

ניתוח חברות הזנק חד-קרן במהלך השנים

שם קורס: ויזאליזציה של נתונים

מבוא

נושא הויזואליזציה הוא חברות הסטארט-אפ מסוג חד-קרן. השאלה העיקרית בתחום שהויזואליזציה הזו שואפת לענות עליה היא לספק סקירה כללית של נוף הסטארט-אפים של חד-קרן על פני סקטורי תעשייה, מדינות וערים שונות. הדאטאסט (עם 937 שורות ו-10 עמודות) נלקח מקאגל והוא מכיל רשימה של חברות הזנק חד-קרן מוכרות ומאפיינים שלהן, הוא מתוחזק על ידי גורמים רשמיים כמו הוול סטריט גורנל ועוד.

השאלות שעליהן הויזואליזציה רוצה לענות הן:

- 1. מהן חברות ההזנק חד-קרן השונות בסקטורי התעשייה?
- 2. כמה חברות הזנק חד-קרן יש בארה"ב יש ואיך המספר שלהם השתנה עם הזמן?
- 3. לאיזה חברות הזנק חד-קרן יש את הערכת השווי הגבוהה ביותר?
- 4. באיזו עיר יש את חברות הזנק חד-קרן הכי מוערכים בעולם?
- 5. מהם המשקיעים השכיחים ביותר בקרב חברות הזנק חד-קרן?

נתונים

הדאטאסט כולל מידע על חברות הזנק חד-קרן, שהן סטארט-אפים פרטיים עם שווי נוכחי של מיליארד דולר או יותר, במרכזי טכנולוגיה ברחבי העולם. המפתחות עבור מערך נתונים זה אינם מוגדרים במפורש, אך המפתח המתאים ביותר עבור מערך נתונים זה יכול להיות שם החברה, וסוג הנתונים הוא טבלאי.

העמודות לאחר עיבוד מקדים להורדת העמודות של השנה, חודש, ויום בחודש (יש כבר עמודה של תאריך) הן:

עמודה	סוג נתונים	קרדינליות	תיאור	הטופולוגיה של מונצנר
חברה	סטרינג	אחד לאחד	שם חברת ההזנק	שמי, קטגוריאלי
שווי	אינטגר	אחד לאחד	שווי החברה הנוכחי בדולרים	כמותי, רציף
תאריך	תאריך	אחד לאחד	תאריך דיווח השווי	זמני, רציף
מדינה	סטרינג	אחד להרבה	המדינה בה החברה ממוקמת	שמי, קטגוריאלי
עיר	סטרינג	אחד להרבה	העיר בה החברה ממוקמת	שמי, קטגוריאלי
תעשייה	סטרינג	אחד להרבה	סקטור התעשייה אליה החברה שייכת	שמי, קטגוריאלי
משקיעים	סטרינג	אחד להרבה	שמות המשקיעים שהקימו את החברה	שמי, קטגוריאלי

מטלות

משימה	פעולה	מטרה
זהה את המדינות עם הכי הרבה חברות הזנק בכל סקטור	Query; Compare	Attributes; Extremes
זהה את מגמות הקמת חברות ההזנק מכל סקטור במשך השנים	Analyze; Consume; Discover	All Data; Trends
מצא את התפלגות המשקיעים הפעילים ביותר על פני כל סקטור	Analyze; Consume; Discover	Attributes; Distribution
מצא את הסקטור עם השווי פר חברת הזנק הגבוה ביותר	Search; Explore	Attributes; Extremes
השווה בין התפלגויות הערכות השווי בין סקטורים שונים על פני הזמן	Analyze; Consume; Discover	Attributes; Distribution

הסבר על העיצוב שנבחר ויישומו
תיאור היישום במונחים של marks & channels

1. מפת צבע של העולם:

- סימנים: הסימנים בויזואליזציה זו הם המצולעים במפה, המייצגים מדינות שונות.
- ערוצים: הערוצים בויזואליזציה זו כוללים:
- ערוץ גיאוגרפי: מיקום המדינות מיוצג על המפה.
- ערוץ צבע: עוצמת הצבע מייצגת את מספר החברות בכל מדינה.

