

תקציר

מסמך הבנת הנתונים מתאר את הניתוח הראשוני שביצענו בפרויקט הגמר, תוך הביקורות בזיהוי עמודות מעניינות, מגמות אפשריות, והשערות על בסיס תצוגה ראשונית של הנתונים שהשגנו. יש להדגיש שכל המסקנות המוצגות, לרבות זיהוי עמודות חשובות ופחות חשובות, הן בגדר השערה בלבד ונשענות על התרשמות אנושית. ייתכן שבשלבי למידת המכונה והפעלת האלגוריתם נמצא תובנות שונות אשר ישענו דווקא על עמודות אחרות לחלוטין. מסמך הבנת הנתונים קריטי לתהליך החקר, זיהוי אתגרים מקדים וביסוס המטרות להמשך פרויקט הגמר.



דוח הבנת הנתונים

השקעות בסיכון נמוך זה אצלנו-NexTrade

מגישים – אריאל קריב ומשי בר

תוכן עניינים

2	1. איסוף נתונים
2	1.1 מקורות נתונים
3	1.2 בדיקה ראשונית
5	2. תיאור הנתונים
5	2.1 כמות הנתונים
5	2.2 סוגי ערכים
6	2.3 סכמות קידוד
7	3. חקר נתונים
13	4. איכות הנתונים

1. איסוף נתונים

1.1 מקורות נתונים

בפרויקט הגמר שלנו נבצע שימוש בשני מקורות נתונים עיקריים המספקים נתונים עדכניים והיסטוריים לצרכי הניתוח המתקדם ולמידת המכונה.

1. נתונים קיימים:

- Alpha Vantage API :

- באמצעות חשבון חינמי, הצלחנו לקבל נתונים של מדד ה-S&P 500, הכוללים 100 ימים אחרונים בלבד של מחירי סגירה יומיים. כמו כן, במסגרת החשבון החינמי הצלחנו לקבל 50 כתבות חדשות כלכליות רלוונטיות בהתאם לנושאים שבחרנו בלבד.

- מגבלת כמות הבקשות לחשבון החינמי מאלצת אותנו להסתפק בנתונים חלקיים לצורך שלבי המחקר הראשוניים.

- בשלבים הבאים של הפרויקט, נרכוש מנוי פרימיום ל-Alpha Vantage שיאפשר גישה נרחבת לנתונים, הן עבור מדד ה-S&P 500 לאורך מספר שנים, והן עבור חדשות כלכליות בקנה מידה רחב יותר לשם הצלבת הנתונים. את מאגרי המידע נסנכרן עלפי תאריכים ובכך נבצע חקר מעמיק להפקת תובנות באמצעות למידת המכונה ואלגוריתמים מתאימים.

- Yahoo Finance (YFinance) :

- YFinance סיפק לנו נתוני מחירים יומיים של מדד ה-S&P 500 בלבד עבור 5 השנים האחרונות, בין התאריכים 1 בינואר 2019 ועד 19 בינואר 2025.

- מקור זה שימש אותנו להמחשה בלבד והבנה מעמיקה של הנתונים על פני תקופת זמן רחבה יותר, כיוון ש-Alpha Vantage לא סיפק נתונים ברמת פריסה זו במגבלות החשבון החינמי.

2. נתונים שנרכשו:

נכון לשלב הבנת הנתונים לא ביצענו רכישת נתונים. בשלב יותר מתקדם נבצע רכישה למנוי הפרמיום של-Alpha Vantage לצורך הרחבת הנתונים על פני שנים רבות הן עבור תחזיות המדד והן עבור נתוני חדשות כלכליות ללא הגבלה.

3. נתונים נוספים:

בשלב הבנת הנתונים לא מצאנו צורך בנתונים נוספים ממקורות חיצוניים. מקורות הנתונים הקיימים מספקים בסיס מספק לצורך התחלת הניתוחים, ועם רכישת המנוי המורחב, הנתונים יספקו כיסוי רחב ועמוק יותר לצורכי הפרויקט.

סיכום:

המגבלות הקיימות בחשבון החינמי מאלצות אותנו להשתמש במקורות משלימים כמו YFinance לצורך המחשה ראשונית. עם זאת, בתכנון הפרויקט העתידי, רכישת המנוי ל-Alpha Vantage תאפשר גישה לנתונים מקיפים יותר ולניתוח מעמיק ומדויק.

