

|||||

פרופ' גלעד רביד



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

דו"ח מסכם – ייזואלייזציה של המידע

נושא: חברת איקומרס Olist

מגישים:

אוהד אשכנזי 318321601

מלacci ויס 314969338

תוכן עניינים

תיאור הנתונים ואופן הטיפול	2
דין בויזואליזציות קיימות	3
ויזואליזציה ראשונה	4
ויזואליזציה שנייה	5
ויזואליזציה שלישית	6
ויזואליזציה רביעית	7
ויזואליזציה חמישית	8
ויזואליזציה ששית	9
ויזואליזציה שביעית	10

תיאור הנתונים ואופן הטיפול

הנתונים בפרויקט מבוססים על מאגר נתונים איקומרס של פלטפורמת Olist וככללים מידע רב-ממדית על הזמן, פרטיים, לקוחות, מוצרים, קטגוריות מוצרים, תשלום ומשלוחים. בסיס הנתונים בניי מספק טבלאות רלוונטיות המחויבות באמצעות מזהים ייחודיים (כגון מזאה הזמןה, מזאה ללקוח ומזהה מוצר), וכן שלב ראשון בתהליך כל מיזוג (join) בין הטבלאות לצורך יצירת טבלאות אנליטיות מאוחדות. לאחר האיחוד בוצע ניקוי נתונים שככל טיפול בערכים חסרים, הסרת כפליות, תיקון טיפוס משתנים (בעיקר שמות תאריך ושעה), והמרת חוות זמן לפורמטים אחידים המאפשרים ארגזיה לפי יום, שבוע, חודש ורבעון. בנוסף, חישבו מדדים נגזרים כגון סך הכנסות להזמנה, מחיר ממוצע לפריט, אחוז צמיחה בין תקופות, שיעור איחורים במשלוחים, מדדי ריכוזיות (N-Top), ומדד מתאם בין משתנים מרכזיים. במקרים בהם זהה ערכים קיצוניים חריגים (outliers), בוצע חיתוך סטטיסטי מבוקר (למשל Trim באחוזים 99%-1%) על מנת למנוע עיוות של התמונה הכוללת. כמו כן בוצעה ארגזיה ברמות שונות – קטגורית מוצר, תקופה זמן, אזור גיאוגרפי ומוכר – בהתאם לשאלת העסקית שנבדקה בכל יזואליזציה. תהליך ההכנה שם דגש על עקבות מדדים, שמירה על פרשנות עסקית נכונה, והבטחת יכולת אינטראקטיבית למשתמש, קיבוץ ושינוי גורנולריות בזמן אמת, כך שהנתונים יהיו לא רק נקיים ואמינים אלא גם מותאמים לניתוח חזותי תומך החלטה.

דין בייזואלייזציה קיימות לאותם נתונים



מתוך האתר Kaggle (מאיפה שלקחנו את הנתונים)

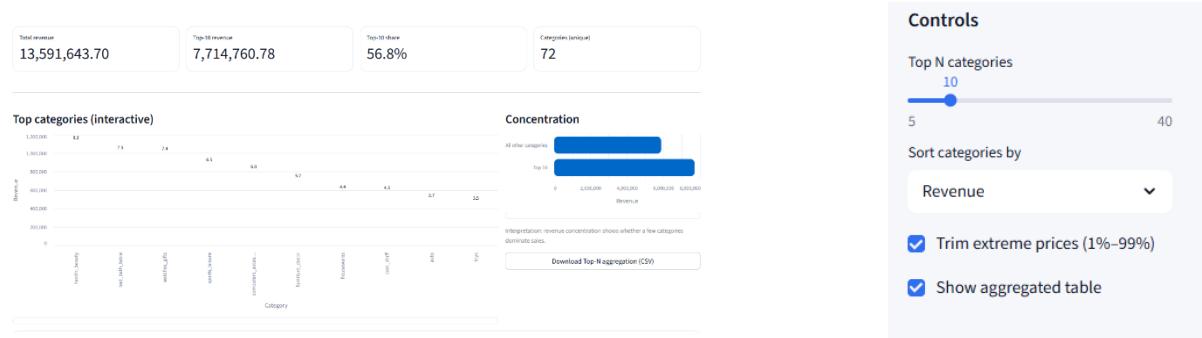
<https://www.kaggle.com/code/philipclemente/olist-e-commerce-analysis-and-dashboard/notebook>