2. תרשים קווים:

- סימנים: הסימנים בויזואליזציה זו הם נקודות וקווים.
- ערוצים: הערוצים בויזואליזציה זו כוללים:
- ערוץ מיקום: ציר ה-x מייצג זמן (שנה), וציר ה-y מייצג את הספירה המצטברת של חברות.
- ערוץ צבע: הקווים והסימנים מקודדים בצבע בהתבסס על הסקטורים שנבחרו.

3. תרשים עמודות:

- סימנים: הסימנים בויזואליזציה זו הם מלבנים.
- ערוצים: הערוצים בויזואליזציה זו כוללים:
- ערוץ מיקום: ציר ה-x מייצג משקיעים, וציר ה-y מייצג את מספר ההשקעות.
- ערוץ צבע: הקווים מקודדים בצבע בהתאם לסקטורים בהם בוצעו ההשקעות.

4. תרשים כינור:

- סימנים: הסימנים בויזואליזציה זו הם תרשימי כינור.
- ערוצים: הערוצים בויזואליזציה זו כוללים:
- ערוץ מיקום: ציר ה-x מייצג סקטורים בתעשייה, וציר ה-y מייצג את הערכת השווי הלוגריתמית של החברות.

הסבר על איך הוויזואליזציה עומדת בקריטריונים של אפקטיביות ואקספרסיביות

1. מפת צבע של העולם:

- אקספרסיביות -

תפוצה: המפה מייצגת את ההפצה הגיאוגרפית של חברות חד-קרבן לפי מדינה בתוך הסקטור הנבחר. כל מדינה מיוצגת ויזואלית, מה שמאפשר למשתמשים להבין את הנוכחות והריכוז של סטארט-אפי חד-קרבן באזורים שונים.

ספירת חברות: עוצמת הצבע במפה מייצגת את מספר החברות בכל מדינה. גוונים בהירים מצביעים על מספר גבוה יותר של חברות, בעוד שגוונים כהים מצביעים על מספר נמוך יותר. נבחר ציר לוגריתמי לצבעים על מנת לראות את ההבדלים בין מדינות בהן יש קצת חברות בסקטור הכללי, וכדי שהארצות המובילות עם הכי הרבה חברות לא יטשטשו את ההבדלים.

הוויזואליזציה מבטאת רק את העובדות הרלוונטיות לסקטור הנבחר ולמדינה של חברות החד-קרבן. הוא מתמקד במספר החברות ובנוכחותן הגיאוגרפית, ומספק מידע תמציתי ורלוונטי למשתמש.

- אפקטיביות -

ייצוג חזותי: מפת choropleth משתמשת בעוצמת צבע כדי לייצג את ספירת החברות בכל מדינה. הקידוד הוויזואלי הזה מקל על המשתמשים לזהות במהירות מדינות עם ריכוז גבוה של סטארט-אפים חד קרבן (גוונים כהים יותר) ואלה עם ריכוז נמוך יותר (גוונים בהירים יותר). השימוש בצבע מאפשר תפיסה יעילה של הבדלים יחסיים בספירת החברה.

הקשר גיאוגרפי: על ידי מיפוי הנתונים לייצוג גיאוגרפי, הוויזואליזציה מנצלת את המודעות המרחבית של המשתמשים כדי להבין את התפוצה של סטארט-אפים חד-קרבן במדינות שונות. משתמשים יכולים לזהות בקלות מדינות של עניין ולצפות בדפוסים או מגמות.

תוויות ומקרא ברורים: הוויזואליזציה מספקת תוויות ברורות למדינות ולסקטורים, ומאפשרת למשתמשים להבין את ההקשר. מקרא הצבעים מספק אינדיקציה ברורה של טווח ספירת החברה, מה שמקל על פירוש עוצמת הצבע במפה.