1.2 בדיקה ראשונית

העמודות המבטיחות ביותר-

מהנתונים שהתקבלו מ- Yahoo Finance על נתוני המדד S&P 500, העמודות Volume ו-Close נראות כמבטיחות ביותר מכיוון שהן מספקות מידע מרכזי על מחירי הסגירה, הנפח והתיקונים במחירים שעשויים להשפיע על הניתוחים. בנוסף עמודות של תוצאות האינדיקטורים של MACD ו-RSI הם חיוניים לניתוח טכני של מגמות השוק.

בנוגע לנתוני החדשות הכלכליות, הסקנו שהעמודות המבטיחות ביותר הן:

"ticker_sentiment", "overall_sentiment_label", "overall_sentiment_score", "topics", "category_within_source", "time_published". לפי עמודות אלו נראה כי נוכל למקד את האלגוריתם במה שחשוב אם זה כיוון השוק בהינתן החדשה או אף לברר את הרלוונטיות של אותה חדשה.

עמודות לא רלוונטיות-

נכון לפרק זה מ-YFinance אנו בחרנו את העמודות שאנו סבורים שיהיו רלוונטיות. מנגד בנתוני החדשות סבורים שהעמודות הלא רלוונטיות לנו הן: "summary", "authors", "URL", "source", "banner_image".

האם הנתונים מספיקים-

בשלב מתקדם של הפרויקט אנו נרכוש את המנוי של Alpha Vantage. עם זאת לפרק הבנת הנתונים קיבלנו מידע היסטורי של 5 שנים עבור נתוני S&P 500 מ-Yahoo Finance. דבר אשר הקצה לנו מספיק נתונים כדי לבצע ניתוחים ראשוניים בלבד ולשער מגמות עם דיוק יחסי.

הספקת נתונים לחיזויים-

עם התחשבות בנתונים שביצענו בהם שימוש משני המקורות, מצאנו כי המידע מספיק לבניית תחזיות ראשוניות. עם רכישת מנוי הפרמיום ל-Alpha Vantage מודל ה-LSTM יוכל לנתח נתונים על המדד ולחזות מגמות קצרות טווח, בעוד ש-BERT ישתמש בחדשות להבנת השפעתם הרגשית והכלכלית.

מיזוג מקורות נתונים: נכון לחלק זה של הפרויקט אנו מבצעים מיזוג נתונים משני מקורות. את הנתונים אנו קיבלנו כקובץ CSV, ואת ההתאמה בניהם לצורך המחשה ראשונית ובסיסית ביותר ביצענו באמצעות עמודה משותפת לשתייהן של Date. חשוב לציין כי בחלקים מתקדמים של בפרויקט נבצע שימוש מ-Alpha Vantage בלבד ולכן לא יתבצע להבא מיזוג מקורות נתונים מאתרים שונים, אלא מיזוג של נתונים של המדד והחדשות הכלכליות מאתר רשמי אחד בלבד.

ערכים חסרים: בקבצי הנתונים שקיבלנו אכן נמצאו ערכים חסרים ממגוון רחב של סיבות. לדוגמא- חוסר בנתוני עבר לצורך חישובים, חוסר של שם של מחבר כתבה ועוד. בסעיף 4 של המסמך נעניק מענה רחב על איך נתמודד עם הערכים החסרים שקיימים בנתונים שלנו בתהליך הכנת הנתונים שעתיד לבוא.

2. תיאור הנתונים

2.1 כמות הנתונים

- ייבוא הנתונים מ-YFinance על מדד ה-S&P 500 לחמש שנים האחרונות מכיל 1,524 שורות של נתונים, כאשר שורה מייצגת יום נפרד בעל – 8 עמודות .
- מאתר Alpha Vantage ייבאנו את נתוני החדשות. מפאת חוסר שימוש במנוי של האתר אנו יכלנו לייבא אך ורק חדשות מהחודש האחרון שהצטברו להיות כ-260 שורות וכ- 13 עמודות של חדשות לפי המסננים שהפעלנו. מסננים אלו הביאו לנו את הכותרות הרצויות ביותר למען השגת יעדי הפרויקט, הכותרות הן :
 - "Core CPI"
 - "Average Hourly Earnings"
 - "S&P Global Services PMI"
 - "ISM Manufacturing PMI"
 - "CPI"
 - "GDP"
 - "Fed Interest Rate Decision"
 - "Core PCE Price Index"
 - "ISM Manufacturing Prices"
 - "Unemployment Rate"
 - "ECB Interest Rate Decision"
 - "CB Consumer Confidence"
 - "JOLTS Job Openings"
 - "Crude Oil Inventories"