התמקדות בייזואלייזיה השנייה והשוואה:

במסגר שլט הצגנו ויזואלייזיה מתקדמת ורב-מדנית אשר משלבת במקביל גם את נפח ההזמנות וגם את סך הרכנותות לאורך זמן, ובכך מאפשרת ניתוח דינמי של הקשר בין פעילות לתוצאות לביצועים פיננסיים. זאת בוגרוף לגרף הבסיסי המופיע ב-Kaggle, אשר מציג מכירות חודשיות בלבד בצורה סטטistica ולא העמיקה אנליטית. היזואלייזיה שלנו כוללת אינטראקטיביות מלאה – אפשרות לבעז Hover לקלט ערכים מדויקים, לסנן טווחי זמן, לשנות רוחניים תצוגה (שבועי/חודש/רבעוני), ולהשווות בין המדדים. בנוסף, באמצעות מבנה של Small Multiples ניתן לבחון כל מdad בסΚאלה טبعית ונפרד, תוך שמירה על אותו ציר זמן להשוואה אינטואיטיבית. גישה זו מאפשרת לא רק לראות "כמה נמכר", אלא להבין האם העלייה במכירות מלאה בעלייה פרופורציונלית בהרכנות, האם קיימות תקופות של שחיקת ערך להזמנה, והאם הדינמיקה מצביעה על שינוי מבני בפלטפורמה – תובנות שאינן ניתנות לylie'i בגרף בסיסי חד-ממד.

יזואליזציה ראשונה

Pareto Analysis



ניתוח תיאורתי WWW

What - הנתונים הם נתונים טבלאים ארגטטיביים בرمת קטגורית מוצר. כל שורה מייצגת קטgorיה 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) ומשתנים כמוותים רציפים.

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

How - הקידוד הויזואלי נעשה באמצעות טרשים עמודות לקידוד הכנסות (עורך אורך – מדוק לתיפסית), קו נוסף לקידוד המחיר הממוצע (ציר משני), ותוויות אוחזים להציג תרומות כל קטgorיה להכנסה הכוללת. בנוסף, מוצג טרשים ריכוזיות משלים המשווה בין Top-N קטגוריות לבין שאר הקטגוריות. האינטראקטיביות מאפשרת שליטה במספר הקטגוריות המציגות, שני קритריון המין, ווינון מחירם 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 מעודות נבעת מהעובדת שערוך האורך מאפשר השוואת בין קטגוריות בדידות. הוספה קו המציג הממוצע מאפשרת דיהוי 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 עיקנון **Overview+Detail**: המשמש רואה גם את הערך האבסולוטי וגם את המשמעות היחסית שלו.

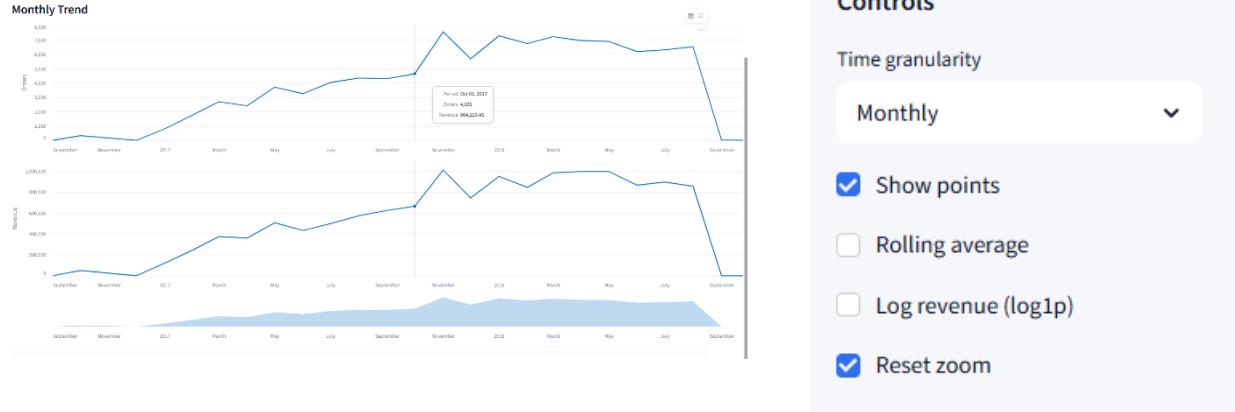
טרשים הריכוזיות המשלים N-Top (מול היותר) מאפשר לזרות האם הפלטפורמה ושונעת על מספר מצומצם של מנועי צמיחה – תובנה קרייטית לניהול סיכון עסקים ולגיון פורטפולוי.