- דיוק: המפה מייצגת במדויק את התפוצה של סטארט-אפים חד-קרבן במדינות שונות, ומספקת ייצוג חזותי מדויק של מספר החברות בכל מדינה. אפשר לומר שהרגישות שלנו לערכים גדולים היא נמוכה יותר, מחמת אופי מקרא הצבע הנבחר. עם זאת, אני מאמין שחלופה זו היא עדיין טובה יותר משום שהיא מאפשרת לראות את ההבדל בין מדינות עם מספר חברות קטן - מקרה רווח יותר בסט הנתונים שלי.

- יכולת הפרדה: עוצמת הצבע של כל מדינה על המפה מבדילה למעשה את המספר המשתנה של חברות, ומאפשרת הבחנה בין מדינות עם ריכוזי סטארט-אפ גבוהים או נמוכים יותר.

- פופ-אאוט: מדינות עם מספר גבוה יותר של סטארט-אפים חד-קרבן יכולים לבלוט חזותית בגלל עוצמת הצבע, למשוך תשומת לב ולאפשר תפיסה מהירה.

- קיבוץ: המפה אינה מספקת קיבוץ מפורש של מדינות אלא מציגה כל מדינה כנקודת נתונים אינדיבידואלית.

2. תרשים קווים:- אקספרסיביות -

ספירה מצטברת: תרשים הקווים מציג את הספירה המצטברת של סטארט-אפי חד-קרן לאורך זמן. ציר ה-y מייצג את ספירת החברה, וציר ה-x מייצג את השנים. כל קו מגמה בתרשים מייצגת סקטור נבחר, מראה את הצמיחה של סטארט-אפים באותו סקטור בטווח הזמן שנבחר.

השוואת סקטורים: כאשר נבחרים סקטורים מרובים, מוצגים קווי מגמה עבור כל סקטור, המאפשרים השוואה קלה של מגמות הצמיחה שלהם. כל סקטור מיוצג על ידי קו בצבע שונה, מה שמקל על ההפרדה וההשוואה ביניהם.

היזואליזציה מבטאת רק את העובדות הקשורות לספירה המצטברת של סטארט-אפי חד-קרן וצמיחתם לאורך זמן עבור הסקטורים הנבחרים. היא מתמקדת בהשוואת המגמות בין הסקטורים ומספקת מידע תמציתי ורלוונטי למשתמש.

- אפקטיביות -

היזואליזציה של מגמות: תרשים הקווים מדגים ביעילות את המגמות בספירה המצטברת של סטארט-אפי חד-קרן לאורך זמן. השימוש בקווים מקל על מעקב אחר מסלול הצמיחה של כל סקטור. משתמשים יכולים לראות את העליות והמורדות, לזהות תקופות של צמיחה מהירה או סטגנציה, ולהשוות את המגמות בין הסקטורים השונים.

תוויות ומקרא ברורים: היזואליזציה מספקת תוויות ברורות עבור הסקטורים וטווח הזמן, ומאפשרת למשתמשים להבין את ההקשר. כל שורה מסומנת בשם הסקטור המתאים, מה שמאפשר זיהוי והשוואה קלה.

בחירה אינטראקטיבית: משתמשים יכולים לבחור סקטורים מרובים ולהתאים את טווח הזמן, מה שמאפשר חקר אינטראקטיבי של שילובים ותקופות שונות. אינטראקטיביות זו משפרת את האפקטיביות של היזואליזציה בכך שהיא מאפשרת למשתמשים להתמקד בסקטורים ספציפיים ובמרווחי זמן מעניינים.

היזואליזציה מציגה ביעילות את המידע בצורה מושכת ויזואלית ואינטואיטיבית, ומקלה על חקירה והשוואה של מגמות סטארט-אפ על פני סקטורים ותקופות זמן שונות.

- דיוק: תרשים הקו מציג במדויק את הספירה המצטברת של סטארט-אפי חד-קרן לאורך זמן, ומאפשר ניתוח מדויק של דפוסי הצמיחה.