כותרות אלו נבחרו באופן מדויק מכיוון שהן התיאור לחדשות הכלכליות הקבועות והרלוונטיות ממקורות מסחר שמסווג כל חדשה לפי החשיבות שלה על השוק (לדוגמא אתר investing).

2.2 סוגי ערכים

הערכים מגיעים כפורמטים שונים :

הנתונים שייבאנו מ-YFinance על מדד ה-S&P 500 הם נתונים מספריים בלבד, לעומת ייבוא נתוני החדשות מהאתר Alpha Vantage המורכב גם ממחרוזות וגם מנתונים מספריים. רוב נתוני החדשות הן מחרוזות , אך העמודה "overall_sentiment_score" שמתארת את הציון של אותה מנייה לאחר החדשה מצוינת כמספר. חשוב לציין כי בשני מקורות הנתונים קיימת עמודת התאריך (date) שמהווה לנו בסיס להתאמת זמן בין שני סוגי הנתונים.

2.3 סכמות קידוד

הנתונים שייבאנו מ-YFinance על מדד ה-S&P 500 הם רק נתונים מספריים ואינם קטגוריאליים, אך נתוני החדשות שייבאנו מאתר Alpha Vantage מתחלקים לעמודות מפורטות וגם לעמודות אשר מתארות קטגוריות.

העמודות הקטגוריאליות הן:

"category_within_source" - עמודה זו מתארת את סוג ההתייחסות של אותה חדשה לכמה קטגוריות: Trading, Business, Markets, General, Mergers, News. הקידוד בעמודה זו חשובה להבנת הקשר לניתוח השפעות שונות של האירועים על המדד S&P 500.

"overall_sentiment_labe" - עמודה זו מתארת את ההשפעה של החדשה על שוק המניות בידי סיווג העמודה לפי "השפעת מגמה". משמע, ערכים בעמודה זו נעים מביטויי רגש ברורים ועד לטונים יותר מופשטים, וכוללים את התיוגים הבאים:

- **"Bullish"** - מייצג רגש חיובי או תחזית שמעלה באופטימיות לגבי השוק. זהו טון שעשוי להצביע על ציפייה לעלייה בשווי הנכסים או המדדים המדוברים.
- **"Somewhat-Bullish"** - מעט פחות חיובי מ-Bullish אך עדיין מתאר גישה אופטימית כלפי השוק, אך עם פחות בטחון.
- **"Neutral"** - טון נייטרלי שאינו מצביע במפורש על ציפיות חיוביות או שליליות, אלא מתאר עובדות.
- **"Somewhat-Bearish"** - טון שמעיד על קצת פסימיות או חשש כלפי השוק או הנכס הספציפי, אך לא באופן חריף.
- **"Bearish"** - טון שלילי במיוחד שמצביע על ציפיות לירידה בשווי הנכסים או בביצועי השוק. מייצג פסימיות רבה לגבי התפתחויות עתידיות.

השימוש בתיוגים אלו במודל החיזוי מאפשר לנו להטמיע נתונים רגשיים ותחזיות מוטעות רגש בתוך התחזיות הכלכליות והטכניות. זה יכול להוביל לתובנות עשירות יותר על איך דעת הקהל או הפרספקטיבה הכלכלית משפיעה על מחירי שוק.

