

פרופ' גלעד רVID



אוניברסיטת בן-גוריון בנגב  
Ben-Gurion University of the Negev

# דוח מסכם – ויזואלייזציה של המידע

נושא: חברת איקומרס Olist

מגישים:

אוהד אשכנזי 016323183

מלacci ויס 33896431

**תוכן עניינים**

|                            |    |
|----------------------------|----|
| תיאור הנתונים ואופן הטיפול | 2  |
| דיון בויזואלייזציות קיימות | 3  |
| <br>                       |    |
| ויזואלייזציה ראשונה        | 4  |
| ויזואלייזציה שנייה         | 5  |
| ויזואלייזציה שלישיית       | 6  |
| ויזואלייזציה רביעית        | 7  |
| ויזואלייזציה חמישית        | 8  |
| ויזואלייזציה ששית          | 9  |
| ויזואלייזציה שביעית        | 10 |

## תיאור הנתונים ואופן הטיפול

הנתונים בפרויקט מבוססים על מאגר נתונים איקומרס של פלטפורמת Olist וכוללים מידע רב-מדדי על הזמן, פרטיים, לקוחות, מוצרים, קטגוריות מוצרים, תשלום ומשלוחים. בסיס הנתונים בוויי מסמך טבלאות רלוונטיות המוחברות באמצעות מזהים ייחודיים (כגון מזהה הזמן, מזהה לקוח ומזהה מוצר), וכן שלב ראשון בתהליך כלל מיזוג (join) בין הטבלאות לצורך יצירת טבלאות אנליטיות מאוחדות. לאחר האיחוד בוצע ניקוי נתונים شامل טיפול בערכים חסרים, הסרת כפליות, תיקון טיפוס משתנים (בעיקר שדות תאריך ושעה), והמרת חוטמוות זמן לפורמטים אחידים המאפשרים ארגזיה לפי יום, שבוע, חודש ורבעון. בנוסף, חושבו מדדים נגזרים כגון סך הכנסות להזמנה, מחיר ממוצע לפריט, אחוז צמיחה בין תקופות, שיור איחורים במשלוחים, מדדי ריכוזיות (N-Top), ועוד מותאם בין משתנים מרכזיים. במקרים בהם זהה ערכים קיצוניים חריגים (outliers), בוצע חיתוך סטטיסטי מבוקר (למשל Trim באחוזונים 99%-1%) על מנת למנוע עיוות של התמונה הכללת. כמו כן בוצעה ארגזיה ברמות שונות – קטגורית מוצר, תקופה זמן, אזור גיאוגרפי ומוכר – בהתאם לשאללה העסקית שנבדקה בכל יזואליזציה. תהליך ההכנה שם דגש על עקביות מדדים, שמירה על פרשנות עסקית נכונה, והבטחת יכולת אינטראקטיבית לסייע, קיבוץ ושינוי גורנולריות בזמן אמיתי, כך שהנתונים יהיו לא רק נקיים ואמינים אלא גם מותאים לניתוח חזותי תומך החלטה.



## דיאן בויזואלייזציה קיימות לאותם נתונים



מתוך האתר Kaggle (מайפה שלקחנו את הנתונים)

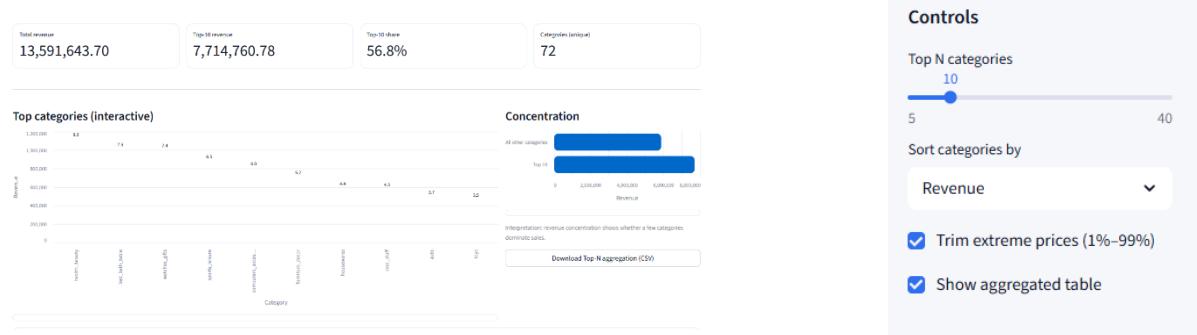
<https://www.kaggle.com/code/philipclemente/olist-e-commerce-analysis-and-dashboard/notebook>