- יכולת הפרדה: סקטורים שונים מיוצגים על ידי תרשימי קווים ברורות, המאפשרות בידול קל והשוואה של מגמות סטארט-אפ.

- פופ-אאוט: קווים המייצגים צמיחה או ירידה משמעותית בספירת הסטארט-אפים יכולים להבליט ויזואלית, ולמשוך את תשומת הלב למגמות בולטות.

- קיבוץ: תרשימי קו מרובים מאפשרים קיבוץ של מגמות סטארט-אפ לפי סקטור, מה שמקל על ניתוח השוואתי.

3. תרשים עמודות:

- אקספרסיביות -

משקיעים מובילים: תרשים העמודות מציג את המשקיעים המובילים בהתבסס על מספר ההשקעות שהם ביצעו בסקטורים הנבחרים. ציר ה-x מייצג את המשקיעים, וציר ה-y מייצג את מספר ההשקעות שבוצעו. כל עמודה בתרשים מייצגת משקיע.

חלוקת השקעות: הקווים מקודדים בצבעים בהתאם לסקטורים בהם בוצעו ההשקעות. זה מאפשר למשתמשים לזהות את סקטורי העניין ולהבין את התפלגות ההשקעות בין סקטורים שונים.

טווח זמן N-ו העליונים: משתמשים יכולים להתאים את טווח הזמן ולבחור את מספר המשקיעים המובילים שיוצגו, מה שמאפשר התאמה אישית והתמקדות בתקופות ספציפיות ובמשקיעים מובילים.

כמו כן, העמודות ממיינות לפי מספר ההשקעות של כל משקיע על מנת לאפשר הבנה קלה יותר של הסדר והיחס בין המשקיעים. בנוסף, כל סקטור נמצא באותה רמה בעמודה בה הוא נמצא ביחס לאחרות, וזאת על מנת להקל על ההשוואה בין כמות ההשקעות של המשקיעים גם ברמת הסקטור.

- אפקטיביות -

השוואה חזותית: תרשים העמודות מאפשר השוואה ויזואלית קלה של מספר ההשקעות שבוצעו על ידי משקיעים שונים. האורכים המשתנים של העמודות מאפשרים לזהות בקלות את המשקיעים המובילים ולקבוע את גודל ההשקעות שלהם.

קידוד צבע: השימוש בצבע לקידוד הסקטורים מוסיף מימד נוסף של מידע. משתמשים יכולים לזהות במהירות לאילו סקטורים שייכות ההשקעות על ידי הסתכלות על הצבעים המשויכים לכל סרגל.

שמות צירים ותוויות ברורות: הויזואליזציה מספקת תוויות ברורות לציר ה-x (המשקיעים) ולציר ה-y (מספר ההשקעות). זה מאפשר למשתמשים לפרש את התרשים ללא עמימות.

בחירה אינטראקטיבית: משתמשים יכולים להתאים את טווח הזמן, לבחור את מספר המשקיעים המובילים שיוצגו ולבחור את הסקטורים המעניינים אותם. אינטראקטיביות זו משפרת את האפקטיביות של הויזואליזציה בכך שהיא מאפשרת למשתמשים להתמקד בתת-קבוצות ספציפיות של נתונים ולחקור תרחישים שונים.

- דיוק: תרשים העמודות מייצג במדויק את מספר ההשקעות שבוצעו על ידי משקיעים מובילים בסקטורים שונים.

- יכולת הפרדה: כל קו מייצג משקיע ספציפי, המאפשר הבחנה קלה בין המשקיעים להשקעותיהם.

- פופ-אאוט: משקיעים עם מספר גבוה יותר של השקעות בולטים חזותית בגלל הקווים הגבוהים יותר, הם מושכים את תשומת הלב.

- קיבוץ: ההשקעות מקובצות לפי סקטור באמצעות קידוד צבע, המאפשר לזהות את הסקטורים בהם פעילים המשקיעים המובילים.