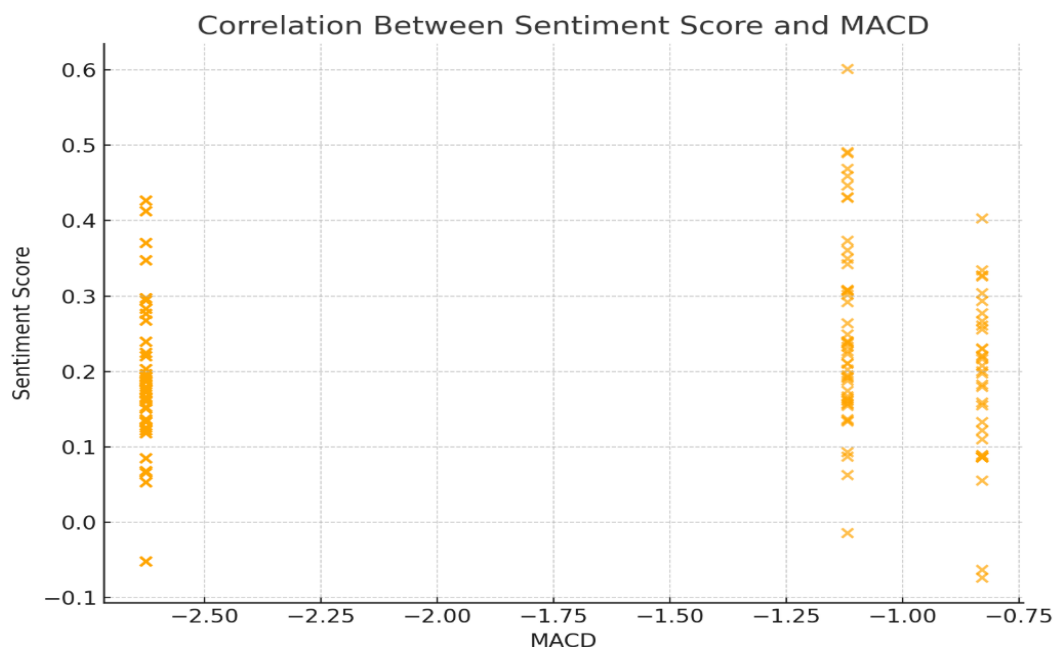
3. חקר נתונים

הצגת הממצאים והגרפים

הניתוחים והגרפים המוצגים מבוססים על חלק קטן מהנתונים הכוללים בפרויקט. מטרת הגרפים היא להמחיש מגמות ראשוניות ולהעלות השערות, אך אין לראות בהם תשובות סופיות או מדויקות בשום שלב. נתונים נוספים עשויים לשנות לחלוטין את המסקנות הנוכחיות.

הגרפים המוצגים כוללים הסברים מפורטים על המבנה שלהם, העמודות שמשתמשים בהם, והמשמעויות שמסתתרות מאחורי הנתונים.

גרף 1: התאמה בין ניקוד רגשות ל-MACD-



הסבר גרף 1-

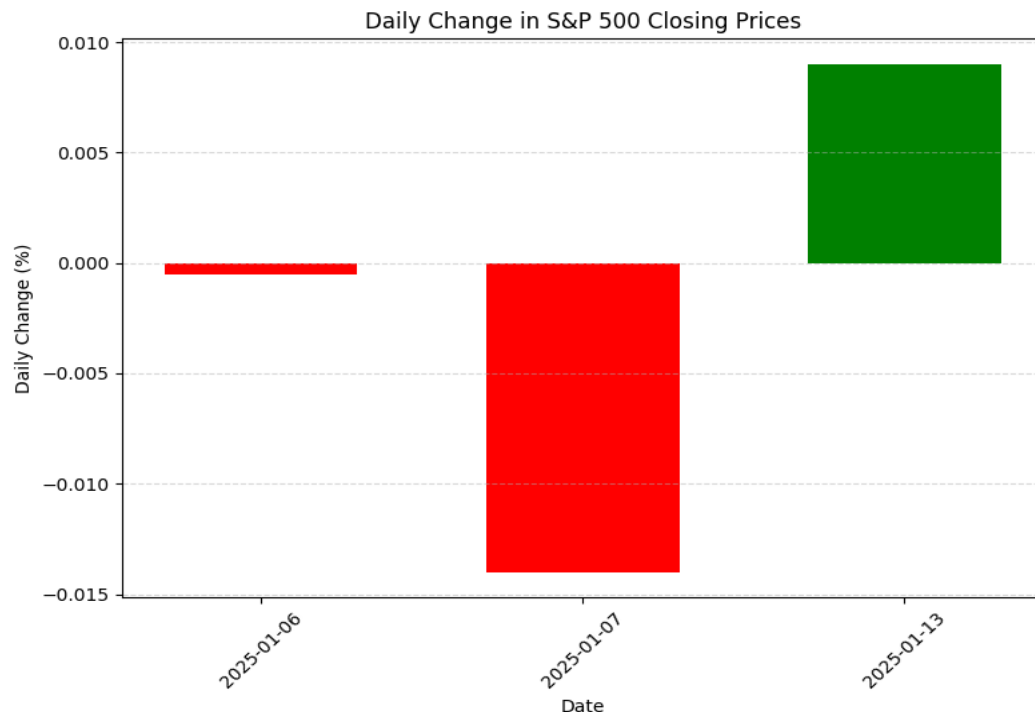
- גרף פיזור המציג את הקשר בין ניקוד רגשות (Sentiment Score) לעמודת MACD אינדיקטור טכני המשמש לזיהוי מגמות בשוק.
- **מסקנות מהגרף:** עלפי הגרף שקיבלנו, נראה שאין קשר ברור או לינארי בין הרגשות שנובעות מהחדשות הכלכליות לתוצאות חישוב הממוצעים ב-MACD. הנקודות מפוזרות בצורה לא אחידה, עם מקבצים ברמות שונות. יש לציין כי תחילה שיערנו שדווקא כן יהיה קשר בין הרגש החיובי או השלילי של כל חדשה לתוצאות הממוצעים הנעים של MACD.

