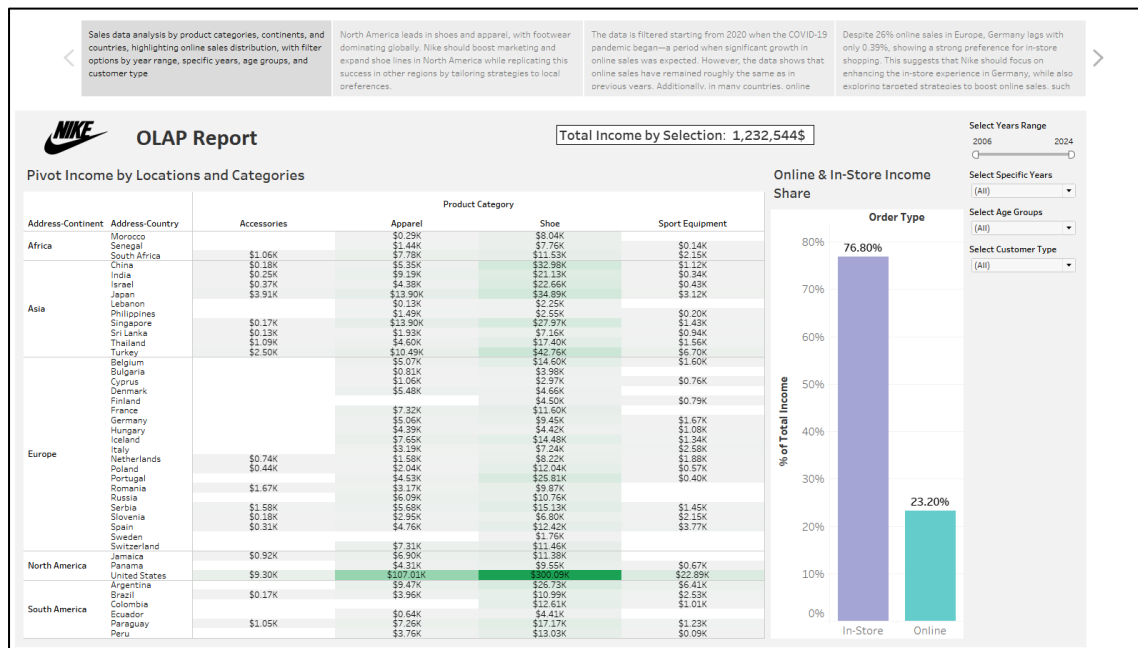
נספח 1.4: KPI Product Share in Order



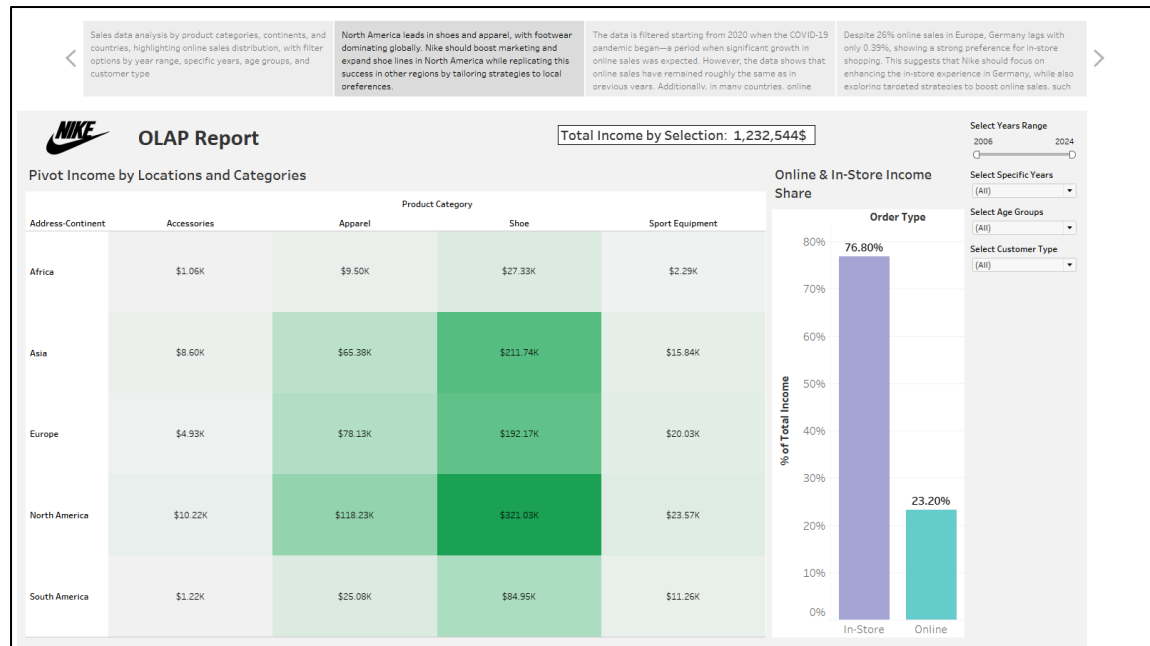


OLAP – Data Storytelling :נספח 2

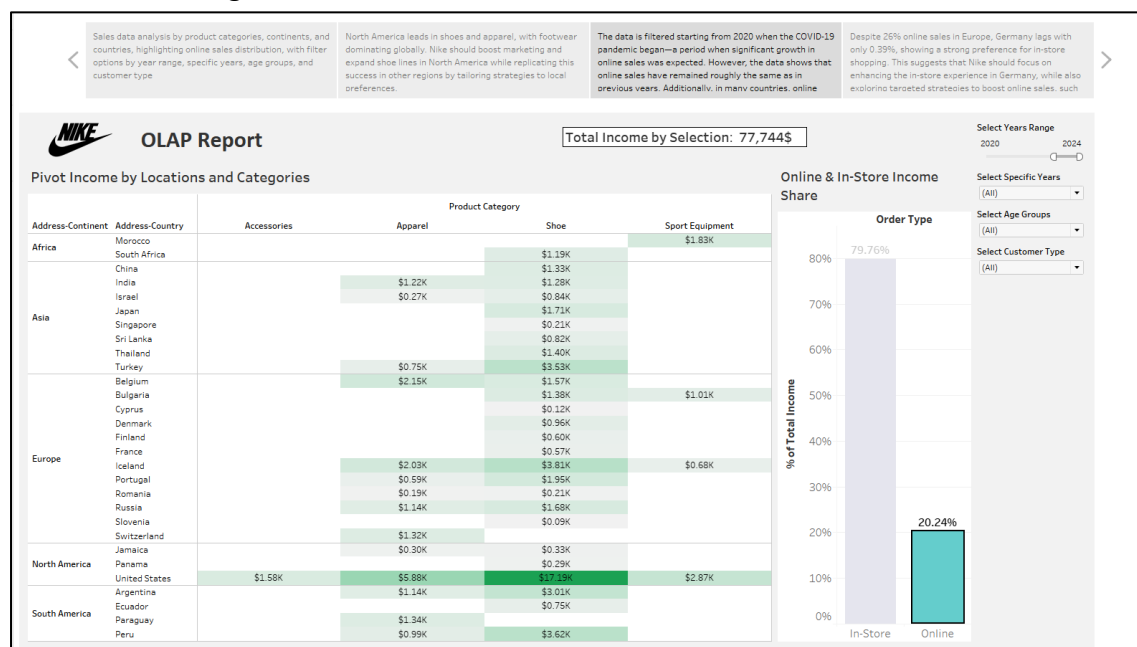
(1) Sales data analysis by product categories, continents, and countries, highlighting online sales distribution, with filter options by year range, specific years, age groups, and customer type.