### התמקדות בויזואלייזציה השנייה והשווואה:

במסגר שולנו הצגנו ויזואלייזציה מתקדמת ורב-מדנית אשר משלבת במקביל גם את נפח ההזמנות וגם את סך הכנסות לאורך זמן, ובכך מאפשרת ניתוח דינמי של הקשר בין פעילות תעשיית לביצועים פיננסיים. זאת בגין גראף הבסיסי המופיע ב-Kaggle, אשר מציג מכירות חודשיות בלבד בצורה סטטיסטית ולא העמיקה אנליטית. הוויזואלייזציה שלנו כוללת אינטראקטיביות מלאה – אפשרות לבצע Hover לקלת ערכים מדויקים, לסנן טווחי זמן, לשנות רוחניים תצוגה (שבועי/חודש/רבעון), ולהשווות בין המדדים. בנוסף, באמצעות מבנה של Small Multiples ניתן לבחון כל ממד בסקירה טبعית ונפרד, תוך שמירה על אותו ציר זמן להשוואה אינטואיטיבית. גישה זו מאפשרת לא רק לראות "כמה נמכר", אלא להבין האם העלייה במכירות מלווה בעלייה פרופורציונלית בהכנסות, האם קיימות תקופות של שחיקת ערך להזמנה, והאם הדינמיקה מצביעה על שינוי מבני בפלטפורמה – תובנות שאינן ניתנות ליהו' בגרף בסיסי חד-מדדי.

# יזואלייזציה ראשונה

## Pareto Analysis



## ניתוח תיאורתי WWW

**What** - הנתונים הם נתונים טבלאיים אגרגטיביים בرمת קטגוריה אחת, והמשתנים המרכזים הם: סך ההכנסות (סכום מחירי הפריטים), מחיר ממוצע לפרט, מספר פריטים שנמכרו ואחוז מההכנסה הכוללת. מדובר בשילוב של משתנה קטגוריאלי (שם קטgorיה) וממשתנים כמוותים רציפים.

**Why** - המשימה האנליטית היא זהוו ריכוזות הכנסות בפלטפורמה, איתור קטגוריות דומיננטיות, והשואאה בין גודל פיננסי לבין המחיר הממוצע. בנוסף, מבקשים לבחון האם הכנסות מפוזרות באופן שוווני יחסית או שמא קיימת תלות במספר קטן של קטגוריות מובילות (עקרון Pareto)

**How** - הקידוד הויזואלי נעשה באמצעות טבלה עם מדדי קידוד הכנסות (עורך אורך – מדijk תפיסתית), קו נוסף לקידוד המחיר הממוצע (ציר משני), ותוויות האחוזים להציג תרומות כל קטגוריה להכנסה הכוללת. בנוסף, מוצג טבלה ריכוזיות משלים המשווה בין N-Top קטגוריות לבין שאר הקטגוריות. האינטרקטיביות מאפשרת שליטה במספר הקטגוריות המציגות, שינוי קритריון המין, ווינון מחירים קיצוניים.

## הסבר בחירות העיצוב והשאלה העסקית

הגרף עונה על השאלה:

האם הכנסות מפוזרות על פני קטגוריות רבות, או שמספר קטן של קטגוריות מהוות את עיקר הפעולות העסקית?

הבחירה בתרשימים מעודדת נובעת מהעובדה שערוך האורך מאפשר השוואת בין קטגוריות בדידות. הוספה קו המהווה הממוצע מאפשרת זיהוי תופעות כגון קטגוריות עם הכנסה גבוהה אך מחיר נמוך (נפח גדול), לעומת קטגוריות פרימיום עם מחיר גבוה אך נפח נמוך יותר.

הוספה אחוז מההכנסה הכוללת מעל כל עמודה מיישמת את עקרון **Overview+Detail**: המשתמש רואה גם את הערך האבסולוטי וגם את המשמעות היחסית שלו.