הסבר על עמודות מתוך הנתונים :

- **overall_sentiment_score** - עמודה זו מספקת לנו מדד מספרי שמאפשר ניתוח רגשות של כתבות כלכליות.
 - סקאלת הציון של overall_sentiment_score נעה בין ערכים שליליים לערכים חיוביים, כאשר :
 - **ערכים שליליים ($0 >$)** - מצביעים על טון שלילי של הכתבה. לדוגמה, אם כתבה מדברת על משבר כלכלי או הפסדים בשוק, הציון יהיה שלילי.
 - **ערך 0 (ניטרלי)** - מצביע על חדשות ללא הטיה רגשית. כלומר, כתבה המתארת עובדות יבשות ללא רגש (כמו "הבנק המרכזי העלה את הריבית ב-0.25%").
 - **ערכים חיוביים ($0 <$)** - מצביעים על טון חיובי בכתבה. לדוגמה, אם כתבה מדברת על צמיחה בשוק או עליות במדד המניות, הציון יהיה חיובי.
- **MACD (Moving Average Convergence Divergence)** - אינדיקטור טכני פופולרי המשמש לניתוח מגמות שוק ומבוסס על הבדל בין שני ממוצעים נעים.
 - **"EMA" קצר-טווח ו "EMA" ארוך-טווח.**
 - **מסייע בזיהוי מגמות עלייה או ירידה בשוק.**
 - **סקאלת MACD יכולה לקבל ערכים חיוביים ושליליים, כאשר :**
 - **ערכים חיוביים ($MACD > 0$)** - מעידים על מומנטום חיובי. מצב זה נתפס כאות "Bullish" ועשוי לרמוז על מגמת עלייה במחירי הנכסים.
 - **ערך 0 ($MACD = 0$)** - מציין שהממוצעים הנעים נפגשו, כלומר אין מגמת יתרון ברורה כלפי מעלה או מטה.
 - **ערכים שליליים ($MACD < 0$)** - מעידים על מומנטום שלילי, כלומר שהממוצע הקצר נמצא מתחת לממוצע הנע הארוך. מצב זה נתפס כאות "Bearish" ועשוי להצביע על מגמת ירידה במחירי הנכסים.
- **השערות מעודכנות :**
 - **השפעה של רגשות קיצוניים:** ייתכן שרגשות קיצוניים (חיוביים מאוד או שליליים מאוד) משפיעים בצורה חזקה על MACD. לדוגמה, חדשות "Bullish" מאוד עשויות להוביל לערכים חיוביים משמעותיים של MACD.
 - **הקשר תלוי הקשר נוסף:** MACD לא מושפע ישירות מהרגשות, אלא בשילוב עם גורמים כמו נפח מסחר גבוה או קטגוריות High-Low.

○ **אזורי רגישות:** ייתכן ש-MACD רגיש רק במצבים של תנודתיות קיצונית בשוק.

גרף 2: שינוי יומי במחירי סגירה-



הסבר גרף 2- עמודות המציינות את השינוי היומי במחירי הסגירה של מדד S&P 500.