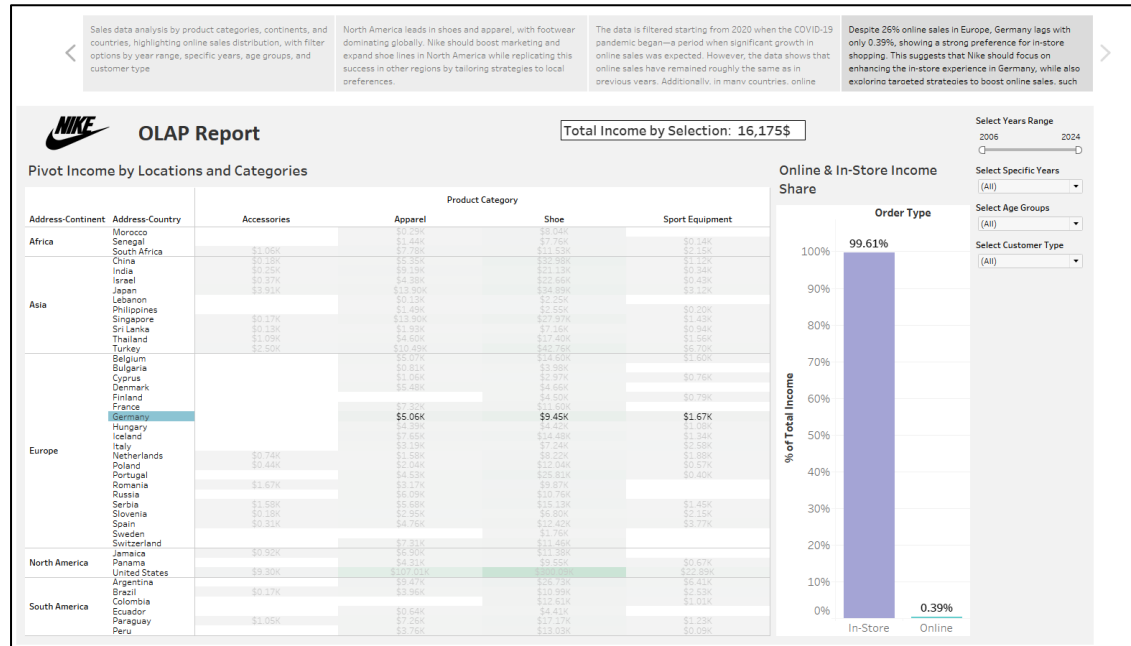
(2) North America leads in shoes and apparel, with footwear dominating globally. Nike should boost marketing and expand shoe lines in North America while replicating this success in other regions by tailoring strategies to local preferences.



(3) The data is filtered starting from 2020 when the COVID-19 pandemic began—a period when significant growth in online sales was expected. However, the data shows that online sales have remained roughly the same as in previous years. Additionally, in many countries, online sales are concentrated in a single product category, often shoes. Given this, Nike should consider enhancing the digital shopping experience and expanding online marketing efforts to encourage purchases across more categories and boost overall online sales.



(4) Despite 26% online sales in Europe, Germany lags with only 0.39%, showing a strong preference for in-store shopping. This suggests that Nike should focus on enhancing the in-store experience in Germany, while also exploring targeted strategies to boost online sales, such as localized promotions or improving the online shopping experience to better meet German consumer preferences.



נספח 3: תחקור אנליטי מאשש מלווה בניתוח השערות סטטיסטיות

נספח 3.1: מבחני קורלציה - בחינת הקשר בין מחיר מוצר לכמות הרכישות שלו בפילוח מגדרי

1. חישוב עלות ממוצעת למוצר וסך הרכישות ממנו:

```
summary_df <- df %>%
  group_by(`Product Name`) %>%
  summarize(AveragePrice = mean(`Version Price`, na.rm = TRUE),
            TotalQuantity = sum(Quantity, na.rm = TRUE))
```