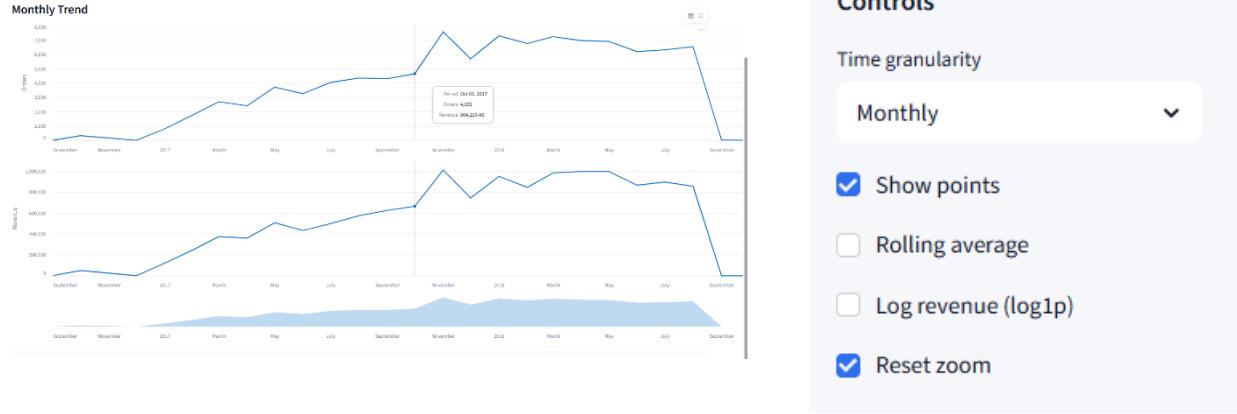
trsheets הריכוזיות המשלים N-Top (מול היתר) מאפשר לזהות האם הפלטפורמה נשענת על מספר מצומצם של מנועי צמיחה – תובנה קריטית לניהול סיכון עסקים ולגיאון פורטפולוי.

היכולת לשולט במספר הקטגוריות (N-Top) מימושה את עקרון "מבט על תחילת, ולאחר מכן זום", בכך שהוא מאפשר לבחון את המערכת ברמת מקורה ולאחר מכן להתמקד בקבוצות ספציפיות.

סינון מחירים קיצוניים (trim 1%-99%) משקף עקרון סטטיסטי חשוב של הפחתת השפעת outliers לצורך קבלת תמונה יציבה ואמינה יותר.

## ויזואלייזציה שנייה

### Small Multiples Time Series



### ניתוח תיאורתי WWH

- **What** - הנתונים הם נתונים סדרתי-זמן אגרגטיביים, כאשר כל רשומה מייצגת יחידת זמן (שבועית או חודשית). המשתנים הקיימים הם: מספר הזמן ייחודי וסך הכנסות הנובעות מפרק הזמן. מדובר בשני משתנים רציפים בעלי סקלנות שונות, אך מבוססים על אותו ציר זמן.

- **Why** - המטרה האנליטית היא זיהוי מגמות לאורך זמן, איתור תקופות האטה בפועלות, ובוחנת הקשר הדינמי בין נפק פעילות (מספר הזמן) לבין ביצוע פיננסי (הכנסות). בנוסף, מבקשים לבחון האם קיימת סינכרוניזציה בין שני המדדים או שמא קיימים פערים המעידים על שינוי במבנה הביקוש (למשל עליה בכמות ללא עליה תואמת בהכנסה).

- **How** - הקידוד הויזואלי מתבצע באמצעות שני תרשימי קו נפרדים (Small Multiples) החולקים ציר זמן משותף. כל ממד מוצג עם ציר Y עצמאי בעל סקללה טבעית ומספרים מלאים.

### הסבר בחירות העיצוב והשאלת העסקית

הגרף עונה על השאלה המרכזית:  
יכיזד מתפתחות הכנסות והכנסות לאורך זמן, והאם קיים קשר עקבי ביניהן?

בחירה ב-Small Multiples במקומן ציר כפול נובעת מעקרונות ויזואלייזציה אקדמיים הממליצים להימנע מציר Y כפויים כאשר קיימים פער סדרי גודל משמעותיים. הפרדה זו מפחיתה עומס קוגניטיבי, משמרת קריאות כמותית מלאה, ומאפשרת השוואה אינטואיטיבית לאורך אותו ציר זמן.

הוסף Brush בתחתיו הגраф מיישמת את עקרון: **Overview first, zoom and filter, details on demand**.

המשמש רואה את התמונה הכלולית, בוחר טווח רלוונטי, ומעמיך בניתו מבלי לעבור דף או לשנות גרפ.