עמודות מתוך הנתונים:

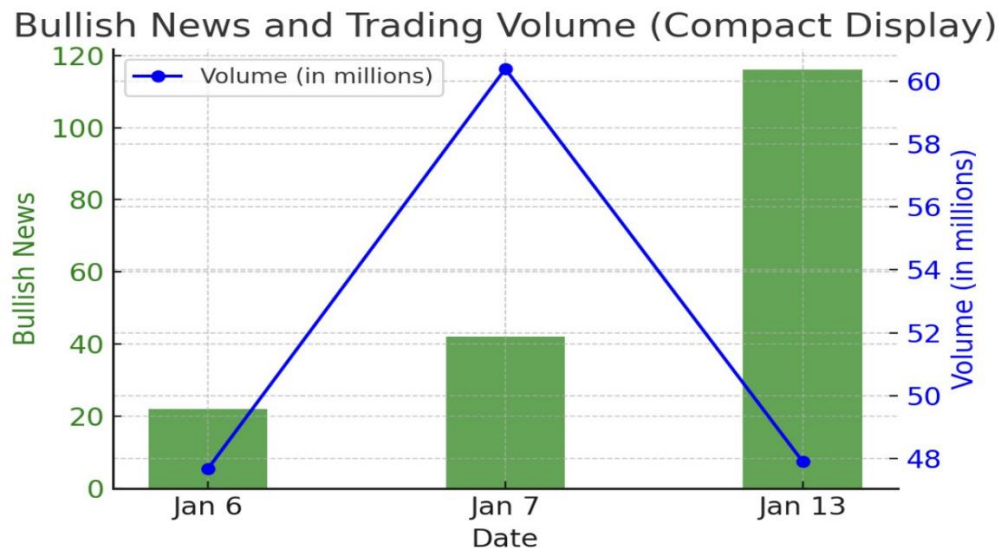
- Close - מחיר הסגירה של המדד בסוף יום המסחר.
- Open - מחיר הפתיחה של המדד בתחילת יום המסחר.
- daily_change - השינוי המחושב בין מחיר הפתיחה למחיר הסגירה, מנורמל כיחס לאחוז.

הסבר מפורט: הגרף מדגים את התנודתיות היומית במדד S&P 500 לאורך זמן. ניתן לראות ימים עם שינויים שליליים חדים (למשל, 7 בינואר 2025) לצד ימים עם שינויים חיוביים משמעותיים (כמו 13 בינואר 2025). התנודות אינן מתואמות תמיד עם ציפיות רגשות חדשות.

השערות מעודכנות:

- **זמן פרסום החדשות:** תנודתיות עשויה להיות תלויה גם בעיתוי הפרסום – למשל, חדשות שפורסמו בשעות הערב עשויות להשפיע על המסחר ביום למחרת.
- **בדיקת הרגשות של החדשות שפורסמו:** אם נתונים רבים יותר נוכל לבדוק האם קיים דפוס לאורך השנים ובהתאם לרגשות של החדשות הכלכליות בהתאמה למחירי הסגירה.

גרף 3: - המחשת הקשר בין נפח המסחר (Volume) לבין החדשות שמסווגות (Bullish) -



הסבר גרף 3-

- הגרף מציג בצורה חזותית האם קיימת התאמה בין כמות החדשות החיוביות ביום מסוים לבין נפח המסחר באותו היום.

עמודות מתוך הנתונים:

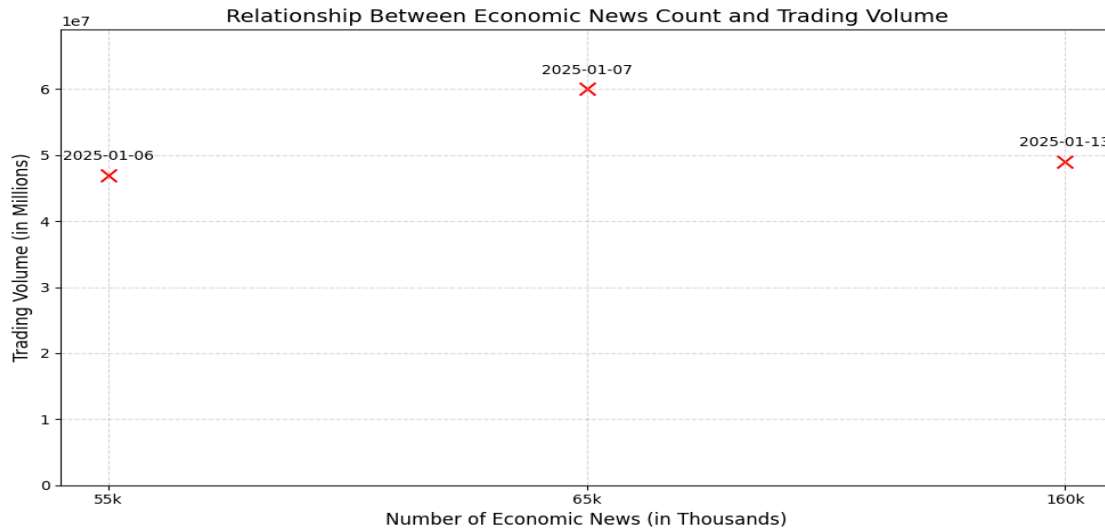
- Date: תאריכים שבהם פורסמו החדשות הכלכליות המסווגות כ-"Bullish" בלבד.
- Volume: נפח המסחר (Volume) של מדד S&P 500 באותם תאריכים.

הסבר מפורט: הגרף מציג בצורה חזותית האם קיימת התאמה בין כמות החדשות החיוביות ביום מסוים לבין נפח המסחר. לפי מה שקיבלנו בהתאם לכמות הנתונים המצומצמת שלנו, אנו יכולים להסיק כי אין קשר או לא יהיה קשר בין נפח המסחר ביום כלשהו וכמות החדשות החיוביות באותו היום. נוכל להוסיף כי תחילה חשבנו שדווקא ביום שיש בו הרבה חדשות חיוביות בהתאם גם נפח המסחר יהיה הרבה יותר גבוהה.

השערות להמשך ניתוח עם מאגר נתונים רחב יותר: כדי לברר אם הבדיקה נכונה, ניתן לבצע חקר נוסף על תקופה ארוכה יותר. השערות נוספות להמשך:

1. **בדיקת איכות החדשות ולא רק הכמות**- אפשר שחדשות עם פעילות כללית יותר על נפח המסחר מאשר חדשות חיוביות.
2. **בדיקת חדשות שליליות** - (Bearish News) **אולי דווקא חדשות שליליות גורמות ליותר נפח מסחר בגלל בהלה בקרב המשקיעים?**
3. **בדיקת זמני יום המסחר**- האם חדשות שמפורסמות לפני פתיחת המסחר משפיעות יותר מאשר חדשות שמתמות באמצע יום המסחר?

גרף 4: כמות חדשות כלכליות היומיות ונפח מסחר היומי -



הסבר גרף 4-

הגרף מציג את הקשר בין מספר החדשות הכלכליות שהתפרסמו ביום מסוים לבין נפח המסחר היומי בשוק ההון (S&P 500). נראה שקיים קשר מסוים בין נפח המסחר לבין כמות החדשות הכלכליות אך לא ניתן לקבוע מתאם ברור מהגרף לבדו.

- ציר ה-X מציג את מספר החדשות הכלכליות שהתפרסמו ביום נתון (באלפים).
- ציר ה-Y מציג את נפח המסחר הכולל של אותו היום (במיליונים).
- כל נקודה מייצגת יום מסחר ספציפי, כאשר תאריך היום מצוין ליד הנקודה.

סקאלת הנתונים-

1. כמות החדשות הכלכליות - הערכים נעים בטווח של 55,000 - 160,000 כתבות כלכליות ביום.
2. נפח המסחר היומי - טווח הערכים נע בין 45 מיליון ל-60 מיליון יחידות מסחר ביום.

השערות להמשך המחקר על מאגר נתונים גדול יותר-

1. בדיקת תוכן החדשות - האם כתבות עם רגשות שליליים או חיוביים משפיעות יותר על נפח המסחר?
2. בדיקה על תקופות זמן ממושכות - האם יש מגמות קבועות בין כמות החדשות לנפח המסחר לאורך שנים?