2. חישוב קורלציה בין עלות המוצר לכמות הרכישות:

```
correlation_result <- cor(summary_df$AveragePrice, summary_df$TotalQuantity, use = "complete.obs")
```

```

              AveragePrice TotalQuantity
AveragePrice  1.00000000  0.02178667
TotalQuantity 0.02178667  1.00000000
  
```

נספח 3.2: השוואת מכירות עונתיות - ANOVA

1. יצירת שדה עונות:

```
df$Season <- ifelse(format(as.Date(df$`Order Date`), "%m") %in% c("12", "01", "02"), "winter",
                    ifelse(format(as.Date(df$`Order Date`), "%m") %in% c("03", "04", "05"), "Spring",
                            ifelse(format(as.Date(df$`Order Date`), "%m") %in% c("06", "07", "08"), "Summer", "Fall"))))
```

חלוקת החודשים לעונות:

- חורף: דצמבר – פברואר
- אביב: מרץ – מאי
- קיץ: יוני – אוגוסט
- סתיו: ספטמבר – נובמבר

2. חישוב ממוצע המכירות בכל עונה:

```
seasonal_sales <- df %>%
  group_by(Season) %>%
  summarize(MeanSales = mean(Income))
```

3. חישוב ANOVA:

```
anova_model <- aov(Income ~ Season, data = df)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Season	3	410363	136788	0.637	0.591
Residuals	2489	534202026	214625		

נספח 3.3: בחינת הקשר בין גיל ומגדר הלקוח לסך עלויות הרכישה - מבחן t ומבחני קורלציה

1. חישוב סך עלויות רכישה לפי גיל הלקוח:

```
income_by_age <- sales_data %>%
  group_by(`Customer Age`, `Customer Gender`) %>%
  summarize(TotalIncome = sum(Income, na.rm = TRUE))
```

2. חישוב קורלציה בין גיל לסך עלויות הרכישה:

```
correlation_result <- cor(income_by_age$`Customer Age`, income_by_age$TotalIncome, use = "complete.obs")
"Correlation between Customer Age and Total Income: -0.581727618124644"
```

3. מבחן t עבור מין וסך עלויות הרכישה:

```
t_test_result <- t.test(Income ~ `Customer Gender`, data = sales_data)
```

```
Welch Two Sample t-test

data: Income by Customer Gender
t = 0.519, df = 1295, p-value = 0.6039
alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
95 percent confidence interval:
 -30.14235  51.82756
sample estimates:
mean in group Female  mean in group Male
        651.4004         640.5578
```

נספח 4: חיזוי מספר היחידות שימכרו בשנה עבור כל מוצר בכל מדינה

1. משתנים שהופקו עבור מודל החיזוי:

- כמות שנתי ממוצעת (Avg_Annual_Quantity)
- שיפוע כמות שנתי (Slope_Annual_Quantity)
- כמות שנתי מינימלית (Min_Annual_Quantity)
- כמות שנתי מקסימלית (Max_Annual_Quantity)
- מחיר ממוצע (Avg_Price)
- גיל ממוצע (Avg_Age)
- קבוצת גיל הלקוח השכיחה (Customer_Age_Group)
- מגדר הלקוח השכיח (Most_Frequent_Gender)
- סוג ההזמנה השכיח (Most_Frequent_Order_Type)
- חברות במועדון השכיחה (Most_Frequent_Membership)
- מספר גברים ממוצע (Avg_Male_Count)
- מספר הזמנות אונליין ממוצע (Avg_Online_Count)
- מספר חברות במועדון ממוצע (Avg_Membership_Count)

2. תוצאות המודל רגרסיה:

```
Call:
lm(formula = .outcome ~ ., data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-2.12393 -0.05382 -0.02456  0.01400  1.11317