היכולת לשולט במספר הקטגוריות (N-Top) ממשתת את עיקנון "מבט על תחילת, ולאחר מכן זום", בכך שהיא מאפשרת לבחון את המערכת ברמת מקור ולאחר מכן להתמקד בקבוצות ספציפיות.

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

יזואלייזציה 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 "חודיות וסך הכנסות המבאות מפרט הזמן. מדובר בשני משתנים רציפים בעלי סקלנות שונות, אך מבוססים על אותו ציר זמן.

- **Why** - המטרה האנלטיטית היא זיהוי מגמות לארוך זמן, איתור תקופות האטה בפועלות, ובוחינת הקשר הדינמי בין נפח 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 עלייה 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:
יכיז מפתוחות הזרנות והכנסות לארוך זמן, והאם קיימם קשר עקבי בינהו?

הבחירה ב-Small Multiples במקומם ציר כפול נובעת מעקרונות ייזואלייזציה אקדמיים הממליצים להימנע מציר Y כפولي אשר קיימים פער סדרי גודל ממשמעותיים. הפרדה זו מפחיתה עומס קוגניטיבי, משמרת קריאות כמותית מלאה, ומאפשרת השוואה אינטואיטיבית לארוך אותו ציר זמן.

Overview first, zoom and filter, details on demand: הוספה Brush בתחרית הגרף מיישמת את עקרון: המستخدم רואה את התמונה הכללית, בוחר טווח רלוונטי, ועמוק בניתוח מבלי לעבור דף או לשנות גרף.

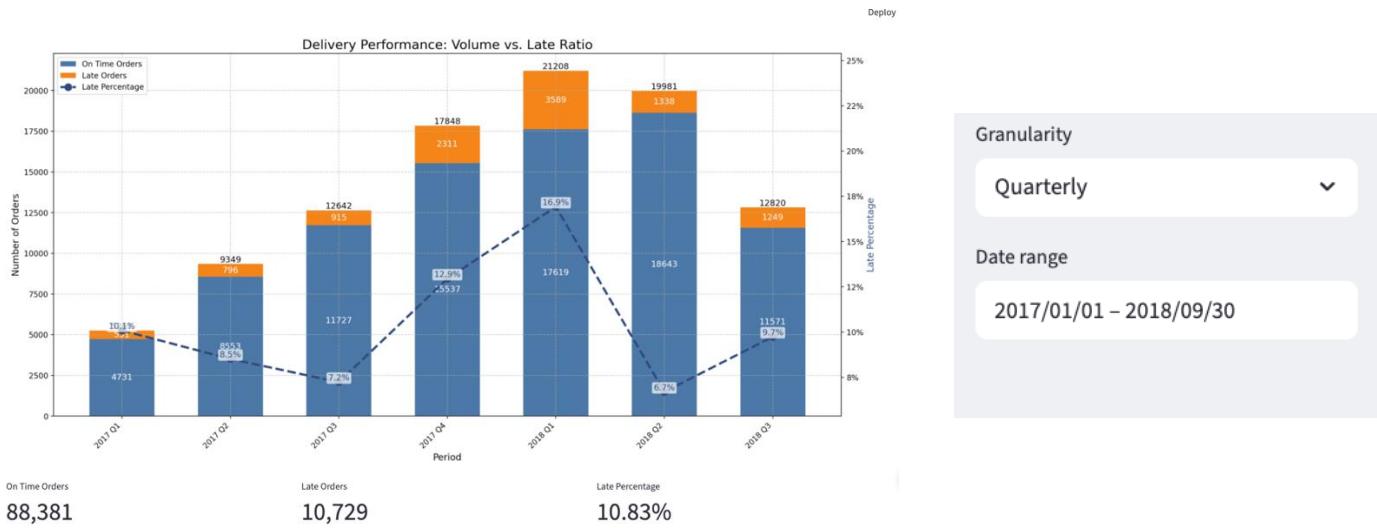
- Hover מאפשר דיקט מספרי נקודתי ללא עומס טקסטואלי, קבוע על הציר, וכך נשמר איזון בין מינימליזם עיצובי לבין דיקט אנליטי.