3. השפעת חדשות כלכליות גדולות – לבדוק אם יש ימים עם חדשות משמעותיות במיוחד (כגון הכרזות של הבנק המרכזי) שמשפיעות על נפח המסחר.
4. ניתוח נוסף על טווחי מסחר שונים – לבדוק האם חדשות מוקדמות ביום משפיעות יותר על המסחר מאשר חדשות שמתפרסמות בסוף יום המסחר

חידוד מטרות מדעי הנתונים בפרויקט-

1. מודלים מבוססי הקשר:
 - מעבר ממודלים שמתמקדים במשתנה בודד (כמו רגשות) למודלים שמשלבים הקשרים בין משתנים כמו נפח מסחר, קטגוריות High-Low ו-MACD.
 2. זיהוי אירועים חריגים:
 - יצירת מנגנון לזיהוי ימים עם נפח מסחר גבוה ורגשות קיצוניים, שמנבאים אירועים משמעותיים בשוק.
 3. ניתוח רמות ניטרליות:
 - חקירת ההשפעה של רגשות ניטרליים על יציבות או תנודתיות בשוק, ובדיקה האם ניתן להשתמש בהם לניבוי שוק רגוע.
- הגרפים והניתוחים הצביעו על מגמות מעניינות, אך חשוב להדגיש כי הנתונים המנותחים מייצגים מדגם קטן בלבד. הרחבת הנתונים ושימוש במידע נוסף תסייע ליצירת מסקנות מבוססות יותר. הגרפים המחישו את החשיבות של שילוב בין רגשות, נפח מסחר וקטגוריות High-Low, מה שמוביל לשינוי מטרות החקירה לכיוון ניתוחים מורכבים יותר.

4. איכות הנתונים

ערכים חסרים

- ערכים חסרים מופיעים כאשר אין מידע זמין עבור עמודה מסוימת. זה עשוי להופיע כתאים ריקים, ערכים כמו- "null", סימנים כמו- "?" או ערכים לא הגיוניים כמו- "[]".

פתרונות:

1. זיהוי מוקדם של ערכים חסרים :

- בחינת כל עמודה כדי להבין כמה ערכים חסרים קיימים ומהי השפעתם על הנתונים.

2. טיפול בערכים חסרים :

השלמה :

בעמודה "banner_image" בנתוני החדשות הכלכליות מצאנו כי ישנם מספר תאים עם ערך "null". בשלב זה אנו סבורים שלעמודה זו אין השפעה קיצונית על המודל, לכן ככל הנראה בשלב מתקדם נבצע על עמודה זו או קידוד או שנסיר אותה לגמרי.

שגיאות נתונים

- בנתוני המדד S&P-500 ובנתוני החדשות לא נתקלנו בשגיאות נתונים.

שגיאות מדידה

- בנתוני המדד S&P-500 ובנתוני החדשות לא נתקלנו בשגיאות מדידה.

אי עקביות בקידוד

זיהוי הבעיה:

- חוסר עקביות בקידוד עשוי להתרחש כאשר עמודות טקסט מכילות ערכים כתובים בדרכים שונות. לדוגמה- Male, male או- Bullish, Positive.

פתרונות:

אחידות קידוד :

- בעמודה "overall_sentiment_label" מנתוני החדשות הכלכליות יש ערכים מקודדים שונים וארוכים. נרצה להחליף ערכים אלו בערכים מקודדים ופשוטים באמצעות אותיות קטנות לדוגמה- "Bullish" נהפוך ל-"b", ואת "Somewhat Bullish" נהפוך ל-"sb". בכך, פעולת הקידוד תסייע לנו בקריאה מהירה של נתוני הרגשות המובהקים של כל חדשה.

מטא-נתונים גרועים

- שמות עמודות או ערכים אינם תואמים את המשמעות שלהם, או ששם השדה אינו מתאר היטב את המידע שהוא מכיל.

פתרונות:

1. שינוי שמות עמודות :

- כדי שיהיה התאמה טובה ואחידה בין הסוגי נתונים אנו החלטנו להחליף את השם של עמודת " time_published " ל- "Date" כדי שהיא תהיה תואמת לנתוני המדד S&P-500 שנוכל לבקש את אותו סוג עמודה משתי הסוגי נתונים בפעם אחת. בנוסף בהמשך הפרויקט ככל הנראה משנה עוד שמות של עמודת עלפי צורך ותדירות שימוש לשמות יותר קצרים ונוחים לשימוש.