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.416e-01  4.212e-02   3.362 0.000801 ***
Slope_Annual_Quantity  4.438e-03  6.880e-03   0.645 0.519069
Min_Annual_Quantity  5.777e-01  4.857e-03 118.950 < 2e-16 ***
Max_Annual_Quantity  4.071e-01  3.990e-03 102.024 < 2e-16 ***
Avg_Price     -3.184e-06  6.393e-05 -0.050 0.960286
Avg_Age      -4.661e-03  1.654e-03 -2.817 0.004937 **
  factor(Customer_Age_Group)2  4.721e-02  1.945e-02  2.427 0.015397 **
  factor(Customer_Age_Group)3  1.001e-01  2.494e-02  4.015 6.39e-05 ***
  factor(Customer_Age_Group)4  1.036e-01  4.698e-02  2.206 0.027591 *
  factor(Customer_Age_Group)5  1.673e-01  5.886e-02  2.842 0.004574 **
  factor(Customer_Age_Group)6  2.165e-01  8.091e-02  2.676 0.007568 **
  factor(Most_Frequent_Gender)2  1.133e-01  3.629e-02  3.123 0.001842 **
  factor(Most_Frequent_Order_Type)2 -7.129e-02  3.720e-02 -1.916 0.055604 .
  factor(Most_Frequent_Membership)Y -4.861e-03  5.080e-02 -0.096 0.923785
Avg_Male_Count -9.727e-02  3.622e-02 -2.685 0.007361 **
Avg_Online_Count  8.325e-02  3.932e-02  2.117 0.034467 *
Avg_Membership_Count  3.484e-02  5.287e-02  0.659 0.510075

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1993 on 1035 degrees of freedom
Multiple R-squared:  0.9794,    Adjusted R-squared:  0.979
F-statistic: 3068 on 16 and 1035 DF, p-value: < 2.2e-16
```

3. תוצאות מדד VIF למולטי-קולינאריות:

ערכי ה-VIF (Variance Inflation Factor) שהתקבלו נמוכים מ-10, מה שמעיד על כך שבעיית המולטי-קולינאריות (תלות גבוהה בין המשתנים המסבירים) אינה חמורה במודל. עם זאת, בבחירת משתנים מובהקים ניתן להסיר מומנטים של משתנים מסוימים ולהפחית עוד יותר את המולטי-קולינאריות.

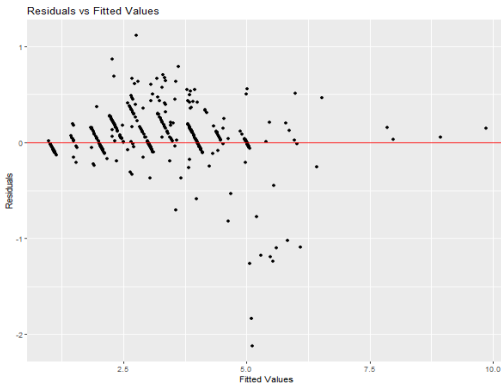
```
> # Variance Inflation Factor (VIF) for Multicollinearity
> vif_values <- vif(model$finalModel)
> vif_table <- data.frame(VIF_Value = vif_values)
> print(vif_table)
```

	VIF_Value
Slope_Annual_Quantity	1.022142
Min_Annual_Quantity	1.329827
Max_Annual_Quantity	1.480444
Avg_Price	1.025774
Avg_Age	6.295961
factor(Customer_Age_Group)2	1.647174
factor(Customer_Age_Group)3	3.725374
factor(Customer_Age_Group)4	2.798042
factor(Customer_Age_Group)5	3.758764
factor(Customer_Age_Group)6	2.757577
factor(Most_Frequent_Gender)2	7.197871
factor(Most_Frequent_Order_Type)2	6.205173
factor(Most_Frequent_Membership)Y	7.159260
Avg_Male_Count	7.197684
Avg_Online_Count	6.239231
Avg_Membership_Count	7.139935

4. בחינת הנחות מודל רגרסיה לינארית: תרשימים ותוצאות מבחנים סטטיסטיים

לינאריות, הומוסקדסטיות ואי תלות השאריות:

בתרשים Residuals vs. Fitted Values ניתן להבחין כי השאריות אינן מפוזרות באופן אקראי ואינן מתרכזות סביב קו האפס, דבר המעלה חשש להפרה של הנחות הלינאריות, הומוסקדסטיות (שוונות שאריות אחידה) ואי תלות השאריות. כדי לבחון אם הנחות אלו מתקיימות, ביצענו שני מבחנים: Breusch-Pagan וגם Durbin-Watson.



- מבחן Breusch-Pagan נועד לבחון האם השוונות של השאריות היא אחידה. ערך p-value נמוך מאוד שהתקבל מצביע על כך שהשוונות אינן אחידה, ולכן נסיק כי הנחה זו אינה מתקיימת.
 - מבחן Durbin-Watson בוחן את הנחת אי תלות השאריות. התקבל ערך p-value נמוך ולכן נקבע כי קיימת תלות בין השאריות שמעיד על הפרת הנחה זו.