הוספה מוצעת נוע מאפשרת להבדיל בין תנודות מוקנית לבין מגמה מבנית, בהתאם לעקרונות ניתוח סדרות זמן. האפשרות לסקללה לוגריתמית רלוונטית במצבים של צמיחה אקספוננציאלית, ומונעת דחיסה של תקופות מוקדמות.

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

יזואלייזציה של שלישיית

Stacked Bar Chart



Granularity

Quarterly

Date range

2017/01/01 – 2018/09/30

ניתוח תיאורתי WWW

What - הנתונים הם נתונים טבלאיים הכוללים משתנה זמן כמפתח ראשי, ומשתנים של סטטוטו המשלוח וכמות ההזמנות.

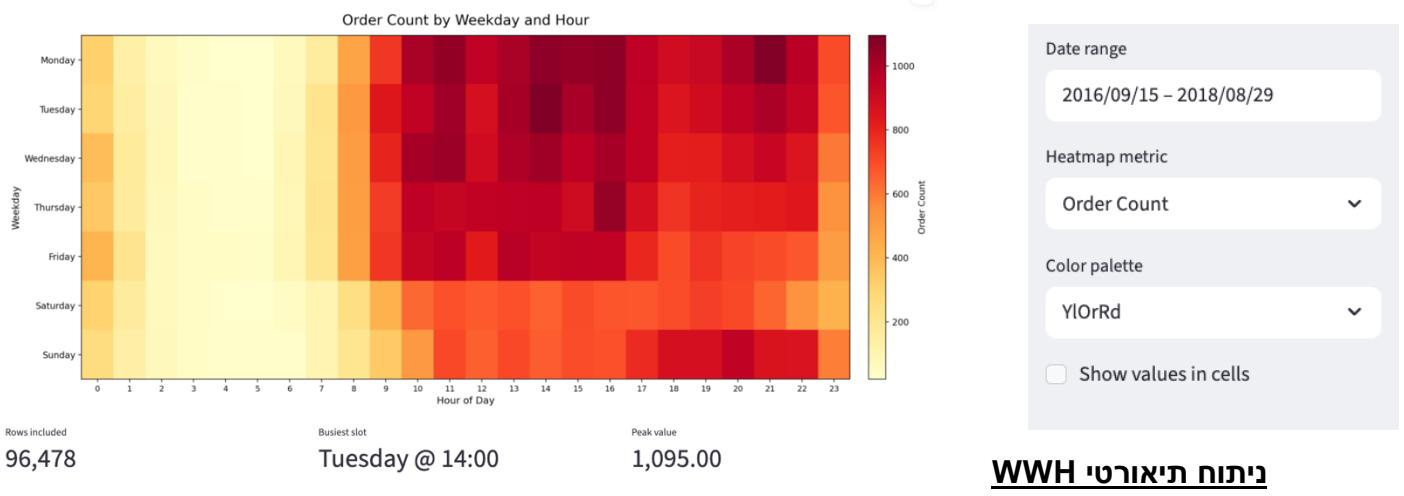
Why - המשימה המרכזית היא דיהוי מגמות עונתיות ביצועי השילוח, איתור תקופות חריגות בהן אחוז האיחורים זינק בצורה קיצונית, והשוואה בין נפח הפעולות לבין איכות הביצוע.

How - הקידוד הויזואלי נעשה באמצעות שימוש בקיים עבור המגמה ובשיטה عمودות עבור הכמהות. ערוץ האורק משמש לקידוד כמות ההזמנות מכיוון שהוא מאפשר תפיסת מדיקת יחסית, בעוד ערוץ הצבע משמש להפרדה בין הקטגוריות. בנוסף, האינטראקטיביות מאפשרת מניפולציה של הנתונים באמצעות סינון וקבוץ לשינוי רצולוציית הזמן.