4. תרשים כינורות:

- אקספרסיביות -

התפלגות שווי לוגריתמית: כל סקטור מיוצג על ידי תרשים כינור, כאשר ציר ה-x מייצג את הסקטור וציר ה-y מייצג את ההערכה הלוגריתמית. תרשים הכינור מספק תובנות לגבי התפלגות הערכות השווי בתוך כל סקטור. רוחב הכינור מעיד על צפיפות הערכות השווי.

השוואת סקטורים: על ידי בחירת סקטורים מעניינים, משתמשים יכולים להשוות את התפלגות הערכת השווי בין הסקטורים הללו. הויזואליזציה מאפשרת לחקור דפוסי הערכה והבדלים בין הסקטורים.

הערכת שווי ממוצעת: הויזואליזציה כוללת גם קו המציג את הערכת השווי הממוצעת עבור כל סקטור. גובה הסרגל מייצג את הערכת השווי הממוצעת.

כמו כן, תרשימי הכינור ממיינים לפי השווי הממוצע על מנת להקל על המשתמשים בהשוואה בין ההתפלגות והממוצע בין הסקטורים השונים.

- אפקטיביות -

סולם לוגריתמי: על ידי שימוש בסולם לוגריתמי להערכות שווי על ציר ה-y, הויזואליזציה יכולה לטפל ביעילות במגוון רחב של הערכות שווי. סולם זה מאפשר ייצוג ברור של התפלגות וצפיפות הערכות השווי, במיוחד כאשר עוסקים בערכים קיצוניים.

תרשים כינור: תרשים הכינור מספק סיכום ויזואלי של התפלגות השווי בתוך כל סקטור. משתמשים יכולים לזהות בקלות את הצפיפות והתפשטות של הערכות שווי, כולל חריגים פוטנציאליים. רוחב הכינור מראה את צפיפות הערכות השווי ברמות שונות, מה שמאפשר השוואה מהירה בין סקטורים.

השוואת שווי ממוצעת: תרשימי הבר המציגים את ההערכת השווי הממוצעת עבור כל סקטור מאפשרת למשתמשים להשוות את הערכות השווי הממוצעות ביניהם. הערכים המספריים על הקווים מספקים מידע מדויק על הערכות השווי הממוצעות, משלימים את הייצוג החזותי.

שמות צירים ותוויות ברורות: הויזואליזציה מספקת תוויות ברורות עבור ציר ה-x (התעשייה) וציר ה-y (הערכת שווי). זה מבטיח שמשתמשים יכולים לפרש את התרשים ללא עמימות.

בחירה אינטראקטיבית: משתמשים יכולים לבחור סקטורים מעניינים, מה שמאפשר להם להתמקד בסקטורים מסוימים או להשוות בין סקטורים רבים. אינטראקטיביות זו משפרת את האפקטיביות של הויזואליזציה על ידי מתן אפשרויות גמישות והתאמה אישית.

- דיוק: תרשים הכינור מראה במדויק את התפלגות הערכות הסטארט-אפ של חד קרן בקנה מידה לוגריתמי עבור כל סקטור.

- יכולת הפרדה: כל סקטור מיוצג על ידי תרשים כינור נפרד, המאפשרת הבחנה קלה של התפלגות הערכות בין סקטורים.