```
> # Breusch-Pagan Test for Homoscedasticity Assumption
> bptest(model$finalModel)

studentized Breusch-Pagan test

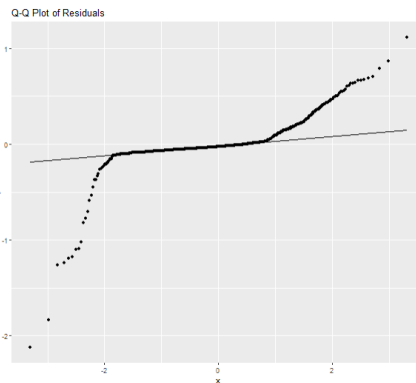
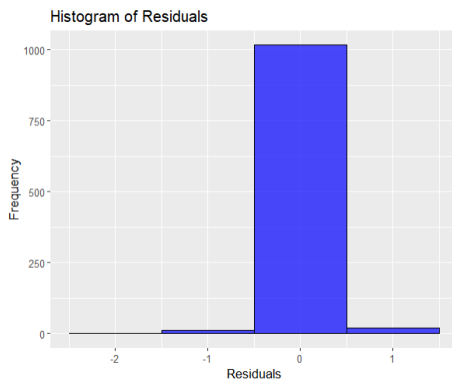
data: model$finalModel
BP = 365.26, df = 16, p-value < 2.2e-16

> # Durbin-Watson Test for Independence Assumption
> dwtest(model$finalModel)

Durbin-Watson test

data: model$finalModel
DW = 1.9043, p-value = 0.04886
alternative hypothesis: true autocorrelation is greater than 0
```

נורמליות השאריות:



בהתבסס על תרשימים ההיסטוגרמה ניתן לראות כי השאריות אינן מתפלגות באופן נורמלי. בנוסף, בתרשימים QQ ניכרות סטיות משמעותיות מהקו האלכסוני המייצג את ההתפלגות הנורמלית התיאורטית, כאשר הסטיות בולטות במיוחד בקצוות. ממצאים אלו מעידים כי השאריות אינן מתפלגות נורמלית.

נבחן את הנחת הנורמליות על ידי שימוש בשני מבחני טיב ההתאמה הבאים: Kolmogorov-Smirnov וגם Shapiro-Wilk. מבחנים אלו משווים בין התפלגות השאריות להתפלגות הנורמלית התיאורטית בהשערת האפס. התוצאות המובהקות שהתקבלו מביצוע שני המבחנים הללו, ערך p-value נמוך מאוד, מאששים כי הנחת הנורמליות מופרת במודל זה.

```
> # Shapiro-Wilk Test for Normality
> shapiro_test <- shapiro.test(residuals)
> print(shapiro_test)

Shapiro-Wilk normality test

data: residuals
W = 0.61164, p-value < 2.2e-16

> # KS Test for Normality
> ks_test <- ks.test(residuals, "pnorm", mean = mean(residuals), sd = sd(residuals))
> print(ks_test)

Asymptotic one-sample Kolmogorov-Smirnov test

data: residuals
D = 0.26147, p-value < 2.2e-16
alternative hypothesis: two-sided
```