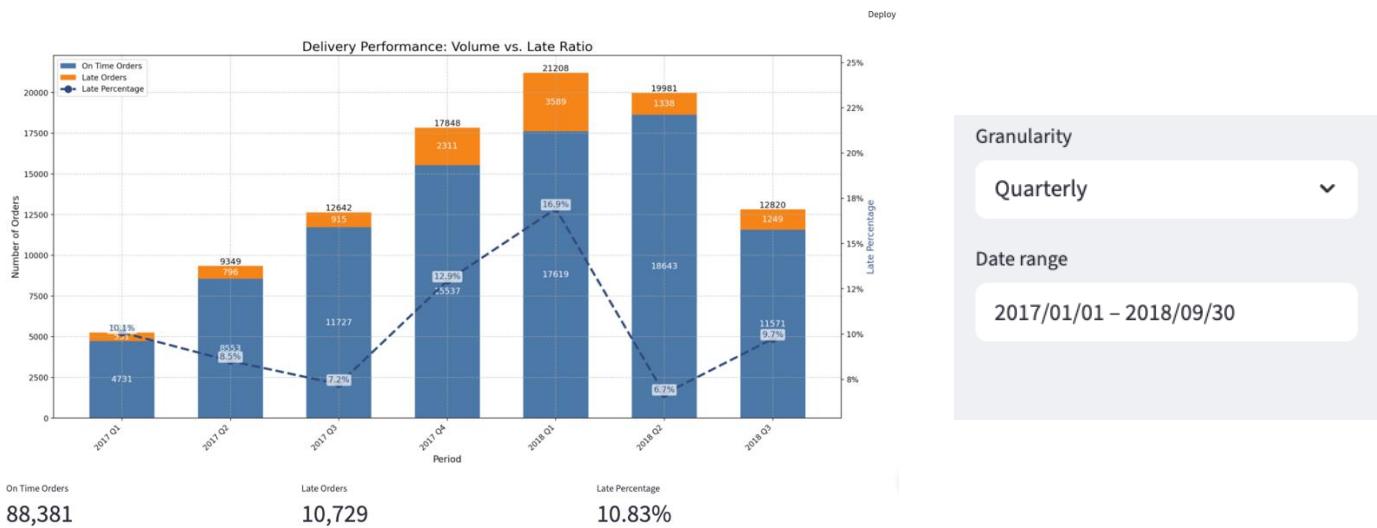
ה- Hover מאפשר דיקט מספרי נקודתי ללא עומס טקסטואלי, קבוע על הציר, וכך נשמר איזון בין מינימליزم עיצובי לבין דיקט אנליטי.

הוסף ממוצע נועד לאפשר להבדיל בין תנודות מקרית לבין מגמה מבנית, בהתאם לעקרונות ניתוח סדרות זמן. האפשרות לסקלה לוגריתמית רלוונטית במצבים של צמיחה אקספוננציאלית, ומונעת דחיסה של תקופות מוקדמות.

מדד הסיכון בראש הדף (סה"כ הכנסות, סה"כ הוצאות, וטווות תאריכים) מספקים הקשר כמותי ראשון לפני בחינת הדינמיקה, ומישמים עקרון Contextual Framing.