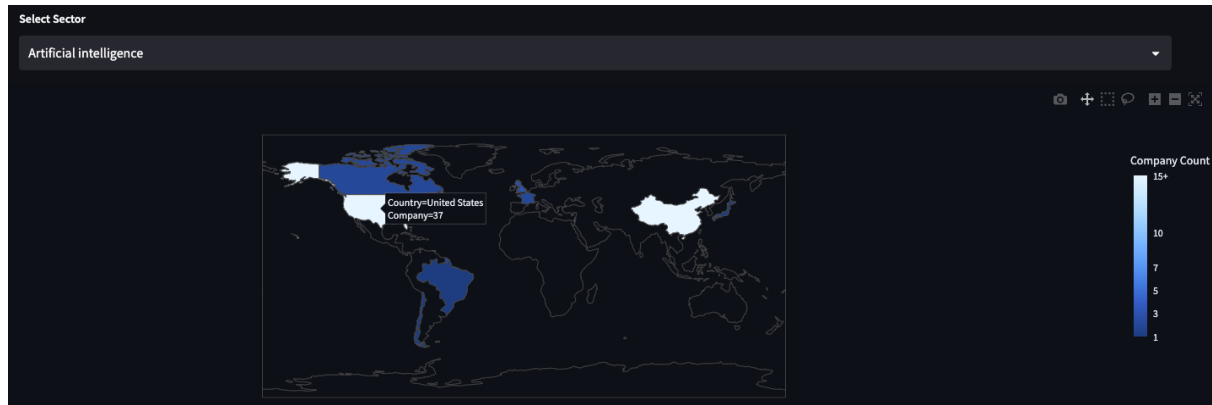
- פופ-אאוט: סקטורים עם התפלגויות הערכת שווי ברורות, כגון אלה עם פריסה רחבה יותר או מצומצמת יותר, יכולים לבלוט ויזואלית, ולסייע בזיהוי וריאציות.

- קיבוץ: הסקטורים מקובצים לפי קידוד ציר ה-x, המספק קיבוץ תפיסתי של התפלגויות הערכת השווי.

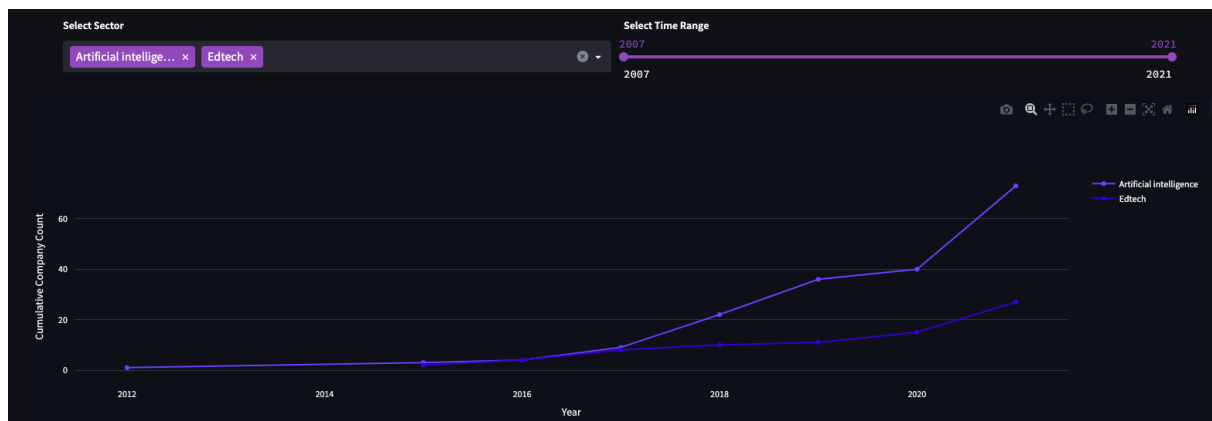
עיבוד מקדים

1. חילוץ המידע על המשקיעים - בכל תא בעמודה של המשקיעים, הייתה מחרוזת שייצגה את המידע על המשקיעים של כל חברה. לכן, היה צריך לחלץ ולעבד את המשקיעים המתאימים עבור כל חברה.
2. שינוי מפתחות שגויים - היו מדינות וסקטורים עם שמות שגויים, שיניתי אותם לשם הנכון.
3. מחיקת עמודות מיותרות - מחקתי עמודות מיותרות כמו שנה, חודש ויום, שכן יש לי עמודת תאריך.