הסבר בחירות העיצוב והשאלה העסקית

הgraf שונה על השאלה האם הגידול בנפח ההזמנות פוגע באיכות השירות הלוגיסטי. ניתן לפטלר את רצולוציית 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 הכמהות לשיעור ההצלחה הוא קריטי. הבחירה בפלטת הצבעים נעשתה בהתאם לעקרונות הסמנטיקה של הצבע. הצבע הכלול נבחר לייצוג משלוחים בזמן מכיוון שהוא מסדר יציבות ותקינות, בעוד הצבע הכתום נבחר לייצוג איחורים בשל בולטות גובהה והיותו צבע התראה המושך את העין לבעה. עברו קו אחוז האיחורים נבחר כחול כהה לציר ניגודיות גובהה. מבחינת אינטראקטיבית, האפשרות לשנתת את רמת הפירוט בין חודשי, רביעוני ושנתי מיישמת את עקרון "מבט על תחילת", והיכולת לסנן טווח תאריכים מאפשרת התקפות רלוונטיות

ויזואלייזציה רביעית (אינטרקטיבי) Heatmap



ניתוח תיאורתי WWH

What - הנתונים הם רשומות של הזמנות עם חותמת זמן. שני משתנים קטגוריאליים-מחזוריים – יום בשבוע ושעה ביממה, ומשתנה כמותי מספר ההזמנות 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 ביממה) ושימוש בערוץ הרויה של הצבע לייצוג עצמת הביקוש

הסבר בחירות העיצוב והשאלה העסקית

השאלה העסקית שבה מתעסקת הויזואלייזציה היא מתי מתרחש שייא הביקוש לצורך תכנון כוח אדם בשירות לקוחות וכדומה. כתוצאה ממדון שנהוג להשתמש במפות חום טופורוליות כדי לחשוף דפוסים מחזוריים של שעות ימים, ניתן שטבלאות רגילים מתקשות להציג צפיפות נתונים גבוהה כזו בצורה אינטואיטיבית. עבור זהוי זמני השיא בחרכן במפת חום מכיוון שהוא מאפשר סריקה מהירה של כל שעות השבוע מבט אחד וambilיה מיידית "אזורים חמימים". במפת החום נעשה שימוש בפלוטות צבעים סדרתיות (כגון צהוב-כתום-אדום) המאפשרות למשתמש לתפוא באופן אינטואיטיבי את המעבר מערכים נמוכים לגבוהים.

יזואליזציה חמישית

(אינטרקטיבי) Proportional Symbol Map

Top Seller Leaderboard by State

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



ניתוח תיאורתי WWH

What - הנתונים הם נתונים מרחביים הכוללים מפתח גיאוגרפי דו-ממדי (קו רוחב/אורך), משתנים קטגוריאליים (מדינה, מזאה מוכר) ומשתנים כמוותים 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 הממוקמוں לפי קוordinטות גיאוגרפיות, תוך שימוש בערך הגדול (size) לייצוג הממד הכמותי הנבחר(הכנסות וכמות הזמן).