## ויזואלייזציה שלישית

### Stacked Bar Chart



### ניתוח תיאורתי WWH

**What** - הנתונים הם נתונים טבלאיים הכלולים משנתנו זמן מפתח ראשי, ומשתנים של סטטוס המשלוח וכמות ההזמנה.

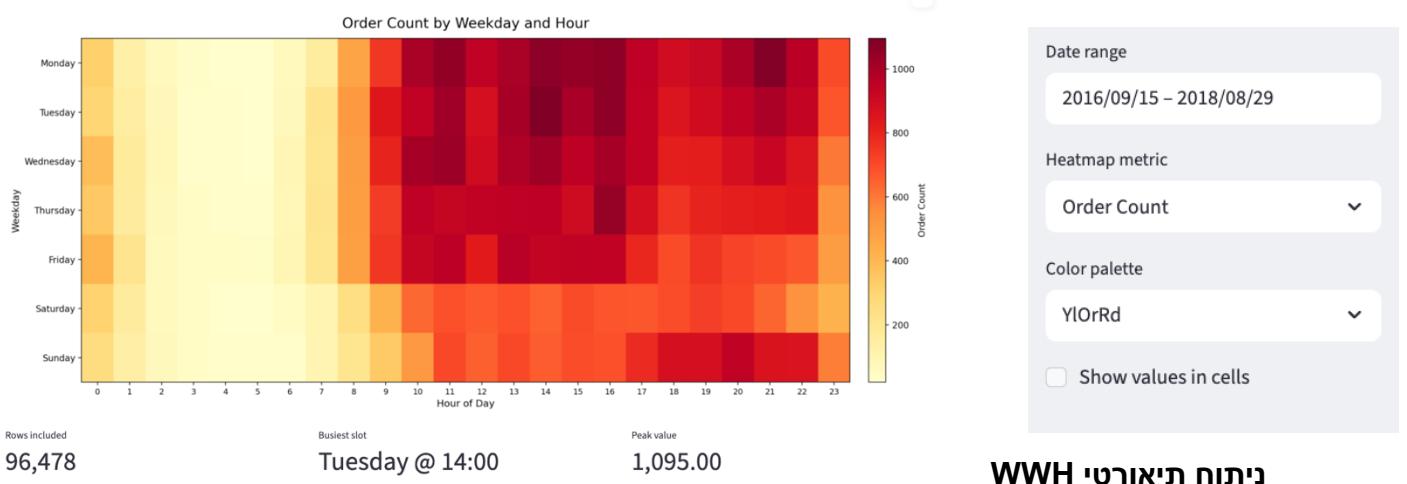
**Why** - המשימה המרכזית היא זיהוי מגמות עונתיות בביוצי השילוח, איתור תקופות חריגות בהן אחוז האיחורים זינק בצורה קיצונית, והשוואה בין נפח הפעולות לבין איכות הביצוע.

**How** - הקידוד הויזואלי נעשה באמצעות שימוש בקיים עבור המגמה ובשיטה عمودות עבור הנקודות. ערוץ האור משמש לקידוד כמות הזמן מכיוון שהוא מאפשר תפיסת מדיקת יחסית, בעוד ערוץ הצבע משמש להפרדה בין הקטגוריות. בנוסף, האינטראקטיבית מאפשרת מניפולציה של הנתונים באמצעות סינון וקבוץ לשוני רוחוצייטת הזמן.

### הסבר בחירות העיצוב והשאלה העסקית

הgraf עונה על השאלה האם הגידול בנפח ההזמנות פוגע באיכות השירות הלוגיסטי. ניתן לפתר את רוחוציות הזמן (חודשי, רביעוני, שנתי) ואת טווח התאריכים כדי לzechות האם בעיות שלוחה הן נקודתיות או כרוניות. נבחר תרשימים משולב בעל ציר קבוע מכיוון שהקשר בין הנקודות לשיעור ההצלחה הוא קריטי. הבחירה בפלטת הצבעים נעשתה בהתאם לעקרונות הסמנטיקה של הצבע. הצבע הכהול נבחר לייצוג משלוחים בזמן מכיוון שהוא מסדר יציבות ותקינות, בעוד הצבע הכתום נבחר לייצוג איחורים בשל בולטות גובהה והיותו צבע התרבות המושך את העין לבעה. עבורקו אחוז האיחורים נבחר כחול כהה לציר ניגודיות גובהה. מבחינת אינטראקטיבית, האפשרות לשנות את רמת הפירוט בין חודשי, רביעוני ושנתי מיישמת את עקרון "מבט על תחילת", והיכולת לסנן טווח תאריכים מאפשרת התמקדות בתקופות רלוונטיות.

## ויזואלייזציה רביעית (אינטראקטיבי) Heatmap



### ניתוח תיאורתי WWH

**What** - הנתונים הם רשומות של חזרנות עם חותמת זמן. שני משתנים קטגוריאליים-מחזוריים – יום בשבוע ושעה ביום, ומשתנה כמותי מספר ההזדמנויות או סף ההכנסות

**Why** - המשימה היא סיכום ויזהו דפוסים חוזרים על פני הזמן, לצד איתור חריגים כמו שעות עומס קיצונית

**How** - הקידוד הויזואלי נעשה באמצעות מיקום דו-ממדי במטריצה (יום בשבוע מול שעה ביום) ושימוש בערוץ הרויה של הצבע לייצוג עצמת הביקוש

### הסבר בחירות העיצוב והשאלת העסקית

השאלה העסקית שבה מתעסקת הויזואלייזציה היא מתי מתרחשシア הביקוש לצורך תכנון כוח אדם בשירות לקוחות וכדומה. כתוצאה מכך שנחגג להשתמש במפות חום טמפורליות כדי לחושף דפוסים מחזוריים של שעות וימים, כיוון שבגלות מתקשות להציג צפיפות נתונים גבוהה כזו בצורה אינטואיטיבית. עבור זיהוי זמני השיא בחרנו בפתח חום מכיוון שהוא מאפשר סריקה מהירה של כל שעות השבוע מבט אחד ומבליטה מיידית "אזורים חמים". בפתח החום נעשה שימוש בפלוטות צבעים סדרתיות (כגון צהוב-כתום-אדום) המאפשרות למשתמש לתפוא באופן אינטואיטיבי את המעבר מערכיים נמוכים לגבוהים.