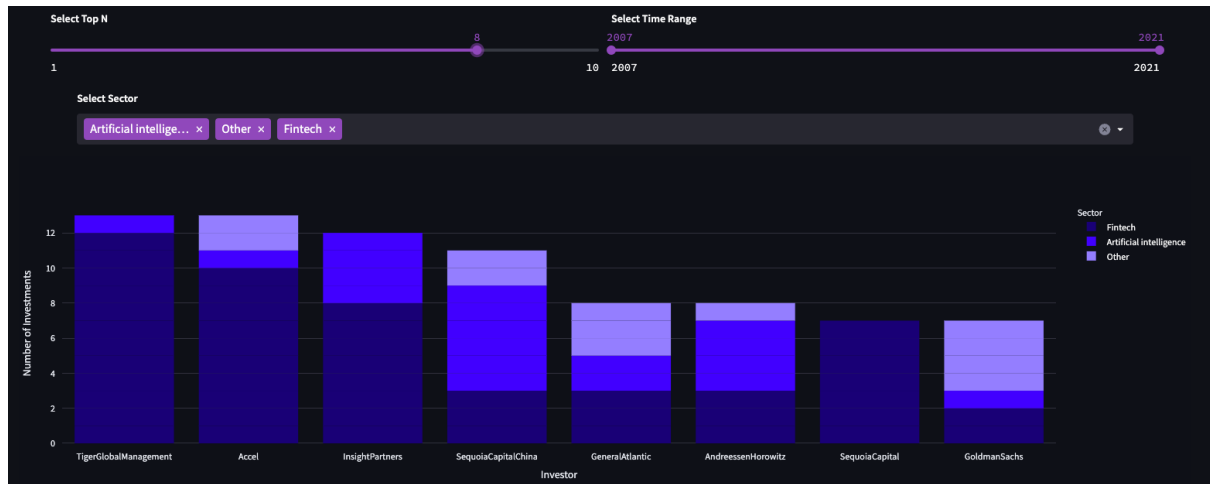
הויזואליזציות



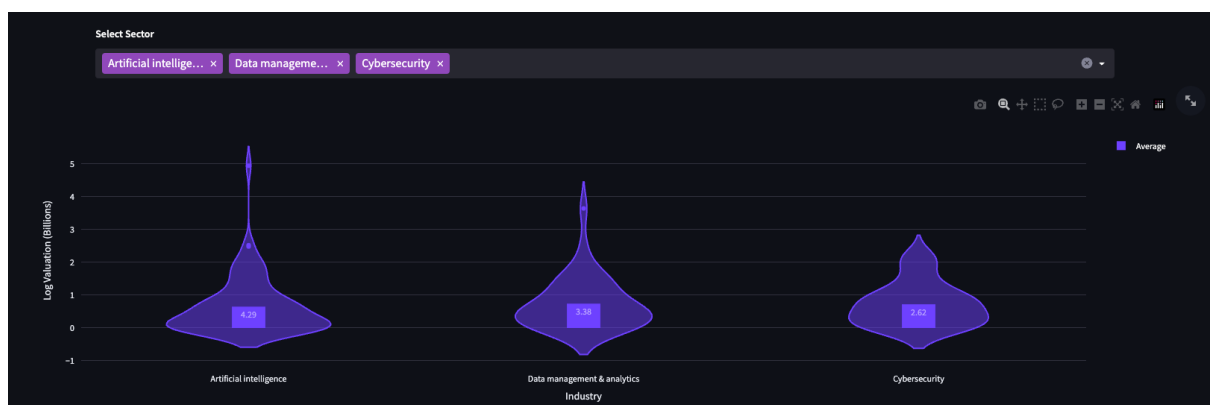
הסבר: צילום המסך מציג מפת צ'ורופל' המייצגת את התפלגות סטארט-אפים חד-קרבן לפי מדינה. כל מדינה מקודדת בצבע בהתאם למספר הסטארט-אפים שיש לה. משתמשים יכולים ליצור אינטראקציה עם המפה על ידי ריחוף מעל מדינה כדי לראות את המספר המדויק של סטארט-אפים ולקבל תובנות לגבי המדינות הנפוצות ביותר עבור כל סקטור.