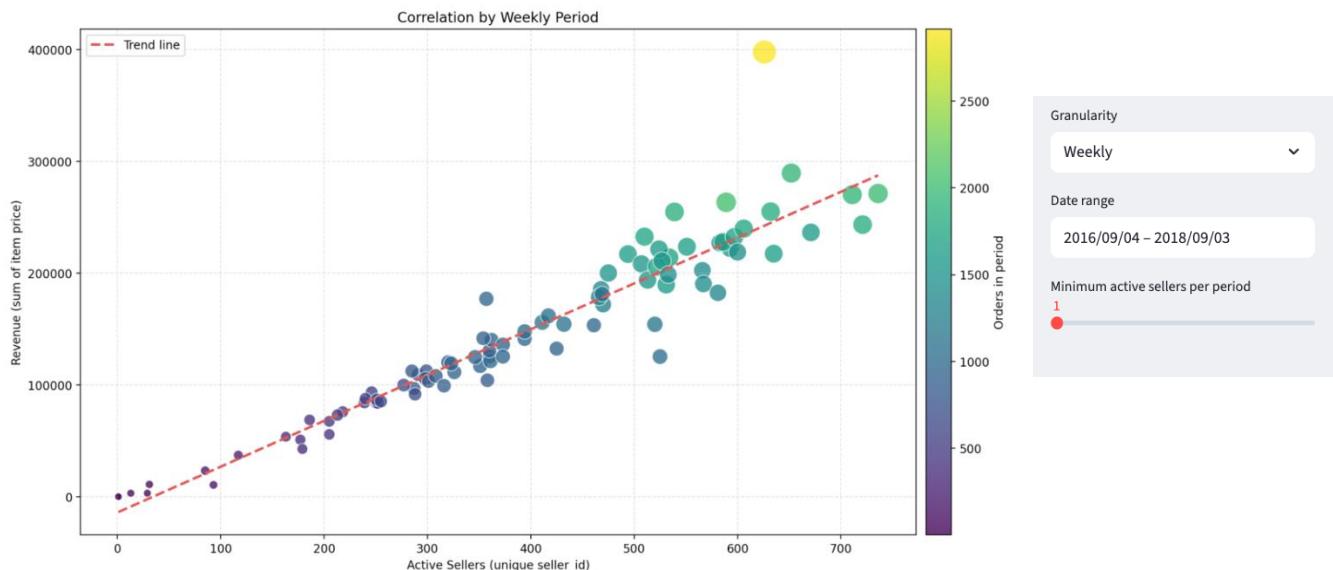
הסבר בחירות העיצוב והשאלת העסקית

המפה עונה על השאלה "מי הם השחקנים הדומיננטיים בכל אזור גיאוגרפי ומהו היקף השימוש". ניתן לפטר איזורים גיאוגרפיים, טווח תאריכים, הערך הכמותי שאוטו רוצים לראות על המפה (הכנסות, כמות מכירות, מספר מוכרים וכו') ומינימום הזמן להנעה למוכר. הבחירה במפת סמלים ולא בטבלה רגילה מבעת 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 נועדה לשרת שתי מטרות: השקיפות מאפשרת לראות את מפת הרקע ומקרה על זיהוי חפיפות במידה וקייםות, בעוד קו המתאר הלבן מבטיח הפרדה בhorra של האובייקט מהרקע. היכולת לשנות את הממד הקובע את גודל הבועה (למשל, מעבר מהכנסות למספר הזמן) מאפשרת למשתמש לבחון את המושג "הצלחה" מזוויות שונות.

ויזואליזציה ששית

Scatterplot (אינטרקטיבי)



ניתוח תיאורטי WWH

What - הנתונים הם נתונים כמותיים ארגטטיביים שבهم כל פריט מייצג יחידת זמן (שבוע/חודש), והמשתנים הם כמותיים בלבד: מספר המוכרים הפעילים, סך ההכנסות ונפח ההזמנות.

Why - המשימה היא גילוי ואפיון של המתאים בין גודל היצוא המוכרים לבין הצמיחה בהכנסות, במטרה לאשש את קיומו של "אפקט רשות" חיובי בפלטפורמה.