## יזואלייזציה חמישית

### Propotional Symbol Map (אינטרקטיבי)

#### **Top Seller Leaderboard by State**

Map shows the top seller in each state after your filters.



### ניתוח תיאורתי WWW

**What** - הנתונים הם נתונים מרחביים הכוללים מפתח גיאוגרפי דו-ממדי (קו רוחב/אורך), משתנים קטגוריאליים (מדינה, מזאה מוכר) ומשתנים כמותיים הכנסות, כמות הזמן.

**Why** - המשימה היא איתור של מוכרים מוביילים לפי אזור, והשוויה של עצמת הפעולות העסקית בין טריטוריות גיאוגרפיות שונות

**How** - הקידוד היזואלי משתמש בנקודות הממוקמו לפיקוורדינטאות גיאוגרפיות, תוך שימוש בערז הגודל (size) לייצוג הממד הכמותי הנבחר(הכנסות וכמות הזמן).

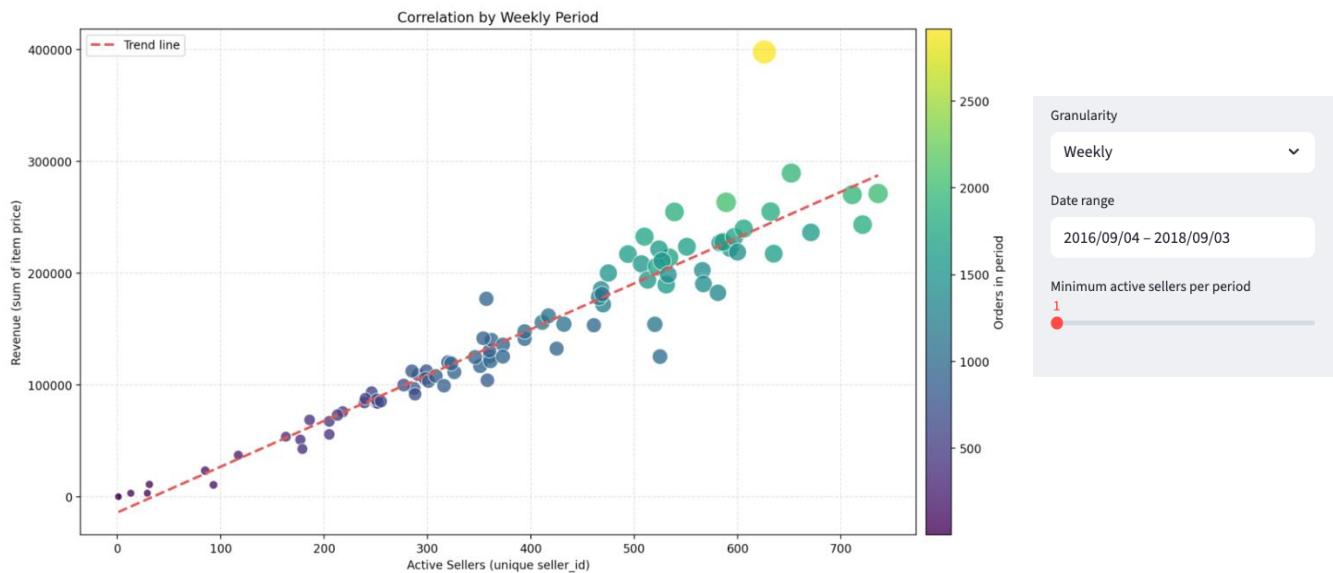
### הסבר בחירות העיצוב והשאלת העסקית

המפה עונה על השאלה "מי הם השחקנים הדומיננטיים בכל אזור גיאוגרפי ומהו היקף השימוש". ניתן לפלטר איזורים גיאוגרפיים, טווח תאריכים, הערך הכמותי שאוטו רוצים לראות על המפה (הכנסות, כמות מכירות, מספר מוכרים וכו') ומינימום הזמן להנחייה. הבחירה במפת סמלים ולא בטבלה רגילה נובעת מהחשיבות של ההקשר המרחבי – היכולת לראות האם ישנם ריכוזים של מוכרים חזקים באזוריים סמוכים (למשל בדרום מזרח ברזיל) לעומת אזוריים מבודדים.