הסבר: צילום המסך מציג תרשים קווי המתאר את הספירה המצטברת של סטארט-אפי חד-קרבן בסקטור מסויים לאורך זמן. כל קו מייצג סקטור מסוים, ומאפשר למשתמשים להשוות את מגמות הצמיחה של סקטורים שונים. על ידי אינטראקציה עם התרשים, משתמשים יכולים לרחף מעל נקודות נתונים כדי לראות את הכמות המצטברת במרווחי זמן ספציפיים ולקבל תובנות לגבי המגמות של סקטורים בודדים.



הסבר: צילום המסך מציג תרשים עמודות הממחיש את המשקיעים המובילים ואת ההשקעות שלהם בסקטורים שונים. כל פס מייצג משקיע, וגובה הפס מציין את מספר ההשקעות שבוצעו. משתמשים יכולים ליצור אינטראקציה עם התרשים על ידי ריחוף מעל פס כדי להציג את שם המשקיע ומספר ההשקעות. הפסים המקודדים בצבע מאפשרים למשתמשים לזהות את הסקטורים בהם המשקיעים פעילים ביותר.



הסבר: צילום המסך מציג תרשימי כינור המייצגים את התפלגות הערכות הסטארט-אפ של חדר בקנה מידה לוגריתמי עבור סקטורים שונים. כל סקטור מתואר על ידי תרשים כינור נפרד, כאשר רוחב התרשים מעיד על צפיפות הערכות השווי. משתמשים יכולים ליצור אינטראקציה עם התרשים על ידי ריחוף מעל חלקים שונים של הכינורות כדי לצפות בפרטי הערכת שווי ספציפיים. ציר ה-x מציג את סקטורי התעשייה, ומאפשר למשתמשים להשוות את התפלגות הערכת השווי בין הסקטורים.

הסבר על הקוד

בקוד נעשה שימוש בספריות הבאות:

```
import pandas as pd
import streamlit as st
import numpy as np
import plotly.express as px
import plotly.graph_objects as go
```

הקוד יוצר אפליקציית אינטרנט אינטראקטיבית באמצעות Streamlit ומציג תרשימים לניתוח חברות הזנק חד קרן.

הקוד מחולק לחלקים שונים, כשכל אחד אחראי על יצירת תרשים ספציפי ומרכיבי ממשק המשתמש המתאימים לו. ארבעת החלקים הם כדלקמן:

1. מפת Choropleth: מציגה מפה המציגה את התפלגות סטארט-אפים חד-קרן על פני מדינות שונות בסקטור הנבחר. משתמשים יכולים לבחור את הסקטור מתפריט נפתח.
 2. ספירה מצטברת של סטארט-אפים חד-קרן לאורך זמן: מציג את הספירה המצטברת של סטארט-אפים חד-קרן לאורך זמן עבור הסקטורים הנבחרים. זה יכול להציג תרשים קו בודד או תרשימי קו מרובים להשוואה. משתמשים יכולים לבחור את הסקטורים ואת טווח הזמן.
 3. משקיעים מובילים והשקעותיהם לפי סקטור: מדגים את התפלגות ההשקעות שבוצעו על ידי המשקיעים המובילים על פני סקטורים שונים. משתמשים יכולים להתאים פרמטרים כגון מספר המשקיעים המובילים, טווח זמן וסקטורים נבחרים.
 4. התפלגות שווי יומן בין סקטורים: ממחיש את התפלגות הערכות שווי של חברות החד-קרן בקנה מידה לוגריתמי עבור הסקטורים הנבחרים באמצעות תרשימי כינור. הוא כולל גם תרשים עמודות המציג את הערכת השווי הממוצעת עבור כל סקטור. משתמשים יכולים לבחור את הסקטורים המעניינים אותם.
- הקוד משלב מניפולציה של נתונים עם Plotly Express ו-Plotly Graph Objects כדי ליצור לוח מחוונים אינטראקטיבי לניתוח סטארט-אפים של חד קרן.