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

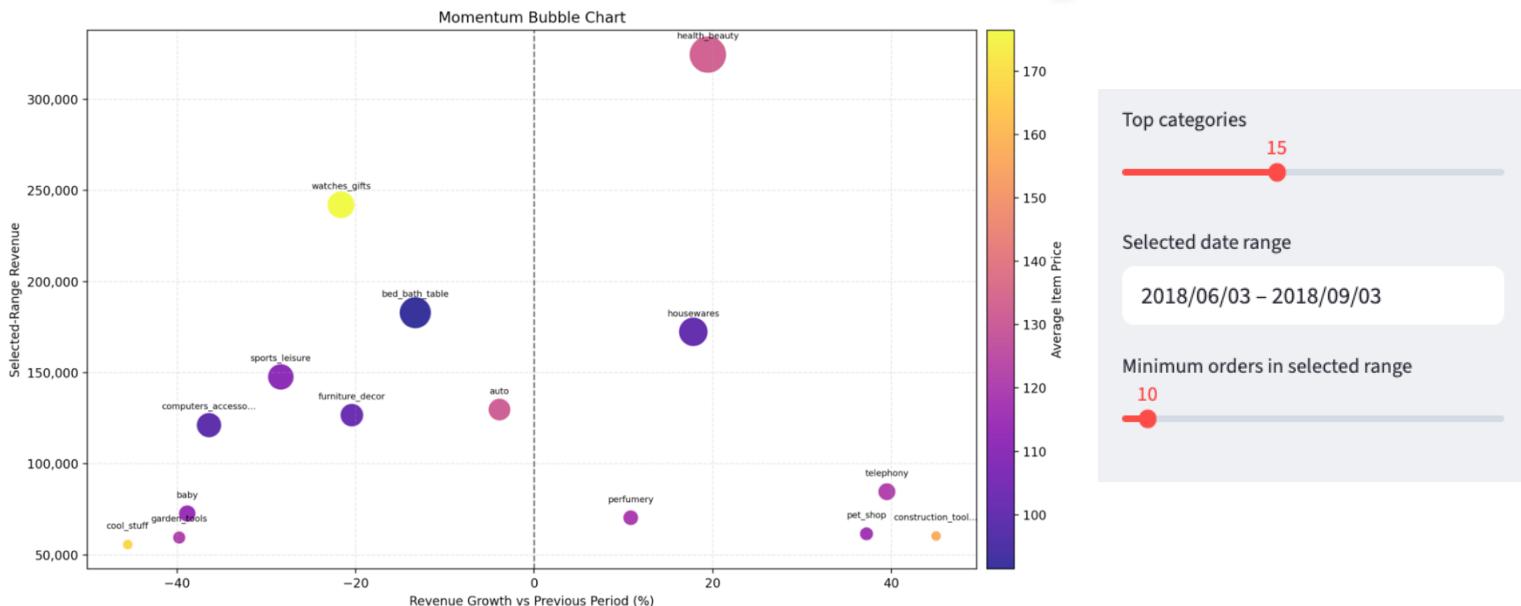
הסבר בחירות העיצוב והשאלת העסקית

התרשימים עונה על השאלה "האם קיים קשר ישיר בין הגדלת היצוא המוכרים לצמיחה בהכנסות הפלטפורמה. ניתן לפולטר את רצולוציית הזמן (שבועי, חודשי, רביעוני ושנתי), טווח התאריכים, ואת מינימום המוכרים הפעילים כדי לנטרל רעשם של תקופות הקמה. הבחירה בתרשימים פיזור מגיעה מהשיעורים שלמדו על כך שגרף פיזור מתאים למציאת מתאים בין המשתנים - היכולת לראות כיצד שינוי במצב ה-X-מוספר מוכרים משפיע ישירות על מצב ה-Y-הכנסות

הבחירה בפלטת הצבעים ובשינוי גודל הבועה לציג כמות ההזמנות נועדה להוסיף מידע שלישי של "נפח פעילות", המאפשר לבדוק בין נקודות זמן דומות אך עצימות פעולה שונה. הוספה זו המוגדרת האדם המקווין מספקת עוגן ויזואלי מיידי לכיוון ועוצמת הקשר הסטטיסטי. היכולת לשנות את הגרנולריות (למשל משבועי לחודשי) מאפשרת למשתמש "להחליק" תנודות קצרת-טווח ולהתמקד במוגמות מוארכות יותר.

ויזואלייזציה שביעית

Bubble Chart



Selected range: 2018-06-03 to 2018-09-03 | Compared against: 2018-03-02 to 2018-06-02 (same number of days, immediately before selected range).

ניתוח תיאורתי WWH

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

Why - המשימה היא גילוי ואפיון של קטגוריות המכניות ביוטר בעסק, איתור קטגוריות חדשות צמיחה או דעיכה קיצונית, והשואה בין גודל הקטגוריה לקצב השינוי שלא לצורך ניהול פורטפולוי מוצרים.

How - הקידוד הויזואלי מ被执行 באמצעות תרשימים בועות. ערכוי המיקום משמשים למשתנים המרכזים (זמןיה והכנסות – ציר x,y), ערוץ הגוד מקודד את נפח הזמן, וערוץ הצבע משמש לקידוד מחיר ממוצע לממכריה של הקטגוריה.

הסבר בחירות העיצוב והשאלה העסקית

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

הוספת קו הייחוס האנכי בערך 0 (צמיחה אפסית) היא קריטית לחלוקת המרחב לאזורי סמנטים בחרומים: צד ימין מיצג הגדמות (צמיחה חיובית) וצד שמאל מיצג סיכון (דעיכה). הבחירה בפפלטת

הצבעים עבור מחיר קטגוריה המוצע מאפשרות לזהות האם הצמיחה בקטגוריה מסוימת מנעט על ידי מוצרים זולים ונגיסים או על ידי מוצרי פריימים יקרים, תובנה שהיא חיונית לתכנון מלאי ומציעים.