הבחירה בצבע כחול שקופה עם קו מתאר לבן לנעה לשורת שתי מטרות: השקיפות מאפשרת לראות את מפת הרקע ומקרה על זיהוי חפיפות במידה וקיים, בעוד קו המתאר הלבן מבטיח הפרדה ברורה של האובייקט מהרקע. היכולת לשנות את הממד הקובע את גודל הבועה (למשל, מעבר מהכנסות למספר הזמן) מאפשרת למשתמש לבחון את המושג "הצלחה" מזוויות שונות.

## ויזואלייזציה ששית

### (אינטראקטיבי) Scatterplot



### כיתוח תיאורתי WWH

**What** - הנתונים הם נתונים מסוימים אגרגטיביים שבהם כל פריט מייצג יחידת זמן (שבוע/חודש), והמשתנים הם כמותיים בלבד: מספר המוכרים הפעילים, סך הרכנות ומספר ההזדמנויות.

**Why** - המשימה היא גילוי ואפיון של המתאם בין גודל היצע המוכרים לבין הצמיחה בהכנסות, במטרה לאישש את קיומו של "אפקט רשות" חיובי בפלטפורמה.

**How** - הקידוד הויזואלי משתמש בנקודות בתוך תרשימים פיזור, כאשר ערכי המיקום המרחבוי מייצגים את הקשר בין המשתנים, וערך הגודל והצבע משמשים לקידוד Redundant Encoding של נפח ההזדמנות להדגשת המשקל של כל תקופה.

### הסבר בחירות העיצוב והשאלה העסקית

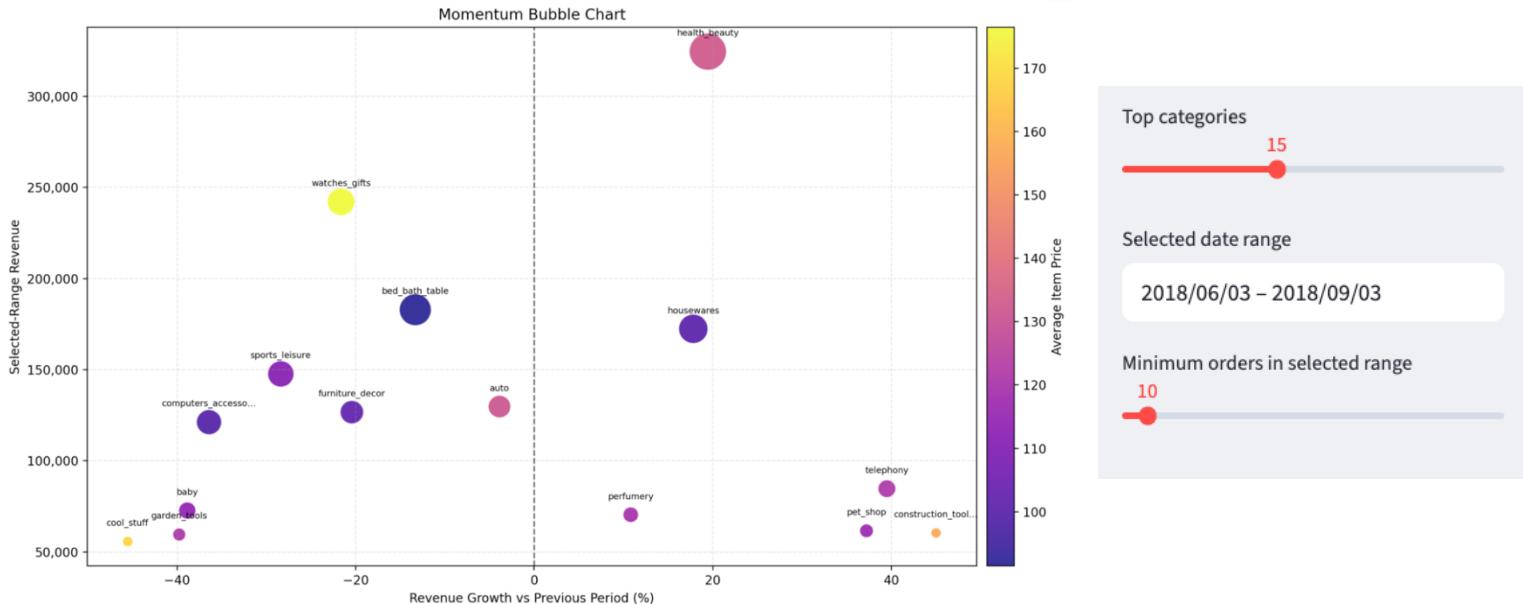
התרשימים עונה על השאלה "האם קיים קשר ישיר בין הגדלת היצע המוכרים לצמיחה בהכנסות הפלטפורמה. ניתן לפולטר את רצולניות הזמן (שבועי, חודשי, רביעוני ושנתי), טווח התאריכים, ואת מינימום המוכרים הפעילים כדי לנטרל רעשיות של תקופות הקמה. הבחירה בתרשימים פיזור מגיעה מהשיעורים שלמדונו על כך שగרפּ פיזור מתאים למציאות מתאמת בין המשתנים - היכולת לראות כיצד שינוי בציר ה-X-מספר מוכרים משפיע ישירות על ציר ה-Y-הכנסות

הבחירה בפלטת הצבעים ובשינוי גודל הבועה לייצוג כמות ההזדמנויות נועדה להוסיף מידע שלישי של "נפח פעילות", המאפשר לבדוק אם נקודות זמן נקודות דומות אך עצימות תפעולית שונה. הוספה זו המוגנה האדום המקווין מספקת עוגן ויזואלי מיידי לכיוון ועוצמת הקשר הסטטיסטי. يمكنك לשנות

את הגרנולריות (למשל משובע לחודשי) מאפשרת למשתמש "להחליק" תונדרתיות קצרת-טוווח ולהתמקד בSEGMENTATION מכך ארכוכות טווח.

## ויזואלייזציה שביעית

### Bubble Chart



### ניתוח תיאורתי WWH

**What** - הנתונים הם טבלה רב-ממדית שבה כל פריט מייצג קטגורית מוצר. המשטנים כוללים משתנה קטגוריאלי (שם הקטגוריה) ומשטנים כמוותים הדורשים חישוב על בסיס חלונות זמן: אחוז הצמיחה, סך ההכנסות, נפח ההזמנות ומהירות פריט ממוצע.

**Why** - המשימה היא גילוי ואפיון של קטגוריות המכניות ביותר ביוטר בעסק, איתור קטגוריות המציגות צמיחה או דעיכה קיצונית, והשואנה בין גודל הקטגוריה לקצב השינוי שלא ניהול פורטפוליו מוצרים.

**How** - הקידוד הויזואלי מתבצע באמצעות תרשימים בוועות. עורך המיקום משמשים למשטנים המרכזים (צמיחה והכנסות – ציר X,Y), עורך הגוד מקודד את נפח ההזמנות, ועורך הצבע משמש לקידוד מחיר ממוצע למכירה של הקטגוריה.

### הסבר בחירות העיצוב והשאלה העסקית

המטריצה עונה על השאלה האסטרטגית "אילו קטגוריות הן מנوعי הצמיחה העתידיים ואילו נמצאות בסיכון?". ניתן לפולטר את מספר הקטגוריות המוצגות, את טווח התאריכים להשוואה ואת מינימום הזמן נתון כדי לסנן קטגוריות נשא לא יציבות. הבחירה בתרשים בוועות, המהווה הרחבה של תרשימים פיזור, נובעת מה הצורך להציג ארבעה ממדיים של מידע בו-זמןית כדי לקבל תמונה הוליסטית של "בריאות הקטgorיה": האם היא גדולה ע, האם היא צומחת א, האם היא פופולרית (גודל), והאם היא יקרה (צבע).

הוסףתו קו הייחוס האנכי בערך 0 (צמיחה אפסית) היא קריטית לחלוקת המרחב לאזורים סמנטיים ברורים: צד ימין מייצג הגדמנות (צמיחה חיובית) וצד שמאל מייצג סיכון (דעיכה). הבחירה בפלטת הצבעים עברו מחיר קטgorיה המוצע מאפשרת לזהות האם הצמיחה בקטgorיה מסוימת מונעת על ידי מוצרים זולים ונגישים או על ידי מוצר פרימיום יקרים, תובנה שהיא חיונית לתכנון מלאי ומציעים.